

ΠΑΝΕΠΙΣΤΗΜΙΟ ΜΑΚΕΔΟΝΙΑΣ
ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
ΤΜΗΜΑΤΟΣ ΕΦΑΡΜΟΣΜΕΝΗΣ ΠΛΗΡΟΦΟΡΙΚΗΣ

ΑΝΑΛΥΣΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ TWEETS ΓΙΑ COVID-19

Διπλωματική Εργασία

ΑΓΓΕΛΟΣ ΠΑΝΑΓΙΩΤΟΥ

Θεσσαλονίκη, Ιούνιος 2022

ΑΝΑΛΥΣΗ ΚΑΙ ΔΙΑΧΕΙΡΙΣΗ TWEETS ΓΙΑ COVID-19

ΑΓΓΕΛΟΣ ΠΑΝΑΓΙΩΤΟΥ

ΜΗΧΑΝΙΚΟΣ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ, ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ

Διπλωματική Εργασία

υποβαλλόμενη για τη μερική εκπλήρωση των απαιτήσεων του

ΜΕΤΑΠΤΥΧΙΑΚΟΥ ΤΙΤΛΟΥ ΣΠΟΥΔΩΝ ΣΤΗΝ ΕΦΑΡΜΟΣΜΕΝΗ ΠΛΗΡΟΦΟΡΙΚΗ

Επιβλέπων/ουσα Καθηγητής/τρια
Ονοματεπώνυμο Καθηγητή/τριας

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την ηη/μμ/εεεε

Ονοματεπώνυμο 1

Ονοματεπώνυμο 2

Ονοματεπώνυμο 3

.....

.....

.....

ΠΑΝΑΓΙΩΤΟΥ ΑΓΓΕΛΟΣ

.....

Περίληψη

Ο ιός Covid-19 εμφανίστηκε τέλη του 2019 και μέσα σε ένα πολύ σύντομο χρονικό διάστημα εξελίχθηκε σε πανδημία. Οι επιπτώσεις της πανδημίας σε διάφορους τομείς ήταν και παραμένουν ακόμα τεράστιες. Η παρούσα διπλωματική εργασία μελετάει δεδομένα που προέρχονται από το Μέσο Κοινωνικής Δικτύωσης Twitter και αφορούν τον ιό Covid-19. Σκοπός της διπλωματικής εργασίας είναι η κατανόηση του τρόπου διαχείρισης της πανδημίας στο Ηνωμένο Βασίλειο, κατά τη διάρκεια της πρώτης φάσης που ξέσπασε ο ιός, από τη 1 Φεβρουαρίου μέχρι τη 1 Μαΐου. Τα tweets είναι γραμμένα στα αγγλικά και αφορούν μόνο το Ηνωμένο Βασίλειο. Σε πρώτη φάση έγινε ανίχνευση των θεμάτων που απασχολούσαν τους Βρετανούς πολίτες τη συγκεκριμένη χρονική περίοδο. Για την επίτευξη αυτού, εφαρμόστηκαν δύο αλγόριθμοι μοντελοποίησης θεμάτων. Αρχικά υλοποιήθηκε ο γνωστός αλγόριθμος LDA. Στη συνέχεια αφού εντοπίστηκαν οι πιο ενδιαφέρουσες θεματικές ενότητες, εφαρμόστηκε ένας άλλος αλγόριθμος μοντελοποίησης θεμάτων, ο GSDMM, για την καλύτερη κατανόηση και αποσαφήνιση των θεμάτων που προέκυψαν. Έπειτα μετά τον καθορισμό των κυριότερων θεμάτων, έγινε χρονική ανάλυση των δεδομένων. Με τη βοήθεια καταγραφής των κυριότερων ειδήσεων, πραγματοποιήθηκε η συσχέτιση τους και μια προσπάθεια κατανόησης της καθημερινότητας που βίωσαν οι Βρετανοί πολίτες.

Λέξεις Κλειδιά: κορονοϊός, Twitter, tweets, Μεγάλη Βρετανία, ανάλυση, μοντελοποίηση θεμάτων, LDA, GSDMM, χρονική ανάλυση

Abstract

The Covid-19 virus appeared in late 2019 and within a very short time, developed into a pandemic crisis. The impact in various fields have been enormous. This dissertation studies data, related to the Covid-19 virus, from the Social Media Platform, Twitter. The aim of this dissertation is to understand how the pandemic in the United Kingdom was managed, during the first phase, from February 1st to May 1st. The tweets are in english and are concerning the United Kingdom only. The issues that concerned British people during the first phase of the pandemic outbreak, were detected. To achieve this topic modeling techniques were used. The well-known LDA algorithm was implemented first. After the identification of the most interesting topics, another topic modeling algorithm was implemented, in order to understand and clarify better, the issues tha arose. After identifying the main issues, a time analysis of the data was performed. With the help of recording the main news and these results, there was an attempt to understand what happened in United Kingdom and how they dealt with the pandemic.

Keywords: covid-19, coronavirus, Twitter, tweets, Uk, topic modelling, LDA, GSDMM, time analysis

Ευχαριστίες

Θα ήθελα να ευχαριστήσω την επιβλέπουσα καθηγήτρια μου κ. Κολωνιάρη Γεωργία για την καθοδήγηση της, τις συμβουλές και την βοήθεια της κατά τη διάρκεια εκπόνησης της διπλωματικής εργασίας. Επιπλέον ένα μεγάλο ευχαριστώ στην οικογένεια μου και σε όλους τους ανθρώπους που ήταν δίπλα μου όλο αυτό το χρονικό διάστημα.

Περίληψη	4
1. Εισαγωγή	2
1.1 Πρόβλημα – Σημαντικότητα του θέματος.....	2
1.2 Σκοπός – Στόχοι.....	2
1.3 Διάρθρωση της μελέτης	2
2. Θεωρητικό Υπόβαθρο	3
2.1 Covid-19	3
2.1.1 Προέλευση	3
2.1.2 Συμπτώματα	4
2.1.3 Δημόσια υγεία.....	5
2.1.4 Πολιτικές αποφάσεις και Covid-19	8
2.1.5 Επιπτώσεις του Covid-19.....	9
2.2 Μέσα Κοινωνικής Δικτύωσης	12
2.2.1 Twitter	13
2.3 Τεχνητή Νοημοσύνη	15
2.3.1 Εξόρυξη Δεδομένων.....	15
2.3.2 Μηχανική Μάθηση.....	17
2.3.3 Μηχανική Μάθηση και Τεχνητή Νοημοσύνη.....	17
2.3.4 Μηχανική Μάθηση και Στατιστική	17
2.3.5 Υπολογιστικό Μοντέλο	18
2.3.6 Κατηγορίες Μηχανικής Μάθησης.....	18
2.4 Εφαρμογή Τεχνικών Ανάλυσης Κειμένου.....	19
2.4.1 Αναπαράσταση Δεδομένων κειμένου.....	20
2.4.2 Μοντελοποίηση θεμάτων.....	21
2.4.3 Latent Dirichlet Allocation (LDA)	22
2.4.4 GSDMM	23
2.4.5 Αξιολόγηση.....	23
2.4.6 Εργαλεία Οπτικοποίησης και Ανάλυσης Αποτελεσμάτων LDA.....	24
2.5 Ανάλυση Δεδομένων στο Twitter	27
2.5.1 Ανάλυση Δεδομένων από το Twitter για τον Covid-19.....	27
2.5.2 Το σύνολο δεδομένων.....	30
3. Μεθοδολογία.....	31
3.1 Συλλογή Δεδομένων	31
3.2 Προ-επεξεργασία των δεδομένων.....	33
3.3 Προσδιορισμός των θεμάτων	34
3.4 Αρχικός διαχωρισμός θεμάτων με LDA	34
3.5 Οπτικοποίηση των αποτελεσμάτων	35
3.6 Εξαγωγή των τελικών θεμάτων με GSDMM	36
3.7 Εξαγωγή διαγραμμάτων χρόνου	36
4. Αποτελέσματα.....	37
4.1 Αποτελέσματα και παρατηρήσεις LDA	37

4.2 Αποτελέσματα GSDMM.....	53
4.3 Χρονική ανάλυση	68
5. Σύνοψη και συμπεράσματα	98
5.1 Όρια και περιορισμοί της έρευνας	99
5.2 Μελλοντικές Επεκτάσεις.....	99

Εικόνα 1 : Παράδειγμα οπτικοποίησης μοντέλου LDA, με pyLDAvis	25
Εικόνα 2 : Παράδειγμα οπτικής απεικόνισης θεμάτων με pyLDAvis	26
Εικόνα 3 : Γράφημα αξιολόγησης των θεμάτων LDA, με βάση το coherence score	38
Εικόνα 4 : Οπτική απεικόνιση των αποτελεσμάτων του LDA.....	39
Εικόνα 5 : Οπτική απεικόνιση του θέματος 1	40
Εικόνα 6 : Οπτική απεικόνιση του θέματος 2	40
Εικόνα 7 : Οπτική απεικόνιση του θέματος 3	41
Εικόνα 8 : Οπτική απεικόνιση του θέματος 4	41
Εικόνα 9 : Οπτική απεικόνιση του θέματος 5	42
Εικόνα 10 : Οπτική απεικόνιση του θέματος 6	42
Εικόνα 11 : Οπτική απεικόνιση του θέματος 7	43
Εικόνα 12 : Οπτική απεικόνιση του θέματος 8	43
Εικόνα 13 : Οπτική απεικόνιση του θέματος 9	44
Εικόνα 14 : Οπτική απεικόνιση του θέματος 10	44
Εικόνα 15 : Οπτική απεικόνιση του θέματος 11	45
Εικόνα 16 : Οπτική απεικόνιση του θέματος 12	45
Εικόνα 17 : Οπτική απεικόνιση του θέματος 13	46
Εικόνα 18 : Οπτική απεικόνιση του θέματος 14	46
Εικόνα 19 : Οπτική απεικόνιση του θέματος 15	47
Εικόνα 20 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 3 του θέματος 0 του LDA	57
Εικόνα 21 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 6 του θέματος 0 του LDA	58
Εικόνα 22 : : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 2 του θέματος 2 του LDA	58
Εικόνα 23 : : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 3 του θέματος 2 του LDA	59
Εικόνα 24 : : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 4 του θέματος 2 του LDA	59
Εικόνα 25 : : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 0 του θέματος 4 του LDA	60
Εικόνα 26 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 2 του θέματος 4 του LDA	60
Εικόνα 27 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 0 του θέματος 6 του LDA	61
Εικόνα 28 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 2 του θέματος 6 του LDA	61
Εικόνα 29 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 3 του θέματος 6 του LDA	62
Εικόνα 30 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 1 του θέματος 8 του LDA	63
Εικόνα 31 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 2 του θέματος 8 του LDA	63
Εικόνα 32 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 7 του θέματος 8 του LDA	64

Εικόνα 33 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 3 του θέματος 10 του LDA	64
Εικόνα 34 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 4 του θέματος 10 του LDA	65
Εικόνα 35 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 2 του θέματος 12 του LDA	65
Εικόνα 36 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 2 του θέματος 13 του LDA	66
Εικόνα 37 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 4 του θέματος 13 του LDA	66
Εικόνα 38 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 0 του θέματος 14 του LDA	67
Εικόνα 39 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 1 του θέματος 14 του LDA	67
Εικόνα 40 : Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 4 του θέματος 14 του LDA	68
Εικόνα 41 : Χρονική ανάλυση του υποθέματος 3 του topic 0 του LDA	69
Εικόνα 42 : Χρονική ανάλυση του υποθέματος 6 του topic 0 του LDA	70
Εικόνα 43 : Χρονική ανάλυση του υποθέματος 2 του topic 2 του LDA	72
Εικόνα 44 : Χρονική ανάλυση του υποθέματος 3 του topic 2 του LDA	73
Εικόνα 45 : Χρονική ανάλυση του υποθέματος 4 του topic 2 του LDA	75
Εικόνα 46 : Χρονική ανάλυση του υποθέματος 0 του topic 4 του LDA	76
Εικόνα 47 : Χρονική ανάλυση του υποθέματος 2 του topic 4 του LDA	77
Εικόνα 48 : Χρονική ανάλυση του υποθέματος 0 του topic 6 του LDA	79
Εικόνα 49 : Χρονική ανάλυση του υποθέματος 2 του topic 6 του LDA	80
Εικόνα 50 : Χρονική ανάλυση του υποθέματος 3 του topic 6 του LDA	81
Εικόνα 51 : Χρονική ανάλυση του υποθέματος 1 του topic 8 του LDA	83
Εικόνα 52 : Χρονική ανάλυση του υποθέματος 2 του topic 8 του LDA	84
Εικόνα 53 : Χρονική ανάλυση του υποθέματος 7 του topic 8 του LDA	86
Εικόνα 54 : Χρονική ανάλυση του υποθέματος 3 του topic 10 του LDA	88
Εικόνα 55 : Χρονική ανάλυση του υποθέματος 4 του topic 10 του LDA	89
Εικόνα 56 : Χρονική ανάλυση του υποθέματος 2 του topic 12 του LDA	90
Εικόνα 57 : Χρονική ανάλυση του υποθέματος 2 του topic 13 του LDA	91
Εικόνα 58 : Χρονική ανάλυση του υποθέματος 4 του topic 13 του LDA	92
Εικόνα 59 : Χρονική ανάλυση του υποθέματος 0 του topic 14 του LDA	93
Εικόνα 60 : Χρονική ανάλυση του υποθέματος 1 του topic 14 του LDA	95
Εικόνα 61 : Χρονική ανάλυση του υποθέματος 4 του topic 14 του LDA	96

Πίνακας 1 : Τεχνικές προ-επεξεργασίας που εφαμόστηκαν σε αντίστοιχες εργασίες	29
Πίνακας 2 : Αποτελέσματα του coherence score για τα υποθέματα του topic 0	53
Πίνακας 3 : Αποτελέσματα του coherence score για τα υποθέματα του topic 2	54
Πίνακας 4 : Αποτελέσματα του coherence score για τα υποθέματα του topic 4	54
Πίνακας 5 : Αποτελέσματα του coherence score για τα υποθέματα του topic 6	54
Πίνακας 6 : Αποτελέσματα του coherence score για τα υποθέματα του topic 8	55
Πίνακας 7 : Αποτελέσματα του coherence score για τα υποθέματα του topic 10	55
Πίνακας 8 : Αποτελέσματα του coherence score για τα υποθέματα του topic 12	55
Πίνακας 9 : Αποτελέσματα του coherence score για τα υποθέματα του topic 13	56
Πίνακας 10 : Αποτελέσματα του coherence score για τα υποθέματα του topic 14	56

1. Εισαγωγή

1.1 Πρόβλημα – Σημαντικότητα του θέματος

Τα τελευταία δύο χρόνια έχουν αλλάξει πάρα πολλά πράγματα στη καθημερινή μας ζωή. Η πανδημία έδειξε πόσο εύθραυστα είναι πολλά από αυτά που θεωρούσαμε δεδομένα. Το αίσθημα της αγωνίας και του φόβου ήταν μέχρι πριν λίγο καιρό κυρίαρχο. Παγκόσμια οι κυβερνήσεις κλήθηκαν να λάβουν μέτρα περιορισμού της πανδημίας. Κάποιες πιο επιτυχημένα, αντιδρώντας πιο γρήγορα, κατάφεραν να πετύχουν σε μεγάλο βαθμό την επιστροφή στη κανονικότητα. Κάποιες άλλες χώρες διαχειρίστηκαν τη πανδημία λιγότερο αποτελεσματικά. Καθίσταται απαραίτητη, λοιπόν η έρευνα, η ανάλυση και η αξιολόγηση του τρόπου διαχείρισης της πανδημίας.

1.2 Σκοπός – Στόχοι

Κατά τη χρονική περίοδο 1 Φεβρουαρίου μέχρι τη 1 Μαΐου το Ηνωμένο Βασίλειο ήταν μια από τις χώρες, που δέχτηκαν έντονη κριτική για τον χρόνο που έλαβαν περιοριστικά μέτρα περιορισμού της πανδημίας. Στόχος της εργασίας η ανάλυση δεδομένων που προέρχονται από το Twitter, και αφορούν το Ηνωμένο Βασίλειο τη συγκεκριμένη χρονική περίοδο, ώστε να εξαχθούν χρήσιμα συμπεράσματα για τον τρόπο διαχείρισης της πανδημίας στη συγκεκριμένη χώρα. Συγκεκριμένα να απαντηθούν ερωτήματα που σχετίζονται με το τι απασχόλησε περισσότερο τους Βρετανούς πολίτες, τι περιοριστικά μέτρα πάρθηκαν, αν το σύστημα υγείας ενισχύθηκε όπως έπρεπε και αν η πολιτική ηγεσία άργησε στο να λάβει αποφασιστικά μέτρα.

1.3 Διάρθρωση της μελέτης

Στο πρώτο κεφάλαιο της εργασίας παρουσιάζεται η σημασία διεξαγωγής της έρευνας και ο στόχος της. Στο δεύτερο κεφάλαιο αναλύεται διεξοδικά το θεωρητικό υπόβαθρο της εργασίας, δηλαδή ο covid-19, τα Μέσα Κοινωνικής Δικτύωσης και συγκεκριμένα το Twitter, οι βασικές έννοιες της μηχανικής μάθησης και οι τεχνικές ανάλυσης των δεδομένων που χρησιμοποιήθηκαν, δηλαδή τεχνικές μοντελοποίησης θεμάτων. Επιπλέον στο δεύτερο κεφάλαιο παρουσιάζονται κάποιες προηγούμενες έρευνες σχετικές με το αντικείμενο της διπλωματικής. Στο τρίτο κεφάλαιο παρουσιάζονται τα στάδια υλοποίησης της ανάλυσης των δεδομένων. Στο τέταρτο κεφάλαιο παρουσιάζονται τα αποτελέσματα των τεχνικών ανάλυσης δεδομένων σε συνδυασμό με την καταγραφή των κυριότερων ειδήσεων που έλαβαν χώρα, το χρονικό διάστημα κατά το οποίο μελετά η εργασία. Στο πέμπτο κεφάλαιο παρουσιάζονται τα συμπεράσματα της εργασίας, τα προβλήματα που εμφανίστηκαν καθώς και οι πιθανές μελλοντικές προεκτάσεις της.

2. Θεωρητικό Υπόβαθρο

2.1 Covid-19

Τον Δεκέμβριο του 2019 εμφανίστηκαν στην πόλη Wuhan της Κίνας αρκετά περίεργα περιστατικά πνευμονίας. Στις 11 Φεβρουαρίου ο Παγκόσμιος Οργανισμός Υγείας ονόμασε Covid-19 μια νέα ασθένεια προερχόμενη από την οικογένεια ιών του SARS-CoV 2. Για τον συγκεκριμένο διαχωρισμό του Covid-19 από την υπόλοιπη ομάδα των coronavirus έπαιξαν ρόλο κάποιες πολύ συγκεκριμένες ιδιαιτερότητες που παρατηρήθηκαν στους ασθενείς που νόσησαν. Τα ιδιάζουσα αυτά χαρακτηριστικά ήταν η πιο εύκολη μετάδοση του ιού από άνθρωπο σε άνθρωπο, οι ενδείξεις ότι κάποιος μπορούσε να μεταδώσει τον ιό πριν ή και χωρίς να έχει εμφανίσει καθόλου συμπτώματα και το τελευταίο είναι ότι ο χρόνος εμφάνισης συμπτωμάτων ήταν πάνω από 5 μέρες. Ως αποτέλεσμα η μετάδοση του ιού ήταν ραγδαία. Μέχρι το τέλος του ίδιου μήνα είχαν καταγραφεί 83.652 περιστατικά σε παγκόσμιο επίπεδο. Στη συνέχεια, λόγω της τρομακτικής μετάδοσης του ιού ο Παγκόσμιος Οργανισμός Υγείας αναγκάστηκε στις 11 Μαρτίου να την χαρακτηρίσει πανδημία (Carvalho et al., 2021).

Η Ιταλία ήταν η πρώτη ευρωπαϊκή χώρα που χτυπήθηκε από τον ιό και ο αντίκτυπος που είχε ήταν τεράστιος. Ο ιός εξαπλώθηκε υπερβολικά γρήγορα με αποτέλεσμα η Ιταλία να μετράει 30.000 ασθενείς και 2.500 θανάτους έως και τις 18 Μαρτίου του 2020. Αυτό είχε ως αποτέλεσμα η ιταλική κυβέρνηση να προβεί σε έκτακτα μέτρα όπως το καθολικό lockdown, με σκοπό να περιορίσει τη διασπορά όσο το δυνατόν πιο αποτελεσματικά. Αργότερα ακολούθησαν και άλλες ευρωπαϊκές χώρες όπως η Ισπανία, η Αγγλία και ανάμεσα σε όλες τις χώρες που άρχισαν να παίρνουν έκτακτα μέτρα διαχείρισης της πανδημίας και η Ελλάδα (Spinelli and Pellino, 2020).

2.1.1 Προέλευση

Από την εμφάνιση του ιού Covid-19 ή αλλιώς SARS-CoV 2 το μεγαλύτερο ίσως θέμα συζήτησης στην επιστημονική αλλά και την πολιτική κοινότητα αποτελεί η προέλευση του. Οι φήμες για την προέλευση του ιού ήταν εξαρχής πολλές, αόριστες, τρομακτικές και οι θεωρίες διάφορες.

Μια θεωρία προέλευσης του ιού είναι ότι πρόκειται για αποτέλεσμα εργαστηριακών πειραμάτων που κάπως ξέφυγε από τον έλεγχο και έτσι ξεκίνησε η διασπορά του. Ωστόσο μέχρι στιγμής κάτι τέτοιο δεν έχει αποδειχθεί με βάση τα διάφορα γενετικά δεδομένα που έχουν μελετηθεί και ως εκ τούτου δεν φαντάζει αρκετά πιθανό σενάριο. Επιπλέον η πρόσφατη έκθεση του Παγκόσμιου Οργανισμού Υγείας (ΠΟΥ) αναφέρει ότι η εργαστηριακή προέλευση της πανδημίας θεωρείται ιδιαίτερα απίθανη (WHO Headquarters (HQ)). Σχετικές γονιδιακές αναλύσεις και συγκρίσεις με προηγούμενα γνωστά γονιδιώματα κορονοϊού, υποδεικνύουν ότι παρουσιάζει χαρακτηριστικά που τον διακρίνουν από άλλα είδη. Ωστόσο βρέθηκε μια αρκετά μεγάλη ομοιότητα με το είδος κορονοϊού που προσβάλλει της νυχτερίδες και ως εκ τούτου οι νυχτερίδες είναι αρκετά πιθανό να είναι ο ξενιστής του ιού (Ciotti et al., 2020). Το σενάριο που εξέτασε ο ΠΟΥ στην πρόσφατη έκθεση του είναι η μετάδοση του ιού από μια ομάδα ζώων σε ένα άλλο ενδιάμεσο ζώο-ξενιστή και στη συνέχεια μετάδοση του ιού

στον άνθρωπο. Το συμπέρασμα ήταν ότι κάτι τέτοιο θεωρείται πιθανό έως πολύ πιθανό. Τέλος το τρίτο σενάριο είναι η μετάδοση μέσω της τροφικής αλυσίδας ή μέσω μόλυνσης της τροφής ή των δοχείων τροφίμων. Αυτό περιλαμβάνει τη πιθανότητα μετάδοσης μέσω των παγωμένων τροφών που πωλούνται σε σουπερ μαρκετ. Η συγκεκριμένη εκδοχή θεωρείται επίσης πιθανή. (WHO Headquarters (HQ))

2.1.2 Συμπτώματα

Οι ασθενείς που προσβάλλονται από τον ιο Covid-19 μπορεί να παρουσιάσουν από ήπια μέχρι πολύ σοβαρά συμπτώματα, ενώ είναι μεγάλος ο αριθμός των ατόμων που είναι ασυμπτωματικοί. Τα άτομα που είναι ασυμπτωματικά μπορούν να μεταδώσουν τον ιό με μεγαλύτερη ευκολία, καθώς δεν γνωρίζουν ότι νοσούν, άρα να έρθουν σε επαφή με άλλα άτομα και να τον μεταδώσουν στον περίγυρο τους.

Τα πιο συνηθισμένα συμπτώματα που έχουν συλλεχθεί από διάφορες μελέτες και ιατρικά δεδομένα, δείχνουν ότι είναι ο πυρετός, ο βήχας και η έλλειψη οξυγόνου. Επιπλέον αυτό που έχει διαπιστωθεί είναι ότι σε μερικές περιπτώσεις εμφανίζονται γαστρεντερικά συμπτώματα όπως είναι ο εμετός, η διάρροια και το κοιλιακό άλγος. Σ' αυτές τις περιπτώσεις έχει παρατηρηθεί ότι εμφανίζονται πριν εμφανιστεί ο πυρετός, ο βήχας και η έλλειψη οξυγόνου. Συνεπώς οι μελέτες έχουν διαπιστώσει ότι παρόλο που η λοίμωξη από τον ιο στοχεύει πρωτίστως το αναπνευστικό σύστημα και πιο συγκεκριμένα τους πνεύμονες, δεν αφήνει ανεπηρέαστα τα υπόλοιπα όργανα καθώς συχνά εμφανίζεται να επηρεάζονται προκαλώντας καρδιολογικά, γαστρεντερικά προβλήματα και σε κάποιες περιπτώσεις προβλήματα στο κεντρικό νευρικό σύστημα (Ciotti et al., 2020). Το ενδιαφέρον σχετικά με τα προβλήματα που εμφανίζονται στο κεντρικό νευρικό σύστημα είναι ότι τα νευρολογικά συμπτώματα που θα αναφερθούν παρακάτω, προηγούνται των τυπικών αναπνευστικών αρκετές μέρες και σε κάποιες περιπτώσεις ασυμπτωματικών φορέων, είναι ίσως οι μόνες ενδείξεις.

Τα πιο συχνά νευρολογικά συμπτώματα είναι οι διαταραχές στην όσφρηση και στη γεύση, τα οποία φαίνεται να εμφανίζονται στις αρχές της νόσησης και αποτελούν μια σημαντική ένδειξη ότι το άτομο είναι θετικό στον Covid-19. Η μυαλγία είναι ένα ακόμα νευρολογικό σύμπτωμα που σύμφωνα με τις έρευνες είναι αρκετά συνηθισμένο αν και σύμφωνα με μελέτες δεν αποτελεί σοβαρό διαγνωστικό δείκτη ούτε είναι προγνωστικός παράγοντας. Επίσης ο πονοκέφαλος σύμφωνα με έρευνες είναι το πιο κοινό μη ειδικό νευρολογικό σύμπτωμα. Η συχνότητα του ποικίλει ανάλογα με τις έρευνες αν και θεωρείται ως το πέμπτο πιο συχνό σύμπτωμα που εμφανίζουν τα άτομα που έχουν προσβληθεί από τον ιο, ενώ σε κάποιες μελέτες έχει αναφερθεί ότι είναι χρονικά το πρώτο σύμπτωμα που εμφανίζεται (Harapan and Yoo, 2021). Μια μελέτη περίπτωσης από την Αγγλία δείχνει ότι η οξεία σύγχυση, το παραλήρημα είναι μια πρωταρχική εκδήλωση και ίσως η μόνη που έχει εκδηλώσει ένα άτομο που έχει προσβληθεί από τον ιο, χωρίς να έχει εμφανίσει κάποιο αναπνευστικό πρόβλημα. Παράλληλα και άλλοι μελετητές έχουν συσχετίσει την οξεία σύγχυση και το παραλήρημα ως μια επιπλοκή του ιου (Butt et al., 2020). Η πρόσφατη βιβλιογραφία δείχνει υψηλότερα ποσοστά θνησιμότητας στα άτομα που νόσησαν από τον ιο και είχαν χρόνιες νευρολογικές διαταραχές. Τέλος είναι αξιοσημείωτο ότι σύμφωνα με τις μελέτες οι ασθενείς με σοβαρή λοίμωξη Covid-19 είναι πιο επιρρεπείς να αναπτύξουν

νευρολογικές επιπλοκές σε σχέση με τους ασθενείς που νόσησαν με πιο ήπια συμπτώματα (Harapan and Yoo, 2021).

2.1.3 Δημόσια υγεία

Σύμφωνα με τη βιβλιογραφική έρευνα κατά την αξιολόγηση της πιθανής επίδρασης μιας μολυσματικής ασθένειας, λαμβάνονται υπόψη δύο παράμετροι. Αυτές είναι η πιθανότητα μετάδοσης της νόσου και η ικανότητα της να σκοτώσει. Συνεπώς η μεταδοτικότητα και το ποσοστό θνησιμότητας είναι οι δύο πιο κρίσιμοι παράγοντες που καθορίζουν τον αντίκτυπο της πανδημίας στη δημόσια υγεία. Ο Covid-19 έχει υψηλό ρυθμό μετάδοσης και σύμφωνα με τους επιστήμονες, το ποσοστό θνησιμότητας είναι πολύ υψηλότερο από την εποχική επιδημία γρίπης, η οποία ήδη από μόνη της είναι αρκετά θανατηφόρα καθώς μεγάλος αριθμός ασθενών κάθε χρόνο απεβιώνει από την εποχική γρίπη. Ως εκ τούτου ο Covid-19 ιός έχει χαρακτηριστεί ως σημαντική απειλή για τη δημόσια υγεία (Binns et al., 2020).

Για τον συγκεκριμένο λόγο μέτρα υγειονομικής προστασίας ξεκίνησαν να λαμβάνονται από σχεδόν όλες τις χώρες παγκοσμίως. Οι κυβερνήσεις, άλλοτε με καθυστέρηση και άλλοτε πιο έγκαιρα, σιγά-σιγά ξεκίνησαν να στρέφονται στην επιστημονική κοινότητα και τον ΠΟΥ προκειμένου να λάβουν τα απαραίτητα υγειονομικά μέτρα με στόχο την μείωση μετάδοσης του ιού και συνεπώς τον περιορισμό των θανάτων ασθενών λόγω Covid-19. Η συγκεκριμένη εργασία θα μελετήσει τον αντίκτυπο που είχαν οι συγκεκριμένες αποφάσεις, μελετώντας την κοινωνικοπολιτική και οικονομική διάσταση του ζητήματος. Ωστόσο στο συγκεκριμένο σημείο καλό θα ήταν να αναφερθούν ποια είναι τα μέτρα προστασίας μετάδοσης του ιού.

- Καθολική χρήση μάσκας: Η συνεχής και σωστή χρήση μάσκας προσώπου είναι ένα κρίσιμο μέτρο υγειονομικής προστασίας και μετάδοσης του ιού. Ιδιαίτερα μετά από τις επιστημονικές έρευνες που ήρθαν στην επιφάνεια και μελετούσαν την μετάδοση του ιού από άτομα που δεν είχαν παρουσιάσει κάποιο σύμπτωμα αλλά ήταν φορείς, όπου σύμφωνα με τις εκτιμήσεις αυτές τουλάχιστον ο μισός αριθμός των μολύνσεων προέρχεται από αυτούς. Υπάρχουν πολλές έρευνες που έχουν αποδείξει την χρησιμότητα και τα πλεονεκτήματα της μάσκας τόσο για την προστασία των άλλων όσο και για τους ίδιους. Η χρήση της είναι απαραίτητη σε εσωτερικούς, κλειστούς χώρους αλλά και εξωτερικούς αφού πολλές φορές υπάρχει συνωστισμός και είναι δύσκολη η διατήρηση μιας ασφαλούς φυσικής απόστασης.
- Φυσική απόσταση και περιορισμός επαφών: Η διατήρηση της φυσικής απόστασης άνω του ενός μέτρου μπορεί να διασφαλίσει την μείωση κινδύνου μόλυνσης μέσω των σταγονιδίων και του αέρα. Η στενή φυσική επαφή και ειδικά σε κλειστούς χώρους έχει συσχετιστεί με αυξημένο κίνδυνο μετάδοσης και μόλυνσης.
- Αποφυγή εσωτερικών και εξωτερικών πολυσύχναστων χώρων: Οι εσωτερικοί χώροι όπου η απόσταση και η χρήση μάσκας δεν είναι δυνατή, όπως για παράδειγμα τα εστιατόρια, έχουν χαρακτηριστεί ως χώροι υψηλού κινδύνου. Αντίστοιχα και τα μαγαζιά εστίασης. Επιπλέον οι εξωτερικοί χώροι όπου

λαμβάνουν χώρα διάφορες εκδηλώσεις, όπως για παράδειγμα συναυλίες, έχουν συνδεθεί με την εξάπλωση του ιού. Φυσικά υπάρχουν μέτρα που οι επιχειρήσεις έλαβαν για την προστασία μετάδοσης. Ο καλός μηχανισμός εξαέρωσης, η απόσταση, η χρήση εξωτερικών χώρων, όπου υπάρχει δυνατότητα, είναι μέτρα προσαρμογής για τις επιχειρήσεις που σχετίζονται με τον τομέα της εστίασης. Παράλληλα η σωματική άσκηση πλέον αναπροσαρμόζεται σε εξωτερικούς χώρους όπου η φυσική απόσταση και η χρήση μάσκας μπορεί να τηρηθεί, καθώς τα γυμναστήρια κρίνονται ως εστίες υπερμετάδοσης. Επίσης η τηλεργασία, όπου είναι εφικτό να εφαρμοστεί, είναι μια πρακτική που ξεκίνησε να εφαρμόζεται σε πολύ μεγάλο βαθμό από την έξαρση της πανδημίας καθώς κρίθηκε ότι ο εργασιακός χώρος είναι επίσης ένα μέρος υπερμετάδοσης του ιού.

- Αυξημένος αριθμός διαγνωστικών τεστ, διάγνωση και απομόνωση: Τα άτομα που έχουν διαγνωστεί ότι νοσούν πρέπει να θέτουν τον εαυτό τους σε καθεστώς καραντίνας και απομόνωσης από τα άτομα που δεν έχουν μολυνθεί. Εξαιτίας της ιδιαιτερότητας του ιού, δηλαδή κάποιος μπορεί να είναι ασυμπτωματικός φορέας, ο έλεγχος των συμπτωμάτων δεν μπορεί να χρησιμοποιείται για τον εντοπισμό μολυσμένων ατόμων. Αντίθετα η αύξηση του αριθμού των τεστ, είτε με την μορφή μοριακών, self test, rapid test, είναι μια πολύ σημαντική στρατηγική για την καταπολέμηση του ιού. Όπου εφαρμόστηκε σωστά, ολοκληρωμένα, στοχευμένα ο ιός αντιμετωπίστηκε αποτελεσματικά και η επιστροφή στην προ-Covid εποχή επανήλθε σε πολύ σημαντικό βαθμό. Άλλωστε αυτή ήταν και η προτροπή του ΠΟΥ σε όλες τις χώρες από την αρχή της πανδημίας, να πραγματοποιήσουν δηλαδή ελέγχους στοχευμένους με συνεχή διαγνωστικά τεστ σε στοχευμένες και τυχαίες ομάδες πληθυσμού.
- Έγκαιρη ιχνηλάτηση επαφών: Η ιχνηλάτηση επαφών είναι μια διαδικασία απόκτησης ολοκληρωμένων πληροφοριών σχετικά με τα άτομα που έχουν διαγνωστεί ότι νοσούν και περιλαμβάνει την αναγνώριση και επικοινωνία με άτομα που έχουν έρθει σε στενή επαφή με τον ασθενή. Η επικοινωνία έχει ως στόχο την ενημέρωση και την παρακολούθηση της υγείας τους καθώς και την απομόνωση τους ώστε να μην έρθουν σε επαφή με άλλα άτομα. Η ιχνηλάτηση επαφών πρέπει να εστιάζεται στα άτομα που έχουν διαγνωστεί θετικοί στον ιό καθώς πρέπει να ταυτοποιούνται τα άτομα του στενού οικογενειακού κύκλου που για παράδειγμα διαμένουν μαζί, ή ήρθαν πρόσφατα σε στενή επαφή. Ειδικότερα αυτά τα άτομα θα πρέπει να μπαίνουν σε καραντίνα καθώς οι πιθανότητες να έχουν τον ιό είναι πολύ υψηλές. Η έγκαιρη ιχνηλάτηση είναι μια σημαντική πρόκληση που επαφίεται τόσο στις κρατικές δομές όσο και στο ίδιο το άτομο που πρέπει να ενημερώσει τα άτομα που ήρθε σε επαφή. Η υποχρεωτική τήρηση της καραντίνας είναι επιβεβλημένη για την προστασία της δημόσιας υγείας και ένας τρόπος για να επιτευχθεί είναι η επιβολή χρηματικών επιπτώσεων με παράλληλη όμως ύπαρξη κατάλληλων υπηρεσιών στήριξης.

- Προστασία των ευπαθών ομάδων: Λόγω της εύκολης και γρήγορης μεταδοτικότητας του ιού, οι ευπαθείς ομάδες κινδυνεύουν περισσότερο καθώς το ποσοστό θνησιμότητας λόγω του ιού είναι αυξημένο σε σύγκριση με τα άτομα που χαρακτηρίζονται υγεί. Μια κατηγορία ατόμων που ανήκουν σε ευπαθή ομάδα είναι οι ηλικιωμένοι, στους οποίους η θνησιμότητα του ιού είναι ιδιαίτερα αυξημένη και ως εκ τούτου είναι η κοινωνική ομάδα που έχει πληγεί περισσότερο. Συνεπώς χρειάζονται μέτρα προστασίας και επίβλεψης σε εγκαταστάσεις μακροχρόνιας φροντίδας ηλικιωμένων και ατόμων με άλλες ιατρικές παθήσεις. Στις συγκεκριμένες μονάδες είναι πολύ σημαντικό να εφαρμόζονται αυστηρά υγειονομικά πρωτόκολλα καθώς και συνεχείς έλεγχοι στο προσωπικό και τους θαμώνες, ώστε να γίνει έγκαιρη ανίχνευση και απομόνωση του ιού. Επιπλέον ο Covid-19 έχει επηρεάσει φυλετικές και εθνοτικές μειονότητες. Αυτό σχετίζεται με την κοινωνική ανισότητα καθώς πολλοί ζουν σε ιδιαίτερα άσχημες συνθήκες και δεν διαθέτουν πρόσβαση σε υγειονομική περίθαλψη.
- Προστασία των εργαζομένων: Οι εργαζόμενοι που η φυσική τους παρουσία είναι επιβεβλημένη λόγω της φύσης του επαγγέλματος, είναι ιδιαίτερα εκτεθειμένοι στον κίνδυνο μόλυνσης. Υγειονομικό και νοσηλευτικό προσωπικό, εργαζόμενοι σε καταστήματα, εργάτες γης, εργοστάσια και βιομηχανίες, είναι μερικά από τα επαγγέλματα που πρέπει να εφαρμόζονται όλα τα υγειονομικά πρωτόκολλα σε απόλυτο βαθμό για την προστασία τους. Επιπλέον από την έναρξη της διάθεσης των εμβολίων, οι πρώτες ομάδες που ξεκίνησαν να εμβολιάζονται ήταν το υγειονομικό και νοσηλευτικό προσωπικό, αφού είναι και οι πιο εκτεθειμένοι στον Covid-19. Άρα ο καθολικός εμβολιασμός όλων των εργαζομένων σε όλες τις ευαίσθητες και κρίσιμες εργασιακές ομάδες θα οδηγήσει στην μεγαλύτερη ασφάλεια και προστασία τους.
- Αναβολή ταξιδιών: Τα ταξίδια και η κάθε είδους μετακίνηση αναπόφευκτα οδηγεί στην μεγαλύτερη διασπορά του ιού από τη μία κοινότητα στην άλλη. Η αναβολή ταξιδιών στις περιπτώσεις που δεν συντρέχει κάποιος ιδιαίτερα σοβαρός λόγος, είναι ο καλύτερος τρόπος να μειωθεί αυτός ο κίνδυνος. Οι περισσότερες χώρες, ανάλογα με τον τόπο έλευσης, ανακοινώνουν κάποια μέτρα που πρέπει να εφαρμόζονται από τον ταξιδιώτη πριν και μετά το ταξίδι. Κάποια από τα μέτρα που είναι απαραίτητα για να ταξιδέψει κάποιος είναι η διενέργεια υποχρεωτικού τεστ και η συμπλήρωση μιας φόρμας υγείας. Επιπλέον κατά την άφιξη σε μια άλλη χώρα, ανάλογα και με τις κυβερνητικές αποφάσεις που ανανεώνονται σε συχνά χρονικά διαστήματα, το πιο συνηθισμένο είναι η υποχρεωτική καραντίνα για δέκα με δεκατέσσερις μέρες. Αυτό το μέτρο δεν είναι καθολικό για όλες τις χώρες παγκοσμίως αλλά σχετίζεται με την χώρα έλευσης και την τελική χώρα-προορισμό.
- Προσωπική υγιεινή και καθαριότητα: Ο συχνός αερισμός του χώρου, η απολύμανση συχνά αγγιζόμενων επιφανειών, ο προσωπικός καθαρισμός, το συχνό πλύσιμο χεριών, βοηθούν στη μείωση μετάδοσης του ιού, καθώς μπορεί να αποτρέψει την μετάδοση μέσω των επιφανειών.

- Εμβολιασμός: Τα εμβόλια αποτελούν τον πιο ασφαλή και αποτελεσματικό τρόπο ελέγχου της πανδημίας. Πλέον το μεγαλύτερο στοίχημα είναι η όσο το δυνατόν μεγαλύτερη εμβολιαστική κάλυψη του πληθυσμού ώστε να επιτευχθεί η μεγαλύτερη δυνατή ανοσία. Όλα τα διαθέσιμα εμβόλια στην χώρα μας, έχουν λάβει την έγκριση του Ευρωπαϊκού Οργανισμού Φαρμάκων (EMA) ο οποίος και συνεχώς επανεξετάζει την αποτελεσματικότητα και την ασφάλεια των εμβολίων που έχουν λάβει έγκριση. Η δικαιότερη διανομή, η οργάνωση της διαδικασίας, η ενημέρωση και η πληροφόρηση των θετικών επιπτώσεων του εμβολιασμού, κρίνεται πολύ σημαντικός για τον εμβολιασμό του μεγαλύτερου μέρους του πληθυσμού, κάθε χώρας. Η Ευρωπαϊκή Ένωση φρόντισε την παραγγελία και την διανομή στις χώρες-μέλη της, όπου σαφώς πήρε μέρος και η χώρα μας σαν κράτος μέλος. Ωστόσο η εμβολιαστική διαδικασία, η ενημέρωση, η οργάνωση και η σειρά προτεραιότητας των εμβολίων γίνεται με βάση τις κυβερνήσεις των χωρών ξεχωριστά. Η εμπιστοσύνη στον εμβολιασμό είναι ένα ιδιαίτερα κρίσιμο στοίχημα που πρέπει να κερδηθεί. (Honein et al., 2020)

2.1.4 Πολιτικές αποφάσεις και Covid-19

Μέχρι στιγμής η πανδημία έχει κοστίσει τη ζωή σε 6.294.969 ανθρώπους μέχρι στιγμής, με το νούμερο αυτό να αυξάνεται μέρα με τη μέρα (“Coronavirus disease (COVID-19) pandemic”). Από την αρχή της πανδημίας όπου δεν υπήρχαν τα εμβόλια, οι κυβερνήσεις παγκοσμίως άρχισαν να λαμβάνουν μέτρα περιοριστικού χαρακτήρα με σκοπό την αναχαίτιση και καλύτερη προστασία των πολιτών. Ωστόσο αυτά τα μέτρα, λόγω της σοβαρότητας της κατάστασης επηρέασαν και άλλαξαν τη ζωή όλων. Η παρούσα εργασία έχει σκοπό να μελετήσει το κοινωνικοπολιτικό και οικονομικό αντίκτυπο που επέφερε η πανδημία, όχι απλα μέσα από βιβλιογραφική ανασκόπηση, αλλά αναλύοντας μεγάλο όγκο δεδομένων από την δημοφιλέστερη πλατφόρμα κοινωνικής δικτύωσης, το Twitter, χρησιμοποιώντας εξόρυξη δεδομένων και διαφόρων τεχνικών μηχανικής μάθησης για τον σκοπό αυτό. Για το τι είναι η εξόρυξη δεδομένων, η μηχανική μάθηση και οι τεχνικές που έχουν χρησιμοποιηθεί για την εκπόνηση της παρούσας εργασίας, το επόμενο κεφάλαιο θα προσφέρει μια αναλυτική και λεπτομερή περιγραφή. Ωστόσο πιο πριν απαιτείται μια περιγραφή των εφαρμοζόμενων μέτρων που επηρέασαν και επηρεάζουν τις ζωές των πολιτών σε παγκόσμια κλίμακα από την αρχή της πανδημίας.

Στο πρώτο κύμα της πανδημίας, λόγω της μη ύπαρξης των εμβολίων και του πρώτου σοκ, οι κυβερνήσεις παγκοσμίως στηρίχτηκαν σε προληπτικά μέτρα που είχαν σαν στόχο τον περιορισμό της μετάδοσης του ιού. Ορισμένα από αυτά είχαν μικρό οικονομικό κόστος και αφορούσαν κυρίως συστάσεις ατομικής προστασίας όπως η χρήση μάσκας και γαντιών. Όμως η εξάπλωση και η επικινδυνότητα του ιού ανάγκασαν τις κυβερνήσεις να λάβουν πιο δραστικά μέτρα. Σχεδόν όλες οι χώρες εφάρμοσαν μέτρα lockdown δηλαδή αναστολή λειτουργίας πολλών δραστηριοτήτων και επιχειρήσεων. Επιχειρήσεις που χαρακτηρίστηκαν ως εστίες υπερμετάδοσης εξαιτίας της φύσης τους διέκοψαν την λειτουργία τους, όπως για παράδειγμα εστιατόρια, επιχειρήσεις εστίασης, κέντρα αισθητικής, γυμναστήρια, εμπορικά μαγαζιά και άλλα. Επιπλέον τα πανεπιστήμια και τα σχολεία έκλεισαν και χρησιμοποιήσαν -

χρησιμοποιούν την μέθοδο της εξ αποστάσεως εκπαίδευσης μέσω της τηλεεκπαίδευσης. Επίσης πολλοί εργασιακοί χώροι, όπου αυτό είναι εφικτό, χρησιμοποίησαν αντίστοιχα την τηλεργασία. Δημόσιες εκδηλώσεις, όπως για παράδειγμα συναυλίες, ακυρώθηκαν. Ακόμη, τέθηκαν περιορισμοί στις μετακινήσεις μεταξύ πόλεων της ίδιας χώρας αλλά και σε επίπεδο χωρών, ώστε μόνο οι πολίτες που έχουν σοβαρούς λόγους επιτρέπεται να μετακινηθούν. Τέλος σαν ένα πολύ σημαντικό περιοριστικό μέτρο, που χαρακτηρίζεται αμφιλεγόμενο από την κοινή γνώμη είναι η απαγόρευση κυκλοφορίας από μια ώρα και μετά. Στην Ελλάδα ένα επιπλέον μέτρο που πάρθηκε ήταν η αποστολή μηνύματος για να βγει κάποιος από το σπίτι του, να κυκλοφορήσει ακόμα και για τις απαραίτητες εργασίες του. Παράλληλα με τα περιοριστικά μέτρα, οικονομικά μέτρα στήριξης ελήφθησαν προκειμένου να στηριχθούν οι επιχειρήσεις και οι άνθρωποι που εξαιτίας της πανδημίας δεν μπορούσαν να εργαστούν και να λειτουργήσουν. Φυσικά οι κυβερνήσεις διεξήγαγαν πολλές καμπάνιες ενημέρωσης για την καλύτερη πληροφόρηση των πολιτών, με σκοπό την προστασία του κοινωνικού συνόλου. Επιπλέον όλες οι χώρες, με διαφορετικούς βέβαια ρυθμούς η κάθε μια, χρησιμοποίησε διαγνωστικά τεστ ανίχνευσης του ιού με σκοπό την καλύτερη εικόνα της επιδημιολογικής κατάστασης.

2.1.5 Επιπτώσεις του Covid-19

Υπάρχουν πολλές ενδείξεις που αποδεικνύουν πως η πανδημία έχει επιφέρει πολύ σημαντικές ψυχολογικές, κοινωνικές και οικονομικές επιπτώσεις, καθώς έχει επηρεάσει τις ζωές όλων μας. Η όλη κατάσταση έχει επιφέρει μια αβεβαιότητα όσον αφορά το μέλλον και πλέον μια αποδοχή της πραγματικότητας, ότι δηλαδή η επαναφορά σε μια κατάσταση η οποία θα μοιάζει, αλλά σίγουρα δεν θα είναι ίδια με την κανονικότητα που γνωρίσαμε πριν την έναρξη της, φαντάζει μάλλον ουτοπική.

- Ψυχολογικές και κοινωνικές επιπτώσεις της πανδημίας.

Από την έναρξη της πανδημίας έχουν πραγματοποιηθεί πολυάριθμες έρευνες και μελέτες με στόχο τον προσδιορισμό και την εξέταση τυχόν ψυχολογικών επιπτώσεων στην ψυχική υγεία του γενικού πληθυσμού, εξαιτίας της πανδημίας. Ένας τομέας που επηρεάστηκε σε υπερβολικά σημαντικό βαθμό από την πανδημία, είναι η κοινωνική ζωή των ανθρώπων. Λόγω της μεταδοτικότητας και της ίδιας της φύσης του ιού, οι κυβερνήσεις σε παγκόσμιο επίπεδο προσπάθησαν να τη θέσουν υπό έλεγχο κλείνοντας διάφορες δραστηριότητες, όπως σχολεία, πανεπιστήμια, εστιατόρια, χώροι διασκέδασης, έχοντας ως συνέπεια η κοινωνική ζωή όλων να έρθει σε δεύτερη μοίρα, καθώς η ατομική προστασία αλλά και η προστασία των ατόμων που απαρτίζουν τον κοινωνικό περίγυρο έγινε ύψιστη προτεραιότητα. Ουσιαστικά η κοινωνική αποστασιοποίηση είναι αναγκαία για την προστασία από τον ιό. Αυτό σημαίνει ότι οι άνθρωποι αναγκάζονται να εργαστούν, να εκπαιδευτούν εξ αποστάσεως, αποφεύγοντας κάθε είδους κοινωνικής συναναστροφής ακόμα και με τα πολύ οικεία τους πρόσωπα. Η συγκεκριμένη πραγματικότητα έρχεται σε σύγκρουση με την ίδια την φύση του ανθρώπου, καθώς ο άνθρωπος είναι ένα ζώο κοινωνικό, που δημιουργεί κοινωνικές σχέσεις και αλληλεπιδρά με το κοινωνικό σύνολο. Μελέτες, στο παρελθόν, διαπίστωσαν ότι μία ριζική και απότομη αλλαγή είναι ικανή να επηρεάσει την κοινωνική ζωή των ανθρώπων, προκαλώντας άγχος, κρίσεις πανικού, κατάθλιψη, ψυχικές

διαταραχές, μοναξιά, απομόνωση, ενώ ο φόβος είναι το συναίσθημα που κυριαρχεί (Jaspreet, 2020). Το συναίσθημα του φόβου σχετίζεται με την ίδια την αρρώστια, τον θάνατο, την απομόνωση, την οικονομική ανασφάλεια, τους κοινωνικούς περιορισμούς και την απώλεια. Ο Covid-19 θεωρείται μαζικό τραύμα το οποίο μπορεί να οδηγήσει σε ψυχολογικά προβλήματα, αλλαγές στη συμπεριφορά, καθώς και ζητήματα εθισμού συμπεριλαμβανομένου της κατανάλωσης αλκοόλ (S. Dubey et al., 2020).

Διάφορες επιστημονικές έρευνες που διεξήχθησαν διαπίστωσαν ότι τα αρνητικά συναισθήματα, όπως το άγχος, η κατάθλιψη, ο θυμός και ο φόβος, αυξήθηκαν σε πολύ μεγάλο βαθμό, σε αντίθεση με τα θετικά συναισθήματα και την ευζωία (Li et al., 2020). Μια άλλη έρευνα διαπίστωσε αυξημένα ποσοστά άγχους και κατάθλιψης στα άτομα που ανήκουν στην ηλικιακή ομάδα από 21 έως 40 χρονών, ενώ παράλληλα παρατηρήθηκε και η αύξηση κατανάλωσης αλκοόλ. Ειδικότερα για την κατανάλωση αλκοόλ παρατηρήθηκε σε πολλές χώρες αύξηση κατανάλωσης όπως έδειξε για παράδειγμα μια σχετική έρευνα στην Αμερική, η οποία διαπίστωσε ότι η κατανάλωση αλκοόλ αυξήθηκε κατά 14% το 2020 σε άτομα που ανήκουν στην ηλικιακή ομάδα από 30 έως 80 ετών (Pollard et al., 2020). Στην Ευρώπη έρευνες δείχνουν αυξητική τάση όσον αφορά την κατανάλωση αλκοόλ η οποία παρατηρείται από τα μέσα του 2020 και συνεχίζεται μέχρι τώρα. Επιπλέον, έρευνες στο Ηνωμένο Βασίλειο διαπίστωσαν την αυξητική τάση όσον αφορά την συχνότητα κατανάλωσης αλκοόλ, καθώς πολλοί ανέφεραν ότι αύξησαν την κατανάλωση αλκοόλ από την έναρξη της πανδημίας και τον αναγκαστικό περιορισμό δραστηριοτήτων. Επίσης διαπιστώθηκε συσχέτιση μεταξύ της αύξησης της κατανάλωσης αλκοόλ και της κακής ψυχικής υγείας, συμπτωμάτων κατάθλιψης και χαμηλότερης ψυχικής ευεξίας (Jacob et al., 2020). Τέλος στην Ελλάδα πραγματοποιήθηκε έρευνα σε δείγμα 705 ενηλίκων, διερευνώντας τις συνήθειες κατανάλωσης πριν και μετά την έναρξη της πανδημίας. Στην συγκεκριμένη μελέτη δεν παρατηρήθηκε ιδιαίτερη μεταβολή, όμως διαπιστώθηκε αύξηση στο ότι περισσότεροι άνθρωποι έπιναν μόνοι τους. Αρκετοί από τους συμμετέχοντες στην έρευνα δήλωσαν περισσότερη κατανάλωση αλκοόλ και ανέφεραν σαν κύριες αιτίες την απομόνωση, τις αλλαγές στην καθημερινή τους ζωή και την προσπάθεια διαχείρισης άγχους ή κατάθλιψης (Panagiotidis et al., 2020).

