



ΕΛΛΗΝΙΚΗ
ΔΗΜΟΚΡΑΤΙΑ

ΠΑΝΕΠΙΣΤΗΜΙΟ
ΜΑΚΕΔΟΝΙΑΣ

ΔΙΑΤΜΗΜΑΤΙΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ ΣΤΑ
ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ
MASTER IN INFORMATION SYSTEMS

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Θέμα Εργασίας:

Επισκόπηση των Συστημάτων Μηχανικής Μάθησης για Πρόβλεψη Πωλήσεων

Review of Machine Learning Systems for Sales Forecasting

Όνομα Φοιτητή: **Χαραυγής Φίλιππος**
Α.Μ.: mis20020

Επιβλέπων Καθηγητής: **Ταραμπάνης Κωνσταντίνος**

Επιτελική Σύνοψη

Η πρόβλεψη πωλήσεων αφορά την εκτίμηση των μελλοντικών πωλήσεων μιας επιχείρισης, για την προληπτική αντιμετώπιση φαινομένων όπως ελλειπή ή περισσευούμενα αποθέματα. Τα τελευταία χρόνια έχουν μελετηθεί και προταθεί διάφορες προσεγγίσεις για την επίλυση αυτού του προβλήματος με τη χρήση μηχανικής μάθησης, είτε με τη χρήση προϋπάρχοντων μοντέλων, είτε με την ανάπτυξη νέων.

Στην παρούσα έρευνα συλλέχθηκαν και μελετήθηκαν 38 διαφορετικά άρθρα που δημοσιεύθηκαν πάνω στο αντικείμενο της χρήσης μηχανικής μάθησης για την πρόβλεψη πωλήσεων σε διάφορους τομείς. Αρχικά συνοψίζονται τα περιεχόμενα και αποτελέσματα της κάθε μελέτης ξεχωριστά, και στη συνέχεια σχηματίζονται συμπεράσματα από το σύνολο των ερευνών που συμπεριλήφθηκαν.

Abstract

Sales forecasting concerns the estimation of the future sales of a company, in order to preemptively deal with phenomena such as insufficient or excess inventory. Various approaches have been studied and proposed to solve this problem in recent years using machine learning, either by using existing models or by developing new ones.

In this research, 38 different published articles on the subject of using machine learning for sales forecasting in a variety of sectors were collected and studied. The contents and results of each study were initially studied and summarized separately, then conclusions were drawn from the sum of all included studies.

Περιεχόμενα

| | |
|---|----|
| Κεφάλαιο 1: Εισαγωγή..... | 5 |
| 1.1 Σκοπός της διπλωματικής..... | 5 |
| 1.2 Περιεχόμενα της μελέτης..... | 5 |
| Κεφάλαιο 2: Μεθοδολογία..... | 6 |
| 2.1 Περιγραφή Μεθοδολογίας..... | 6 |
| 2.2 Σύνοψη αποτελεσμάτων..... | 7 |
| 2.3 Συναφείς έρευνες..... | 8 |
| Κεφάλαιο 3: Θεωρητικό Υπόβαθρο..... | 9 |
| Κεφάλαιο 4: Συγγραφο-κεντρική Ανάλυση..... | 11 |
| 4.1 Άρθρα που δημοσιεύτηκαν σε επιστημονικά περιοδικά..... | 11 |
| 4.2 Άρθρα που δημοσιεύτηκαν σε συνέδρια..... | 20 |
| Κεφάλαιο 5: Εννοιο-κεντρική Ανάλυση..... | 42 |
| Κεφάλαιο 6: Συμπεράσματα - Προβλήματα που αντιμετωπίστηκαν..... | 49 |
| 6.1 Συμπεράσματα..... | 49 |
| 6.2 Προβλήματα που αντιμετωπίστηκαν..... | 53 |
| Βιβλιογραφία..... | 54 |

Πίνακες

| | |
|---|-------|
| Πίνακας 3.1: Ποιοτικά μέτρα μοντέλων μηχανικής μάθησης [Studer, S. et al, 2021]..... | 9 |
| Πίνακας 4.1: Ορισμοί μεταβλητών πρόβλεψης [Wang, S., Yang, Y., 2021]..... | 11 |
| Πίνακας 4.2: Παράμετροι και ρυθμίσεις του μοντέλου XGBOOST [Wang, S., Yang, Y., 2021]..... | 12 |
| Πίνακας 4.3: Μεταβλητές πρόβλεψης που επιλέχθηκαν από τα δυο σετ δεδομένων [Chen, T., Yin, H., Chen, H., Wang, H., Zhou, X., Li, X., 2020]..... | 16 |
| Πίνακας 4.4: Εβδομαδιαίες προβλέψεις από τα διάφορα μοντέλα [Dai, Y., Huang, J., 2021]..... | 24 |
| Πίνακας 4.5: Σύγκριση απόδοσης των μοντέλων [Wang, J., 2020]..... | 25 |
| Πίνακας 4.6: Σύγκριση δεικτών σφάλματος για τα διάφορα μοντέλα [Behera, G., Nain, N,2020]...31 | 31 |
| Πίνακας 4.7: Σύγκριση δεικτών αποτελεσματικότητας και σφάλματος για τα διάφορα μοντέλα [Tekin, A.T., Cebi, F., 2020]..... | 32 |
| Πίνακας 4.8: Σύγκριση αποτελεσμάτων των μοντέλων (ποσοστιαία) [Silva, J., Villa, J.V., Cabrera, D., 2020]..... | 33 |
| Πίνακας 4.9: Σημαντικότητα των μεταβλητών και σύγκριση δεικτών σφάλματος πριν και μετά την εφαρμογή της GSO [Behera, G., Nain, N., 2019]..... | 36 |
| Πίνακας 4.10: Σύγκριση δεικτών σφάλματος για τα διάφορα μοντέλα [Punam, K., Pamula, R., Jain, P.K., 2019]..... | 38 |
| Πίνακας 4.11: Σύγκριση δεικτών σφάλματος για τα διάφορα μοντέλα [Chen, T., Yin, H., Chen, H., Wu, L., Wang, H., Zhou, X., Li, X., 2018]..... | 39 |
| Πίνακας 6.1: Συγκεντρωτικά αποτελέσματα της έρευνας..... | 50-52 |

Εικόνες

| | |
|---|-------|
| Εικόνα 4.1: Σημαντικότητα πρόβλεψης των διαφόρων μεταβλητών [Liu, Y., Feng, L., Jin, B., 2020]..... | 13 |
| Εικόνα 4.2: Αρχιτεκτονική του συστήματος ForeXGBoost [Xia, Z., Xue, S., Wu, L., Sun, J., Chen, Y., Zhang, R., 2020]..... | 14 |
| Εικόνα 4.3: Διάγραμμα ροής της δημιουργίας του μοντέλου [Wang, H., 2020]..... | 15 |
| Εικόνα 4.4: Εικονική αναπαράσταση του μοντέλου πρόβλεψης πωλήσεων του Big Mart [Theresa, I., Medikonda, V.R., Reddy, K.V.N., 2020]..... | 17 |
| Εικόνα 4.5: Διάγραμμα ροής λειτουργίας μοντέλου και συγκριτικό διάγραμμα αποτελεσμάτων. [Dong, X.,2016]..... | 19 |
| Εικόνα 4.6: Προβλέψεις πωλήσεων ανα κατηγορία ηλικίας, φύλου και επαγγέλματος [Ramachandra, H.V., Balaraju, G., Rajashekar, A., Patil, H., 2021]..... | 22 |
| Εικόνα 4.7: Συγκριτικά διαγράμματα προβλέψεων για τρόφιμα, χόμπυ και οικιακά είδη στην Καλιφόρνια [Huo, Z., 2021]..... | 23 |
| Εικόνα 4.8: Αποτελέσματα προβλέψεων για τα μοντέλα γραμμικής παλινδρόμησης, SVR και Stacking [Liao, W., Ye, G., Yin, Y., Yan, W., Ma, Y., Zuo, D., 2020]..... | 26 |
| Εικόνα 4.9: Σημαντικότητα πρόβλεψης των μεταβλητών [Edwardo, T.O., Ruldeviyani, Y., 2020].. | 27 |
| Εικόνα 4.10: Σημαντικότητα πρόβλεψης των μεταβλητών & αξιολόγηση των μοντέλων [Singh, A., Srivastava, S., 2020]..... | 30 |
| Εικόνα 4.11: Αποτελέσματα πρόβλεψης του Convolutional Neural Network για 40 βδομάδες [Buyar, V., Abdel-Raouf, A., 2019]..... | 34 |
| Εικόνα 4.12: Σύγκριση προβλεπόμενου και πραγματικού αριθμού πωλήσεων για τα έτη 2011-2014 [Gopalakrishnan, T., Choudhary, R., Prasad, S., 2018]..... | 41 |
| Εικόνα 5.1: Περιγραφή της διαδικασίας επεξεργασίας δεδομένων στις έρευνες..... | 42 |
| Εικόνα 5.2: Συχνότητα ανάπτυξης νέων αλγορίθμων στις έρευνες..... | 44 |
| Εικόνα 5.3: Διαστήματα στα οποία γίνεται η πρόβλεψη πωλήσεων στις έρευνες..... | 45 |
| Εικόνα 5.4: Προελεύσεις σετ δεδομένων που χρησιμοποιήθηκαν στις έρευνες..... | 45 |
| Εικόνα 5.5: Συχνότητα χρήσης των αλγορίθμων στα συμπεριλαμβανόμενα άρθρα..... | 47-48 |

Κεφάλαιο 1: Εισαγωγή

1.1 Σκοπός της διπλωματικής

Η ακριβής πρόβλεψη πωλήσεων για την βέλτιστη διαχείριση μιας επιχείρησης είναι ζωτικό κομμάτι για την επιτυχία της. Μπορεί να οδηγήσει σε τόσο βραχυχρόνιες όσο και μακροχρόνιες αποφάσεις για την ελαχιστοποίηση των απωλειών και την αύξηση των πωλήσεων, και κατά συνέπεια ήδη επιχειρήσεις έχουν επενδύσει σε αυτόν τον τομέα. Η μηχανική μάθηση έχει ήδη χρησιμοποιηθεί με επιτυχία προς αυτόν τον σκοπό [Tsoumakas, 2019 & Xia et al, 2020]. Δεν υπάρχει όμως μόνο μια λύση που επαρκεί για όλα τα προβλήματα, το οποίο παρατηρείται από το γεγονός ότι διάφορες έρευνες που συμπεριλήφθηκαν εφαρμόζουν διαφόρους αλγορίθμους για να επιλέξουν τον πιο κατάλληλο, ενώ άλλες επιχειρούν να προτείνουν νέες λύσεις, αναπτύσσοντας το δικό τους μοντέλο.

Σκοπός της παρούσας διπλωματικής εργασίας είναι η συλλογή και σύνοψη των ερευνών που έχουν δημοσιευθεί πάνω στο θέμα τα τελευταία χρόνια. Η εργασία θα επικεντρωθεί συγκεκριμένα στις έρευνες που αφορούν τη χρήση μηχανικής μάθησης για την πρόβλεψη πωλήσεων.

1.2 Περιεχόμενα της μελέτης

Τα ακόλουθα περιεχόμενα της παρούσας μελέτης έχουν ως εξής:

Το κεφάλαιο 2 περιέχει επισκόπηση της μεθοδολογίας και των προτύπων που χρησιμοποιήθηκαν.

Το κεφάλαιο 3 περιέχει μια αρχική σύνοψη των βασικών εννοιών που θα χρησιμοποιηθούν στην μελέτη.

Το κεφάλαιο 4 περιέχει μια πρώτη συγγραφο-κεντρική (author-centric) ανάλυση, στην οποία παρουσιάζεται μια σύνοψη του κάθε άρθρου που συμπεριλήφθηκε στη μελέτη.

Το κεφάλαιο 5 περιέχει στη συνέχεια μια εννοιο-κεντρική (concept-centric) ανάλυση, στην οποία παρουσιάζονται τα κοινά σημεία που προέκυψαν από το σύνολο των άρθρων που συμπεριλήφθηκαν στη μελέτη.

Τέλος, το κεφάλαιο 6 περιέχει τα συμπεράσματα της μελέτης, προβλήματα που αντιμετωπίστηκαν κατά την διεξαγωγή της, και επιτελική σύνοψη της μελέτης με πρόταση για βελτίωσή της στο μέλλον.

Κεφάλαιο 2: Μεθοδολογία

2.1 Περιγραφή Μεθοδολογίας

Για την παρούσα έρευνα χρησιμοποιήθηκε η μεθοδολογία Webster – Watson, η οποία αναλύει πώς γίνεται μία βιβλιογραφική επισκόπηση. Αρχικά γίνεται μια συγγραφο-κεντρική ανάλυση, και στη συνέχεια ακολουθεί εννοιο-κεντρική ανάλυση στο βαθμό που επιτρέπει η ποιότητα των άρθρων που συμπεριλήφθηκαν. Σαν μηχανή αναζήτησης για το στάδιο συλλογής των άρθρων χρησιμοποιήθηκε το Scopus.

Αναζητήθηκαν άρθρα που περιέχουν τον όρο “sales prediction” στον **τίτλο** και εκδόθηκαν από το 2016 μέχρι και τις 31 Μαΐου 2022.

Η αναζήτηση αυτή επέφερε 167 αποτελέσματα. Αξίζει επίσης να σημειωθεί επίσης ότι η συντριπτική πλειοψηφία των αποτελεσμάτων εκδόθηκε από το 2019 και μετά, το οποίο ήταν αναμενόμενο.

Σημειωτέον ότι κάποια άρθρα προέρχονταν από συνέδρια (Conference Papers) ενώ κάποια από επιστημονικά περιοδικά (Articles). Από τα αποτελέσματα της αναζήτησης, τα 65 προέρχονταν από επιστημονικά περιοδικά, ενώ τα 102 προέρχονταν από συνέδρια.

Στη συνέχεια μελετήθηκαν οι περιλήψεις (abstracts), λέξεις-κλειδιά (keywords) και βιβλιογραφικές αναφορές (references) από αυτά τα αποτελέσματα για να κρατηθούν μόνο όσα ήταν σχετικά με το ζητούμενο αντικείμενο. Πιο συγκεκριμένα συμπεριλαμβάνονται μόνο όσα άρθρα αφορούν τη χρήση κάποιου μοντέλου μηχανικής μάθησης για την πρόβλεψη πωλήσεων, συνεπώς σε αυτό το στάδιο κρατήθηκαν τα άρθρα στα οποία αναφερόταν ο όρος “μηχανική μάθηση” (“machine learning”) οπουδήποτε στις περιλήψεις, λέξεις-κλειδιά και βιβλιογραφικές αναφορές, ή εναλλακτικά περιείχαν ονομαστική αναφορά κάποιου μοντέλου μηχανικής μάθησης (π.χ. Gradient Boosting). Σε αυτό το σημείο απορρίφθηκαν 92 άρθρα και κρατήθηκαν τα 75.

Ένα επιπλέον κριτήριο απόρριψης ήταν η χρήση δεδομένων που προέρχονται από κοινωνικά δίκτυα.

Μετά την αρχική διαλογή μελετήθηκαν τα σώματα των κειμένων των άρθρων σε ένα τελευταίο στάδιο διαλογής. Σε αυτό το στάδιο απορρίφθηκαν 37 επιπλέον άρθρα, καθώς κατά την ανάγνωση του κυρίως κειμένου έγινε εμφανές ότι δεν ήταν απολύτως συμβατά με το ζητούμενο της παρούσας εργασίας. Σημειώνεται ότι δύο από τα άρθρα απορρίφθηκαν λόγω του ότι ήταν έρευνες επισκόπησης πάνω στο αντικείμενο όπως η παρούσα, και δεν έκαναν δικιά τους εφαρμογή μηχανικής μάθησης, αλλά θα αναφερθούν ξεχωριστά στην παρούσα ενότητα ως συναφείς έρευνες.

2.2 Σύνοψη αποτελεσμάτων

Συνολικά αποτελέσματα: 167

Αριθμός άρθρων από επιστημονικά περιοδικά: 65

Αριθμός άρθρων από συνέδρια: 102

Αριθμός άρθρων που απορρίφθηκαν στον πρώτο έλεγχο: 92

Αριθμός άρθρων που απορρίφθηκαν στον τελικό έλεγχο: 37

Αριθμός άρθρων που συμπεριλήφθηκαν: 38

Από τα οποία άρθρα από επιστημονικά περιοδικά: 10

Από τα οποία άρθρα από συνέδρια: 28

2.3 Συναφείς έρευνες

2.3.1 A survey of machine learning techniques for food sales prediction by Tsoumakas, G. (2019)

Η συγκεκριμένη έρευνα κάνει αρχικά αναφορά σε προκλήσεις που αντιμετωπίζονται στη χρήση μηχανικής μάθησης για την πρόβλεψη πωλήσεων στον τομέα τροφίμων, όπως για παράδειγμα η ανεπαρκής ποιότητα καταγραφής δεδομένων. Στη συνέχεια παρουσιάζει συνήθεις μεταβλητές πρόβλεψης και τύπους μεταβλητών που χρησιμοποιήθηκαν στις 4 συμπεριλαμβανόμενες έρευνες, καθώς και μια αναφορά στα μοντέλα που χρησιμοποιήσαν, τα οποία ήταν Moving Average (MA), Radial Basis Function (RBF), 24 διαφορετικούς συνδυασμούς 8 αλγορίθμων, και Deep Learning.

Τέλος, γίνεται σύντομη περιγραφή των συνήθων τρόπων αξιολόγησης μοντέλων μηχανικής μάθησης, και παρατίθενται οι τύποι υπολογισμών ορισμένων συναρτήσεων απώλειας.

2.3.2 A survey on retail sales forecasting and prediction in fashion markets by Beheshti-Kashi, S., Karimi, H.R., Thoben, K.-D., Lütjen, M. Teucke, M. (2014)

Η έρευνα αυτή παρουσιάζει συνοπτικά τις συνεισφορές που παρείχαν προηγούμενα άρθρα για την πρόβλεψη πωλήσεων στον τομέα της μόδας. Παρόλο που δεν επικεντρώνεται συγκεκριμένα σε προβλέψεις με τη χρήση μηχανικής μάθησης, γίνεται αναφορά σε μεθόδους όπως ARIMA και Artificial Neural Networks που χρησιμοποιήθηκαν για πρόβλεψη από ορισμένες έρευνες.

Κεφάλαιο 3: Θεωρητικό Υπόβαθρο

Η κεντρική έννοια της παρούσας έρευνας είναι η Μηχανική Μάθηση, ή Machine Learning (ML), ένας τομέας του πεδίου της τεχνητής νοημοσύνης και επιστήμης των υπολογιστών, που επικεντρώνεται στη χρήση πληροφοριών και αλγορίθμων για να μιμηθεί τον τρόπο που μαθαίνουν οι άνθρωποι, βελτιώνοντας σταδιακά την ακρίβεια. Με τη χρήση στατιστικών μεθόδων, οι αλγόριθμοι εκπαιδεύονται για να κάνουν κατηγοριοποιήσεις ή προβλέψεις, και να παρέχουν πληροφορίες που μπορούν να χρησιμοποιηθούν για τη λήψη αποφάσεων σε εφαρμογές και επιχειρήσεις.

Οι περιπτώσεις όπου οι αλγόριθμοι χρησιμοποιούνται για να κάνουν πρόβλεψη, στις οποίες ανήκουν και τα άρθρα που αφορούν την παρούσα έρευνα, εμπίπτουν συνήθως στην κατηγορία της επιβλεπόμενης μηχανικής μάθησης (supervised machine learning). Από την άλλη πλευρά, οι περιπτώσεις στις οποίες οι αλγόριθμοι χρησιμοποιούνται για να κάνουν κατηγοριοποίηση ανήκουν στην κατηγορία της μη-επιβλεπόμενης μηχανικής μάθησης (unsupervised machine learning). [IBM Cloud Education, 2020]

Η πιο βασική έννοια, που αποτελεί το άλφα και το ωμέγα της μηχανικής μάθησης, είναι τα features, τα οποία σε κάποιες περιπτώσεις θα αποκαλούνται μεταβλητές πρόβλεψης, ή predictor variables. Οι μεταβλητές πρόβλεψης στη μηχανική μάθηση είναι ανεξάρτητες μεταβλητές που χρησιμοποιούνται ως εισαγόμενες τιμές του αλγορίθμου, και τις οποίες το μοντέλο χρησιμοποιεί για να κάνει προβλέψεις. Πέρα από τις μεταβλητές πρόβλεψης που δίνονται έτοιμες από κάποια πηγή δεδομένων, μπορούν να προκύψουν και νέες μέσα από μια διαδικασία επεξεργασίας των ακατέργαστων δεδομένων που είναι γνωστή ως feature engineering. Αυτή η διαδικασία είναι συχνά απαραίτητη για τη βελτιστοποίηση της δυνατότητας πρόβλεψης του μοντέλου, είτε με την ανάδειξη των πιο σημαντικών μεταβλητών πρόβλεψης, είτε με την απομάκρυνση των λιγότερο χρήσιμων, είτε με την τυποποίηση των στοιχείων ώστε να είναι πιο εύκολα επεξεργάσιμα από το μοντέλο. Η ποιότητα αυτής της διαδικασίας σε πολλές περιπτώσεις αποτελεί τη βασική διαφορά ανάμεσα σε ένα καλό και ένα κακό μοντέλο. [Cogito Tech – Roger Brown, 2019]

Πέρα από το τι είναι η μηχανική μάθηση, καλό είναι να αναφερθούν και κάποια κριτήρια αξιολόγησης ενός μοντέλου μηχανικής μάθησης. Υπάρχουν πολλά κριτήρια με τα οποία μπορεί να αξιολογηθεί η ποιότητα ενός μοντέλου μηχανικής μάθησης, όπως για παράδειγμα τα πέντε που αναφέρονται στον παρακάτω πίνακα.

| Performance | The Model's Performance on Unseen Data |
|------------------|---|
| Robustness | Ability of the ML application to maintain its level of performance under defined circumstances (ISO/IEC technical report 24029) |
| Scalability | The model's ability to scale to high data volume in the production system. |
| Explainability | The model's direct or post hoc explainability. |
| Model Complexity | The model's capacity should suit the data complexity. |
| Resource Demand | The model's resource demand for deployment. |

Πίνακας 3.1: Ποιοτικά μέτρα μοντέλων μηχανικής μάθησης [Studer, S. et al, 2021]

Τα άρθρα που συμπεριλήφθηκαν στην παρούσα έρευνα όμως επικεντρώνονται αποκλειστικά σε

καθαρά αντικειμενικούς μαθηματικούς δείκτες. Οι δείκτες αυτοί ονομάζονται συναρτήσεις απώλειας (loss functions), και παρόλο που υπολογίζονται με διαφορετικό τρόπο κι έχουν διαφορετικά πλεονεκτήματα και μειονεκτήματα, ακολουθούν παρόμοια λογική. Στόχος τους είναι να αποδώσουν μια συνολική εικόνα των αποκλίσεων των προβλέψεων του αλγορίθμου από τις πραγματικές τιμές, εκφράζοντάς το είτε αθροιστικά, είτε ποσοστιαία, είτε κατά μέσο όρο. Όσο μεγαλύτερη είναι η τιμή του εκάστοτε δείκτη σφάλματος, τόσο χαμηλότερη είναι η ακρίβεια των αποτελεσμάτων του μοντέλου.