Οι επιπτώσεις της πανδημίας στην ψυχική υγεία, δυστυχώς φαίνεται από το γεγονός ότι υπάρχουν πολλές αναφορές περιπτώσεων ανθρώπων που αυτοκτόνησαν και σχετίζονται άμεσα με την πανδημία. Στις Ηνωμένες Πολιτείες, την Γερμανία, την Ιταλία, την Ινδία, το Ηνωμένο Βασίλειο και σε άλλες χώρες υπάρχουν πολλές αναφορές στα μέσα μαζικής ενημέρωσης και στη διεθνή βιβλιογραφία που περιγράφουν αντίστοιχα περιστατικά και το πως σχετίζονται με την πανδημία. Ένα τέτοιο περιστατικό ήταν μια δεκαεπταετής σερβιτόρα από την Αγγλία η οποία κατέληξε έπειτα από την προσπάθεια της να βάλει τέλος στη ζωή της λόγω του φόβου των συνεπειών που θα είχε στη ψυχική της υγεία η απομόνωση (Miller, 2020). Επίσης μια άλλη περίπτωση ενός τριανταεξάχρονου ανθρώπου από το Μπαγκλαντές όπου ο ίδιος και οι συγγωριανοί του πίστευαν ότι νοσεί από τον ιό εξαιτίας των συμπτωμάτων που εμφάνισε. Η μεταθανάτια εξέταση έδειξε ότι ήταν αρνητικός (Mamun and Griffiths, 2020). Επιπλέον ένα ακόμα περιστατικό ήταν μιας σαρανταεπταετής γυναίκας από την Νέα Υόρκη η οποία ήταν επικεφαλής του τμήματος επειγόντων περιστατικών, και αφού διηγήθηκε στην οικογένεια της περιστατικά νοσηλείας, φροντίδας και απώλειας ασθενών που φρόντιζε, αυτοκτόνησε (Rosner and Sheehy, 2020). Στο

συγκεκριμένο σημείο είναι ιδιαίτερα σημαντικό να αναφερθεί η προσπάθεια των ανθρώπων που εργάζονται στον τομέα της υγείας και από την πρώτη στιγμή προσπαθούν να σώσουν ζωές. Οι επιπτώσεις της πανδημίας στην ψυχική υγεία των εργαζόμενων του υγειονομικού τομέα είναι σοβαρό αντικείμενο μελέτης και έρευνας καθώς και οι συνέπειες μετά την επιστροφή στην καθημερινότητα. Συνεπώς έρευνες έχουν διαπιστώσει ότι η κοινωνική απομόνωση, το άγχος, ο φόβος μετάδοσης, η αβεβαιότητα και οι οικονομικές δυσκολίες μπορούν να οδηγήσουν στην ανάπτυξη ή επιδείνωση διαταραχών που σχετίζονται με το στρες και με αυτοκτονικές συμπεριφορές. Άτομα με ιστορικό ψυχιατρικών διαταραχών και άτομα που έχουν βιώσει την απώλεια συγγενικών, κοντινών τους ατόμων εξαιτίας του ιού είναι ιδιαίτερα ευάλωτα (Sher, 2020).

Από το πρώτο κύμα της πανδημίας οι κυβερνήσεις σε παγκόσμια κλίμακα ξεκίνησαν να λαμβάνουν μέτρα καραντίνας και περιορισμού των ανθρώπων στα σπίτια τους. Ενώ η πρόθεση ήταν και παραμένει η προσπάθεια ελαχιστοποίησης της μετάδοσης του ιού, πέρα από τις επιπτώσεις στην ψυχική υγεία των ανθρώπων που αναφέρθηκαν παραπάνω, ξεκίνησαν να υπάρχουν ανησυχίες για την αύξηση της ενδοοικογενειακής βίας και ειδικότερα για την κακοποίηση γυναικών και παιδιών. Σε μεγάλο βαθμό οι ανησυχίες βασίζονταν στο γεγονός ότι οι γονείς και τα παιδιά πλέον είναι περιορισμένα στα σπίτια τους χωρίς καμία πρόσβαση σε αυτούς που θα μπορούσαν να διαπιστώσουν σημάδια κακοποίησης, ή και να απευθυνθούν για βοήθεια ώστε να ξεφύγουν από αυτές τις καταστάσεις (Piquero et al., 2021). Ουσιαστικά ο κατ' οίκων περιορισμός, ο οικονομικός αντίκτυπος της πανδημίας, η ανεργία, η ανασφάλεια, η παιδική μέριμνα είναι ένας συνδυασμός που αυξάνουν τον κίνδυνο κακοποίησης ή κλιμάκωσης της ενδοοικογενειακής βίας, τόσο από ήδη κακοποιητικούς συζύγους-συντρόφους όσο και από προηγούμενως μη βίαιους. Συνεπώς είναι σαφές ότι η οποιαδήποτε μείωση των αναφορών που σχετίζονται με την ενδοοικογενειακή βία, δεν αντικατοπτρίζει μια ρεαλιστική κατάσταση, αλλά αντίθετα υποδεικνύει μεγαλύτερο κίνδυνο για το θύμα, αφού είναι παγιδευμένο χωρίς την δυνατότητα να ζητήσει εξωτερική βοήθεια (Sacco et al., 2020). Σύμφωνα όμως με τη βιβλιογραφική έρευνα υπάρχουν πολλές μελέτες που αναφέρουν ότι έχει αυξηθεί η ενδοοικογενειακή βία και υπάρχει ισχυρή συσχέτιση με την πανδημία και την καραντίνα. Σε αντίθεση με τον φόβο ότι το θύμα θα είναι πιο δύσκολο να αναφέρει και να καταγγείλει τέτοια περιστατικά, αναφορές στην βιβλιογραφία που βασίζονται σε αστυνομικές αναφορές, οργανώσεις και δίκτυα έδειξαν ότι πιθανότατα η καραντίνα έπαιξε καταλυτικό ρόλο στο να πάρουν την απόφαση να καταγγείλουν τα περιστατικά αυτά, ειδικά στις περιπτώσεις όπου η κακοποιητική συμπεριφορά προϋπήρχε της πανδημίας (Piquero et al., 2021). Τέτοια συγκλονιστικά περιστατικά ήρθαν στο φως της επικαιρότητας από την έναρξη της πανδημίας ενώ πολλά από αυτά κατέληξαν σε εγκλήματα, όπως για παράδειγμα ένα περιστατικό ενός άνδρα από της Ηνωμένες πολιτείες που σκότωσε την κοπέλα του και στη συνέχεια αυτοκτόνησε λόγω του φόβου ότι είχαν προσβληθεί από κορονοϊό (Garger, 2020). Αντίστοιχο περιστατικό έλαβε χώρα στην Ιταλία όπου ένας εικοσιοκτάχρονος σκότωσε την κοπέλα του επειδή νόμιζε ότι του μετέδωσε τον ιό, ενώ στη συνέχεια προσπάθησε να αυτοκτονήσει ανεπιτυχώς. (ΣΚΑΙ, 2020)

- Οικονομικές επιπτώσεις

Το ξέσπασμα της πανδημίας έχει μειώσει και σταματήσει διάφορες οικονομικές δραστηριότητες παγκοσμίως λόγω της επιβολής περιοριστικών μέτρων για τον περιορισμό

της εξάπλωσης του ιού. Ο κόσμος είχε πολλές δεκαετίες να βιώσει αντίστοιχη εμπειρία και αυτό ουσιαστικά έχει επιφέρει σημαντική επιβάρυνση σε διάφορες οικονομίες χωρών, οι οποίες υπέστησαν σοκ εξαιτίας της πρωτόγνωρης υγειονομικής κρίσης. Άλλωστε, μπορεί η κρίση να είναι υγειονομική αλλά ταυτόχρονα είναι κοινωνική και οικονομική εξαιτίας των επιπτώσεων που έχει προκαλέσει στις ζωές όλων. Φυσικά ο οικονομικός αντίκτυπος διαφέρει από χώρα σε χώρα καθώς η οικονομική κατάσταση κάθε χώρας ήταν διαφορετική πριν το ξέσπασμα της πανδημικής κρίσης. Αυτός ο φόβος είναι ιδιαίτερα ρεαλιστικός και υπάρχει ιδιαίτερη ανησυχία ότι θα εντείνει την οικονομική ανισότητα και παγκόσμια φτώχεια, της οποίας το μέγεθος θα φανεί μετά το πέρας της πανδημίας, ή και όσο θα συνεχίζεται η παρούσα κατάσταση.

Σε οικονομικό επίπεδο η πανδημία προκάλεσε αναστάτωση τόσο στην ζήτηση όσο και στην προσφορά σε παγκόσμια κλίμακα. Από την πλευρά της προσφοράς, οι διαταραχές προέκυψαν εξαιτίας των μολύνσεων, οι οποίες μείωσαν την παραγωγικότητα και την προσφορά εργασίας. Τα περιοριστικά μέτρα που εφαρμόστηκαν στην κινητικότητα, την κοινωνική απόσταση, το αναγκαστικό κλείσιμο επιχειρήσεων και διαφόρων κλάδων της οικονομίας αναστάτωσε την προμήθεια αγαθών και υπηρεσιών. Από την πλευρά της ζήτησης η απώλεια εισοδήματος, οι απολύσεις, η ανεργία, η οικονομική ανασφάλεια και η αβεβαιότητα είχαν σαν αποτέλεσμα τη μείωση των επενδύσεων και της κατανάλωσης.

Παρά τις προσπάθειες των κυβερνήσεων να ανακόψουν τις οικονομικές επιπτώσεις της πανδημίας, το αποτέλεσμα ήταν να χαθούν 6,3 τρισεκατομμύρια δολάρια από το παγκόσμιο πραγματικό ΑΕΠ το 2020 και το πρώτο τρίμηνο του 2021. Αυτό σημαίνει ότι έχουμε παγκόσμια ύφεση της τάξεως του 4,5% σύμφωνα με τους διεθνείς οργανισμούς, και αυτό παρά τις διάφορες κυβερνητικές παρεμβάσεις για την αντιμετώπιση των φορολογικών και νομισματικών επιπτώσεων. Παράλληλα υπολογίζεται ότι οι εισαγωγές-εξαγωγές μειώθηκαν κατά 9,2% σε παγκόσμια κλίμακα. Όσον αφορά την εργασία, καθώς λόγω των περιοριστικών μέτρων πολλοί έχασαν τις δουλειές τους, ενώ άλλοι τέθηκαν σε αναστολή εργασίας, υπολογίζεται ότι χάθηκαν 114 εκατομμύρια θέσεις εργασίας (Βέργου et al., 2021). Όσον αφορά τον τραπεζικό τομέα, η μακροοικονομική επιβράδυνση έχει ως αποτέλεσμα την αύξηση των μη εξυπηρετούμενων δανείων. Ειδικότερα οι ιδιωτικές τράπεζες είναι αυτές που έχουν την υψηλότερη έκθεση στον πιστωτικό κίνδυνο. Επιπλέον λόγω της γενικευμένης καραντίνας υπήρξε μια μείωση του όγκου των τραπεζικών συναλλαγών και μείωση της χρήσης ATM, τα κέρδη των τραπεζών μειώθηκαν. Από την άλλη πλευρά αυξήθηκαν οι ηλεκτρονικές αγορές και το ηλεκτρονικό εμπόριο, καθώς πάρα πολύ κόσμος στράφηκε προς αυτού του είδους αγορών (Ozili and Thankom, 2020).

2.2 Μέσα Κοινωνικής Δικτύωσης

Τα Μέσα Κοινωνικής Δικτύωσης (social media) αποτελούν πλέον αναπόσπαστο κομμάτι της καθημερινότητας για ένα μεγάλο τμήμα του παγκόσμιου πληθυσμού. Στο πλαίσιο αυτό, η συμμετοχή στα μέσα κοινωνικής δικτύωσης αυξάνεται καθημερινά. Πρόσφατες στατιστικές έρευνες υποδεικνύουν ότι πάνω από τα δύο τρίτα των χρηστών του διαδικτύου είναι ενεργοί χρήστες σε κάποια πλατφόρμα κοινωνικής δικτύωσης (Kemp, 2017).

Η ραγδαία εξέλιξή τους έχει επέμβει στη μορφή της διαπροσωπικής επικοινωνίας και την ανθρώπινη αλληλεπίδραση. Η χρήση τους αφορά τόσο την προσωπική επικοινωνία όσο και την επιχειρηματικότητα, γι' αυτό τα τελευταία χρόνια αποτελούν αντικείμενο μελέτης ενός ευρύτερου φάσματος επιστημών, όπως η τεχνολογία, το μάρκετινγκ, η δημοσιογραφία, η πολιτική (Pantic, 2014). Η παρατήρηση της ανθρώπινης συμπεριφοράς στα social media έχει απασχολήσει τους ερευνητές κοινωνικών επιστημών ήδη από το 1998.

Ο όρος social media είναι δύσκολος να οριστεί διότι με την έννοια αυτή, άλλες φορές αναφερόμαστε σε μια πλατφόρμα, άλλες φορές σε ένα λογισμικό και άλλες σε μια δραστηριότητα. Επιπλέον συνεχώς εξελίσσεται και αλλάζει, με αποτέλεσμα να προστίθενται συνέχεια καινούργιοι ορισμοί και έννοιες. Παρόλα αυτά υπάρχουν πολλοί ορισμοί που αποσαφηνίζουν την έννοια αυτή καθ' αυτή. Ένας πολύ σημαντικός ορισμός είναι από τους Kaplan και Haenlein, οι οποίοι ορίζουν τα μέσα κοινωνικής δικτύωσης, ως μια ομάδα εφαρμογών βασισμένες στο διαδίκτυο, φτιαγμένες στα ιδεολογικά και τεχνολογικά θεμέλια του Web 2.0. Οι συγκεκριμένες εφαρμογές επιτρέπουν την δημιουργία και την ανταλλαγή περιεχομένου από τους χρήστες (Kaplan and Haenlein, 2010). Ένας άλλος ορισμός είναι ότι τα μέσα κοινωνικής δικτύωσης αποτελούν εργαλεία που αυξάνουν την ικανότητα μας να μοιραζόμαστε, να συνεργαζόμαστε και να πραγματοποιούμε συλλογικές δράσεις (Shirky, 2008). Ένας επιπλέον ορισμός είναι των Lon Safko και David K. Brake στο βιβλίο τους *The Social Media Bible – Tactics, Tools, and Strategies for Business Success*, όπου χαρακτηριστικά λένε πως τα social media αναφέρονται σε δραστηριότητες, πρακτικές και συμπεριφορές ανάμεσα σε κοινότητες ατόμων που συγκεντρώνονται στο διαδίκτυο, με σκοπό να μοιραστούν πληροφορίες, γνώσεις και απόψεις χρησιμοποιώντας μέσα συζήτησης (conversational media). Conversational media είναι εφαρμογές που καθιστούν δυνατή τη δημιουργία και την εύκολη μετάδοση περιεχομένου με τη μορφή λέξεων, εικόνων, βίντεο και ηχογραφημένων μηνυμάτων (Safko, 2010).

Πλεον οι άνθρωποι μπορούν εύκολα να επικοινωνούν μεταξύ τους σε όλο τον κόσμο και να μοιραστούν προσωπικές απόψεις για διάφορα θέματα όπως η πολιτική, η υγεία, η παιδεία. Η μετάδοση ειδήσεων είναι ένα χαρακτηριστικό παράδειγμα που απεικονίζει την ταχύτατη ανταλλαγή και μετάδοση πληροφοριών σε όλο τον κόσμο. Αυτό έχει επιτρέψει τους ανθρώπους να ξεφύγουν από τους παραδοσιακά μέσα ενημέρωσης, τα οποία πολλές φορές έχουν κατηγορηθεί ως μη αντικειμενικά και αξιόπιστα. Τα μέσα κοινωνικής δικτύωσης αποτελούν τα μέσα διάδοσης ενός μεγάλου όγκου δεδομένων, που σχετίζονται με πληροφορίες, προσωπικές εμπειρίες, απόψεις, ιδέες (Khan et al., 2021).

2.2.1 Twitter

Ένα χαρακτηριστικό μέσο κοινωνικής δικτύωσης ιδιαίτερα δημοφιλές και ευρέως διαδεδομένο είναι το Twitter. Αποτελεί το τρίτο πιο δημοφιλές μέσο κοινωνικής δικτύωσης μετά το Facebook και το Instagram. Σε σύγκριση με τα υπόλοιπα το Twitter έχει ένα απλό μοντέλο δεδομένων και ένα απλό API ανάκτησης δεδομένων, το οποίο χρησιμοποιήθηκε στην παρούσα εργασία και θα αναλυθεί παρακάτω. Αυτα τα χαρακτηριστικά καθιστούν το Twitter ιδανικό για μελέτες κοινωνικών δικτύων που προσπαθούν να αναλύσουν, να ερευνήσουν και να επεξεργαστούν πρότυπα διαδικτυακής συμπεριφοράς, τη δομή του κοινωνικού γραφήματος, το συναίσθημα για διάφορες οντότητες και τη φύση των

κακόβουλων επιθέσεων σε ένα ζωντανό δίκτυο με εκατομμυρια χρήστες. Δεν είναι λοιπόν τυχαίο ότι έχει καθιερωθεί ως μια πολύ σημαντική ερευνητική πλατφόρμα (Antonakaki et al., 2021).

Μια ανάρτηση στο Twitter είναι ένα σύντομο κείμενο το οποίο ονομάζεται tweet και περιλαμβάνει μέχρι 280 χαρακτήρες. Σαν εφαρμογή, επιτρέπει την κοινή χρήση σύντομων κειμένων, εικόνες, βίντεο, υπερσυνδέσμους προς άλλους ιστότοπους (Erskine and Hendricks, 2021). Όλοι οι λογαριασμοί είναι δημόσιοι εξ' αρχής, πράγμα που σημαίνει ότι οποιοσδήποτε χρήστης μπορεί να διαβάσει το tweet του συγκεκριμένου λογαριασμού, εκτός και αν ο ίδιος αλλάξει τις αρχικές ρυθμίσεις του λογαριασμού και το καταστήσει ιδιωτικό.

Ιδιαίτερα κατά τη διάρκεια κρίσιμων γεγονότων όπως έξαρση ασθενειών, φυσικών και ανθρώπινων καταστροφών το twitter λαμβάνει ένα τεράστιο αριθμό μηνυμάτων. Για το λόγο αυτό αποτελεί ένα σημαντικό εργαλείο που μπορεί να χρησιμοποιηθεί με σκοπό την ανάλυση και επεξεργασία διάφορων σημαντικών γεγονότων συμβάλλοντας στην αντιμετώπιση διαφόρων κρίσεων. Πάρα πολλές εργασίες και έρευνες έχουν χρησιμοποιήσει το Twitter σε διάφορα ερευνητικά θέματα και πεδία. Συνεπώς δεν θα μπορούσε να μην χρησιμοποιηθεί και αξιοποιηθεί για ένα τόσο σημαντικό και παγκοσμιο φαινόμενο όπως είναι η πανδημία. Ο σκοπός της συγκεκριμένης εργασίας είναι η ανάλυση του Covid-19 μέσα από τη πλατφόρμα κοινωνικής δικτύωσης Twitter. Η δυνατότητα συλλογής και αξιοποίησης τεράστιων όγκων δεδομένων για ερευνητικούς σκοπούς είναι διαθέσιμη και πιο εύκολη στη συγκεκριμένη πλατφόρμα, τηρώντας κάποιους κανόνες και ικανοποιώντας τις απαιτήσεις που θέτει.

Ωστόσο το Twitter API θέτει περιορισμούς, καθώς οι όροι παροχής υπηρεσιών του εμποδίζουν την αναπαραγωγή μεγάλης ποσότητας δεδομένων. Τα δεδομένα αυτά μπορούν να χρησιμοποιηθούν για σκοπούς έρευνας και ατομικής χρήσης αλλά δεν επιτρέπεται η δημόσια έκθεση τους. Για τον λόγο αυτό, τα περισσότερα dataset, που είναι διαθέσιμα από άλλες εργασίες, δίνουν πρόσβαση μόνο σε έναν αριθμό, το tweet id, το οποίο αντιστοιχεί σε ένα tweet. Ένας τρόπος ανάκτησης ολόκληρης της πληροφορίας του tweet είναι μέσα από ένα εργαλείο, που ονομάζεται twarc. Δημιουργήθηκε από μια ομάδα ατόμων, οι οποίοι βρίσκονται πίσω από το Documenting Now, το οποίο είναι ένα έργο που στοχεύει στη συλλογή και αρχειοθέτηση δεδομένων προερχόμενα από τα Μέσα Κοινωνικής Δικτύωσης, μέσα από ηθικές και διαφανείς διαδικασίες ("Twitter Data Collection Using Twarc"). Είναι ένα εργαλείο εντολών και μια βιβλιοθήκη της Python, το οποίο συλλέγει δεδομένα tweet μέσω του Twitter API και τα αρχειοθετεί σε json αρχεία. Κάθε tweet αντιπροσωπεύει ένα αντικείμενο json, που είναι ακριβώς αυτό που επιστρέφεται από το Twitter API, με όλη τη διαθέσιμη πληροφορία. Το twarc εκτελεί τρεις λειτουργίες: την αναζήτηση (search), τη ροή (stream) και ενυδάτωση (hydration) (Ruest and Milligan, 2016). Η εντολή hydrate διαβάζει το αρχείο με τους αναγνωριστικούς αριθμούς, το tweet id. Στη συνέχεια χρησιμοποιώντας το Twitter API, παράγει το json αρχείο. Στη παρούσα εργασία χρειάστηκε το twarc και συγκεκριμένα η εντολή hydrate, καθώς από το dataset που επιλέχθηκε, χρειάστηκε πρόσβαση σε όλες τις πληροφορίες που δεν ήταν διαθέσιμες.

2.3 Τεχνητή Νοημοσύνη

Ο όρος Τεχνητή Νοημοσύνη (TN, AI) αναφέρεται στον κλάδο της πληροφορικής, ο οποίος ασχολείται με τη σχεδίαση και υλοποίηση υπολογιστικών συστημάτων που μιμούνται την ανθρώπινη συμπεριφορά. (McCarthy, 1989) Τα συστήματα αυτά χαρακτηρίζονται ως ευφυή διότι καθιστούν τις μηχανές, ικανές να εκτελούν ανθρωπομορφικά έργα. Τα υπολογιστικά συστήματα που χρησιμοποιούνται ονομάζονται ευφυείς πράκτορες στους οποίους εφαρμόζονται μέθοδοι αναπαράστασης γνώσης ή επίλυσης προβλημάτων. Η TN χρησιμοποιεί ευφυείς πράκτορες, οι οποίοι αντιλαμβάνονται το περιβάλλον τους μέσω «αισθητήρων» και δρουν μέσα σε αυτό ως μέσα διέγερσης. Ένας ανθρώπινος πράκτορας χρησιμοποιεί σαν αισθητήρες βιολογικά όργανα (μάτια, αυτιά, μύτη, κτλ). Αντίστοιχα ένας ρομποτικός πράκτορας μπορεί να χρησιμοποιεί κάμερες ή άλλου είδους κινητήρες, ενώ ένας πράκτορας λογισμικού μπορεί να λαμβάνει εισόδους από το πληκτρολόγιο, περιεχόμενα αρχείων ή διαδικτυακά πακέτα σαν αισθητήρες και αντιδρά στο περιβάλλον αντίστοιχα, με εμφάνιση των αποτελεσμάτων στην οθόνη ή με την αποστολή πακέτων (McCarthy, 1989).

Η Τεχνητή Νοημοσύνη είναι ένα δυναμικά ισχυρό εργαλείο για την καταπολέμηση της πανδημίας. Διάφορες έρευνες έχουν ασχοληθεί με πιθανές εφαρμογές της Τεχνητής Νοημοσύνης, ιδίως την ικανότητα να αναγνωρίζει, να προβλέπει, να εξηγεί αλλά και να διαχειρίζεται πιθανές οικονομικές, κοινωνικές επιπτώσεις της πανδημίας. Από την έξαρση της πανδημίας υπήρξε μια μεγάλη κινητοποίηση αξιοποίησης της TN και άλλων εργαλείων ανάλυσης δεδομένων για αυτούς τους σκοπούς.

2.3.1 Εξόρυξη Δεδομένων

Η Εξόρυξη Δεδομένων είναι η διερεύνηση-ανάλυση μεγάλου όγκου δεδομένων χρησιμοποιώντας μαθηματικούς αλγόριθμους και τεχνικές εκμάθησης υπολογιστών, με σκοπό την ανίχνευση υποθετικών και προηγούμενων άγνωστων σχέσεων και μοτίβων. Συχνά ο στόχος είναι η πρόβλεψη συμπεριφορών και αποτελεσμάτων, η ταυτοποίηση συσχετίσεων και η ομαδοποίηση παρόμοιων περιπτώσεων. Η Εξόρυξη Δεδομένων, αλλιώς γνωστή ως Ανάλυση Δεδομένων, είναι ένας όρος που περιγράφει ένα σύνολο τεχνικών ανάλυσης που μπορούν να χρησιμοποιηθούν για τους παραπάνω σκοπούς και έχει τις ρίζες της στις επιστήμες της στατιστικής, της τεχνητής νοημοσύνης και της μηχανικής μάθησης (Fayyad et al., 1996).

Η Εξόρυξη Δεδομένων επιτρέπει την αξιολόγηση των δεδομένων ουτως ώστε να διαπιστωθούν πολύπλοκα μοτίβα, να δημιουργηθούν κανόνες, να αποκαλυφθούν προηγούμενες σχέσεις, να προβλέψει, να αναπτύξει ή να επαληθεύσει θεωρίες. Για τον σκοπό αυτό εκμεταλλεύεται μεγάλες ποσότητες δεδομένων και επιτρέπει στα δεδομένα να προτείνουν ποιες σχέσεις αξίζουν προσοχή και χρήζουν περισσότερης ανάλυσης (Murray et al., 2009).

Οι τεχνικές ανάλυσης δεδομένων που χρησιμοποιούνται στην Εξόρυξη Δεδομένων αποτελούνται από εργαλεία ανάλυσης όπως η δημιουργία δέντρων αποφάσεων, συστάδες, κανόνες συσχέτισης, ανάλυση παλινδρόμησης, ανάλυση χρονοσειρών, λειτουργίες κινδύνου και ανάλυση επιβίωσης. Όλες οι παραπάνω τεχνικές μπορούν να εφαρμοστούν με πολλούς διαφορετικούς τρόπους χρησιμοποιώντας διαφορετικές αλγοριθμικές μεθοδολογίες. Σε γενικές γραμμές η Εξόρυξη Δεδομένων είναι κατάλληλη για:

- Αναγνώριση ετερογενών επιδράσεων ή κοινών χαρακτηριστικών μεταξύ ομάδων ή ατόμων, που συμπεριφέρονται με όμοιο τρόπο και στη συνέχεια αναλύονται οι διαφορές τους.
- Ανάλυση συσχέτισης και σύνδεσης. Πιο συγκεκριμένα η κατανόηση ποιων στοιχείων, χαρακτηριστικών, συμπεριφορών ή κινδύνων εμφανίζονται μαζί.
- Ανάλυση σπάνιων συμπεριφορών, ανθρώπων, ή γεγονότων που είναι πιθανό να συμπεριφέρονται αντίθετα από το αναμενόμενο.
- Ανάλυση της τάσης. Δηλαδή, η περιγραφή των ενεργειών ή των διαφορών, μεταξύ των ατόμων με τη πάροδο του χρόνου. (Murray and Scime, 2015)

Η Εξόρυξη Δεδομένων έχει εφαρμοστεί σε πολλά επιστημονικά πεδία, όπως για παράδειγμα στην κοινωνιολογία, στη ψυχολογία, στις πολιτικές επιστήμες, στη ποινική δικαιοσύνη αλλά και στην επιστήμη της υγείας, με αξιοσημείωτα και ιδιαίτερα χρήσιμα αποτελέσματα που μπορούν μελλοντικά να αξιοποιηθούν. Στη παρούσα εργασία θα χρησιμοποιηθούν τεχνικές εξόρυξης δεδομένων με σκοπό την ανάλυση tweets που αφορούν τον ιο Covid-19, προκειμένου να εξαχθούν χρήσιμες πληροφορίες που αφορούν την πανδημία.

Οι τεχνικές Εξόρυξης Δεδομένων κατηγοριοποιούνται σε επιβλεπόμενη ανάλυση (supervised analysis) και σε μη επιβλεπόμενη ανάλυση (unsupervised analysis). Η διαφορά μεταξύ τους είναι ότι στην επιβλεπόμενη ανάλυση ανιχνεύεται η σχέση μιας μεταβλητής με τις υπόλοιπες μεταβλητές στα δεδομένα, ενώ αντίθετα στη μη επιβλεπόμενη ανάλυση ανιχνεύονται οι σημαντικές σχέσεις μεταξύ των μεταβλητών και των τιμών τους στα δεδομένα. Ένα παράδειγμα κατανόησης των διαφορών τους είναι ότι αν ένας ερευνητής επιθυμεί να προβλέψει μια συμπεριφορά ψηφοφορίας, μπορεί να χρησιμοποιήσει επιβλεπόμενη ανάλυση για να κατηγοριοποιήσει τους πολίτες με βάση το κόμμα που θα ψηφίσουν. Αντίθετα αν επιθυμεί να ανιχνεύσει τα χαρακτηριστικά των πολιτών που σχετίζονται με το κάθε κόμμα θα χρησιμοποιήσει μη επιβλεπόμενη ανάλυση (Murray and Scime, 2015). Φυσικά δεν είναι καθόλου ασυνήθιστο να χρησιμοποιούνται και οι δύο μέθοδοι ανάλυσης σε ένα σύνολο δεδομένων (Rajasethupathy, 2009).

Οι πιο διαδεδομένες τεχνικές είναι η ανάλυση ταξινόμησης (classification), η συσχέτιση (association), η ομαδοποίηση (clustering).

- Ανάλυση ταξινόμησης: Είναι επιβλεπόμενη τεχνική και λειτουργεί δημιουργώντας ένα μοντέλο δέντρου ταξινόμησης ανιχνεύοντας ένα μονοπάτι σε μια προκαθορισμένη εξαρτώμενη ή μεταβαλλόμενη μεταβλητή για κάθε περίπτωση. Μπορούν να χρησιμοποιηθούν και συνεχείς μεταβλητές. (Murray and Scime, 2015)
- Συσχέτιση: Είναι μια μη επιβλεπόμενη τεχνική, η οποία ανιχνεύει μοτίβα από διακριτές μεταβλητές στα δεδομένα. (Murray and Scime, 2015)
- Ομαδοποίηση: Χρησιμοποιείται για την εύρεση ομοιοτήτων μεταξύ μεταβλητών που είναι είτε συνεχείς είτε διακριτές. Ουσιαστικά ομαδοποιεί ένα σύνολο ετερογενών μεταβλητών σε ομοιογενείς υποομάδες χρησιμοποιώντας φίλτρα ομοιότητας. Συχνά προηγείται άλλων τεχνικών ανάλυσης επειδή βοηθάει στον εντοπισμό χρήσιμων μεταβλητών, που μπορούν στην πορεία να αξιοποιηθούν περαιτέρω. (Murray and Scime, 2015)

2.3.2 Μηχανική Μάθηση

Η Μηχανική Μάθηση (Machine Learning) ορίζεται ευρέως ως μία υπολογιστική στρατηγική που προσδιορίζει αυτόματα (δηλαδή μαθαίνει) μεθόδους και παραμέτρους για την επίτευξη βέλτιστης λύσης σε ένα πρόβλημα χωρίς να απαιτεί να είναι προγραμματισμένο από έναν άνθρωπο για να παραδώσει μια εμπειριστατωμένη λύση .

Στόχος μιας μεθόδου μηχανικής μάθησης είναι η υλοποίηση ενός αλγορίθμου, ο οποίος θα κάνει μια εκτίμηση σε σχέση με την άγνωστη υπάρχουσα συσχέτιση μεταξύ των δεδομένων. Εφ' όσον έχει γίνει η εκτίμηση του βαθμού εξάρτησης, τότε μπορεί να γίνει και πρόβλεψη εξόδων για γνωστά δεδομένα εισόδου.

2.3.3 Μηχανική Μάθηση και Τεχνητή Νοημοσύνη

Η MM αποτελεί υποκατηγορία της Τεχνητής Νοημοσύνης, η οποία σχετίζεται άμεσα με τη δημιουργία ευφύων προγραμμάτων. Για να θεωρηθεί ένα σύστημα ευφύς οφείλει να έχει δυνατότητες που σχετίζονται με κάποια από τις ακόλουθες κατηγορίες (Russell and Norvig, 2002):

- Επεξεργασία φυσικής γλώσσας - Natural Language Processing (NLP), για την επικοινωνία.
- Αναπαράσταση Γνώσης - Knowledge Representation, για το διαχωρισμό μεταξύ των δεδομένων, πληροφοριών και γνώσης.
- Αυτόματος Συμπερασμός- Automated Reasoning, για τη χρήση των αποθηκευμένων δεδομένων.
- Μηχανική Μάθηση - Machine Learning, για την προσαρμογή του συστήματος σε νέες συνθήκες, την ανίχνευση και εξαγωγή νέων μοτίβων.

Οι πράκτορες της TN μπορούν να βελτιώσουν τη συμπεριφορά τους μέσα από την μελέτη των εμπειριών τους. Ένας πράκτορας μαθαίνει, εάν η απόδοσή του σε μια μελλοντική διεργασία είναι καλύτερη, εφόσον πρώτα έχει παρατηρήσει τον κόσμο (Russell and Norvig 2002).

Η εκμάθηση ενός γενικού κανόνα ή συνάρτησης από ζεύγη δεδομένων εισόδου-εξόδου ονομάζεται Επαγωγική Μάθηση. Εύκολα μπορεί κάποιος να αντιληφθεί ότι η ανάπτυξη μεθόδων μηχανικής μάθησης είναι η βέλτιστη προσέγγιση για τη λύση του καθορισμένου προβλήματος, καθώς περιορίζει τους παράγοντες σφάλματος, βελτιστοποιεί τον χρόνο που χρειάζεται για την διαδικασία και προσφέρει αυτοματοποιημένη δυνατότητα προσαρμογής κατά την μεταβολή των συνθηκών (Russell and Norvig, 2002).

2.3.4 Μηχανική Μάθηση και Στατιστική

Βασικός άξονας της επιστήμης της στατιστικής είναι η περιγραφή και κατανομή των συσχετίσεων μεταξύ των δεδομένων. Εφαρμογές της στατιστικής στη MM αποτελεί η ανάπτυξη μοντέλων για :

- την παρατήρηση της κατανομής μιας μεταβλητής,
- τον χαρακτηρισμό των σχέσεων μεταξύ δύο μεταβλητών,

- τον εντοπισμό ομοιοτήτων μεταξύ των πληθυσμών που παρατηρούνται.

Επιπλέον, σχετίζεται με τον τομέα της ανίχνευσης επαναλαμβανόμενων μοτίβων που εμφανίζονται στα δεδομένα.

Πολλές μέθοδοι που εφαρμόζονται στη στατιστική για την κατανόηση και ερμηνεία των δεδομένων χρησιμοποιούνται και στη ΜΜ. Οι μέθοδοι εξόρυξης γνώσης από τα δεδομένα που βασίζονται στη στατιστική ανάλυση είναι σε θέση να αναγνωρίσουν, να διατηρήσουν τη δομή των δεδομένων και να την αναπαράξουν σε διαφορετικά δείγματα τα οποία εξάγονται από τον ίδιο πληθυσμό.

2.3.5 Υπολογιστικό Μοντέλο

Οι αλγόριθμοι της μηχανικής μάθησης έχουν συνήθως στόχο την διαμόρφωση ενός υπολογιστικού μοντέλου κατάλληλο για την ανάλυση των δεδομένων που εξετάζονται. Το μοντέλο αυτό μπορεί να είναι είτε περιγραφικό, είτε προβλεπτικό. Το περιγραφικό μοντέλο αποσκοπεί στην αναγνώριση των προτύπων και συσχετίσεων μεταξύ των δεδομένων, καθώς και στην διερεύνηση των ιδιοτήτων των δεδομένων. Σε αυτό συναντάμε μεθόδους συσταδοποίησης, την ανάλυση συσχετίσεων και ανίχνευση ακολουθιών. Το προβλεπτικό μοντέλο στοχεύει στην πρόβλεψη της τιμής δεδομένων, βασιζόμενο σε στοιχεία ιστορικού. Σε αυτό συναντάμε αλγόριθμους κατηγοριοποίησης, μεθόδους παλινδρόμησης και ανάλυσης χρονοσειρών.

Ταυτόχρονα, ένα υπολογιστικό μοντέλο μπορεί να είναι είτε παραμετρικό ή μη παραμετρικό. Κατά τα παραμετρικά μοντέλα προσδιορίζεται η συσχέτιση που υπάρχει ανάμεσα σε δεδομένα εισόδου και εξόδου με τη χρήση αλγεβρικών εξισώσεων. Μέσα σε αυτές τις εξισώσεις υπάρχουν παράμετροι, οι οποίες παίρνουν τιμές μέσω της εκπαίδευσης που έχει προηγηθεί κατά την είσοδο των δεδομένων. Πρόκειται για ένα προϋπάρχον συγκεκριμένο μοντέλο, το οποίο λαμβάνει γνώση από τα δεδομένα πριν ξεκινήσει η διαδικασία. (Russell and Norvig, 2002)

Στα μη παραμετρικά μοντέλα δεν περιέχονται παράμετροι, καθώς αυτά καθοδηγούνται από τα δεδομένα και προσαρμόζονται σ' αυτά. Πιο συγκεκριμένα, διαμορφώνεται ένα μοντέλο το οποίο βασίζεται στα δεδομένα εισόδου και εμπεριέχει τεχνικές ΜΜ που το καθιστούν δυναμικό και εύκολα προσαρμόσιμο σε νέα δεδομένα. Σε αυτήν την κατηγορία ανήκουν τα νευρωνικά δίκτυα, οι γενετικοί αλγόριθμοι, τα δέντρα αποφάσεων κ.α. (Russell and Norvig, 2002)

2.3.6 Κατηγορίες Μηχανικής Μάθησης

Η μηχανική μάθηση χωρίζεται σε τρεις βασικές κατηγορίες ανάλογα με τον τύπο της μάθησης που απαιτείται κάθε φορά. Αυτές είναι (Jun Chin, 2015):

- Επιβλεπόμενη Μάθηση – Supervised Learning.

Στόχος είναι η πρόβλεψη της τιμής μιας καινούριας περίπτωσης ως προς ένα χαρακτηριστικό, με δεδομένες τις τιμές αυτής της περίπτωσης σε κάποια άλλα

χαρακτηριστικά, τα οποία ονομάζονται μεταβλητές πρόβλεψης (predictive variables). Για να γίνει αυτό, απαιτείται η εκπαίδευση ενός μοντέλου μηχανικής μάθησης, η οποία πραγματοποιείται σε ένα σύνολο εκπαίδευσης (training set) με μια διαδικασία που ορίζεται από κάποιον αλγόριθμο μηχανικής μάθησης. Στο σύνολο εκπαίδευσης είναι γνωστές οι τιμές των παραδειγμάτων τόσο για τις μεταβλητές πρόβλεψης όσο και για τη μεταβλητή – στόχο (target variable). Στόχος οποιουδήποτε αλγορίθμου μηχανικής μάθησης είναι η εκμάθηση μιας συνάρτησης μεταξύ των μεταβλητών πρόβλεψης και της μεταβλητής – στόχου. Η εξαγόμενη συνάρτηση ονομάζεται μοντέλο μηχανικής μάθησης και αυτό εφαρμόζεται σε καινούρια δεδομένα (περιπτώσεις) με σκοπό την πρόβλεψη της τιμής αυτών σε ότι αφορά τη μεταβλητή – στόχο.

Τα προβλήματα επιβλεπόμενης μάθησης διακρίνονται σε προβλήματα κατηγοριοποίησης (classification) και σε προβλήματα παλινδρόμησης (regression):

- ❖ Κατηγοριοποίηση : Προβλέπει την κατηγορία στην οποία ανήκουν τα δεδομένα, όταν τιμών η μεταβλητή – στόχος έχει διακριτό εύρος τιμών.
- ❖ Παλινδρόμηση (Regression) : Προβλέπει την τιμή της μεταβλητής – στόχου όταν αυτή είναι συνεχής (δηλαδή όταν μπορεί να λάβει οποιαδήποτε τιμή από το σύνολο των πραγματικών αριθμών).
- Μη επιβλεπόμενη Μάθηση – Unsupervised Learning.

Κατά τη μη επιβλεπόμενη μάθηση οι αλγόριθμοι προσπαθούν να εντοπίσουν συγκεκριμένα μοτίβα στα υπάρχοντα δεδομένα, χωρίς να υπάρχει κάποια προηγούμενη γνώση σχετικά με αυτά. Σε αντίθεση με την επιβλεπόμενη μάθηση, δεν υπάρχει το στάδιο της εκπαίδευσης κάποιου μοντέλου. Ωστόσο, οι μέθοδοι μη επιβλεπόμενης μάθησης μας επιτρέπουν να εξάγουμε χρήσιμες πληροφορίες για τη δομή και τις υποβόσκουσες σχέσεις μεταξύ των δεδομένων.

- Ενισχυτική Μάθηση – Reinforcement Learning .

Η ενισχυτική μάθηση στο πλαίσιο της τεχνητής νοημοσύνης εκπαιδεύει τους αλγόριθμους μέσα από το σύστημα τιμωρίας – επιβράβευσης. Οι ίδιοι αποφασίζουν για τις ενέργειές τους και καθορίζουν ποια πράξη είναι υπεύθυνη για το αποτέλεσμα κάθε φορά. Εφαρμογές της ενισχυτικής μάθησης είναι ο έλεγχος κίνησης των ρομπότ ή η αυτοματοποίηση εργασιών ή η εκμάθηση ενός παιχνιδιού από τον υπολογιστή.

Η βασική διαφορά που εντοπίζεται στους παραπάνω τύπους αφορά τη διαδικασία μάθησης.

2.4 Εφαρμογή Τεχνικών Ανάλυσης Κειμένου

Τα δεδομένα με τα οποία ασχοληθήκαμε είναι δεδομένα κειμένου που προέρχονται από το Twitter. Σ' αυτά εφαρμόστηκαν τεχνικές μοντελοποίησης θεμάτων (topic modeling). Τα δεδομένα κειμένου, για να είναι εφικτό να αναλυθούν μέσω αλγορίθμων μηχανικής μάθησης, πρέπει πρώτα να αναπαρασταθούν με έναν κατάλληλο τρόπο. Στο συγκεκριμένο

κεφάλαιο περιγράφονται τεχνικές αναπαράστασης των δεδομένων, καθώς και οι αλγόριθμοι θεματικής μοντελοποίησης που χρησιμοποιήθηκαν.

2.4.1 Αναπαράσταση Δεδομένων κειμένου

Ένας πολύ συνηθισμένος τρόπος αναπαράστασης δεδομένων κειμένου είναι με το λεγόμενο μοντέλο BoW (Bag of Words). Σύμφωνα με αυτό, κάθε έγγραφο (tweet) αναπαρίσταται ως ένα διάνυσμα τιμών, όπου κάθε τιμή αφορά μια λέξη και στην ουσία εκφράζει τη σημαντικότητα της λέξης στο κείμενο. Η διαδικασία της αναπαράστασης με το μοντέλο BoW αρχίζει με το διαχωρισμό των κειμένων σε λέξεις (tokenization) και το σχηματισμό ενός λεξιλογίου $\{t_1, t_2, \dots, t_v\}$ που αποτελείται από τις λέξεις που εμφανίζονται στο σύνολο των εγγράφων. Οι λέξεις του λεξιλογίου είναι τα χαρακτηριστικά που αναπαριστούν τα κείμενα. Επομένως, ένα έγγραφο d_j αναπαρίσταται ως ένα διάνυσμα $\{tw_{1j}, tw_{2j}, \dots, tw_{vj}\}$ όπου tw_{ij} είναι το βάρος (η σημαντικότητα) της λέξης t_i στο συγκεκριμένο έγγραφο. Στην απλή περίπτωση, την οποία ακολουθούμε στη συγκεκριμένη εργασία, η τιμή για μια λέξη-χαρακτηριστικό για ένα συγκεκριμένο έγγραφο είναι η συχνότητα αυτής της λέξης στο έγγραφο. Η συγκεκριμένη αναπαράσταση χρησιμοποιήθηκε λόγω του ότι αποτελεί τη συνηθισμένη πρακτική όταν η ανάλυση των κειμένων γίνεται με Θεματική Μοντελοποίηση.

Συχνά, ως κομμάτι της αναπαράστασης των κειμένων, πριν τη διαδικασία του tokenization, μπορούν να εφαρμοστούν κάποιες τεχνικές προ-επεξεργασίας των κειμένων, με σκοπό τη βελτίωση των αποτελεσμάτων.

Η προεπεξεργασία των δεδομένων είναι ένα πολύ σημαντικό και κρίσιμο στάδιο καθώς καθορίζει σε πολύ μεγάλο βαθμό την επιτυχή εφαρμογή των αλγορίθμων. Λόγω του περιορισμού των 280 χαρακτήρων που επιτρέπει το Twitter, συχνά χρησιμοποιούνται συντομογραφίες που θα μπορούσαν να ερμηνευτούν διαφορετικά, ανάλογα με το πλαίσιο που διατυπώνονται. Επιπλέον η χρήση λέξεων σε μια πιο εκλαϊκευμένη και νεανική διάλεκτο (αργκό), όπως επίσης και η μη ορθή γραμματική αυξάνουν την ανάγκη να προηγηθεί ένα στάδιο προεπεξεργασίας των δεδομένων. Ένα tweet μπορεί να περιέχει διάφορα σύμβολα όπως είναι # (hashtag), @ (mention), συνδέσμους, φωτογραφίες, εικόνες και emojis. Όλα τα παραπάνω δυσχεραίνουν την ανάλυση, καθώς μπορούν να επηρεάσουν τα αποτελέσματα μειώνοντας την απόδοση των τεχνικών ανάλυσης κειμένου. Συνεπώς είναι μια ιδιαίτερα σημαντική διαδικασία η οποία απαιτεί προσοχή, λαμβάνοντας υπόψη το είδος της ανάλυσης, τους αλγόριθμους που θα χρησιμοποιηθούν και τις μεθόδους υλοποίησης, καθώς δεν είναι συγκεκριμένα και σίγουρα τα στάδια προεπεξεργασίας σε κάθε περίπτωση. Για παράδειγμα, διαφορετική προεπεξεργασία συνίσταται στην ανίχνευση συναισθήματος και διαφορετική στην επεξεργασία κειμένου.

Οι πιο συνηθισμένες τεχνικές προεπεξεργασίας κειμένου είναι:

- Εξαγωγή URLs: πολλά tweet περιέχουν URL για να μοιράζονται περισσότερο περιεχόμενο εξαιτίας του περιορισμού στο όριο των χαρακτήρων που επιτρέπεται να κοινοποιήσει κάποιος. Επιπλέον συνήθως περιέχει επιπλέον πληροφορίες που ο χρήστης επιθυμεί να μοιραστεί ή να υποστηρίξει την γνώμη του και τις σκέψεις του. Ωστόσο επειδή δεν είναι ιδιαίτερα πρακτική η ανίχνευση του περιεχομένου κάθε

διεύθυνσης URLs θεωρείται ιδιαίτερα τετριμμένο και βασικό στάδιο της προεπεξεργασίας που σκοπό έχει την μείωση του μεγέθους των δεδομένων.

- Αφαίρεση των stopwords: η αφαίρεση λέξεων που συναντώνται αρκετά συχνά, είναι επίσης μια πολύ συχνή διαδικασία, καθώς δεν προσφέρουν καμία αξία και καμία χρήσιμη πληροφορία.
- Αφαίρεση συμβόλων @, #: αυτά τα σύμβολα χρησιμοποιούνται για να τονίσουν μια συγκεκριμένη λέξη και να επισημάνουν το όνομα ενός χρήστη αντίστοιχα. Δεν συμβάλλουν όμως στην ανάλυση και επεξεργασία των δεδομένων.
- Μετατροπή από κεφαλαία σε μια μικρά γράμματα: σε ένα tweet είναι αρκετά συνηθισμένο να υπάρχει εναλλαγή μεταξύ κεφαλαίων και μικρών χαρακτήρων. Επίσης μια πολύ συνηθισμένη τεχνική είναι η μετατροπή όλων των χαρακτήρων σε μικρούς ώστε η σύγκριση και η ανάλυση να είναι ευκολότερη.
- Lemmatization: προκειται για μια αλγοριθμική διαδικασία κατά την οποία γίνεται ο προσδιορισμός του λήμματος μιας λέξης με βάση τη προβλεπόμενη σημασία της (“Lemmatization”). Επαναφέρει κάθε λέξη στην βασική της μορφή που είναι η λέξη η οποία συναντάται σε ένα λεξικό (“Stemming and Lemmatization in Python”).
- Stemming: είναι μια διαδικασία μείωσης των λέξεων στη μορφή της βάσης ή της ρίζας τους. Ουσιαστικά θα μπορούσε κάποιος να σκεφτεί ότι πρόκειται για μια παρόμοια τεχνική με το Lemmatization, αφού και οι δύο αλγόριθμοι επαναφέρουν την λέξη στη ρίζα της. Ωστόσο, ακριβώς το αντίθετο. Πιο απλοϊκά κατά το stemming, η κατάληξη μιας λέξης είναι σαν κόβεται, αντίθετα με το lemmatization, όπου η λέξη επαναφέρεται στην αρχική της μορφή, η οποία είναι κανονική λέξη στη πραγματικότητα. Αντίθετα στο stemming δεν είναι απαραίτητο ότι η λέξη υπάρχει στο λεξικό (Johnson, 2022). Ανάλογα με το είδος της ανάλυσης και τους αλγόριθμους που εφαρμόζονται, επιλέγεται μία από τις δύο τεχνικές.
- Αφαίρεση αριθμών

2.4.2 Μοντελοποίηση θεμάτων

Η μοντελοποίηση θεμάτων (topic modeling) είναι μια δημοφιλής τεχνική μηχανικής μάθησης που χρησιμοποιείται ιδιαίτερα συχνά στην εξόρυξη δεδομένων και την επεξεργασία κειμένου. Πρόκειται για ένα σύνολο αλγορίθμων που αποκαλύπτουν τις θεματικές δομές που κρύβονται σε μια συλλογή από έγγραφα. Αυτοί οι αλγόριθμοι μας βοηθούν να αναπτύξουμε νέους τρόπους αναζήτησης και να συνοψίσουμε μεγάλες συλλογές κειμένων (Yang and Zhang, 2018).

Η βασική ιδέα είναι ότι το έγγραφο αντιμετωπίζεται συνήθως ως μια μίξη από θέματα. Κάθε θέμα θεωρείται μια πιθανή κατανομή των λέξεων. Όταν η μοντελοποίηση θεμάτων χρησιμοποιείται ως εργαλείο εξόρυξης κειμένου, κάθε θέμα θεωρείται μια συλλογή από λέξεις και κάθε έγγραφο μπορεί να θεωρηθεί ως ένα σύνολο θεμάτων με διαφορετικές αναλογίες, ανάλογα με τη συχνότητα των όρων (Yang and Zhang, 2018). Η μοντελοποίηση θεμάτων είναι μια τεχνική μη επιβλεπόμενης μάθησης, καθώς δεν είναι απαραίτητο κανένα είδος εκπαίδευσης με ήδη ταξινομημένα δεδομένα. Ως συνέπεια, δεν υπάρχουν συγκεκριμένα χαρακτηριστικά των θεμάτων που πρέπει να εξεταστούν από τον αλγόριθμο πριν από το στάδιο της ανάλυσης (Dimitriadis, 2020).

Η έννοια της μοντελοποίησης θεμάτων αποτελείται από οντότητες δηλαδή “λέξεις” (word), “έγγραφα” (documents) και “σώματα” (corpora). Οι λέξεις (words) είναι η βασική μονάδα διακριτών δεδομένων σε ένα έγγραφο, που ορίζονται ως λεξιλογικά στοιχεία τα οποία είναι καταχωρημένα για κάθε μοναδική λέξη στο έγγραφο. Ένα έγγραφο είναι μια διάταξη από N λέξεις. Corpus είναι μια συλλογή εγγράφων και ένα θέμα (topic) είναι η διανομή ενός σταθερού λεξιλογίου. Πιο απλοϊκά, κάθε έγγραφο στο corpus περιέχει τις δικές του θεματικές αναλογίες σύμφωνα με τις λέξεις που περιέχονται σ’ αυτά (Negara et al., 2019).

2.4.3 Latent Dirichlet Allocation (LDA)

Είναι η πιο δημοφιλής τεχνική μοντελοποίησης θεμάτων (topic modelling). Εμφανίστηκε ως μια από τις επιλεγμένες μεθόδους για την ανάλυση πολύ μεγάλων εγγράφων (Negara et al., 2019). Οι David Blei, Andrew Ng και Michael I. Jordan ήταν οι πρώτοι που πρότειναν το μοντέλο LDA ως ένα θεματικό ερευνητικό γραφικό μοντέλο το 2003 (Yang and Zhang, 2018). Μπορεί να χρησιμοποιηθεί για να συνοψίσει, να ομαδοποιήσει, να συνδέσει και να επεξεργαστεί πολύ μεγάλα δεδομένα. Ο LDA αλγόριθμος εξαρχής βασίζεται σε δύο υποθέσεις (Blei 2003):

- Ένα έγγραφο αποτελείται από διάφορα θέματα. Ο αλγόριθμος παρέχει ως έξοδο την πιθανότητα κατανομής ότι ένα έγγραφο περιλαμβάνει ορισμένα θέματα. Τα έγγραφα σε ένα corpus αντιπροσωπεύονται ως κατανομές πιθανοτήτων σε θέματα, που σημαίνει πιο απλά ότι τα θέματα έχουν μια πιθανότητα να αποτελούν μέρος ενός εγγράφου.
- Ένα θέμα είναι μια συλλογή από λέξεις. Επιστρέφει ως έξοδο την πιθανότητα κατανομής όπου ένα θέμα περιλαμβάνει ορισμένες λέξεις. Τα θέματα στο corpus αντιπροσωπεύονται ως κατανομές πιθανοτήτων σε λέξεις, το οποίο πρακτικά σημαίνει ότι σε όλες τις λέξεις ενός εγγράφου μπορεί να ανατεθεί μια πιθανότητα να ανήκουν σε ένα θέμα.