Οι δείκτες που απαντώνται στην παρούσα έρευνα είναι οι παρακάτω:

- Mean Square Error (MSE)
- Mean Absolute Error (MAE)
- Mean Absolute Percentage Error (MAPE)
- Symmetric Mean Absolute Percentage Error (SMAPE)
- Root Mean Square Error (RMSE)
- Root Mean Square Prediction Error (RMSPE)

Κεφάλαιο 4: Συγγραφο-κεντρική Ανάλυση

4.1 Άρθρα που δημοσιεύτηκαν σε επιστημονικά περιοδικά

4.1.1 M-GAN-XGBOOST model for sales prediction and precision marketing strategy making of each product in online stores
by Wang, S., Yang, Y. (2021)

Στο συγκεκριμένο άρθρο οι συγγραφείς αποσκοπούν να παρουσιάσουν ένα μοντέλο για την αντιμετώπιση των προκλήσεων που απέφερε η ραγδαία ανάπτυξη των διαδικτυακών πωλήσεων. Παρόλο που η πρακτική αποτελεί μεγάλη ευκολία τόσο για τους αγοραστές όσο και τους πωλητές, προβλήματα όπως ανεπαρκείς ή πλεονάζουσες προμήθειες έγιναν συχνό φαινόμενο σε καταστήματα που δεν μπορούσαν να προβλέψουν αποτελεσματικά τον όγκο των μελλοντικών πωλήσεων.

Το προτεινόμενο μοντέλο, με την ονομασία M-GAN-XGBOOST, αναπτύχθηκε με βάση τα LSTM (Long Short-Term Memory), GAN (Generative Adversarial Networks) και XGBOOST (eXtreme Gradient Boosting), υλοποιήθηκε με τη χρήση της πλατφόρμας Tensorflow, και κληρονομεί τα πλεονεκτήματα τις κάθε μεθόδου ώστε να παράξει πιο ακριβή αποτελέσματα. Τα δεδομένα προήλθαν από το online store Jingdong Mall και περιέχουν 12068 εγγραφές στο διάστημα 01/02/2018 με 15/04/2018. Οι μεταβλητές, που επιλέχθηκαν με τη χρήση Self-Determination Theory, παρουσιάζονται στον παρακάτω πίνακα.

| Name | Symbol | Meaning |
|-----------------------------|---------------|--|
| External Factors | <i>ecause</i> | Consumers receive external stimuli that can affect purchasing behavior |
| Consumer Traffic | <i>Ctra</i> | Changes in related variables caused by consumers visiting pages of the store |
| Consumer Perception | <i>Cper</i> | Related product information obtained by consumers |
| Potential Purchase Behavior | <i>Cpot</i> | Consumer hesitation |
| Sales | <i>Y</i> | Merchandise sales |

Πίνακας 4.1: Ορισμοί μεταβλητών πρόβλεψης [Wang, S., Yang, Y., 2021]

Επίσης παρουσιάζονται οι παράμετροι που χρησιμοποιήθηκαν στο XGBOOST τμήμα του μοντέλου.

| Name | Setting | Meaning | |
|-------------------------|-------------------|--|---|
| <i>learning_rate</i> | 0.0001 | Learning rate, which can control the step size of each weight when updating | |
| <i>n_estimators</i> | 50000 | Training times | |
| <i>max_depth</i> | 9 | The maximum depth of the tree | |
| <i>min_child_weight</i> | 1 | The smallest sum of sample weights in child nodes | |
| <i>gamma</i> | 0.0001 | The value of the minimum loss function | |
| <i>subsample</i> | 0.7 | The proportion of the sub-sample of the training model to the entire sample | |
| <i>colsample_bytree</i> | 0.5 | The proportion of random sampling of features when building the tree | |
| <i>eta</i> | 0.1 | Shrink the step size, reduce the feature weight in the promotion calculation | |
| <i>silent</i> | 1 | Print runtime information | |
| <i>eval_metric</i> | <i>rmse</i> | The evaluation index of the verification data, the root mean square error | The parameters and setting of XGBOOST model |
| <i>objective</i> | <i>reg:linear</i> | Objective function, linear regression problem | |

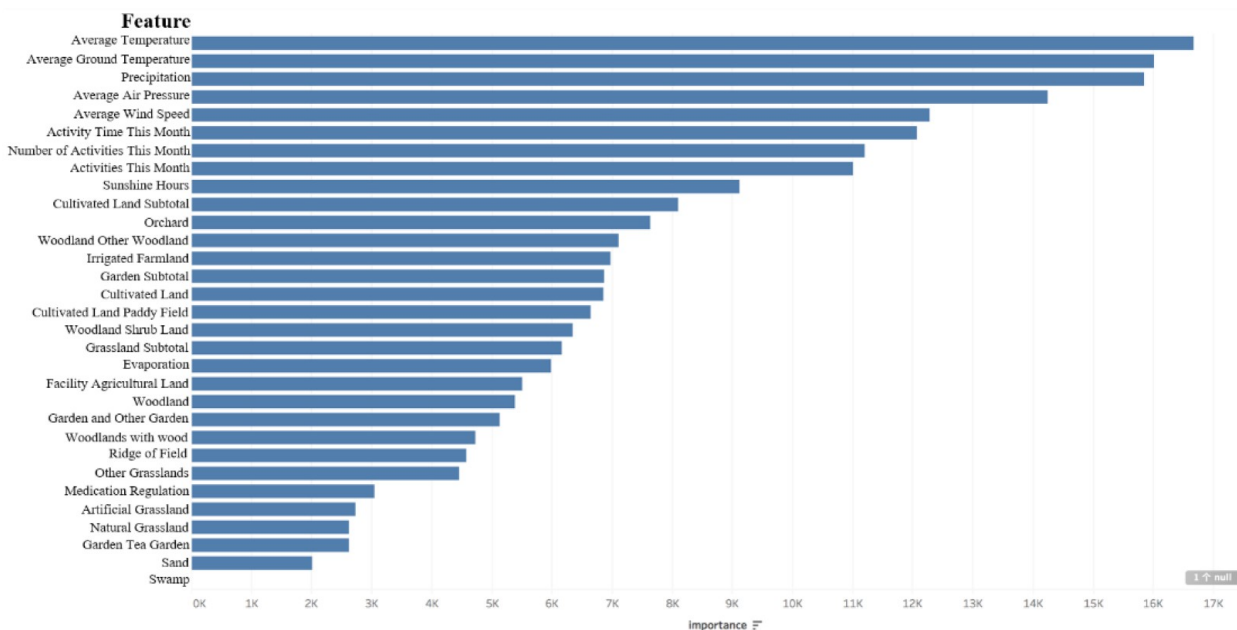
Πίνακας 4.2: Παράμετροι και ρυθμίσεις του μοντέλου XGBOOST [Wang, S., Yang, Y., 2021]

Το Mean Square Error (MSE) του τελικού μοντέλου ήταν 8,2 σε σύγκριση με το 11,03 του μοντέλου σύγκρισης, αλλά η ακριβεία διαφοροποιούταν από προϊόν σε προϊόν, ανάλογα με το αν υπήρχαν ραγδαίες διακυμάνσεις πωλήσεων στα ιστορικά στοιχεία.

4.1.2 Future-Aware Trend Alignment for Sales Predictions by Liu, Y., Feng, L., Jin, B. (2020)

Οι συγγραφείς σε αυτό το άρθρο επέλεξαν να εφαρμόσουν μέθοδο Deep Learning (Βαθιά Μάθηση) για την ακριβή πρόβλεψη πωλήσεων, που είναι κρίσιμη για την αλυσίδα παραγωγής. Για αυτό τον σκοπό ανέπτυξαν μια παραλλαγή της μεθόδου Trend Alignment with Dual-Attention Multi-task Recurrent Neural Networks (TADA) που αναπτύχθηκε από προηγούμενους ερευνητές, και την ονόμασαν Future-Aware TADA (F-TADA). Η διαφορά σε σύγκριση με την προηγούμενη εφαρμογή της μεθόδου είναι ότι λαμβάνεται υπόψη η επίγνωση παραγόντων σε μελλοντικό χρόνο, για παράδειγμα ο καιρός, οι μη-εργάσιμες μέρες και εκπτώσεις.

Για μακροπρόθεσμες προβλέψεις λήφθηκαν δεδομένα από αλυσίδες supermarket που προήλθαν από διαγωνισμό Kaggle. Για βραχυπρόθεσμες προβλέψεις λήφθηκαν δεδομένα πωλήσεων φυτοφαρμάκου από φαρμακευτική εταιρία, από τα οποία εξάχθηκαν μεταβλητές με τη χρήση XGBoost, οι οποίες φαίνονται παρακάτω.



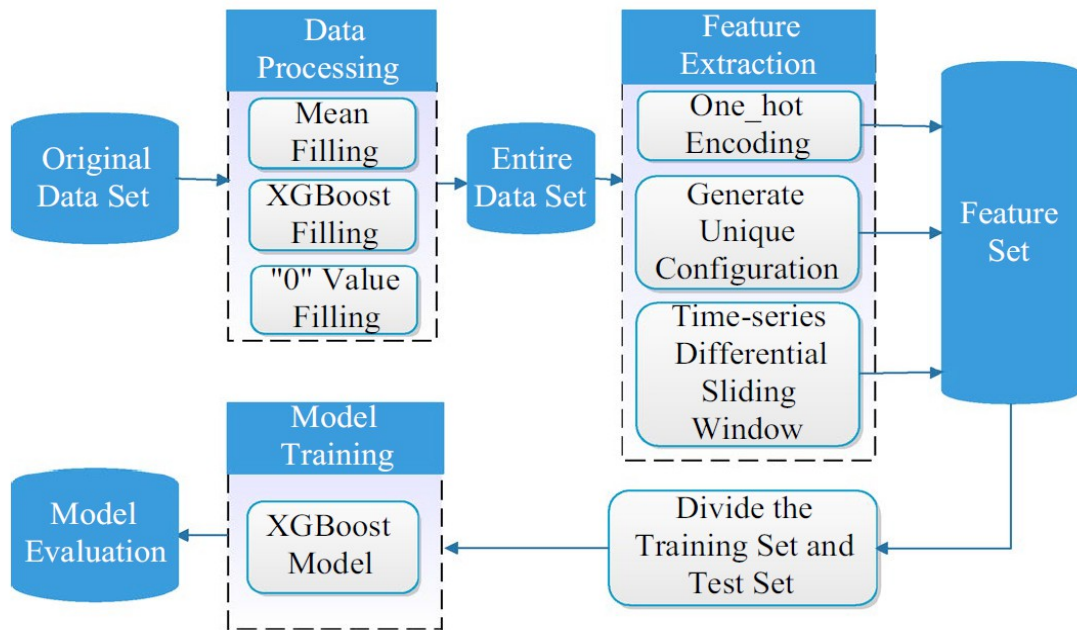
Εικόνα 4.1: Σημαντικότητα πρόβλεψης των διαφόρων μεταβλητών [Liu, Y., Feng, L., Jin, B., 2020]

Τα τελικά αποτελέσματα έδειξαν ότι η εφαρμογή της F-TADA βελτίωσε την ακρίβεια των προβλέψεων σε σύγκριση με την προηγούμενη μέθοδο, τόσο για τις μακροπρόθεσμες όσο και για τις βραχυπρόθεσμες προβλέψεις, όμως υπάρχουν επιπρόσθετοι παράγοντες που πρέπει να ληφθούν υπόψη, όπως το ότι η μέθοδος εφαρμόστηκε μόνο σε δεδομένα στα οποία οι μελλοντικοί παράγοντες που αναφέρθηκαν παραπάνω ήταν γνωστοί.

4.1.3 ForeXGBoost: passenger car sales prediction based on XGBoost by Xia, Z., Xue, S., Wu, L., Sun, J., Chen, Y., Zhang, R. (2020)

Το παρόν άρθρο αφορά ένα μοντέλο μηχανικής μάθησης που αναπτύχθηκε στα πλαίσια διαγωνισμού που οργάνωσε το 2018 η Alibaba Cloud κι η κινέζικη τοπική αυτοδιοίκηση Yancheng, και στον οποίο κέρδισε την πρώτη θέση. Το συγκεκριμένο μοντέλο, ονόματι ForeXGBoost, είναι βασισμένο στο μοντέλο XGBoost.

Η έρευνα ασχολείται συγκεκριμένα με την πώληση επιβατικών οχημάτων, και χρησιμοποιεί δεδομένα που προμηθεύτηκαν από την Alibaba Cloud και αφορούν το διάστημα μεταξύ του Ιανουαρίου 2011 και Δεκεμβρίου 2017, περιέχοντας περισσότερο από 5 εκατομμύρια δείγματα. Η μέθοδος, που ονόμασαν ForeXGBoost, αφορά το σύνολο των δράσεων επεξεργασίας δεδομένων και εξαγωγής μεταβλητών πρόβλεψης που εφαρμόσαν πριν την εφαρμογή της τυπικής μεθόδου XGBoost, η συνολική διαδικασία της οποίας απεικονίζεται παρακάτω.



The system architecture of ForeXGBoost

Εικόνα 4.2: Αρχιτεκτονική του συστήματος ForeXGBoost [Xia, Z., Xue, S., Wu, L., Sun, J., Chen, Y., Zhang, R., 2020]

Με τη διαδικασία επιλογής μεταβλητών πρόβλεψης, από τις 33 που περιείχε το αρχικό σετ δεδομένων επιλέχθηκαν οι 9, οι οποία σε σειρά από την πιο δυνατή στην πιο αδύναμη δυνατότητα πρόβλεψης ήταν ως εξής:

Brand, Displacement, Department, Compartment, Price, Gearbox Type, TypeID, Cylinder Number, Rated Passenger

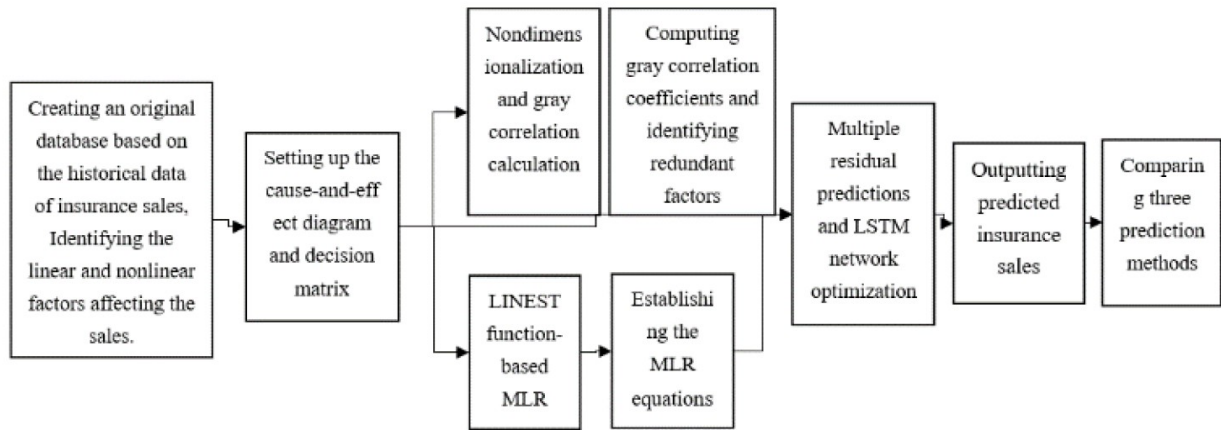
Στη συνέχεια με τη διαδικασία ρύθμισης των παραμέτρων, κατέληξαν ότι οι βέλτιστες τιμές παραμέτρων για την εφαρμογή του μοντέλου ήταν $\eta = 0.1$, $d_{max} = 9$, $w_{min} = 50$, $\gamma = 0$, $sr = 0.7$, $tc = 0.7$, $\lambda = 50$. Το τελικό μοντέλο που προέκυψε παρουσίασε αισθητή βελτίωση της ακρίβειας των προβλέψεων σε σύγκριση με της μεθόδους γραμμικής παλινδρόμησης και GBDT που χρησιμοποιήθηκαν ως σημεία αναφοράς.

4.1.4 An Insurance Sales Prediction Model Based on Deep Learning by Wang, H. (2020)

Το συγκεκριμένο άρθρο χρησιμοποιεί Deep Learning για την πρόβλεψη πωλήσεων στον τομέα των υπηρεσιών, και συγκεκριμένα στον τομέα ασφάλισης. Η χρήση μοντέλων και μηχανικής μάθησης για πρόβλεψη στην πώληση υπηρεσιών δεν έχει μελετηθεί τόσο εκτεταμένα όσο στην πώληση προϊόντων, και συνεπώς αποτελεί μια σχετικά νέα πρόκληση.

Ο συγγραφέας χρησιμοποιεί μια βελτιωμένη μορφή του Long Short-Term Memory (LSTM) network, την οποία ακολουθεί με Multiple Linear Regression (MLR) βασισμένη στην εντολή LINEST, η οποία χρησιμοποιεί τη μέθοδο των ελαχίστων τετραγώνων. Τα δεδομένα που χρησιμοποιήθηκαν προήλθαν από τα 4 τελευταία έτη λειτουργίας (Ιανουαρίου 2016 με Δεκεμβρίου 2019) μια γνωστή ασφαλιστικής εταιρίας, η οποία όμως δεν ονοματίζεται στο άρθρο.

Στο στάδιο της μοντελοποίησης χρησιμοποιήθηκε Gray Correlation Analysis για να βρεθούν οι παράγοντες που επηρεάζουν μη γραμμικά το αποτέλεσμα, και βρέθηκε ότι ήταν κυρίως εποχιακή παράγοντες, όπως διαφήμιση τις υπηρεσίας, προφορες και αργίες. Η τελική μορφή της μοντελοποίησης απεικονίζεται παρακάτω.



The flow of model construction

Εικόνα 4.3: Διάγραμμα ροής της δημιουργίας του μοντέλου [Wang, H., 2020]

Το προτεινόμενο μοντέλο συγκρίθηκε με άλλα προϋπάρχοντα μοντέλα πρόβλεψης, και βρέθηκε να παράγει πιο ακριβή αποτελέσματα. Σε σύγκριση με το GA-optimized Backpropagation Neural Network (BPNN) μειώνει το ποσοστό σφάλματος από 26,77% σε 8,25%, ενώ σε σχέση με το απλό LSTM network μειώνει το ποσοστό σφάλματος κατά 12,5%.

4.1.5 Online sales prediction via trend alignment-based multitask recurrent neural networks by Chen, T., Yin, H., Chen, H., Wang, H., Zhou, X., Li, X. (2020)

Οι συγγραφείς του παρόντος άρθρου προτείνουν τη χρήση ενός νευρωνικού δικτύου που ονομάζουν Trend Alignment with Dual-Attention Multitask Recurrent Neural Network, ή για συντομία TADA+, για την πρόβλεψη πωλήσεων διαδικτυακών καταστημάτων. Με τη χρήση ιστορικών στοιχείων πωλήσεων, στοχεύουν να δημιουργήσουν ένα γενικευμένο μοντέλο που να μπορεί να ανταπεξέλθει σε ένα συνεχώς εξελισσόμενο online περιβάλλον.

Το μοντέλο TADA+ αποτελεί εξέλιξη του προηγούμενου μοντέλου TADA που αναπτύχθηκε από την ίδια ομάδα το 2018, το οποίο βασιζόταν σε ανάλυση LSTM, και το οποίο τώρα προσαρμόζεται για να μπορεί να εφαρμοστεί στη δυναμικώς εξελισσόμενη δομή του διαδικτυακού περιβάλλοντος. Το μοντέλο λαμβάνει υπόψη τόσο τα εσωτερικά χαρακτηριστικά ενός προϊόντος, όσο και τα εξωτερικά, όπως ο καιρός, αργίες, προσφορές κτλ.

Για την εφαρμογή του μοντέλου χρησιμοποιούνται δυο ξεχωριστές πηγές σετ δεδομένων. Το πρώτο προέρχεται από το Kaggle και ονομάζεται Favorita, και περιέχει μεταβλητές και όγκο πωλήσεων όλων των προϊόντων από 56 καταστήματα βασισμένα στο Εκουαδόρ. Το αρχικό σετ δεδομένων καλύπτει όλο το διάστημα από την 1 Ιανουαρίου 2013 ως τις 15 Αυγούστου 2017, αλλά λαμβάνεται

υπόψη μόνο διάστημα ενός έτους από 15 Αυγούστου 2016 έως 15 Αυγούστου 2017, ώστε να αποφευχθεί η ασυνήθης επιρροή που είχε ένας έντονος σεισμός νωρίτερα το έτος 2016. Το δεύτερο προέρχεται από το One Stop Warehouse (OSW), έναν από τους μεγαλύτερους προμηθευτές συσκευών ηλιακής ενέργειας στην Αυστραλία, και καλύπτει το διάστημα από τις 22 Φεβρουαρίου 2016 έως 4 Μαρτίου 2017.

Στον παρακάτω πίνακα αναφέρονται όλες οι μεταβλητές πρόβλεψης που εξήχθησαν και από τα δυο σετ δεδομένων.

| Summarization of features extracted from datasets | | | | | |
|---|------------------|---------------------------------|-----------|----|---|
| Dataset | Type | Feature | Dimension | | |
| Favorita | Internal feature | City of store ^c | 3* | 17 | |
| | | State of store ^c | 2* | | |
| | | Store type ^c | 2* | | |
| | | Store group ^c | 2* | | |
| | | Item family ^c | 3* | | |
| | | Item class ^c | 5* | | |
| | External feature | Promotion state ^b | 1 | 11 | |
| | | Date ^c | 5* | | |
| | | Store transaction ⁿ | 1 | | |
| | | Oil price ⁿ | 1 | | |
| | | Local holiday ^b | 1 | | |
| OSW | Internal feature | National holiday ^b | 1 | 16 | |
| | | Pay day ^b | 1 | | |
| | | Item index ^c | 5* | | |
| | | City of store ^c | 2* | | |
| | | Item category ^c | 5* | | |
| | External feature | Battery type ^c | 3* | | 9 |
| | | Item price ⁿ | 1 | | |
| | | Week number ^c | 4* | | |
| | | Discontinued state ^b | 1 | | |
| | | Solar exposure ⁿ | 1 | | |
| | | Temperature ⁿ | 1 | | |
| Week(s) after last holiday ⁿ | 1 | 1 | | | |
| Week(s) to next holiday ⁿ | 1 | | | | |

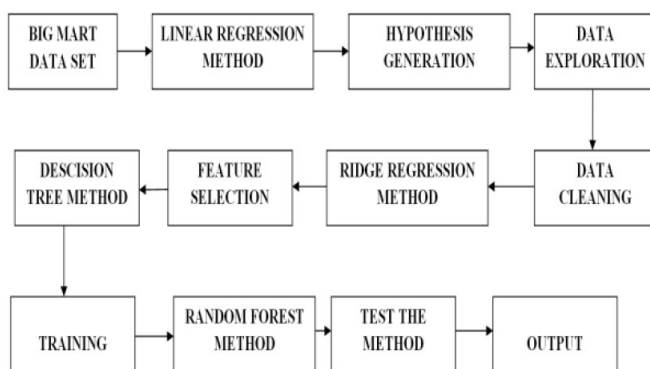
Πίνακας 4.3: Μεταβλητές πρόβλεψης που επιλέχθηκαν από τα δυο σετ δεδομένων [Chen, T., Yin, H., Chen, H., Wang, H., Zhou, X., Li, X., 2020]

Το νέο μοντέλο συγκρίνεται με 4 προϋπάρχοντα μοντέλα πρόβλεψης, και πιο συγκεκριμένα τα Random Forest (RF), XGBoost, SAE-LSTM και Attention Recurrent Neural Network (A-RNN). Επιπροσθέτως, για να μελετηθεί πόσο συνεισφέρει κάθε εξειδικευμένο κομμάτι του μοντέλου τους στο σύνολο, οι μελετητές επαναλαμβάνουν την εφαρμογή για το κάθε κομμάτι, αντικαθιστώντας το με ένα πιο απλουστευμένο.

Για τη σύγκριση των αποτελεσμάτων χρησιμοποιήθηκαν σαν κριτήρια τα Mean Absolute Error (MAE) και Symmetric Mean Absolute Percentage Error (SMAPE). Τα τελικά αποτελέσματα έδειξαν ότι το μοντέλο TADA είχε την χαμηλότερη απώλεια ακρίβειας καθώς αυξάνεται το διάστημα πρόβλεψης, και ανταπεξήλθε πιο αποτελεσματικά στη μετάβαση από μη-διαδικτυακό σε διαδικτυακό περιβάλλον πωλήσεων. Τέλος, επιδεικνύουν ότι το μοντέλο μπορεί να χρησιμοποιηθεί για μεγάλα σετ δεδομένων, καθώς ο χρόνος εφαρμογής αυξάνεται γραμμικά με τον όγκο των δεδομένων.