Ο αλγόριθμος LDA είναι, εκτός από τον πιο διαδεδομένο αλγόριθμο μοντελοποίησης θεμάτων, ιδιαίτερα αποτελεσματικός. Ωστόσο έχουν αναφερθεί κάποιοι περιορισμοί, οι οποίοι επηρεάζουν την απόδοση και τα αποτελέσματα του αλγορίθμου. Κάποιες αδυναμίες του LDA είναι :

- Ο αριθμός των θεμάτων πρέπει να οριστεί εξαρχής μέσα από την παράμετρο k .
- Η Dirichlet κατανομή πιθανότητας δεν είναι ικανή να συσχετίσει θέματα.
- Δεν είναι ιεραρχικός. Αυτό σημαίνει ότι δεν επιτρέπει την κοινή χρήση δεδομένων, σε αντίθεση με τους αλγόριθμους που χαρακτηρίζονται ιεραρχικοί.
- Στατικός αλγόριθμος, δηλαδή δεν περιλαμβάνει την έννοια του χρόνου.
- Βασίζεται στην τεχνική bag of words (BoW), η οποία υποθέτει ότι οι λέξεις είναι ανταλλάξιμες και ότι η δομή μιας πρότασης δεν διαμορφώνεται.
- Ο LDA αλγόριθμος είναι πιο αποδοτικός σε μεγάλα κείμενα, ενώ αντίθετα υστερεί στα πιο μικρά, όπως είναι για παράδειγμα ένα tweet (Qiang, 2020).

2.4.4 GSDMM

Ο Gibbs Sampling Dirichlet Multinomial Mixture (GSDMM) είναι ένας αλγόριθμος ομαδοποίησης μικρών κειμένων. Οι Jianhua Yin και Jianyong Wang (Yin and Wang, 2014) το 2014 σε εργασία τους πρότειναν τον συγκεκριμένο αλγόριθμο, υποστηρίζοντας ότι λύνει το πρόβλημα της μη ακριβούς μοντελοποίησης της γλώσσας που οφείλεται στην μη παρατήρηση αρκετών δεδομένων (sparsity problem) όταν πρόκειται για σύνολα μικρών κειμένων, η οποία προκύπτει από το ότι τα μικρά κείμενα αποτελούνται από λίγες λέξεις του συνολικού λεξιλογίου. Για το λόγο αυτό, ο GSDMM εξειδικεύεται στην ανάλυση κειμένων μικρού μεγέθους (short text topic modeling). Ουσιαστικά, πρόκειται για έναν αλγόριθμο που αποτελεί τροποποίηση του LDA, υποθέτοντας ότι ένα έγγραφο ανήκει σε ένα μόνο θέμα, αντίθετα με τον LDA που υποθέτει ότι μπορεί να ανήκει σε πολλά θέματα (Bowers, 2022). Ως αποτέλεσμα, ο GSDMM υποθέτει ότι οι λέξεις ενός εγγράφου δημιουργούνται χρησιμοποιώντας ένα μοναδικό θέμα και όχι από μια μείξη θεμάτων, όπως στο LDA (Amrouche, 2022).

Η βασική φιλοσοφία του GSDMM περιγράφεται από τους Yin και Wang (Yin and Wang, 2014) με ένα απλό αναλογικό παράδειγμα που ονομάζεται “Movie Group Process”. Σύμφωνα με το παράδειγμα υποθέτουμε ότι υπάρχει μια ομάδα ατόμων που τους αρέσουν οι ταινίες σε ένα εστιατόριο και κάθονται τυχαία σε K τραπέζια. Τους ζητάμε να γράψουν σε ένα χαρτί μια λίστα με τις αγαπημένες τους ταινίες. Στη συνέχεια τους μεταφέρουμε σε K τραπέζια, με στόχο να τους ομαδοποιήσουμε, έτσι ώστε αυτοί με την ίδια αγαπημένη ταινία να κάτσουν στο ίδιο τραπέζι. Για να πετύχει τους ζητάμε να κάνουν μια νέα επιλογή τραπέζιου, με τη σειρά ο ένας μετά τον άλλον, εφαρμόζοντας δύο κανόνες.

1. Επιλέξτε τραπέζι με περισσότερα άτομα. Αυτός ο κανόνας βελτιώνει την πληρότητα. Σε όλους όσους αρέσει η ίδια ταινία, θα κάτσουν στο ίδιο τραπέζι.
2. Επιλέξτε ένα τραπέζι με άτομα παρόμοιου ενδιαφέροντος ταινίας. Αυτός ο κανόνας αποσκοπεί στην αύξηση της ομοιογένειας.

Μετά την επανάληψη αυτών των βημάτων, κάποια τραπέζια αναμένονται να μείνουν άδεια, ενώ άλλα να συγκεντρώσουν μεγαλύτερες ομάδες ατόμων. Αυτή είναι η διαδικασία που ο GSDMM αλγόριθμος ακολουθεί (Yin and Wang, 2014).

Πριν την υλοποίηση του αλγόριθμου, πρέπει να οριστούν κάποιες παράμετροι. Αυτές είναι (Yin and Wang, 2014):

- K : είναι ο αριθμός των θεμάτων
- α : αφορά την πληρότητα, η προεπιλεγμένη τιμή είναι 0,1
- β : αφορά την ομοιογένεια, η προεπιλεγμένη τιμή είναι 0,1
- n_iters : είναι ο αριθμός των επαναλήψεων που πρόκειται να εκτελέσει ο αλγόριθμος

2.4.5 Αξιολόγηση

Η αξιολόγηση των αλγορίθμων μοντελοποίησης θεμάτων είναι πρόκληση εξαιτίας της μη επιβλεπόμενης εκπαίδευσης του μοντέλου. Ωστόσο είναι πολύ σημαντικό να προσδιοριστεί αν ένα μοντέλο είναι αντικειμενικά επιτυχημένο ή όχι, καθώς και αν μπορεί να συγκριθεί με διαφορετικά μοντέλα και μεθόδους. Παραδοσιακά μια πρακτική που χρησιμοποιείται είναι με βάση την αντίληψη και την προσωπική κρίση του καθένα. Δηλαδή εμπειρικά, με βάση τη λογική και τη γνώση, μπορεί να διαπιστωθεί αν τα αποτελέσματα είναι

έγκυρα. Επιστημονικά όμως δεν είναι ορθός αυτός ο τρόπος και για αυτό τον λόγο υπάρχουν μετρήσεις που μπορεί κάποιος να λάβει υπόψη του για να αξιολογήσει το μοντέλο του. Στην μοντελοποίηση θεμάτων δύο είναι οι πιο συχνές μετρήσεις που εφαρμόζονται για να αξιολογηθεί η επιτυχία του μοντέλου. Αυτές είναι:

- **Perplexity:** Είναι μια μέτρηση που χρησιμοποιείται ευρέως για την αξιολόγηση μοντέλων. Καταγράφει τον βαθμό έκπληξης του μοντέλου από νέα δεδομένα που δεν έχει δει στο παρελθόν και μετριέται ως η κανονικοποιημένη πιθανότητα καταγραφής ενός συνόλου δοκιμών σε αναμονή. Η πιθανότητα καταγραφής αφορά το πόσο πιθανό είναι να δοθούν νέα δεδομένα στο μοντέλο. Δηλαδή πόσο καλά το μοντέλο αναπαραγάγει και αντιπροσωπεύει τα στατιστικά στοιχεία των δεδομένων (Karadia, 2019).
- **Topic Coherence:** Είναι ο βαθμός συνεκτικότητας. Δηλαδή μετρά τον βαθμό σημασιολογικής ομοιότητας λέξεων με υψηλή βαθμολογία που βρίσκονται στο θέμα. Βοηθά στη διάκριση των θεμάτων που είναι σημασιολογικά ερμηνεύσιμα από τα θέματα που προκύπτουν μέσα από στατιστικά συμπεράσματα (Karadia, 2019).

Υπάρχουν διαφορετικοί τρόποι υπολογισμού για την μέτρηση του coherence. Επιγραμματικά αναφέρονται κάποιοι από αυτούς (Karadia, 2019).

- **C_v:** Βασίζεται σε ένα συρόμενο παράθυρο, την τμηματοποίηση των κορυφαίων λέξεων, σε ένα μέτρο έμμεσης επιβεβαίωσης που χρησιμοποιεί κανονικοποιημένες σειριακές αμοιβαίες πληροφορίες (NPMI) καθώς και την ομοιότητα του συνημιτόνου. Προτάθηκε στο (Roder 2015). Η συγκεκριμένη μετρική παίρνει τιμές από 0 ως 1 και μεγαλύτερες τιμές σημαίνουν καλύτερη ποιότητα μοντέλου.
- **C_p:** Βασίζεται σε ένα συρόμενο παράθυρο, σε μια προηγούμενη τμηματοποίηση των κορυφαίων λέξεων και στο μέτρο επιβεβαίωσης της συνοχής με βάση τον Fitchelson.
- **C_{ucc}:** Βασίζεται σε ένα συρόμενο παράθυρο και στις σημειακές αμοιβαίες πληροφορίες όλων των ζευγών των κορυφαίων λέξεων (PMI).
- **C_{umass}:** Βασίζεται σε μια προηγούμενη τμηματοποίηση και σε μία λογαριθμική υπό όρους πιθανότητα ως μέτρο επιβεβαίωσης. Μπορεί να πάρει μόνο αρνητικές τιμές. Μεγαλύτερη τιμή σημαίνει καλύτερο μοντέλο.
- **C_{npmi}:** Είναι μια βελτιωμένη εκδοχή της C_{ucc} που χρησιμοποιεί τις κανονικοποιημένες σημειακές αμοιβαίες πληροφορίες NPMI.
- **C_a:** Βασίζεται σε ένα παράθυρο περιβάλλοντος, μια σύγκριση των κορυφαίων λέξεων κατά ζεύγη και ένα μέτρο έμμεσης επιβεβαίωσης που χρησιμοποιεί κανονικοποιημένες σημειακές αμοιβαίες πληροφορίες (NPMI) καθώς και την ομοιότητα του συνημιτόνου.

2.4.6 Εργαλεία Οπτικοποίησης και Ανάλυσης Αποτελεσμάτων LDA

Η μοντελοποίηση θεμάτων είναι ένα πολύ χρήσιμο εργαλείο, αλλά η ερμηνεία των αποτελεσμάτων δεν είναι πάντα εύκολη υπόθεση. Για τον λόγο αυτό η οπτικοποίηση των δεδομένων συνεισφέρει στην απάντηση μερικών κρίσιμων ερωτήσεων που προκύπτουν μετά την ανάλυση. Μερικά τέτοια κρίσιμα ερωτήματα είναι (Sievert and Shirley, 2014):

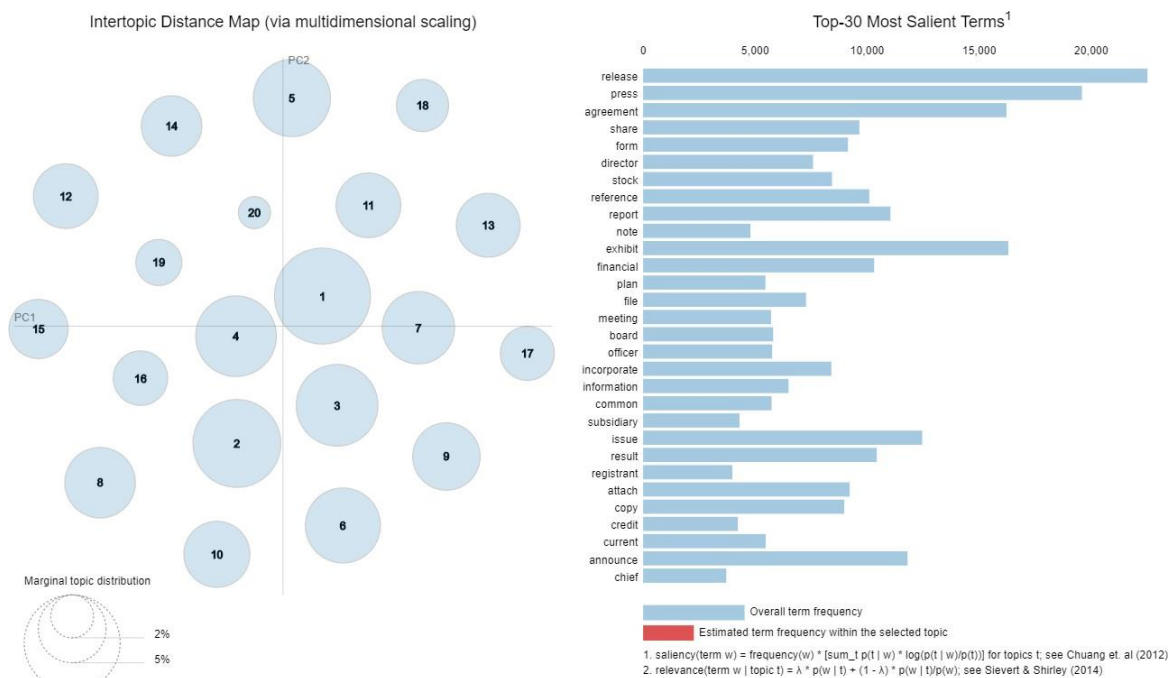
α) Ποια είναι η σημασία κάθε θέματος.

β) Πόσο σχετικό είναι κάθε θέμα.

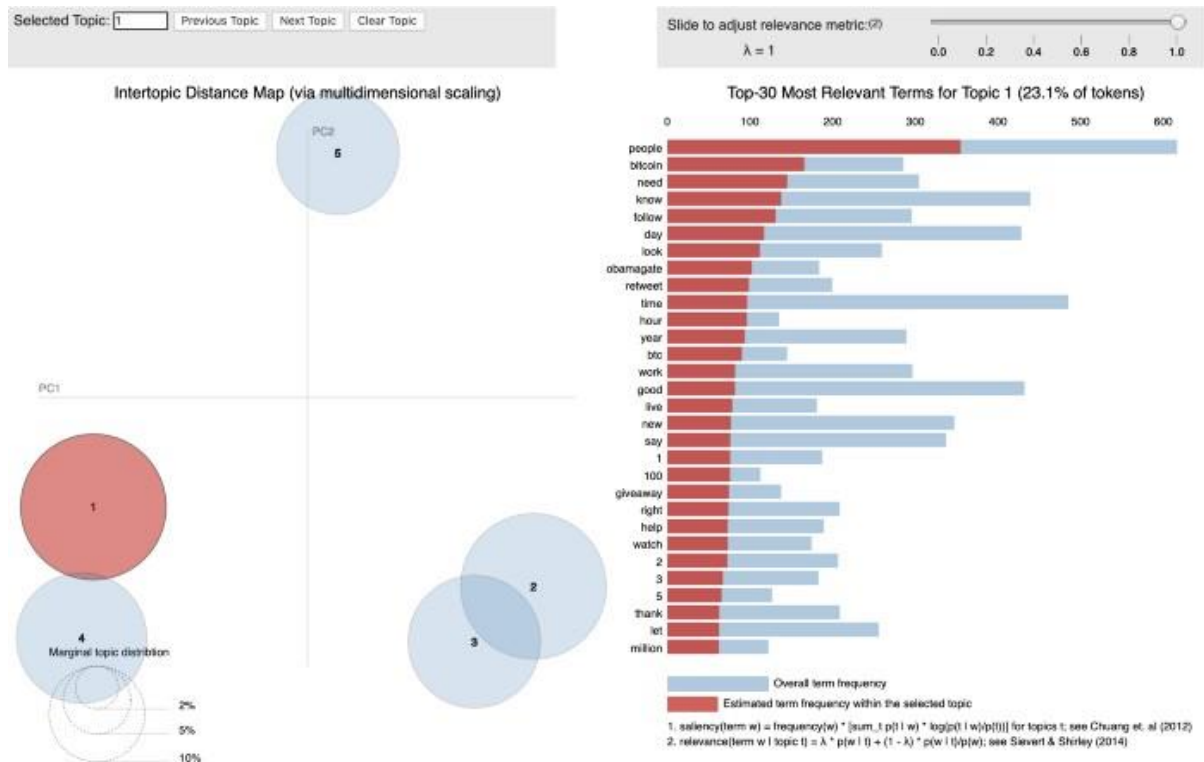
γ) Πως σχετίζονται μεταξύ τους.

Τα τελευταία χρόνια έχουν αναπτυχθεί πολλά συστήματα οπτικοποίησης για τη μοντελοποίηση θεμάτων. Αυτό σχετίζεται με τα νέα εργαλεία και τις νέες τεχνικές που έχουν εξελιχθεί και πλέον χρησιμοποιούνται για τη συγκεκριμένη ανάλυση. Η πιο απλή απεικόνιση είναι γραφήματα, διαγράμματα και word clouds τα οποία χρησιμοποιούν λίστες με τις πιο συχνές λέξεις που βρίσκονται σε κάθε θέμα.

Ένα εργαλείο οπτικοποίησης αποτελεσμάτων είναι το pyLDAvis. Είναι ιδιαίτερα διαδεδομένο και θεωρείται το πιο αποτελεσματικό εργαλείο ειδικά για τα αποτελέσματα που προκύπτουν από το LDA. Πρόκειται για μια βιβλιοθήκη της rython, μέσω της οποίας, διαδραστικά οπτικοποιούνται τα αποτελέσματα της μοντελοποίησης θεμάτων. Οι Kenneth Shirley και Sievert Carson είναι οι δημιουργοί του, οι οποίοι προσπάθησαν να δημιουργήσουν ένα αποτελεσματικό και εύκολο τρόπο οπτικής απεικόνισης για τη διευκόλυνση της ερμηνείας των αποτελεσμάτων του LDA. Το pyLDAvis επιτρέπει την ερμηνεία των θεμάτων, εξάγοντας πληροφορίες από ένα προσαρμοσμένο μοντέλο θεμάτων, για να ενημερώσει μια διαδραστική απεικόνιση βασισμένη στο διαδίκτυο. Μπορεί να χρησιμοποιηθεί από ένα σημειωματάριο IPython, αλλά υπάρχει η δυνατότητα αποθήκευσης σε ένα αυτόνομο αρχείο html για πιο εύκολη κοινή χρήση (“pyLDAvis 2.1.2 documentation”). Στην εικόνα 1 και στην εικόνα 2 φαίνονται δύο ενδεικτικά παραδείγματα μοντέλων οπτικοποίησης του pyLDAvis.



Εικόνα 1: Παράδειγμα οπτικοποίησης μοντέλου LDA, με pyLDAvis



Εικόνα 2: Παράδειγμα οπτικής απεικόνισης θεμάτων με pyLDAvis

Κάθε φυσαλίδα της γραφικής παράστασης που βρίσκεται στην αριστερή μεριά αντιπροσωπεύει ένα θέμα. Όσο μεγαλύτερη είναι η φούσκα τόσο πιο διαδεδομένο είναι το συγκεκριμένο θέμα, δηλαδή τόσο μεγαλύτερο είναι το ποσοστό του αριθμού των tweets για αυτό το θέμα. Ανάλογα με το πόσο επιτυχημένο είναι το μοντέλο, οι φυσαλίδες θα είναι αρκετά μεγάλες, μη επικαλυπτόμενες και διάσπαρτες σε κάθε τεταρτημόριο. Αντίθετα ένα μοντέλο με πολλά θέματα θα έχει μικρότερες φυσαλίδες που συχνά η μία θα επικαλύπτεται της άλλης και θα βρίσκονται συγκεντρωμένες σε ένα τεταρτημόριο (Azad, 2020). Οι μπλε γραμμές, στη δεξιά πλευρά, αντιπροσωπεύουν τη συνολική συχνότητα κάθε λέξης στο σώμα (corpus). Αν δεν επιλεγεί κάποιο θέμα εμφανίζονται οι μπλε γραμμές των λέξεων που συναντώνται πιο συχνά. Αν επιλεγθεί μια οποιαδήποτε λέξη θα δούμε μόνο τον κύκλο για το θέμα που περιέχει τη λέξη. Μία λέξη μπορεί να υπάρχει σε περισσότερα από ένα θέματα, τα οποία θα μεγαλώσουν σε αντίθεση με τις υπόλοιπες φυσαλίδες οι οποίες θα μικρύνουν. Αν επιλεγθεί ένα θέμα θα εμφανιστεί στα δεξιά μια λίστα με τους πιο συχνούς όρους που βρίσκονται σ' αυτό το θέμα, ενώ παράλληλα θα εμφανιστεί μια κόκκινη γραμμή. Οι κόκκινες γραμμές δείχνουν τον εκτιμώμενο αριθμό των φορών που δημιουργήθηκε ένας όρος από το θέμα. Όσο μεγαλύτερη κόκκινη γραμμή έχει μια λέξη, αυτό συνεπάγεται ότι η συχνότητα εμφάνισης της στο συγκεκριμένο θέμα είναι μεγαλύτερη από' ότι στα υπόλοιπα θέματα, δηλαδή ότι η συγκεκριμένη λέξη χρησιμοποιήθηκε πιο συχνά από τα tweets του συγκεκριμένου θέματος (Tran, 2021).

Τέλος, στην δεξιά στήλη πάνω από τη λίστα με τις πιο συνηθισμένες λέξεις που αφορούν κάθε θέμα, υπάρχει μια μέτρηση λ . Το λ παίρνει τιμές από 0 έως 1. Ορίζεται ως η

συνάφεια του όρου w με το θέμα k , δίνοντας του μια παράμετρο βάρους λ . Το λ καθορίζει το βάρος που δίνεται στην πιθανότητα του όρου w στο θέμα k . Υπολογίζεται ως εξής:

$$\text{relevance}(\text{term } w \mid \text{topic } t) = \lambda * p(w \mid t) + (1 - \lambda) * p(w \mid t)/p(w)$$

Για $\lambda=1$, έχουμε ως αποτέλεσμα τη κατάταξη όρων κατά φθίνουσα σειρά της πιθανότητας για το συγκεκριμένο θέμα. Ενώ για $\lambda=0$ οι όροι κατατάσσονται μόνο με βάση το “υψος” τους. Το λ μπορεί να πάρει τιμές ενδιάμεσες του 1 και του 0. Όσο μειωνουμε τη τιμή του λ , εμφανίζονται κάποιες διαφορετικές λέξεις στη λίστα, οι οποίες είναι πιο περιγραφικές για το θέμα. Σύμφωνα με τους Kenneth Shirley και Sievert Carson (Sievert and Shirley 2014) έπειτα από διάφορες περιπτώσεις μελέτης και πειραμάτων κατέληξαν στο ότι οι όροι κατάταξης κατά φθίνουσα σειρά πιθανότητας είναι υποβέλτιστη για την ερμηνεία ενός θέματος. Με βάση αυτό το δεδομένο για να γίνει πιο εύκολη η ερμηνεία των θεμάτων, η τιμή του λ δεν ρυθμίστηκε στην τιμή 1 αλλά κάπου στο ενδιάμεσο μεταξύ του 0,4 και 0,6, όπου κρίθηκε ότι θα μπορούσε το κάθε θέμα να γίνει πιο εύκολα ερμηνεύσιμο.

2.5 Ανάλυση Δεδομένων στο Twitter

2.5.1 Ανάλυση Δεδομένων από το Twitter για τον Covid-19

Σε πρώτο στάδιο έγινε μια έρευνα σχετικά με προηγούμενα ερευνητικά άρθρα που μέχρι εκείνο το διάστημα είχαν ήδη δημοσιευθεί και αφορούν την πανδημία. Υπήρχε μια πληθώρα εργασιών που ήδη είχαν ολοκληρωθεί καθώς εκείνο το χρονικό διάστημα ήταν, και φυσικά ακόμα παραμένει, ένα πεδίο που απασχολεί ολόκληρη την επιστημονική κοινότητα με συνεχή έρευνες τόσο σε ιατρικό όσο και σε επίπεδο τεχνολογικό, ψυχολογικό, οικονομικό, καθώς το αντίκτυπο που έχει αφορά όλους τους τομείς της ζωής μας σε παγκόσμια κλίμακα. Η κατανόηση, η ανάλυση αντίστοιχων εργασιών, είναι μια διαδικασία που απαιτεί προσήλωση και προσοχή. Για την συγκεκριμένη εργασία μελετήθηκαν αρκετές έρευνες με στόχο την καλύτερη κατανόηση του θέματος, τον καθορισμό των πεδίων που κινείται η εργασία, τον στόχο της και την απόφαση των μεθόδων που εφαρμόζονται. Παρακάτω γίνεται μια αναφορά των ερευνών που μελετήθηκαν, αναφέροντας περιληπτικά τους στόχους και τις μεθόδους που χρησιμοποιήθηκαν.

Ιδιαίτερα ενδιαφέρουσα εργασία είναι του Naseem, Usman (Usman, 2021). Σκοπός της είναι η αξιολόγηση της ροής των πληροφοριών κατά τη διάρκεια της πανδημίας. Προσπαθεί να αναγνωρίσει τις θεματικές ενότητες που συζητούνται περισσότερο στο Twitter και τα συναισθήματα της κοινής γνώμης που αφορούν την πανδημία. Συλλέγει τα tweets χρησιμοποιώντας το Twitter API και στη συνέχεια γίνεται η επισήμανση κάθε tweet ανάλογα με το αν είναι θετικό, αρνητικό ή ουδέτερο, χρησιμοποιώντας τεχνικές ανίχνευσης συναισθήματος. Έπειτα για την εύρεση των θεμάτων χρησιμοποιείται ο LDA αλγόριθμος, ένας από τους πιο γνωστούς αλγορίθμους μοντελοποίησης θεμάτων. Παρουσιάζει ένα dataset μεγάλης κλίμακας που περιέχει 90.000 tweets και τα οποία έχουν χαρακτηριστεί ως προς το αν είναι θετικά, αρνητικά ή ουδέτερα.

Μια άλλη εργασία παρόμοια με τη προηγούμενη είναι των Gupta, R. K., A. Vishwanath και Y. Yang (Gupta et al., 2021) οι οποίοι παρουσιάζουν ένα μεγάλο, επισημασμένο ως προς το συναισθημα, σύνολο δεδομένων, συγκεντρωμένο από τις 28

Ιανουαρίου 2020 μέχρι τις 1 Ιανουαρίου 2021, με πάνω από 132 εκατομμύρια tweets. Χρησιμοποιείται ένα σύνολο από διαφορετικούς αλγόριθμους για την ανάλυση συναισθήματος και τον προσδιορισμό της έντασης του καθενός. Στη συνέχεια εφαρμόζεται ο LDA αλγόριθμος για την κατανόηση των θεμάτων. Το dataset είναι δημόσια διαθέσιμο.

Επίσης ενδιαφέρουσα εργασία είναι των Chandrasekaran, Ranganathan, et al (Chandrasekaran and Ranganathan, 2020). Σκοπός της εργασίας είναι η αναγνώριση των κυρίαρχων θεμάτων, τα συναισθήματα που σχετίζονται με αυτά και την αλλαγή τάσεων που σχετίζεται με την πανδημία. Συγκέντρωσαν ένα σύνολο από 13.9 εκατομμύρια αγγλικά tweets από τις 1 Ιανουαρίου 2020 μέχρι τις 9 Μαΐου 2020. Στη συνέχεια εφαρμόστηκε ο LDA αλγόριθμος για την εξαγωγή των πιο βασικών θεμάτων. Αμέσως μετά χρησιμοποιώντας το VADER έγινε η ανάλυση συναισθήματος και η εξέταση της τάσης των θεμάτων στο συγκεκριμένο χρονικό διάστημα.

Ένα άλλο paper των Ordun Catherine, Sanjay Purushotham, and Edward Raff (Ordun et al., 2020) αναφέρει πέντε διαφορετικές τεχνικές για την αξιολόγηση των θεμάτων, των πιο συχνών όρων, των χαρακτηριστικών και την ταχύτητα διάδοσης πληροφοριών που σχετίζονται με τον covid-19. Αρχικά χρησιμοποιείται μια τεχνική ελέγχου, η αντιστοίχιση προτύπων. Στη συνέχεια εφαρμόζεται ο LDA αλγόριθμος για την εξαγωγή θεμάτων. Έπειτα χρησιμοποιείται μια άλλη μέθοδος η UMAP για τον καλύτερο προσδιορισμό και την κατανόηση των θεμάτων. Αμέσως μετά υπολογίζεται ο χρόνος αναμετάδοσης των tweets, με στόχο την κατανόηση της ταχύτητας του χρόνου μετάδοσης της πληροφορίας. Τέλος γίνεται η οπτικοποίηση των χρηστών οι οποίοι συνδέονται μεταξύ τους μέσω των retweets.

Μια εξίσου ενδιαφέρουσα εργασία είναι της Rabindra Lamsal (Rabindra, 2021), το οποίο παρουσιάζει ένα σύνολο δεδομένων με πάνω από 310 εκατομμύρια αγγλικά tweets. Παράλληλα παρουσιάζεται και ένα ακόμα dataset, το οποίο περιέχει γεωγραφικές πληροφορίες. Στη συνέχεια για την ανάλυση συναισθήματος χρησιμοποιήθηκε το TextBlob, μία βιβλιοθήκη της Python. Τέλος στο dataset με τις γεωγραφικές πληροφορίες πραγματοποιήθηκε ανάλυση κοινωνικού δικτύου, δηλαδή η διερεύνηση κοινωνικών δομών, χρησιμοποιώντας τη σχέση μεταξύ χώρας και #. Τα δύο επιμέρους σύνολα δεδομένων είναι διαθέσιμα για χρήση.

Οι Man Hung, Evelyn Lauren και Eric S. Hon στην εργασία τους (Hung et al., 2020) εφάρμοσαν τεχνικές μηχανικής μάθησης για την ανάλυση tweets αποκλειστικά από τις Ηνωμένες Πολιτείες από τις 20 Μαρτίου μέχρι τις 19 Απριλίου του 2020. Χρησιμοποιώντας τον LDA αλγόριθμο έγινε η εξαγωγή των θεμάτων και στη συνέχεια με το VADER καθορίστηκε ποια tweets εκφράζουν θετικό, αρνητικό ή ουδέτερο συναίσθημα.

Μια διαφορετική προσέγγιση είναι των Abeer Abuzayed και Hend Al- Khalifa (Abuzayed and Khalifa, 2021) οι οποίοι πειραματίζονται με διαφορετικές τεχνικές μοντελοποίησης θεμάτων και συγκεκριμένα τον LDA, τον NMF και τον BERTopic. Δεν σχετίζεται με την ανάλυση tweets αναφορικά με την πανδημία, αλλά αναφέρει μια νέα τεχνική μοντελοποίησης θεμάτων, τον BERTopic. Ουσιαστικά συγκρίνει την απόδοση των τριών τεχνικών σε tweets γραμμένα στα αραβικά.

Επίσης μια διαφορετική προσέγγιση ανάλυσης tweets για covid-19 είναι η δημοσίευση των Regis Ebeling, Carlos Abel Cordova Saenz και Jefferson Nobre (Ebeling et al., 2021). Ανέπτυξαν μια μελέτη περίπτωσης που αφορά τον covid-19 στη Βραζιλία, διερευνώντας την πολιτική επιρροή των δύο κυρίαρχων πολιτικών ομάδων της Βραζιλίας,

όσον αφορά τα μέτρα κοινωνικής απομόνωσης. Συνδυάζοντας δύο τεχνικές μοντελοποίησης έγινε η προσπάθεια κατανόησης των θέσεων κάθε ομάδας. Οι τεχνικές μοντελοποίησης θεμάτων που εφαρμόστηκαν είναι ο LDA και ο BERTopic. Στη συνέχεια εφαρμόστηκαν τεχνικές ανάλυσης του δικτύου με σκοπό να προσδιοριστεί αν η πολιτική πόλωση επηρεάζει τη κάθε ομάδα.

Για την επιλογή των πιο κατάλληλων τεχνικών προεπεξεργασίας έγινε αντίστοιχη έρευνα άλλων δημοσιευμένων εργασιών, ώστε να γίνει καλύτερη κατανόηση τους και να αποφασιστεί ποιες τεχνικές είναι πιο κατάλληλες για αυτή την εργασία. Παρακάτω στο πίνακα 1 ακολουθεί μια αναφορά αντίστοιχων εργασιών και των τεχνικών προεπεξεργασίας που εφάρμοσαν.

Πίνακας 1: Τεχνικές προ-επεξεργασίας που εφαρμόστηκαν σε αντίστοιχες εργασίες

.Αρθρο	Τεχνικές Προεπεξεργασίας
"COVIDSenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis." Naseem, Usman <i>IEEE Transactions on Computational Social Systems</i> 8.4 (2021): 1003-1015	<ul style="list-style-type: none"> • αφαίρεση # (όχι τη συνοδευτική λέξη) • Μετατροπή από κεφαλαία σε μια μικρά γράμματα • word segmentation • Αφαίρεση των stopwords • Lemmatization • Αφαίρεση ειδικών χαρακτήρων, αριθμών και σημεία στίξης
"Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal infoveillance study." Chandrasekaran, Ranganathan <i>Journal of medical Internet research</i> 22.10 (2020): e22624.	<p>Για την προεπεξεργασία χρησιμοποιήθηκαν το Natural Language Toolkit (NLTK), (RegEx) και η βιβλιοθήκη gensim της Python. Επιπλέον:</p> <ul style="list-style-type: none"> • Αφαίρεση των stopwords • Αφαίρεση συμβόλων @ • Εξαγωγή URLs • Lemmatization
"Exploratory analysis of covid-19 tweets using topic modeling, umap, and digraphs." Ordun, Catherine, Sanjay Purushotham, and Edward Raff <i>arXiv preprint arXiv:2005.03082</i> (2020).	<ul style="list-style-type: none"> • Αφαίρεση αριθμών • Αφαίρεση λέξεων με λιγότερους από 3 χαρακτήρες • Εξαγωγή URLs • Αφαίρεση συμβόλων @ • Μετατροπή από κεφαλαία σε μια μικρά γράμματα • Stemming • Tokenization • Αφαίρεση λατινικών χαρακτήρων
"Design and analysis of a large-scale COVID-19 tweets dataset." Lamsal, Rabindra <i>applied intelligence</i> 51.5 (2021): 2790-2804.	<ul style="list-style-type: none"> • Αφαίρεση συμβόλων @, #: • Εξαγωγή URLs • Αφαίρεση κενών

"Social network analysis of COVID-19 sentiments: Application of artificial intelligence." Hung, Man <i>Journal of medical Internet research</i> 22.8 (2020): e22590.	<ul style="list-style-type: none"> • Εξαγωγή URLs • Αφαίρεση συμβόλων @, #: • Αφαίρεση των stopwords • Αφαίρεση λέξεων με λιγότερους από 3 χαρακτήρες • Lemmatization
"The effect of political polarization on social distance stances in the brazilian covid-19 scenario." Ebeling, Régis <i>Journal of Information and Data Management</i> 12.1 (2021).	<ul style="list-style-type: none"> • Αφαίρεση σημείων στίξης • Αφαίρεση ειδικών χαρακτήρων • Αφαίρεση συμβόλων #,@ • Εξαγωγή URLs • Αφαίρεση λέξεων με λιγότερους από 3 χαρακτήρες
"Fuzzy rule based unsupervised sentiment analysis from social media posts." Vashishtha, Srishti, and Seba Susan. <i>Expert Systems with Applications</i> 138 (2019): 112834.	<ul style="list-style-type: none"> • Εξαγωγή URLs • Αφαίρεση συμβόλων @ • Αφαίρεση συμβόλων # • Αφαίρεση των stopwords • Αφαίρεση σημείων στίξης • Lemmatization

2.5.2 Το σύνολο δεδομένων

Το δεύτερο βασικό στάδιο της εργασίας ήταν η εύρεση ενός dataset, το οποίο θα αφορά την πανδημία και τη ζωή εκατομμυρίων ανθρώπων τη χρονική περίοδο της πρώτης φάσης της έξαρσης του ιού. Κατά τη διάρκεια μελέτης διαφόρων δημοσιευμένων εργασιών, μία που ξεχώρισε ήταν των Umair Qazi, Muhammad Imran και Ferda Ofli (Umair et al., 2020) με τίτλο “GeoCoV19: A Dataset of Hundreds of Millions of Multilingual COVID-19 Tweets with Location Information”. Η εργασία παρουσιάζει ένα σύνολο δεδομένων από το Twitter πολύ μεγάλης κλίμακας που σχετίζεται με τη συνεχιζόμενη πανδημία. Συγκεκριμένα αποτελείται από περισσότερα από 524 εκατομμύρια tweets τα οποία είναι γραμμένα σε πολλές διαφορετικές γλώσσες. Το σύνολο δεδομένων συλλέχθηκε σε μια περίοδο 90 ημερών από τις 1 Φεβρουαρίου έως τις 1 Μαΐου του 2020, στην πρώτη δηλαδή φάση της πανδημίας όπου πολλά πράγματα ήταν ακόμα άγνωστα και η κατάσταση πρωτόγνωρη.

Η συλλογή των δεδομένων έγινε μέσω του Twitter API, χρησιμοποιώντας λέξεις κλειδιά εκείνης της περιόδου. Στη συνέχεια χρησιμοποιώντας μια προσέγγιση που βασίζεται σε gazeteer τεχνικές και διάφορες άλλες μεθόδους έγινε ο γεωγραφικός προσδιορισμός των tweets. Στο τέλος η εργασία παρουσιάζει μια μεγάλη βάση δεδομένων η οποία περιέχει πληροφορίες για την τοποθεσία του χρήστη και το περιεχόμενο των tweets. Όσον αφορά τη γεωγραφική κάλυψη, το σύνολο των δεδομένων εκτείνεται σε 218 χώρες και ειδικότερα 47 χιλιάδες πόλεις. Οι πληροφορίες γεωγραφικής θέσης είναι πολύ χρήσιμες σε πολλές ερευνητικές εργασίες και ανάλογα με την ανάλυση των δεδομένων μπορούν να αξιοποιηθούν με διαφορετικό τρόπο. Τα tweets στο σύνολο δεδομένων προέρχονται από περισσότερους από 43 εκατομμύρια χρήστες σε 62 διαφορετικές γλώσσες και 491 εκατομμύρια tweets με μία τουλάχιστον γεωγραφική πληροφορία.

Η συγκεκριμένη εργασία λόγω του ότι περιέχει tweets για κάθε χώρα θεωρήθηκε ιδιαίτερα ενδιαφέρουσα καθώς το dataset της μπορεί να αξιοποιηθεί με πάρα πολλούς τρόπους, όπως είναι για παράδειγμα η κατανόηση του τρόπου με τον οποίο διαφορετικές κοινωνίες αντιμετωπίζουν την πρωτοφανή κρίση της πανδημίας (Umair et al., 2020). Η γεωγραφική ανάλυση και ο προσδιορισμός της θέσης καθιστά το συγκεκριμένο dataset ιδανικό στην περίπτωση που κάποιος επιθυμεί να εστιάσει την έρευνα του σε μία συγκεκριμένη χώρα. Ο παραπάνω λόγος καθόρισε την επιλογή του dataset.

Το dataset της εργασίας αυτής, προσφέρεται δημόσια για ερευνητική και μη κερδοσκοπική χρήση. Βρίσκεται διαθέσιμο στον παρακάτω σύνδεσμο <https://crisisNLP.qcri.org/covid19>. Ωστόσο τηρώντας τις πολιτικές αναδιανομής του Twitter δεν παρέχει το πλήρες περιεχόμενο των tweets, αλλά μόνο το tweet id, το user id και επιπλέον πληροφορίες σχετικές με τη γεωγραφική θέση. Συγκεκριμένα είναι διαθέσιμα αρχεία με τα tweets γραμμένα στα αγγλικά και τα tweets γραμμένα σε όλες τις γλώσσες. Τα αρχεία είναι ταξινομημένα ανά ημέρα και περιέχουν το tweet id και το used id. Τα αρχεία είναι της μορφής .tsv. Παράλληλα αντίστοιχα είναι τα αρχεία που περιλαμβάνουν τις γεωγραφικές πληροφορίες των tweets σε όλες τις γλώσσες και τα αρχεία με τις γεωγραφικές πληροφορίες των αγγλικών tweets. Επίσης είναι ταξινομημένα ανά ημέρα. Τα αρχεία με τις γεωγραφικές πληροφορίες είναι της μορφής `{tweet_id: xx, user_id: xx, created_at: xx, geo_source: xx, user_location: {one json}, geo: {one json}, tweet_locations: [array of jsons], place: {one json}}`

Για την διεξαγωγή της συγκεκριμένης εργασίας επιλέχθηκε το Ηνωμένο Βασίλειο. Οι λόγοι που οδήγησαν σ' αυτήν την επιλογή ήταν αρχικά η γλώσσα, καθώς τα αγγλικά είναι παγκόσμια γλώσσα, στη συνέχεια διότι το δείγμα των αγγλικών tweets ήταν πολύ μεγάλο ενώ σε άλλες γλώσσες, όπως για παράδειγμα τα ελληνικά, το δείγμα θα ήταν αισθητά μειωμένο και ίσως να μην μπορούσαν να εξαχθούν ασφαλή συμπεράσματα. Παράλληλα το Ηνωμένο Βασίλειο κατά τη διάρκεια της πρώτης φάσης της πανδημίας, απασχόλησε τον διεθνή τύπο καθώς ήταν μια από τις χώρες που άργησαν να λάβουν περιοριστικά μέτρα, ενώ υπήρχαν εισηγήσεις για το αντίθετο. Συνεπώς την καθιστά ένα ενδιαφέρον δείγμα για έρευνα καθώς από την ανάλυση των tweets θα μπορούσαν να εξαχθούν συμπεράσματα σε σχέση με το κατά πόσο η κοινή γνώμη στο Ηνωμένο Βασίλειο επιθυμούσε την λήψη περιοριστικών μέτρων πιο έγκαιρα, ή σε ένα πιο γενικό πλαίσιο πληροφορίες για την πανδημία, την κατάσταση που επικράτησε εκείνο το διάστημα και τους χειρισμούς από πλευράς πολιτικής διαχείρισης.

3. Μεθοδολογία

Στο συγκεκριμένο κεφάλαιο παρουσιάζονται τα στάδια της συλλογής των δεδομένων, της προ-επεξεργασίας τους και υλοποίησης των τεχνικών που εφαρμόστηκαν.

3.1 Συλλογή Δεδομένων

Τα δεδομένα που επιλέχθηκαν όπως αναφέρθηκε παραπάνω προέρχονται από την εργασία “GeoCoV19: A Dataset of Hundreds of Millions of Multilingual COVID-19 Tweets

with Location Information”. Για την συλλογή των δεδομένων ακολουθήθηκαν οι οδηγίες και τα βήματα που δόθηκαν από τους ίδιους τους συγγραφείς της εργασίας αυτής.

Αρχικά κατέβηκαν τα αρχεία με τα αγγλικά tweets. Το δεύτερο στάδιο είναι η ανάκτηση του περιεχομένου των αρχείων. Τα βήματα που ακολουθήθηκαν είναι τα εξής:

- Εγκατάσταση της python και της Pycharm
- Εγκατάσταση της βιβλιοθήκης twarc

Πριν την χρήση του εργαλείου twarc, απαιτείται η αποστολή ενός αιτήματος στο Twitter API. Το Twitter τα τελευταία χρόνια έχει δυσκολέψει τη χορήγηση άδειας πρόσβασης στο API και απαιτεί κάποια συγκεκριμένα βήματα. Σε πρώτη φάση δημιουργήθηκε ένα προφίλ στο Twitter. Στη συνέχεια στην ιστοσελίδα <https://developer.twitter.com/en/apply-for-access> ξεκινάει η διαδικασία εγγραφής και αίτησης. Αρχικά υπήρξε προσπάθεια απόκτησης άδειας για εκπαιδευτικούς λόγους, ωστόσο λόγω της αυστηροποίησης των πολιτικών του Twitter, απαιτήθηκαν στοιχεία που ήταν αδύνατο να βρεθούν και να χρησιμοποιηθούν. Γι’ αυτό το λόγο επιλέχθηκε η βασική (standard) χρήση. Η διαφορά τους έγκειται στο ότι το Twitter παρέχει περιορισμένη πρόσβαση στα δεδομένα του στη μία περίπτωση, ενώ αντίθετα για ακαδημαϊκή χρήση σου παρέχει τη δυνατότητα να έχεις πρόσβαση σε πολύ μεγαλύτερο αριθμό tweets, καθώς και κάποιους επιπλέον πόρους. Στην παρούσα εργασία δεν έπαιξε κάποιο ρόλο και γι’ αυτό τελικά προτιμήθηκε η βασική χρήση του API. Στη συνέχεια συμπληρώθηκαν τα προσωπικά στοιχεία, κάποιες φόρμες ερωτήσεων που αφορούν τους λόγους χρήσης, που θα χρησιμοποιηθούν, τι μέθοδοι θα χρησιμοποιηθούν, λεπτομέρειες σχετικές με την εργασία και διαβεβαιώσεις ότι δεν θα αξιοποιηθούν για εμπορικούς ή άλλους σκοπούς τα δεδομένα.

Μόλις το αίτημα έγινε δεκτό, στάλθηκαν τέσσερις κωδικοί. Οι τέσσερις κωδικοί υποβάλλονται στο περιβάλλον συστήματος με την παρακάτω μορφή (--consumer-key, --consumer-secret, --access-token, --access-token-secret). Μόλις γίνονται δεκτοί το twarc είναι έτοιμο για χρήση. Από τις εντολές του twarc χρησιμοποιείται το hydrate για να γίνει η ανάκτηση του περιεχομένου των tweets. Για κάθε ένα αρχείο εκτελείται η εντολή hydrate και επιστρέφεται ένα άλλο αρχείο σε json μορφή. Το μέγεθος των αρχείων ήταν ιδιαίτερα μεγάλο και χρειάστηκαν πολλές μέρες ώστε να ανακτηθούν όλα τα δεδομένα χρησιμοποιώντας αυτή τη διαδικασία. Ενδεικτικά πολλά από τα JSON αρχεία υπερέβαιναν τα 30GB

Τα κατεβασμένα δεδομένα στην αρχική τους μορφή είναι σε μορφή αρχείων κειμένου και για κάθε tweet συμπεριλαμβάνονται, εκτός από το κείμενο του tweet (που προσδιορίζεται με την επισήμανση "full_text") διάφορες άλλες πληροφορίες, όπως το id του (μοναδικό για κάθε tweet), το αν είναι απάντηση σε κάποιο άλλο tweet ή χρήστη, το αν είναι retweet, τυχόν πληροφορίες σχετικές με την τοποθεσία όπου γράφτηκε ή τοποθεσίες στις οποίες αναφέρεται, καθώς και πληροφορίες σχετικές με το χρήστη που το έγραψε όπως για παράδειγμα το id του, ο αριθμός των ακολούθων του και άλλα. Από αυτές τις πληροφορίες, για την παρούσα εργασία ενδιαφερόμαστε μόνο για το κείμενο του tweet, για την ημερομηνία του, καθώς και για το id του ώστε να ταυτοποιείται γρήγορα και αποτελεσματικά. Επομένως, με δεδομένο και τον τεράστιο όγκο του dataset, το πρώτο βήμα ήταν να κρατηθούν σε ξεχωριστά αρχεία τα συγκεκριμένα στοιχεία. Για κάθε ημερομηνία, δημιουργήθηκαν δύο τέτοια αρχεία (ένα για τα tweets και ένα για τα id τους). Τέλος, αξίζει

να σημειωθεί ότι κρίθηκε σκόπιμο να μην χρησιμοποιηθούν τα retweets, οπότε αυτά αφαιρέθηκαν.

Το δεύτερο βήμα ήταν η επιλογή μόνο εκείνων των tweets που αφορούν τη Μεγάλη Βρετανία, ώστε να επικεντρωθεί η ανάλυση εκεί. Το φιλτράρισμα έγινε με τον εξής τρόπο:

(1) Διάβασμα των αρχείων «en_geo_'date'.json» τα οποία περιέχουν τις γεωγραφικές πληροφορίες για κάθε tweet του dataset και φιλτράρισμα των id των tweets που είναι από τη Μ. Βρετανία.

(2) Με βάση αυτά τα id ανάκτηση εκείνων μόνο των tweet που έχουν σχέση με τη Μ. Βρετανία. Αποθήκευση τους σε ένα csv με τα πεδία «id» και «tweet», καθώς και σε ένα DataFrame, που είναι μια ευρέως χρησιμοποιούμενη δομή της Python και πιο συγκεκριμένα της βιβλιοθήκης Pandas. Το DataFrame αποθηκεύτηκε σε ένα αρχείο τύπου pickle με όνομα «final_unprocessed_data_df.pkl». Ο αριθμός των tweets που περιέχονται σε αυτό το DataFrame και τελικά χρησιμοποιήθηκαν στην ανάλυση είναι 10.715.563.

3.2 Προ-επεξεργασία των δεδομένων

Σε αυτό το στάδιο τα tweets που συλλέχθηκαν και φιλτραρίστηκαν έπρεπε να επεξεργαστούν με κατάλληλο τρόπο ώστε να αναλυθούν στη συνέχεια με τους αλγόριθμους topic modeling. Πιο συγκεκριμένα, πέρασαν από τα εξής στάδια:

- (1) Μετατροπή όλων των κεφαλαίων σε μικρά γράμματα
- (2) Αφαίρεση των υπερσυνδέσμων
- (3) Αφαίρεση του συμβόλου «@»
- (4) Αφαίρεση των αριθμών
- (5) Αφαίρεση των σημείων στίξης και του τονισμού
- (6) Αφαίρεση των πολύ μικρών (< 2 χαρακτήρες) ή πολύ μεγάλων (> 20 χαρακτήρες) λέξεων
- (7) Λημματοποίηση (lemmatization) των λέξεων
- (8) Αφαίρεση stopwords με τη χρήση λίστας. Πιο συγκεκριμένα, χρησιμοποιήθηκε η λίστα «stopwords» της βιβλιοθήκης nltk.

Το εξαγόμενο της συγκεκριμένης διαδικασίας είναι μια λίστα από λίστες. Κάθε εσωτερική λίστα αντιστοιχεί σε ένα tweet και αποτελείται από τις λέξεις που αυτό περιλαμβάνει, μετά την παραπάνω επεξεργασία (tokenization). Το εξαγόμενο αποθηκεύτηκε ως pickle αρχείο, με το όνομα «tokenized_tweets_stopw_rem.pkl».

3.3 Προσδιορισμός των θεμάτων

Ο προσδιορισμός των θεμάτων έγινε με τεχνικές topic modeling. Συνδυάστηκαν δύο τεχνικές, ο LDA και ο GSDMM, με σκοπό να επωφεληθούμε από τα προτερήματα και των δύο και να καλυφθούν οι αδυναμίες τους. Πιο συγκεκριμένα, όπως αναφέρθηκε και στο Κεφάλαιο 2, ο LDA είναι ο πιο κλασικός αλγόριθμος για topic modeling, όμως παρουσιάζει αδυναμία όταν χρησιμοποιείται για την ανάλυση μικρών κειμένων. Παρ' όλα αυτά, είναι μια λύση προσιτή από την άποψη των υπολογιστικών πόρων που χρειάζεται για την ανάλυση δεδομένων πολύ μεγάλου όγκου, ενώ είναι και αποδοτικός από άποψη χρόνου. Αντίθετα, ο GSDMM, λόγω της εξειδίκευσής του στην ανάλυση κειμένων μικρού μεγέθους (short text topic modeling), θα ήταν ιδανική λύση για τη συγκεκριμένη εργασία που ασχολείται με δεδομένα από το Twitter, όμως ήταν αδύνατο να εφαρμοστεί σε όλα τα δεδομένα με τους διαθέσιμους πόρους, καθώς απαιτεί περισσότερη μνήμη και χρόνο.

Για αυτούς τους λόγους, ο προσδιορισμός των θεμάτων έγινε ως εξής:

(1) Εφαρμογή του LDA στο σύνολο των δεδομένων (εξαγόμενο της προ-επεξεργασίας) ώστε να εξαχθούν κάποια αρχικά topics.

(2) Οπτικοποίηση των θεμάτων ώστε να κατανοηθούν με βάση τις λέξεις – κλειδιά που χαρακτηρίζουν κάθε θέμα. Ανάλυσή τους και επιλογή ορισμένων για περισσότερη ανάλυση. Πιο συγκεκριμένα, ανακτήθηκαν τα tweets που είχαν ως βασικό θέμα κάποιο από τα επιλεγμένα θέματα.

(3) Εφαρμογή του GSDMM σε κάθε ένα από τα σύνολα των tweets που ανακτήθηκαν ως αποτέλεσμα του βήματος (2), με σκοπό τον περαιτέρω διαχωρισμό τους σε θέματα με μεγαλύτερη συνοχή και ποιότητα.

3.4 Αρχικός διαχωρισμός θεμάτων με LDA

Για την υλοποίηση του LDA χρησιμοποιήθηκε η βιβλιοθήκη gensim της Python, και πιο συγκεκριμένα η κλάση LDAMulticore η οποία επιτρέπει τη γρήγορη εκτέλεση του αλγορίθμου με multithreading. Η εφαρμογή του LDA απαιτεί τον προσδιορισμό μιας κατάλληλης τιμής για την παράμετρο του αριθμού των θεμάτων (n). Επομένως η διαδικασία που ακολουθήθηκε περιλαμβάνει δύο στάδια, την πειραματική εφαρμογή του αλγορίθμου με διάφορες τιμές για τη συγκεκριμένη παράμετρο με σκοπό την εύρεση της πιο κατάλληλης και στη συνέχεια την εφαρμογή του αλγορίθμου με αυτήν την παραμετροποίηση.

Για το πρώτο στάδιο δοκιμάστηκαν μοντέλα με αριθμό θεμάτων από 3 έως 21 με βήμα 3 (δηλαδή $n=\{3, 6, 9, \dots, 21\}$). Η ποιότητα του κάθε μοντέλου μετρήθηκε με τη μετρική coherence c_v . Η υψηλότερη τιμή της ήταν για $n=15$, επομένως αυτός ήταν ο αριθμός των θεμάτων που τελικά επιλέχθηκε. Η συγκεκριμένη παραλλαγή της μετρικής coherence εφαρμόστηκε καθώς αποτελεί την default επιλογή στη μέθοδο `get_coherence` της κλάσης `CoherenceModel` της gensim, η οποία χρησιμοποιήθηκε για την αξιολόγηση. Στη συνέχεια εφαρμόστηκε ο LDA με $n=15$ στο σύνολο των tweets και μετρήθηκε η ποιότητα του μοντέλου με τις μετρικές coherence c_v και perplexity.

Η εφαρμογή του αλγορίθμου και στα δύο στάδια (στο πειραματικό και στο τελικό) απαιτεί τη δημιουργία ενός index που αντιστοιχεί κάθε λέξη που εμφανίζεται στο σύνολο των tweets με ένα μοναδικό id. Το index αυτό στη gensim ονομάζεται Dictionary. Από αυτό, αφαιρέθηκαν οι λέξεις που εμφανίζονται σε λιγότερο από τρία tweets. Επίσης, απαιτείται η αναπαράσταση του συνόλου των tweets με τη μορφή BoW (Bag of Words). Στα πλαίσια της gensim, το σύνολο δεδομένων στη μορφή αυτή ονομάζεται πλέον Corpus.

Ακολουθεί ο κώδικας της πειραματικής εφαρμογής του αλγορίθμου με σκοπό την εύρεση της πιο κατάλληλης τιμής για την παράμετρο n.

```
if __name__ == '__main__':  
    # get the data  
    input_data = 'tokenized_tweets_stopw_rem.pkl'  
    with open(input_data, "rb") as rb1:  
        tweets_tokenized = pickle.load(rb1)  
    id2word = corpora.Dictionary(tweets_tokenized)  
    # filter out words that appear in less than 3 documents  
    id2word.filter_extremes(no_below=3)  
    corpus = [id2word.doc2bow(tweet) for tweet in tweets_tokenized]  
    print(corpus[:5])  
    freeze_support()  
    coherence_values = []  
    for num_topics in range(3, 21, 3):  
        model = LdaMulticore(corpus=corpus, id2word=id2word,  
num_topics=num_topics)  
        coherencemodel = CoherenceModel(model=model, texts=tweets_tokenized,  
dictionary=id2word, coherence='c_v')  
        coherence_values.append(coherencemodel.get_coherence())
```

3.5 Οπτικοποίηση των αποτελεσμάτων

Για την οπτικοποίηση των αποτελεσμάτων χρησιμοποιήθηκε η βιβλιοθήκη pyLDAvis της Python. Η λύση αυτή δίνει τη δυνατότητα να αναπαρασταθεί με γραφικά η απόσταση μεταξύ των θεμάτων που παρήχθησαν για να κατανοηθεί το πόσο διακριτά είναι μεταξύ τους. Επίσης, για κάθε θέμα εμφανίζονται οι κυριότερες λέξεις που το περιγράφουν.

3.6 Εξαγωγή των τελικών θεμάτων με GSDMM

Αρχικά, για κάθε θέμα που παράχθηκε από τον LDA και επιλέχθηκε για περαιτέρω ανάλυση, ανακτούμε τα tweets τα οποία σύμφωνα με τον LDA έχουν το θέμα αυτό ως κυρίαρχο από το αποθηκευμένο αντικείμενο «tokenized_tweets_storw_rem.pkl» το οποίο περιέχει τα προ-επεξεργασμένα tweets. Στη συνέχεια, για τα tweets ενός συγκεκριμένου LDA θέματος, δημιουργήθηκε ένα Dictionary και ένα BoW corpus. Στη συνέχεια εφαρμόστηκε ο GSDMM με παραμέτρους $\alpha = 0,1$, $\beta = 0.1$, $n_iters = 20$ και με διαφορετικές τιμές για τον αριθμό των θεμάτων $K = \{5,6,7,8,9,10\}$. Η υλοποίηση έγινε με τη χρήση της βιβλιοθήκης `gsdmm`¹. Παρακάτω παρατίθεται ένα απόσπασμα του σχετικού κώδικα, για τα tweets του θέματος 14, τα οποία είχαν εξαχθεί σύμφωνα με τον LDA, για $K=10$.

```
dictionary = gensim.corpora.Dictionary(tweets_topic14)

# create variable containing length of dictionary/vocab

vocab_length = len(dictionary)

# create BOW dictionary, to be used later for the calculation of the
Coherence

bow_corpus = [dictionary.doc2bow(tweet) for tweet in tweets_topic14]

gsdmm_topic14 = MovieGroupProcess(K=10, alpha=0.1, beta=0.1,
n_iters=20)

y = gsdmm_topic14.fit(tweets_topic14, vocab_length)
```

Για κάθε GSDMM μοντέλο που παράχθηκε μετρήθηκε η ποιότητά του με τη μετρική coherence u_mass . Η συγκεκριμένη μετρική προτιμήθηκε από τη c_v γιατί σύμφωνα με τους ίδιους τους συγγραφείς της δημοσίευσης όπου προτάθηκε η c_v , αυτή τελικά παρουσιάζει κάποια προβλήματα² οπότε προτείνεται να μην χρησιμοποιείται. Με βάση αυτήν τη μέτρηση, για κάθε υποσύνολο από tweet (τα tweets ενός συγκεκριμένου LDA θέματος) επιλέχθηκε το καλύτερο GSDMM μοντέλο (δηλαδή το GSDMM μοντέλο με τη βέλτιστη τιμή για την παράμετρο K). Στη συνέχεια για κάθε τέτοιο υποσύνολο tweets εξήχθησαν σε μια λίστα τα θέματα των tweets σύμφωνα με τα αντίστοιχα GSDMM μοντέλα και αποθηκεύτηκαν.

3.7 Εξαγωγή διαγραμμάτων χρόνου

Για κάθε επιλεγμένο LDA θέμα, δημιουργήθηκε ένα Dataframe με τα πεδία «text» και «date», τα οποία περιέχουν για κάθε tweet το κείμενό του και την ημερομηνία της δημοσίευσής του αντίστοιχα. Χρησιμοποιώντας αυτά σε συνδυασμό με τις αποθηκευμένες αναθέσεις θέματος για κάθε tweet σύμφωνα με τον GSDMM, δημιουργήθηκε ένα διάγραμμα

¹ <https://github.com/rwalk/gsdmm>

² <https://palmetto.demos.dice-research.org/>

χρόνου για κάθε GSDMM θέμα. Κάθε τέτοιο διάγραμμα δείχνει τον αριθμό των tweets που μιλάνε για κάθε θέμα που παράχθηκε από ένα GSDMM μοντέλο.

Παρακάτω παρατίθεται ο κώδικας. Αρχικά, βλέπουμε το πώς ομαδοποιήθηκαν κατά ημερομηνία τα δεδομένα του DataFrame που περιείχε τις παραπάνω πληροφορίες (με το όνομα “subtopic_data”), καθώς και τα αποτελέσματα εκτέλεσης του κώδικα για τα tweets ενός συγκεκριμένου θέματος που είχε εξαχθεί μέσω του GSDMM.

```
daily_subtopic_data = subtopic_data.groupby(['date']).size()
```

```
date
2020-02-01    48
2020-02-02    84
2020-02-03    74
2020-02-04    60
2020-02-05    69
...
2020-04-27   1119
2020-04-28    464
2020-04-29    229
2020-04-30   1032
2020-05-01    935
Length: 91, dtype: int64
```

Με τη χρήση των παραπάνω δεδομένων, τα διαγράμματα χρόνου εξήχθησαν με τον παρακάτω κώδικα:

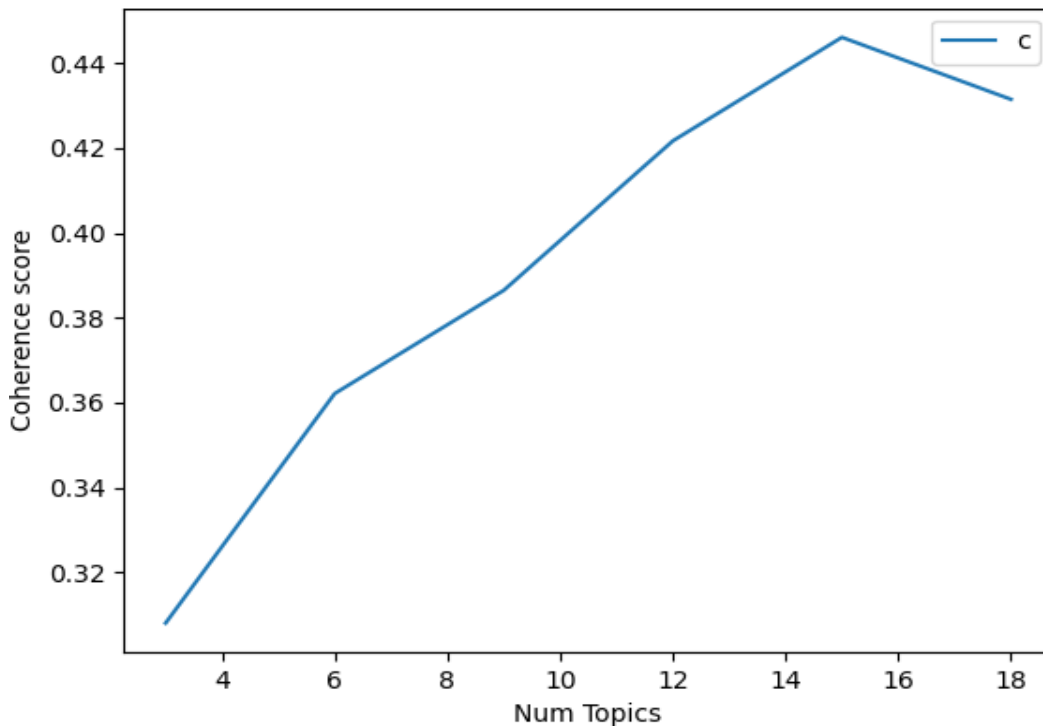
```
a = plt.figure(figsize=(20,10))
lnplot = sns.lineplot(data=daily_subtopic_data)
lnplot.set(ylabel='number of tweets')
labels = plt.xticks()[1]
plt.setp(labels, rotation=90)
```

4. Αποτελέσματα

Στο συγκεκριμένο κεφάλαιο παρουσιάζονται τα αποτελέσματα την ανάλυσης των δεδομένων. Συγκεκριμένα παρουσιάζονται τα αποτελέσματα των αλγόριθμων LDA και GSDMM που υλοποιήθηκαν, και στη συνέχεια παρουσιάζονται τα αποτελέσματα της χρονικής ανάλυσης.

4.1 Αποτελέσματα και παρατηρήσεις LDA

Στο συγκεκριμένο κεφάλαιο θα γίνει ανάλυση των αποτελεσμάτων της εργασίας. Αρχικά θα παρουσιαστούν τα αποτελέσματα που προέκυψαν από την υλοποίηση του LDA. Ο τελικός αριθμός των θεμάτων που επιλέχθηκε είναι δεκαπέντε διότι φάνηκε στην αξιολόγηση να έχει καλύτερο coherence score σε σχέση με τα υπόλοιπα. Το γράφημα αξιολόγησης του coherence score με βάση το οποίο έγινε η τελική επιλογή του αριθμού των θεμάτων φαίνεται στην εικόνα 3.

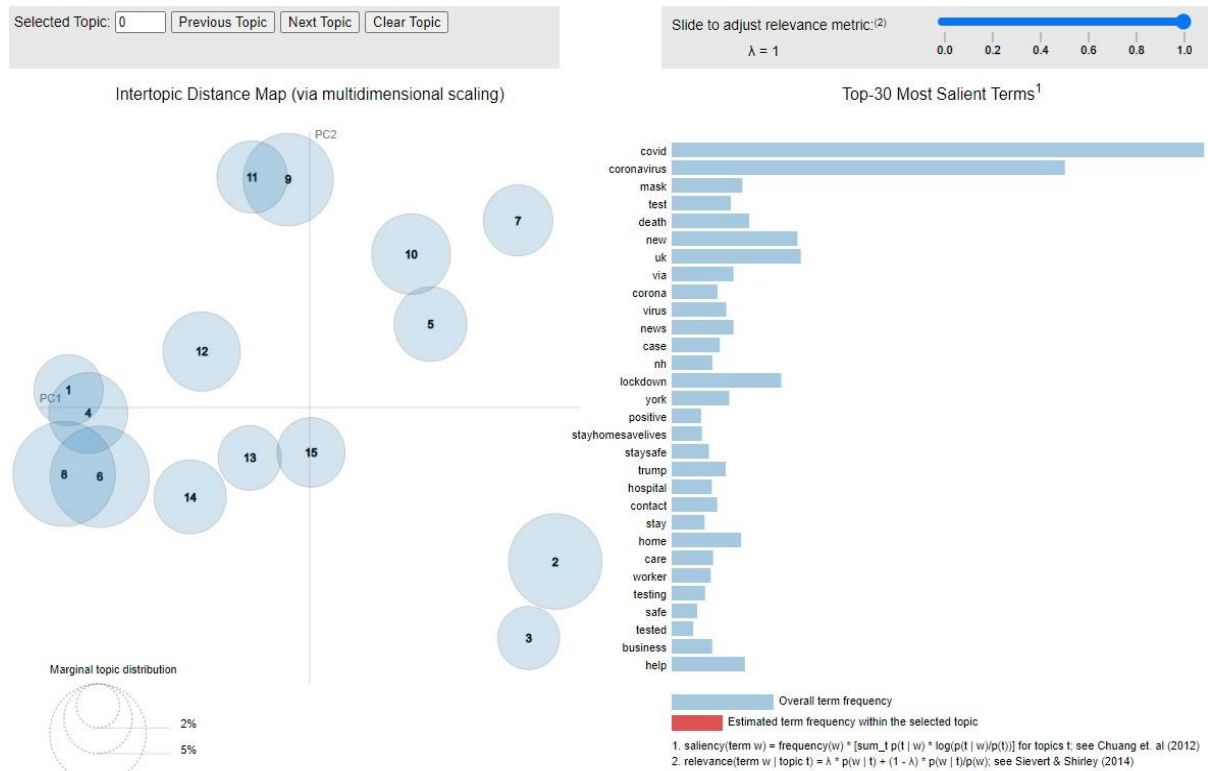


Εικόνα 3: Γράφημα αξιολόγησης των θεμάτων LDA, με βάση το coherence score.

Ο χρόνος υλοποίησης του αλγορίθμου ήταν 561.1017384529114 δευτερόλεπτα. Το coherence score ήταν 0.41861 και το perplexity -9.436801255673561. Γενικότερα δεν υπάρχει ένα συγκεκριμένο όριο πάνω από το οποίο θεωρείται καλή μια τιμή. Ωστόσο μπορεί να συγκριθεί με τα αποτελέσματα άλλων εργασιών και ερευνών. Αυτό που παρατηρήθηκε είναι ότι σε αρκετές έρευνες δεν αξιολογούνται τα αποτελέσματα της μοντελοποίησης των θεμάτων, όπως για παράδειγμα στις δημοσιεύσεις των (Chandrasekaran and Ranganathan, 2020), (Usman et al., 2021) και (Man et al., 2020). Στις συγκεκριμένες δημοσιεύσεις δεν χρησιμοποιούν κάποιο εργαλείο αξιολόγησης των μοντέλων που εφάρμοσαν για τη μοντελοποίηση θεμάτων, αν και χρησιμοποίησαν τον LDA αλγόριθμο. Κάποιες άλλες έρευνες ενώ αξιολόγησαν τα αποτελέσματα του μοντέλου που εφάρμοσαν δεν αναφέρουν τις τιμές του coherence αλλά δηλώνουν ότι επιλέχθηκε ο αριθμός των θεμάτων με βάση την υψηλότερη τιμή και σε συνδυασμό με την προσωπική τους εκτίμηση και εμπειρία. Για παράδειγμα οι έρευνες των (Koyel et al., 2020) και (Abeer and Hend Al-Khalifa, 2021). Υπάρχουν όμως δημοσιεύσεις οι οποίες χρησιμοποιούν το coherence score για την αξιολόγηση του μοντέλου όπως (Catherine et al., 2020) και (Gupta et al., 2021). Στην πρώτη εργασία χρησιμοποιώντας τον Gensim LDA MultiCore αλγόριθμο εξήχθησαν 20 θέματα και η τιμή του coherence ήταν 0.428 (Catherine et al., 2020), ενώ στη δεύτερη εργασία εξήχθησαν 10 θέματα με coherence score 0.575 (Gupta et al., 2021). Συνεπώς μετά από ενδελεχή έρευνα διαπιστώθηκε ότι οι περισσότεροι επιλέγουν να μην αξιολογήσουν τα αποτελέσματα της μοντελοποίησης θεμάτων με βάση το coherence score αλλά με βάση την εμπειρία τους και την προσωπική τους εκτίμηση. Ωστόσο σε αυτές που τα αξιολόγησαν οι

τιμές κυμαίνονται ανάμεσα στο 0.4 και 0.6. Με βάση αυτή τη σύγκριση και αυτά τα δεδομένα έγινε η αξιολόγηση του LDA στη συγκεκριμένη εργασία.

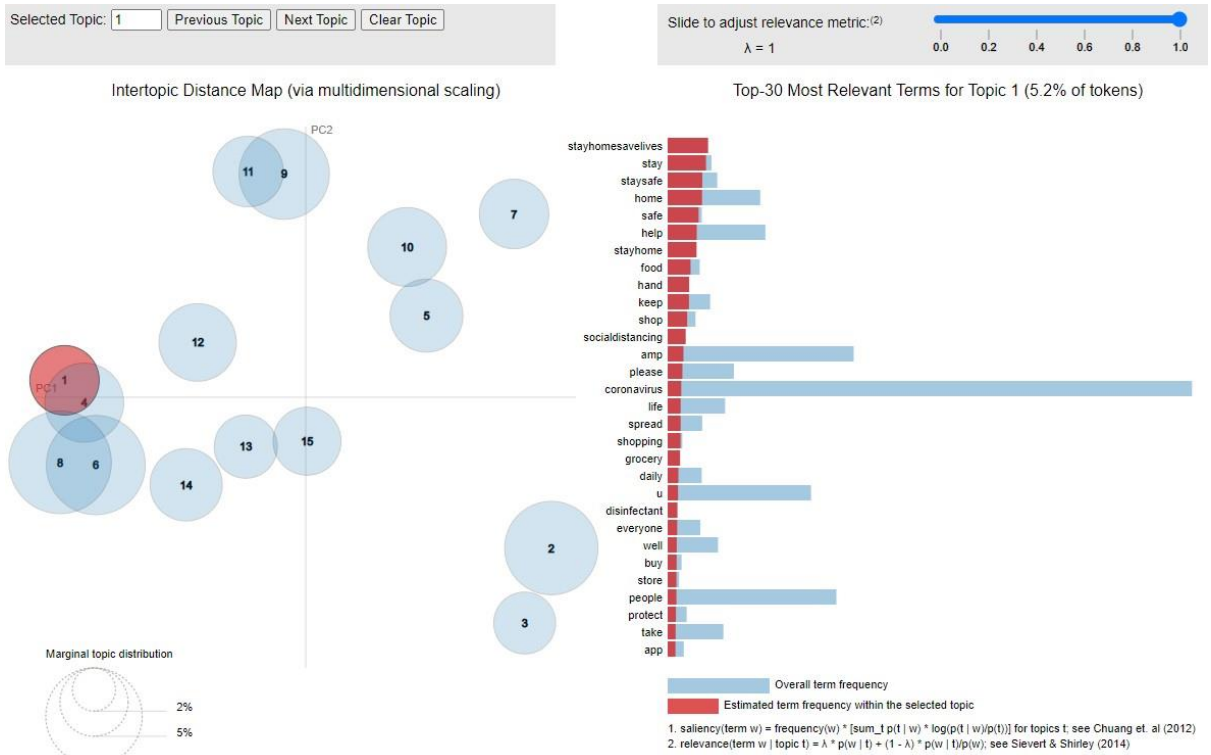
Ο συνολικός αριθμός των θεμάτων που εξήχθησαν μετά την υλοποίηση του LDA αλγορίθμου, όπως αναφέρθηκε παραπάνω είναι δεκαπέντε. Παρακάτω ακολουθεί η λίστα που απεικονίζει τα κορυφαία θέματα που εντοπίστηκαν στο σύνολο των δεδομένων και τις δέκα πιο αντιπροσωπευτικές λέξεις που σχετίζονται με κάθε ένα από αυτά. Παράλληλα για να γίνουν πιο κατανοητά, ακολουθεί η οπτική απεικόνιση όλων των θεμάτων στην εικόνα 4 και στη συνέχεια, για κάθε θέμα ξεχωριστά, όπως αυτή πραγματοποιήθηκε με το pyLDAvis.



Εικόνα 4: Οπτική απεικόνιση των αποτελεσμάτων του LDA.

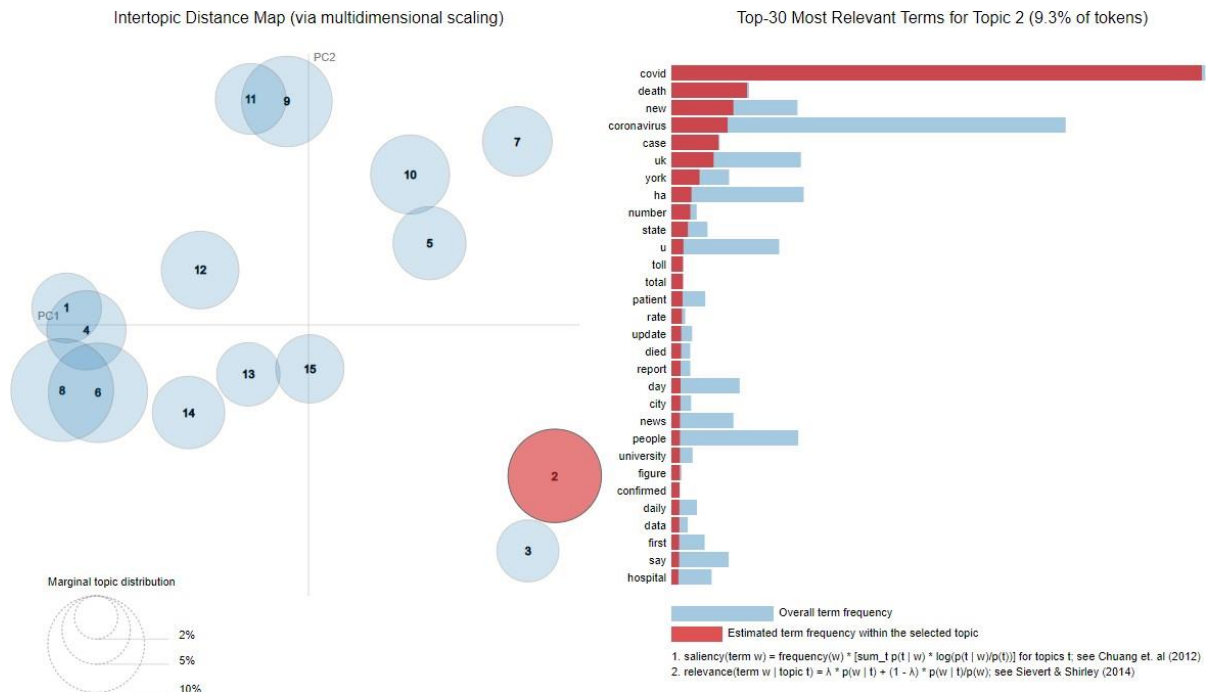
Στην αριστερή πλευρά φαίνονται οι φυσαλίδες, κάθε μία από τις οποίες αναπαριστά ένα θέμα. Στη δεξιά πλευρά βρίσκονται οι 30 πιο συνηθισμένες λέξεις σε όλο το δείγμα. Ακολουθεί η λίστα με τα θέματα ξεχωριστά.

- Topic 1: stayhomesavelives, stay, staysafe, home, safe, help, stayhome, food, hand, keep



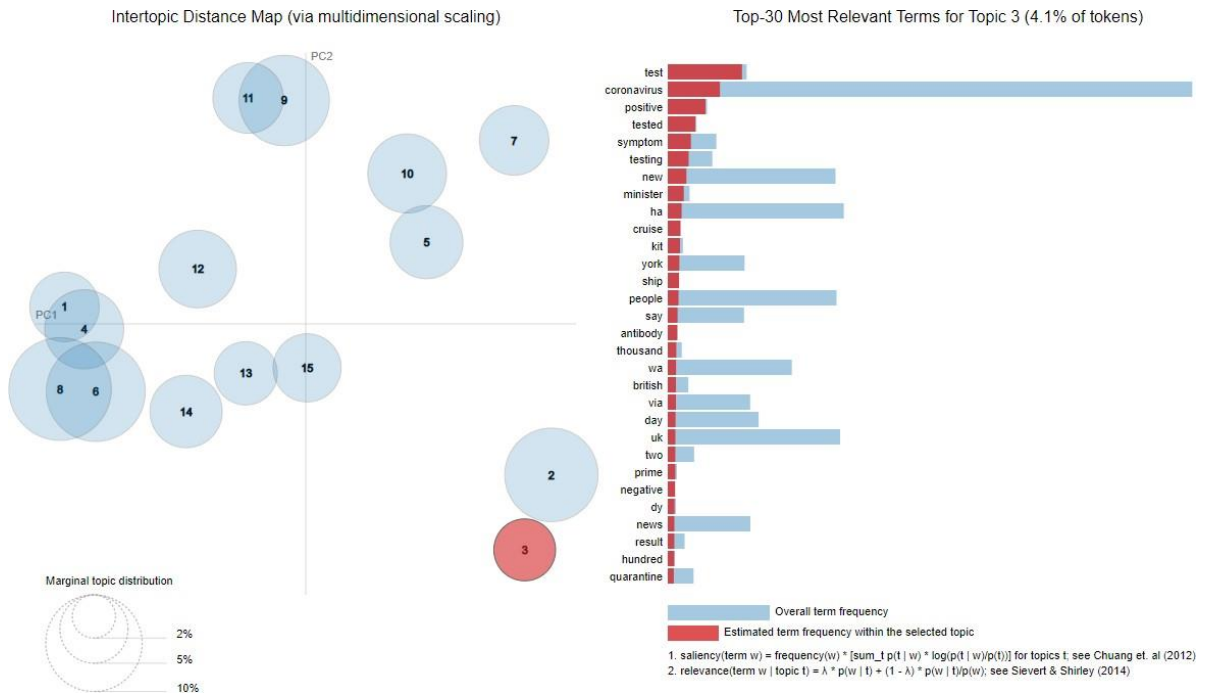
Εικόνα 5: Οπτική απεικόνιση του θέματος 1

- Topic 2: covid, death, new, coronavirus, case, uk, york, ha, number, state



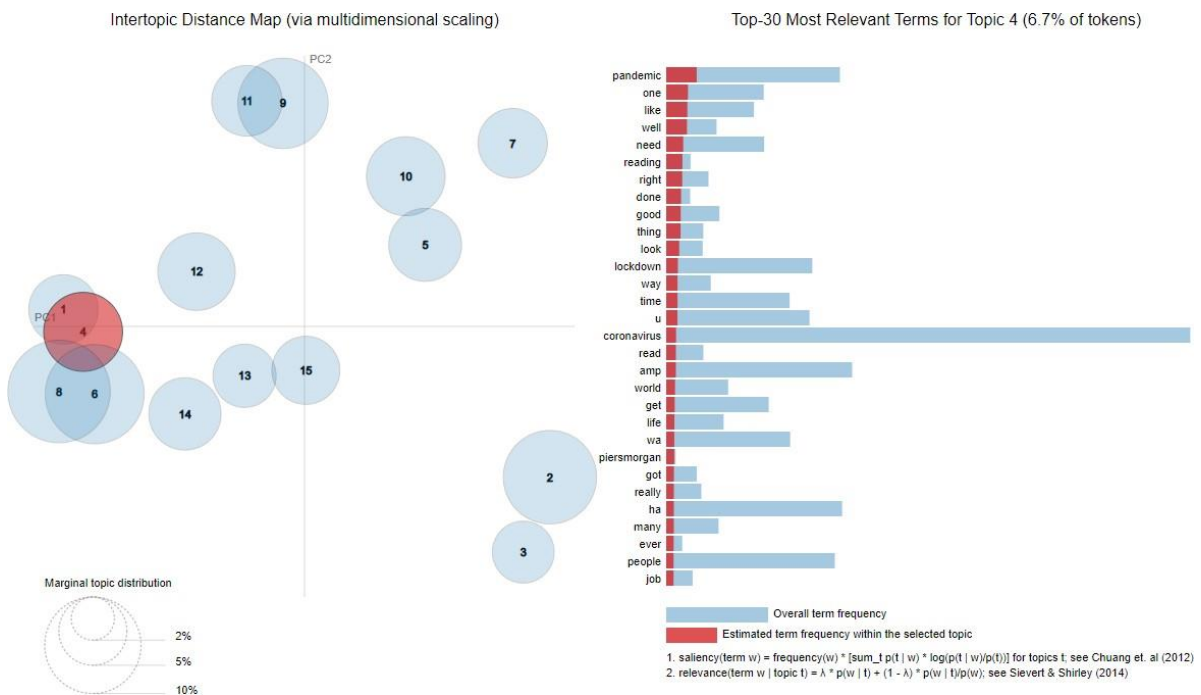
Εικόνα 6: Οπτική απεικόνιση του θέματος 2

- Topic 3: test, coronavirus, positive, tested, symptom, testing, new, minister, ha, cruise



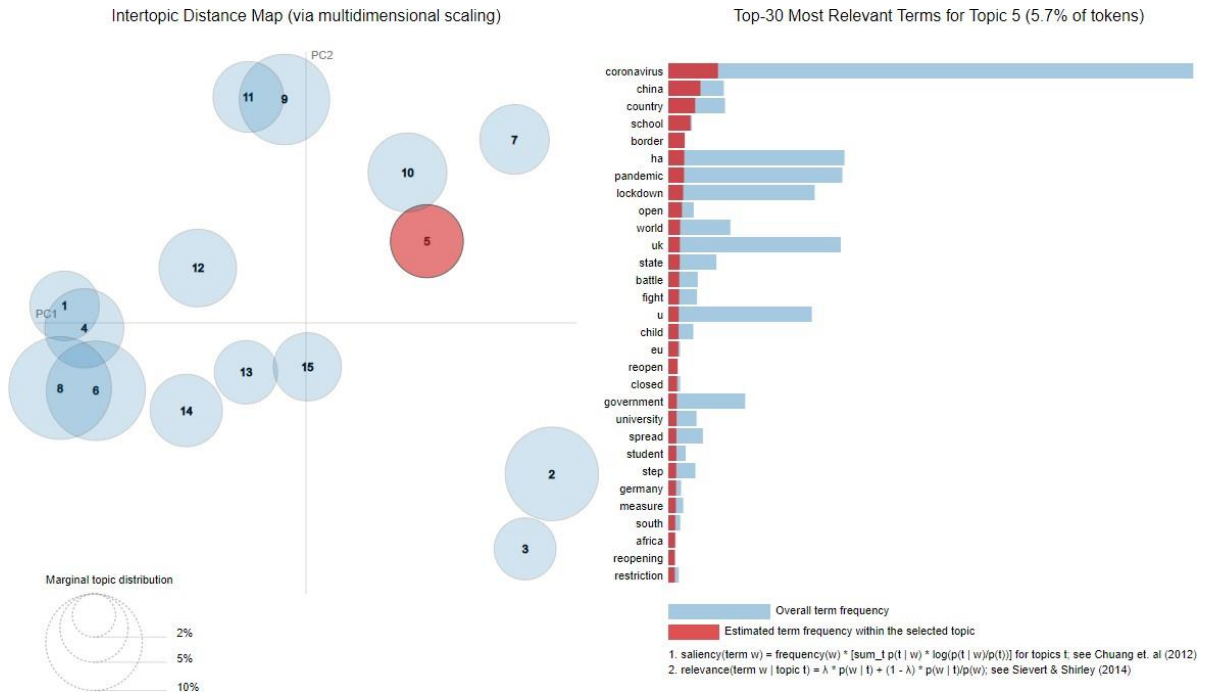
Εικόνα 7: Οπτική απεικόνιση του θέματος 3

- Topic 4: pandemic, one, like, well, need, reading, right, done, good, thing



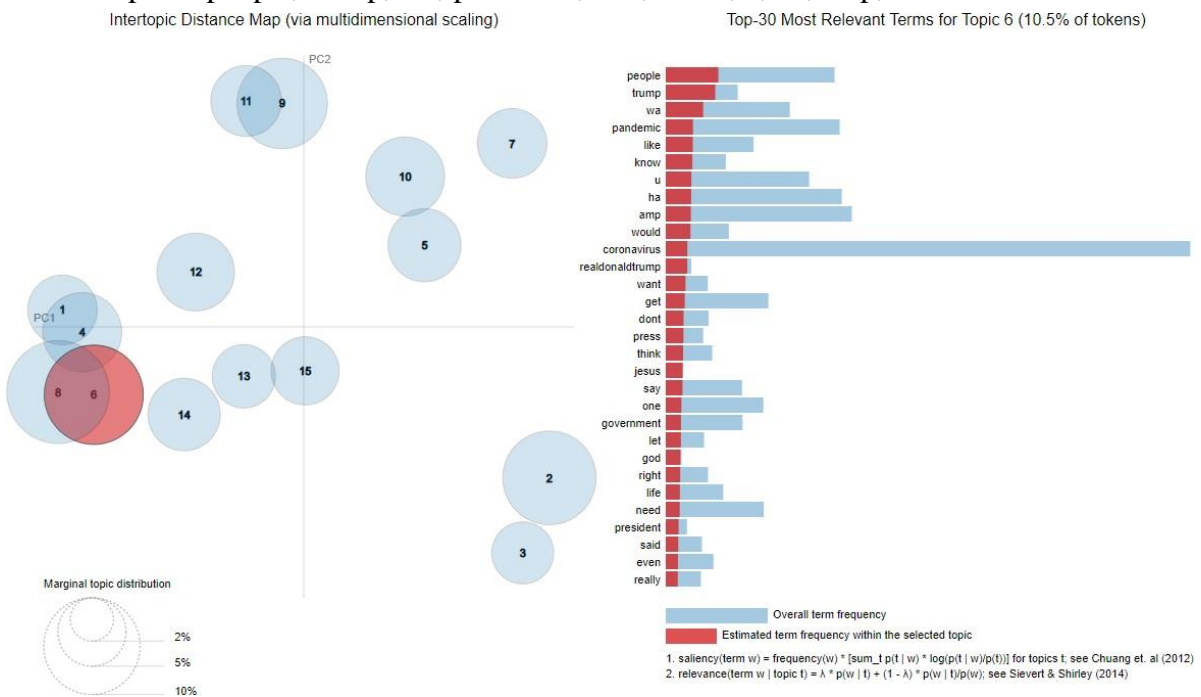
Εικόνα 8: Οπτική απεικόνιση του θέματος 4

- Topic 5: coronavirus, china, country, school, border, ha, pandemic, lockdown, open, world



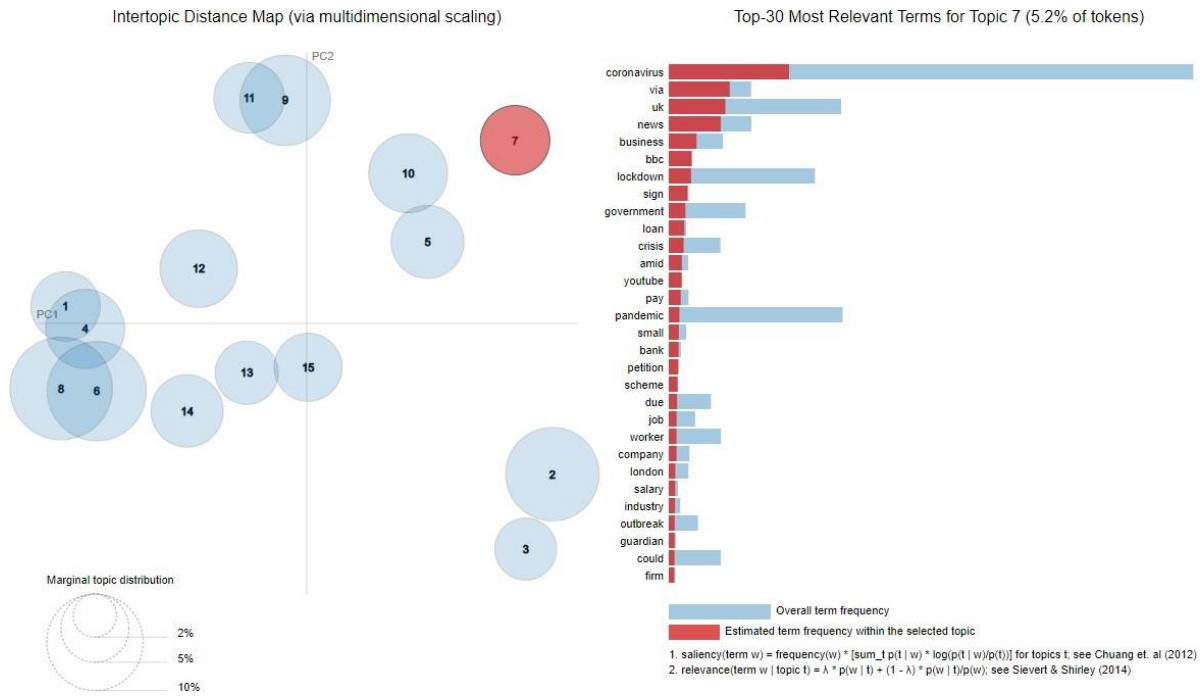
Εικόνα 9: Οπτική απεικόνιση του θέματος 5

- Topic 6: people, trump, wa, pandemic, like, know, u, ha, amp, would



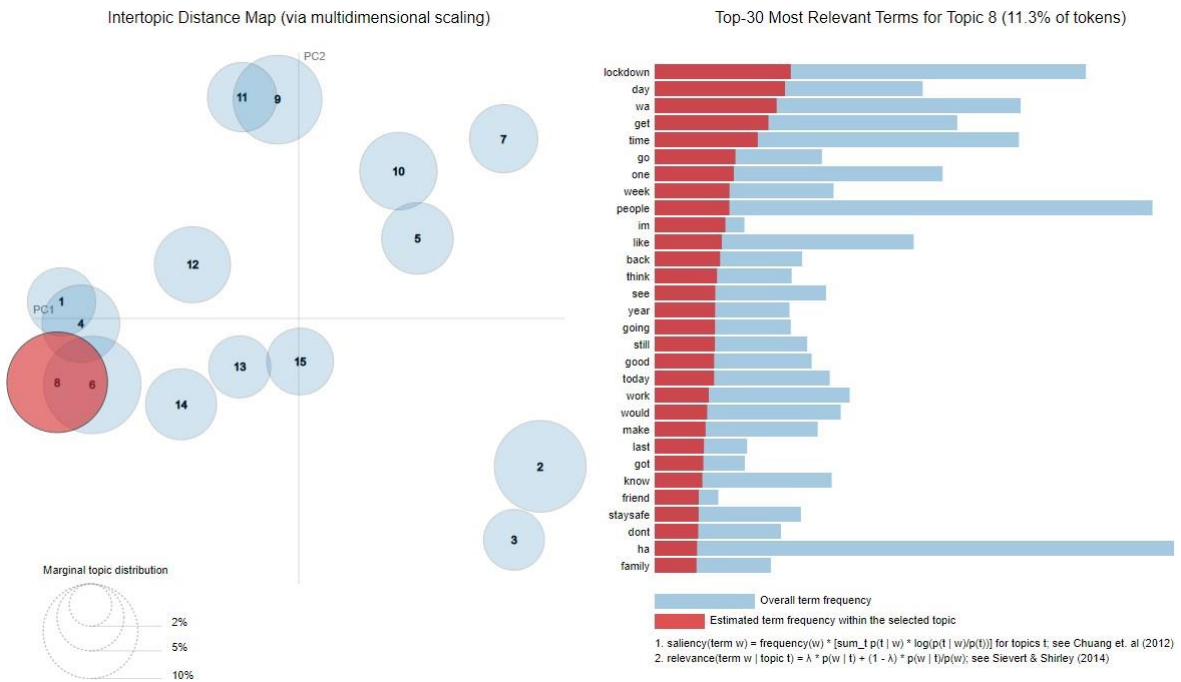
Εικόνα 10: Οπτική απεικόνιση του θέματος 6

- Topic 7: coronavirus, via, uk, news, business, bbc, lockdown, sign, government, loan



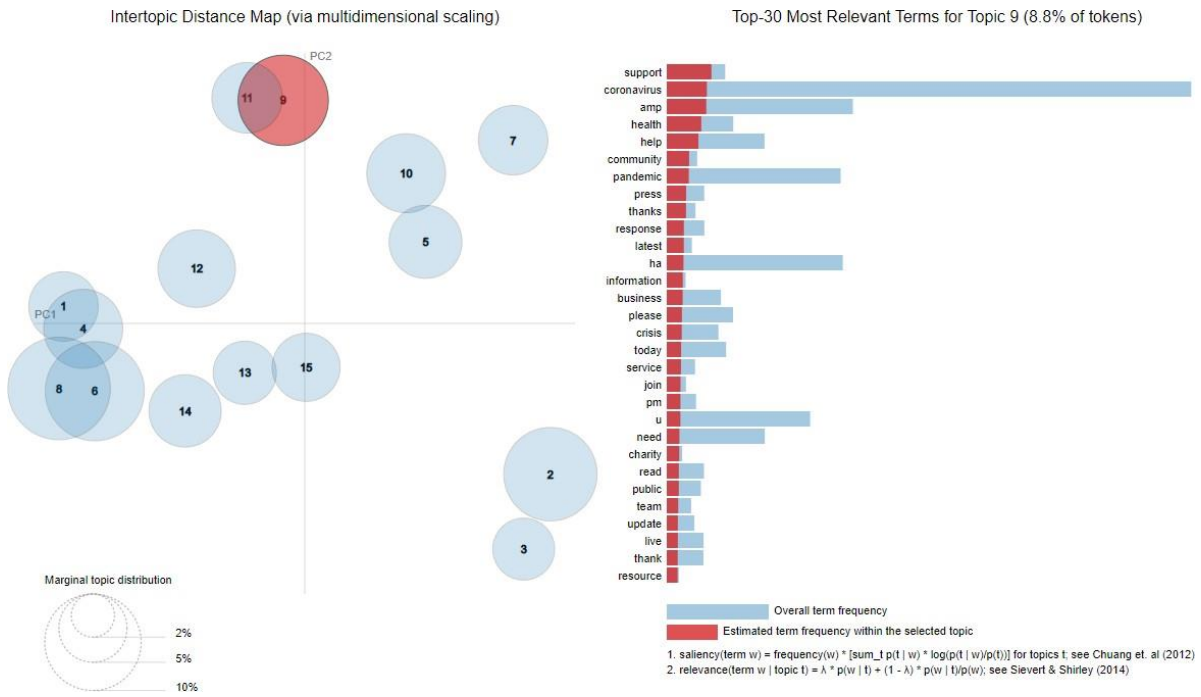
Εικόνα 11: Οπτική απεικόνιση του θέματος 7

- Topic 8: lockdown, day, wa, get, time, go, one, week, people, in



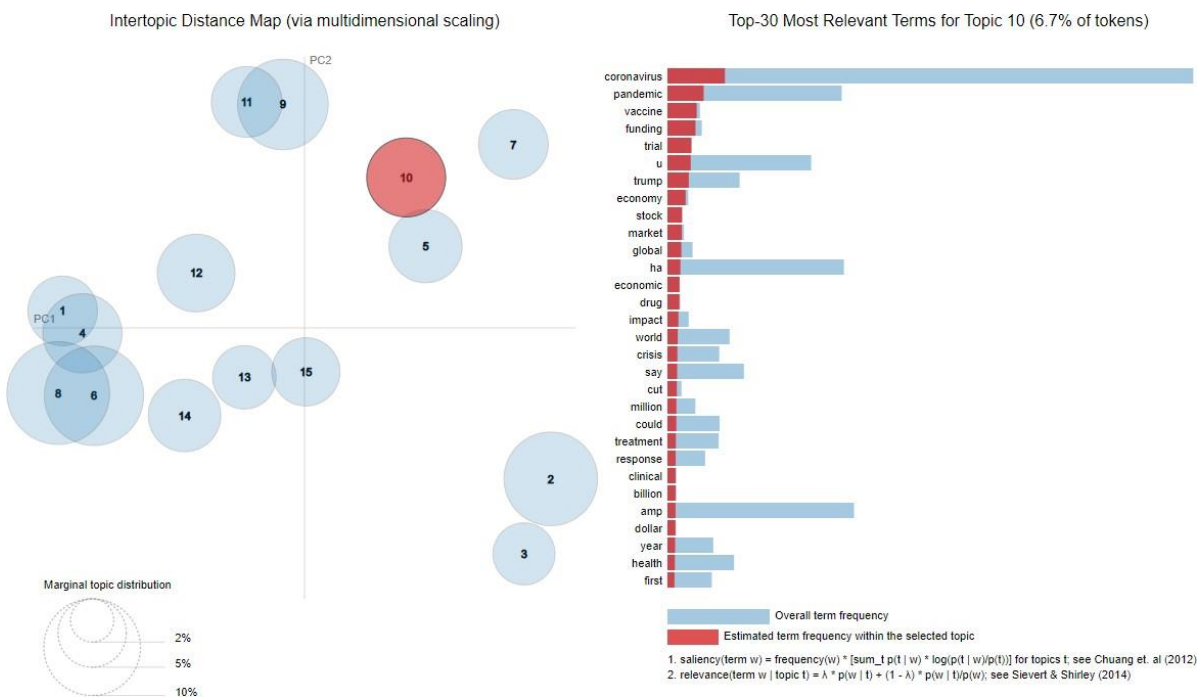
Εικόνα 12: Οπτική απεικόνιση του θέματος 8

- Topic 9: support, coronavirus, amp, health, help, community, pandemic, press, thanks, response



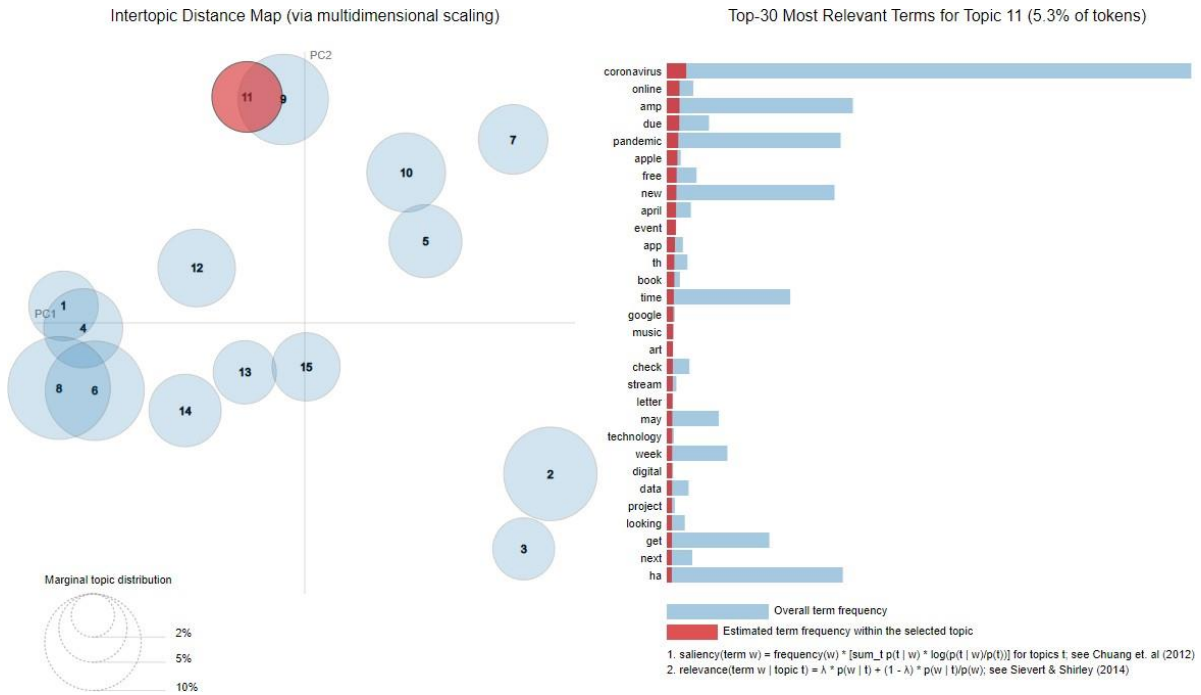
Εικόνα 13: Οπτική απεικόνιση του θέματος 9

- Topic 10: coronavirus, pandemic, vaccine, funding, trial, u, trump, economy, stock, market



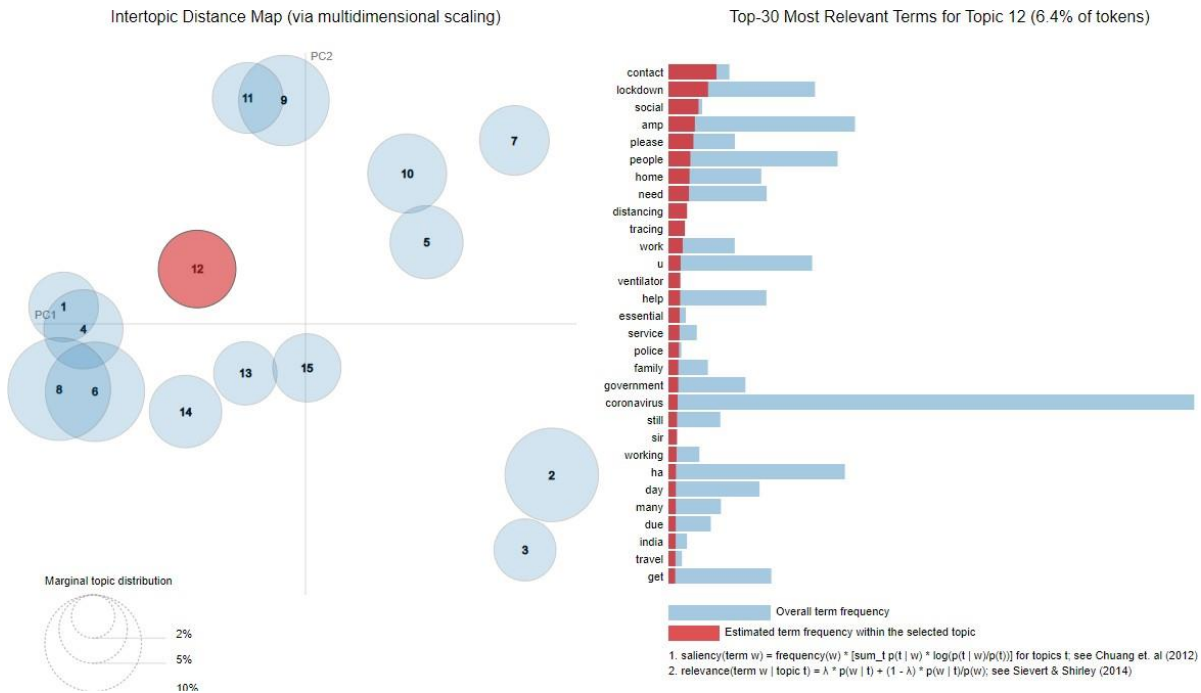
Εικόνα 14: Οπτική απεικόνιση του θέματος 10

- Topic 11: coronavirus, online, amp, due, pandemic, apple, free, new, april, event



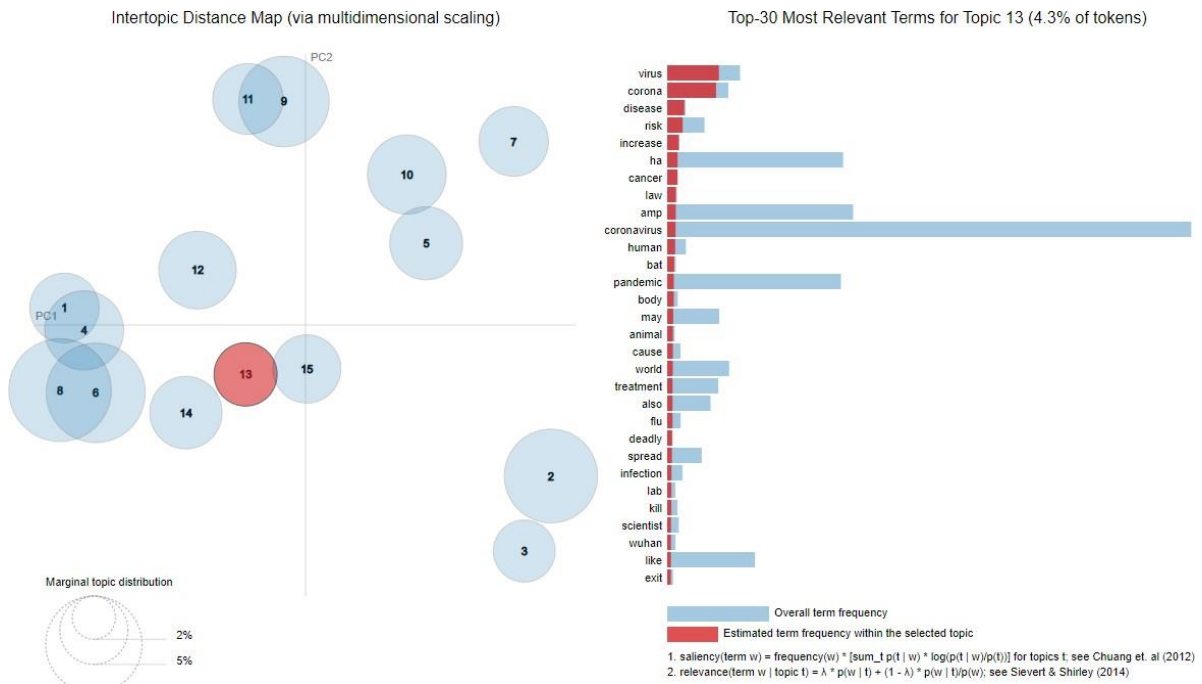
Εικόνα 15: Οπτική απεικόνιση του θέματος 11

- Topic 12: contact, lockdown, social, amp, please, people, home, need, distancing, tracing



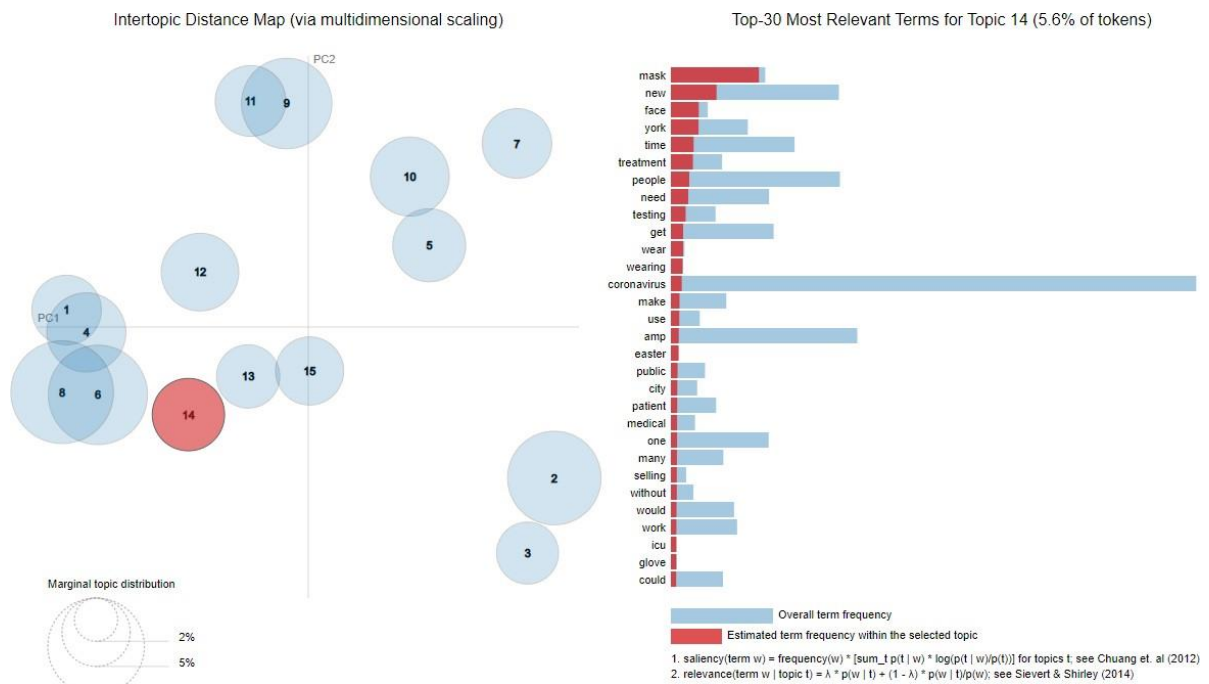
Εικόνα 16: Οπτική απεικόνιση του θέματος 12

- Topic 13: virus, corona, disease, risk, increase, ha, cancer, law, amp, coronavirus



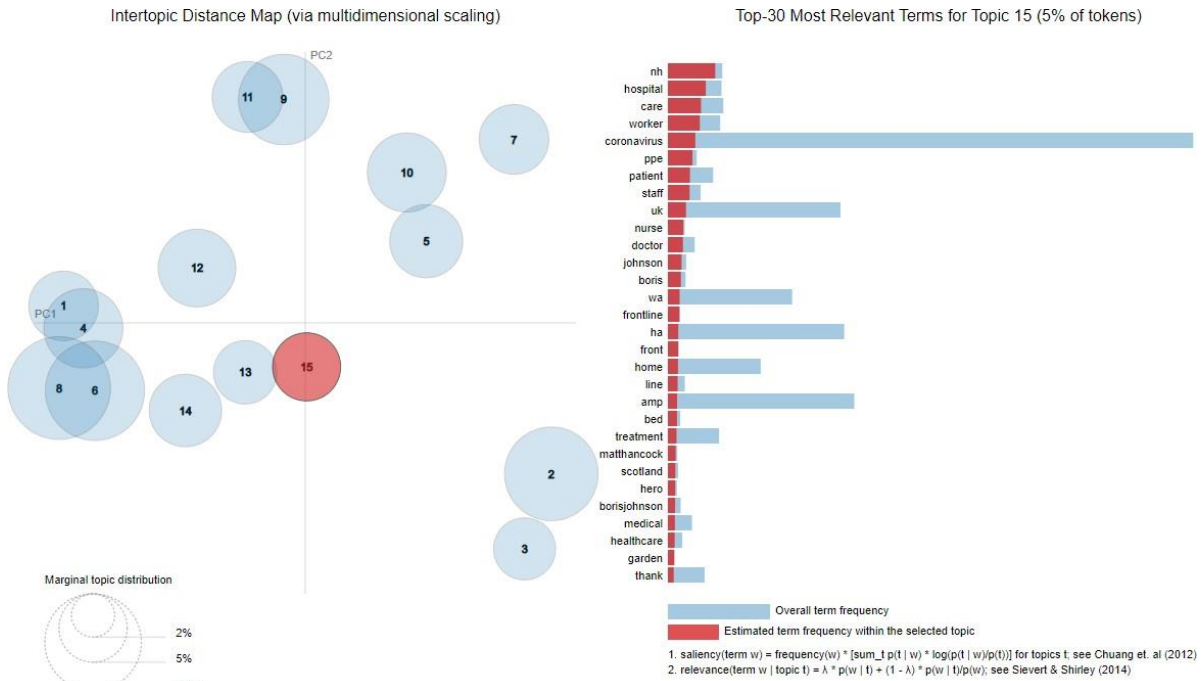
Εικόνα 17: Οπτική απεικόνιση του θέματος 13

- Topic 14: mask, new, face, york, time, treatment, people, need, testing, get



Εικόνα 18: Οπτική απεικόνιση του θέματος 14

- Topic 15: nh, hospital, care, worker, coronavirus, ppe, patient, staff, nurse



Εικόνα 19: Οπτική απεικόνιση του θέματος 15

Από τα αποτελέσματα του LDA, με βάση το οποίο έγινε η μοντελοποίηση θεμάτων, σε συνδυασμό με την οπτική τους απεικόνιση, χρησιμοποιώντας το pyLDAvis, μπορούν να εξαχθούν ορισμένα πολύ χρήσιμα συμπεράσματα. Αρχικά διακρίνοντας και παρατηρώντας το γράφημα που αποτυπώνει το σύνολο των θεμάτων μπορούμε να εξάγουμε πιο γενικές παρατηρήσεις. Αυτές είναι:

- Εύκολα παρατηρεί κάποιος ότι ορισμένες φυσαλίδες, οι οποίες αναπαριστούν τις θεματικές ενότητες, επικαλύπτονται μεταξύ τους. Αυτό υποδηλώνει ότι το πεδίο που πραγματεύονται, ορισμένα θέματα, είναι κοντά μεταξύ τους. Δηλαδή μπορεί να αφορούν ένα κοντινό θέμα και άρα να χρειάζονται παραπάνω ανάλυση, διότι δεν μπορεί να γίνουν εύκολα αντιληπτές οι διαφορές τους. Τα θέματα 1, 4, 6, 8 παρατηρείται ότι επικαλύπτονται σε μεγάλο βαθμό μεταξύ τους, όπως επίσης το θέμα 11 με το θέμα 9, σε πολύ μικρότερο βαθμό το θέμα 10 με το θέμα 5 και το θέμα 13 με το θέμα 15. Ο διαχωρισμός και η εύρεση διαφορών στα θέματα που επικαλύπτονται σε πιο έντονο βαθμό είναι πιο δύσκολος. Αυτό αποτυπώνεται, εκτός από το γράφημα, και από τις αντιπροσωπευτικές λέξεις των συγκεκριμένων θεμάτων, στην δεξιά στήλη του γραφήματος.
- Παρατηρείται ότι η απόσταση ορισμένων θεμάτων είναι αρκετά μεγάλη. Αυτό σημαίνει ότι είναι πιο ευδιάκριτες οι διαφορές μεταξύ αυτών των θεμάτων και ότι είναι πιθανό να αφορούν διαφορετικά πεδία. Για παράδειγμα το θέμα 7 φαίνεται να είναι πιο απομονωμένο από τα υπόλοιπα θέματα. Αυτό ίσως σημαίνει ότι είναι πιο εύκολη η διάκριση του από τα υπόλοιπα θέματα.