4.1.6 Prediction of Big Mart Sales Using Exploratory Machine Learning Techniques by Theresa, I., Medikonda, V.R., Reddy, K.V.N. (2020)

Σκοπός του παρόντος άρθρου είναι η απλή δοκιμή ενός μοντέλου δυο επιπέδων για την πρόβλεψη πωλήσεων στην αλυσίδα καταστημάτων Big Mart. Το εν λόγω μοντέλο αναπαριστάται εικονικά στο παρακάτω σχεδιάγραμμα.



BLOCK DIAGRAM OF BIG MART SALE PREDICTION BASED ON MACHINE LEARNING TECHNIQUES

Εικόνα 4.4: Εικονική αναπαράσταση του μοντέλου πρόβλεψης πωλήσεων του Big Mart [Theresa, I., Medikonda, V.R., Reddy, K.V.N., 2020]

Το άρθρο δεν αναλύει σε βάθος τη λειτουργία της μεθόδου, και απλώς παρουσιάζει τα αποτελέσματα σύγκρισης με συνήθη μοντέλα ενός επιπέδου, η οποία σύγκριση δείχνει ότι το μοντέλο δυο επιπέδων έχει μεγαλύτερη ακρίβεια.

4.1.7 Forecasting Techniques for Sales Prediction by Yellamma, P., Abhinav, B., Jaya Vaishnavi, B., Ushaswini, G., Srinivas, M. (2020)

Σε αυτό το άρθρο οι συγγραφείς δοκιμάζουν δυο διαφορετικά προϋπάρχοντα μοντέλα πρόβλεψης πάνω σε ένα σετ δεδομένων που προμηθεύτηκαν από την αλυσίδα Walmart, το οποίο περιέχει τις εβδομαδιαίες πωλήσεις 45 καταστημάτων Walmart σε όλη την Αμερική από το 2010 ως το 2012. Τα μοντέλα που επιλέγουν να εφαρμόσουν είναι τα Auto Regressive Integrated Moving Average (ARIMA) και Feed Forward Neural Network (FFNN).

Τα σετ δεδομένων περιείχαν τις παρακάτω τιμές:

- Store: store number
- Department: department number
- Date: Weekly date
- Weekly Sales: Sales of a given department of a particular store
- IsHoliday: Whether the week has a special holiday
- Temperature: Average Temperature in that region
- Fuel Price: cost of fuel in the region
- Markdown (1-5): Pricing strategy on items
- CPI (Consumer per index): measures changes in price level of goods
- Unemployment: The Unemployment rate
- Store Type: the type of the store (A, B or C)
- Size: Size of the store is based on the number of items in the store

Τα τελικά αποτελέσματα έδειξαν ότι το μοντέλο FFNN είχε καλύτερη δυνατότητα πρόβλεψης, καθώς το μοντέλο ARIMA δεν μπορούσε να κάνει προβλέψεις για κινητές εορτές όπως το Πάσχα.

4.1.8 Sales prediction using machine learning algorithm

by Mohanapriya, S., Mohana Saranya, S. (2020)

Το κατάστημα Big Market στην Ινδία αναζήτησε μελετητές για να δοκιμάσουν και να προτείνουν αλγόριθμους για πρόβλεψη πωλήσεων. Οι συγγραφείς σε αυτό το άρθρο συγκρίνουν τα αποτελέσματα που παράγουν δυο διαφορετικοί προϋπάρχοντες αλγόριθμοι μηχανικής μάθησης, το Gradient Boost (GBT) και το XGBoost, εφαρμόζοντάς τα σε ένα σετ δεδομένων που τους προμήθευσε το Big Market, και περιείχε 8525 εγγραφές με 12 στοιχεία η καθεμία.

Η έρευνά τους διεξήχθη με απλή εφαρμογή των αλγορίθμων μετά από την οποία προέκυψε ότι το GBT είχε Root Mean Square Error 0,8070, Explained Variance Regression Score 0,5594 και Mean Absolute Error 0,7252, ενώ το XGBoost αντίστοιχα είχε RMSE 0,7381, EVRS 0,6519 και MAE 0,7242.

4.1.9 Car Sales Prediction Using Machine Learning Algorithms

by Madhuvanthi, K., Nallakaruppan, M.K., Senthilkumar, N.C., Siva Rama Krishnan, S. (2019)

Οι συγγραφείς του παρόντος άρθρου χρησιμοποιούν μοντέλα μηχανικής μάθησης για την πρόβλεψη πωλήσεων και τον εντοπισμό τάσεων στην βιομηχανία αυτοκινήτων. Η πηγή των δεδομένων που χρησιμοποιήθηκαν για την εφαρμογή των μοντέλων δεν διευκρινίζεται.

Αρχικά εφάρμοσαν τη μέθοδο Analytic Hierarchy Process (AHP) για την επιλογή και βαθμολόγηση των μεταβλητών πρόβλεψης, μετά από την οποία κατέληξαν στην παρακάτω κατάταξη:

1. Τιμή
2. Διάρκεια ντεπόζιτου στην πόλη
3. Διάρκεια ντεπόζιτου στον αυτοκινητόδρομο
4. Είδος καυσίμου
5. Ιπποδύναμη
6. Πλάτος
7. Ύψος
8. Μάρκα μοντέλου

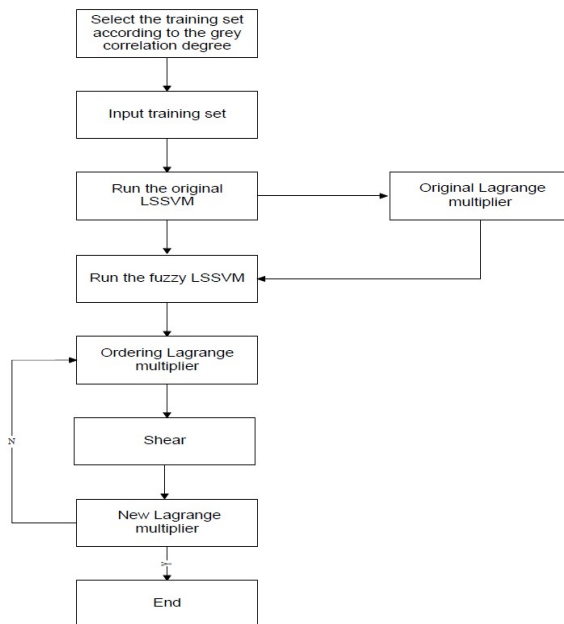
Στη συνέχεια εφάρμοσαν διάφορα μοντέλα μηχανικής μάθησης, και πιο συγκεκριμένα Γραμμική Παλινδρόμηση, Random Tree και Random Forest. Σε ένα επιπλέον παρεμφερές κομμάτι που συμπεριέλαβαν στην έρευνα, εφάρμοσαν και τη μέθοδο Technique for Order of Preference by Similarity to Ideal Solution (TOPSIS) σαν μέτρο αναφοράς για αγοραστές που θέλουν να επιλέξουν αμάξι για αγορά.

Τα τελικά αποτελέσματα της έρευνας έδειξαν ότι η τιμή είναι ο παράγοντας που έχει τον μεγαλύτερο αντίκτυπο της στις πωλήσεις, και σαν επιπλέον παρατήρηση οι συγγραφείς ανέφεραν ότι η Chevrolet φαίνεται να έχει τον υψηλότερο αριθμό πωλήσεων από το σετ δεδομένων τους, με δεύτερη τη Honda.

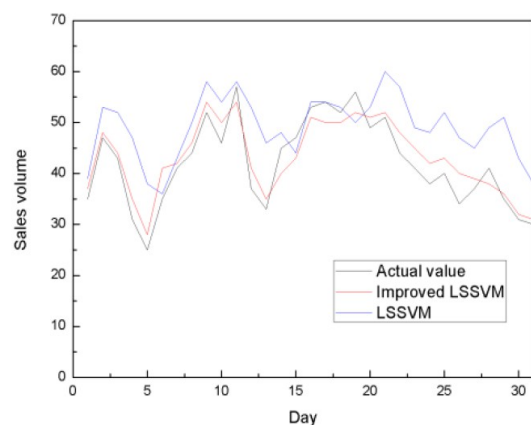
4.1.10 The Research on the Electronic Commerce Sales Prediction based on the Improved LSSVM Algorithm by Dong, X. (2016)

Σε αυτό το άρθρο ο συγγραφέας χρησιμοποιεί μια βελτιωμένη μορφή του μοντέλου Least Squares Support Vector Machine (LSSVM) για την πρόβλεψη διαδικτυακών πωλήσεων. Για την έρευνα χρησιμοποιεί στοιχεία πωλήσεων από τον Ιούλιο του 2012 έως τον Ιούλιο του 2015, χωρίς όμως να διευκρινίζει την προέλευσή τους.

Ομοίως δε δίνει πολλά άλλα στοιχεία, πέρα από τη γενική ροή του μοντέλου, και ένα συγκριτικό διάγραμμα των αποτελεσμάτων, όπως φαίνεται στην παρακάτω εικόνα.



Flow chart of improved LSSVM algorithm



Comparison of experimental results

Εικόνα 4.5: Διάγραμμα ροής λειτουργίας μοντέλου και συγκριτικό διάγραμμα αποτελεσμάτων. [Dong, X.,2016]

Το συγκριτικό διάγραμμα δείχνει ότι σε γενικό βαθμό η βελτιωμένη μορφή του μοντέλου ανταποκρίνεται πιο κοντά στην πραγματικότητα από την απλή μορφή.

4.2 Άρθρα που δημοσιεύτηκαν σε συνέδρια

4.2.1 Adaptive GRU With Regularization Item for Book Sales Prediction

by Xie, H., Liu, L.

Στόχος της έρευνας είναι η ανάπτυξη νέου μοντέλου για την πρόβλεψη πωλήσεων βιβλίων. Το μοντέλο είναι βασισμένο σε Recurrent Neural Networks (RNN) και ονομάστηκε Adaptive Gated Recurrent Unit (AdaGRU). Για την έρευνα χρησιμοποιήθηκαν δυο διαφορετικά σετ δεδομένων, ένα που λήφθηκε ευθέως από μεγάλη αλυσίδα παντοπωλείων, κι ένα από την ανοιχτή βάση δεδομένων GitHub.

Η έρευνα διακρίνεται στο ότι στοχεύει στην πρόβλεψη ωριαίων πωλήσεων. Το μοντέλο συγκρίθηκε με το μοντέλο LSTM και βρέθηκε να έχει πιο ακριβή αποτελέσματα με βάση τους δείκτες MAE και RMSE.

4.2.2 Smart Sales Prediction of Pharmaceutical Products

by Dutta, S.R., Das, S., Chatterjee, P.

Οι συγγραφείς της παρούσας έρευνας εφαρμόζουν 5 διαφορετικά μοντέλα με στόχο να καταλήξουν στο αποτελεσματικότερο για την πρόβλεψη πωλήσεων νέων φαρμάκων. Το φαρμακευτικό σετ δεδομένων που χρησιμοποιήθηκε το προμεθεύτηκαν από την ανοιχτή βάση δεδομένων του GitHub, ενώ τα μοντέλα που δοκιμάστηκαν, για την πρόβλεψη εβδομαδιαίων, μηνιαίων και ετησίων πωλήσεων, ήταν τα:

1. Seasonal Naive
2. Holt-Winters
3. ARIMA
4. Seasonal ARIMA
5. Γραμμική Παλινδρόμηση

Σαν δείκτες αποτελεσματικότητας χρησιμοποιήθηκαν οι RMSE, MAE και MAPE, και από τα αποτελέσματα βρέθηκε το χαμηλότερο σφάλμα σε όλους τους δείκτες να προκύπτει από το μοντέλο της Γραμμικής Παλινδρόμησης.

4.2.3 New Energy Vehicles Sales Prediction Model Combining the Online Reviews Sentiment

Analysis: A Case Study of Chinese New Energy Vehicles Market

by Du, Y., Wei, K., Wang, Y., Jia, J. (2022)

Η παρούσα έρευνα αφορά την πρόβλεψη πωλήσεων “οχημάτων νέας ενέργειας”, τα οποία ορίζει ως οχήματα που είτε χρησιμοποιούν μη συμβατικές μορφές καυσίμων για την λειτουργία τους, είτε ενσωματώνουν νέες τεχνολογίες για την πιο αποτελεσματική διαχείριση της ενέργειας, καθώς οι πωλήσεις τους αυξήθηκαν σημαντικά από το 2015 έως το 2020. Τα στοιχεία που χρησιμοποιήθηκαν για την έρευνα συλλέχθηκαν από ανοιχτές βάσεις δεδομένων.

Αρχικά γίνεται απλή χρήση μοντέλου SARIMA μέσω της συνάρτησης `auto_arima` της Python σαν μέτρο σύγκρισης, και χρησιμοποιείται για να γίνει πρόβλεψη πωλήσεων για διαστήματα 7

διαφορετικών μηνών που επιλέχθηκαν τυχαία. Στη συνέχεια γίνεται εκπαίδευση του μοντέλου Random Forest με ιστορικά στοιχεία, κάνοντας διαχωρισμό εκπαίδευσης και ελέγχου 9:1. Αφού έχει γίνει η επιλογή των μεταβλητών πρόβλεψης και η εκπαίδευση του μοντέλου, χρησιμοποιείται το μοντέλο SARIMA για πρόβλεψη των τιμών που θα παρουσιάσουν οι μεταβλητές πρόβλεψης κατά το επιθυμητό διάστημα. Σαν τελευταίο βήμα εισάγονται οι προβλεπόμενες τιμές στο μοντέλο Random Forest για να παραχθεί η τελική πρόβλεψη πωλήσεων.

Για την αξιολόγηση των μοντέλων χρησιμοποιήθηκαν οι δείκτες RMSE και R2, οι οποίοι έδειξαν ότι το μοντέλο που προτάθηκε είχε καλύτερη απόδοση από το αρχικό μοντέλο SARIMA. Σε μια επιπρόσθετη ενότητα επαναλαμβάνεται η διαδικασία συμπεριλαμβάνοντας στο μοντέλο και σχόλια χρηστών, το οποίο βελτιώνει επιπλέον το τελικό αποτέλεσμα.

4.2.4 Medical Equipment Sales Management Prediction System Based on LSTM Algorithm by Hu, Y., Xu, B.

Σε αυτήν την έρευνα οι συγγραφείς χρησιμοποιούν τον αλγόριθμο LSTM για την πρόβλεψη πωλήσεων ιατρικού εξοπλισμού. Στόχος τους είναι η ενσωμάτωση του αλγορίθμου σε υπάρχον σύστημα διαχείρισης αγορών, πωλήσεων και αποθήκευσης ιατρικού εξοπλισμού. Κατά συνέπεια, στα πλαίσια της έρευνας δοκιμάζουν την αποτελεσματικότητα του αλγορίθμου στην πρόβλεψη των μηνιαίων πωλήσεων για την εξυπηρέτηση της λειτουργίας του συστήματος.

4.2.5 Identifying Leading Indicators for Tactical Truck Parts' Sales Predictions Using LASSO by Gerritsen, D., Reshadat, V.

Πρωταρχικός στόχος της παρούσας έρευνας είναι η αναγνώριση των βέλτιστων μεταβλητών πρόβλεψης για τις πωλήσεις εξαρτημάτων οχημάτων, προτού χρησιμοποιηθούν στην πράξη για την ίδια την πρόβλεψη. Πιο συγκεκριμένα, μετά την επιλογή των βέλτιστων μεταβλητών εφαρμόζει και συγκρίνει 4 διαφορετικούς αλγορίθμους για την πρόβλεψη μηνιαίων πωλήσεων μεταχειρισμένων εξαρτημάτων φορτηγών. Τα δεδομένα που χρησιμοποιήθηκαν προήλθαν από εταιρίες, αλλά και από διάφορες στατιστικές υπηρεσίες. Συνολικά τα σετ δεδομένων περιείχαν 33 διαφορετικές μεταβλητές πρόβλεψης, από τις οποίες τελικά διακρίθηκαν 7, οι οποίες όμως δεν διευκρινίζονται για λόγους εμπιστευτικότητας.

Στη συνέχεια εφαρμόστηκαν οι μέθοδοι LASSO, Additive Holt-Winters, Multiplicative Holt-Winters, και Season ARIMA για τη χρήση των μεταβλητών αυτών στην πρόβλεψη των πωλήσεων. Για την αξιολόγηση των αποτελεσμάτων χρησιμοποιήθηκε ο δείκτης MAPE, και βρέθηκε ότι εάν η πρόβλεψη αφορούσε του 5 επόμενου μήνες η μέθοδος Additive Holt-Winters έδινε τα πιο αξιόπιστα αποτελέσματα, ενώ για πιο μακροπρόθεσμες προβλέψεις την ξεπερνούσαν οι LASSO και Seasonal ARIMA.

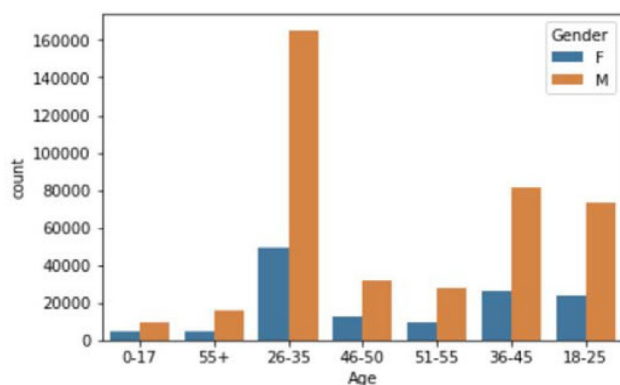
4.2.6 Machine Learning Application for Black Friday Sales Prediction Framework by Ramachandra, H.V., Balaraju, G., Rajashekar, A., Patil, H. (2021)

Σε αυτό το άρθρο οι συγγραφείς παίρνουν μια πελατοκεντρική προσέγγιση για την πρόβλεψη του αριθμού πωλήσεων κατά τη διάρκεια της Black Friday, που λόγω του μεγάλου αριθμού προσφορών βλέπει αντίστοιχα μεγάλες διακυμάνσεις στον όγκο πωλήσεων. Στόχος της έρευνας, παράλληλα με την πρόβλεψη πωλήσεων, είναι και η δημογραφική ανάλυση της κατανομής αυτών των πωλήσεων στην πελατεία.

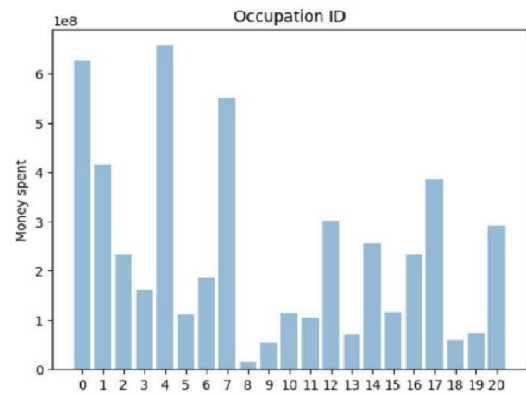
Για την έρευνα χρησιμοποιείται ένα σετ δεδομένων που αποτελείται από 8523 εγγραφές και 12 μεταβλητές, οι οποίες είναι:

1. User ID
2. Product ID
3. Ηλικία
4. Φύλο
5. Επάγγελμα
6. Κατηγορία πόλης
7. Χρόνος διαμονής στην τρέχουσα πόλη
8. Έγγαμη κατάσταση
9. Κατηγορία προϊόντος 1
10. Κατηγορία προϊόντος 2
11. Κατηγορία προϊόντος 3
12. Αγορές

Τα μοντέλα που επιλέχθηκαν για την διεξαγωγή της έρευνας, καθώς φαινόταν να δίνουν πιο ακριβή αποτελέσματα, ήταν το Random Forest Regressor & XGBoost, με ακρίβεια της τάξης του 81%. Σαν παράδειγμα της δημογραφικής ανάλυσης των παραγόμενων του μοντέλου δίνονται τα παρακάτω διαγράμματα, τα οποία δίνουν την πρόβλεψη πωλήσεων ανά κατηγορία ηλικίας και φύλου, καθώς και ανά κατηγορία επαγγέλματος.



Graphical analysis of no. of customers vs Age group



"Money spent" analysis with respect to the the people's occupation ID's

Εικόνα 4.6: Προβλέψεις πωλήσεων ανα κατηγορία ηλικίας, φύλου και επαγγέλματος [Ramachandra, H.V., Balaraju, G., Rajashekar, A., Patil, H., 2021]

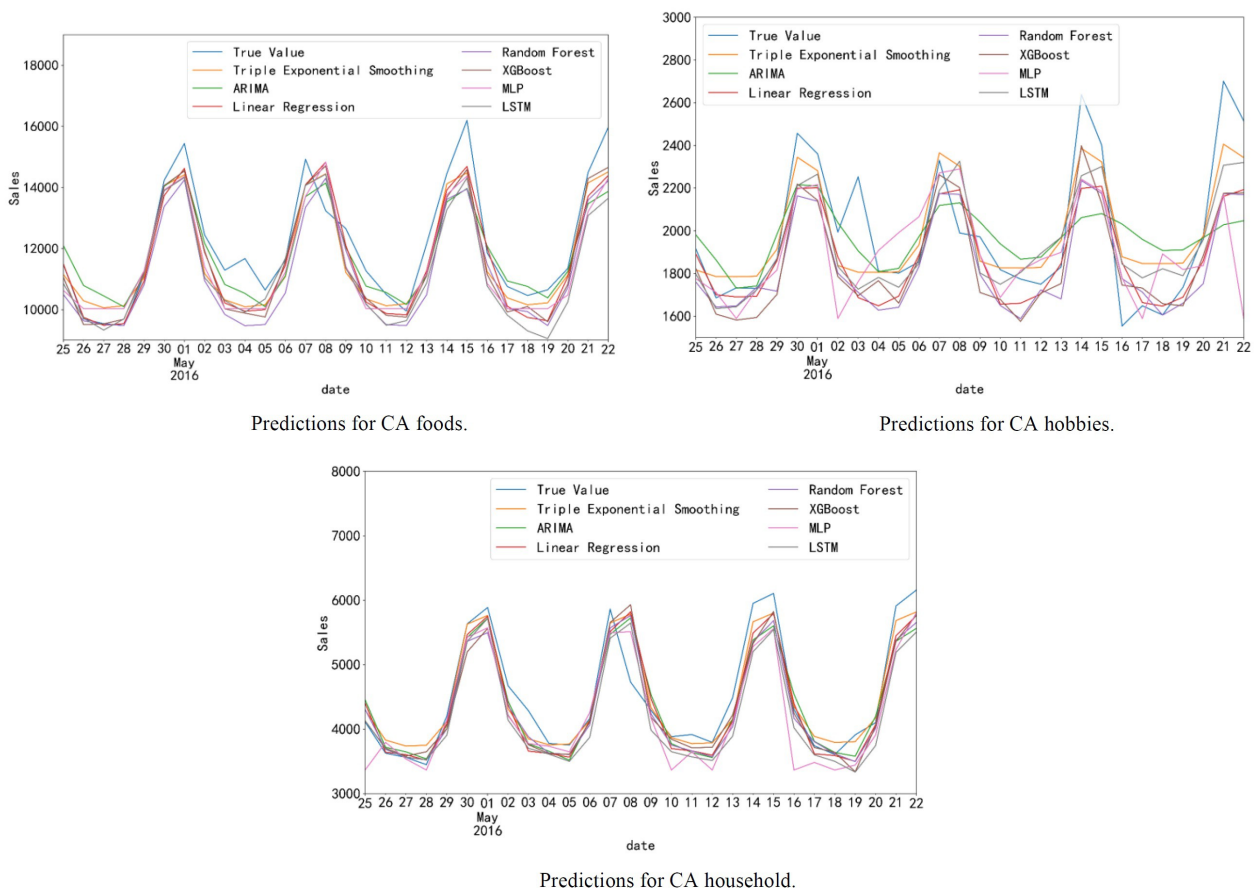
Από τα παραπάνω προκύπτει για παράδειγμα ότι ο μεγαλύτερος αριθμός πωλήσεων προέρχεται από πελάτες στην ηλικιακή ομάδα των 26-35 ετών.