Με τη βοήθεια του pyLDAvis και με βάση τους πιο συχνούς όρους που εμφανίζονται σε κάθε θέμα, έγινε η προσπάθεια ερμηνείας των θεμάτων. Η συγκεκριμένη διαδικασία απαιτεί κριτική σκέψη και αντιληπτική ικανότητα, καθώς η ερμηνεία των θεμάτων στη

συγκεκριμένη φάση είναι πιο υποκειμενική. Παρακάτω αναλύεται η λογική με την οποία έγινε η ερμηνεία των θεμάτων.

- Στο θέμα 1 παρατηρείται ότι οι δέκα πιο συχνοί όροι που εμφανίζονται σχετίζονται με το μένουμε σπίτι (stay home). Ειδικότερα οι όροι με την λέξη stay (stayhome, staysafe, stayhomesavelives κτλ) παρατηρείται ότι εμφανίζονται κατα κύριο λόγο μόνο στο συγκεκριμένο θέμα. Πειράζοντας τις τιμές του λ στην δεξιά στήλη, αντί για 1 0.8 ή 0.6, εμφανίζονται κάποιες επιπλέον λέξεις όπως για παράδειγμα το “water”, “healthy”, “bleach” οι οποίες σε συνδυασμό με τις υπόλοιπες λέξεις διαμορφώνουν μια πρώτη αντίληψη σε σχέση με το τι αφορά το πρώτο θέμα. Παράλληλα παρατηρούνται και κάποιες άλλες λέξεις που αφορούν την αγορά προϊόντων, όπως για παράδειγμα “shop”, “food”, “store”. Το σύνολο των λέξεων αφορά όρους καθαριότητας, προστασίας, πλυσίματος και κατανάλωση. Συνεπώς ύστερα από την ανάγνωση των αποτελεσμάτων και με βάση καθαρά την υποκειμενική κρίση και την προσωπική αντίληψη το πρώτο θέμα αφορά τα μέτρα ατομικής προστασίας κατά του Covid-19 και θέματα κατανάλωσης αγαθών-προϊόντων.
- Στο θέμα 2 οι δύο πρώτοι όροι είναι το “covid” και “death”. Αυτές οι λέξεις από μόνες τους παραπέμπουν σε μία συγκεκριμένη κατεύθυνση. Παρατηρώντας τους υπόλοιπους όρους διακρίνεται μια συνοχή των λέξεων. Δηλαδή “new”, “case”, “confirmed”, “uk”, “number”. Όλοι οι όροι παραπέμπουν σε καταγραφές θανάτων από τον Covid-19. Άρα το θέμα δύο, με βάση τα παραπάνω που πάντα επαφίεται στην υποκειμενική κρίση, αφορά την καταγραφή και τον αριθμό των θανάτων που οφείλεται στον Covid-19.
- Το θέμα 3 παρατηρώντας τα αποτελέσματα του pyLDavis, βρίσκεται πολύ κοντά με το θέμα 2, χωρίς ωστόσο να επικαλύπτονται. Αυτό αρχικά προδιαθέτει στο ότι ίσως τα δύο αυτά θέματα έχουν ορισμένα κοινά και με κάποιο τρόπο να σχετίζονται μεταξύ τους. Οι πιο συχνοί όροι που εμφανίζονται σχετίζονται με το “test” και “testing”. Συνοδεύονται από όρους όπως “positive” “negative” “symptom” “kit” οι οποίες παραπέμπουν στην εξέταση και την ανίχνευση του ιού. Υπάρχουν όμως κάποιες επιπλέον λέξεις οι οποίες μοιάζουν σε πρώτη φάση κάπως παράταιρες και δεν υπάρχει μια συνοχή με τις υπόλοιπες. Για παράδειγμα “cruise”, “ship”, “board”, “crew”. Ενδεχομένως οι συγκεκριμένες λέξεις να αφορούν το πλαίσιο εξέτασης του ιού μεταξύ επιβατών και εργαζόμενου προσωπικού σε πλοία, αεροπλάνα και πιο συγκεκριμένα ένα συμβάν που εκείνο το διάστημα ήταν επίκαιρο και αφορούσε την ανίχνευση, διασπορά του ιού μέσω ενός ταξιδιού. Το συγκεκριμένο θέμα μπορεί να διερευνηθεί περαιτέρω ώστε να εξαχθούν πιο ασφαλή συμπεράσματα. Όσον αφορά τη σχέση του θέματος 3 με το θέμα 2, κρίνονται σχετικά καθώς αφορούν καταγραφές θανάτων και κρουσμάτων σε ένα πιο γενικό πλαίσιο.
- Το θέμα 4 όπως εύκολα διαπιστώνεται από το γράφημα είναι επικαλυπτόμενο με το θέμα 1, το θέμα 6 και το θέμα 8. Αυτό συνεπάγεται ότι η κατανόηση του και ο διαχωρισμός του από τα υπόλοιπα θέματα είναι πιο περίπλοκος και δύσκολος. Στο συγκεκριμένο θέμα εμφανίζονται όροι όπως για παράδειγμα “reading”, “article”, “worth”, “lockdown”, “pet” και “piercemorgan” ο οποίος είναι γνωστός Άγγλος παρουσιαστής, δημοσιογράφος και συγγραφέας. Με βάση καθαρά την υποκειμενική

κρίση και αντίληψη, φαίνεται το θέμα 4 να αφορά τη ζωή μέσα στην καραντίνα, δηλαδή συνήθειες και ασχολίες στο σπίτι όπως για παράδειγμα διάβασμα, παρακολούθηση μιας τηλεοπτικής εκπομπής, φροντίδα κατοικίδιου ζώου. Συμπερασματικά, το θέμα 4 αφορά την προσωπική ζωή και τα ενδιαφέροντα του ατόμου μέσα στην καραντίνα.

- Το θέμα 5 επικαλύπτεται σε ένα αρκετά μικρό βαθμό με το θέμα 10, άρα όπως και στα προηγούμενα θέματα υπάρχει η πιθανότητα αυτά τα δύο θέματα να σχετίζονται μεταξύ τους. Παρατηρώντας προσεκτικά τους όρους που εμφανίζονται διαπιστώνουμε ότι υπάρχουν λέξεις που συνδέονται νοηματικά μεταξύ τους όπως για παράδειγμα “teacher”, “student”, “university”, “school” οι οποίες σαφώς παραπέμπουν στην εκπαίδευση. Επιπλέον υπάρχουν όροι όπως “reopen”, “measure”, “lockdown”, “open”, “closed”, “spread” αλλά και “border”, “country”, “world”, “government”, “restriction” που παραπέμπουν σε ένα σύνολο από πολιτικές περιορισμού μετάδοσης του ιού. Το κλείσιμο σχολείων, πανεπιστημίων, περιορισμός μετακινήσεων μέσα από το κλείσιμο συνόρων, αποτελούσαν κύριες πολιτικές αποφάσεις των περισσότερων χωρών που είχαν σαν στόχο τον περιορισμό της πανδημίας. Σαν θέμα είναι πολύ ενδιαφέρον και αφορά τα μέτρα περιορισμού της πανδημίας σε επίπεδο πολιτικών αποφάσεων. Ενδεχομένως να χρειάζεται παραπάνω ανάλυση για να γίνει ένας διαχωρισμός μεταξύ των αποφάσεων που αφορούν την εκπαίδευση και της συνοριακής πολιτικής που εφαρμόστηκε.
- Το θέμα 6 επικαλύπτεται με το θέμα 4 και το θέμα 8. Οι πιο συνηθισμένοι όροι που εμφανίζονται στο θέμα 6 είναι “trump”, “realdonaldtrump”, “people”, “jesus”, “god”, “american”. Όλες οι παραπάνω λέξεις προφανώς έχουν σαν κοινό παρονομαστή τον Donald Trump πρώην πρόεδρο των Ηνωμένων Πολιτειών, ο οποίος θεωρείται και ως ένα από τις πιο αμφιλεγόμενες προσωπικότητες παγκοσμίως. Κάθε δήλωση του εκείνο το χρονικό διάστημα αποτελούσε είδηση και είναι ένα από τα πιο πολυσυζητημένα πολιτικά πρόσωπα εκείνης της περιόδου για τον τρόπο που διαχειρίστηκε την πανδημία, τις δηλώσεις του και τις ψευδείς μη επιστημονικά τεκμηριωμένες ειδήσεις που διακινούσε. Συνεπώς δεν αποτελεί έκπληξη που το θέμα 6 αφορά τον Donald Trump. Παράλληλα όμως παρεμβάλλεται το θρησκευτικό στοιχείο, καθώς εμφανίζονται όροι που αναφέρονται σε θέματα πίστης. Φαινομενικά μπορεί να φαίνονται δύο μη συσχετιζόμενα πεδία και άρα δύο ξεχωριστά θέματα, όμως η αλήθεια είναι ότι ο Donald Trump στις πολιτικές του ομιλίες και σε πολλές δηλώσεις του έχει χρησιμοποιήσει ιδιαίτερα τη θρησκεία, καθώς το κοινό που απευθύνεται έχει χαρακτηριστεί από όλους σχεδόν τους πολιτικούς αναλυτές, ως υπερσυντηρητικό. Με μια πρόχειρη αναζήτηση στο διαδίκτυο μπορεί κάποιος να βρει πολλές δηλώσεις του όπου συνδέει την πανδημία με τον Θεό (Thomas). Συνεπώς δεν είναι τόσο ασύνδετα μεταξύ τους. Το θέμα 6 αναφέρεται στον Donald Trump και στο θρησκευτικό συναίσθημα.
- Το θέμα 7 βρίσκεται αρκετά απομονωμένο από τα υπόλοιπα θέματα. Παρατηρώντας τις πιο συχνές λέξεις, κάπως το νόημα παραπέμπει σε πιο οικονομικά θέματα. Πιο συγκεκριμένα λέξεις όπως “loan”, “business”, “sign”, “petition”, “bank”, “crisis”, “company”, “industry” οδηγούν προς μια συγκεκριμένη νοηματική κατεύθυνση η οποία αφορά περισσότερο τον οικονομικό τομέα.

- Το θέμα 8 επικαλύπτεται σε μεγάλο βαθμό με το θέμα 6, σε λίγο λιγότερο βαθμό με το θέμα 4 και σε αρκετά μικρότερο βαθμό με το θέμα 1. Οι όροι που εμφανίζονται εδώ δεν μπορούν εύκολα να ερμηνευθούν από μόνοι τους, καθώς σαν μεμονωμένες λέξεις νοηματικά δεν παραπέμπουν εύκολα σε μια συγκεκριμένη θεματική ενότητα. Ωστόσο αλλάζοντας τις τιμές του λ, εμφανίζονται κάποιες λέξεις όπως “happy”, “birthday”, “lovely”, “wonderful”, “feeling” οι οποίες παραπέμπουν ενδεχομένως σε πιο συναισθηματικές καταστάσεις. Σε γενικότερο πλαίσιο το θέμα 8 είναι πιο αόριστο και γενικό και κάπως δύσκολο στο να προσδιοριστεί με σαφήνεια.
- Το θέμα 9 επικαλύπτεται με το θέμα 11. Παρατηρώντας τους πιο συχνούς όρους γίνεται αντιληπτό ότι αφορά την ευρεία έννοια της κοινότητας. Πιο συγκεκριμένα λέξεις όπως “support”, “community”, “help”, “information”, “charity”, “service”, “advice”, “guidance” “webinar” υποδεικνύουν δράσεις υποστήριξης στο πλαίσιο της κοινότητας. Επειδή η περίοδος της καραντίνας ήταν μια ιδιαίτερα επώδυνη κατάσταση, πέρα από το υγειονομικό κόστος, η κοινωνική απομόνωση στα πλαίσια της καραντίνας επέφερε βαρύ πλήγμα στην ψυχολογία όλων. Ιδιαίτερο πλήγμα δέχτηκαν διαφορές κοινωνικές δράσεις οι οποίες αποτελούσαν μεγάλη βοήθεια και στήριξη για πολλούς ανθρώπους. Όμως μέσα από τη νέα αυτή πραγματικότητα οι δράσεις αυτές βρήκαν νέες μορφές εκδήλωσης κυρίως με τη χρήση του διαδικτύου. Με αυτόν τον τρόπο στηρίχθηκαν ψυχολογικά και υλικά άνθρωποι ευάλωτοι οι οποίοι έμειναν ιδιαίτερα εκτεθειμένοι λόγω της πανδημίας. Το θέμα 9 αφορά ακριβώς αυτό το κομμάτι, δηλαδή τις συλλογικές δράσεις στο πλαίσιο της κοινότητας.
- Το θέμα 10 όπως αναφέρθηκε παραπάνω, επικαλύπτεται σε μικρό βαθμό με το θέμα 5 το οποίο αφορά τα μέτρα περιορισμού της πανδημίας σε επίπεδο πολιτικών αποφάσεων. Στο συγκεκριμένο θέμα εμφανίζονται όροι όπως “vaccine”, αλλά και άλλες λέξεις οι οποίες παραπέμπουν σε πιο χρηματοοικονομικά πεδία, όπως για παράδειγμα “funding”, “economy”, “market”, “dollar”, “economic”, “stock”, “debt”, “unemployment”, “oil”. Ο συνδυασμός αυτών των λέξεων οδηγεί σε οικονομικά ζητήματα που αφορούν την οικονομία, τις αγορές, πιθανότατα σε συνδυασμό με τις εξαγωγές και τα αποθέματα αγαθών. Φυσικά είναι γνωστό πως αυτά τα ζητήματα ήταν καίριας σημασίας τη χρονική περίοδο που μελετάται, καθώς οι συνοριακοί περιορισμοί και η καραντίνα είχαν τεράστιο αντίκτυπο σε όλους τους τομείς της οικονομίας. Είναι επίσης γνωστό ότι όλες οι χώρες αναγκάστηκαν να πάρουν μέτρα ενίσχυσης των οικονομιών τους, χρηματοδοτώντας διάφορους νευραλγικούς τομείς με σκοπό να αποφευχθεί μια μεγάλης κλίμακας κατάρρευση. Παράλληλα ο όρος “vaccine” συναντάται πολύ έντονα σε αυτό το θέμα, και ενδεχομένως να συνδέεται με τη χρηματοδότηση και την οικονομική ενίσχυση που δέχτηκαν οι φαρμακευτικές εταιρείες με σκοπό την ανακάλυψη ενός εμβολίου που θα έδινε τέλος στην πανδημία. Συνοψίζοντας το συγκεκριμένο θέμα με βάση τα παραπάνω και με μια προσωπική εκτίμηση, αφορά τις οικονομικές πολιτικές και το εμβόλιο.
- Το θέμα 11 έχει ήδη αναφερθεί ότι επικαλύπτεται με το θέμα 9. Παρατηρώντας προσεκτικά μπορεί εύκολα κάποιος να διαπιστώσει ότι πρόκειται για ένα θέμα που αφορά την τεχνολογία. Λέξεις όπως “digital”, “online”, “technology”, “apple”, “google”, “stream” είναι φανερό ότι παραπέμπουν σε τεχνολογικά πεδία και πιο ειδικά στο ψηφιακό περιβάλλον. Παράλληλα εμφανίζονται όροι που αφορούν πιο

πολύ ενδιαφέροντα και δραστηριότητες όπως για παράδειγμα “event”, “book”, “music”, “art”, “project”, “game”, “sport”. Μια πολύ απλή εξήγηση του πως μπορούν να συνδέονται αυτοί οι όροι μεταξύ τους και με ποιον τρόπο μπορεί να γίνει η συσχέτιση τους είναι το γεγονός ότι πολλές δραστηριότητες και όλες οι εκδηλώσεις που ήταν προγραμματισμένες εκείνη τη χρονική περίοδο, αναστάλησαν ή ακυρώθηκαν. Συναυλίες, εκθέσεις, παρουσιάσεις, αθλητικές εκδηλώσεις δεν ήταν εφικτό να πραγματοποιηθούν λόγω των περιορισμών της πανδημίας. Ωστόσο χάρη στις νέες τεχνολογίες πολλά από αυτά πραγματοποιήθηκαν ψηφιακά. Καινούργιες εφαρμογές και ψηφιακές πλατφόρμες επέτρεψαν τη μεταφορά πολλών δραστηριοτήτων στο ψηφιακό περιβάλλον μέσω του διαδικτύου.

- Το θέμα 12 αποτελείται από όρους όπως “lockdown”, “distancing”, “contact”, “social”, “tracing”, “home”, “police”, “work”, “isolation”. Παρατηρώντας τις λέξεις εύκολα γίνονται συνειρμοί σε σχέση με το θέμα και παραπέμπουν στο περιοριστικό μέτρο μετακινήσεων που εφαρμόστηκε δηλαδή το lockdown. Είναι ένα περιοριστικό μέτρο που πάρθηκε από αρκετές χώρες, όπου το πλαίσιο εφαρμογής του ήταν λιγότερο ή περισσότερο αυστηρό.
- Το θέμα 13 με μία πρώτη ματιά αφορά ιατρικά θέματα. Πιο συγκεκριμένα οι όροι που συναντώνται είναι: “virus”, “corona”, “disease”, “cancer”, “flu”. Πρόκειται καθαρά για ασθένειες. Επιπλέον στο συγκεκριμένο θέμα εμφανίζονται όροι όπως “bat”, “animal”, “wuhan”, “lab” οι οποίες εύκολα παραπέμπουν στις θεωρίες προέλευσης του ιού. Προήλθε από κάποιο ζώο όπως οι νυχτερίδες, που στην συνέχεια μεταδόθηκε στον άνθρωπο, ή πρόκειται για αποτέλεσμα πειραμάτων στο εργαστήριο της Γιουχαν; Πολλά έχουν γραφτεί, πολλές οι υποθέσεις και ο ΠΟΥ έχει συντάξει διάφορες εκθέσεις οι οποίες αναφέρουν τα πιο πιθανά σενάρια. Συνοπτικά και ύστερα από προσωπική εκτίμηση το θέμα 13 αφορά την προέλευση του Covid-19.
- Το θέμα 14 αποτελείται από όρους όπως “mask”, “face”, “wearing”, “testing”, “gloves”. Για ακόμα μια φορά προκειται για μέτρα προστασίας από τον ιό. Ωστόσο στη συγκεκριμένη περίπτωση αφορά στο μέτρο της χρήσης μάσκας. Η χρήση μάσκας σε κλειστούς χώρους ακόμα και τώρα σε πολλές χώρες εφαρμόζεται καθώς είναι αποδεδειγμένο ότι περιορίζει τη διασπορά του ιού στην κοινότητα. Επίσης τους πρώτους μήνες της πανδημίας συνοδεύονταν από τη χρήση γαντιών που όμως στην πορεία δεν κρίθηκε απαραίτητο κυρίως λόγω της μη ορθής εφαρμογής τους.
- Το θέμα 15 επικαλύπτεται σε μικρό βαθμό με το θέμα 13. Οι όροι που το συνοδεύουν είναι πάλι ιατρικοί αλλά αφορούν και την πολιτική. Πιο συγκεκριμένα οι λέξεις που εμφανίζονται είναι “hospital”, “care”, “worker”, “staff”, “nurse”, “patient”, “healthcare”, “doctor”, οι οποίες προφανώς αφορούν το ιατρικό και νοσηλευτικό προσωπικό, τους ασθενείς, τη λειτουργία των νοσοκομείων και συνεπώς το σύστημα υγείας. Παράλληλα όμως υπάρχουν και λέξεις που αναφέρονται σε πρόσωπα όπως είναι ο πρωθυπουργός του Ηνωμένου Βασιλείου, Boris Johnson και ο πρώην υπουργός Υγείας και Κοινωνικής Μέριμνας, Matt Hancock. Σίγουρα πρόκειται για δύο πολυσυζητημένα πρόσωπα της επικαιρότητας, οι οποίοι ήταν καθαρά υπεύθυνοι για την λήψη πολιτικών αποφάσεων ώστε να αντιμετωπιστεί η πανδημία και να ενισχυθεί το σύστημα υγείας στο Ηνωμένο Βασίλειο. Συνεπώς το θέμα 15 αφορά το σύστημα υγείας σε συνδυασμό με την πολιτική.

Στη συνέχεια είναι σημαντικό να γίνει διάκριση των θεμάτων με βάση το ενδιαφέρον που παρουσιάζουν για την μελέτη και την ανάλυση της παρούσας εργασίας. Κάποια από τα παραπάνω θέματα παρουσιάζουν πιο έντονο ενδιαφέρον, κάποια λιγότερο και κάποια άλλα χρήζουν περαιτέρω ανάλυση και διευκρίνιση.

Αρχικά το θέμα 1, το θέμα 4, το θέμα 6 και το θέμα 8 είναι εύκολο να διαπιστώσει κάποιος ότι αφορά πιο πολύ προσωπικά το άτομο. Δηλαδή το θέμα 1 αφορά τα ατομικά μέτρα προστασίας κατά της πανδημίας, το θέμα 4 αφορά τα ατομικά ενδιαφέροντα και την ζωή κατά την διάρκεια της πανδημίας, το θέμα 6 την θρησκεία, και το θέμα 8 ενώ είναι πιο αόριστο παραπέμπει σε πιο προσωπικές και συναισθηματικές στιγμές. Από τα παραπάνω θέματα το θέμα 1 κρίνεται ότι είναι το πιο ενδιαφέρον καθώς περιλαμβάνει μέτρο ατομικής προστασίας, υγιεινής, καθαριότητας αλλά και μέτρα αποτροπής της μετάδοσης του ιού. Μπορούν να εξαχθούν συμπεράσματα ειδικά για την πρώτη περίοδο της πανδημίας, η οποία μελετάται, κατά την διάρκεια της οποίας οι επιστημονικές και ιατρικές γνώσεις ήταν λίγες, ο κόσμος πιο ανήσυχος και φοβισμένος. Τα υπολοιπα θέματα κρίθηκαν πως δεν παρουσιάζουν κάποιο ιδιαίτερο ενδιαφέρον για τα ερωτήματα και τους ερευνητικούς σκοπούς της συγκεκριμένης εργασίας. Γι' αυτό το λόγο δεν θα γίνει περαιτέρω ανάλυση τους.

Το θέμα 2 σχετίζεται με την καταγραφή κρουσμάτων και θανάτων που προκλήθηκαν από τον Covid-19 και αφορούν το Ηνωμένο Βασίλειο. Οι αριθμοί και τα στοιχεία είναι γενικότερα γνωστά και καταγεγραμμένα. Κρίθηκε ότι το θέμα 2 δεν προσφέρεται για περαιτέρω ανάλυση. Αντίθετα το θέμα 3 που βρίσκεται πολύ κοντά με το θέμα 2 παρουσιάζει ενδιαφέρον και μάλιστα μπορεί να διερευνηθεί παραπάνω, καθώς ενδέχεται να προκύπτουν υποομάδες θεμάτων. Το θέμα 3 αφορά την ανίχνευση του ιού και περιστατικά διασποράς του.

Το θέμα 5, το θέμα 10 αλλά και το θέμα 7 είναι θέματα με ιδιαίτερο ενδιαφέρον καθώς σχετίζονται με πεδία που η παρούσα εργασία θέλει να εστιάσει και να μελετήσει. Πιο ειδικά το θέμα 5 αφορά την εκπαίδευση, δηλαδή τα σχολεία, τα πανεπιστήμια, και τους περιορισμούς των μετακινήσεων μεταξύ των χωρών που τέθηκαν σε εφαρμογή. Το θέμα 10 σχετίζεται με τον εμβολιασμό αλλά και με οικονομικά θέματα που σχετίζονται με τις αγορές. Τέλος το θέμα 7 επίσης αφορά την οικονομία σε ένα πιο γενικό πλαίσιο. Οι περιορισμοί που τέθηκαν εξαιτίας της πανδημίας επηρέασαν εκτός από όλα τα άλλα, την οικονομία παγκόσμια, το βιοτικό επίπεδο εκατομμυρίων ανθρώπων. Από τα τρία θέματα, το μεγαλύτερο ενδιαφέρον παρουσιάζει το θέμα 5 και το θέμα 7, καθώς αφορούν την εκπαίδευση, και την οικονομία. Το θέμα 9 δεν επιλέχθηκε καθώς φαίνεται να μην αφορά αποκλειστικά το Ηνωμένο Βασίλειο, αντίθετα αφορά θέματα με παγκόσμια διάσταση. Αυτά τα δύο θέματα συνεπώς παρουσιάζουν εξαιρετικά μεγάλο ενδιαφέρον και χρειάζεται να γίνει περαιτέρω ανάλυση καθώς αφορούν την οικονομία, την εκπαίδευση και τις πολιτικές αποφάσεις που πάρθηκαν εκείνο το διάστημα και αφορούν αυτούς τους τομείς.

Το θέμα 9 και το θέμα 11 σχετίζονται σε ένα βαθμό μεταξύ τους καθώς αφορούν κυρίως τη ψηφιακή ζωή και το διαδίκτυο. Λόγω περιορισμών πολλά πράγματα που υπό κανονικές συνθήκες θα γινόντουσαν δια ζώσης δεν μπορούσαν να πραγματοποιηθούν. Το διαδίκτυο και η τεχνολογία κλήθηκαν να καλύψουν το κενό είτε για πιο κοινωνικούς σκοπούς, είτε για πιο ψυχαγωγικούς. Το θέμα 9 αφορά τη χρήση του διαδικτύου για να καλυφθούν κοινωνικές ανάγκες υποστήριξης πιο ευάλωτων ανθρώπων, ενώ το θέμα 11 τη χρήση του διαδικτύου για να καλυφθούν πιο πολύ οι ψυχολογικές ανάγκες ψυχαγωγίας. Είναι

δύο θέματα που παρουσιάζουν ενδιαφέρον και σχετίζονται με την κοινωνική διάσταση της πανδημίας.

Το θέμα 12 είναι ένα θέμα με μη σαφή προσανατολισμό. Αφορά το lockdown, το οποίο υπήρξε ένα από τα πιο σκληρά περιοριστικά μέτρα που εφαρμόστηκαν αλλά παράλληλα αφορά και την Ινδία, καθώς υπάρχουν διάφοροι όροι που σχετίζονται άμεσα. Το θέμα 12 δεν θα αναλυθεί παραπάνω καθώς κρίθηκε ότι αφορά ένα πολύ ειδικό μέρος των μέτρων προστασίας κατά της πανδημίας και καλύπτεται από την ανάλυση άλλων θεμάτων με πιο σαφή προσανατολισμό.

Τα θέματα 13, 14 και 15 αφορούν σε ένα πολύ γενικότερο πλαίσιο το πεδίο της υγείας. Πιο συγκεκριμένα το θέμα 13 αφορά τον ιο Covid-19, την προέλευση του αλλά και άλλες ασθένειες. Το θέμα 14 αφορά το μέτρο προστασίας για τη μετάδοση του ιού και πιο ειδικά την μάσκα προστασίας. Ενώ το θέμα 15 αφορά το σύστημα υγείας, όχι μόνο τους ανθρώπους που το στελεχώνουν, αλλά το συσχετίζει την πολιτική. Είναι τρία θέματα που αφορούν ένα εξαιρετικά ενδιαφέρον πεδίο, τα οποία αναλύονται μαζί για να αποσαφηνιστούν περαιτέρω και να μπορούν να εξαχθούν χρήσιμα συμπεράσματα.

4.2 Αποτελέσματα GSDMM

Στη συνέχεια μετά τη διαλογή των θεμάτων που παρουσιάζουν μεγαλύτερο ενδιαφέρον, τα συγκεκριμένα θέματα αναλύθηκαν περαιτέρω με έναν άλλο αλγόριθμο μοντελοποίησης θεμάτων τον GSDMM. Στο συγκεκριμένο κεφάλαιο παρουσιάζονται η αξιολόγηση του μοντέλου με τη μετρική coherence (συγκεκριμένα την εκδοχή u_mass), τα θέματα που προέκυψαν και τέλος ποια από αυτά παρουσιάζουν το μεγαλύτερο ενδιαφέρον ανάλυσης.

Όπως αναφέρεται παραπάνω ο GSDMM αλγόριθμος εφαρμόστηκε σε επιλεγμένα θέματα που προέκυψαν μετά την υλοποίηση του LDA. Κάθε θέμα μετά την υλοποίηση του GSDMM εξήγαγε καινούργια θέματα. Ο αριθμός των θεμάτων που προέκυψε για κάθε ένα είναι διαφορετικός. Αυτό οφείλεται στο ότι ο αλγόριθμος εφαρμόστηκε σε κάθε θέμα πέντε φορές αλλάζοντας τον αριθμό των εξαγόμενων υποθεμάτων. Με άλλα λόγια κάθε θέμα εξήγαγε από 5 έως 10 υποθέματα. Στη συνέχεια έγινε ο υπολογισμός του coherence score για κάθε αριθμό υπο-θεμάτων που προέκυψε. Στο τέλος κρατήθηκε το θέμα με τον αριθμό υπο-θεμάτων που σημείωσε το καλύτερο coherence score. Παρακάτω φαίνεται αναλυτικά στους πίνακες το coherence score όλων των θεμάτων.

Πίνακας 2: Αποτελέσματα του coherence score για τα υποθέματα του topic 0

Topic	K = αριθμός θεμάτων	Coherence score
0	10	-2.2407463861347
	9	-1.981004063332718
	8	-1.8724274677365693
	7	-1.852179059071645
	6	-2.049138334318777

	5	-1.976312390373166
--	---	--------------------

Το καλύτερο coherence score είναι για $k = 7$

Πίνακας 3: Αποτελέσματα του coherence score για τα υποθέματα του topic 2

Topic	K = αριθμός θεμάτων	Coherence score
2	10	-2.2064
	9	-2.3389
	8	-2.2050
	7	-2.21002
	6	-2.13821
	5	-2.044151

Το καλύτερο coherence score είναι για $k = 5$

Πίνακας 4: Αποτελέσματα του coherence score για τα υποθέματα του topic 4

Topic	K = αριθμός θεμάτων	Coherence score
4	10	-2.77962
	9	-2.70504
	8	-2.585161
	7	-2.54627
	6	-2.555480
	5	-2.49460

Το καλύτερο coherence score είναι για $k = 5$

Πίνακας 5: Αποτελέσματα του coherence score για τα υποθέματα του topic 6

Topic	K = αριθμός θεμάτων	Coherence score
6	10	-2.883513
	9	-2.605113
	8	-2.6687093
	7	-2.6650304
	6	-2.71722281
	5	-2.4713731

Το καλύτερο coherence score είναι για $k = 5$

Πίνακας 6: Αποτελέσματα του coherence score για τα υποθέματα του topic 8

Topic	K = αριθμός θεμάτων	Coherence score
8	10	-2.5134199
	9	-2.522546759
	8	-2.55122872
	7	-2.611693025
	6	-2.828213489
	5	-2.52857372

Το καλύτερο coherence score είναι για $k = 10$

Πίνακας 7: Αποτελέσματα του coherence score για τα υποθέματα του topic 10

Topic	K = αριθμός θεμάτων	Coherence score
10	10	-2.87738006
	9	-2.9610822
	8	-2.823978091
	7	-2.92894444
	6	-2.7356264
	5	-2.769555376

Το καλύτερο coherence score είναι για $k = 6$

Πίνακας 8: Αποτελέσματα του coherence score για τα υποθέματα του topic 12

Topic	K = αριθμός θεμάτων	Coherence score
12	10	-2.65151756
	9	-2.9936612521
	8	-2.6712238
	7	-2.651896137
	6	-2.883596
	5	-2.5887745

Το καλύτερο coherence score είναι για $k = 5$

Πίνακας 9: Αποτελέσματα του coherence score για τα υποθέματα του topic 13

Topic	K = αριθμός θεμάτων	Coherence score
13	10	-2.53906100
	9	-2.7568474
	8	-2.50421315
	7	-2.7792535
	6	-2.604239
	5	-2.552623

Το καλύτερο coherence score είναι για $k = 8$

Πίνακας 10: Αποτελέσματα του coherence score για τα υποθέματα του topic 14

Topic	K = αριθμός θεμάτων	Coherence score
14	10	-2.28315725
	9	-2.4367
	8	-2.204841
	7	-2.204841
	6	-2.29130
	5	-2.43573

Το καλύτερο coherence score είναι για $k = 7$

Ανακεφαλαιώνοντας τα θέματα με το καλύτερο coherence score είναι:

- topic 0: 7 θέματα με coherence score -1.852179059071645
- topic 2: 5 θέματα με coherence score -2.04415135869612
- topic 4: 5 θέματα με coherence score -2.4946090164766455
- topic 6: 5 θέματα με coherence score -2.471373159707395
- topic 8: 10 θέματα με coherence score -2.513419930664066
- topic 10: 6 θέματα με coherence score -2.7356264734061106
- topic 12: 5 θέματα με coherence score -2.5887745713875883
- topic 13: 8 θέματα με coherence score -2.5042131554408047
- topic 14: 7 θέματα με coherence score -2.2048419717348233

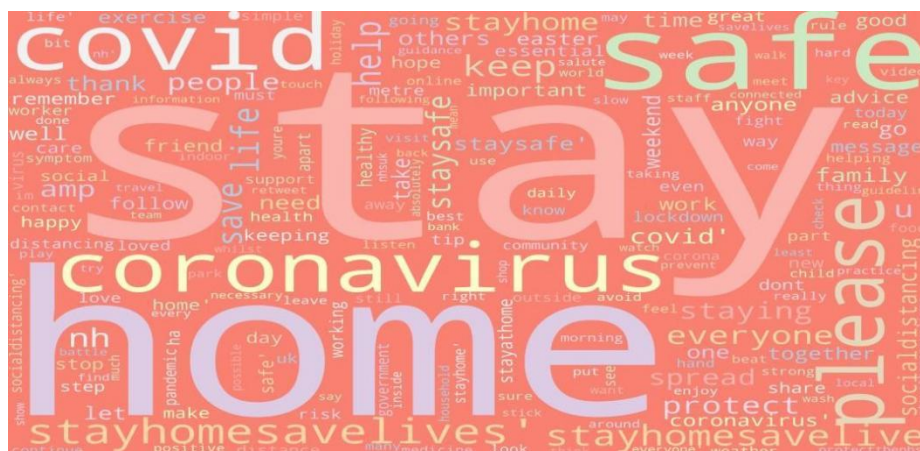
Από τα παραπάνω θέματα δεν παρουσιάζουν όλα το ίδιο ενδιαφέρον. Κάποια δεν σχετίζονται με το αντικείμενο μελέτης της παρούσας εργασίας, ενώ κάποια άλλα έχουν μεταξύ τους κοινή νοηματική κατεύθυνση. Συνεπώς στο συγκεκριμένο σημείο γίνεται ένας επιπλέον διαχωρισμός με βάση το πόσο ενδιαφέρον είναι ένα θέμα για την εργασία, όπως

έγινε αντίστοιχα στα αποτελέσματα του LDA αλγορίθμου. Παράλληλα για τα θέματα που επιλέγονται, παρουσιάζεται και η οπτική απεικόνιση των όρων που αποτελούν το καθ' ένα από αυτά και υλοποιήθηκε με το wordcloud, ένα πολύ γνωστό εργαλείο οπτικής αναπαράστασης. Όσο πιο έντονη και μεγάλη είναι ένας όρος τόσο πιο συχνός είναι στο συγκεκριμένο θέμα.

- Από το topic 0 προέκυψαν 7 θέματα. Από αυτά το θέμα 3 και το θέμα 6 κρίνονται αυτά με το μεγαλύτερο ενδιαφέρον.
 - Το θέμα 3 αναφέρεται σε θέματα κατανάλωσης και αγοράς προϊόντων, κάτι που ειδικά τη πρώτη περίοδο της πανδημίας όπου αποτελεί το αντικείμενο μελέτης της παρούσα εργασίας, ήταν πολύ επίκαιρο. Ο αριθμός των tweets στο θέμα 3 είναι 72568.
 - Το θέμα 6 αφορά το μένουμε σπίτι. Ένα μέτρο κοινωνικής απομόνωσης που σκοπό είχε την ανάσχεση και τον περιορισμό μετάδοσης της πανδημίας. Ο αριθμός των tweets στο θέμα 6 αντίστοιχα είναι 140760.



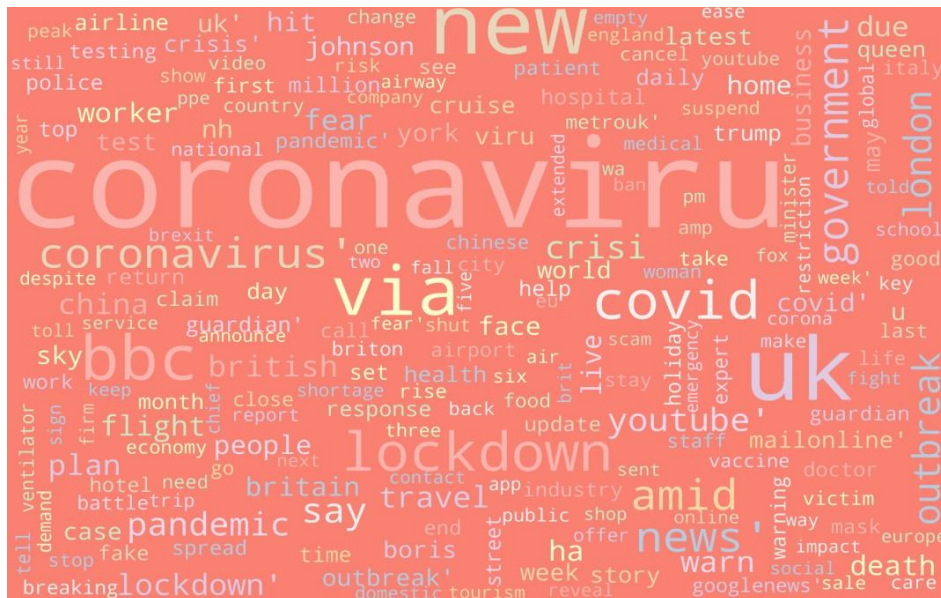
Εικόνα 20: Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 3 του θέματος 0 του LDA.



- Το θέμα 2 αφορά την μετάδοση ειδήσεων της πανδημίας, της κυβέρνησης και το bbc, ένα από τα μεγαλύτερα δημοσιογραφικά δίκτυα παγκοσμίως. Ο αριθμός των tweets είναι 264137.
- Το θέμα 3 σχετίζεται με τις επιχειρήσεις και τα δάνεια. Πρόκειται για ένα οικονομικό θέμα για το οποίο υπεύθυνος ήταν ο Rishi Sunak, καθώς έλαβε τις πολιτικές αποφάσεις που σχετίζονται με το συγκεκριμένο πεδίο. Ο αριθμός των tweets είναι 150477.



Εικόνα 27: Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 0 του θέματος 6 του LDA.



Εικόνα 28: Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 2 του θέματος 6 του LDA.



Εικόνα 34: Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 4 του θέματος 10 του LDA.

- Από το θέμα 12 προέκυψαν 5 θέματα εκ των οποίων το θέμα 2 εμφανίζει το μεγαλύτερο ενδιαφέρον. Αφορά την προέλευση του ιού και όλες τις θεωρίες που αναπτύχθηκαν γύρω του. Ο αριθμός των tweets είναι 43191.



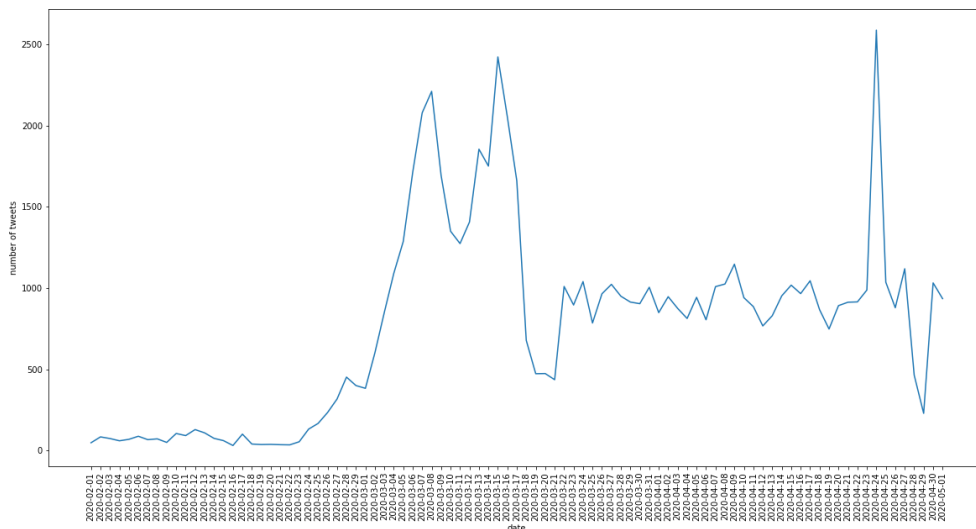
Εικόνα 35: Οπτική απεικόνιση των πιο συχνών όρων του υποθέματος 2 του θέματος 12 του LDA.

- Από το θέμα 13 προέκυψαν 8 θέματα. Τα πιο σημαντικά είναι το θέμα 2 και το θέμα 4.
 - Το θέμα 2 σχετίζεται με την μάσκα προστασίας. Ο αριθμός των tweets αντίστοιχα είναι 96714
 - Το θέμα 4 με τα test ανίχνευσης του ιού και ενδεχομένως με την έλλειψη τους. Ο αριθμός των tweets αντίστοιχα είναι 140836.

και με βάση αυτά προσπάθησε να γίνει ένας συνδυασμός χρονικής και ειδησεογραφικής ανάλυσης.

Παρακάτω παρουσιάζονται τα διαγράμματα που δείχνουν πως κυμαίνεται ο αριθμός των tweets για κάθε topic που επιλέχθηκε, σε συνάρτηση με τον χρόνο, από τη 1 Φεβρουαρίου μέχρι τη 1 Μαΐου. Στη συνέχεια παρουσιάζονται για κάθε ένα θέμα οι σημαντικότερες ειδήσεις που σχετίζονται με αυτό. Τέλος γίνονται οι παρατηρήσεις και σχόλια που βασίζονται στην παρατήρηση τους.

Από το topic 0 του LDA μοντέλου στη συνέχεια επιλέχθηκαν το θέμα 3 και το θέμα 6.



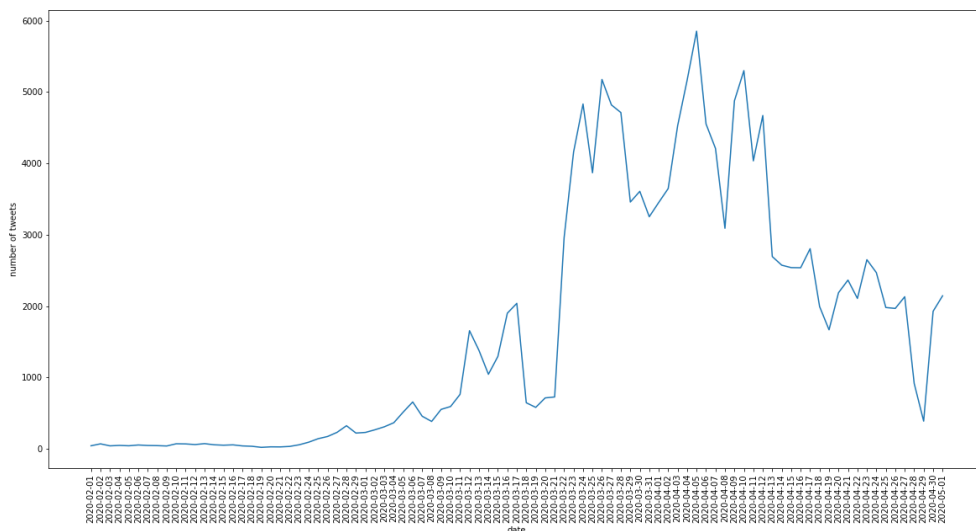
Εικόνα 41: Χρονική ανάλυση του υποθέματος 3 του topic 0 του LDA.

Το θέμα 3 αφορά την κατανάλωση και την αγορά προϊόντων (Εικόνα 41). Οι πιο σημαντικές ημερομηνίες που σχετίζονται με το θέμα 3 είναι:

- 14 Μαρτίου: Σε κοινή επιστολή τους εκπρόσωποι σουπερ-μαρκετ υπενθυμίζουν στους καταναλωτές να είναι προσεκτικοί στις αγορές τους, ούτως ώστε να μην υπάρξει έλλειψη αγαθών και προϊόντων (Austin, 2020). Ο αριθμός των tweets είναι περίπου 1.500 σύμφωνα με το διάγραμμα και είναι υψηλός. Επίσης παρατηρείται ότι ο αριθμός των tweets συνεχίζει να αυξάνεται τις επόμενες μέρες.
- 21 Μαρτίου: Ο Υπουργός Περιβάλλοντος, Τροφίμων και Αγροτικής Πολιτικής του Ηνωμένου Βασιλείου George Eustice επισημαίνει ότι υπάρχει διαθεσιμότητα τροφίμων και αγαθών, όμως τα καταστήματα παλεύουν να διατηρήσουν τα ράφια τους γεμάτα. Επιπλέον αναφέρει ότι οι καταναλωτές πρέπει να λειτουργήσουν υπεύθυνα και να σταματήσουν την αγορά προϊόντων λόγω πανικού (“Coronavirus: Shoppers told to buy responsibly”). Ο αριθμός των tweets είναι σχετικά χαμηλός, δεν ξεπερνούν τα 500. Ωστόσο παρατηρείται αυξητική τάση και άνοδος τις επόμενες ημέρες.

Με βάση το διάγραμμα από τις 23 Φεβρουαρίου αρχίζει να αυξάνεται ο αριθμός των tweets που αφορά το συγκεκριμένο θέμα. Αυτή είναι η ημερομηνία που σύμφωνα με τις ειδήσεις

εκείνης της εποχής, το Ηνωμένο Βασίλειο κατέγραψε 13 νέα περιστατικά covid-19 (“Coronavirus: Four new UK cases among ship evacuees”). Στις 8 Μαρτίου, σύμφωνα με το διάγραμμα παρατηρείται ένας μεγάλος αριθμός tweets. Η συγκεκριμένη ημερομηνία σημαδεύτηκε με την ανακοίνωση 3 θανάτων και 273 νέων κρουσμάτων (“Coronavirus: UK to remain in 'containment' phase of response”). Στη συνέχεια, στις 16 Μαρτίου παρατηρείται ότι ο αριθμός tweets είναι ιδιαίτερα αυξημένος. Αν γίνει μια ανασκόπηση των ειδήσεων αυτής της ημερομηνίας, θα διαπιστωθεί ότι ο αριθμός των νεκρών από την πανδημία αυτή τη μέρα ανέρχεται συνολικά στους 55, ενώ καταγράφονται πάνω από 1500 κρούσματα. Παράλληλα την ίδια μέρα ο πρωθυπουργός της Αγγλίας, Boris Johnson κάνει μια ισχυρή σύσταση στους πολίτες να αποφύγουν τα μη απαραίτητα ταξίδια καθώς και την επαφή με τους άλλους (“Coronavirus: PM says everyone should avoid office, pubs and travelling”). Στις 25 Απριλίου παρατηρείται ότι ο αριθμός των tweets είναι ιδιαίτερα υψηλός. Την συγκεκριμένη ημέρα καταγράφονται 813 θάνατοι και το Ηνωμένο Βασίλειο ξεπερνάει το όριο των 20.000 θανάτων.



Εικόνα 42: Χρονική ανάλυση του υποθέματος 6 του topic 0 του LDA.