4.2.7 Sales Prediction based on Machine Learning by Huo, Z. (2021)

Σε αυτό το άρθρο ο συγγραφέας δοκιμάζει και παρουσιάζει τα αποτελέσματα από έναν μεγαλύτερο αριθμό μοντέλων συγκριτικά με άλλα άρθρα, καθώς επιλέγει να ακολουθήσει συγκριτική προσέγγιση σε μια ποικιλία προϋπάρχοντων μοντέλων αντί να αναπτύξει καινούριο. Πιο συγκεκριμένα, χρησιμοποιεί τα μοντέλα ARIMA, Regression, XGBoost, Random Forest, Multi-Layer Perception (MLP) και Long Short-Term Memory (LSTM).

Τα δεδομένα του προέρχονται από το M5 σετ δεδομένων της αλυσίδας καταστημάτων WalMart, το οποίο περιέχει δεδομένα πωλήσεων 3049 προϊόντων, τα οποία μοιράζονται σε 3 κατηγορίες: Χόμπυ, τρόφιμα και οικιακά. Το σετ αφορά πωλήσεις από 10 καταστήματα μοιρασμένα σε 3 πολιτείες της Αμερικής, την Καλιφόρνια, το Τέξας και το Γουισκόνσιν. Συνεπώς για την ανάλυση ο συγγραφέας μοιράζει τα δεδομένα σε 9 συνολικές κατηγορίες, ανά το είδος προϊόντος και την πολιτεία. Οι μεταβλητές πρόβλεψης που περιέχουν τα σετ δεδομένων δεν αναφέρονται συγκεκριμένα, αλλά αναφέρεται ότι συμπεριλαμβάνουν πληροφορίες για τα είδη και τις τιμές προϊόντων, καθώς και τις ημερήσιες πωλήσεις.

Το άρθρο δίνει τα συγκριτικά διαγράμματα προβλέψεων για τα καταστήματα στην πολιτεία της Καλιφόρνια, και σχολιάζει ότι περισσότερες πληροφορίες σχετικά με το χρόνο και την τιμή βοηθούν στην πρόβλεψη πωλήσεων.



Εικόνα 4.7: Συγκριτικά διαγράμματα προβλέψεων για τρόφιμα, χόμπυ και οικιακά είδη στην Καλιφόρνια [Huo, Z., 2021]

4.2.8 A Sales Prediction Method Based on LSTM with Hyper-Parameter Search by Dai, Y., Huang, J. (2021)

Στο παρόν άρθρο οι συγγραφείς συγκρίνουν ένα εξειδικευμένο μοντέλο Long Short-Term Memory (LSTM) με διάφορα μοντέλα μηχανικής μάθησης που εφαρμόζουν με τη χρήση του εργαλείου AutoML. Το μοντέλο που προτείνεται προσθέτει ειδικής συνάρτησης απώλειας και τη χρήση Hyper-Parameter Search στο κλασσικό μοντέλο LSTM, και στη συνέχεια συγκρίνεται με μοντέλα όπως KNN, Random Forest, CatBoost, XGBoost, LightBGM και Stacking.

Η ακρίβεια των αποτελεσμάτων όλων των μοντέλων παρουσιάζεται στον παρακάτω πίνακα:

Result of weekly predictions of different models

| Models | First week | | Second week | | Third week | | Fourth week | |
|-------------------------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|--------------|
| | MAPE | RMSPE | MAPE | RMSPE | MAPE | RMSPE | MAPE | RMSPE |
| KNN | 3.281 | 4.252 | 3.624 | 4.812 | 3.823 | 4.568 | 2.912 | 3.253 |
| Random Forest | 4.235 | 3.950 | 3.598 | 4.132 | 3.722 | 4.232 | 3.974 | 4.135 |
| AutoML | | | | | | | | |
| CatBoost | 0.392 | 0.473 | 0.741 | 0.987 | 0.823 | 0.971 | 0.412 | 0.435 |
| XGBoost | 0.353 | 0.413 | 0.212 | 0.133 | 0.498 | 0.593 | 0.653 | 0.857 |
| LightBGM | 0.217 | 0.438 | 0.587 | 0.633 | 0.227 | 0.396 | 0.261 | 0.356 |
| Stacking | 0.302 | 0.343 | 0.498 | 0.689 | 0.579 | 0.987 | 0.210 | 0.341 |
| LSTM | | | | | | | | |
| LSTM+ | 0.043 | 0.056 | 0.088 | 0.178 | 0.184 | 0.133 | 0.594 | 0.727 |
| Hyper-Parameter Search | 0.024 | 0.032 | 0.125 | 0.183 | 0.035 | 0.087 | 0.327 | 0.382 |

Πίνακας 4.4: Εβδομαδιαίες προβλέψεις από τα διάφορα μοντέλα [Dai, Y., Huang, J., 2021]

Οι συγγραφείς σχολιάζουν ότι το μοντέλο τους προσαρμόζεται καλύτερα στα δεδομένα για πιο ακριβή αποτελέσματα πρόβλεψης, με το κυριότερο ελάττωμα του να είναι η αδυναμία εφαρμογής του σε νέα προϊόντα.

4.2.9 E-Commerce System for Sale Prediction Using Machine Learning Technique by Singh, K., Booma, P.M., Eaganathan, U. (2020)

Σκοπός των συγγραφέων σε αυτό το άρθρο είναι να δοκιμάσουν διαφορετικά μοντέλα μηχανικής μάθησης για την επίλυση ενός προβλήματος, ώστε με τη δοκιμή να καταλήξουν στο μοντέλο με τη μεγαλύτερη ακρίβεια το οποίο και θα χρησιμοποιήσουν για τη δημιουργία εφαρμογής.

Το πρόβλημα με το οποίο επέλεξαν να ασχοληθούν στη συγκεκριμένη περίπτωση ήταν η πρόβλεψη ηλεκτρονικών πωλήσεων, χρησιμοποιώντας σαν σετ δεδομένων στοιχεία που ανέβασε στο Kaggle το βραζιλιάνικο E-Commerce site Olist Store. Τα δεδομένα περιλαμβάνουν περίπου 100.000 εγγραφές από το 2016 ως το 2018. Τα μοντέλα τα οποία δοκιμάστηκαν ήταν τα Random Forest, Gradient Boosting, ARIMA και SARIMA.

Για τα δυο πρώτα μοντέλα, παρόλο που το Gradient Boosting είχε καλύτερη προσαρμογή στο σετ δεδομένων εκπαίδευσης, με ακρίβεια 99,99% σε σύγκριση με 90,99%, υστερούσε όσον αφορά την προβλεψιμότητα για το σετ δεδομένων ελέγχου, με ακρίβεια 87,39% σε σύγκριση με 86,88%. Ομοίως το Random Forest είχε και χαμηλότερα RSME και MAE, συνεπώς προκύπτει ότι είναι το

ακριβέστερο μοντέλο. Τα άλλα δυο μοντέλα έδωσαν σχετικά κοντινά αποτελέσματα, αλλά τελικά επιλέχθηκε το μοντέλο SARIMA, καθώς είχε χαμηλότερο Mean Absolute Percentage Error (MAPE), το οποίο θεωρήθηκε το πιο αξιόπιστο κριτήριο. Το SARIMA επίσης επιτρέπει τον εντοπισμό εποχιακών τάσεων.

4.2.10 A hybrid machine learning model for sales prediction by Wang, J. (2020)

Στο παρόν άρθρο ο συγγραφέας προτείνει ένα υβριδικό μοντέλο μηχανικής μάθησης βασισμένο στο συνδυασμό των προϋπάρχοντων μοντέλων LightGBM και XGBoost. Τα σετ δεδομένων που χρησιμοποιήθηκαν για την έρευνα προέρχονται από τις πωλήσεις της αλυσίδας καταστημάτων Wal-Mart σε 3 πολιτείες της Αμερικής, την Καλιφόρνια, το Τέξας, και το Γουισκόνσιν. Μεταξύ άλλων, τα σετ δεδομένων περιείχαν τις μεταβλητές:

- State ID
- Store ID
- time
- item level
- department
- product category
- price
- promotion
- weekday
- special events

Οι πρώτες 1413 ημέρες πωλήσεων χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου, το οποίο δοκιμάστηκε ύστερα για την πρόβλεψη των επόμενων 28 ημερών. Τα αποτελέσματα του προτεινόμενου μοντέλου συγκρίθηκαν με τα αντίστοιχα των μοντέλων γραμμικής παλινδρόμησης, SVM, καθώς και με τα μεμονομένα μοντέλα LightGBM και XGBoost, και η απόδοσή τους παρουσιάζεται στον παρακάτω πίνακα.

PERFORMANCE COMPARISON OF MODELS

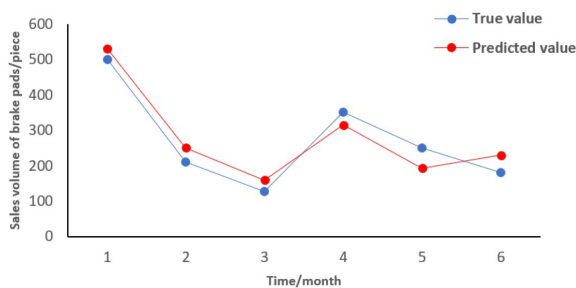
| Models | RMSE |
|-------------------|-------------|
| Linear Regression | 3.35 |
| SVM | 2.88 |
| LightGBM | 2.09 |
| Xgboost | 2.11 |
| Ensembled | 2.07 |

Πίνακας 4.5: Σύγκριση απόδοσης των μοντέλων [Wang, J., 2020]

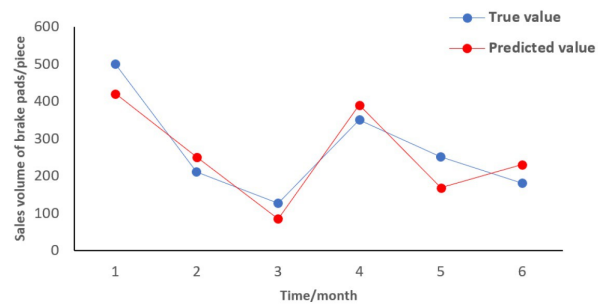
Προκύπτει συνεπώς ότι το συνδυαστικό μοντέλο παρουσιάζει βελτίωση σε σύγκριση με τα μεμονομένα μοντέλα που περιλαμβάνει.

4.2.11 Auto Parts Sales Prediction Based on Machine Learning for small data and a long replacement cycle by Liao, W., Ye, G., Yin, Y., Yan, W., Ma, Y., Zuo, D. (2020)

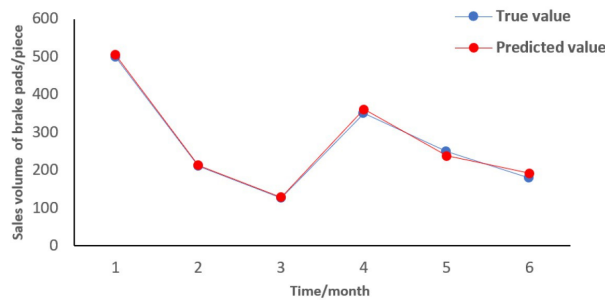
Στο συγκεκριμένο άρθρο γίνεται απλή δοκιμή προϋπάρχοντων μοντέλων μηχανικής μάθησης, χρησιμοποιώντας δημόσιο σετ δεδομένων από το UCI Machine Learning Repository. Το αντικείμενο αφορά την πρόβλεψη πωλήσεων εξαρτημάτων αυτοκινήτου με μικρό όγκο δεδομένων, ενώ τα μοντέλα που δοκιμάζονται είναι η Γραμμική Παλινδρόμηση, το SVR και το Stacking. Δε δίνονται συγκεκριμένες πληροφορίες σχετικά με τις μεταβλητές πρόβλεψης που χρησιμοποιήθηκαν, απλά παρουσιάζονται τα αποτελέσματα που φαίνονται στα παρακάτω διαγράμματα.



Prediction results of the Linear regression model.



Prediction results of the SVR model.



Prediction results obtained using the stacking model.

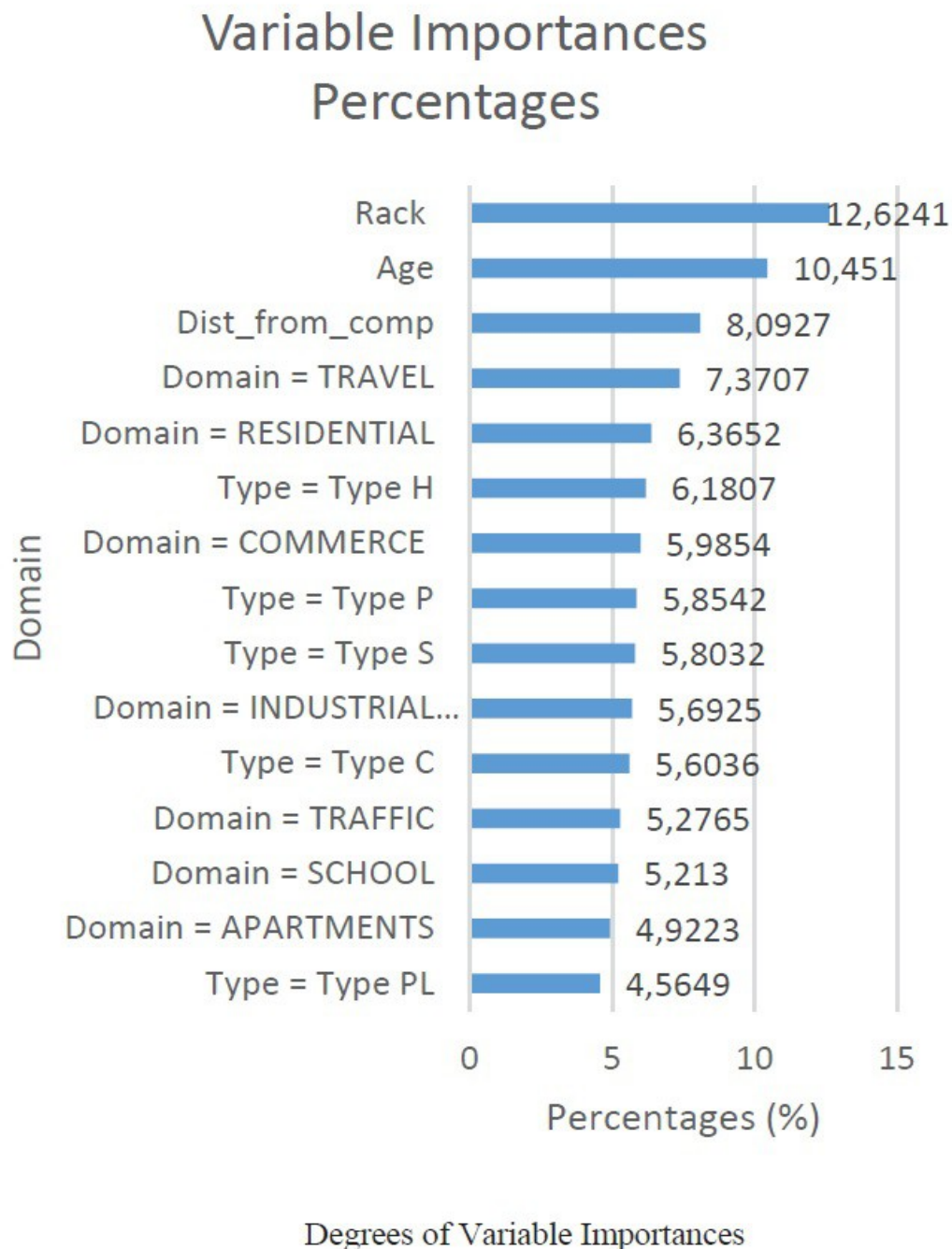
Εικόνα 4.8: Αποτελέσματα προβλέψεων για τα μοντέλα γραμμικής παλινδρόμησης, SVR και Stacking [Liao, W., Ye, G., Yin, Y., Yan, W., Ma, Y., Zuo, D., 2020]

Τα αποτελέσματα δείχνουν ότι ενώ τα μοντέλα Γραμμικής Παλινδρόμησης και SVR ακολουθούν τη σωστή τάση στο διάγραμμα, και τα δυο παρουσιάζουν αποκλίσεις. Το μοντέλο Stacking αντίθετα φαίνεται να κάνει πιο ακριβή πρόβλεψη με μικρό όγκο δεδομένων.

4.2.12 Sales Prediction Analysis in Determining New Minimarket Stores by Edwardo, T.O., Ruldeviyani, Y. (2020)

Η παρούσα έρευνα ξεκίνησε με στόχο την ανάπτυξη μεθόδου για την πρόβλεψη επιτυχίας στις πωλήσεις νέων καταστημάτων της αλυσίδας mini-market PT XYZ στην Ινδονησία. Οι συγγραφείς επέλεξαν τις μεθόδους Data Mining και Deep Learning για την ανάπτυξη του μοντέλου τους, ενώ το σετ δεδομένων το προμήθευσε η εταιρεία PT XYZ και περιείχε 2639 εγγραφές.

Το μοντέλο έδειξε τα παρακάτω επίπεδα επιρροής των μεταβλητών πρόβλεψης στις πωλήσεις.



Εικόνα 4.9: Σημαντικότητα πρόβλεψης των μεταβλητών [Edwardo, T.O., Ruldeviyani, Y., 2020]

Τελικά οι συγγραφείς συνέστησαν στην αλυσίδα να χρησιμοποιήσει τη μέθοδο Deep Learning για την απόφαση ίδρυσης νέων καταστημάτων, με την πρόταση να εγκρίνεται η ίδρυση εφόσον το μοντέλο δείξει πωλήσεις τουλάχιστον όσο τις θεωρητικές πωλήσεις στην πρόταση ίδρυσης, με ανοχή τουλάχιστον 20%. Επίσης προτείνουν την ενσωμάτωση δημογραφικών στοιχείων σε τυχόν επέκταση του μοντέλου.

4.2.13 Prediction Analysis Sales for Corporate Services Telecommunications Company using Gradient Boost Algorithm by Wisesa, O., Adriansyah, A., Khalaf, O.I. (2020)

Στην παρούσα έρευνα οι συγγραφείς δοκιμάζουν το μοντέλο Gradient Boosted Trees για την πρόβλεψη πωλήσεων στον τομέα των τηλεπικοινωνιών. Το σετ δεδομένων που χρησιμοποιήθηκε περιείχε στοιχεία πωλήσεων για τα έτη 2016-2018 από εταιρεία τηλεπικοινωνιών που δεν ονοματίζεται στο άρθρο, η οποία λειτουργεί υπό μοντέλο Business-to-Business (B2B), ενώ οι μεταβλητές πρόβλεψης που περιείχε ήταν: category, region, item type and opportunity ID, quarter, product name, product sub-component, service product (MIDI) & sales revenue.

Οι συγγραφείς κατέληξαν ότι η δυνατότητα πρόβλεψης του μοντέλου ήταν ικανοποιητική.

4.2.14 A knowledge-intensive methodology for explainable sales prediction by Cristani, M., Pasetto, L., Tomazzoli, C. (2020)

Με αφορμή τα προβλήματα που προξένησε σε ορισμένους εμπορικούς τομείς η πανδημία του COVID-19, οι συγγραφείς του παρόντος άρθρου επέλεξαν να επικεντρωθούν στις μεθόδους πρόβλεψης πωλήσεων στον τομέα τροφίμων, και να επιδείξουν ότι οι δεδομένο-κεντρικές (data-driven) μέθοδοι που χρησιμοποιούνται συνήθως είναι λιγότερο αποτελεσματικές από γνωσεο-κεντρικές (knowledge-driven).

Σαν δεδομένα χρησιμοποίησαν τα στοιχεία πωλήσεων εμπόρου που δεν ονοματίζεται στο κείμενο. Για την εκπαίδευση των μοντέλων χρησιμοποιήθηκαν στοιχεία πωλήσεων από τον Μάιο του 2018 έως τον Μάιο του 2019, και αναφέρεται ότι περιέχουν στοιχεία τόσο σταθερών (Continuous Customer Table) όσο και περιοδικών (Promotional Customer Table) πελατών, κι επίσης ότι αφορούν τόσο φέσκα τρόφιμα με κοντινή ημερομηνία λήξης όσο και κατεψυγμένα με μακρύτερη. Για την επαλήθευση των αποτελεσμάτων χρησιμοποιήθηκαν τα στοιχεία πωλήσεων του διαστήματος από τον Σεπτέμβριο έως Δεκέμβριο 2019 ώστε να συγκριθούν με τα εξαγόμενα των μοντέλων.

Αφού ανέλυσαν το πρόβλημα με το οποίο επιθυμούσαν να ασχοληθούν, οι συγγραφείς έτρεξαν την έρευνα τους σε 3 στάδια:

- Στο πρώτο στάδιο δοκίμασαν διάφορα προϋπάρχοντα μοντέλα πρόβλεψης, όπως Νευρωνικά Δίκτυα, Associative Rules, ILP & Random Forest. Το Random Forest αναδείχθηκε ως η πιο χρήσιμη μέθοδος.
- Στο δεύτερο στάδιο δοκίμασαν να κάνουν βελτιώσεις στη μέθοδο Random Forest, και τελικά λόγω περιορισμών στο μοντέλο καταλήγουν στην ανάπτυξη μεθόδου βασισμένη στο μοντέλο Bayes, την οποία ονομάζουν Sales Forecasting Belief Network (SFBN), η οποία ξεπερνάει σε αποτελεσματικότητα την Random Forest.

- Στο τρίτο και τελευταίο στάδιο εισάγουν βελτιώσεις στο μοντέλο SFBN για να παράξουν το τελικό μοντέλο SFNB+.

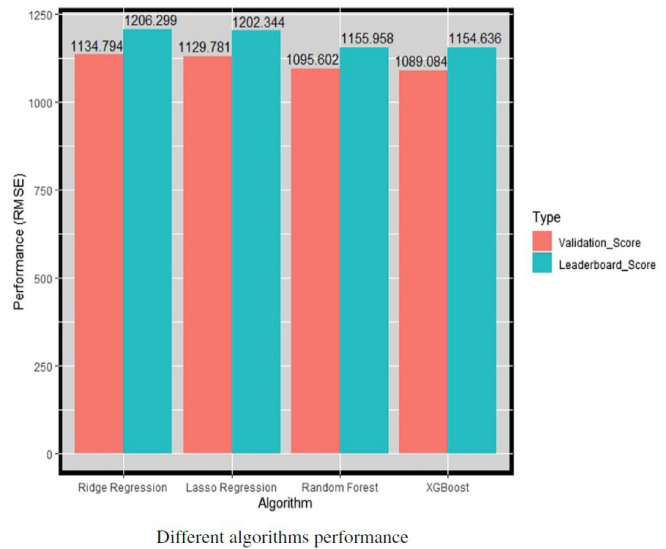
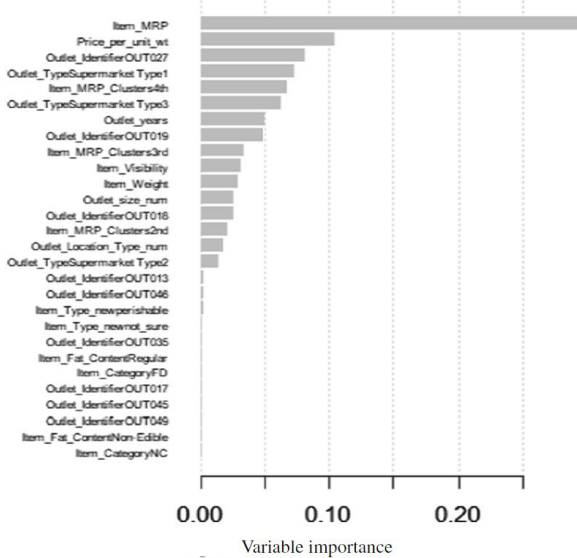
Το τελικό συμπέρασμα των συγγραφέων είναι ότι γνωστο-κεντρικές μέθοδοι είναι πιο αξιόπιστες και πιο γρήγορες από τις κοινές μεθόδους μηχανικής μάθησης.

4.2.15 Model Performance Evaluation: Sales Prediction by Singh, A., Srivastava, S. (2020)

Σκοπός της παρούσας έρευνας ήταν να δοκιμαστούν ένας αριθμός μοντέλων μηχανικής μάθησης ώστε να εντοπιστεί το πιο αποτελεσματικό. Για την έρευνά τους, οι συγγραφείς χρησιμοποιούν δεδομένα από ένα διαγωνισμό Hackathon που οργάνωσε η ιστοσελίδα Analytics Vidya. Το σετ δεδομένων περιέχει δεδομένα πωλήσεων της αλυσίδας καταστημάτων Big Mart, το οποία περιέχει 8523 εγγραφές με 12 μεταβλητές που αφορούν 1559 διαφορετικά προϊόντα που πωλούνται από 10 διαφορετικά καταστήματα. Οι 12 προαναφερόμενες μεταβλητές είναι οι εξής:

- Item_Identifier
- Item_Weight
- Item_Fat_Content
- Item_Visibility
- Item_Type
- Item_MRP
- Outlet_Identifier
- Outlet_Establishment_Year
- Outlet_Size Size of Store
- Outlet_Location_Type
- Outlet_Type
- Item_Outlet_Sales

Τα μοντέλα που επέλεξαν οι συγγραφείς να χρησιμοποιήσουν για την έρευνά τους ήταν η Γραμμική Παλινδρόμηση, Ridge Regression, Lasso Regression, Random Forest & XGBoost. Στα παρακάτω διαγράμματα παρουσιάζεται τόσο η σημαντικότητα που βρέθηκε σε καθεμία από τις παραπάνω μεταβλητές, όσο και η αξιολόγηση του κάθε μοντέλου.



Εικόνα 4.10: Σημαντικότητα πρόβλεψης των μεταβλητών & αξιολόγηση των μοντέλων [Singh, A., Srivastava, S., 2020]

Προέκυψε λοιπόν ότι το XGBoost έδωσε τα πιο αξιόπιστα αποτελέσματα, ακόμα και σε σύγκριση με το Random Forest, το οποίο οι συγγραφείς σχολιάζουν ότι έως τότε θεωρούνταν το καλύτερο μοντέλο.

4.2.16 A Comparative Study of Big Mart Sales Prediction by Behera, G., Nain, N. (2020)

Στο παρόν άρθρο οι συγγραφείς δοκιμάζουν μια πληθώρα μοντέλων μηχανικής μάθησης για την πρόβλεψη πωλήσεων της αλυσίδας καταστημάτων Big Mart, ώστε να καταλήξουν στο πιο αποτελεσματικό και να προτείνουν τη χρήση του, το οποίο στην συγκεκριμένη περίπτωση προέκυψε να είναι το XGBoost. Σαν δεδομένα χρησιμοποιούν στοιχεία πωλήσεων του Big Mart που αφορούν πωλήσεις 1559 διαφορετικών προϊόντων από 10 διαφορετικά καταστήματα κατά το έτος 2013, ενώ για την αξιολόγηση των μοντέλων χρησιμοποιούν τα Mean Absolute Error (MAE) και Root Mean Square Error (RMSE).

Τα σετ δεδομένων που τους δόθηκαν περιείχαν 8523 εγγραφές και αποτελούνταν από τις παρακάτω 12 μεταβλητές:

- Item Fat
- Item Type
- Item MRP
- Outlet Type
- Item Visibility
- Item Weight
- Outlet Identifier
- Outlet Size
- Outlet Establishment Year
- Outlet Location Type
- Item Identifier
- Item Outlet Sales

Για την εκπαίδευση και επαλήθευση των μοντέλων έγινε διαμοιρασμός των εγγραφών 80-20 αντίστοιχα. Τα μοντέλα που χρησιμοποιήθηκαν πέρα από το XGBoost για σύγκριση ήταν η Γραμμική Παλινδρόμηση, Decision Tree & Ridge Regression. Τα τελικά MEA & RMSE που προέκυψαν φαίνονται στον παρακάτω πίνακα.

Comparison of MAE and RMSE of proposed model with other model

| Model | MAE | RMSE |
|-------------------|---------------|-------------|
| Linear regression | 836.1 | 1127 |
| Decision tree | 751.6 | 1068 |
| Ridge regression | 836 | 1129 |
| Xgboost | 749.03 | 1066 |

Πίνακας 4.6: Σύγκριση δεικτών σφάλματος για τα διάφορα μοντέλα [Behera, G., Nain, N., 2020]

Προκύπτει λοιπόν ότι όντως η μέθοδος XGBoost που κατέληξαν να προτείνουν οι συγγραφείς είχε το χαμηλότερο σφάλμα και στις δυο κατηγορίες.

4.2.17 Sales Prediction of Svijany Slovakia, Ltd. Using Microsoft Azure Machine Learning and ARIMA by Haffner, O., Kucera, E., Moravcik, M. (2020)

Σε αυτή την έρευνα οι συγγραφείς χρησιμοποίησαν το εργαλείο Microsoft Azure Machine Learning, το οποίο προσφέρει ένα εικονικό περιβάλλον για τη μοντελοποίηση μηχανικής μάθησης, καθώς και το μοντέλο ARIMA. Σκοπός της έρευνας ήταν η πρόβλεψη πωλήσεων της εταιρείας Svijany Slovakia, Ltd., η οποία εμπορεύεται τσέχικες μπύρες στη Σλοβακία.

Τα δεδομένα που χρησιμοποιήθηκαν αφορούσαν τις πωλήσεις της εταιρείας για το χρονικό διάστημα από τις 01.09.2014 μέχρι και τις 31.05.2016. Οι συγγραφείς τα εξήγαγαν από το σύστημα Helios Green, και ανάμεσα σε άλλες πληροφορίες περιείχαν ημερήσιες πωλήσεις του κάθε προϊόντος της εταιρείας.

Οι συγγραφείς σχολιάζουν ότι τα αποτελέσματα της ανάλυσής τους, που διεξήχθη το 2016, χρησιμοποιήθηκαν το επόμενο έτος για την υποστήριξη της απόφασης κατασκευής νέας αποθήκης μεγαλύτερης χωρητικότητας.

4.2.18 A Q-Learning Approach for Sales Prediction in Heterogeneous Information Networks by Kodali, S., Dabbiru, M., Rao, B.T. (2020)

Στο παρόν άρθρο οι συγγραφείς χρησιμοποιούν μια τεχνική που ονομάζεται Q-Learning, την οποία αποκαλούν “model-free reinforcement learning”, και η οποία μαθαίνει από τα αποτελέσματα προηγούμενων δράσεων. Στην έρευνα γίνεται απλή εφαρμογή της τεχνικής χωρίς να σχολιαστούν τα αποτελέσματα, πάνω σε ένα σετ δεδομένων που περιέχει στοιχεία πωλήσεων κι εσόδων από iPhone κατά τα έτη 2007-2009, και παράγεται πρόβλεψη για την εξέλιξη του αριθμού των πωλήσεων για τα επόμενα έτη έως το 2016.

4.2.19 Click and Sales Prediction for Digital Advertisements: Real World Application for OTAs by Tekin, A.T., Cebi, F. (2020)

Ενώ στις περισσότερες περιπτώσεις τα μοντέλα μηχανικής μάθησης εφαρμόζονται σε τομείς που αφορούν φυσικά αγαθά, η συγκεκριμένη έρευνα ασχολείται με υπηρεσίες, και συγκεκριμένα τον τομέα του τουρισμού. Η έρευνα διεξάχθηκε με τη χρήση διαφόρων αλγορίθμων που εφαρμόστηκαν πάνω σε δεδομένα που αφορούσαν τουρκικό διαδικτυακό ταξιδιωτικό πρακτορείο, και στόχο είχε να επιτύχει την πρόβλεψη μελλοντικών πωλήσεων του πρακτορείου μέσω του Click-Through Rate (CTR) των διαδικτυακών διαφημίσεών του, που είναι το ποσοστό κατά το οποίο χρήστες αλληλεπιδρούν με διαφημίσεις του πρακτορείου σε άλλες ιστοσελίδες. Η έρευνα συνεπώς συνδυάζει τόσο την πρόβλεψη των πωλήσεων του πρακτορείου, όσο και την πρόβλεψη της αποτελεσματικότητας των διαφημίσεων.

Τα μοντέλα που χρησιμοποιήθηκαν για την έρευνα ήταν τα Random Forest, Gradient Boosting, AdaBoost, Support Vector Regression & XGBoost. Όπως αναφέρθηκε η έρευνα περιέχει τόσο πρόβλεψη πωλήσεων όσο και CTR. Παρακάτω παρατίθενται τα συγκριτικά αποτελέσματα των μοντέλων για την πρόβλεψη πωλήσεων συγκεκριμένα. Οι συγγραφείς σχολιάζουν ότι τα πιο αξιόπιστα αποτελέσματα πετυχαίνονται σε κάθε περίπτωση με το μοντέλο XGBoost.

Comparison of algorithms for sales prediction

| Model | RSQ | MSE | MAE |
|----------------------------|-------|---------|-------|
| eXtreme gradient boosting | 0.574 | 2267.02 | 9.45 |
| Gradient boosting machines | 0.564 | 2159.98 | 9.54 |
| Random forest | 0.572 | 2121.13 | 9.40 |
| Generalized linear models | 0.443 | 2760.80 | 16.18 |
| Deep neural network | 0.507 | 2441.98 | 15.15 |

Πίνακας 4.7: Σύγκριση δεικτών αποτελεσματικότητας και σφάλματος για τα διάφορα μοντέλα [Tekin, A.T., Cebi, F., 2020]

4.2.20 Sale Forecast for Basic Commodities Based on Artificial Neural Networks Prediction by Silva, J., Villa, J.V., Cabrera, D. (2020)

Σε αυτό το σχετικά σύντομο άρθρο οι συγγραφείς αναφέρουν ότι δοκίμασαν να κάνουν πρόβλεψη πωλήσεων για τις πωλήσεις τροφίμων και άλλων βασικών αγαθών με τη χρήση μιας ευρείας ποικιλίας μοντέλων, συμπεριλαμβανομένων και της Γραμμικής Παλινδρόμησης, ARIMA και Support Vector Regression. Για την έρευνά τους χρησιμοποίησαν δεδομένα που προμήθευσε το National Administrative Department of Statistics of Colombia (DANE - Departamento Administrativo Nacional de Estadísticas), και περιείχε στοιχεία από 1257 δημοφιλή καταστήματα στην Κολομβία κατά τα έτη 2015 έως 2018.

Λόγω του μικρού μεγέθους του άρθρου δε δίνονται ιδιαίτερες πληροφορίες σχετικά με την έρευνα, αλλά παρατίθενται τα αποτελέσματα της ακρίβειας τις κάθε μεθόδου στον παρακάτω πίνακα.

Table 2. Comparison of general results (percentages)

| Methods | Cluster 1 | Cluster 2 | Cluster 3 | Cluster 4 | Cluster 5 | Cluster 6 | Cluster 7 | Cluster 8 | MAPE |
|---------------------------------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-----------|-------------|
| Simple moving average (3) | 10.85 | 11.84 | 12.54 | 9.54 | 12.5 | 14.8 | 13.5 | 11.14 | 12.58 |
| Weighted moving average | 42.5 | 43.52 | 44.25 | 44.58 | 42.58 | 45.5 | 44.5 | 44.69 | 44.12 |
| Smoothing simple exponential | 58.12 | 59.54 | 57.25 | 56.25 | 58.14 | 59.4 | 57.32 | 58.14 | 57.94 |
| Linear regression | 25.2 | 24.5 | 24.6 | 23.9 | 24.87 | 25.24 | 25.9 | 24.9 | 24.68 |
| Seasonal or cyclical variation | 52.2 | 54.2 | 56.2 | 54.8 | 54.9 | 51.2 | 57.2 | 54.2 | 54.84 |
| Box-Jenkins - ARIMA (1,0,0) | 14.5 | 14.2 | 14.2 | 16.5 | 14.5 | 14.98 | 14.5 | 12.98 | 14.91 |
| ANN | 9.58 | 8.25 | 8.64 | 8.34 | 8.14 | 8.95 | 8.26 | 8.24 | 8.64 |
| (SVR) Support Vector Regression | 12.5 | 12.4 | 12.6 | 12.8 | 12.7 | 12.5 | 12.9 | 12.4 | 12.35 |

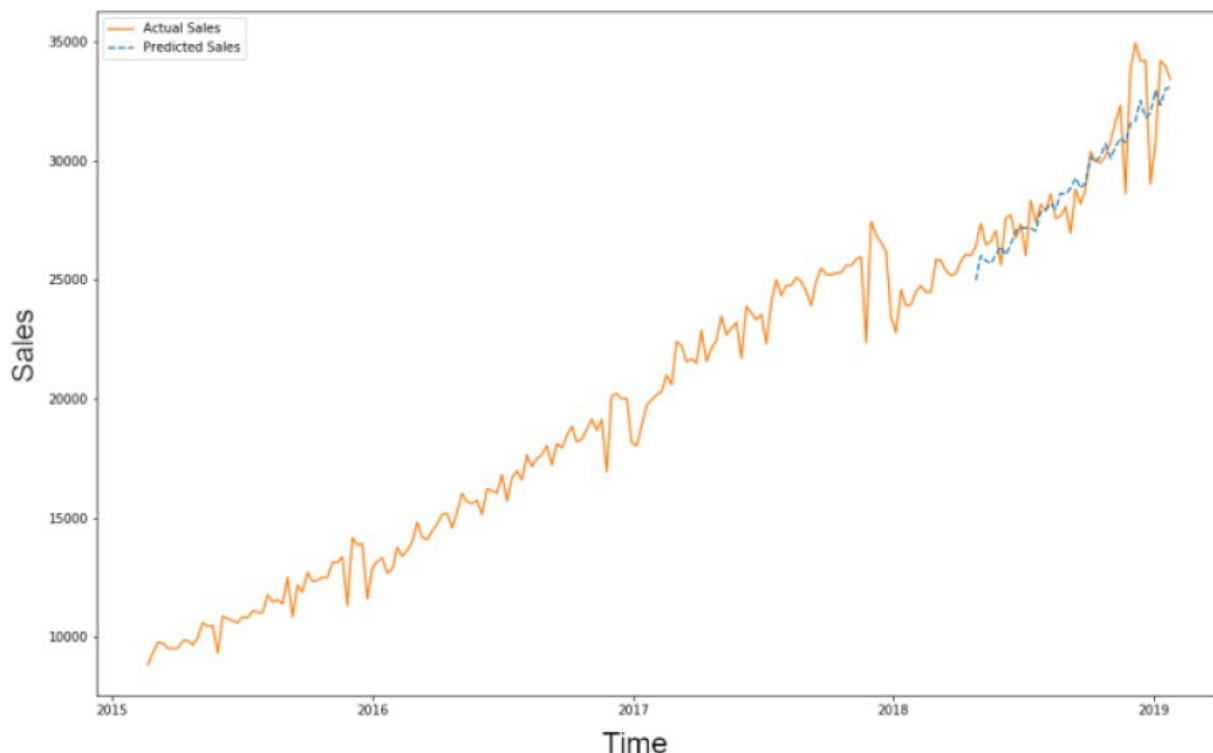
Πίνακας 4.8: Σύγκριση αποτελεσμάτων των μοντέλων (ποσοστιαία) [Silva, J., Villa, J.V., Cabrera, D., 2020]

Φαίνεται λοιπόν ότι το Artificial Neural Network (ANN) παρουσίασε το χαμηλότερο σφάλμα στο σύνολο των μεθόδων. Οι συγγραφείς σχολιάζουν το γεγονός ότι στα ANN δε χρειάζεται να προσδιορισθούν υποθέσεις, σε αντίθεση με τις υπόλοιπες συμβατικές μεθόδους, καθώς η μέθοδος μαθαίνει της συσχετίσεις από τα δεδομένα που της προμηθεύονται.

4.2.21 A Convolutional Neural Networks-based Model for Sales Prediction by Buyar, V., Abdel-Raouf, A. (2019)

Σε έρευνα που ζητήθηκε από φαρμακευτική εταιρεία που αποσκοπούσε να επεκταθεί, οι συγγραφείς του παρόντος άρθρου εφάρμοσαν μια υποκατηγορία Neural Networks που ονομάζεται Convolutional Neural Networks, τα οποία συνήθως χρησιμοποιούνται στην ανάλυση εικόνων. Η εταιρεία που προμήθευσε τα δεδομένα αφήνεται ανώνυμη και αναφέρεται μόνο ως Y LLC, ενώ τα δεδομένα καλύπτουν τη χρονική περίοδο από τον Φεβρουάριο του 2015 μέχρι τον Ιανουάριο του 2019.

Από το παραπάνω διάστημα, οι 40 τελευταίες εβδομάδες παραμερίστηκαν για επαλήθευση, ενώ οι υπόλοιπες χρησιμοποιήθηκαν για την εκπαίδευση του μοντέλου (διαμοιρασμός περίπου 80%/20%). Οι μεταβλητές πρόβλεψης που χρησιμοποιήθηκαν στο μοντέλο δεν εξηγούνται πλήρως. Τα τελικά αποτελέσματα έδειξαν ότι η πρόβλεψη του μοντέλου είχε Mean Absolute Percent Error (MAPE) 3,54%, ενώ σε μια επιπλέον δοκιμή όπου χρησιμοποιήθηκαν μόνο οι 120 πρώτες βδομάδες αντί για 166, το MAPE αυξήθηκε στο 3,86%, το οποίο οι συγγραφείς χαρακτήρισαν ως μη σημαντική αλλοίωση. Η πρόβλεψη του μοντέλου σε σύγκριση με τα πραγματικά στοιχεία των 40 τελευταίων βδομάδων φαίνεται και στο παρακάτω διάγραμμα.



Forty week sales prediction, using CNN, trained on 166 weeks.

Εικόνα 4.11: Αποτελέσματα πρόβλεψης του Convolutional Neural Network για 40 βδομάδες [Buyar, V., Abdel-Raouf, A., 2019]

Οι συγγραφείς προσθέτουν επίσης ότι η μέθοδός τους αποδείχθηκε πιο αποτελεσματικά από τη χρήση ARIMA, η οποία είχε προταθεί σε άλλη έρευνα.

4.2.22 Multi-task based Sales Predictions for Online Promotions

by Shen, X., Yao, C., Ester, M., Li, Z., Bu, J., Zhou, X., Ye, Y., Wang, C. (2019)

Το παρόν άρθρο αποσκοπεί στην επίλυση ενός προβλήματος που ονοματίζεται Sales Predictions for Online Promotions (SPOP), το οποίο όπως φαίνεται κι από το όνομα αφορά την πρόβλεψη πωλήσεων κατά τη διάρκεια περιόδων προσφορών, όπως η Black Friday και Cyber Monday, που οργανώνει το διαδικτυακό κατάστημα Alibaba. Τα δεδομένα που χρησιμοποιήθηκαν προέρχονται από το Alibaba 11.11 Global Shopping Festival, που λαμβάνει χώρα στις 11 Νοεμβρίου κάθε χρόνο από το 2009, και εξελίχθηκε από ημερήσια εκδήλωση στο να διαρκεί μια περίοδο 23 ημερών, από τις οποίες οι 22 είναι περίοδος προπωλήσεων.

Οι συγγραφείς προτείνουν μια μέθοδο βασισμένη στα Neural Networks, την οποία ονομάζουν Deep Item Network for Online Promotions (DINOP). Η μέθοδος αυτή χρησιμοποιεί και στατικές και δυναμικές μεταβλητές πρόβλεψης, ενώ λαμβάνει επίσης υπόψη και τα αθροιστικά δεδομένα που διατίθενται σχετικά με τα προφίλ των χρηστών του Alibaba. Το δυναμικό κομμάτι επιτυγχάνεται με την ενσωμάτωση μιας επιπλέον μεθόδου την οποία οι συγγραφείς ονομάζουν Target Users Controlled Gated Recurrent Unit (TUC-GRU).

Οι μεταβλητές που χρησιμοποιήθηκαν διακρίνονται στις εξής κατηγορίες:

- Static Item Inherent profiles (SII), όπως κατηγορία, μάρκα κτλ.
- Static Item Sales data (SIS), συνήθεις πωλήσεις των αντικειμένων υπό κανονικές συνθήκες.
- Static target Users profiles (SU), στατικά στοιχεία των πελατών, όπως ηλικία, φύλο, ωροσκόπιο
- Dynamic Item Sales (DISt), οι ημερήσιες “προπωλήσεις” ενός αντικειμένου κατά την ημέρα t των προπωλήσεων.
- Dynamic target Users (Dut), η ημερήσια κατανομή των στοιχείων των χρηστών που αγοράζουν ένα αντικείμενο κατά την ημέρα t των προπωλήσεων.

Η μέθοδος συγκρίθηκε με συμβατικά μοντέλα όπως Deep Neural Networks, LSTM, Gradient Boosting Decision Tree (GBDT) και XGBoost, και βρέθηκε να τις ξεπερνάει σε όλα τα ζητούμενα. Οι συγγραφείς σχολιάζουν επίσης ότι η μέθοδος τους μπορεί να χρησιμοποιηθεί και για την επίλυση περαιτέρω προβλημάτων, όπως η πρόβλεψη Click-Through Rate και ο εντοπισμός απατηλών συναλλαγών.

4.2.23 Grid Search Optimization (GSO) Based Future Sales Prediction For Big Mart

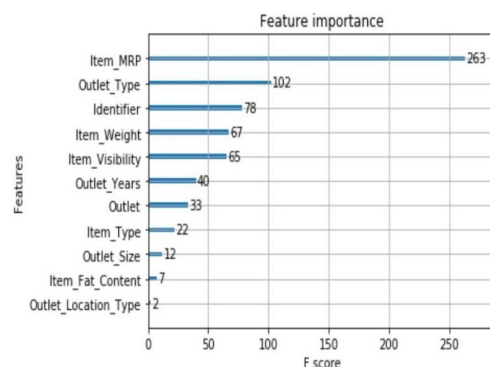
by Behera, G., Nain, N. (2019)

Ο στόχος της έρευνας στο συγκεκριμένο άρθρο είναι η πρόβλεψη των μελλοντικών πωλήσεων κάθε ξεχωριστού καταστήματος της αλυσίδας Big Mart. Για την επίτευξη αυτού του στόχου οι συγγραφείς προτείνουν τη μέθοδο Grid Search Optimization (GSO) σε συνδυασμό με το μοντέλο XGBoost. Η μέθοδος GSO σε αυτή την περίπτωση χρησιμοποιείται για την επιλογή των βέλτιστων παραμέτρων για την εφαρμογή του μοντέλου XGBoost.

Οι παρακάτω εικόνες δείχνουν τα περιεχόμενα των αρχικών σετ δεδομένων, τη σημαντικότητα που προέκυψε για το καθένα, καθώς και την ακρίβεια του μοντέλου πριν και μετά την εφαρμογή της μεθόδου GSO.