Το θέμα 6 αφορά το μένουμε σπίτι (Εικόνα 42). Οι πιο σημαντικές ημερομηνίες που σχετίζονται με το θέμα 6 είναι:

- 12 Μαρτίου: Η Βρετανική κυβέρνηση συστήνει σε όσους έχουν συμπτώματα βήχα και υψηλό πυρετό, να απομονωθούν για 7 μέρες, καθώς τίθεται σε εφαρμογή το λεγόμενο “σχέδιο καθυστέρησης” της πανδημίας (“Coronavirus: People with fever or 'continuous' cough told to self-isolate”). Ο αριθμός των tweets είναι περίπου 1500. Παρατηρείται μια ανοδική τάση.
- 15 Μαρτίου: Ο Υπουργός Υγείας, Matt Hancock, δηλώνει ότι τις επόμενες μέρες κάθε Βρετανός άνω των 70 ετών θα κληθεί να μείνει στο σπίτι για ένα μεγάλο χρονικό διάστημα, με σκοπό την προστασία του από τον ιό (“Coronavirus: Isolation

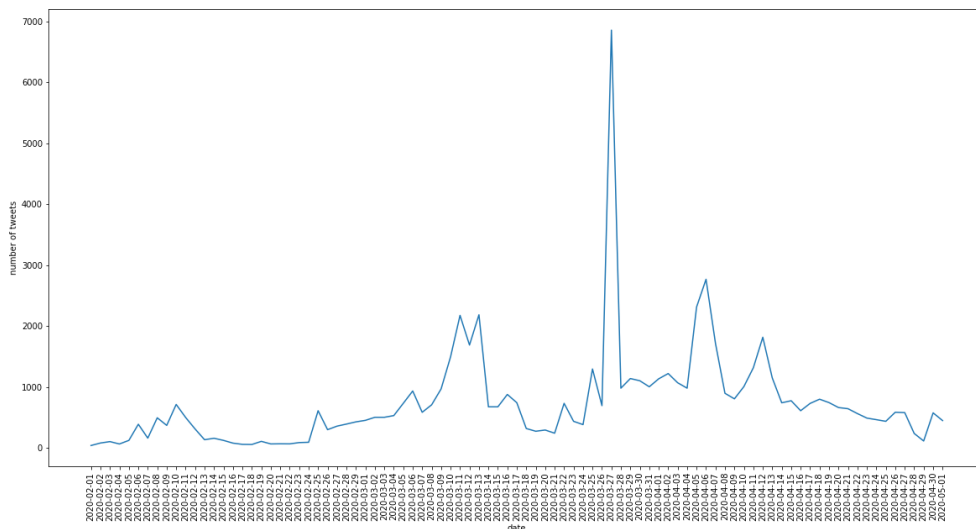
for over-70s 'within weeks'"). Μείωση του αριθμού των tweets, κυμαίνονται περίπου στα 1.000.

- 16 Μαρτίου: Ο Boris Johnson πραγματοποιεί ισχυρές συστάσεις στους πολίτες να δουλέψουν από το σπίτι, όπου αυτό είναι εφικτό. Παράλληλα οι πολίτες άνω των 70 ετών και οι έγκυες γυναίκες θα πρέπει να λάβουν πολύ σοβαρά υπόψη τις συστάσεις αυτές, καθώς τις επόμενες μέρες θα ζητηθεί από πολίτες που ανήκουν σε ευπαθείς ομάδες, να μείνουν σπίτι για 12 εβδομάδες ("Coronavirus: PM says everyone should avoid office, pubs and travelling"). Σημαντική αύξηση του αριθμού των tweets σε διάστημα μιας ημέρας, καθώς τα tweets πλησιάζουν τα 2.000 σε αριθμό.
- 23 Μαρτίου: Ο Boris Johnson σε διαγγελμα του ανακοινώνει την επιβολή lockdown. Δηλώνει ότι όλοι οι πολίτες πρέπει να μείνουν σπίτι τους. Οι πολίτες μπορούν να φεύγουν από το σπίτι τους για λόγους άσκησης μία φορά την ημέρα, για λόγους εργασίας μόνο όταν είναι απαραίτητο και για την αγορά αγαθών πρώτης ανάγκης ("Coronavirus: Strict new curbs on life in UK announced by PM"). Μεγάλος αριθμός tweets, πλησιάζουν τα 3.000, με ιδιαίτερα απότομη αυξητική τάση.
- 24 Μαρτίου: Μέσω των δικτύων κινητής τηλεφωνίας η κυβέρνηση στέλνει μήνυμα σε όλους τους πολίτες να μείνουν σπίτι, ασφαλείς και να τηρήσουν τα νέα περιοριστικά μέτρα ("Coronavirus: Mobile networks send 'stay at home' text"). Συνεχιζόμενη αυξητική πορεία, με πάνω από 3.500 tweets.
- 3 Απριλίου: Λόγω των καλών καιρικών προβλέψεων τις επόμενες μέρες, υπάρχει έντονη ανησυχία ότι πολλοί Βρετανοί θα αψηφήσουν τα μέτρα. Για τον λόγο αυτό ο Matt Hancock απευθύνει έκκληση στους Βρετανούς να μείνουν σπίτι, παρόλο που αναφέρει ότι η παραμονή στο σπίτι αυτό το Σαββατοκύριακο είναι οδηγία και όχι αίτημα της κυβέρνησης ("Coronavirus: Staying home this weekend 'not a request', UK told"). Ο αριθμός των tweets είναι περίπου 3.500. Τις προηγούμενες ημέρες παρατηρείται πτωτική τάση, όμως μετά από αυτήν την ημερομηνία ξεκινάει να αυξάνεται.
- 9 Απριλίου: Ο Υπουργός Εξωτερικών, Dominic Raab, προτρέπει τους πολίτες να μείνουν σπίτι και να τηρήσουν τα μέτρα το Σαββατοκύριακο του Πάσχα ("Coronavirus: Raab urges UK public not to ruin lockdown progress"). Ύστερα από μία σημαντική μείωση του αριθμού των tweets τις προηγούμενες ημερομηνίες, στις 9 Απριλίου ο αριθμός των tweets είναι 3.000 περίπου. Την επόμενη ημέρα, 10 Απριλίου, ο αριθμός των tweets ξεπερνάει τα 5.000.
- 30 Απριλίου: Ο Boris Johnson δηλώνει ότι θα ανακοινώσει ένα ολοκληρωμένο σχέδιο σταδιακής χαλάρωσης του lockdown την επόμενη εβδομάδα, το οποίο θα περιλαμβάνει άνοιγμα σχολείων και επανεκκίνηση της οικονομίας ("Coronavirus: Boris Johnson says UK is past the peak of outbreak"). Το προηγούμενο διάστημα παρατηρείται ότι ο αριθμός των tweets για το συγκεκριμένο θέμα έχει αρχίσει να φθίνει. Αυτό φαίνεται από το ότι τα tweets αυτή την ημέρα είναι περίπου 1.000 σε αριθμό. Ωστόσο παρατηρείται αυξημένη τάση το επόμενο διάστημα.

Στις 25 Μαρτίου, σύμφωνα με το γράφημα ο αριθμός των tweets είναι πολύ υψηλός φτάνοντας τις 5.000. Είναι η δεύτερη μέρα εφαρμογής του lockdown και παράλληλα η αστυνομία εξουσιοδοτείται να ασκήσει βία στη περίπτωση που υπάρξει άρνηση εφαρμογής των μέτρων από πλευράς των πολιτών (Dodd, 2020). Στις 28 Μαρτίου ο αριθμός των tweets

ξεπερνάει τις 5.000. Είναι η μέρα καταγραφής 260 νέων θανάτων, ξεπερνώντας το όριο των 1.000 θανάτων συνολικά από την αρχή της πανδημίας. Επιπλέον 17.089 νέα ημερήσια κρούσματα (“Coronavirus: Number of UK deaths rises above 1000”). Στις 5 Απριλίου παρατηρείται ο μεγαλύτερος αριθμός tweets και συγκεκριμένα φτάνουν τα 6.000. Είναι η ημέρα όπου μεταδίδεται το μαγνητοσκοπημένο μήνυμα της Βασίλισσας Ελισάβετ, απευθυνόμενη στον λαό, τους ευχαριστεί για τη τήρηση των μέτρων και τους προτρέπει να συνεχίσουν να μένουν στο σπίτι. Ενώ στο τέλος εκφράζει την απόλυτη πεποίθηση της ότι το Ηνωμένο Βασίλειο θα νικήσει την μάχη με την πανδημία. Ουσιαστικά πρόκειται για μια σπάνια κίνηση της Βασίλισσας, καθώς είναι μόλις το τέταρτο μήνυμα που απευθύνει στον βρετανικό λαό κατά τη διάρκεια της ζωής της (“Coronavirus: Queen tells UK 'we will succeed' in fight”).

Από το topic 2 του LDA, όπως αναφέρθηκε παραπάνω επιλέγονται τα subtopic 2, 3, 4.

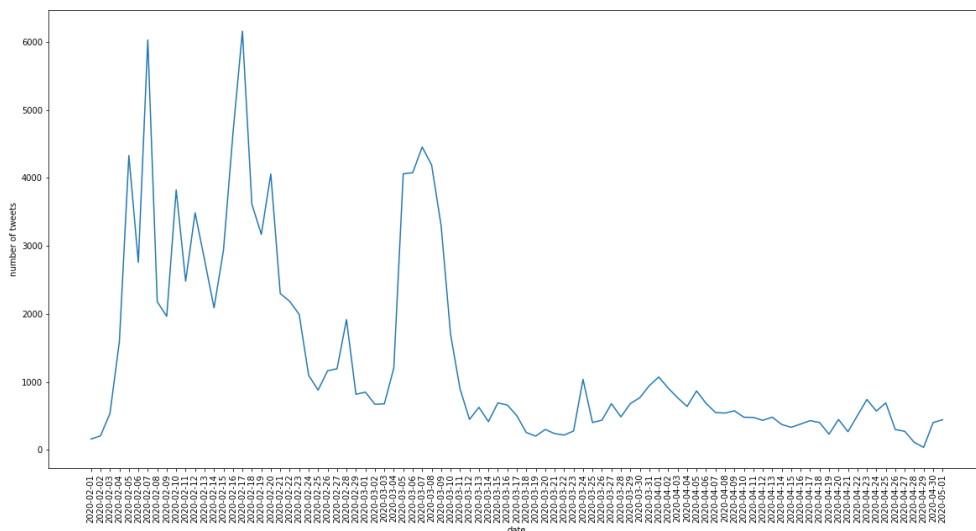


Εικόνα 43: Χρονική ανάλυση του υποθέματος 2 του topic 2 του LDA.

Το θέμα 2 αφορά τον Boris Johnson και το ότι νόσησε από covid-19 κατά τη διάρκεια της πρώτης φάσης της πανδημίας (Εικόνα 43). Οι σημαντικές ειδήσεις που σχετίζονται με το θέμα είναι:

- 10 Μαρτίου: Η Υπουργός Υγείας, Nadine Dorries, γίνεται η πρώτη υπουργός που βρίσκεται θετική στον covid-19. Εκφράζονται ανησυχίες για πιθανή μετάδοση του ιού στην Downing Street, όπου είχε παρευρεθεί σε δεξίωση μια εβδομάδα νωρίτερα (Walker, 2020). Ο αριθμός των tweets είναι σχετικά υψηλός ξεπερνώντας τα 2.000.
- 27 Μαρτίου: Ο Πρωθυπουργός, Boris Johnson και ο Υπουργός Υγείας Matt Hancock βρίσκονται θετικοί στον κορονοϊό, με τον Boris Johnson να απομονώνεται στην Downing Street και τον Matt Hancock σπίτι του (“Coronavirus: Prime Minister Boris Johnson tests positive”), (“Coronavirus: Health secretary Matt Hancock tests positive | The Independent”). Ο αριθμός των tweets είναι πάρα πολύ υψηλός αγγίζοντας τα 7.000.

- 5 Απριλίου: Ο Boris Johnson δέκα ημέρες μετά το θετικό τεστ, εισάγεται στο νοσοκομείο για εξετάσεις (“Coronavirus: Boris Johnson admitted to hospital over virus symptoms”). Αύξηση του αριθμού των tweets, μετά από ένα διάστημα πτωτικής σταθεροποιητικής τάσης. Τα tweets φτάνουν τα 2.500.
- 6 Απριλίου: Ο Boris Johnson μεταφέρεται στην εντατική, μετά από επιδείνωση της υγείας του (“Coronavirus: Boris Johnson moved to intensive care as symptoms worsen”). Αύξηση των tweets από τη προηγούμενη ημέρα, 5 Απριλίου, με τον αριθμό να αγγίζει τις 3.000.
- 12 Απριλίου: Ο Boris Johnson παίρνει εξιτήριο από το νοσοκομείο και ευχαριστεί το υγειονομικό προσωπικό που του έσωσε τη ζωή (“Coronavirus: Boris Johnson says 'it could have gone either way’”). Ο αριθμός των tweets αγγίζει τα 2.000.



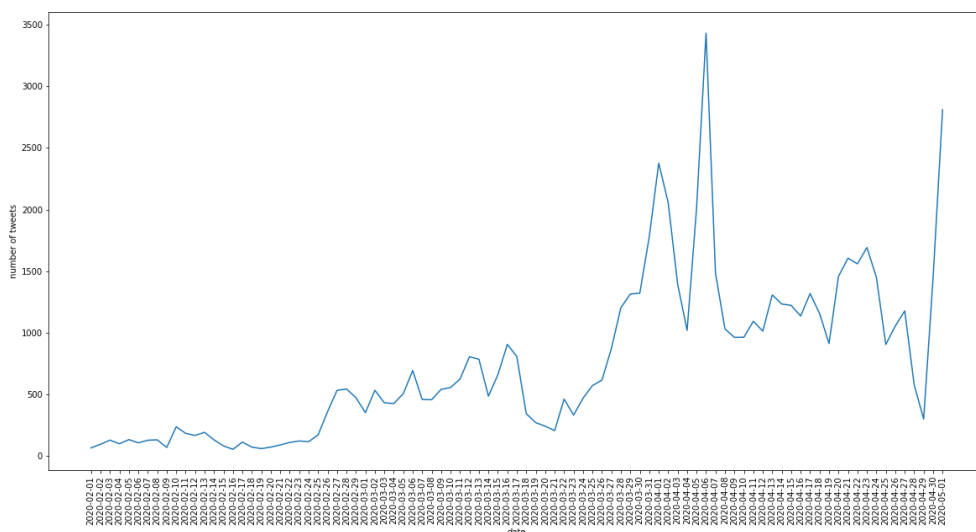
Εικόνα 44: Χρονική ανάλυση του υποθέματος 3 του topic 2 του LDA.

Το θέμα 3 αφορά την κρουαζιέρα Diamond Princess (Εικόνα 44). Παρακάτω ακολουθούν οι σημαντικότερες ημερομηνίες και ειδήσεις:

- 20 Ιανουαρίου: Το Diamond Princess αναχωρεί από το λιμάνι της Yokohama με 2.666 επιβάτες και 1.045 άτομα προσωπικό (“TIMELINE: Coronavirus epidemic (January and February)”).
- 1 Φεβρουαρίου: Το πλοίο φτάνει στο λιμάνι Naha και τίθεται σε καραντίνα. Παράλληλα ένας επιβάτης του πλοίου που αποβιβάστηκε πέντε μέρες νωρίτερα στο Hong Kong, εξετάζεται και βρίσκεται θετικός στον ιό. Το Υπουργείο Υγείας ειδοποιεί αμέσως τον πράκτορα εκμετάλλευσης του πλοίου, ωστόσο οι επιβάτες δεν ενημερώνονται παρά μόνο μετά από δύο ημέρες. Στο ενδιάμεσο στο κρουαζιερόπλοιο έγιναν κανονικά εκδηλώσεις και παραστάσεις (“COVID-19 pandemic on Diamond Princess”).
- 5 Φεβρουαρίου: Οι αρχές ανακοινώνουν δέκα θετικά κρούσματα. Παράλληλα ανακοινώνουν την ακύρωση της κρουαζιέρας και ότι το πλοίο τίθεται σε καθεστώς

καραντίνας για δεκατέσσερις μέρες, με βάση τις οδηγίες του ΠΟΥ (“Diamond Princess Updates - Notices & Advisories”). Ο συνολικός αριθμός των ανθρώπων που τίθενται σε καθεστώς καραντίνας από το ιαπωνικό υπουργείο Υγείας, είναι 3.700 (McCurry and Ratcliffe, 2020). Παρατηρώντας το διάγραμμα, ο αριθμός των tweets την συγκεκριμένη ημέρα ξεπερνάει τις 4.000.

- 7 Φεβρουαρίου: Ο συνολικός επιβεβαιωμένος αριθμός κρουσμάτων ανέρχεται σε 61, εκ των οποίων ο ένας είναι Βρετανός (Sim). Ο αριθμός των tweets εκτινάσσεται φτάνοντας τις 6.000.
- 17 Φεβρουαρίου: Δύο αεροσκάφη ναυλωμένα από την κυβέρνηση των ΗΠΑ, αναχωρούν για τις Ηνωμένες Πολιτείες μεταφέροντας 328 επιβάτες του πλοίου. Αρχικά οι ΗΠΑ ζήτησαν από την Ιαπωνία να τους κρατήσει για το χρονικό διάστημα των δεκατεσσάρων ημερών σύμφωνα με τις οδηγίες του CDC παρά την αντίθετη πρόταση της Ιαπωνίας. Ωστόσο η πολιτική τους άλλαξε στις 15 Φεβρουαρίου οπότε και αποφασίστηκε να επιστρέψουν στην πατρίδα τους. Τις επόμενες μέρες ακολουθούν το παράδειγμα των ΗΠΑ και άλλες χώρες όπως ο Καναδάς, η Αυστραλία και η Ιταλία (“Coronavirus: Japan cruise ship's US passengers home for further quarantine”). Ο αριθμός των tweets σύμφωνα με το γράφημα ξεπερνάει τις 6.000 τη συγκεκριμένη μέρα και αποτελεί την ημερομηνία με τον μεγαλύτερο αριθμό tweets που αφορά αυτό το θέμα.
- 23 Φεβρουαρίου: Τέσσερις Βρετανοί υπήκοοι βρίσκονται θετικοί και μεταφέρονται σε νοσοκομεία του Ηνωμένου Βασιλείου (“Coronavirus: Four new UK cases among ship evacuees”). Ο αριθμός των tweets σύμφωνα με το διάγραμμα είναι περίπου 2.000.
- 28 Φεβρουαρίου: Επιβεβαιώνεται ο πρώτος θάνατος Βρετανού πολίτη στο κρουαζιερόπλοιο Diamond Princess από το ιαπωνικό υπουργείο Υγείας (Weaver, 2020). Ο αριθμός των tweets είναι περίπου 2.000.
- 7 Μαρτίου: Παρατηρώντας το διάγραμμα διαπιστώνεται ότι ο αριθμός των tweets είναι αρκετά υψηλός και συγκεκριμένα ξεπερνάει τις 4.000. Η αύξηση των tweets τη συγκεκριμένη ημερομηνία δεν μπορεί να συσχετιστεί με κάποια είδηση που να αφορά τη κρουαζιέρα Diamond Princess. Ωστόσο με κάποια έρευνα βρέθηκε ότι μπορεί να αφορά ένα άλλο κρουαζιερόπλοιο το Grand Princess, που ταξίδεψε στα μέσα Φεβρουαρίου στην Καλιφόρνια, το Μεξικό και τη Χαβάη. Όμως δεν μπορεί να συσχετιστεί με το Ηνωμένο Βασίλειο αλλά σαν είδηση αναφέρεται στον αγγλικό τύπο και συγκεκριμένα στον Independent (“Coronavirus cases in UK rise to 206 today”).



Εικόνα 45: Χρονική ανάλυση του υποθέματος 4 του topic 2 του LDA.

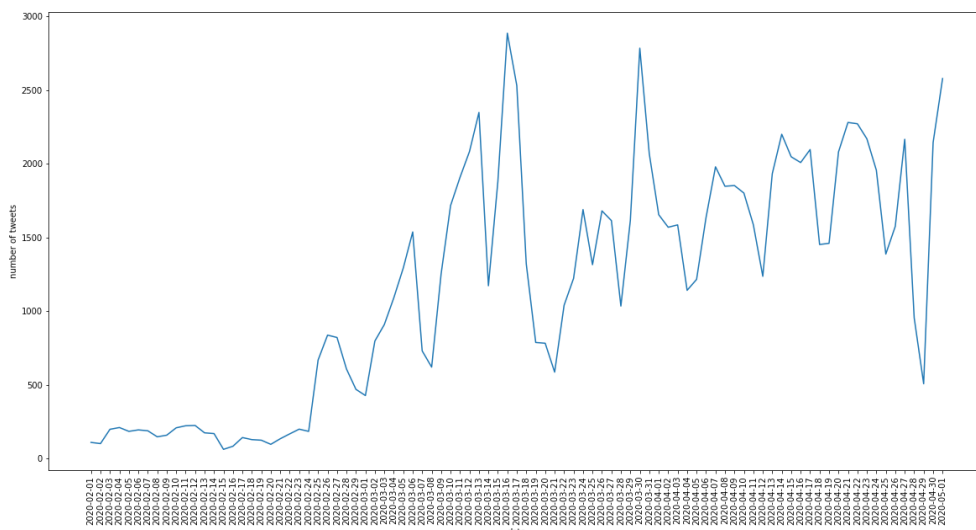
Το θέμα 4 αφορά τα διαγνωστικά test και τα συμπτώματα του covid-19 (Εικόνα 45).

- 23 Μαρτίου: Ο Boris Johnson ανακοινώνει την επιβολή lockdown (“Coronavirus: Strict new curbs on life in UK announced by PM”). Ο αριθμός των tweets που αφορά το θέμα είναι χαμηλός και συγκεκριμένα δεν ξεπερνούν τα 500, όμως στη συνέχεια παρατηρείται μια πολύ σημαντική αυξητική τάση με τον αριθμό των tweets συνεχώς να μεγαλώνει.
- 1 Απριλίου: Η κυβέρνηση επιβεβαιώνει ότι έχουν ελεγχθεί 2.000 μέλη ιατρονοσηλευτικού προσωπικού από την έναρξη της πανδημίας. Ο Υπουργός Προεδρίας του Ηνωμένου Βασιλείου Michael Gove δηλώνει ότι εξαιτίας της έλλειψης χημικών αντιδραστήρων που είναι απαραίτητα στα διαγνωστικά covid-test, δεν είναι δυνατός ο έλεγχος όλων των εργαζομένων υγείας, ο αριθμός των οποίων ανέρχεται σε 1.2 εκατομμύρια (“Coronavirus: Boris Johnson vows more virus tests as UK deaths exceed 2000”). Η Ένωση Χημικών Βιομηχανιών ανταπαντά διαψεύδοντας την ύπαρξη έλλειψης και αναφέροντας συγκεκριμένα ότι σε προηγούμενη συνάντηση τους με εκπρόσωπο της Βρετανικής κυβέρνησης εκείνος δεν προσπάθησε καν να μάθει αν υπάρχουν πιθανά προβλήματα εφοδιασμού (Peston, 2020). Ο αριθμός των tweets είναι υψηλός καθώς αγγίζει τις 2.500.
- 2 Απριλίου: Ο Matt Hancock θέτει ως στόχο την διενέργεια 100.000 test την ημέρα μέχρι το τέλος του μήνα (“Coronavirus: Matt Hancock sets aim of 100000 tests a day by end of April”). Ο αριθμός των tweets συνεχίζει να παραμένει υψηλός και είναι περίπου ίδιος με τον αριθμό της προηγούμενης μέρας.
- 6 Απριλίου: Ο αριθμός των θυμάτων από covid-19 ξεπερνάει τους 5.000. Τα κρούσματα που αναφέρονται είναι 52.000 (Walker and Weaver, 2020). Καταγράφεται ο υψηλότερος αριθμός tweets για το συγκεκριμένο θέμα, αγγίζοντας τις 3.500.
- 23 Απριλίου: Ο Matt Hancock ανακοινώνει ότι τα ημερήσια test έχουν φτάσει τις 51.000 και παράλληλα ανακοινώνει ότι όλοι οι εργαζόμενοι πρώτης γραμμής και οι οικογένειες τους θα μπορούν να υποβληθούν σε διαγνωστικά test δηλώνοντας το από

την επόμενη μέρα στη κυβερνητική σελίδα που θα είναι διαθέσιμη (“Coronavirus: UK to remain in 'containment' phase of response”). Ο αριθμός των tweets έχει αυξηθεί σε σχέση με τις προηγούμενες μέρες και πλησιάζει τις 2.000.

- 24 Απριλίου: Ο ιστότοπος που ανακοινώθηκε τη προηγούμενη μέρα ότι θα είναι διαθέσιμος για κράτηση διαγνωστικών test έπεσα εξαιτίας της τεράστιας συμμετοχής και μεγάλης ζήτησης. Τα πρώτα δύο λεπτά 5.000 δοκιμαστικά kit παραγγέλνονται. Η κυβέρνηση δηλώνει ότι θα αυξήσει ακόμα παραπάνω τα test (“Coronavirus: Test website closes after 'significant demand’”). Συνεχίζεται να είναι αυξημένος ο αριθμός των tweets όπως την προηγούμενη μέρα.
- 1 Μαΐου: Ο Matt Hancock επιβεβαιώνει σε δηλώσεις του ότι ο στόχος της κυβέρνησης για παροχή 100.000 test επιτεύχθηκε, καθώς 122.347 δόθηκαν μόλις σε ένα εικοσιτετράωρο (“Hancock confirms 100,000 coronavirus testing target met”). Ραγδαία αύξηση του αριθμού των tweets αγγίζοντας σχεδόν τις 3.000 ύστερα από μια χρονική περίοδο όπου υπήρχε πτωτική τάση.

Από το topic 4 του LDA, όπως αναφέρθηκε παραπάνω επιλέγονται τα subtopic 0 και 2.

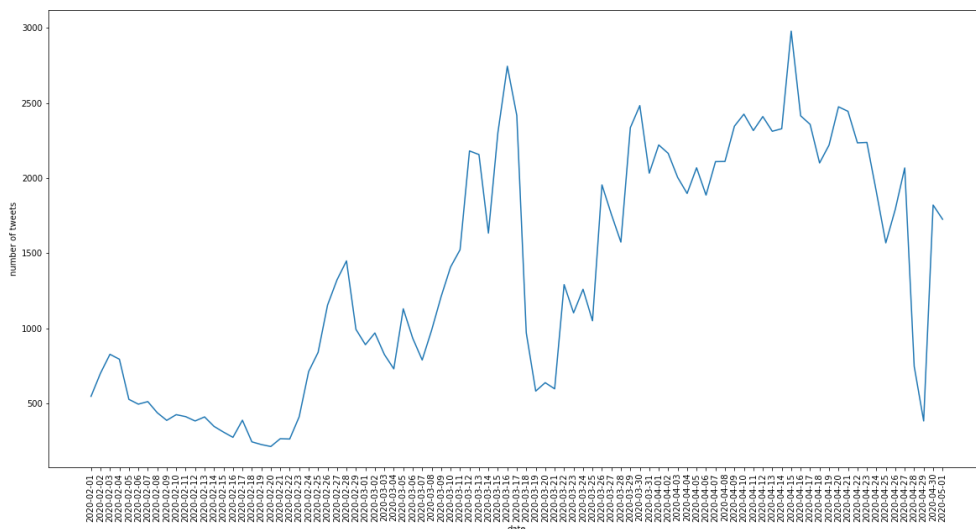


Εικόνα 46: Χρονική ανάλυση του υποθέματος 0 του topic 4 του LDA.

Το θέμα 0 αφορά τα σχολεία, τα πανεπιστήμια και γενικότερα την εκπαίδευση (Εικόνα 46).

- 12 Μαρτίου: Τα σχολεία καλούνται να ακυρώσουν προγραμματισμένα ταξίδια στο εξωτερικό (Mahase, 2020). Υψηλός αριθμός tweets ξεπερνά τις 2.000.
- 17 Μαρτίου: Ο αριθμός των tweets τη συγκεκριμένη ημέρα είναι πολύ υψηλός αγγίζοντας σχεδόν τις 3.000 και είναι ο υψηλότερος αριθμός tweets. Δεν υπάρχει κάποια συγκεκριμένη είδηση που αφορά τα σχολεία, ωστόσο την επόμενη ακριβώς ημέρα ανακοινώνεται το κλείσιμο σχολείων και πανεπιστημίων. Αν μπορεί να γίνει μια υπόθεση είναι ότι λόγω της αυξημένης ανησυχίας, της ανοδικής τάσης κρουσμάτων και θανάτων, υπήρχαν πολλές εισηγήσεις και γνώμες που θεωρούσαν απαραίτητη την αναστολή λειτουργίας των σχολικών μονάδων.

- 18 Μαρτίου: Η κυβέρνηση ανακοινώνει το κλείσιμο των σχολείων από τη Παρασκευή 20 Μαρτίου, εκτός από αυτά που φροντίζουν τα παιδιά εργαζομένων πρώτης γραμμής (Richardson, 2020). Ο Υπουργός Παιδείας Gavin Williamson επιβεβαιώνει τη μη διεξαγωγή εξετάσεων για φέτος (McGuinness and Heffer, 2020). Παρατηρείται πτωτική τάση του αριθμού των tweets ωστόσο παραμένει περίπου στα 1.500.
- 31 Μαρτίου: Παρατηρείται σημαντική αύξηση των tweets με αριθμό λίγο λιγότερο από 3.000. Δεν υπάρχει κάποια είδηση που να αφορά τη λειτουργία των σχολείων αλλά δημοσιεύεται μια έρευνα του πανεπιστημίου του Sheffield και του Ulster που διαπιστώνει αύξηση του άγχους και της κατάθλιψης. Συγκεκριμένα αναφέρει ότι τα άτομα που ανέφεραν ότι αγχώνονται αυξήθηκαν από 17% σε 36%, ενώ τα άτομα με κατάθλιψη από 16% σε 38% (Mohdin, 2020).
- 19 Απριλίου: Ο Υπουργός Παιδείας Gavin Williamson διαψεύδει δημοσίευμα των Times, που αναφέρουν ως πιθανή ημερομηνία ανοίγματος των σχολικών μονάδων στις 11 Μαΐου. Ο ίδιος δηλώνει ότι δεν μπορεί να αναφέρει συγκεκριμένη ημερομηνία και ότι η προσπάθεια εστιάζεται στην τηλεεκπαίδευση (“Coronavirus: No date for when schools will reopen”). Ο αριθμός των tweets είναι περίπου 1.500.
- 30 Απριλίου: Ο Boris Johnson δηλώνει ότι θα ανακοινώσει ένα ολοκληρωμένο σχέδιο σταδιακής χαλάρωσης του lockdown την επόμενη εβδομάδα, το οποίο θα περιλαμβάνει άνοιγμα σχολείων (“Coronavirus: Boris Johnson says UK is past the peak of outbreak”). Μεγάλη αυξητική τάση παρατηρείται στον αριθμό των tweets, σε αντίθεση με τις προηγούμενες μέρες που ο αριθμός των tweets είχε μειωθεί αισθητά.

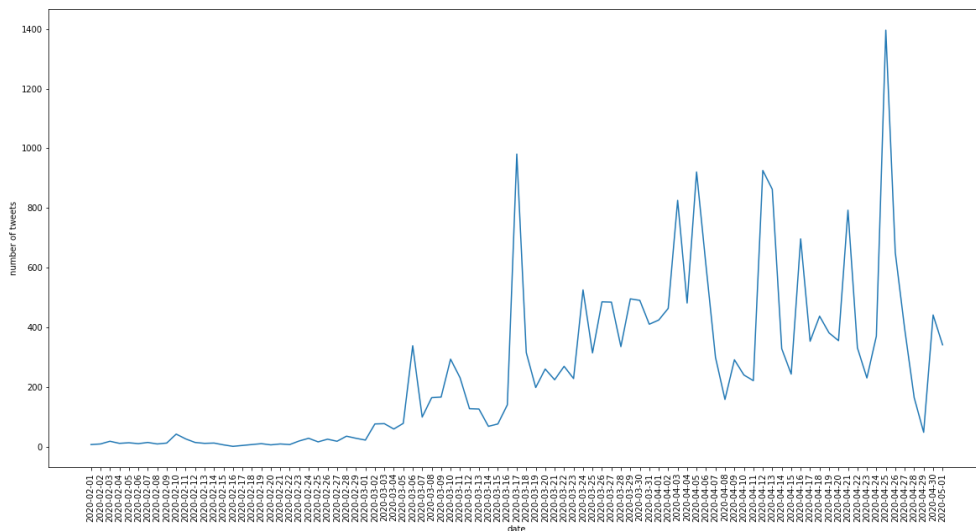


Εικόνα 47: Χρονική ανάλυση του υποθέματος 2 του topic 4 του LDA.

Το θέμα 2 αφορά τους ταξιδιωτικούς περιορισμούς και τη συνοριακή πολιτική που ακολουθήθηκε (Εικόνα 47).

- 25 Φεβρουαρίου: Κυβερνητικές συστάσεις για απομόνωση όσων ταξιδεύουν από το Χουμπέι, Ιραν και ορισμένες περιοχές της Νότιας Κορέας, ακόμα και αν δεν έχουν εμφανίσει κάποιο σύμπτωμα. Ο αριθμός των tweets παρατηρείται ότι αυτό το χρονικό διάστημα αυξάνεται με αυτή την ημέρα να ξεπερνούν τα 1.000
- 28 Φεβρουαρίου: Επιβεβαιώνεται ο πρώτος θάνατος Βρετανού πολίτη στο κρουαζιερόπλοιο Diamond Princess από το ιαπωνικό υπουργείο Υγείας (Weaver, 2020). Ο αριθμός των tweets είναι περίπου 2.000. Η ανησυχία είναι λογικό να εντείνεται και για αυτόν τον λόγο παρατηρείται ο αριθμός των tweets να έχει αυξηθεί σημαντικά και να είναι πάνω από 1.500.
- 12 Μαρτίου: Οι ταξιδιωτικές οδηγίες που είχαν εκδοθεί στις 25 Φεβρουαρίου, αίρονται. Ισχύει ότι και για όλους τους υπόλοιπους (“[Withdrawn] COVID-19: specified countries and areas”). Υψηλός αριθμός tweets, πάνω από 2.000.
- 16 Μαρτίου: Ο Boris Johnson πραγματοποιεί ισχυρές συστάσεις για αποφυγή ταξιδιών, εκτός και αν κρίνεται απαραίτητο (“Coronavirus: PM says everyone should avoid office, pubs and travelling”). Πάνω από 2.500 tweets και σημαντική αύξηση σε σχέση με το προηγούμενο διάστημα.
- 30 Μαρτίου: Ο Υπουργός Εξωτερικών Dominic Raab ανακοινώνει την συμφωνία μεταξύ της κυβέρνησης και των μεγάλων αεροπορικών εταιρειών του Ηνωμένου Βασιλείου, να επαναπατριστούν Βρετανοί υπήκοοι που έχουν αποκλειστεί σε άλλες χώρες (“Coronavirus: Foreign Office links up with airlines to fly stranded Britons home”). Ο αριθμός των tweets είναι υψηλός και συγκεκριμένα 2.500.
- 16 Απριλίου: Παρατηρείται μεγάλη αύξηση, με τα tweets να πλησιάζουν σε αριθμό τις 3.000. Ειδησεογραφικά δεν υπάρχει κάτι συγκεκριμένο που αφορά τους ταξιδιωτικούς περιορισμούς. Η μόνη είδηση που θα μπορούσε να συνδεθεί με τον μεγάλο αριθμό των tweets είναι οι δηλώσεις του Υπουργού Εξωτερικών Dominic Raab ότι το lockdown παρατείνεται για άλλες τρεις εβδομάδες, καθώς τα επιβεβαιωμένα ημερήσια κρούσματα ξεπερνούν τις 100.000 (“Coronavirus: UK lockdown extended for 'at least' three weeks”).
- 30 Απριλίου: Ο Boris Johnson δηλώνει ότι θα ανακοινώσει ένα ολοκληρωμένο σχέδιο σταδιακής χαλάρωσης του lockdown την επόμενη εβδομάδα, το οποίο θα επανεκκίνηση της οικονομίας και θα αίρει κάποιους περιορισμούς στις μετακινήσεις (“Coronavirus: Boris Johnson says UK is past the peak of outbreak”). Απότομη αυξητική τάση του αριθμού των tweets, πλησιάζοντας τα 2.000

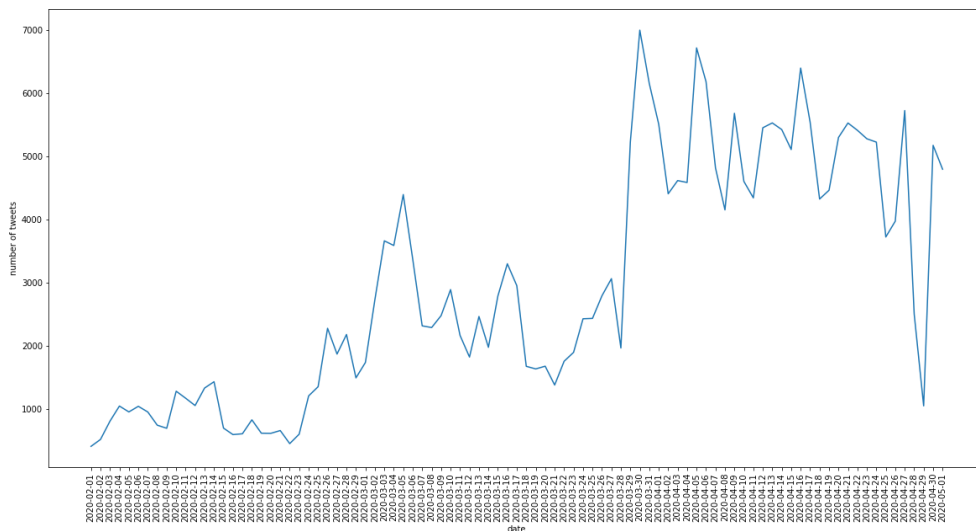
Από το topic 6 του LDA, όπως αναφέρθηκε παραπάνω επιλέγονται τα subtopic 0, 2 και 3.



Εικόνα 48: Χρονική ανάλυση του υποθέματος 0 του topic 6 του LDA.

Το θέμα 0 αφορά των Υπουργό Οησαυροφυλακίου Rishi Sunak (Εικόνα 48).

- 6 Μαρτίου: Ο Rishi Sunak σκοπευει να ανακοινώσει έκτακτο προϋπολογισμό με στόχο να βοηθήσει τις επιχειρήσεις να αντιμετωπίσουν τη πανδημία (Stewart, 2020). Παρατηρείται μια μικρή αύξηση του αριθμού των tweets με τον αριθμό τους να φτάνει τα 400.
- 17 Μαρτίου: Ο Rishi Sunak ανακοινώνει ότι 330 δισεκατομμύρια £ θα διατεθούν ως εγγυήσεις δανείων για τις επιχειρήσεις που έχουν πληγεί από τον ιό (“Coronavirus: Chancellor unveils £350bn lifeline for economy”). Πολύ σημαντική αύξηση του αριθμού των tweets με τον αριθμό να φτάνει τα 1.000.
- 3 Απριλίου: Ανακοινώνεται το CLBILS (Coronavirus Large Business Interruption Loan Scheme), ένα πρόγραμμα οικονομικής στήριξης των επιχειρήσεων (“Timeline of UK government coronavirus economic support for businesses, March 2020 to March 2022”). Αυξητική τάση του αριθμού των tweets. Κυμαίνονται γύρω στα 800.
- 13 Απριλίου: Ο Rishi Sunak ανακοινώνει ότι τριπλασιάζεται το ταμείο αντιμετώπισης του κορονοϊού για το νοσηλευτικό σύστημα και τις δημόσιες υπηρεσίες (Goodley, 2020). Πάνω από 800 tweets.
- 27 Απριλίου: Ο Rishi Sunak ανακοινώνει πρόγραμμα μικροδανείων για τις μικρές επιχειρήσεις με όριο τις 50.000 £ (“Rishi Sunak announces micro-loan scheme for businesses with £50000 limit – video”). Παρατηρείται μια μειωτική τάση στον αριθμό των tweets αν και την προηγούμενη μέρα, 26 Απριλίου, ο αριθμός των tweets είναι ιδιαίτερα υψηλός φτάνοντας τα 1.400.

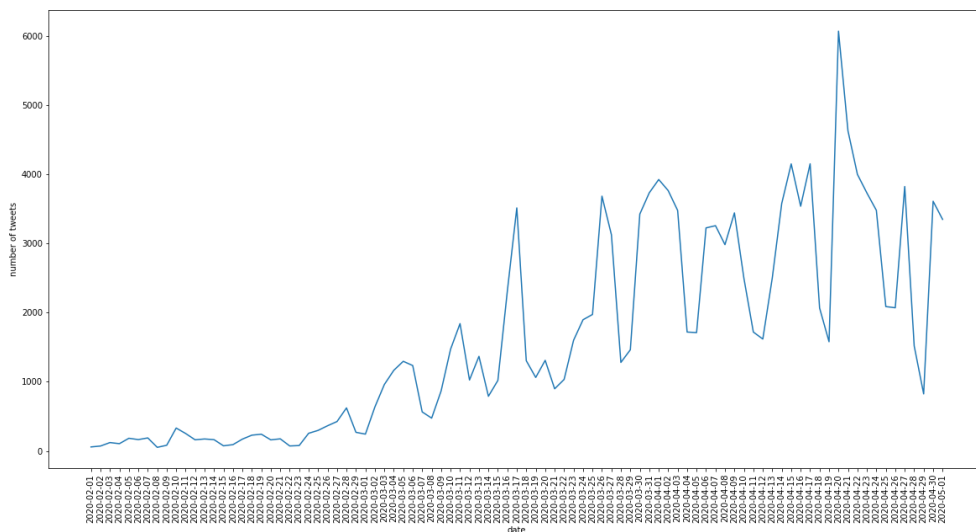


Εικόνα 49: Χρονική ανάλυση του υποθέματος 2 του topic 6 του LDA.

Το θέμα 2 αφορά την μετάδοση ειδήσεων της πανδημίας, της κυβέρνησης και το bbc (Εικόνα 49).

- 5 Μαρτίου: Επιβεβαιώνεται ο πρώτος θάνατος ασθενή από covid-19. Ο αριθμός των κρουσμάτων φτάνει 116 (“Coronavirus: Woman in 70s becomes first virus fatality in UK”). Το Ηνωμένο Βασίλειο πλέον περνά από την φάση περιορισμού στην φάση αντιμετώπισης της πανδημίας (Marsh and Campbell, 2020). Ο αριθμός των tweets ξεπερνάει τις 4.000. Είναι ένας σημαντικά υψηλός αριθμός.
- 16 Μαρτίου: Ο αριθμός των νεκρών φτάνει τους 55, ενώ τα κρούσματα ξεπερνούν τα 1.500 (“Coronavirus: PM says everyone should avoid office, pubs and travelling”). Ο πρωθυπουργός Boris Johnson σε διάγγελμα που απευθύνει στους Βρετανούς πολίτες μέσω του BBC, συμβουλεύει τον περιορισμό επαφών, την εργασία από το σπίτι, αποφυγή κοινωνικών χώρων που έχουν χαρακτηριστεί ως εστίες μετάδοσης (“Coronavirus: PM says everyone should avoid office, pubs and travelling”). Με βάση το διάγραμμα, ο αριθμός των tweets είναι πάνω από 3.000, αισθητά αυξημένος σε σχέση με τις προηγούμενες ημέρες.
- 31 Μαρτίου: Καταγράφονται 381 θάνατοι, ανάμεσα τους και ένα 13χρονο αγόρι. Αποτελεί τον μεγαλύτερο ημερήσιο αριθμό από την έναρξη της πανδημίας, ανεβάζοντας το σύνολο των ασθενών που κατέληξαν από covid-19 στους 1.789. Σύμφωνα με το διάγραμμα ο αριθμός των tweets αγγίζει τις 7.000. Πρόκειται για έναν πολύ υψηλό αριθμό tweets. Η αύξηση είναι ιδιαίτερα μεγάλη συγκριτικά με τις άλλες μέρες.
- 6 Απριλίου: Ο αριθμός των κρουσμάτων ξεπερνά τις 50.000 και παράλληλα ο αριθμός των νεκρών τις 5.000 (Walker and Weaver, 2020). Σύμφωνα με το διάγραμμα υψηλός παραμένει ο αριθμός των tweets πλησιάζοντας τις 7.000.
- 16 Απριλίου: Ο Υπουργός Εξωτερικών Dominic Raab ανακοινώνει την παράταση του lockdown για άλλες τρεις εβδομάδες, καθώς ο αριθμός των κρουσμάτων ξεπερνάει τις 10.000 (“Coronavirus: UK lockdown extended for 'at least' three weeks”). Συνεχίζει ο αριθμός των tweets να είναι υψηλός και συγκεκριμένα πάνω από 6.000.

- 23 Απριλίου: Το BBC μεταδίδει το Big Night In, μία Βρετανική εκπομπή με ψυχαγωγικό χαρακτήρα, κατά την διάρκεια της οποίας οι θεατές μπορούν να δωρίσουν ότι ποσό επιθυμούν σε ένα ταμείο που ενισχύει τοπικές φιλανθρωπικές οργανώσεις. Η εκδήλωση καταφέρνει να συγκεντρώσει 27 £ εκατομμύρια. Η κυβέρνηση δηλώνει τη στήριξη της στον σκοπό και δεσμεύεται να το διπλασιάσει (McDonald, 2020). Ο αριθμός των tweets είναι υψηλός, πάνω από 5.000.
- 27 Απριλίου: Ο Boris Johnson πραγματοποιεί την πρώτη του δημόσια δήλωση μετά το εξιτήριο που έλαβε από το νοσοκομείο και προτρέπει τους πολίτες να παραμείνουν ψύχραιμοι με τα περιοριστικά μέτρα. Ο αριθμός των tweets είναι αρκετά υψηλός, αγγίζοντας τις 6.000.



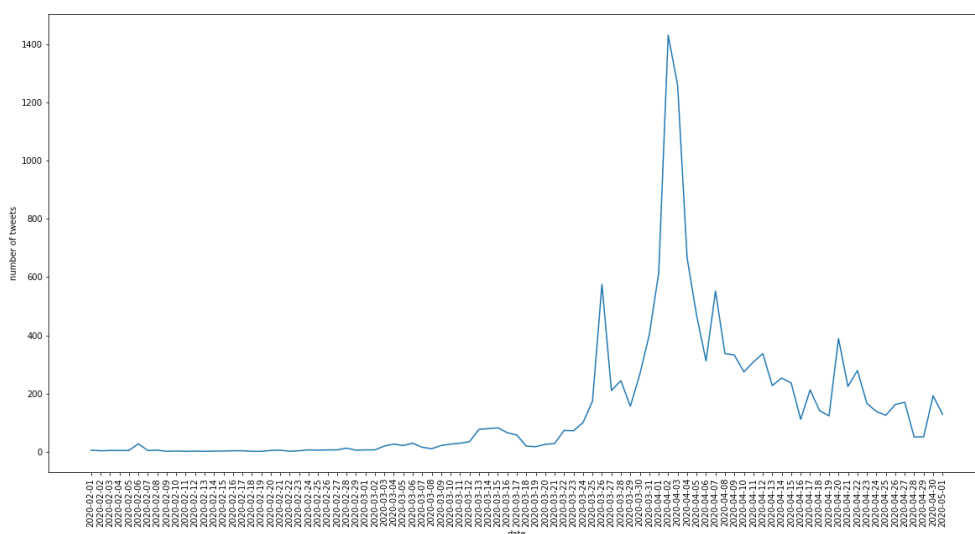
Εικόνα 50: Χρονική ανάλυση του υποθέματος 3 του topic 6 του LDA.

Το θέμα 3 αφορά τις επιχειρήσεις την οικονομία (Εικόνα 50).

- 11 Μαρτίου: Ο Υπουργός Rishi Sunak παρουσιάζει τον προϋπολογισμό της κυβέρνησης, ο οποίος περιλαμβάνει 30 δισεκατομμύρια £ για τη στήριξη και την προστασία της οικονομίας (Partington and Mason, 2020). Ο αριθμός των tweets είναι 2.000 τη συγκεκριμένη ημέρα.
- 17 Μαρτίου: Ο Υπουργός Rishi Sunak ανακοινώνει ότι 330 δισεκατομμύρια £ θα είναι διαθέσιμα ως εγγυήσεις δανείων στις επιχειρήσεις (“Coronavirus: Chancellor unveils £350bn lifeline for economy”). Σημαντική μεταβολή στην αύξηση του αριθμού των tweets. Ξεπερνούν τα 4.000.
- 26 Μαρτίου: Ο Rishi Sunak ανακοινώνει εκ μέρους της κυβέρνησης, ότι οι αυτοαπασχολούμενοι εργαζόμενοι μπορούν να υποβάλλουν αίτηση επιχορήγησης ύψους έως το 80% των μέσων μηνιαίων κερδών τους, ώστε να μπορέσουν να ανταπεξέλθουν στις επιπτώσεις της πανδημίας (“Coronavirus: UK government unveils aid for self-employed”). Παρατηρώντας το διάγραμμα, ο αριθμός των tweets παραμένει υψηλός συνεχίζοντας να είναι πάνω από 4.000.

- 1 Απριλίου: Οι επιχειρήσεις ξεκινούν να δέχονται επιχορηγήσεις μέσα από το Ταμείο Επιχορηγήσεων Μικρών Επιχειρήσεων και Λιανικού Εμπορίου, καθώς και από το Ταμείο Επιχορήγησης Φιλοξενίας και Αναψυχής (“Timeline of UK government coronavirus economic support for businesses, March 2020 to March 2022”). Ο αριθμός των tweets είναι υψηλός και συγκεκριμένα ξεπερνάει τις 4.000. Ουσιαστικά το χρονικό διάστημα μεταξύ 31 Μαρτίου και 3 Απριλίου ο αριθμός των tweets κυμαίνεται σε αυτούς τους ρυθμούς. Εκτιμώντας τη συνολική κατάσταση, με βάση τις ειδήσεις προηγούμενων και επόμενων ημερών η ανάγκη στήριξης των επιχειρήσεων και των εργαζομένων ήταν επιτακτική.
- 16 Απριλίου: Ο Rishi Sunak προσπαθώντας να κατευνάσει τους φόβους της παράτασης του lockdown και κυρίως τις οικονομικές επιπτώσεις μιας τέτοιας πολιτικής απόφασης, ανακοινώνει την επέκταση του προγράμματος χορήγησης εγγυήσεων δανείων ώστε να καλυφθούν επιχειρήσεις που κινδυνεύουν να μείνουν εκτός του προγράμματος (Elliott, 2020). Ο αριθμός των tweets πλησιάζει τις 4.000, ακολουθώντας μια πτωτική τάση σε αντίθεση με το προηγούμενο διάστημα.
- 20 Απριλίου: Στις 20 Μαρτίου ανακοινώνεται η δημιουργία του CJRS(Coronavirus Job Retention Scheme). Είναι ένα πρόγραμμα επιχορήγησης των εργοδοτών, κατά το οποίο καλύπτεται το 80% του μισθού των εργαζομένων και το κόστος απασχόλησης μέχρι 2.500 £ το μήνα (“Coronavirus: Government to pay up to 80% of workers' wages”). Η ηλεκτρονική πλατφόρμα αιτήσεων είναι πλέον διαθέσιμη και μέσα σε 30 λεπτά έχουν ήδη γίνει 67.000 αιτήσεις (“Coronavirus: More than 140000 firms claim wage bill help”). Με βάση το διάγραμμα χρόνου αποτυπώνεται πλήρως η σημασία αυτής της είδησης, καθώς ο αριθμός των tweets είναι 6.000, ο μεγαλύτερος που καταγράφεται και αφορά το θέμα.
- 27 Απριλίου: Ο Rishi Sunak ανακοινώνει ένα επιπλέον πρόγραμμα χορήγησης δανείων στις επιχειρήσεις με όριο τις 50.000 £ (“Rishi Sunak announces micro-loan scheme for businesses with £50000 limit – video”). Μετά από ένα διάστημα πτωτικής τάσης του αριθμού των tweets, τη συγκεκριμένη ημερομηνία παρατηρείται σημαντική αύξηση με 4.000 tweets.
- 30 Απριλίου: Δημοσιεύματα κάνουν λόγο για μη συμφωνία μεταξύ του Υπουργού Rishi Sunak και των τραπεζών σχετικά με το πρόγραμμα που ανακοινώθηκε στις 27 Απριλίου. Οι τελικοί όροι του προγράμματος, σύμφωνα με τα δημοσιεύματα, δεν έχουν ακόμα εγκριθεί και από τις δύο πλευρές με τις διαπραγματεύσεις να είναι έντονες παρασκηνακά (Makortoff, 2020). Τα tweets τη συγκεκριμένη μέρα είναι πάνω από 3.500.

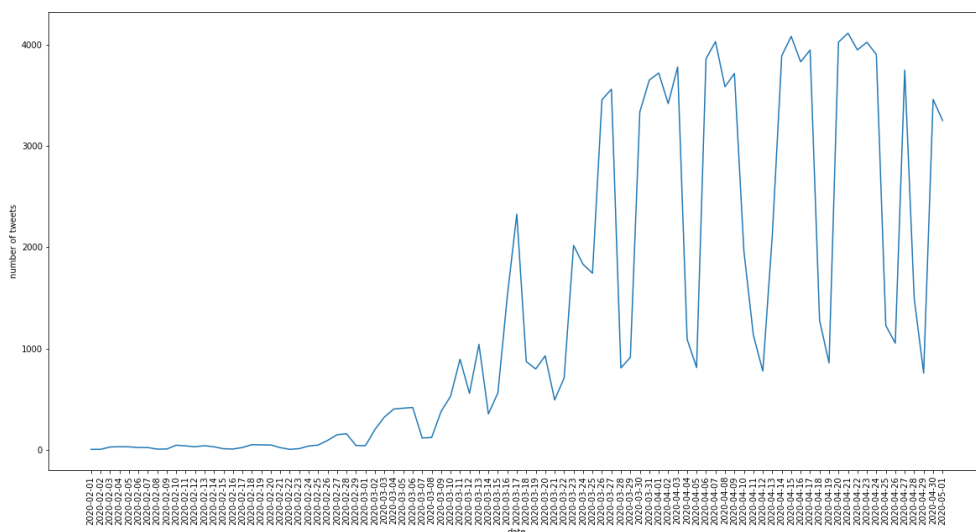
Από το θέμα 8 του LDA επιλέγονται το θέμα 1, το θέμα 2, το θέμα 7 και το θέμα 8.



Εικόνα 51: Χρονική ανάλυση του υποθέματος 1 του topic 8 του LDA.

Το θέμα 1 αφορά το αίτημα του Antonio Guterres, Γενικό Γραμματέα των Ηνωμένων Εθνών, για κατάπαυση του πυρός ανάμεσα στις εμπόλεμες χώρες (Εικόνα 51).

- 23 Μαρτίου: Ο Γενικός Γραμματέας του ΟΗΕ, Antonio Guterres απηύθυνε έκκληση σε όλες τις χώρες που εμπλέκονται σε πόλεμο να καταθέσουν τα όπλα τους, να σταματήσουν τις εχθροπραξίες και να υποστηρίξουν την παγκόσμια προσπάθεια αναχαίτισης της πανδημίας. Όπως τονίζει ο Γενικός Γραμματέας, ο ιός δεν ενδιαφέρεται για εθνότητες ή άλλες διαφορές. Επιπλέον τα συστήματα υγείας στις εμπόλεμες χώρες έχουν φτάσει στο όριο της ολικής κατάρρευσης, ενώ παράλληλα το ιατρικό και νοσηλευτικό προσωπικό θεωρούνται στόχοι. Με αυτόν τον τρόπο παρακινεί τις χώρες να μπουν σε ένα ουσιαστικό διάλογο μεταξύ τους, αναλαμβάνοντας να παίξει πρωταγωνιστικό ρόλο στη διπλωματική επίλυση των προβλημάτων. Το μήνυμα του μεταδόθηκε ζωντανά μέσω διαδικτύου, και πραγματοποιήθηκε στα κεντρικά γραφεία του ΟΗΕ στην Νέα Υόρκη (Dickinson, 2020). Ο αριθμός των tweets μέχρι το διαδικτυακό διάγγελμα του Γενικού Γραμματέα είναι ελάχιστος. Ωστόσο το επόμενο χρονικό διάστημα παρατηρείται μια σημαντική μεταβολή στον αριθμό των tweets. Μετά τις δηλώσεις του Antonio Guterres ξεκίνησε μια παγκόσμια προσπάθεια από πολίτες και ηγέτες χωρών να δηλώνουν δημόσια την στήριξη τους σ' αυτή τη πρωτοβουλία.
- 3 Απριλίου: Ο Γενικός Γραμματέας του ΟΗΕ, Antonio Guterres, ανακοινώνει ότι έντεκα, εμπλεκόμενες σε πόλεμο, χώρες ανταποκρίθηκαν στο αίτημα του ΟΗΕ για κατάπαυση του πυρός. Επιπλέον δήλωσε ότι θα συνεχίσει τις προσπάθειες τις επόμενες ημέρες (Wintour, 2020). Ο αριθμός των tweets είναι ο πιο υψηλός στο συγκεκριμένο θέμα και ξεπερνάει τα 1.400, έπειτα από μια χρονική περίοδο αυξητικής τάσης του αριθμού των tweets από τις 23 Μαρτίου.



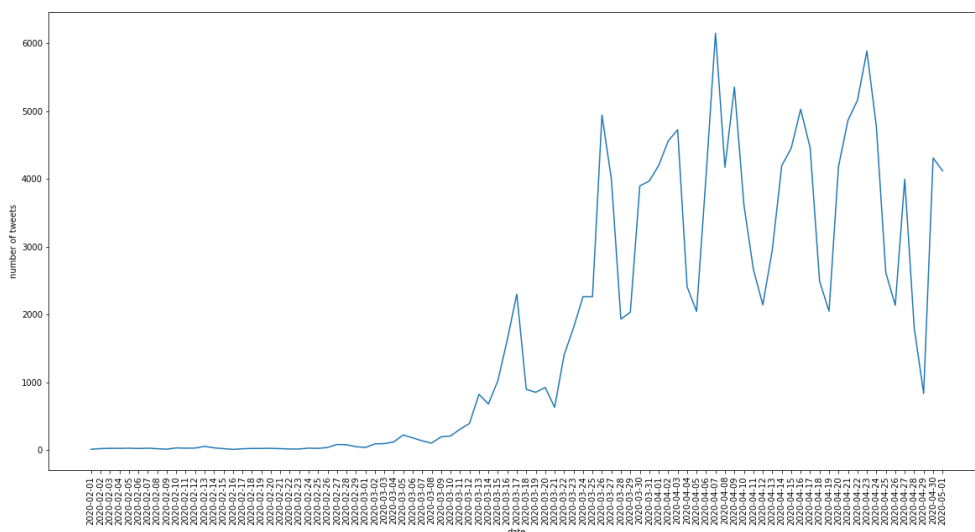
Εικόνα 52: Χρονική ανάλυση του υποθέματος 2 του topic 8 του LDA.

Το θέμα 2 αφορά τις επιχειρήσεις και το αίτημα-έκκληση για οικονομική βοήθεια και χρηματοδότηση (Εικόνα 52).

- 16 Μαρτίου: Ο Boris Johnson προτρέπει και συστήνει όλους τους Βρετανούς πολίτες να αποφύγουν τις συγκεντρώσεις και όλους τους πολυσύχναστους χώρους, όπως για παράδειγμα επιχειρήσεις εστίασης, θέατρα. Επιπλέον προτρέπει όσους είναι δυνατόν να εργαστούν από το σπίτι και να αποφύγουν τα ταξίδια (“Coronavirus: PM says everyone should avoid office, pubs and travelling”). Ουσιαστικά προαναγγέλλει το lockdown, χωρίς όμως ακόμα να προβεί σε καθολική απαγόρευση και αναστολή λειτουργίας των επιχειρήσεων. Παρατηρώντας το διάγραμμα ο αριθμός των tweets σημειώνει αυξητική τάση, ξεπερνώντας τα 2.000. Οι εργαζόμενοι και οι επιχειρήσεις αντιλαμβανόμενες την οικονομική ζημία, είναι λογικό να κάνουν έκκληση για οικονομική στήριξη.
- 26 Μαρτίου: Οι νέοι κανόνες προστασίας της δημόσιας υγείας τίθενται σε εφαρμογή, επεκτείνοντας το φάσμα των επιχειρήσεων που υποχρεούνται σε αναστολή λειτουργίας, συμπεριλαμβάνοντας όλες τις επιχειρήσεις λιανικής που δεν διαθέτουν την απαραίτητη κυβερνητική άδεια (“The Health Protection (Coronavirus, Restrictions) (England) Regulations 2020”). Ο αριθμός των tweets αυξάνεται κατακόρυφα φτάνοντας τις 3.500.
- 29 Μαρτίου μέχρι 3 Απριλίου: Παρατηρώντας το γράφημα του χρόνου διαπιστώνεται ότι ο αριθμός των tweets κυμαίνεται κατά μέσο όρο στις 3.500. Είναι ένα διάστημα αρκετών ημερών το οποίο μπορεί να αιτιολογηθεί παρατηρώντας πιο γενικά τις ειδήσεις εκείνης της περιόδου. Εκείνο το διάστημα η κυβέρνηση δεν έχει προβεί σε επιπλέον ανακοινώσεις στήριξης των πληττόμενων επιχειρήσεων. Η ανησυχία και η έκκληση των επιχειρήσεων για επιπλέον οικονομική στήριξη φαίνεται από τα ειδησεογραφικά άρθρα. Συγκεκριμένα σε άρθρο του ο Guardian στις 2 Απριλίου (Partington and Partridge, 2020) αποτυπώνει την έκδηλη ανησυχία πολλών

επιχειρήσεων και εργαζομένων. Τα Βρετανικά Εμπορικά Επιμελητήρια προειδοποιούν πως η σημαντική πτώση στις πωλήσεις, έχει ως αποτέλεσμα να φέρει πολλές επιχειρήσεις στο χείλος της κατάρρευσης, που θα έχει ως αποτέλεσμα πολλές χιλιάδες θέσεις εργασίας να χαθούν. Σύμφωνα με έρευνα έξι στις δέκα επιχειρήσεις έχουν αποθέματα μετρητών που φτάνουν μόνο για τρεις μήνες παρά τις προσπάθειες της Τράπεζας της Αγγλίας να μειώσει τα επιτόκια και να παράσχει φθηνή χρηματοδότηση.

- 5 Απριλίου μέχρι 9 Απριλίου: Ο αριθμός των tweets αυτή τη χρονική περίοδο συνεχίζει να είναι ιδιαίτερα υψηλός καθώς κυμαίνεται κατά μέσο όρο στις 4.000. Ανατρέχοντας σε δημοσιευμένα άρθρα της συγκεκριμένης χρονικής περιόδου η ανησυχία και η εκκλήσεις για ουσιαστική οικονομική στήριξη των επιχειρήσεων είναι χαρακτηριστική. Ενδεικτικά σύμφωνα με δημοσίευμα του Guardian στις 8 Απριλίου (Elliott, 2020) τα Βρετανικά Εμπορικά Επιμελητήρια δήλωσαν ότι με βάση την εβδομαδιαία τους έρευνα σχετικά με το πως διαχειρίζονται οι επιχειρήσεις τις επιπτώσεις της πανδημίας, υπήρξε αύξηση του αριθμού των επιχειρήσεων που σκοπεύουν να προβούν σε απολύσεις ως αποτέλεσμα του lockdown. Επιπλέον με βάση τις δηλώσεις οργανώσεων εργοδοτών όλο και περισσότερες επιχειρήσεις προβαίνουν σε απολύσεις εργαζομένων διότι έχουν αρχίσει να ξεμένουν από μετρητά και προσπαθούν να αποκτήσουν πρόσβαση στο πρόγραμμα έκτακτης οικονομικής ανάγκης της κυβέρνησης.
- 12 Απριλίου μέχρι 17 Απριλίου: Συνεχίζει ο αριθμός των tweets να κυμαίνεται σε υψηλά επίπεδα και να ξεπερνάει τις 4.000. Στις 15 Απριλίου ο Guardian σε νέο του άρθρο αναφέρει με στοιχεία ότι μόνο το 1/5 των επιχειρήσεων, που έχουν υποβάλλει αίτημα χορήγησης δανείων υποστηριζόμενο από την κυβέρνηση, έχουν λάβει την έκτακτη χρηματοδότηση. Το άρθρο αναφέρει τις ανησυχίες που προκύπτουν σε σχέση με την επιτυχία του προγράμματος (Makortoff, 2020).
- 19 Απριλίου μέχρι 25 Απριλίου: Ο αριθμός των tweets ξεπερνάει τις 4.000, παραμένοντας υψηλός. Σύμφωνα με δημοσίευμα του Guardian στις 22 Απριλίου, πάνω από το 70% των ιδιωτικών εταιρειών έχουν προβεί σε απολύσεις ανθρώπινου δυναμικού ως αποτέλεσμα των περιοριστικών μέτρων. Επιπλέον ενώ πάνω από 140.000 επιχειρήσεις υπέβαλαν αίτηση να ενταχθούν στο πρόγραμμα επιχορήγησης CJRS (Coronavirus Job Retention Scheme) που άνοιξε στις 20 Απριλίου. Ωστόσο όπως αναφέρει σε έρευνα του το BBC πολλές ακόμα επιχειρήσεις χρειάζεται να ενταχθούν στο συγκεκριμένο πρόγραμμα (Inman, 2020).



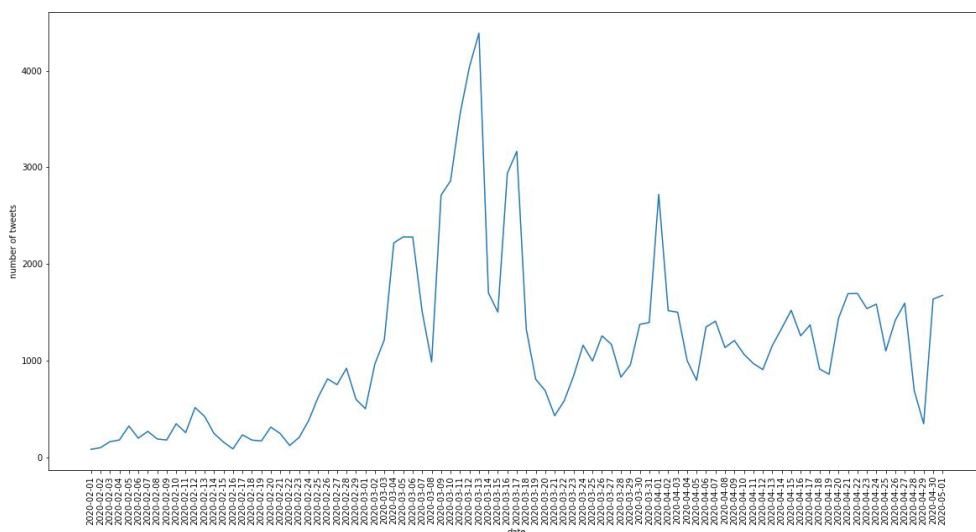
Εικόνα 53: Χρονική ανάλυση του υποθέματος 7 του topic 8 του LDA.

Το θέμα 7 αφορά την κοινωνική βοήθεια και στήριξη στο πλαίσιο της κοινότητας (Εικόνα 53).

- 17 Μαρτίου: Η κυβέρνηση ανακοινώνει οικονομικό πακέτο μέτρων ύψους 3.2 εκατομμύρια £ για την στήριξη απόρων και αστεγών πολιτών (Faleiro, 2020). Ο αριθμός των tweets κορυφώνεται τη συγκεκριμένη μέρα ανακοίνωσης του συγκεκριμένου μέτρου ξεπερνώντας τις 2.000, ύστερα από μια αυξητική τάση των προηγούμενων ημερών. Το πρόβλημα της στέγασης είναι σε έξαρση στο Ηνωμένο Βασίλειο. Σύμφωνα με έρευνα του 2019 περίπου 320.000 άνθρωποι είναι άστεγοι, ενώ τουλάχιστον 84.000 νοικοκυριά ζουν σε προσωρινά καταλύματα (“How many people are homeless in the UK? And what can you do about it?”).
- 27 Μαρτίου: Η αλυσίδα φαγητών Leon συγκεντρώνει εστιατόρια, διανομείς τροφίμων και προμηθευτές για να προσφέρουν δωρεάν καθημερινά ζεστά γεύματα στο υγειονομικό προσωπικό (Butler, 2020). Ο αριθμός των tweets που καταγράφεται είναι ιδιαίτερα υψηλός καθώς αγγίζει τις 5.000. Είναι λογικό με τα μέτρα περιορισμού πολλές κοινωνικές ομάδες που ήταν ήδη οικονομικά ασθενέστερες, πλέον με το lockdown το πρόβλημα να εντείνεται και να διογκώνεται.
- 31 Μαρτίου: Παρατηρείται σημαντική αύξηση των tweets με αριθμό λίγο λιγότερο από 5.000. Δημοσιεύεται μια έρευνα του πανεπιστημίου του Sheffield και του Ulster που διαπιστώνει αύξηση του άγχους και της κατάθλιψης. Συγκεκριμένα αναφέρει ότι τα άτομα που ανέφεραν ότι αγχώνονται αυξήθηκαν από 17% σε 36%, ενώ τα άτομα με κατάθλιψη από 16% σε 38% (Mohdin, 2020). Η δημοσίευση της έρευνας επισημαίνει την ανάγκη ψυχολογικής στήριξης και βοήθειας στους ανθρώπους που έχουν επηρεαστεί ιδιαίτερα από τις συνέπειες της πανδημίας και είναι μια υπενθύμιση του ψυχολογικού αντίκτυπου του ιού.
- 8 Απριλίου: Ο Rishi Sunak δηλώνει ότι οι φιλανθρωπικές οργανώσεις δίνουν τρομερή μάχη στήριξης κατά τη διάρκεια της πανδημίας και γι’ αυτόν τον λόγο θα στηρίξει την προσπάθειά τους με 750 εκατομμύρια λίρες επιπλέον χρηματική ενίσχυση (Perraudin, 2020). Ο αριθμός των tweets είναι υψηλός, περίπου 4.500.

- 16 Απριλίου: Ένας 99χρονος βετεράνος πολέμου ο Tom Moore περπατάει 100 γύρους γύρω από το σπίτι του, με σκοπό να συγκεντρώσει χρήματα που θα δοθούν στην ενίσχυση του συστήματος υγείας. Καταφέρνει να συγκεντρώσει 17 εκατομμύρια λίρες. Στη σελίδα του έκαναν δωρεές 800.000 άνθρωποι καταφέροντας να μαζέψει αυτό το πολύ σημαντικό ποσό (“Coronavirus: Capt Tom Moore's NHS fundraiser hits £17m”). Ο αριθμός των tweets είναι αρκετά υψηλός, αγγίζοντας τα 5.000.
- 18 Απριλίου: Ο Υπουργός Οικισμού, Κοινοτήτων και Τοπικής Αυτοδιοίκησης του Ηνωμένου Βασιλείου, Robert Jenrick ανακοινώνει επιπλέον 1.6 δισεκατομμύρια λίρες οικονομική ενίσχυση στις τοπικές αρχές. Ο αριθμός των tweets παραμένει υψηλός, κοντά στις 5.000. Οι δημοτικές αρχές έχοντας επωμιστεί μεγάλη οικονομική επιβάρυνση για την στήριξη ασθενέστερων οικονομικά πολιτών χρειάζονται περισσότερους οικονομικούς πόρους ώστε να μπορούν να παρέχουν βασικές υπηρεσίες.
- 23 και 24 Απριλίου: Παρατηρείται σημαντική αύξηση του αριθμού των tweets και συγκεκριμένα περίπου 6.000. Η ανάγκη για οικονομική ενίσχυση και στήριξη γίνεται ολοένα και μεγαλύτερη. Για τον λόγο αυτό, αυτές τις δύο μέρες γίνεται η διοργάνωση δύο σημαντικών φιλανθρωπικών δράσεων. Στις 23 Απριλίου το BBC μεταδίδει την εκπομπή The Big Night In με ψυχαγωγικό περιεχόμενο. Η εκπομπή γιορτάζει τις πράξεις αλληλεγγύης και συνεισφοράς, με το κοινό να έχει την ευκαιρία να συμμετέχει δωρίζοντας το ποσό που επιθυμεί σε ένα ταμείο το οποίο θα αξιοποιηθεί από φιλανθρωπικές οργανώσεις σε όλη τη χώρα. Η συμμετοχή είναι μεγάλη και το συνολικό ποσό που συγκεντρώνεται είναι 27 εκατομμύρια λίρες, με την κυβέρνηση επίσημα να υποστηρίζει τη συγκεκριμένη πρωτοβουλία και να δηλώνει την πρόθεση της να διπλασιάσει το ποσό (“BBC charity partners to celebrate the UK's community spirit and resilience with The Big Night In on BBC One”). Στις 24 Απριλίου ο βετεράνος πολεμιστής Tom Moore ηχογραφεί μαζί με τον Michael Ball την έκδοση του “You will never walk alone” με σκοπό την συγκέντρωση χρημάτων για το ταμείο ενίσχυσης του συστήματος υγείας (Savage).
- 30 Απριλίου: Ο βετεράνος Tom Moore γίνεται 100 χρονών και επίτιμος συνταγματάρχης της Βασίλισσας. Η έκκληση του για οικονομική στήριξη του συστήματος υγείας, συγκεντρώνει 32 εκατομμύρια λίρες (Patterson, 2020). Ο αριθμός των tweets είναι πάνω από 4.000.

Από το θέμα 10 του LDA αυτά που παρουσιάζουν το μεγαλύτερο ενδιαφέρον είναι το θέμα 3 και το θέμα 4.



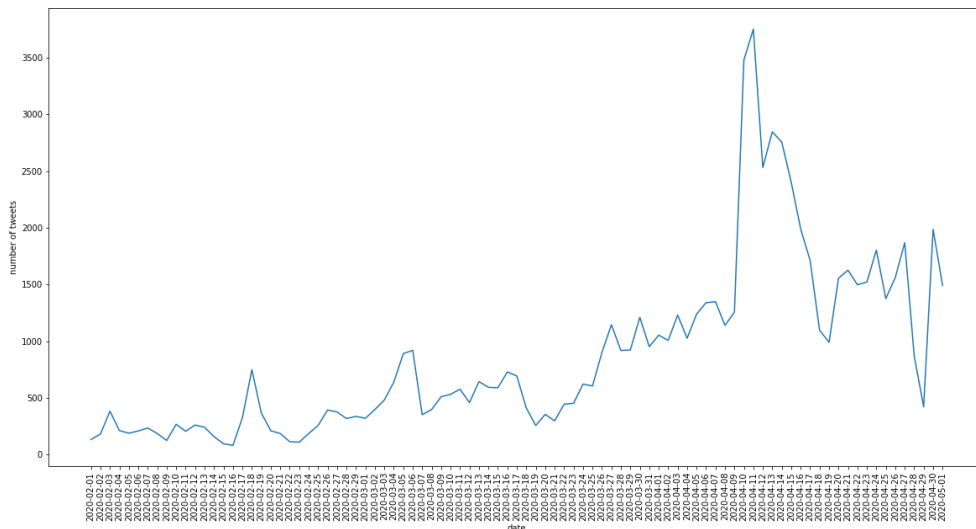
Εικόνα 54: Χρονική ανάλυση του υποθέματος 3 του topic 10 του LDA.

Το θέμα 3 αφορά την ακύρωση διάφορων εκδηλώσεων που ήταν προγραμματισμένες να πραγματοποιηθούν και ακυρώθηκαν λόγω της πανδημίας (Εικόνα 54).

- 5 Μαρτίου: Επιβεβαιώνεται ο πρώτος θάνατος ασθενή από covid-19. Ο αριθμός των κρουσμάτων φτάνει 116 (“Coronavirus: Woman in 70s becomes first virus fatality in UK”). Το Ηνωμένο Βασίλειο πλέον περνά από την φάση περιορισμού στην φάση αντιμετώπισης της πανδημίας (Marsh and Campbell, 2020). Ο αριθμός των tweets εκείνες τις μέρες είναι υψηλός και ξεπερνάει τις 2.000. Εκείνο το χρονικό διάστημα το Ηνωμένο Βασίλειο δεν έχει προβεί σε κάποιο δραστικό μέτρο περιορισμού της πανδημίας. Όλες οι εκδηλώσεις πραγματοποιούνται κανονικά και αυτό έχει ως αποτέλεσμα να τίθεται δημόσια η συζήτηση γύρω από την αναβολή ή μη όλων των τύπων εκδηλώσεων.
- 10 με 13 Μαρτίου: Πραγματοποιείται το Cheltenham festival, ιδιαίτερα γνωστή εκδήλωση και αφορά τις ιπποδρομίες, παρά το γεγονός ότι ο ΠΟΥ στις 11 Μαρτίου κηρύσσει πανδημία. Παρόλα αυτά το τριήμερο εκδηλώσεων πραγματοποιείται κανονικά και συγκεντρώνει 251.684 κόσμο, μόλις 2.000 λιγότερους από τη προηγούμενη χρονιά. Στη πορεία εξελίχθηκε ένας δημόσιος διάλογος σε σχέση με το αν έπρεπε το festival να πραγματοποιηθεί ή να ακυρωθεί. Υπάρχουν διάφορες δημοσιογραφικές έρευνες που αναφέρουν ότι η διεξαγωγή του κόστισε σε ζωές και επιτάχυνε την εξάπλωση της πανδημίας. Η κυβέρνηση και οι διοργανωτές επικρίθηκαν έντονα για το γεγονός από τα μέσα μαζικής ενημέρωσης (Humphries, 2020). Το BBC αντίθετα στις 13 Μαρτίου ακυρώνει τη διεξαγωγή του μουσικού festival Big Weekend, προγραμματισμένο να λάβει χώρα τέλη Μαΐου (“Coronavirus: Coachella, Radio 1's Big Weekend, BTS and other music events affected - CBBC Newsround”). Ο αριθμός των tweets το συγκεκριμένο διάστημα είναι ιδιαίτερα υψηλό, περίπου 4.500.
- 18 Μαρτίου: Το 50ο επετειακό Φεστιβάλ Glastonbury ακυρώνεται εξαιτίας της πανδημίας (Sparrow and Campbell, 2020). Ο αριθμός των tweets δεν είναι ιδιαίτερα

μεγάλος, περίπου 1.000 και ακολουθεί την πτωτική τάση που παρατηρείται εκείνο το χρονικό διάστημα.

- 27 Μαρτίου: Ακυρώνεται το καλοκαιρινό Cambridge Folk festival (Ryder, 2020). Ο αριθμός των tweets είναι λίγο παραπάνω από 1.000.
- 15 Απριλίου: Το Love Supreme Jazz festival που πραγματοποιείται τον Ιούλιο, ακυρώνεται (“Love Supreme Jazz Festival 2020 has been postponed | Music News - Jazz FM”). Ο αριθμός των tweets εξακολουθεί να βρίσκεται σε χαμηλά επίπεδα, πάνω από 1.500.

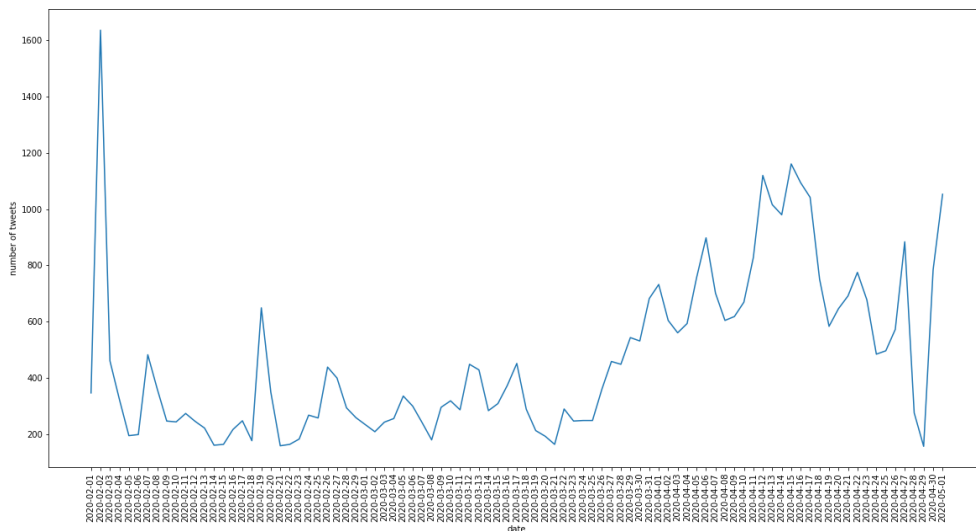


Εικόνα 55: Χρονική ανάλυση του υποθέματος 4 του topic 10 του LDA.

Το θέμα 4 αφορά την εφαρμογή της Apple σε συνεργασία με την Google, η οποία θα ανιχνεύει τις επαφές των ατόμων που νόσησαν (Εικόνα 55).

- 10 Απριλίου: Η εταιρεία Apple ανακοινώνει ότι σε συνεργασία με την Google, θα συνεργαστούν ώστε να αξιοποιήσουν από κοινού την smartphone τεχνολογία, με σκοπό την βοήθεια εντοπισμού και ανίχνευσης του ιού. Η συνεργασία θα επιτρέψει τη δημιουργία προηγμένης τεχνολογίας για την ιχνηλάτηση των επαφών. Η ιδέα είναι οι κυβερνήσεις να εντοπίσουν πιο εύκολα νέες εστίες μετάδοσης και ατομικά όσους έχουν εκτεθεί να απομονωθούν πριν προλάβουν να τον μεταδώσουν με τη σειρά τους. Ωστόσο προκύπτουν θέματα προστασίας προσωπικών δεδομένων και υπάρχουν αντιδράσεις αντίθετες με τη δημιουργία αντίστοιχων λογισμικών (Hern and Paul, 2020). Ο αριθμός των tweets είναι ιδιαίτερα υψηλός ξεπερνώντας τις 3.500, αποδεικνύοντας τον δημόσιο διάλογο και την αντίθετη επιχειρηματολογία των δύο πλευρών.

Από το θέμα 12 του LDA το θέμα 2 εμφανίζει το μεγαλύτερο ενδιαφέρον.

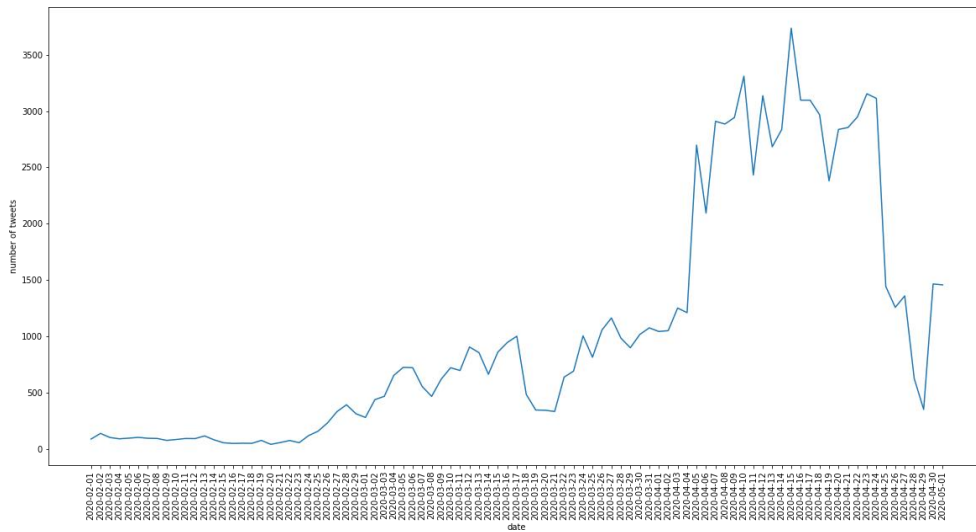


Εικόνα 56: Χρονική ανάλυση του υποθέματος 2 του topic 12 του LDA.

Το θέμα 2 αφορά την προέλευση του ιού και όλες τις θεωρίες που αναπτύχθηκαν γύρω από αυτό (Εικόνα 56).

- 30 Ιανουαρίου: Επιβεβαιώθηκαν τα δύο πρώτα κρούσματα απο covid-19. Πρόκειται για δύο Κινέζους υπηκόους που διαμένουν στο York (“Coronavirus: Two cases confirmed in UK”).
- 1 Φεβρουαρίου: Παρατηρώντας το γράφημα τα tweets, από την πρώτη μέρα είναι σε πολύ υψηλό αριθμό, φτάνοντας τα 1.600. Πρόκειται για έναν άγνωστο ιό μέχρι στιγμής με λίγες πληροφορίες, ενώ η δημόσια συζήτηση γύρω από τον ιό είναι ακόμα σε πολύ πρώιμο στάδιο, οι φήμες, τα ερωτήματα και οι πληροφορίες σε σχέση με τον κορονοϊό ήδη έχουν ξεκινήσει.
- 6 Φεβρουαρίου: Επιβεβαιώνεται τρίτο περιστατικό στο Ηνωμένο Βασίλειο (“Coronavirus: Third UK patient 'caught coronavirus in Singapore”).
- 30 Μαρτίου: Ο ΠΟΥ δημοσιεύει τη παγκόσμια έρευνα που διεξήγαγε στην Κίνα για την προέλευση του ιού και δημοσιοποιεί τα συμπεράσματά του (“WHO-convened global study of origins of SARS-CoV-2: China Part”). Παρατηρείται αύξηση της τάσης του αριθμού των tweets, αν και τη συγκεκριμένη ημερομηνία ο αριθμός παραμένει σε σχετικά χαμηλά επίπεδα. Πιο συγκεκριμένα περίπου 600 σε αριθμό.

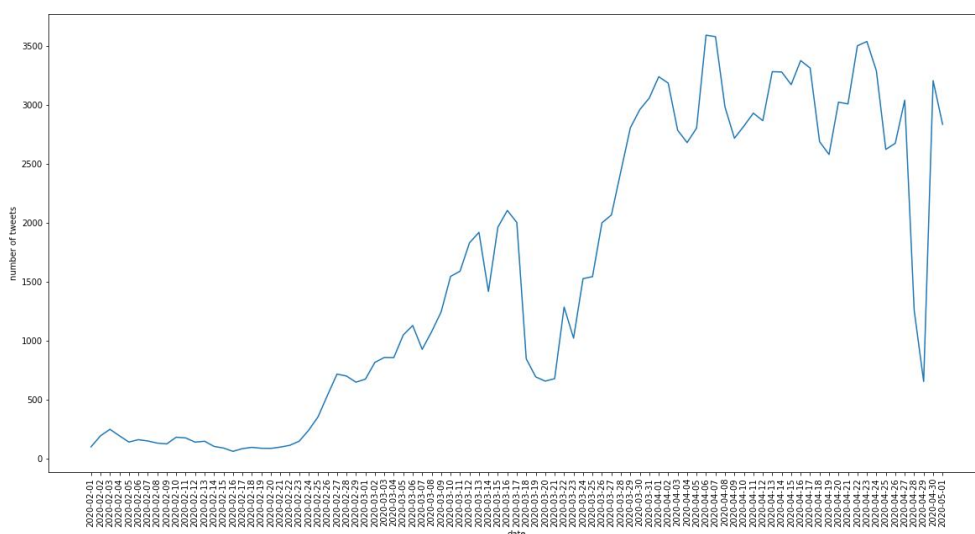
Από το θέμα 13 του LDA το θέμα 2 και το θέμα 4 εμφανίζουν το μεγαλύτερο ενδιαφέρον.



Εικόνα 57: Χρονική ανάλυση του υποθέματος 2 του topic 13 του LDA.

Το θέμα 2 αφορά τις μάσκες προστασίας (Εικόνα 57).

- 4 Απριλίου μέχρι 25 Απριλίου: Παρατηρώντας το διάγραμμα, το συγκεκριμένο χρονικό διάστημα παρατηρείται μια πολύ σημαντική αύξηση του αριθμού των tweets κατά τη διάρκεια της οποίας ο αριθμός παραμένει σταθερά υψηλός, και κυμαίνεται κατά μέσο όρο στις 3.000. Σύμφωνα με τις ειδήσεις και τα περιοριστικά μέτρα που τέθηκαν από την κυβέρνηση εκείνο το χρονικό διάστημα, η απόφαση οι μάσκες προστασίας να είναι υποχρεωτικές είτε σε εξωτερικούς είτε σε εσωτερικούς χώρους, ως μέτρο περιορισμού της πανδημίας, δεν λήφθηκε. Ωστόσο ήταν έντονη η δημόσια συζήτηση γύρω από το αν πρέπει να είναι υποχρεωτικές οι μάσκες προστασίας. Αυτό διαπιστώνεται από διάφορα άρθρα γνώμης εκείνης της περιόδου. Ενδεικτικά ένα από αυτά είναι δημοσιευμένο στον Guardian στις 24 Απριλίου (Javid, 2020). Ο συγγραφέας του άρθρου αναρωτιέται έντονα τον λόγο που δεν έχει παρθεί το συγκεκριμένο μέτρο, σε αντίθεση με άλλες Ευρωπαϊκές χώρες, όπου είχε ληφθεί ήδη από πολύ νωρίς και παράλληλα αναφέρει την κυβερνητική θέση ότι ακόμα δεν έχει αποδειχθεί η χρησιμότητα της.



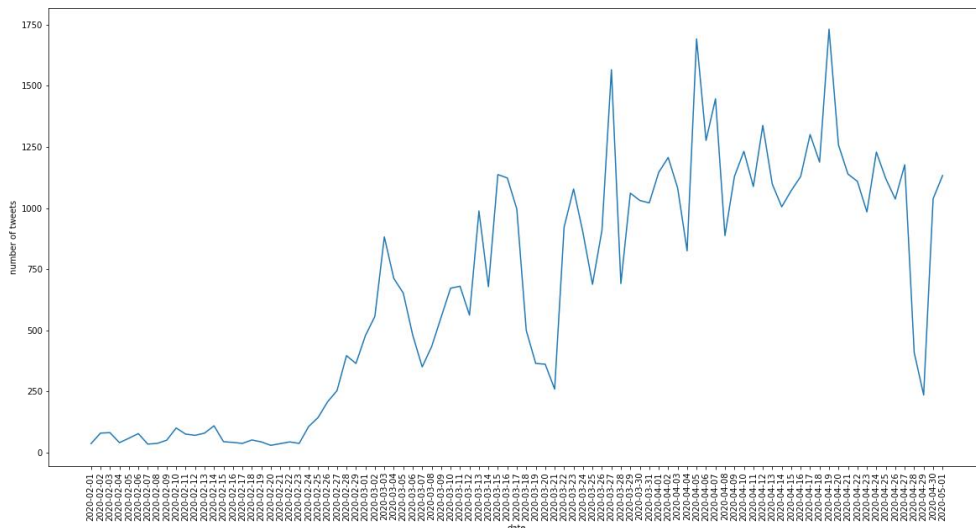
Εικόνα 58: Χρονική ανάλυση του υποθέματος 4 του topic 13 του LDA.

Το θέμα 4 αφορά τα test ανίχνευσης του ιού και ενδεχομένως την έλλειψη τους (Εικόνα 58).

- 12 Μαρτίου - 13 Μαρτίου: Η κυβέρνηση στις 12 Μαρτίου δηλώνει ότι δεν θα παρακολουθεί-ανιχνεύει όσους έχουν υποψίες νόσησης. Αντίθετα τα διαγνωστικά test θα περιορίζονται στους ασθενείς των νοσοκομείων με αναπνευστικά προβλήματα. Την επόμενη μέρα, 13 Μαρτίου, ο ΠΟΥ προτρέπει όλες τις χώρες να να βρουν και να υποβάλλουν σε test ανίχνευσης κάθε κρούσμα, ώστε να περιοριστεί όσο το δυνατόν γίνεται, η πανδημία (Boseley, 2020). Σύμφωνα με το διάγραμμα ο αριθμός των tweets είναι υψηλός, περίπου 2.000, αποδεικνύοντας τη δημόσια συζήτηση και ίσως τα αντιφατικά μηνύματα που προκύπτουν από τις παραπάνω δηλώσεις.
- 18 Μαρτίου: Σύμφωνα με τον Boris Johnson ο αριθμός των test ανίχνευσης του κορονοϊού θα αυξηθεί σε 25.000 τη μέρα, με το προσωπικό υγείας να έχει άμεση προτεραιότητα. Η δήλωση ακολουθεί τις πιέσεις του ιατρονοσηλευτικού προσωπικού, οι οποίοι ζητάνε να υποβάλλονται σε test, καθώς είναι οι πιο άμεσα εκτεθειμένοι (Boseley, 2020). Ο αριθμός των tweets είναι χαμηλός, και συγκεκριμένα λίγο παραπάνω από 500.
- 23 Μαρτίου μέχρι 27 Απριλίου: Παρατηρώντας το διάγραμμα διαπιστώνεται ότι από τις 23 Μαρτίου ο αριθμός των tweets ακολουθεί μια σημαντικά μεγάλη αυξητική τάση, η οποία για όλο το υπόλοιπο διάστημα παραμένει σε πολύ υψηλά επίπεδα. Ο αριθμός των tweets κατά μέσο όρο κυμαίνεται στις 3.000, ενώ κάποιες ημέρες ξεπερνάει τα 3.500. Στη Βρετανία η συζήτηση για test ανίχνευσης ήταν πολύ έντονη, καθώς ασκήθηκαν έντονες πιέσεις ώστε να έχουν πρόσβαση και να μπορούν να ελεγχθούν περισσότεροι πολίτες. Αυτό αποδεικνύεται από δημοσιεύματα εκείνης της χρονικής περιόδου. Συνεπώς με βάση αυτό δεδομένο, είναι απολύτως λογικό ο αριθμός των tweets να είναι υψηλός το συγκεκριμένο χρονικό διάστημα. Ακολουθούν οι σημαντικότερες επιμέρους ειδήσεις ανάμεσα σ' αυτό το χρονικό διάστημα.

- 30 Μαρτίου: Σύμφωνα με δημοσίευμα του Guardian οι αρμόδιοι υπουργοί είναι εκτεθειμένοι καθώς ο στόχος για 10.000 test όχι μόνο δεν επετεύχθη, αλλά έγινε προσπάθεια διαστρέβλωσης της πραγματικότητας. Συγκεκριμένα ο Υπουργός Οικισμού, Κοινοτήτων και Τοπικής Αυτοδιοίκησης του Ηνωμένου Βασιλείου, Michael Gove, δήλωσε την επίτευξη του στόχου, ενώ στη πραγματικότητα πραγματοποιήθηκαν μόλις 9.114 test, εκ των οποίων πολλά από αυτά στους ίδιους ασθενείς. Τέλος το άρθρο αναφέρεται στην αδυναμία του Ηνωμένου Βασιλείου να ακολουθήσει το παράδειγμα άλλων χωρών, όπως η Γερμανία η οποία διεξάγει 70.000 test την ημέρα (Mason, 2020).
- 2 Απριλίου: Ο Υπουργός Υγείας Matt Hancock, θέτει ως κυβερνητικό στόχο τη διενέργεια 100.000 test την ημέρα μέχρι το τέλος του μήνα. Η δήλωση γίνεται ύστερα από έντονες επικρίσεις που δέχτηκε τις προηγούμενες ημέρες (Syal, 2020).
- 3 Απριλίου: Ο Guardian δημοσιεύει άρθρο, το οποίο παραθέτει το ιστορικό των δηλώσεων όλων των αρμόδιων υπουργών, και υποστηρίζει ότι η κυβέρνηση υποσχέθηκε πάρα πολλά και έκανε πολύ λίγα (Perraudin and Duncan, 2020).
- 1 Μαΐου: Ο Matt Hancock ισχυρίζεται ότι ο στόχος των 100.000 test έχει επιτευχθεί, εν μέσω αντιδράσεων που ισχυρίζονται ότι ο αριθμός είναι πλασματικός. Ο αριθμός των tweets παραμένει υψηλός, καθώς είναι λίγο λιγότερος από 3.000.

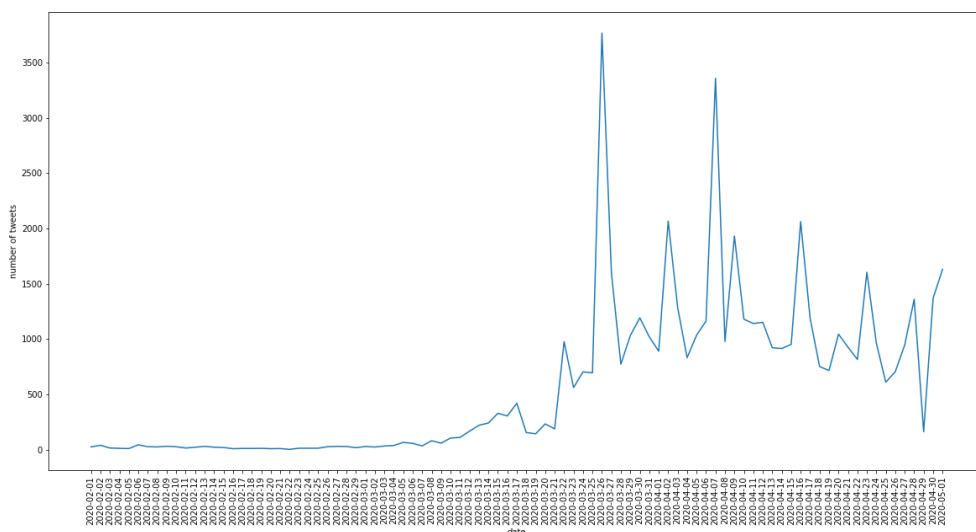
Από το θέμα 14 του LDA το θέμα 0, το θέμα 1 και το θέμα 4 εμφανίζουν το μεγαλύτερο ενδιαφέρον.



Εικόνα 59: Χρονική ανάλυση του υποθέματος 0 του topic 14 του LDA.

Το θέμα 0 αφορά τον Boris Johnson (Εικόνα 59).

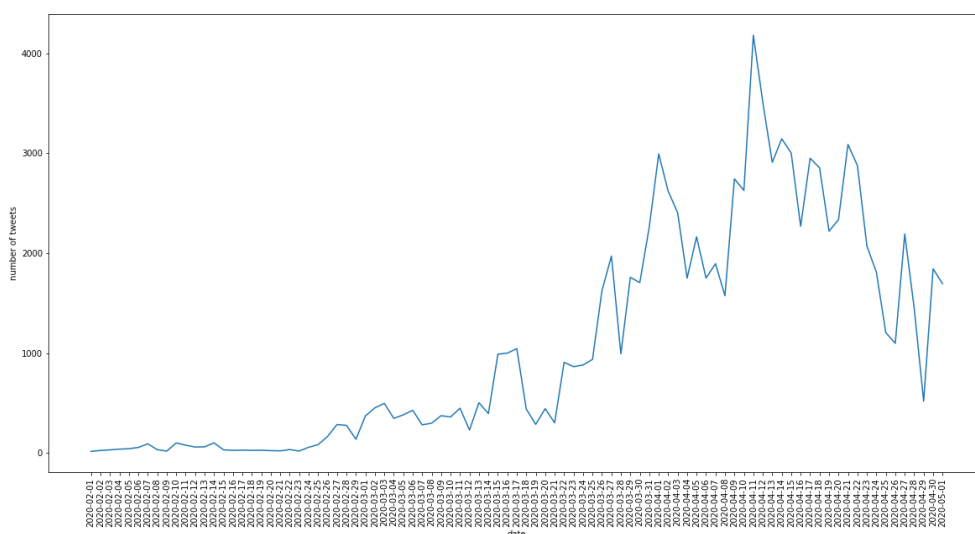
- 2 Μαρτίου: Ο Boris Johnson παρευρίσκεται σε έκτακτη συνεδρίαση της επιτροπής Cobra, για να οριστικοποιήσει τις κυβερνητικές αποφάσεις, εν μέσω ανακοίνωσης 36 νέων κρουσμάτων (Patterson, 2020). Ο αριθμός των tweets παρουσιάζει αυξητική τάση αν και συγκεντρώνει λιγότερα από 750 tweets.
- 16 Μαρτίου: Ο πρωθυπουργός Boris Johnson σε διάγγελμα που απευθύνει στους Βρετανούς πολίτες μέσω του BBC, συμβουλεύει τον περιορισμό επαφών, την εργασία από το σπίτι και την αποφυγή κοινωνικών χώρων που έχουν χαρακτηριστεί ως εστίες μετάδοσης (“Coronavirus: PM says everyone should avoid office, pubs and travelling”). Ο αριθμός των tweets αυξάνεται ξεπερνώντας τα 1.000.
- 23 Μαρτίου: Ο Boris Johnson ανακοινώνει την επιβολή lockdown (“Coronavirus: Strict new curbs on life in UK announced by PM”). Ο αριθμός των tweets είναι περίπου 1.000.
- 27 Μαρτίου: Ο Πρωθυπουργός, Boris Johnson ανιχνεύεται θετικός στον κορονοϊό, με τον Boris Johnson να απομονώνεται στην Downing Street (“Coronavirus: Prime Minister Boris Johnson tests positive”). Ο αριθμός των tweets εκτοξεύεται πάνω από 1.500.
- 5 Απριλίου: Ο Boris Johnson δέκα ημέρες μετά το θετικό τεστ, εισάγεται στο νοσοκομείο για εξετάσεις (“Coronavirus: Boris Johnson admitted to hospital over virus symptoms”). Ο αριθμός των tweets αυξάνεται και φτάνει περίπου τα 1.750.
- 6 Απριλίου: Ο Boris Johnson μεταφέρεται στην εντατική, μετά από επιδείνωση της υγείας του (“Coronavirus: Boris Johnson moved to intensive care as symptoms worsen”). Ο αριθμός των tweets έχει λίγο μειωθεί και είναι περίπου 1.250.
- 12 Απριλίου: Ο Boris Johnson παίρνει εξιτήριο από το νοσοκομείο και ευχαριστεί το υγειονομικό προσωπικό που του έσωσε τη ζωή (“Coronavirus: Boris Johnson says 'it could have gone either way'”). Ο αριθμός των tweets παραμένει περίπου 1.250
- 30 Απριλίου: Ο Boris Johnson δηλώνει ότι θα ανακοινώσει ένα ολοκληρωμένο σχέδιο σταδιακής χαλάρωσης του lockdown την επόμενη εβδομάδα, το οποίο θα περιλαμβάνει άνοιγμα σχολείων (“Coronavirus: Boris Johnson says UK is past the peak of outbreak”). Μετά από ένα διάστημα μείωσης του αριθμού των tweets, παρατηρείται αύξηση, φτάνοντας περίπου τις 1250.



Εικόνα 60: Χρονική ανάλυση του υποθέματος 1 του topic 14 του LDA.

Το θέμα 1 αφορά την ευγνωμοσύνη στο προσωπικό του συστήματος υγείας (Εικόνα 60).

- 26 Μαρτίου: Άνθρωποι σε όλο το Ηνωμένο Βασίλειο συμμετέχουν σε μια κίνηση συμβολισμού, βγαίνοντας στα μπαλκόνια των σπιτιών τους στις 20:00 για να χειροκροτήσουν όλο το ιατρικό και νοσηλευτικό προσωπικό. Μια ένδειξη ευγνωμοσύνης και εκτίμησης (Patterson, 2020). Σύμφωνα με το διάγραμμα, ο αριθμός των tweets είναι πολύ υψηλός ξεπερνώντας τις 3.500, δείχνοντας ότι η συμμετοχή είναι πολύ υψηλή.
- 2 Απριλίου: Για δεύτερη φορά οι πολίτες του Ηνωμένου Βασιλείου εκφράζουν την ευγνωμοσύνη τους σε όλους τους εργαζόμενους του συστήματος υγείας (“Clap for Carers: UK applauds NHS staff and key workers”). Ο αριθμός των tweets αυξάνεται φτάνοντας τα 2.000.
- 9 Απριλίου: Για τρίτη φορά οι πολίτες του Ηνωμένου Βασιλείου δίνουν ραντεβού στις 20:00 χειροκροτώντας όλο το προσωπικό υγείας (“Nation applauds NHS workers fighting coronavirus for third week running”). Ο αριθμός των tweets αγγίζει περίπου τις 2.000.
- 16 Απριλίου: Τέταρτη φορά συμμετέχουν οι πολίτες εκφράζοντας τη συνεχή στήριξη τους, σε μια προσπάθεια να τους εμψυχώσουν (“Nation applauds NHS workers fighting coronavirus for third week running”). Ο αριθμός των tweets είναι περίπου 2.000
- 23 Απριλίου: Πέμπτη συνεχόμενη βδομάδα οι πολίτες δείχνουν την ευγνωμοσύνη τους στο καθιερωμένο ραντεβού (“UK thanks key coronavirus workers in nationwide Clap For Our Carers”). Ο αριθμός των tweets είναι 1.500.
- 30 Απριλίου: Έκτη συνεχόμενη εβδομάδα όπου οι πολίτες δίνουν την ίδια ώρα το παρόν, ευχαριστώντας το προσωπικό υγείας (“Coronavirus: UK's weekly clap for carers”). Ο αριθμός των tweets είναι λίγο λιγότερος από 1.500.



Εικόνα 61: Χρονική ανάλυση του υποθέματος 4 του topic 14 του LDA.

Το θέμα 4 αφορά τα νοσοκομεία και το εθνικό σύστημα υγείας της Βρετανίας (Εικόνα 61).

- 28 Μαρτίου μέχρι 22 Απριλίου: Παρατηρείται ότι από τις 28 Μαρτίου ο αριθμός των tweets ακολουθεί απότομη ανοδική τάση. Ο αριθμός διατηρείται σε υψηλά επίπεδα ως επί των πλείστων καθ' όλη τη διάρκεια, πάνω από 2.000 ενώ μεμονωμένες ημερομηνίες ξεπερνάει ή πλησιάζει τις 4.000. Αυτό το χρονικό διάστημα επικρατεί μια συζήτηση σχετικά με το πόσο επαρκής ποσοτικά είναι ο ιατρικός εξοπλισμός. Υπάρχουν αντιδράσεις από το προσωπικό υγείας, ότι υπάρχουν πολλές βασικές ελλείψεις και για τους ασθενείς και για τους ίδιους. Παρακάτω ακολουθούν οι σημαντικότερες επιμέρους ειδήσεις.
 - 1 Απριλίου: Ο Michael Gove δηλώνει ότι εξαιτίας της έλλειψης χημικών αντιδραστήρων που είναι απαραίτητα στα διαγνωστικά covid-test, δεν είναι δυνατός ο έλεγχος όλων των εργαζομένων υγείας, ο αριθμός των οποίων ανέρχεται σε 1.2 εκατομμύρια (“Coronavirus: Boris Johnson vows more virus tests as UK deaths exceed 2000”). Η Ένωση Χημικών Βιομηχανιών ανταπαντά διαψεύδοντας την ύπαρξη έλλειψης και αναφέροντας συγκεκριμένα ότι σε προηγούμενη συνάντησή τους με εκπρόσωπο της Βρετανικής κυβέρνησης εκείνος δεν προσπάθησε καν να μάθει αν υπάρχουν πιθανά προβλήματα εφοδιασμού (Peston, 2020). Ο αριθμός των tweets πλησιάζει τις 3.000.
 - 5 Απριλίου: Ο Matt Hancock δηλώνει ότι το σύστημα υγείας θα χρειαστεί 18.000 αναπνευστήρες με βάση το χειρότερο σενάριο, μειώνοντας κατά 1/3 τις προηγούμενες εκτιμήσεις (“/”).
 - 8 Απριλίου: Ο Matt Hancock ζητάει από τη βιομηχανία να βοηθήσει στην παραγωγή βιώσιμων test ανίχνευσης του ιού (Mason and Stewart, 2020).
 - 10 Απριλίου: Ο Matt Hancock δηλώνει εκ μέρους της κυβέρνησης, ότι οι εργαζόμενοι στο σύστημα υγείας θα λαμβάνουν καθημερινά τον απαραίτητο

ατομικό εξοπλισμό προστασίας (PPE). Σχετικά με τις αντιδράσεις και την κριτική των προηγούμενων ημερών, που αφορά τη προστασία του υγειονομικού προσωπικού, δηλώνει ότι καταβάλλεται ηράκλεια προσπάθεια (Roxby). Επιπλέον τονίζει ότι είναι ο ατομικός εξοπλισμός προστασίας (PPE) είναι πολύτιμος πόρος και πρέπει να χρησιμοποιηθεί υπεύθυνα. Η δήλωση του Hancock, περί προειδοποιήσεων μη υπερβολικής χρήσης των PPE, προκάλεσε την έντονη αντίδραση των εργαζομένων του συστήματος υγείας, απαιτώντας να ληφθούν όλα τα απαραίτητα μέτρα προστασίας τους (Stewart and Campbell, 2020).