FEATURES OF DATASET

| Name Of the Attributes | Types | Total Count |
|---------------------------|---------|-------------|
| Item_Identifier | object | 8523 |
| Item_Weight | float64 | 7060 |
| Item_Fat_Content | object | 8523 |
| Item_Visibility | float64 | 8523 |
| Item_Type | object | 8523 |
| Item_MRP | float64 | 8523 |
| Outlet_Identifier | object | 8523 |
| Outlet_Establishment_Year | int64 | 8523 |
| Outlet_Size | object | 6113 |
| Outlet_Location_Type | object | 8523 |
| Outlet_Type | object | 8523 |
| Item_Outlet_Sales | float64 | 8523 |



COMPARISON OF XGBOOST MODEL WITH AND WITHOUT PARAMETER TUNING FOR TRAINING SET

| Models | RMSE | MAE |
|-----------------------------------|------|--------|
| Xgboost(Before tuning parameter) | 1066 | 749.78 |
| Xgboost(After tuning parameter) | 1052 | 739.03 |

COMPARISON OF XGBOOST MODEL WITH AND WITHOUT PARAMETER TUNING FOR TEST SET

| Models | RMSE | MAE |
|-----------------------------------|-------|--------|
| Xgboost(Before tuning parameter) | 180.2 | 134.08 |
| Xgboost(After tuning parameter) | 178.7 | 129.90 |

Πίνακας 4.9: Σημαντικότητα των μεταβλητών και σύγκριση δεικτών σφάλματος πριν και μετά την εφαρμογή της GSO [Behera, G., Nain, N., 2019]

Από τους πίνακες προκύπτει ότι η ακρίβεια του μοντέλου όντως βελτιώνεται μετά την εφαρμογή της μεθόδου GSO.

4.2.24 Online Retail Sales Prediction with Integrated Framework of K-mean and Neural Network by Lingxian, Y., Jiaqing, K., Shihuai, W. (2019)

Σε αυτό το άρθρο οι συγγραφείς προτείνουν το συνδυασμό ενός συνόλου μεθόδων για την πρόβλεψη διαδικτυακών πωλήσεων. Οι εν λόγω μέθοδοι είναι το k-mean Clustering και το Long Short Term Memory μοντέλο των Recurrent Neural Network (RNN). Για το σετ δεδομένων που χρησιμοποιήθηκε για την έρευνα, αναφέρεται αόριστα ότι αφορά διαδικτυακή δραστηριότητα στο Ηνωμένο Βασίλειο από τις 01/12/2010 έως τις 09/12/2011 και ότι περιέχει τις εξής 8 μεταβλητές:

- Invoice No.
- Stock code
- Description
- Quantity
- Date
- Unit price
- Customer ID
- Country

Η μέθοδος k-mean Clustering χρησιμοποιείται για την αρχική επεξεργασία των δεδομένων πριν την εισαγωγή τους στο μοντέλο LSTM. Η δοκιμή του μοντέλου σε διάφορα χρονικά διαστήματα έδειξε ότι τα πιο ακριβή αποτελέσματα προκύπτουν όταν τα δεδομένα των 10 προηγούμενων ωρών χρησιμοποιούνται για την πρόβλεψη των 20 επομένων.

4.2.25 A Two-Level Statistical Model for Big Mart Sales Prediction by Punam, K., Pamula, R., Jain, P.K. (2019)

Σε αυτήν την έρευνα οι συγγραφείς προτείνουν τη χρήση μοντέλου 2 επιπέδων για την πρόβλεψη πωλήσεων. Το σετ δεδομένων που χρησιμοποιήσαν προέρχεται από τις πωλήσεις της αλυσίδας καταστημάτων Big Mart για το έτος 2013 και περιέχει 12 μεταβλητές:

- Item Identifier
- Item Fat
- Item Visibility
- Item Type
- Outlet Type
- Item MRP
- Outlet Identifier
- Item Weight
- Outlet Size
- Outlet Establishment Year
- Outlet Location Type
- Item Outlet Sales

Ο συνδυασμός δυο μοντέλων “το ένα πάνω στο άλλο” για την κατασκευή του μοντέλου δυο επιπέδων αποκαλείται Stacking, και σε αυτή τη μέθοδο τα εξαγόμενα του μοντέλου του πρώτου επιπέδου χρησιμοποιούνται σαν τα δεδομένα για το μοντέλο του δεύτερου επιπέδου. Στη συγκεκριμένη έρευνα οι συγγραφείς χρησιμοποίησαν συνδυασμούς των μεθόδων γραμμικής παλινδρόμησης, Support Vector Regression, και Cubist για την κατασκευή του μοντέλου δυο επιπέδων, και μετά σύγκριναν τα αποτελέσματα με συνήθη μοντέλα ενός επιπέδου, όπως γραμμική παλινδρόμηση, Regression Tree, Cubist, Support Vector Regression, και k-nearest Neighbour.

Η αξιολόγηση των μοντέλων γίνεται με τη χρήση του Mean Absolute Error, τα αποτελέσματα της οποίας παρουσιάζονται στον παρακάτω πίνακα.

CROSS-VALIDATION MAE AND TEST MAE OF
MACHINE LEARNING MODELS

| Machine Learning Models | Cross-Validation MAE | Test MAE |
|--------------------------------|-----------------------------|-----------------|
| k-NN | 0.4573 | 0.4699 |
| Linear Regression | 0.4191 | 0.4142 |
| Regression Tree | 0.4184 | 0.413 |
| Support Vector Regression | 0.4144 | 0.4102 |
| Cubist | 0.4016 | 0.4000 |
| Two-level Statistical Model | 0.3975 | 0.3917 |

Πίνακας 4.10: Σύγκριση δεικτών σφάλματος για τα διάφορα μοντέλα [Punam, K., Pamula, R., Jain, P.K., 2019]

Φαίνεται λοιπόν ότι το συνδυαστικό μοντέλο δύο επιπέδων παρουσιάζει βελτίωση σε σχέση με τα συνήθη μοντέλα ενός επιπέδου.

4.2.26 Intelligent Sales Prediction Using Machine Learning Techniques by Cheriyan, S., Ibrahim, S., Mohanan, S., Treesa, S. (2019)

Στο παρόν άρθρο οι συγγραφείς δοκιμάζουν μεθόδους μηχανικής μάθησης για την πρόβλεψη πωλήσεων στον τομέα της μόδας, και πιο συγκεκριμένα διαδικτυακών πωλήσεων. Για την έρευνά τους χρησιμοποιούν σαν δεδομένα στοιχεία πωλήσεων ηλεκτρονικής πλατφόρμας μόδας για τα έτη 2015 έως 2017. Το σετ δεδομένων συμπεριλάμβανε τις παρακάτω μεταβλητές:

- Category,
- City,
- Type of items and description,
- number of items,
- Quantity,
- Quarter,
- Sales Revenue,
- Year,
- SKU description,
- Week,
- Year

Επέλεξαν να δοκιμάσουν 3 διαφορετικούς μεθόδους μηχανικής μάθησης, ένα Γενικευμένο Γραμμικό Μοντέλο (GLM), Decision Tree & Gradient Boosted Trees. Τα αποτελέσματα έδειξαν ότι η πιο αποτελεσματική μέθοδος στο συγκεκριμένο πρόβλημα ήταν η Gradient Boosted Trees, με το υψηλότερο ποσοστό ακρίβειας στο 98%, ακολουθούμενη από την Decision Trees με ποσοστό 71%, και το GLM με ποσοστό 64%.

4.2.27 TADA: Trend Alignment with Dual-Attention Multi-Task Recurrent Neural Networks for Sales Prediction by Chen, T., Yin, H., Chen, H., Wu, L., Wang, H., Zhou, X., Li, X. (2018)

Οι συγγραφείς του παρόντος άρθρου ανέπτυξαν ένα μοντέλο πρόβλεψης πωλήσεων, τόσο σε παντοπωλεία όσο και στον τομέα συσκευών ενέργειας, το οποίο ονόμασαν Trend Alignment with Dual-Attention, ή για συντομία TADA. Το μοντέλο είναι βασισμένο σε Recurrent Neural Network (RNN), και πιο συγκεκριμένα στην ανάλυση Long Short-Term Memory (LSTM).

Για την εφαρμογή του μοντέλου χρησιμοποιούνται δυο ξεχωριστές πηγές σετ δεδομένων. Το πρώτο προέρχεται από το Kaggle και ονομάζεται Favorita, και περιέχει μεταβλητές και όγκο πωλήσεων όλων των προϊόντων από 56 καταστήματα βασισμένα στο Εκουαδόρ. Το αρχικό σετ δεδομένων καλύπτει όλο το διάστημα από την 1 Ιανουαρίου 2013 ως τις 15 Αυγούστου 2017, αλλά λαμβάνεται υπόψη μόνο διάστημα ενός έτους από 15 Αυγούστου 2016 έως 15 Αυγούστου 2017, ώστε να αποφευχθεί η ασυνήθης επιρροή που είχε ένας έντονος σεισμός νωρίτερα το έτος 2016. Το δεύτερο προέρχεται από το One Stop Warehouse (OSW), έναν από τους μεγαλύτερους προμηθευτές συσκευών ηλιακής ενέργειας στην Αυστραλία, και καλύπτει το διάστημα από τις 22 Φεβρουαρίου 2016 έως 4 Μαρτίου 2017.

Το μοντέλο συγκρίνεται με μια πληθώρα προϋπάρχοντων μοντέλων πρόβλεψης, και πιο συγκεκριμένα με τα μοντέλα Random Forest (RF), XGBoost, Stacked Auto Encoder LSTM, Attention RNN, DA-RNN & LSTNet. Συγκρίνεται επίσης και με υπομοντέλα του εαυτού του, σε καθένα από τα οποία λείπει ένα από τα υποτιμήματα που το συνιστούν, για να διερευνηθεί η συνεισφορά του κάθε υποτιμήματος στο συνολικό μοντέλο. Στον παρακάτω πίνακα φαίνεται η σύγκριση των αποτελεσμάτων όλων των παραπάνω μοντέλων.

SALES PREDICTION RESULTS. NUMBERS IN BOLD FACE ARE THE BEST RESULTS WITHIN EACH COLUMN.

| Method | Favorita | | | | | | OSW | | | | | |
|----------------------|--------------|---------------|--------------|---------------|--------------|---------------|---------------|---------------|---------------|---------------|---------------|---------------|
| | $\Delta = 2$ | | $\Delta = 4$ | | $\Delta = 8$ | | $\Delta = 2$ | | $\Delta = 4$ | | $\Delta = 8$ | |
| | MAE | SMAPE(%) | MAE | SMAPE(%) | MAE | SMAPE(%) | MAE | SMAPE(%) | MAE | SMAPE(%) | MAE | SMAPE(%) |
| RF | 32.483 | 200(max) | 35.507 | 200(max) | 41.329 | 200(max) | 29.147 | 89.482 | 35.576 | 137.892 | 43.096 | 200(max) |
| XGBoost [32] | 16.705 | 87.433 | 19.833 | 91.230 | 22.547 | 158.461 | 21.496 | 49.556 | 24.916 | 53.243 | 30.322 | 82.633 |
| SAE-LSTM [33] | 7.364 | 39.447 | 8.033 | 44.384 | 8.116 | 46.932 | 17.828 | 44.241 | 19.805 | 46.887 | 20.823 | 49.873 |
| A-RNN [21] | 11.610 | 60.781 | 12.226 | 62.397 | 13.005 | 65.812 | 17.391 | 44.635 | 18.823 | 44.603 | 22.129 | 49.180 |
| DA-RNN [4] | 7.816 | 43.859 | 8.234 | 44.704 | 8.566 | 46.281 | 17.634 | 44.215 | 19.578 | 47.139 | 20.693 | 48.365 |
| LSTNet [9] | 7.419 | 43.523 | 7.982 | 45.662 | 8.729 | 48.469 | 16.625 | 42.317 | 18.989 | 45.782 | 21.246 | 49.191 |
| TADA-SE | 9.995 | 58.715 | 11.076 | 60.332 | 10.955 | 60.257 | 19.635 | 53.017 | 20.884 | 49.370 | 21.687 | 51.685 |
| TADA-SA ₁ | 8.152 | 46.732 | 8.273 | 43.951 | 8.968 | 49.079 | 16.585 | 42.620 | 18.624 | 44.331 | 21.699 | 51.195 |
| TADA-SA ₂ | 7.635 | 42.883 | 8.247 | 44.942 | 8.626 | 48.609 | 17.087 | 42.199 | 18.643 | 45.219 | 21.190 | 49.825 |
| TADA | 6.955 | 38.770 | 7.323 | 40.588 | 7.422 | 43.675 | 15.418 | 41.354 | 17.572 | 43.265 | 19.618 | 47.782 |

Πίνακας 4.11: Σύγκριση δεικτών σφάλματος για τα διάφορα μοντέλα [Chen, T., Yin, H., Chen, H., Wu, L., Wang, H., Zhou, X., Li, X., 2018]

Για τη σύγκριση των αποτελεσμάτων χρησιμοποιήθηκαν σαν κριτήρια τα Mean Absolute Error (MAE) και Symmetric Mean Absolute Percentage Error (SMAPE). Σε κάθε στήλη το μοντέλο με το μικρότερο σφάλμα είναι έντονα χρωματισμένα, το οποίο για κάθε περίπτωση και συνδυασμό παραμέτρων ήταν το TADA. Οι συγγραφείς καταλήγουν συνεπώς στο συμπέρασμα ότι τα

αποτελέσματα αποδεικνύουν την υπεροχή του μοντέλου τους σε σύγκριση με τα προϋπάρχοντα μοντέλα πρόβλεψης. Φαίνεται επίσης και ότι καθένα από τα υποτιμήματά του συνεισφέρει όντως θετικά στο τελικό αποτέλεσμα, καθώς το σφάλμα στο τελικό μοντέλο ήταν μικρότερο σε σύγκριση και με τα υποτιμήματα του εαυτού του.

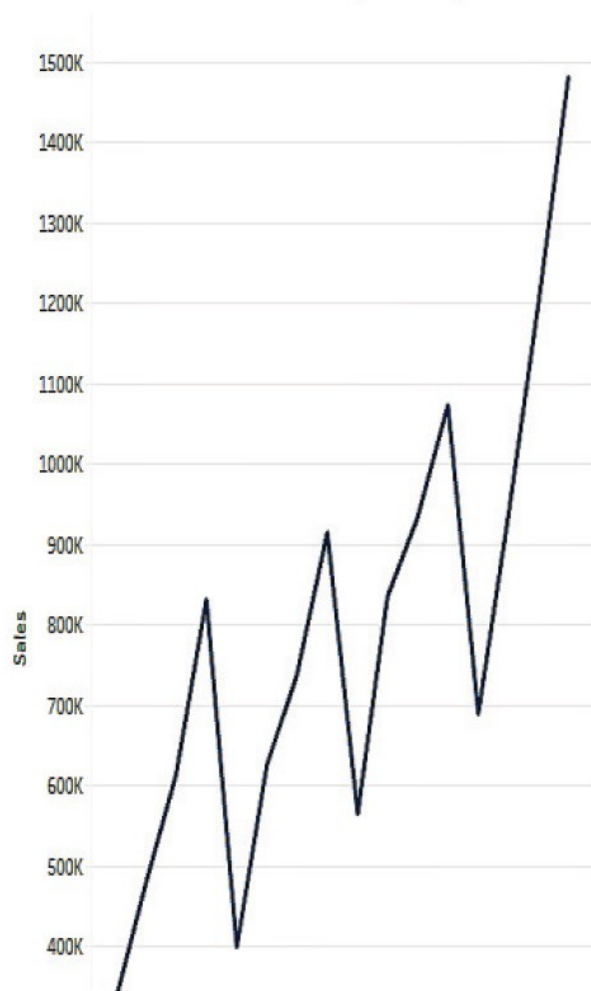
Τέλος, επιδεικνύουν ότι το μοντέλο μπορεί να χρησιμοποιηθεί για μεγάλα σετ δεδομένων, καθώς ο χρόνος εφαρμογής αυξάνεται σχεδόν γραμμικά με τον όγκο των δεδομένων.

4.2.28 Prediction of Sales Value in online shopping using Linear Regression by Gopalakrishnan, T., Choudhary, R., Prasad, S. (2018)

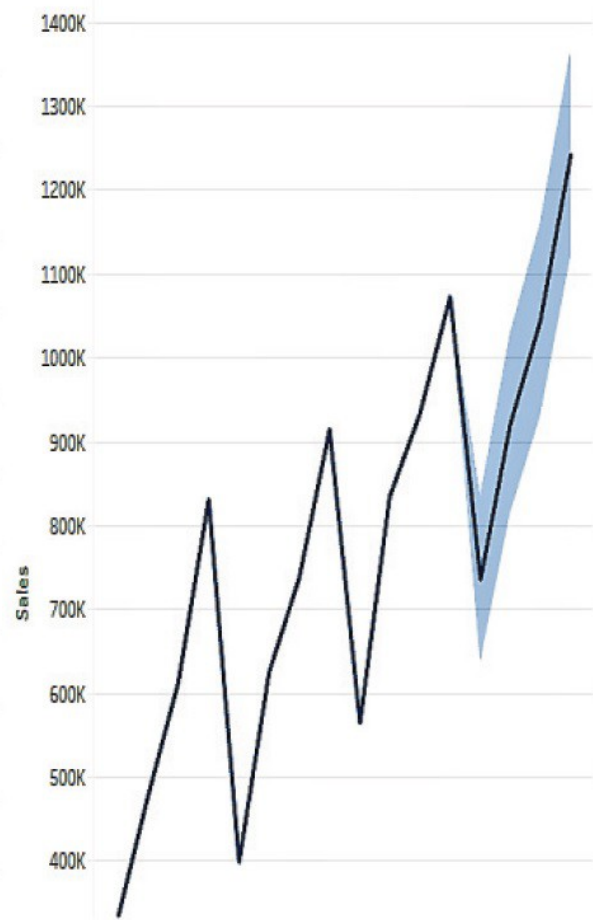
Στο συγκεκριμένο άρθρο οι συγγραφείς δοκιμάζουν να εφαρμόσουν μηχανική μάθηση για την πρόβλεψη πωλήσεων μεγάλου διαδικτυακού καταστήματος με τη μέθοδο της απλής γραμμικής παλινδρόμησης. Το εν λόγω κατάστημα από το οποίο προήλθαν τα δεδομένα είναι το ZMart, και τα δεδομένα αφορούν την περίοδο 2011 έως 2013 για την εκπαίδευση του μοντέλου, καθώς και το έτος 2014 για την δοκιμή της αξιοπιστίας του. Το σετ δεδομένων περιείχε 25 στήλες στοιχείων, χωρίς να διευκρινίζονται περαιτέρω λεπτομέρειες για τα περιεχόμενά τους.

Με την εφαρμογή του μοντέλου της έρευνας προέκυψε το παρακάτω διάγραμμα. Το αριστερό τμήμα δείχνει τους πραγματικούς αριθμούς πωλήσεων κατά τα έτη 2011-2014, ενώ το δεξιό τμήμα δίνει ένα έυρος πρόβλεψης πωλήσεων για το έτος 2014 και μια γραμμή που αντιπροσωπεύει το μέσο όρο.

Actual Quarters vs Sales Data (2011-14)



Forecasted (2014) Quarters vs Sales Data (2011-13)



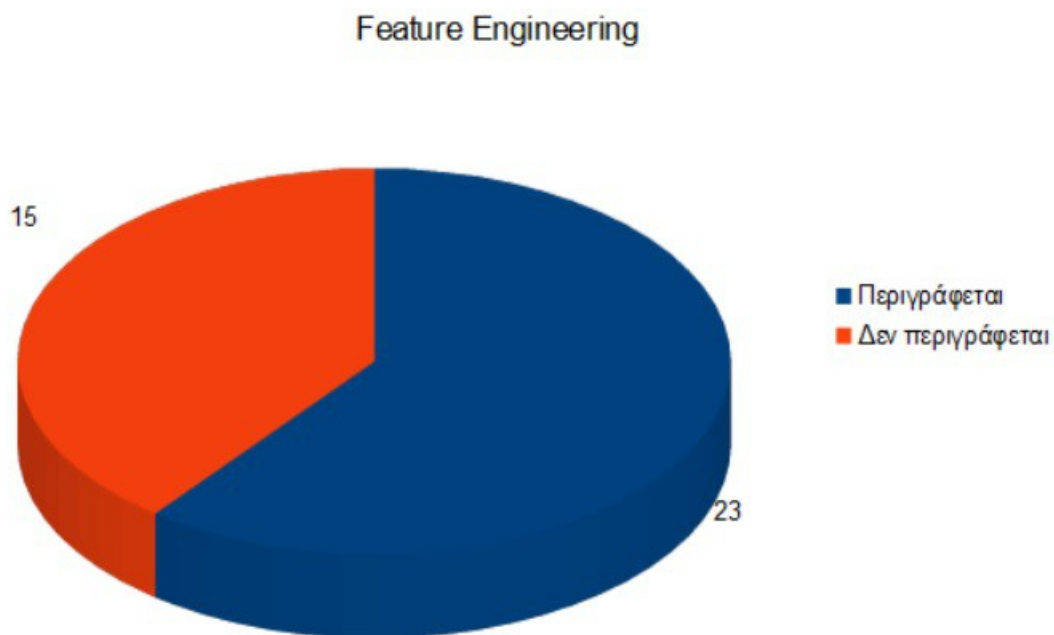
Dashboard containing comparison of actual vs predicted quarters vs sales data for the years 2011-14

Εικόνα 4.12: Σύγκριση προβλεπόμενου και πραγματικού αριθμού πωλήσεων για τα έτη 2011-2014 [Gopalakrishnan, T., Choudhary, R., Prasad, S., 2018]

Ενώ η πρόβλεψη φαίνεται να έχει παρόμοια τάση με την πραγματικότητα, το έτος 2014 παρουσίασε πιο απότομη αύξηση των πωλήσεων σε σύγκριση με τα προηγούμενα έτη. Παρόλα αυτά οι συγγραφείς σχολιάζουν ότι το αποτέλεσμα ήταν ικανοποιητικό, υπολογίζοντας ένα ποσοστό σφάλματος της πρόβλεψης περίπου 16%.

Κεφάλαιο 5: Εννοιο-κεντρική Ανάλυση

Στο παρόν κεφάλαιο παρατίθενται συμπεράσματα τα οποία προκύπτουν μελετώντας τις έρευνες που συμπεριλήφθηκαν ως ένα σύνολο. Αρχικά παρατηρείται ότι δεν ακολουθούν όλες οι έρευνες κάποια τυποποιημένη μέθοδο προσέγγισης του προβλήματος της πρόβλεψης πωλήσεων με τη χρήση μηχανικής μάθησης, ούτε και παρουσίασης της διαδικασίας της έρευνας και των αποτελεσμάτων της. Για παράδειγμα, σε 15 έρευνες δεν περιγράφεται η διαδικασία επεξεργασίας και καθαρισμού των ακατέργαστων δεδομένων που ακολουθήθηκε, εάν ακολουθήθηκε κάποια, και σε ορισμένες δεν αναφέρεται το μέγεθος του διαστήματος για το οποίο γίνεται πρόβλεψη πωλήσεων (ωριαίες, ημερήσιες, εβδομαδιαίες, μηνιαίες ή τριμηνιαίες πωλήσεις).



Εικόνα 5.1: Περιγραφή της διαδικασίας επεξεργασίας δεδομένων στις έρευνες

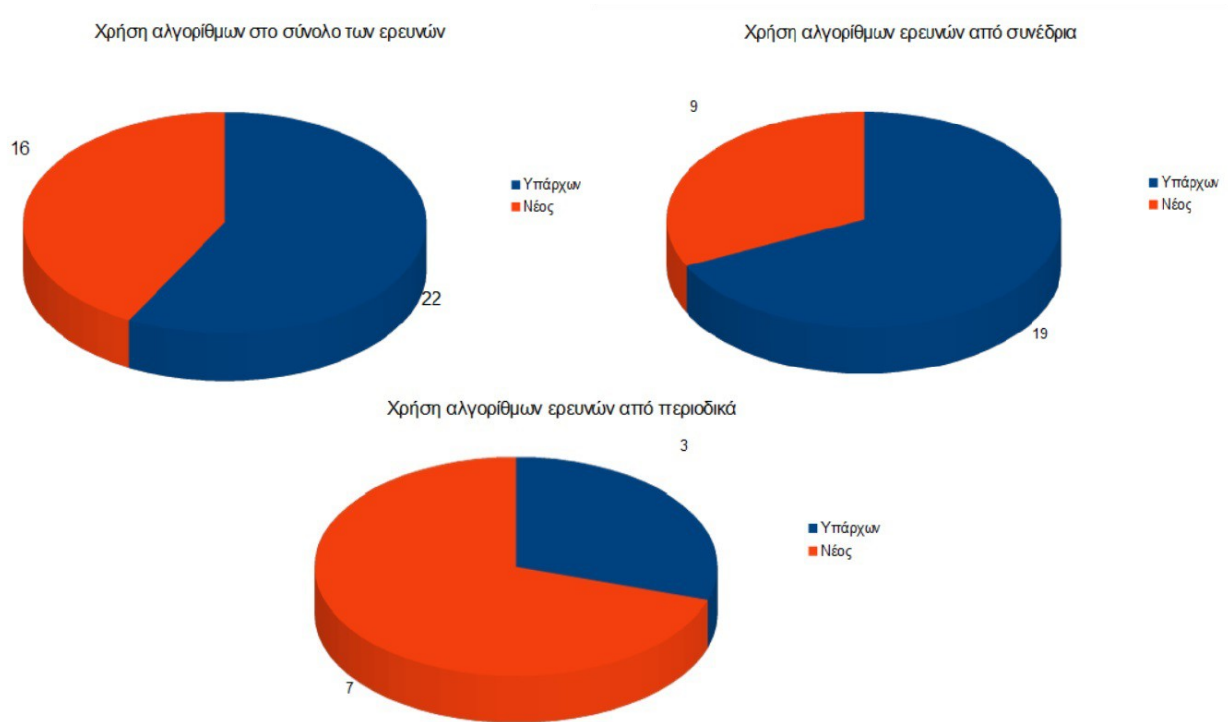
Η αρχική βασική παρατήρηση που μπορεί να γίνει είναι το ποιοι αλγόριθμοι μηχανικής μάθησης χρησιμοποιήθηκαν σε πόσα και ποια άρθρα, το οποίο και παρατίθεται στον πίνακα και το διάγραμμα της εικόνας 5.5 στο τέλος του κεφαλαίου. Ο αριθμός που αντιστοιχεί σε κάθε άρθρο φαίνεται στον πίνακα 6.1 του επόμενου κεφαλαίου. Στον ίδιο πίνακα δίνεται και ο σύνδεσμος DOI του κάθε άρθρου, όπου αυτός υφίσταται, για εύκολη και άμεση πρόσβαση σε αυτά.

Στα πλαίσια αυτής της ανάλυσης, οι περιπτώσεις στις οποίες οι συγγραφείς ενός άρθρου ανέπτυξαν δικό τους αλγόριθμο αντί να αρκεστούν σε κάποιον προϋπάρχοντα ομαδοποιούνται στην κατηγορία Custom. Σε περίπτωση που αυτός ο αλγόριθμος χρησιμοποιεί σαν βάση κάποιον προϋπάρχοντα, σημειώνεται και αυτό αντίστοιχα. Επίσης γίνεται διαχωρισμός στο εάν ο κάθε αλγόριθμος εφαρμόστηκε ως αναπόσπαστο κομμάτι της έρευνας, στην οποία περίπτωση σημειώνεται με X, ή εάν εφαρμόστηκε απλώς σαν μέτρο σύγκρισης του κυρίου αλγορίθμου, στην οποία περίπτωση σημειώνεται με O.

Παρατηρήθηκαν συνολικά 23 προϋπάρχοντες αλγόριθμοι που είτε χρησιμοποιήθηκαν, είτε βασίστηκε πάνω της η ανάπτυξη νέου αλγορίθμου. Αυτοί ήταν οι:

1. Gradient Boosting,
2. eXtreme Gradient Boosting (XGBOOST),
3. AdaBoost,
4. Long Short Term Memory (LSTM),
5. Generative Adversarial Networks (GAN),
6. Decision Tree,
7. Random Tree,
8. Random Forest,
9. Least Squares Support Vector Machine (LSSVM),
10. Auto Regressive Integrated Moving Average (ARIMA),
11. Artificial Neural Network (ANN),
12. Convolutional Neural Network (CNN),
13. Fast Forward Neural Network (FFNN),
14. Microsoft Azure Machine Learning,
15. Q Learning,
16. Cubist,
17. Linear Regression,
18. Multiple Linear Regression,
19. Ridge Regression,
20. Lasso Regression,
21. Support Vector Regression (SVR),
22. Deep Learning και
23. Holt-Winters

Από τον πίνακα προκύπτει ότι 16 στις 38, δηλαδή λιγότερες από τις μισές περιπτώσεις συνολικά επέλεξαν να αναπτύξουν δικό τους αλγόριθμο, όμως κοιτάζοντας συγκεκριμένα τα άρθρα που δημοσιεύθηκαν σε επιστημονικά περιοδικά, αυτές οι περιπτώσεις αποτελούν την πλειοψηφία, όπως απεικονίζεται και στα διαγράμματα της εικόνας 5.2.

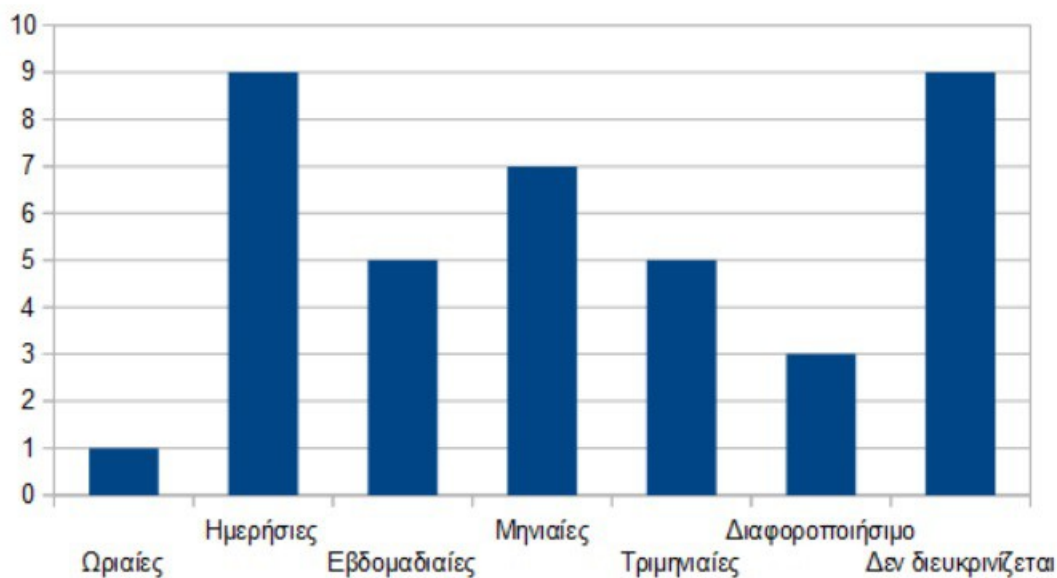


Εικόνα 5.2: Συχνότητα ανάπτυξης νέων αλγορίθμων στις έρευνες

Επίσης προκύπτει ότι οι πιο συχνά χρησιμοποιούμενες μέθοδοι ήταν οι XGBOOST και Γραμμική Παλινδρόμηση, οι οποίες χρησιμοποιήθηκαν σε σχεδόν ένα τρίτο των περιπτώσεων η καθεμία, όταν συμπεριλαμβάνουμε και τις περιπτώσεις αλγορίθμων που βασίστηκαν πάνω τους. Πιο συγκεκριμένα ο XGBOOST χρησιμοποιήθηκε 10 φορές και σε 1 έρευνα συμπεριλήφθηκε ως μέτρο σύγκρισης, ενώ η Γραμμική Παλινδρόμηση χρησιμοποιήθηκε 11 φορές και σε 2 έρευνες συμπεριλήφθηκε ως μέτρο σύγκρισης.

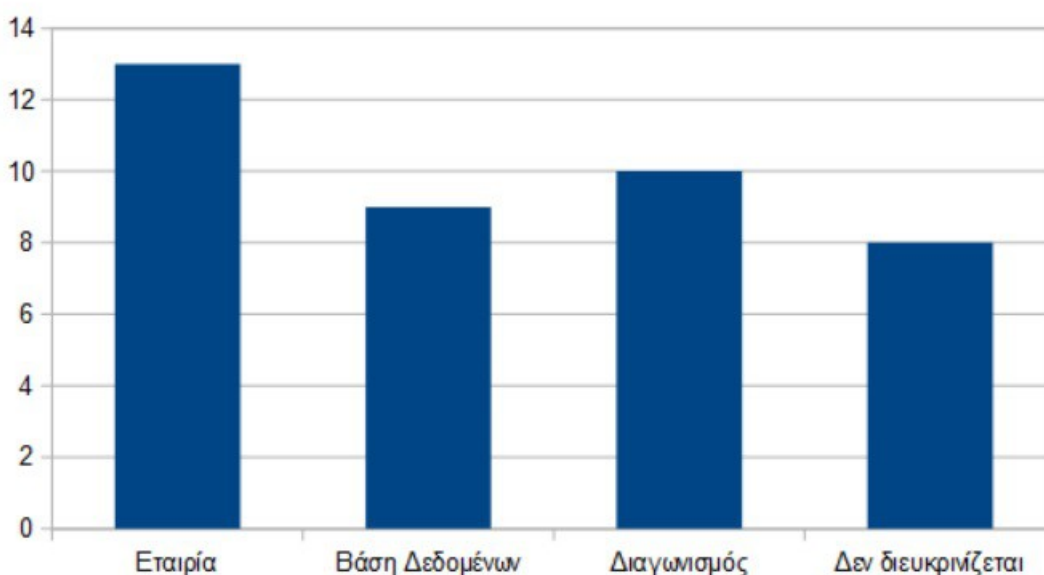
Παρατηρείται επίσης ότι σε άρθρα όπου γίνεται σύγκριση προϋπάρχοντων αλγορίθμων, ο αλγόριθμος που στις περισσότερες περιπτώσεις βρέθηκε να έχει την υψηλότερη αξιοπιστία ήταν ο XGBOOST, γεγονός το οποίο πιθανά επεξηγεί και την παραπάνω προτίμηση στη χρήση του.

Μια ακόμη σημαντική παρατήρηση αφορά την ταξινόμηση των ερευνών με βάση το διάστημα στο οποίο γίνεται πρόβλεψη πωλήσεων, αν δηλαδή η πρόβλεψη αφορά ημερήσιες, εβδομαδιαίες, μηνιαίες ή τριμηνιαίες πωλήσεις. Όπως φαίνεται στο παρακάτω διάγραμμα, περισσότερες έρευνες φαίνεται να αφορούν ημερήσιες προβλέψεις. Οι περιπτώσεις αυτές συνήθως αφορούσαν είτε βραχυπρόθεσμες προβλέψεις, είτε προβλέψεις που αφορούσαν ένα συγκεκριμένο διάστημα εκπτώσεων, για παράδειγμα Black Friday. Στο αντίθετο άκρο, μόνο σε μία έρευνα επέλεξαν οι συγγραφείς να αναλύσουν στο επίπεδο ωριαίων πωλήσεων για πολύ βραχυπρόθεσμες προβλέψεις.



Εικόνα 5.3: Διαστήματα στα οποία γίνεται η πρόβλεψη πωλήσεων στις έρευνες

Ένας επιπρόσθετος παράγοντας είναι η προέλευση των δεδομένων που χρησιμοποίησε η κάθε έρευνα. Λόγω της φύσης του αντικειμένου, τα δεδομένα σε όλες τις έρευνες αφορούν ιστορικό πωλήσεων επιχειρήσεων και εταιριών. Όμως σε άλλες περιπτώσεις δόθηκαν ευθέως από την εταιρία στους ερευνητές, ενώ σε άλλες τα προμηθεύτηκαν εμμέσως, είτε από κάποια ανοιχτή βάση δεδομένων, είτε από κάποιον διαγωνισμό, για παράδειγμα της ιστοσελίδας Kaggle. Σε ορισμένες περιπτώσεις δημοσιεύθηκαν και περισσότερες από μία έρευνες χρησιμοποιώντας το ίδιο σετ δεδομένων από διαγωνισμό. [Theresa et al, 2020. Mohanapriya & Mohana Saranya, 2020. Singh & Srivastava, 2020. Behera & Nain, 2020. Behera & Nain, 2019. Punam et al, 2019]

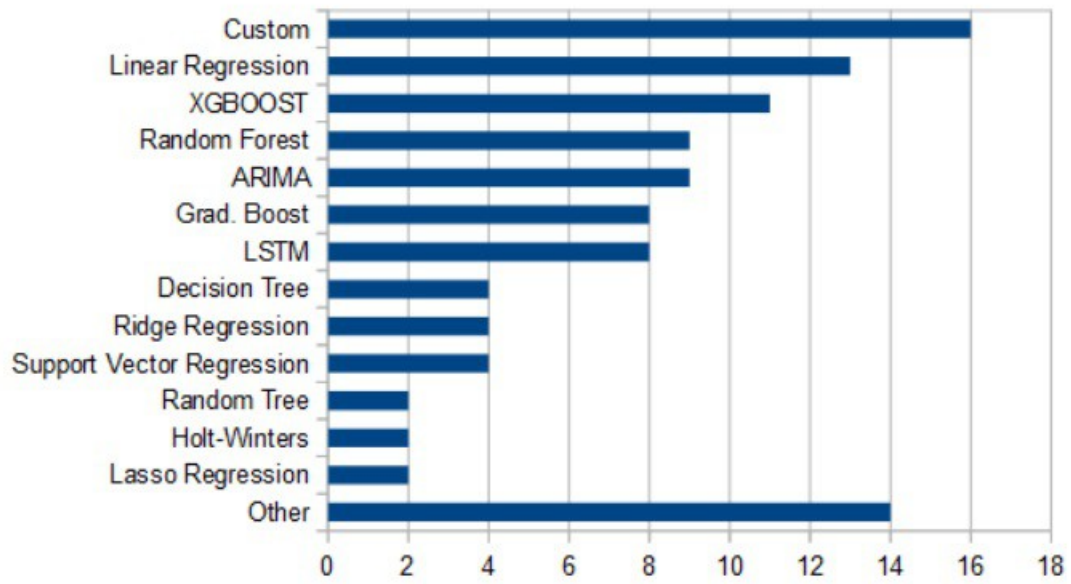


Εικόνα 5.4: Προελεύσεις σετ δεδομένων που χρησιμοποιήθηκαν στις έρευνες

Αξίζει να σημειωθεί πως το γεγονός ότι σε λίγες περιπτώσεις φαίνεται να χρησιμοποιούνται αλγόριθμοι ως μέτρο σύγκρισης, οι περιπτώσεις δηλαδή που σημειώνονται στον πίνακα με Ο, μπορεί να εξηγηθεί από το ότι στις περισσότερες περιπτώσεις οι συγγραφείς δεν ξεκίνησαν την έρευνά τους έχοντας επιλέξει έναν συγκεκριμένο αλγόριθμο, αλλά θεωρούσαν όλους τους αλγόριθμους αναπόσπαστο κομμάτι της έρευνάς τους με σκοπό να καταλήξουν στον πιο ακριβή για την δική τους περίπτωση.

| Apbpo | Ayojobi | | | | | | | | | | | | | | | | | | | | | | | | | | |
|-------|---------|------------|----------|------|-----|---------------|-------------|---------------|-------|-------|--------------|-----|-----|-----|------|----------|---------|--------|--------------|-------------|----------|----------|-----|-----------------|------------|--------|---|
| | XGBOOST | Grad Boost | AdaBoost | LSTM | GAN | Decision Tree | Random Tree | Random Forest | LSSVM | ARIMA | Holt-Winters | ANN | CNN | RNN | FFNN | Azure ML | Q Learn | Cubist | Season Naive | Linear Regr | Ridge R. | Lasso R. | SVR | Multi Linear R. | Deep Learn | Custom | |
| 1 | X | | | | X | X | | | | | | | | | | | | | | | | | | | | | X |
| 2 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 3 | X | 0 | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 4 | | | | | | X | | | | | | | | | | | | | | | | | | | | | X |
| 5 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 6 | | | | | | | | | | X | | | | | | | | | | | | | | | | | X |
| 7 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 8 | X | X | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 9 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 10 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 11 | | | | | | | | 0 | | | | | | | | | | | | | | | | | | | X |
| 12 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 13 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 14 | | | | | | | | X | | | | | | | | | | | | | | | | | | | X |
| 15 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 16 | X | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 17 | X | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 18 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 19 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 20 | X | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 21 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 22 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 23 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 24 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 25 | X | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 26 | X | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 27 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 28 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 29 | X | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 30 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 31 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 32 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 33 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 34 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 35 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 36 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 37 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |
| 38 | | | | | | | | | | | | | | | | | | | | | | | | | | | X |

Η συχνότητα χρήσης του κάθε αλγορίθμου παρατίθεται και παρακάτω σε μορφή διαγράμματος.



Εικόνα 5.5: Συχνότητα χρήσης των αλγορίθμων στα συμπεριλαμβανόμενα άρθρα

Κεφάλαιο 6: Συμπεράσματα - Προβλήματα που αντιμετωπίστηκαν

6.1 Συμπεράσματα

Στον παρακάτω πίνακα παρουσιάζονται συνοπτικά τα συγκεντρωτικά στοιχεία του συνόλου των άρθρων που συμπεριλήφθηκαν στην παρούσα έρευνα. Ακολουθεί πρώτα μια σύντομη περιγραφή των περιεχομένων των στηλών του πίνακα.

Αρχικά στην πρώτη σελίδα δίνονται τα βασικά αναγνωριστικά στοιχεία των άρθρων που συμπεριλήφθηκαν. Στην πρώτη στήλη δίνεται ο αύξων αριθμός του κάθε άρθρου. Όπως αναφέρθηκε, αυτός αντιστοιχεί με τον αριθμό που αντιπροσωπεύει το καθένα στον πίνακα της εννοιο-κεντρικής ανάλυσης του προηγούμενου κεφαλαίου. Γίνεται διαχωρισμός ανάμεσα στα άρθρα που δημοσιεύθηκαν σε επιστημονικά περιοδικά και αυτά που παρουσιάστηκαν σε συνέδρια, και στη συνέχεια είναι αριθμημένα με χρονολογική σειρά, από το πιο πρόσφατο στο παλαιότερο. Στη δεύτερη στήλη δίνεται ο σύνδεσμος DOI, όπου αυτός είναι διαθέσιμος, για την διευκόλυνση της άμεσης πρόσβασης στην κάθε έρευνα. Στις επόμενες στήλες ακολουθούν με σειρά τα ονόματα των συγγραφέων, το έτος δημοσίευσης και οι τίτλοι των άρθρων.

Παρατηρείται αύξηση του αριθμού των ερευνών πάνω στο αντικείμενο κατά το έτος 2020, καθώς επίσης και προτίμηση παρουσιάσης των άρθρων σε συνέδρια αντί για δημοσίευσή τους σε επιστημονικά περιοδικά στα νωρίτερα έτη, ενώ κατά το έτος 2020 εμφανίζεται μια πιο εξισορροπημένη κατανομή ανάμεσα στα δύο.

Οι στήλες της δεύτερης σελίδας παρουσιάζουν τα περιεχόμενα των άρθρων. Στην πρώτη στήλη αναφέρεται ο τομέας της αγοράς στον οποίο επιχειρείται η πρόβλεψη πωλήσεων. Στην δεύτερη παρουσιάζονται οι αλγόριθμοι μηχανικής μάθησης που χρησιμοποιήθηκαν, είτε σαν κομμάτι της έρευνας, είτε για να αναπτυχθούν νέοι αλγόριθμοι, είτε ως μέτρο σύγκρισης. Στη συνέχεια ακολουθεί η πηγή των σετ δεδομένων που χρησιμοποιήθηκαν, όπου αυτή γνωστοποιείται. Στην τελευταία στήλη συμπεριλαμβάνονται ό,τι πληροφορίες παρέχονταν για τις μεταβλητές πρόβλεψης που χρησιμοποιήθηκαν. Σε ορισμένες έρευνες ονοματίζονταν όλα οι μεταβλητές σε άλλες δινόταν απλά ο αριθμός τους και κάποια κατηγοριοποίηση, ενώ σε άλλες δεν αναφέρονταν καθόλου, με αποτέλεσμα να παρουσιάζεται έλλειψη συνοχής στα περιεχόμενα αυτής της στήλης σε σύγκριση με τις υπόλοιπες.

Τέλος, οι δυο στήλες της τρίτης σελίδας αφορούν τα αποτελέσματα των άρθρων. Πιο συγκεκριμένα, η πρώτη στήλη αφορά το είδος των εξαγόμενων αποτελεσμάτων που παρείχαν οι αλγόριθμοι, και αν αφορούσαν ωριαίες, ημερήσιες, εβδομαδιαίες, μηνιαίες ή τριμηνιαίες πωλήσεις, όπου αυτό διευκρινιζόταν, ενώ η τελευταία στήλη αναφέρει τα συμπεράσματα στα οποία κατέληξαν οι συγγραφείς μετά το πέρας της έρευνας τους.