- 12 Απριλίου: Σύμφωνα με τον Υπουργό Επιχειρήσεων, Alok Sharma, παραδέχτηκε ότι το ιατρονοσηλευτικό προσωπικό δεν πρέπει να αποφασίζει αν θα θεραπεύσει τους ασθενείς με κορονοϊό χωρίς τα απαραίτητα προστατευτικά μέσα (Mason, 2020).
- 17 Απριλίου: Σύμφωνα με δημοσίευμα του Guardian, η διοίκηση των νοσοκομείων ζητά από το ιατρονοσηλευτικό προσωπικό να εργαστεί χωρίς τις προστατευτικές ολόσωμες στολές, καθώς οι προμήθειες τελείωσαν (Campbell, 2020).
- 18 Απριλίου: Η κυβέρνηση παραδέχεται την έλλειψη σε ορισμένα είδη ατομικής προστασίας και υπόσχεται την άμεση παραλαβή υλικού από την Τουρκία (Blackall, 2020).
- 22 Απριλίου: Ο Υπουργός Υγείας επιβεβαίωσε ότι έφτασε η αποστολή ατομικού προστατευτικού υλικού από την Τουρκία (Mason and Sabbagh, 2020).

5. Σύνοψη και συμπεράσματα

Το πρώτο χρονικό διάστημα που ξέσπασε ο ιός, σημαδεύτηκε από πρωτόγνωρες καταστάσεις. Η καθημερινότητα όλων άλλαξε ριζικά, καθώς για λόγους υγείας, προσωπικής αλλά και των υπολοίπων, έπρεπε να εφαρμοστούν μέτρα, σκληρά αλλά απαραίτητα για τον περιορισμό της πανδημίας. Η κρισιμότητα της κατάστασης εντάθηκε από το ότι η παγκόσμια κοινότητα αιφνιδιάστηκε, καθώς κανένας δεν περίμενε τη ταχύτητα και τη σοβαρότητα με την οποία εξελίχθηκε η πανδημία. Παρ' όλα αυτά κάποιες κυβερνήσεις διέθεταν καλύτερα αντανακλαστικά από κάποιες άλλες.

Στη παρούσα εργασία επιλέχθηκε να μην πραγματοποιηθεί έρευνα γενικά για τον covid-19 αλλά να μελετήσει συγκεκριμένα τι έγινε στο Ηνωμένο Βασίλειο. Ο λόγος που επιλέχθηκε αφορά ότι στο πρώτο χρονικό διάστημα που ξέσπασε η πανδημία, η πολιτική ηγεσία της χώρας δέχθηκε δριμυία κριτική όσον αφορά τον τρόπο διαχείρισης της. Η παρούσα διπλωματική εργασία έρχεται να συμβάλει στη προσπάθεια κατανόησης του τι έγινε εκείνη τη χρονική περίοδο στο Ηνωμένο Βασίλειο. Για να απαντηθούν τα συγκεκριμένα ερωτήματα χρησιμοποιήθηκαν δεδομένα από το Twitter. Το Twitter είναι μια από τις δημοφιλέστερες πλατφόρμες κοινωνικής δικτύωσης, εξαιρετικά κατάλληλο για τέτοιου είδους αναλύσεις. Τα τελευταία χρόνια χρησιμοποιείται κατα κόρον για τη διεξαγωγή ποικίλων ερευνητικών θεμάτων, ιδιαίτερα σε περιόδους κρίσης. Η προσφορά τέτοιων αναλύσεων και ερευνών είναι ιδιαίτερα σημαντική και μπορεί να αξιοποιηθεί για μελλοντική διαχείριση κρίσεων.

Η παρούσα διπλωματική εργασία έρχεται να συμβάλει στο παραπάνω κενό της βιβλιογραφίας. Έπειτα από εκτεταμένη βιβλιογραφική μελέτη σχετικά με τον κορονοϊό, των τεχνικών που χρησιμοποιούνται στην ανάλυση κειμένου, προηγούμενων ερευνών που διεξήχθησαν, αποφασίστηκε ποιο σύνολο δεδομένων είναι κατάλληλο για τη διεξαγωγή της εργασίας. Στη συνέχεια πραγματοποιήθηκε η συλλογή και η ανάκτηση των δεδομένων μέσω του `twarc` και του Twitter API. Έπειτα από το σύνολο δεδομένων επιλέχθηκαν τα tweets που αφορούν το Ηνωμένο Βασίλειο, στα οποία υλοποιήθηκαν συγκεκριμένες τεχνικές προεπεξεργασίας, ώστε να μπορούν να εφαρμοστούν συγκεκριμένες μέθοδοι ανάλυσης και επεξεργασίας κειμένου. Στο σύνολο δεδομένων εφαρμόστηκαν δύο αλγόριθμοι θεματικής μοντελοποίησης, ο LDA και ο GSDMM. Σε πρώτο στάδιο εφαρμόστηκε ο LDA, εξάγοντας 15 θέματα. Ύστερα, καθαρά με βάση την υποκειμενική κρίση έγινε εκτίμηση των θεματικών πεδίων που κάθε θέμα αφορά. Αυτά που κρίθηκαν πιο ενδιαφέροντα αφορούν τα πεδία που σχετίζονται με την ατομική προστασία, την ανίχνευση του ιού, την εκπαίδευση, τις πολιτικές περιορισμού που εφαρμόστηκαν για τις μετακινήσεις, την οικονομία και το σύστημα υγείας. Για την καλύτερη κατανόηση τους και την αποσαφήνισή τους, εφαρμόστηκε ένας δεύτερος αλγόριθμος θεματικής μοντελοποίησης, ο GSDMM, στα συγκεκριμένα μόνο θέματα. Ο αριθμός των υπο-θεμάτων που εξήχθησαν ήταν από πέντε έως δέκα. Με τη βοήθεια ενός συγκεκριμένου τύπου μέτρησης του coherence score, επιλέχθηκε ο αριθμός των υπο-θεμάτων με το καλύτερο coherence score. Στη πορεία για κάθε θέμα του GSDMM, επιλέχθηκαν τα θέματα με το μεγαλύτερο ενδιαφέρον, αφού πρώτα κατανοήθηκαν και αποσαφηνίστηκαν. Έπειτα, πραγματοποιήθηκε μια χρονική ανάλυση. Πιο συγκεκριμένα παρουσιάστηκαν τα διαγράμματα, που αποτυπώνουν την εξέλιξη του αριθμού των tweets ενός θέματος σε συνάρτηση με τον χρόνο. Παράλληλα γίνεται μια καταγραφή των

κυριότερων ειδήσεων που σχετίζονται με κάθε θεματικό πεδίο ξεχωριστά, για να βρεθεί η ύπαρξη ή μη συσχετίσεων.

Με βάση τα παραπάνω εξήχθησαν ιδιαίτερα χρήσιμα συμπεράσματα. Συγκεκριμένα τα θέματα που απασχόλησαν τους Βρετανούς πολίτες πριν την ανακοίνωση του lockdown από την κυβέρνηση, ήταν η έλλειψη προϊόντων και βασικών αγαθών, ενώ στη συνέχεια φάνηκε να μην είναι ιδιαίτερα. Το κρουαζιερόπλοιο Diamond Princess, μονοπώλησε τον διεθνή τύπο και όπως όλα δείχνουν και το Ηνωμένο Βασίλειο, καθώς φαίνεται να συζητήθηκε πολύ έντονα στις αρχές. Κατά το χρονικό διάστημα μετά την ανακοίνωση του lockdown, ένα από τα βασικά θέματα συζήτησης ήταν η οικονομία, και ειδικότερα η ανάγκη στήριξης επιχειρήσεων και εργαζομένων. Επιπλέον η έλλειψη διαγνωστικών test ανίχνευσης και ατομικού προστατευτικού υλικού για το υγειονομικό προσωπικό ήταν πεδία έντονης συζήτησης, που αποτυπώθηκε και στον αγγλικό τύπο, γεγονός που δείχνει ότι το σύστημα υγείας δεν ενισχύθηκε όσο θα έπρεπε. Η μάσκα προστασίας υπήρξε αντικείμενο συζήτησης αν και όχι μόνο δεν ήταν υποχρεωτική αλλά δεν υπήρχαν καν συστάσεις για τη χρήση της. Το θέμα των σχολείων και της εκπαίδευσης φαίνεται να ήταν ένα από τα πεδία που απασχόλησαν τους Βρετανούς καθ' όλη τη διάρκεια. Όσον αφορά την παρότρυνση για προσωπική απομόνωση στο σπίτι, επίσης, παρατηρήθηκε ότι αρχικά ήταν πιο έντονη η συζήτηση, ενώ μετά από ένα χρονικό διάστημα, ξεκίνησε να φθίνει. Απ' αυτό συμπεραίνεται ότι οι Βρετανοί στην αρχή, ίσως ήταν πιο πρόθυμοι να απομονωθούν σπίτι τους, ώστε να περιοριστεί η πανδημία και να προστατευθούν, και στην πορεία μεγάλη μερίδα των πολιτών ξεκίνησε να δυσανασχετεί με τη μεγάλη διάρκεια του εγκλεισμού. Ο πρωθυπουργός της Αγγλίας, Boris Johnson ήταν ένας από τους πρωταγωνιστές της συγκεκριμένης χρονικής περιόδου όχι μόνο εξαιτίας της θέσης του, αλλά και γιατί εκείνο το χρονικό διάστημα, νόσησε.

5.1 Όρια και περιορισμοί της έρευνας

Στη παρούσα διπλωματική εργασία ένα από τα βασικά προβλήματα ήταν ότι εξαιτίας του πολύ μεγάλου όγκου των δεδομένων, η επεξεργασία τους πολλές φορές δεν μπορούσε να ολοκληρωθεί, λόγω των απαιτήσεων του συστήματος. Πιο συγκεκριμένα μια εναλλακτική ιδέα ήταν η εφαρμογή ενός άλλου αλγόριθμου θεματικής μοντελοποίησης, του BERTopic, αντί για τον GSDMM. Ωστόσο επειδή ο όγκος των δεδομένων ήταν πάρα πολύ μεγάλος, και το υπολογιστικό σύστημα που χρησιμοποιήθηκε δεν μπορούσε να ανταπεξέλθει, το αρχικό πλάνο δεν υλοποιήθηκε. Επιπλέον μια σημαντική καθυστέρηση προέκυψε στην ανάκτηση ολόκληρου του περιεχομένου των δεδομένων, μέσα από τη χρήση της βιβλιοθήκης `twarc` κατά το στάδιο συλλογής των δεδομένων, εξαιτίας του μεγέθους των δεδομένων.

5.2 Μελλοντικές Επεκτάσεις

Μια πολύ σημαντική μελλοντική προέκταση της εργασίας, θα ήταν η ανάλυση συναισθήματος. Πιο συγκεκριμένα, μελλοντικά μπορούν να χρησιμοποιηθούν διάφορες τεχνικές είτε επιβλεπόμενης, είτε μη επιβλεπόμενης μηχανικής μάθησης για την ανίχνευση συναισθημάτων σε συγκεκριμένα θεματικά πεδία. Η επιπλέον προσθήκη αυτού του είδους της ανάλυσης θα προσφέρει ακόμα περισσότερα εργαλεία στην απάντηση κρίσιμων

ερωτημάτων, όπως για παράδειγμα αν οι Βρετανοί πολίτες ήταν θετικά ή αρνητικά προσκείμενοι σε μια σειρά από περιοριστικά μέτρα, παρέμειναν σταθεροί, ή άλλαξε με βάση το χρόνο; Αυτά είναι κάποια ενδεικτικά ερωτήματα που αξίζει να απαντηθούν στο μέλλον, ως συνέχιση της συγκεκριμένης εργασίας.

Βιβλιογραφία

- Documenting the Now*, <https://www.docnow.io/>. Accessed 17 June 2022. “//.” // -
Wikipedia, <https://www.hsj.co.uk/workforce/nhs-needs-a-third-fewer-ventilators-than-forecast-says-hancock/7027322.article>. Accessed 16 June 2022.
- ΣΚΑΙ. “Ιταλίδα γιατρός δολοφονήθηκε από τον σύντροφό της γιατί νόμιζε ότι τον κόλλησε κορωνοϊό.” <https://www.skai.gr/news/world/italida-giatros-dolofonithike-apo-ton-syntrofo-tis-giati-nomize-oti-ton-kollise-koronoio>, <https://www.skai.gr/news/world/italida-giatros-dolofonithike-apo-ton-syntrofo-tis-giati-nomize-oti-ton-kollise-koronoio>.
- Σολομωνίδου, Χριστίνα. *Νέες τάσεις στην εκπαιδευτική τεχνολογία. Επικοινωνιακός και σύγχρονα περιβάλλοντα μάθησης*. Μεταίχμιο, 2006. Τριλιανός, Θανάσης. *Μεθοδολογία της Σύγχρονης Διδασκαλίας: Καινοτόμες επιστημονικές προσεγγίσεις στη διδακτική πράξη*. 2008.
- Βέργου, Ντάνι, et al. “Σοκάρουν τα νούμερα της ύφεσης.” *Η Εφημερίδα των Συντακτών*, 21 03 2021, https://www.efsyn.gr/themata/thema-tis-efsyn/286452_sokaroygn-ta-noymera-tis-yfesis.
- Γεωργακώδης, Φ, and Κ Τσίμπος. “Περιγραφική και διερευνητική στατιστική ανάλυση δεδομένων.” Σταμούλης, 1999.
- Abeer, Abuzayed, and Hend Al-Khalifa. “BERT for Arabic Topic Modeling: An Experimental Study on BERTopic Technique.” *Procedia Computer Science*, vol. 189, 2021.
- Abuzayed, Abeer, and Hend Al Khalifa. “BERT for Arabic topic modeling: an experimental study on BERTopic technique.” *Procedia Computer Science*, vol. 189, 2021.
- Ahuja, S., and D. Dubey. “Clustering and sentiment analysis on Twitter data.” *2nd International Conference on Telecommunication and Networks*, 2017, pp. 1-5.
- Amrouche, Matyas. “Short Text Topic Modeling. Intuition and (some) maths to... | by Matyas Amrouche.” *Towards Data Science*, 22 August 2019, <https://towardsdatascience.com/short-text-topic-modeling-70e50a57c883>. Accessed 18 June 2022.
- Andrea, Eleonora D., et al. “Monitoring the public opinion about the vaccination topic from

- tweets analysis.” *Expert Systems with Applications*, vol. 116, 2019.
- Antonakaki, Despoina, et al. “A survey of Twitter research: Data model, graph structure, sentiment analysis and attacks.” *Expert Systems with Applications*, vol.164, February 2021.
- Austin, Katy. “Coronavirus: Supermarkets ask shoppers to be 'considerate' and stop stockpiling.” *BBC*, 15 March 2020, <https://www.bbc.com/news/business-51883440>. Accessed 13 June 2022.
- Azad, Amin. “Twitter Topic Modeling. Using Machine Learning (Gensim Linear...|by Amin Azad.” *Towards Data Science*, 12 July 2020, <https://towardsdatascience.com/twitter-topic-modeling-e0e3315b12e2>. Accessed 1 April 2022.
- “BBC charity partners to celebrate the UK's community spirit and resilience with TheBig Night In on BBC One.” *BBC*, 7 April 2020, <https://www.bbc.co.uk/mediacentre/latestnews/2020/the-big-night-in>. Accessed 16 June 2022.
- Binns, Colin, et al. “The COVID-19 Pandemic: Public Health and Epidemiology.” *AsiaPacific Journal of Public Health*, May 19, 2020, pp. 140-144.
- Blackall, Molly. “PPE including gowns and masks running out, admits UK government.” *The Guardian*, 18 April 2020, <https://www.theguardian.com/world/2020/apr/18/nhs-frontline-staff-may-refuse-to-work-over-lack-of-coronavirus-ppe-says-union-unison>. Accessed 16 June 2022.
- Boseley, Sarah. “UK coronavirus testing will increase to 25000 a day, says Boris Johnson.” *The Guardian*, 18 March 2020. <https://www.theguardian.com/politics/2020/mar/18/uk-coronavirus-testing-will-increase-to-25000-a-day-says-boris-johnson>. Accessed 16 June 2022.
- Boseley, Sarah. “WHO urges countries to 'track and trace' every Covid-19 case.” *The Guardian*, 13 March 2020, <https://www.theguardian.com/world/2020/mar/13/who-urges-countries-to-track-and-trace-every-covid-19-case>. Accessed 16 June 2022.
- Bowers, Brittany. “A Unique Approach to Short Text Clustering (Algorithmic Theory).” *Towards Data Science*, 24 January 2020, <https://towardsdatascience.com/a-unique->

approach-to-short-text-clustering-part-1-algorithmic-theory-4d4fad0882e1. Accessed 18 June 2022.

Butler, Sarah. "Leon sets up initiative to deliver free meals to NHS critical care staff." *The Guardian*, 27 March 2020, <https://www.theguardian.com/business/2020/mar/27/leon-sets-up-initiative-to-deliver-free-hot-meals-to-nhs-staff-coronavirus>. Accessed 16 June 2022.

Butt, Isabel, et al. "Prolonged confusional state as first manifestation of COVID-19." vol. *Annals of Clinical and Translational Neurology*, 20 May 2020, pp. 1450-1452.

Campbell, Denis. "NHS staff told 'wear aprons' as protective gowns run out." *The Guardian*, 17 April 2020, <https://www.theguardian.com/world/2020/apr/17/nhs-staff-to-be-asked-to-treat-coronavirus-patients-without-gowns>. Accessed 16 June 2022.

Carvalho, Thiago, et al. "The first 12 months of COVID-19: a timeline of immunological insights." *Nature Reviews Immunology*, 15 March 2021, p. 12.

Catherine, Ordun, et al. "Exploratory analysis of covid-19 tweets using topic modeling, umap, and digraphs." *arXiv preprint arXiv*, 2020.

Chandrasekaran, and Ranganathan. "Topics, trends, and sentiments of tweets about the COVID-19 pandemic: Temporal infoveillance study." *Journal of medical Internet research*, vol. 22.10, 2020.

Ciotti, Marco, et al. "The COVID-19 pandemic." *Invited Review Articles*, 09 Jul 2020, pp. 365-388.

"Clap for Carers: UK applauds NHS staff and key workers." *BBC*, <https://www.bbc.com/news/av/uk-52143223>. Accessed 16 June 2022.

"Coronavirus: Boris Johnson admitted to hospital over virus symptoms." *BBC*, 6 April 2020, <https://www.bbc.com/news/uk-52177125>. Accessed 14 June 2022.

"Coronavirus: Boris Johnson moved to intensive care as symptoms worsen." *BBC*, 7 April 2020, <https://www.bbc.com/news/uk-52192604>. Accessed 14 June 2022.

"Coronavirus: Boris Johnson says 'it could have gone either way.'" *BBC*, 12 April 2020, <https://www.bbc.com/news/uk-politics-52262012>. Accessed 14 June 2022.

"Coronavirus: Boris Johnson says UK is past the peak of outbreak." *BBC*, 30 April 2020,

- <https://www.bbc.com/news/uk-52493500>. Accessed 14 June 2022.
- “Coronavirus: Boris Johnson vows more virus tests as UK deaths exceed 2000.” *BBC*, 1 April 2020, <https://www.bbc.com/news/uk-52122761>. Accessed 15 June 2022.
- “Coronavirus: Capt Tom Moore's NHS fundraiser hits £17m.” *BBC*, 17 April 2020, <https://www.bbc.com/news/uk-england-beds-bucks-herts-52303859>. Accessed 16 June 2022.
- “Coronavirus cases in UK rise to 206 today.” *The Independent*, 7 March 2020, <https://www.independent.co.uk/news/health/coronavirus-uk-cases-number-206-locations-latest-a9384586.html>. Accessed 15 June 2022.
- “Coronavirus: Chancellor unveils £350bn lifeline for economy.” *BBC*, 17 March 2020, <https://www.bbc.com/news/business-51935467>. Accessed 15 June 2022.
- “Coronavirus: Coachella, Radio 1's Big Weekend, BTS and other music events affected - CBBC Newsround.” *BBC*, <https://www.bbc.co.uk/newsround/51672257>. Accessed 16 June 2022
- “Coronavirus disease (COVID-19) pandemic.” *WHO | World Health Organization*, <https://www.who.int/emergencies/diseases/novel-coronavirus-2019>. Accessed 5 June 2022.
- “Coronavirus: Foreign Office links up with airlines to fly stranded Britons home.” *BBC*, 30 March 2020, <https://www.bbc.com/news/uk-52099252>. Accessed 15 June 2022.
- “Coronavirus: Four new UK cases among ship evacuees.” *BBC*, 23 February 2020, <https://www.bbc.com/news/uk-51606368>. Accessed 13 June 2022.
- “Coronavirus: Government to pay up to 80% of workers' wages.” *BBC*, 20 March 2020, <https://www.bbc.com/news/business-51982005>. Accessed 15 June 2022.
- “Coronavirus: Health secretary Matt Hancock tests positive | The Independent.” *The Independent*, 27 March 2020, <https://www.independent.co.uk/news/uk/politics/coronavirus-matt-hancock-boris-johnson-test-positive-covid-19-symptoms-a9430031.html>. Accessed 14 June 2022.
- “Coronavirus: Isolation for over-70s 'within weeks.'” *BBC*, 15 March 2020,

<https://www.bbc.com/news/uk-51895873>. Accessed 14 June 2022.

“Coronavirus: Japan cruise ship's US passengers home for further quarantine.” *BBC*, 17 February 2020, <https://www.bbc.com/news/world-asia-china-51524460>. Accessed 15 June 2022.

“Coronavirus: Matt Hancock sets aim of 100000 tests a day by end of April.” *BBC*, 2 April 2020, <https://www.bbc.com/news/uk-52140376>. Accessed 15 June 2022.

“Coronavirus: Mobile networks send 'stay at home' text.” *BBC*, 24 March 2020, <https://www.bbc.com/news/technology-52017451>. Accessed 14 June 2022.

“Coronavirus: More than 140000 firms claim wage bill help.” *BBC*, 20 April 2020, <https://www.bbc.com/news/business-52346685>. Accessed 15 June 2022.

“Coronavirus: No date for when schools will reopen.” *BBC*, 19 April 2020, <https://www.bbc.com/news/uk-52332903>. Accessed 15 June 2022.

“Coronavirus: Number of UK deaths rises above 1000.” *BBC*, 28 March 2020, <https://www.bbc.com/news/uk-52077997>. Accessed 14 June 2022.

“Coronavirus: People with fever or 'continuous' cough told to self-isolate.” *BBC*, 12 March 2020, <https://www.bbc.com/news/uk-51857856>. Accessed 14 June 2022.

“Coronavirus: PM says everyone should avoid office, pubs and travelling.” *BBC*, 16 March 2020, <https://www.bbc.com/news/uk-51917562>. Accessed 14 June 2022.

“Coronavirus: Prime Minister Boris Johnson tests positive.” *BBC*, 27 March 2020, <https://www.bbc.com/news/uk-52060791>. Accessed 14 June 2022.

“Coronavirus: Queen tells UK 'we will succeed' in fight.” *BBC*, 5 April 2020, <https://www.bbc.com/news/uk-52176222>. Accessed 14 June 2022.

“Coronavirus: Raab urges UK public not to ruin lockdown progress.” *BBC*, 9 April 2020, <https://www.bbc.com/news/uk-52235492>. Accessed 14 June 2022.

“Coronavirus: Shoppers told to buy responsibly.” *BBC*, 21 March 2020, <https://www.bbc.com/news/uk-51989721>. Accessed 13 June 2022.

“Coronavirus: Staying home this weekend 'not a request', UK told.” *BBC*, 3 April 2020, <https://www.bbc.com/news/uk-52155430>. Accessed 14 June 2022.

“Coronavirus: Strict new curbs on life in UK announced by PM.” *BBC*, 24 March 2020, <https://www.bbc.com/news/uk-52012432>. Accessed 14 June 2022.

“Coronavirus testing extended to all essential workers in England who have symptoms.” *GOV.UK*, 23 April 2020, <https://www.gov.uk/government/news/coronavirus-testing-extended-to-all-essential-workers-in-england-who-have-symptoms>. Accessed 15 June 2022.

“Coronavirus: Test website closes after 'significant demand.'” *BBC*, 24 April 2020, <https://www.bbc.com/news/uk-52405852>. Accessed 15 June 2022.

“Coronavirus: Third UK patient 'caught coronavirus in Singapore.’” *BBC*, 6 February 2020, <https://www.bbc.com/news/uk-51398039>. Accessed 16 June 2022.

“Coronavirus: Two cases confirmed in UK.” *BBC*, 31 January 2020, <https://www.bbc.com/news/health-51325192>. Accessed 16 June 2022.

“Coronavirus: UK government unveils aid for self-employed.” *BBC*, 26 March 2020, <https://www.bbc.com/news/uk-52053914>. Accessed 15 June 2022.

“Coronavirus: UK lockdown extended for 'at least' three weeks.” *BBC*, 16 April 2020, <https://www.bbc.com/news/uk-52313715>. Accessed 15 June 2022.

“Coronavirus: UK's weekly clap for carers.” *BBC*, <https://www.bbc.com/news/av/uk-52494308>. Accessed 16 June 2022.

“Coronavirus: UK to remain in 'containment' phase of response.” *BBC*, 9 March 2020, <https://www.bbc.com/news/uk-51796072>. Accessed 13 June 2022.

“Coronavirus: Woman in 70s becomes first virus fatality in UK.” *BBC*, 5 March 2020, <https://www.bbc.com/news/uk-51759602>. Accessed 15 June 2022.

Corsi, Alana, et al. “Big data analytics as a tool for fighting pandemics: a systematic review of literature.” *Journal of Ambient Intelligence and Humanized Computing*, 2020.

“COVID-19 pandemic on Diamond Princess.” *Wikipedia*,
https://en.wikipedia.org/wiki/COVID-19_pandemic_on_Diamond_Princess.
Accessed 14 June 2022.

“Diamond Princess Updates - Notices & Advisories.” *Princess Cruises*,
https://www.princess.com/news/notices_and_advisories/notices/diamond-princess-update.html. Accessed 15 June 2022.

Dickinson, Daniel. “COVID-19: UN chief calls for global ceasefire to focus on ‘the true fight of our lives.’” *UN News*, 23 March 2020,
<https://news.un.org/en/story/2020/03/1059972>. Accessed 15 June 2022.

Dimitriadis, Nikolaos S. “Applying Topic Modelling Algorithms on Twitter messages in Greek language.” *Graduate Thesis*, vol. Aristotle University of Thessaloniki, 2020.

Dodd, Vikram. “England: police to get power to use force to impose coronavirus lockdown.” *The Guardian*, 25 March 2020, <https://www.theguardian.com/uk-news/2020/mar/25/coronavirus-lockdown-will-be-imposed-by-force-in-england>.
Accessed 14 June 2022.

Dwyer, Dominic B, et al. “Machine learning approaches for clinical psychology and psychiatry.” *Annual review of clinical psychology*, 2018, pp. 91-118.

Ebeling, Regis, et al. “The effect of political polarization on social distance stances in the Brazilian COVID-19 scenario.” *Journal of Information and Data Management*, 2021.

Elia, Francesco. “Stemming vs Lemmatization.” *Baeldung*, 24 June 2020,
<https://www.baeldung.com/cs/stemming-vs-lemmatization>. Accessed 19 March 2022.

Elliott, Larry. “Government extends loan guarantee scheme to cover bigger firms.” *The Guardian*, 16 April 2020,
<https://www.theguardian.com/world/2020/apr/16/government-extends-loan->

guarantee-scheme-to-cover-bigger-firms. Accessed 15 June 2022.

Elliott, Larry. "UK business survey finds firms struggling to access emergency cash."

The Guardian, 7 April 2020, [https://www.theguardian.com/business/2020/apr/08/uk-](https://www.theguardian.com/business/2020/apr/08/uk-business-survey-finds-firms-struggling-to-access-emergency-cash)

[business-survey-finds-firms-struggling-to-access-emergency-cash](https://www.theguardian.com/business/2020/apr/08/uk-business-survey-finds-firms-struggling-to-access-emergency-cash). Accessed 15

June 2022.

Erskine, Natalie, and Sharief Hendricks. "The Use of Twitter by Medical Journals:

Systematic Review of the Literature." *Journal of Medical Internet Research*, vol. 23, no. 7,

2021.

Faleiro, Sonia. "How do you self-isolate when you live on the street?" *POLITICO*, 21 March

2020, <https://www.politico.eu/article/coronavirus-homeless-self-isolation/>. Accessed 16

June 2022.

Fayyad, Usama M., et al. "Knowledge Discovery and Data Mining: Towards a Unifying

Framework." *KDD.*, vol. Vol. 96, 1996.

Feldman, Ronen. "Techniques and applications for sentiment analysis." *Communications of*

the ACM, vol. 56, no. 4, 2013, pp. 82-89.

Garger, K. "Illinois couple dead in murder-suicide after man feared they had coronavirus."

<https://nypost.com/2020/04/07/illinois-couple-dead-after-man-feared-they-had-covid-19/>,

New York Post, 7 April 2020. [https://nypost.com/2020/04/07/illinois-couple-dead-after-](https://nypost.com/2020/04/07/illinois-couple-dead-after-man-feared-they-had-covid-19/)

[man-feared-they-had-covid- 19/](https://nypost.com/2020/04/07/illinois-couple-dead-after-man-feared-they-had-covid-19/). Accessed 7 April 2020.

Goodley, Simon. "UK triples coronavirus response fund for NHS and public services." *The*

Guardian, 13 April 2020, [https://www.theguardian.com/society/2020/apr/13/uk-](https://www.theguardian.com/society/2020/apr/13/uk-triples-coronavirus-response-fund-for-nhs-and-public-services)

[triples-](https://www.theguardian.com/society/2020/apr/13/uk-triples-coronavirus-response-fund-for-nhs-and-public-services)

[coronavirus-response-fund-for-nhs-and-public-services](https://www.theguardian.com/society/2020/apr/13/uk-triples-coronavirus-response-fund-for-nhs-and-public-services). Accessed 15 June 2022.

Gupta, et al. "Global reactions to covid-19 on twitter: A labelled dataset with latent topic,

sentiment and emotion attributes." 2021.

"Hancock confirms 100,000 coronavirus testing target met." *ITV*, 1 May 2020,

[https://www.itv.com/news/2020-05-01/coronavirus-daily-briefing-matt-hancock-steve-](https://www.itv.com/news/2020-05-01/coronavirus-daily-briefing-matt-hancock-steve-powis-testing-tracing)

[powis-testing-tracing](https://www.itv.com/news/2020-05-01/coronavirus-daily-briefing-matt-hancock-steve-powis-testing-tracing). Accessed 15 June 2022.

Harapan, Biyan Nathanael, and Hyeon Joo Yoo. "Neurological symptoms, manifestations,

and complications associated with severe acute respiratory syndrome coronavirus 2

- (SARS- CoV- 2) and coronavirus disease 19 (COVID- 19).” *Journal of Neurology*, 12 January 2021.
- “The Health Protection (Coronavirus, Restrictions) (England) Regulations 2020.” *The Health Protection (Coronavirus, Restrictions) (England) Regulations 2020*, <https://www.legislation.gov.uk/ukSI/2020/350/regulation/6/2020-03-26>. Accessed 15 June 2022.
- Hern, Alex, and Kari Paul. “Apple and Google team up in bid to use smartphones to track coronavirus spread.” *The Guardian*, 10 April 2020, <https://www.theguardian.com/world/2020/apr/10/apple-google-coronavirus-us-app-privacy>. Accessed 16 June 2022.
- Honein, Margaret A., et al. “Summary of Guidance for Public Health Strategies to Address High Levels of Community Transmission of SARS-CoV-2 and Related Deaths, December 2020.” *MMWR Morb Mortal Wkly Rep.*, 2020 Dec 11, 1860-1867.
- “How many people are homeless in the UK? And what can you do about it?” *The Big Issue*, 29 April 2022, <https://www.bigissue.com/news/housing/how-many-people-are-homeless-in-the-uk-and-what-can-you-do-about-it/>. Accessed 16 June 2022. Humphries, Will.
- “Fears that Cheltenham Festival may have spread coronavirus throughout country.” *The Times*, 3 April 2020, <https://www.thetimes.co.uk/article/cheltenham-festival-spread-coronavirus-across-country-vbzmn5p9q>. Accessed 16 June 2022.
- Hung, Man, et al. “Social network analysis of COVID-19 sentiments: Application of artificial intelligence.” *Journal of medical Internet research*, vol. 22.8, 2020.
- Iniesta, Raquel, et al. “Machine learning, statistical learning and the future of biological research in psychiatry.” *Psychological medicine*, vol. 46.12, 2016, pp.2455-2465.
- Inman, Phillip. “Seven out of 10 UK firms have furloughed staff, survey reveals.” *The Guardian*, 21 April 2020, <https://www.theguardian.com/business/2020/apr/21/uk-unemployment-coronavirus-figures-jobs>. Accessed 16 June 2022.
- Jacob, L., et al. “Alcohol use and mental health during COVID-19 lockdown: a cross-sectional study in a sample of UK adults.” *Drug Alcohol Depend*, 2020.
- Jaspreet, Singh. “COVID-19 and Its Impact on Society.” *Electronic Research Journal of*

Social Sciences and Humanities, vol. Vol 2, no. Issue I, 7 Apr 2020.

Javid, Babak. "Across the world, face masks are becoming mandatory. Why not in the UK? |

Babak Javid." *The Guardian*, 24 April 2020,

<https://www.theguardian.com/commentisfree/2020/apr/24/face-masks-mandatory-spread-coronavirus-government>. Accessed 16 June 2022.

Johnson, Daniel. "Stemming and Lemmatization in Python NLTK with Examples." *Guru99*,

14 May 2022, <https://www.guru99.com/stemming-lemmatization-python-nltk.html>.

Accessed 18 June 2022.

Kapadia, Shashank. "Evaluate Topic Models: Latent Dirichlet Allocation (LDA) | by

Shashank Kapadia." *Towards Data Science*, 19 August 2019,

<https://towardsdatascience.com/evaluate-topic-model-in-python-latent-dirichlet-allocation-lda-7d57484bb5d0>. Accessed 28 March 2022.

Kaplan, Andreas M, and Michael Haenlein. "Users of the world, unite! The challenges

and opportunities of Social Media." *Business horizons*, vol. 53.1, 2010, pp. 59-68.

Kemp, Simon. "Digital in 2017 global overview." *We Are Social & Hootsuite*, 2017.

Khan, Hikmat Ullah, et al. "Twitter trends: A ranking algorithm analysis on real time

data." *Expert Systems with Applications*, vol. 164, February 2021.

Koyel, Chakraborty, et al. "Sentiment Analysis of COVID-19 tweets by Deep Learning

Classifiers—A study to show how popularity is affecting accuracy in social media."

Applied Soft Computing 97, 2020.

"Lemmatisation." *Wikipedia*, <https://en.wikipedia.org/wiki/Lemmatisation>. Accessed 18

March 2022.

Li S., et al. "The impact of COVID-19 epidemic declaration on psychological

consequences: a study on active Weibo users." *Int J Environ Res Public Health*,

2020.

"Love Supreme Jazz Festival 2020 has been postponed | Music News - Jazz FM."

Planet Radio, 15 April 2020, [https://planetradio.co.uk/jazz-fm/news/music-](https://planetradio.co.uk/jazz-fm/news/music-news/love-supreme-festival-2020/)

[news/love-supreme-festival-2020/](https://planetradio.co.uk/jazz-fm/news/music-news/love-supreme-festival-2020/). Accessed 16 June 2022.

- Mahase, Elisabeth. "Covid-19: UK holds off closing schools and restricts testing to people in hospital." *BMJ: British Medical Journal (Online)*, 2020.
- Makortoff, Kalyeena. "Just £1.1bn of Covid-19 bailout loans have been issued to small UK firms." *The Guardian*, 15 April 2020,
<https://www.theguardian.com/business/2020/apr/15/covid-19-bailout-loans-issued-uk-firms-banks>. Accessed 16 June 2022.
- Makortoff, Kalyeena. "UK Treasury rushes to finalise coronavirus bounce back loan scheme." *The Guardian*, 30 April 2020,
<https://www.theguardian.com/business/2020/apr/30/uk-treasury-rushes-to-finalise-coronavirus-bounce-back-loan-scheme>. Accessed 15 June 2022.
- Mamun, M. A., and M. D. Griffiths. "First COVID-19 suicide case in Bangladesh due to fear of COVID-19 and xenophobia: possible suicide prevention strategies." *Asian J Psychiatr*, vol. 51:102073, 2020.
- Man, Hung, et al. "Social network analysis of COVID-19 sentiments: Application of artificial intelligence." *Journal of medical Internet research*, vol. 22.8, 2020.
- Marsh, Sarah, and Denis Campbell. "Coronavirus: first UK death confirmed as cases surge to 116." *The Guardian*, 5 March 2020,
<https://www.theguardian.com/world/2020/mar/05/more-than-a-hundred-people-in-uk-infected-with-coronavirus>. Accessed 15 June 2022.
- Mason, Rowena. "Ministers must get PPE to NHS frontline, says business secretary." *The Guardian*, 12 April 2020,
<https://www.theguardian.com/society/2020/apr/12/coronavirus-ministers-must-get-ppe-to-nhs-frontline-says-business-secretary>. Accessed 16 June 2022.
- Mason, Rowena. "UK ministers accused of overstating scale of coronavirus testing." *The Guardian*, 30 March 2020,
<https://www.theguardian.com/world/2020/mar/30/uk-ministers-accused-of-overstating-scale-of-coronavirus-testing>. Accessed 16 June 2022.

Mason, Rowena, and Dan Sabbagh. "Plane with PPE from Turkey has landed in UK, minister confirms." *The Guardian*, 22 April 2020, <https://www.theguardian.com/world/2020/apr/22/delayed-plane-carrying-ppe-from-turkey-lands-in-uk-coronavirus>. Accessed 16 June 2022.

Mason, Rowena, and Heather Stewart. "Matt Hancock asks industry to help find viable coronavirus test." *The Guardian*, 8 April 2020, <https://www.theguardian.com/society/2020/apr/08/health-minister-defends-uk-coronavirus-testing-approach-chris-whitty-germany-covid-19>. Accessed 16 June 2022.

McCarthy, John. *Artificial intelligence, logic and formalizing common sense*. Dordrecht, 1989.

McCurry, Justin, and Rebecca Ratcliffe. "Newborn among 28000 coronavirus cases as death toll passes 550." *The Guardian*, 4 February 2020, <https://www.theguardian.com/world/2020/feb/05/coronavirus-cruise-ship-carrying-3700-quarantined-in-japan-after-10-test-positive>. Accessed 15 June 2022.

McDonald, Liz. "Big Night In: 11 highlights from BBC's £27m fundraiser." *BBC*, 24 April 2020, <https://www.bbc.com/news/entertainment-arts-52379802>. Accessed 15 June 2022.

McGuinness, Alan, and Greg Heffer. "Coronavirus: UK schools to close from Friday until further notice." *Sky News*, 19 March 2020, <https://news.sky.com/story/coronavirus-uk-schools-to-close-from-friday-11959769>. Accessed 15 June 2022.

Medhat, Walaa, et al. "Sentiment analysis algorithms and applications: A survey." *Ain Shams Engineering Journal*, vol. Volume 5, no. 4, December 2014, pp. 1093-1113.

Miller, J. R. "R. British teen dies after suicide attempt due to coronavirus fears." <https://nypost.com/2020/03/25/british-teen-dies-after-suicide-attempt-due-to-coronavirus-fears/>, *New York Post*, 25 March 2020, <https://nypost.com/2020/03/25/british-teen-dies-after-suicide-attempt-due-to-coronavirus-fears/>. Accessed 25 March 2020.

- Mohdin, Aamna. "Rise in depression and anxiety day after UK lockdown announced—study." *The Guardian*, 31 March 2020, <https://www.theguardian.com/world/2020/mar/31/rise-in-depression-and-anxiety-day-after-uk-lockdown-announced-study-coronavirus>. Accessed 15 June 2022.
- Murray, Gregg R., et al. "Pre-election polling: Identifying likely voters using iterative expert data mining." *Public Opinion Quarterly*, vol. 73.1, 2009, pp. 159-171.
- Murray, Gregg R., and Anthony Scime. "Data Mining." *Methods of Research*, 15 May 2015.
- Murray, Gregg R., and Anthony Scime. "Microtargeting and electorate segmentation: data mining the American National Election Studies." *Journal of Political Marketing*, vol. 9.3, 2010, 143-166.
- "Nation applauds NHS workers fighting coronavirus for third week running." *The Independent*, 9 April 2020, <https://www.independent.co.uk/news/uk/home-news/coronavirus-nhs-clap-for-our-carers-video-charity-time-a9458606.html>. Accessed 16 June 2022.
- "Nation applauds NHS workers fighting coronavirus for third week running." *The Independent*, 9 April 2020, <https://www.independent.co.uk/news/uk/home-news/coronavirus-nhs-clap-for-our-carers-video-charity-time-a9458606.html>. Accessed 16 June 2022.
- Negara, Edi Surya, et al. "Topic Modelling Twitter Data with Latent Dirichlet Allocation Method." *International Conference on Electrical Engineering and Computer Science (ICECOS)*, 2019, pp. 386-390.
- Ordun, Catherine, et al. ""Exploratory analysis of covid-19 tweets using topic modeling, umap, and digraphs."" *arXiv preprint arXiv:2005.03082*, 2020.
- Ozili, Peterson K., and Arun Thankom. "Spillover of COVID-19: impact on the Global Economy." *SSRN 3562570*, 2020.
- Panagiotidis, P., et al. "Changes in alcohol use habits in the general population, during the COVID-19 lockdown in Greece." *Alcohol and Alcohol. (Oxford, Oxfordshire)*, 2020, pp.

702-704.

Pantic, Igor. "Online social networking and mental health." *Cyberpsychology, Behavior, and Social Networking*, 2014.

Partington, Richard, and Rowena Mason. "Key points from budget 2020 - at a glance| Budget 2020." *The Guardian*, 11 March 2020, <https://www.theguardian.com/uk-news/2020/mar/11/key-points-from-budget-2020-at-a-glance>. Accessed 15 June 2022.

Partington, Richard, and Joanna Partridge. "Six in 10 UK firms have no more than three months of cash left." *The Guardian*, 1 April 2020, <https://www.theguardian.com/world/2020/apr/02/coronavirus-six-in-10-british-firms-have-no-more-than-three-months-of-cash-left>. Accessed 15 June 2022.

Patterson, James. "Captain Tom Moore's NHS appeal tops £32m on 100th birthday." *BBC*, 30 April 2020, <https://www.bbc.com/news/uk-england-beds-bucks-herts-52472132>. Accessed 16 June 2022.

Patterson, James. "Clap for Carers: UK in 'emotional' tribute to NHS and care workers." *BBC*, 27 March 2020, <https://www.bbc.com/news/uk-52058013>. Accessed 16 June 2022.

Patterson, James. "Coronavirus: Widespread transmission in UK 'highly likely.'" *BBC*, 2 March 2020, <https://www.bbc.com/news/uk-51700604>. Accessed 16 June 2022.

Perraudin, Frances. "Rishi Sunak heeds calls to help charities with £750m extra funding." *The Guardian*, 8 April 2020, <https://www.theguardian.com/politics/2020/apr/08/rishi-sunak-heeds-calls-to-help-charities-with-750m-extra-funding>. Accessed 16 June 2022.

Perraudin, Frances, and Pamela Duncan. "Britain's coronavirus testing scandal: a timeline of mixed messages." *The Guardian*, 3 April 2020, <https://www.theguardian.com/politics/2020/apr/03/coronavirus-testing-in-uk-timeline-of-ministers-mixed-messages>. Accessed 16 June 2022.

Peston, Robert. "Is Michael Gove right there is a shortage of coronavirus test kit ingredients?" *ITV*, 31 March 2020, <https://www.itv.com/news/2020-03-31/robert->

peston-is-gove-right-that-there-is-a-shortage-of-test-kit-ingredients. Accessed 15 June 2022.

Piquero, Alex R., et al. "Domestic violence during the COVID-19 pandemic - Evidence from a systematic review and meta-analysis." *Journal of Criminal Justice*, vol. 74, May-June 2021.

Pollard, M. S., et al. "Changes in adult alcohol use and consequences during the COVID-19 pandemic in the US." *JAMA Network Open*, 2020. "pyLDavis 2.1.2 documentation." *pyLDavis's documentation!*, <https://pyldavis.readthedocs.io/en/latest/readme.html>. Accessed 1 April 2022.

Qiang, Jipeng. "Short text topic modeling techniques, applications, and performance: a survey." *IEEE Transactions on Knowledge and Data Engineering*, vol. 34.3, 2020.

Rabindra, Lamsal. "Design and analysis of a large-scale COVID-19 tweets dataset." *applied intelligence*, vol. 51.5, 2021.

Rajasethupathy, Karthik. "Finding persistent rules: Combining association and classification results." *Expert Systems with Applications*, vol. 36.3, 2009, pp. 6019-6024.

Richardson, Hannah. "Coronavirus: UK schools, colleges and nurseries to close from Friday." *BBC*, 18 March 2020, <https://www.bbc.com/news/uk-51952314>. Accessed 15 June 2022.

"Rishi Sunak announces micro-loan scheme for businesses with £50000 limit – video." *The Guardian*, 27 April 2020, <https://www.theguardian.com/politics/video/2020/apr/27/rishi-sunak-announces-micro-loan-scheme-for-businesses-with-50000-limit-video>. Accessed 15 June 2022.

Rosenfeld,, A., et al. "Combining psychological models with machine learning to better predict people's decisions." *Synthese*, 2012, pp. 81-93.

Rosner, E., and K. Sheehy. "Top Manhattan ER doc commits suicide, shaken by coronavirus onslaught." <https://nypost.com/2020/04/27/manhattan-er-doc-lorna->

- breen-commits-suicide-shaken-by-coronavirus/*, New York Post, 27 April 2020, <https://nypost.com/2020/04/27/manhattan-er-doc-lorna-breen-commits-suicide-shaken-by-coronavirus/>. Accessed 27 April 2020.
- Roxby, Philippa. "Coronavirus: 'Herculean effort' to provide NHS protective gear." *BBC*, 11 April 2020, <https://www.bbc.com/news/uk-52248423>. Accessed 16 June 2022.
- Ruest, Nick, and Ian Milligan. "An open-source strategy for documenting events: The case study of the 42nd Canadian federal election on Twitter." *Code4Lib Journal*, no. 32, 2016.
- Russell, Stuart, and Peter Norvig. *Artificial intelligence: a modern approach*. 2002.
- Ryder, Alistair. "Cambridge Folk Festival has been cancelled this summer due to coronavirus - Cambridgeshire Live." *Cambridge News*, 27 March 2020, <https://www.cambridge-news.co.uk/whats-on/whats-on-news/cambridge-folk-festival-been-cancelled-17997939>. Accessed 16 June 2022.
- Sacco, Matteo Antonio, et al. "The impact of the Covid-19 pandemic on domestic violence: The dark side of home isolation during quarantine." *Medico-Legal Journal*, vol. 88(2), 2020, pp. 71-73.
- Safko, Lon. *The social media bible: tactics, tools, and strategies for business success*. John Wiley & Sons, 2010.
- Savage, Mark. "Captain Tom tops the charts at the age of 99." *BBC*, 24 April 2020, <https://www.bbc.com/news/entertainment-arts-52415966>. Accessed 16 June 2022.
- S. Dubey, et al. "Psychosocial impact of COVID-19." *Diabetes Metab. Syndr.*, 2020, pp. 779-788.
- Sher, Leo. "The impact of the COVID-19 pandemic on suicide rates." *QJM: An International Journal of Medicine*, vol. 113, no. 10, 30 June 2020, pp. 707-712.
- Shirky, Clay. *Here comes everybody: The power of organizing without organizations*. Penguin, 2008.

Sievert, Carson, and Kenneth Shirley. "LDAvis: A method for visualizing and interpreting topics." *Proceedings of the workshop on interactive language learning, visualization, and interfaces.*, 2014.

Sim, Walter. "Coronavirus: Japan reports 41 more positive cases on cruise ship, total cases on Diamond Princess now 61." *The Straits Times*, 7 February 2020, <https://www.straitstimes.com/asia/east-asia/coronavirus-another-41-on-japan-cruise-tested-positive-nhk>. Accessed 15 June 2022.

"Social Statistics." The University of Manchester, <https://www.socialsciences.manchester.ac.uk/social-statistics/study/why-study-social-statistics/>.

Sparrow, Andrew, and Lucy Campbell. "All schools to close from Friday; GCSE and A-level exams cancelled – UK Covid-19, as it happened." *The Guardian*, 18 March 2020, <https://www.theguardian.com/politics/live/2020/mar/18/uk-coronavirus-live-boris-johnson-pmq-cbi-urges-government-pay-businesses-directly-saying-350bn-loan-grant-package-not-enough>. Accessed 16 June 2022.

Spinelli, A., and G. Pellino. "COVID-19 pandemic: perspectives on an unfolding crisis." *British Journal of Surgery*, vol. Volume 107, no. Issue 7, 23 March 2020, pp.785-787.

"Stemming and Lemmatization in Python." *DataCamp*, 23 October 2018, <https://www.datacamp.com/community/tutorials/stemming-lemmatization-python>. Accessed 19 March 2022.

Stewart, Heather. "Sunak to unveil budget aimed at helping firms deal with coronavirus." *The Guardian*, 6 March 2020, <https://www.theguardian.com/uk-news/2020/mar/06/sunak-to-unveil-budget-aimed-at-helping-firms-deal-with-coronavirus>. Accessed 15 June 2022.

Stewart, Heather, and Denis Campbell. "NHS workers angered at Hancock's warning not to overuse PPE." *The Guardian*, 10 April 2020,

<https://www.theguardian.com/society/2020/apr/10/matt-hancock-urges-public-not-to-overuse-ppe>. Accessed 16 June 2022.

Syal, Rajeev. "Matt Hancock sets target of 100000 coronavirus tests a day." *TheGuardian*, 2 April 2020, <https://www.theguardian.com/world/2020/apr/02/matt-hancock-sets-target-of-100000-coronavirus-tests-a-day>. Accessed 16 June 2022.

Syed, Shaheen, and Marco Spruit. "Full-text or abstract? Examining topic coherence scores using latent dirichlet allocation." *IEEE International conference on data science and advanced analytics*, 2017, pp. 165-174.

Thomas, Katie. "Trump Says Contracting Covid-19 Was 'a Blessing From God.'" *TheNew York Times*, 3 November 2020, <https://www.nytimes.com/2020/10/07/us/politics/trump-coronavirus-blessing.html>. Accessed 15 May 2022.

"TIMELINE: Coronavirus epidemic (January and February)." *BNO News*, 23 January 2020, <https://bnonews.com/index.php/2020/01/timeline-coronavirus-epidemic/>. Accessed 14 June 2022.

"Timeline of UK government coronavirus economic support for businesses, March 2020 to March 2022." *The Institute for Government* |, <https://www.instituteforgovernment.org.uk/charts/coronavirus-support-businesses>. Accessed 15 June 2022.

Tran, Khuyen. "pyLDavis: Topic Modelling Exploration Tool That Every NLP Data Scientist Should Know - neptune.ai." *Neptune.ai*, 15 November 2021, <https://neptune.ai/blog/pyldavis-topic-modelling-exploration-tool-that-every-nlp-data-scientist-should-know>. Accessed 1 April 2022.

"Twitter Data Collection Using Twarc." *UNLV Libraries*, <https://www.library.unlv.edu/sites/default/files/inline-images/fy9jEgQjTudzio3zy68UnbyPBnNJBbJp1kmnPjVR5KaFiv2QTq.pdf>. Accessed 17 June 2022.

“UK thanks key coronavirus workers in nationwide Clap For Our Carers.” *ITV*, 23 April 2020, <https://www.itv.com/news/2020-04-23/uk-thanks-key-coronavirus-workers-in-nationwide-clap-for-our-carers>. Accessed 16 June 2022.

Umair, Qazi, et al. “GeoCoV19: a dataset of hundreds of millions of multilingual COVID-19 tweets with location information.” *SIGSPATIAL Special*, vol. 12.1, 2020, pp. 6-15.

Usman, Naseem. “COVIDSenti: A Large-Scale Benchmark Twitter Data Set for COVID-19 Sentiment Analysis.” *IEEE Transactions on Computational Social Systems*, 2021.

Usman, Naseem, et al. “COVIDSenti: A large-scale benchmark Twitter data set for COVID-19 sentiment analysis.” *IEEE Transactions on Computational Social Systems*, vol. 8.4, 2021.

Walker, Amy, and Matthew Weaver. “Government says it's 'too early' to say when UK lockdown will end as death toll reaches 5373 – as it happened.” *The Guardian*, <https://www.theguardian.com/politics/live/2020/apr/06/uk-coronavirus-live-boris-johnson-spends-night-in-hospital?page=with:block-5e8af4188f080bdb9f534ee8&filterKeyEvents=false>. Accessed 15 June 2022.

Walker, Peter. “Health minister Nadine Dorries diagnosed with coronavirus | Coronavirus.” *The Guardian*, 10 March 2020, <https://www.theguardian.com/world/2020/mar/10/health-minister-nadine-dorries-diagnosed-with-coronavirus>. Accessed 14 June 2022.

Weaver, Matthew. “First British coronavirus death confirmed | Coronavirus.” *The Guardian*, 28 February 2020, <https://www.theguardian.com/world/2020/feb/28/first-briton-dies-from-coronavirus-japanese-media-report-diamond-princess>. Accessed 15 June 2022.

“WHO-convened global study of origins of SARS-CoV-2: China Part.” *WHO | World Health Organization*, 30 March 2021, <https://www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part>. Accessed 16 June 2022.

WHO Headquarters (HQ). “WHO-convened global study of origins of SARS-CoV-2: China Part.” <https://www.who.int/publications/i/item/who-convened-global-study-of-origins-of-sars-cov-2-china-part>.

Wintour, Patrick. "Coronavirus: UN says warring countries have responded to ceasefire call." *The Guardian*, 3 April 2020,

<https://www.theguardian.com/world/2020/apr/03/coronavirus-threat-prompts-un-to-redouble-efforts-to-end-long-term-wars>. Accessed 15 June 2022.

"[Withdrawn] COVID-19: specified countries and areas." *GOV.UK*, 25 February 2020,

<https://www.gov.uk/government/publications/covid-19-specified-countries-and-areas>. Accessed 15 June 2022.

Yang, Sidi, and Haiyi Zhang. "Text Mining of Twitter Data Using a Latent Dirichlet Allocation Topic Model and Sentiment Analysis." *International Journal of Computer and Information Engineering*, vol. 12.7, 2018, pp. 525-529.

Yin, Jianhua, and Jianyong Wang. "A dirichlet multinomial mixture model-based approach for short text clustering." *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, 2014.

Zhang, Lei, and Liu Bing. *Mining Text Data*. Springer, Boston, MA, 2012.