| # | DOI | Author(s) | Year | Title |
|----------|---|---|------|---|
| 1 | https://doi.org/10.1108/DTA-11-2020-0286 | Wang, S., Yang, Y., Liu, Y., Feng, L., Jin, B. | 2021 | M-GAN-XGBOOST model for sales prediction and precision marketing strategy making of each product in online stores |
| 2 | https://doi.org/10.3390/mh11120558 | Xia, Z., Xue, S., Wu, L., Sun, J., Chen, Y., Zhang, R. | 2020 | Future-aware trend alignment for sales predictions |
| 3 | https://doi.org/10.1007/s10619-020-07294-y | Wang, H. | 2020 | ForeXGBoost: passenger car sales prediction based on XGBoost |
| 4 | https://doi.org/10.18280/na.340309 | Chen, T., Yin, H., Chen, H., Wang, H., Zhou, X., Li, X. | 2020 | An insurance sales prediction model based on deep learning |
| 5 | https://doi.org/10.1007/s10115-019-01404-8 | Theresa, L., Medikonda, V.R., Reddy, K.V.N. | 2020 | Online sales prediction via trend alignment-based multitask recurrent neural networks |
| 6 | - | Yellamma, P., Abhinav, B., Jaya Vashnavi, B., Ushaavini, G., Srinivas, M. | 2020 | Prediction of big mart sales using exploratory machine learning techniques |
| 7 | - | Mohanapriya, S., Mohana Saranya, S. | 2020 | Forecasting techniques for sales prediction |
| 8 | - | Madhuvanithi, K., Nallakaruppan, M.K., Senthikumar, N.C., Siva Rama Krishna, S. | 2019 | Sales prediction using machine learning algorithm |
| 9 | - | Dong, X. | 2019 | Car sales prediction using machine learning algorithms |
| 10 | https://doi.org/10.3303/CET1651154 | | 2016 | The research on the electronic commerce sales prediction based on the improved LSSVM algorithm |
| Zuvédpuc | https://doi.org/10.1109/HPCC-DSS-SmartCity-DependSys3894.2021.000058 | Xie, H., Liu, L. | 2022 | Adaptive GRU With Regularization Item for Book Sales Prediction |
| 11 | https://doi.org/10.1109/ICSSSS4381.2022.9782271 | Dutta, S.R., Das, S., Chatterjee, P. | 2022 | Smart Sales Prediction of Pharmaceutical Products |
| 12 | https://doi.org/10.1007/978-981-16-9423-3_53 | Du, Y., Wei, K., Wang, Y., Jia, J. | 2022 | New Energy Vehicles Sales Prediction Model Combining the Online Reviews Sentiment Analysis |
| 13 | https://doi.org/10.1007/978-981-16-7466-2_17 | Hu, Y., Xu, B. | 2022 | Medical Equipment Sales Management Prediction System Based on LSTM Algorithm |
| 14 | https://doi.org/10.1007/978-3-030-82196-8_38 | Gertsen, D., Reshadat, V. | 2022 | Identifying Leading Indicators for Tactical Truck Parts Sales Predictions Using LASSO |
| 15 | https://doi.org/10.1099/ECS0559.2021.9396994 | Ramachandra, H.V., Balaraju, G., Rajashekar, A., Patil, H. | 2021 | Machine learning application for black friday sales prediction framework |
| 16 | https://doi.org/10.1088/1742-6596/1756/1/012015 | Huo, Z. | 2021 | Sales Prediction based on Machine Learning |
| 17 | https://doi.org/10.1088/1742-6596/1712/012042 | Singh, K., Boorna, P.M., Eaganathan, U. | 2020 | A Sales Prediction Method Based on LSTM with Hyper-Parameter Search |
| 18 | https://doi.org/10.1016/j.procs.2020.09.114 | Singh, A., Srivastava, S. | 2020 | E-Commerce System for Sale Prediction Using Machine Learning Technique |
| 19 | https://doi.org/10.1109/ICCSA50499.2020.9316540 | Liao, W., Ye, G., Yin, Y., Yan, W., Ma, Y., Zuo, D. | 2020 | A hybrid machine learning model for sales prediction |
| 20 | https://doi.org/10.1109/ICITSI50517.2020.9264911 | Edwarado, T.O., Rudevyan, Y. | 2020 | Auto Parts Sales Prediction based on Machine Learning for Small Data and a Long Replacement Cycle |
| 21 | https://doi.org/10.1109/ICITSI50517.2020.9264911 | Wisessa, O., Adriansyah, A., Khaif, O.I. | 2020 | Sales prediction analysis in determining new minimarket stores |
| 22 | https://doi.org/10.1016/j.procs.2020.09.114 | Cristiani, M., Pasetto, L., Tomazzoli, C. | 2020 | Prediction Analysis Sales for Corporate Services Telecommunications Company using Gradient Boost Algorithm |
| 23 | https://doi.org/10.1007/978-3-030-23756-1_26 | Tekin, A.T., Cebik, F. | 2020 | A knowledge-intensive methodology for explainable sales prediction |
| 24 | https://doi.org/10.1007/978-981-15-5827-6_3 | Behera, G., Nain, N. | 2020 | Model Performance Evaluation: Sales Prediction |
| 25 | https://doi.org/10.1007/978-981-15-4015-8_37 | Hafner, O., Kucera, E., Moravcik, M. | 2020 | A comparative study of big mart sales prediction |
| 26 | https://doi.org/10.1007/978-981-13-9042-5_72 | Kodali, S., Dabhru, M., Rao, B.T. | 2020 | Sales prediction of Synergy Slovaki, Ltd. using microsoft azure machine learning and ADMLA |
| 27 | https://doi.org/10.1007/978-3-030-23756-1_26 | Silva, J., Villa, J.V., Cabrera, D. | 2020 | A Q-Learning Approach for Sales Prediction in Heterogeneous Information Networks |
| 28 | https://doi.org/10.1007/978-3-030-23756-1_26 | Buyar, V., Abdel-Raouf, A. | 2020 | Click and sales prediction for digital advertisements: Real world application for OTAs |
| 29 | https://doi.org/10.1145/3388218.3388228 | Shen, X., Yao, C., Ester, M., Li, Z., Bu, J., Zhou, X., Ye, Y., Wang, C. | 2019 | Sale forecast for basic commodities based on artificial neural networks prediction |
| 30 | https://doi.org/10.1145/3388218.3388228 | Behera, G., Nain, N. | 2019 | A Convolutional Neural Networks-based Model for Sales Prediction |
| 31 | https://doi.org/10.1145/3345035.3345048 | Lingxian, Y., Jiaqing, K., Shihual, W. | 2019 | Multi-task based sales predictions for online promotions |
| 32 | https://doi.org/10.1109/GUCON.2018.8675060 | Punam, K., Pamula, R., Jain, P.K. | 2019 | Grid search optimization (GSO) based future sales prediction for big mart |
| 33 | https://doi.org/10.1109/ICCECOHIE.2018.8659115 | Cheriyen, S., Ibrahim, S., Mohanan, S., Treasa, S. | 2019 | Online retail sales prediction with integrated framework of K-mean and neural network |
| 34 | https://doi.org/10.1109/ICM.2018.00020 | Chen, T., Yin, H., Chen, H., Wu, L., Wang, H., Zhou, X., Li, X. | 2018 | A two-level statistical model for big mart sales prediction |
| 35 | https://doi.org/10.1109/ICCAA.2018.8777620 | Gopalakrishnan, T., Choudhary, R., Prasad, S. | 2018 | Intelligent Sales Prediction Using Machine Learning Techniques |
| 36 | | | | TADA: Trend Alignment with Dual-Attention Multi-task Recurrent Neural Networks for Sales Prediction |
| 37 | | | | Prediction of sales value in online shopping using linear regression |
| 38 | | | | |

| Domain | ML Technique | Dataset | Features |
|---------------------|---|--|---|
| Online commerce | M-GAN, XGBOOST | Unspecified | Divided into external factors, consumer traffic, consumer perception, potential purchase behaviors |
| General commerce | Future-Aware Trend Alignment with Dual Attention (F-TADA) | Kaggle contest and pharmaceutical company | Kaggle data for long-period supermarket sales and XGBoost extracted features for short-period pesticide sales |
| Automotive industry | Fore(X)GBoost (custom XGBoost) | Alibaba Cloud dataset from January 2011 to December 2017 | Brand, Displacement, Department, Compartment, Price, Gearbox Type, Typed, Cylinder Number, Rated Passenger |
| Insurance industry | Deep Learning based LSTM Neural Network & LNEST based MLR | Jan 2016-Dec 2019 billing data of insurance company | Promotions, holidays, seasons etc. |
| Online commerce | Trend Alignment with Dual-Attention Multitask RNN (TADA+) | Favorita & One Stop Warehouse commercial datasets | 13 variables for Favorita dataset, 11 variables for OSW dataset |
| Supermarket sales | Linear regression, ridge regression, decision tree, random forest | Big Mart supermarket sales | 12 unspecified attributes |
| Retail | ARIMA and Feed-Forward-NeuralNetwork (FFNN) | Weekly sales of 45 Walmart stores | Markdown, consumer per index, IsHoliday, store size, unemployment, store type, fuel price, temperature |
| Supermarket sales | Gradient Boost (GBT) and XG Boost | Big Mart supermarket sales | Unspecified |
| Automotive industry | Linear, Random Tree, Random Forest | Automobile dataset from unspecified source | Make (Brand), Fuel type, Width, Height, Horsepower, City mileage, Highway mileage, Price |
| Electronics | Improved LSSVM algorithm | Unspecified sales data from July 2012 to July 2015 | Unspecified |
| Books | AdaGRU, LSTM | Chinese convenience store chain & GitHub open dataset | Unspecified |
| Pharmaceutical | Seasonal Naive, Holt-Winters, SARIMA, Linear Regression | GitHub open dataset | Unspecified |
| Automotive industry | Random Forest, SARIMA | Wind Database, Pcauto Website, JProP Website | Changing piles, Manufacturing capacity, Cost of materials, Gas price, Vehicle demand, Number of patents, Seasonal Index |
| Medical industry | LSTM | Unspecified | Unspecified |
| Automotive industry | LASSO, ARIMA, Holt-Winters | Unnamed case company & Publicly available databases | 7 features were chosen from the pool of 33 potential predictor variables |
| Retail | Random Forest Regressor & XGBoost | Unspecified Black Friday sales data | product ID, User ID, Age, Gender, occupation, stay_in_current_city_years, marital status, Product, category_1-3, and Purchase |
| Online commerce | ARIMA, Regression, XGB, Random Forest, MLP, LSTM | Walmart's H5 data set | Product category, state, calendar and price-related information |
| Pharmaceutical | LSTM with special loss function, compared to AutoML models | Kaggle contest (Rossmann Store Sales) | 5 Sales features (weekly/monthly), 2 Store features, 4 external features (weekly/monthly) |
| Online commerce | Random Forest & Gradient Boosting | California, Texas & Wisconsin Wal-mart stores | Unspecified |
| Retail | LightGBM & XGBoost hybrid | Kaggle Brazilian E-Commerce Public Dataset by Olist Store | State ID, Store ID, time, item level, department, product category, price, promotion, weekday, special events and more |
| Automotive industry | Linear Regression, Support Vector Regression & stacking model | UCI machine learning repository public dataset | Unspecified |
| Retail | Data Mining & Deep Learning | PT XYZ retail in Indonesia | store type, rack size, distance to competitors, age of the store, location domain, and average sales per day |
| Telecommunications | Gradient Boosted Trees | Unspecified telecommunications company sales 2016-2018 | category, region, item type and opportunity ID, quarter, product name, product sub-component, service product (MID) & sales revenue |
| Food production | Random Forest, Bayes Model, SFBM, Knowledge-Intensive | Past promotional sales from unspecified source | Multiple configurations tested ranging from 5 to 12 features |
| Retail | Linear R, Ridge R, Lasso R, Random Forest & XG Boost | Big-Mart sales data from Analytics Vidya site | 13 item related features, 15 outlet related features |
| Retail | XGBoost, Linear R, Decision Tree, Ridge R | Big Mart retail 2013 sales data | Item ID, It Fat, It Visibility, It Type, Outlet Type, It MRP, Out ID, It Weight, Out Size, Out Establish Year, Out Location Type & It Out Sales |
| Alcohol | Microsoft Azure ML & ARIMA | Svijany Slovakia sales for 09/2014-05/2016 | Date, Item number, Name of goods, Number of sold pieces |
| Mobile | Q-Learning | Apple iPhone unit sales 2007-2009 | Year/Quarter, Units sold, YOY growth |
| Tourism | Gradient Boost, XGBoost, Random Forest, GLM, Deep Neural Net | Unspecified Online Travel Agency (OTA) | Year/Quarter, Units sold, YOY growth |
| Food/Retail | ARIMA, ANN, Linear R, Support Vector Regression | National Administrative Department of Statistics of Colombia | Definite/Pre-/Canceled Reservations, temporal moving averages and standard deviations |
| Pharmaceutical | Convolutional Neural Networks | Unamed pharmaceutical company referred to as Y LLC | Unspecified |
| Online commerce | Deep Item Network for Online Promotions (DINOP) | Alibaba's 11.11 Global Shopping Festival | Dataset contained 19 features which are only partially explained |
| Retail | XGBoost | Big Mart retail sales data | Static item inherent profiles, Static item Sales data, Static target Users profiles, Dynamic item Sales data, Dynamic target Users |
| Online commerce | LSTM, K-mean, Neural Network | UK-based online retail activity 01/12/2010 to 09/12/2011 | Item MRP, Outlet type, ID, Item weight, Item visibility, Outlet years, Outlet, Item type, Outlet size, Item fat content, Outlet location type |
| Retail | Linear Regression, Support Vector Regression, Cubist | Big Mart retail 2013 sales data | Invoice No, Stock Code, Description, Quantity, Date, Unit Price, Customer ID, and Country |
| Fashion | Generalized Linear Model, Decision Tree & Gradient Boosted Trees | e-Fashion store sales from 2015 to 2017 | Item ID, It Fat, It Visibility, It Type, Outlet Type, It MRP, Out ID, It Weight, Out Size, Out Establish Year, Out Location Type & It Out Sales |
| Retail & Energy | Trend Alignment with Dual-Attention (TADA) Multi-Task RNN | Favorita & OSW commercial datasets | Category, City, Type of items and description, number of items, Quantity, Quarter, Sales Revenue, Year, SKU description, Week, Year |
| Online commerce | Linear Regression | Various Zmart superstore department sales data 2011-2014 | 6 internal & 7 external features for Favorita dataset, 5 internal & 6 external features for OSW dataset |
| | | | 25 unspecified attributes |

| Output | Conclusion |
|----------------------------|--|
| Daily sales volume | Effectiveness demonstrated |
| Monthly sales volume | Effectiveness demonstrated |
| Quarterly sales volume | Outperforms benchmark algorithms |
| Sales volume | Outperforms traditional LSTM networks |
| Sales for specific items | Outperforms conventional predictive methods |
| Weekly sales volume | Outperforms previous single level model |
| Sales for specific items | FFNN outperformed ARIMA |
| Car sales by brand | XG Boost outperforms GBT |
| Monthly sales volume | Price greatest impact, Chevrolet leading brand |
| | Outperforms traditional LSVM algorithm |
| Hourly sales volume | Outperforms other machine learning models |
| Sales volume | Linear Regression outperforms other methods |
| Monthly sales volume | Effectiveness demonstrated |
| Monthly sales volume | Method tested |
| Monthly sales volume | LASSO outperforms other methods |
| Daily sales per customer | Methods tested |
| Daily sales volume | Date and price information helps prediction |
| Weekly sales volume | Outperforms AutoML tool models |
| Sales volume | Methods tested |
| Daily sales volume | Outperforms individual machine learning models |
| Monthly sales volume | Methods tested |
| Daily sales volume | Effectiveness demonstrated |
| Quarterly sales volume | Methods tested |
| Promotional sales volume | Knowledge-intensive approaches outperform |
| Sales for specific items | XGBoost outperformed the others |
| Sales for specific items | XGBoost outperformed the others |
| Weekly sales for sp. items | Methods tested |
| Quarterly sales volume | Methods tested |
| Daily sales volume | Methods tested |
| Sales volume | Methods tested |
| Weekly sales volume | Effectiveness demonstrated, outperforms ARIMA |
| Promotional sales volume | Outperforms previous practical methodologies |
| Sales volume | Methods tested |
| Sales volume | Effectiveness demonstrated |
| Sales for specific items | Outperforms individual machine learning models |
| Quarterly sales volume | Methods tested |
| Daily/Weekly sales volume | Outperforms other machine learning models |
| Quarterly sales volume | Methods tested |

Πίνακας 6.1: Συγκριτικά αποτελέσματα της έρευνας

6.2 Προβλήματα που αντιμετωπίστηκαν

Όπως αναφέρθηκε και στο προηγούμενο κεφάλαιο, δεν ακολουθείται από όλες τις έρευνες κάποια τυποποιημένη μεθοδολογία για την προσέγγιση της έρευνας του αντικειμένου της πρόβλεψης πωλήσεων με τη χρήση μηχανικής μάθησης. Κάποια άρθρα έκαναν πλήρη, λεπτομερή ανάλυση του αντικειμένου και της μεθοδολογίας που ακολούθησαν, ενώ άλλα ήταν συνοπτικά και δεν παρείχαν πολλές πληροφορίες που μπορούσαν να αναφερθούν κατά τη συμπερίληψή τους στο παρόν κείμενο.

Η έλλειψη συνοχής ανάμεσα στην ποιότητα των άρθρων είχε ως αποτέλεσμα να προκύπτουν περιορισμένα συμπεράσματα κατά την εννοιο-κεντρική ανάλυση στο σύνολο των συγκεντρωμένων άρθρων. Πιθανά στο μέλλον όταν θα έχει καθιερωθεί μια πιο ευρέως διαδεδομένη μεθοδολογία έρευνας στο συγκεκριμένο αντικείμενο, να υπάρχει η δυνατότητα να προκύψουν πιο ουσιαστικά αποτελέσματα.

Συνεπώς, θα ήταν σκόπιμο να επαναληφθεί η έρευνα στο μέλλον, όταν το αντικείμενο θα έχει επεκταθεί αρκετά ώστε να υπάρχει η δυνατότητα επιλεκτικής συμπερίληψης μόνο άρθρων που ανταποκρίνονται σε πιο αυστηρά κριτήρια επιλογής όσον αφορά την ποιότητα καταγραφής των μεθόδων και αποτελεσμάτων τους.

Βιβλιογραφία

- Behera, G., Nain, N. "Grid Search Optimization (GSO) Based Future Sales Prediction for Big Mart," 2019 15th International Conference on Signal-Image Technology & Internet-Based Systems (SITIS), 2019, pp. 172-178.
- Behera, G., Nain, N. (2020). A Comparative Study of Big Mart Sales Prediction. In: Nain, N., Vipparthi, S., Raman, B. (eds) Computer Vision and Image Processing. CVIP 2019. Communications in Computer and Information Science, vol 1147. Springer, Singapore.
- Beheshti-Kashi, S., Karimi, H.R., Thoben, K.-D., Lütjen, M. Teucke, M. (2014). "A survey on retail sales forecasting and prediction in fashion markets". Systems Science & Control Engineering, 3:1, 154-161.
- Buyar, V., and Abdel-Raouf, A. "A Convolutional Neural Networks-based Model for Sales Prediction". AIRC '19: Proceedings of the 2019 International Conference on Artificial Intelligence, Robotics and Control, December 2019. Pages 61–67.
- Chen, T. et al., "TADA: Trend Alignment with Dual-Attention Multi-task Recurrent Neural Networks for Sales Prediction," 2018 IEEE International Conference on Data Mining (ICDM), 2018, pp. 49-58.
- Chen, T., Yin, H., Chen, H. et al. "Online sales prediction via trend alignment-based multitask recurrent neural networks". Knowl Inf Syst 62, 2139–2167 (2020).
- Cheriyana, S., Ibrahim, S., Mohanan, S. and Treesa, S., "Intelligent Sales Prediction Using Machine Learning Techniques," 2018 International Conference on Computing, Electronics & Communications Engineering (iCCECE), 2018, pp. 53-58.
- Choudhary, G. T., R. and Prasad, S., "Prediction of Sales Value in Online shopping using Linear Regression," 2018 4th International Conference on Computing Communication and Automation (ICCCA), 2018, pp. 1-6.
- Cogito Tech – Roger Brown. "What are Features in Machine Learning and Why it is Important?". 15 July 2019.
<https://cogitotech.medium.com/what-are-features-in-machine-learning-and-why-it-is-important-e72f9905b54d>
- Cristani, M., Pasetto, L., Tomazzoli, C. "A knowledge-intensive methodology for explainable sales prediction". Procedia Computer Science 176, pp. 1180-1187 (2020).
- Dai, Y., Huang, J., "A Sales Prediction Method Based on LSTM with Hyper-Parameter Search". J. Phys.: Conf. Ser. 1756 012015 (2021).
- Dong, X. "The research on the electronic commerce sales prediction based on the improved LSSVM algorithm". Chemical Engineering Transactions 51, pp. 919-924 (2016).
- Du, Y., Wei, K., Wang, Y., Jia, J. "New Energy Vehicles Sales Prediction Model Combining the Online Reviews Sentiment Analysis: A Case Study of Chinese New Energy Vehicles Market". Lecture Notes in Electrical Engineering, Volume 854 LNEE, pp. 424 - 431, 3rd International Conference on Artificial Intelligence, 2022, Kunming
- Dutta, S.R., Das, S., Chatterjee, P. "Smart Sales Prediction of Pharmaceutical Products". 8th International Conference on Smart Structures and Systems, ICSSS 2022, Chennai.
- Edwardo, T. O. and Ruldeviyani, Y., "Sales Prediction Analysis in Determining New Minimarket Stores," 2020 International Conference on Information Technology Systems and Innovation (ICITSI), 2020, pp. 79-84.
- Gerritsen, D., Reshadat, V. "Identifying Leading Indicators for Tactical Truck Parts' Sales Predictions Using LASSO". Lecture Notes in Networks and Systems, Volume 295, pp. 518 – 535, 2022, Intelligent Systems Conference, IntelliSys 2021.
- Haffner, O., Kučera, E. and Moravčík, M., "Sales Prediction of Svijany Slovakia, Ltd. Using

- Microsoft Azure Machine Learning and ARIMA," 2020 Cybernetics & Informatics (K&I), 2020, pp. 1-9.
- Hu, Y., Xu, B. "Medical Equipment Sales Management Prediction System Based on LSTM Algorithm". Lecture Notes on Data Engineering and Communications Technologies, Volume 102, pp. 157 - 164, 2022
- Huo, Z., "Sales Prediction based on Machine Learning," 2021 2nd International Conference on E Commerce and Internet Technology (ECIT), 2021, pp. 410-415.
- IBM Cloud Education. "What is Machine Learning?". 15 July 2020.
<https://www.ibm.com/cloud/learn/machine-learning>
- Kodali, S., Dabbiru, M., Rao, B.T. (2020). A Q-Learning Approach for Sales Prediction in Heterogeneous Information Networks. In: Das, A., Nayak, J., Naik, B., Pati, S., Pelusi, D. (eds) Computational Intelligence in Pattern Recognition. Advances in Intelligent Systems and Computing, vol 999. Springer, Singapore.
- Liao, W., Ye, G., Yin, Y., Yan, W., Ma, Y. and Zuo, D., "Auto Parts Sales Prediction based on Machine Learning for Small Data and a Long Replacement Cycle," 2020 IEEE/ACS 17th International Conference on Computer Systems and Applications (AICCSA), 2020, pp. 1-5
- Liu Y, Feng L, Jin B. "Future-Aware Trend Alignment for Sales Predictions". Information. 2020; 11(12):558.
- Madhuvanthi, K., Nallakaruppan, M.K., Senthilkumar, N.C., Siva Rama Krishnan, S. "Car sales prediction using machine learning algorithms". International Journal of Innovative Technology and Exploring Engineering 8(5), pp. 1039-1050 (2019).
- Mohanapriya, S., Mohana Saranya, S. "Sales prediction using machine learning algorithm". International Journal of Advanced Science and Technology 29(3 Special Issue), pp. 1049-1055 (2020).
- Punam, K., Pamula, R. and Jain, P. K., "A Two-Level Statistical Model for Big Mart Sales Prediction," 2018 International Conference on Computing, Power and Communication Technologies (GUCON), 2018, pp. 617-620.
- Ramachandra, H. V., Balaraju, G., Rajashekar, A. and Patil, H., "Machine Learning Application for Black Friday Sales Prediction Framework," 2021 International Conference on Emerging Smart Computing and Informatics (ESCI), 2021, pp. 57-61.
- Silva, J., Villa, J.V., Cabrera, D. (2020). Sale Forecast for Basic Commodities Based on Artificial Neural Networks Prediction. In: Herrera, F., Matsui, K., Rodríguez-González, S. (eds) Distributed Computing and Artificial Intelligence, 16th International Conference. DCAI 2019. Advances in Intelligent Systems and Computing, vol 1003. Springer, Cham.
- Singh, A., Srivastava, S. (2020). "Model Performance Evaluation: Sales Prediction." In: Batra, U., Roy, N., Panda, B. (eds) Data Science and Analytics. REDSET 2019. Communications in Computer and Information Science, vol 1229. Springer, Singapore.
- Singh, K. et al. "E-Commerce System for Sale Prediction Using Machine Learning Technique". J. Phys.: Conf. Ser. 1712 012042 (2020).
- Studer, S., Bui, T.B., Drescher, C., Hanuschkin, A., Winkler, L., Peters, S., Müller, K.-R. "Towards CRISP-ML(Q): A Machine Learning Process Model with Quality Assurance Methodology". Mach. Learn. Knowl. Extr. 2021, 3, 392–413.
- Tekin, A.T., Cebi, F. (2020). Click and Sales Prediction for Digital Advertisements: Real World Application for OTAs. In: Kahraman, C., Cebi, S., Cevik Onar, S., Oztaysi, B., Tolga, A., Sari, I (eds) Intelligent and Fuzzy Techniques in Big Data Analytics and Decision Making. INFUS 2019. Advances in Intelligent Systems and Computing, vol 1029. Springer, Cham.
- Theresa, I., Medikonda, V.R., Reddy, K.V.N. "Prediction of big mart sales using exploratory machine learning techniques". International Journal of Advanced Science and Technology 29(6), pp. 2906-2911 (2020).
- Tsoumakas, G. (2019). "A survey of machine learning techniques for food sales prediction". Artif

Intell Rev (2019) 52:441–447

- Wang, H.P. (2020). “An insurance sales prediction model based on deep learning”. *Revue d'Intelligence Artificielle*, Vol. 34, No. 3, pp. 315-321.
- Wang, J., "A hybrid machine learning model for sales prediction," 2020 International Conference on Intelligent Computing and Human-Computer Interaction (ICHCI), 2020, pp. 363-366.
- Wang, S. and Yang, Y. (2021), "M-GAN-XGBOOST model for sales prediction and precision marketing strategy making of each product in online stores", *Data Technologies and Applications*, Vol. 55 No. 5, pp. 749-770.
- Webster, J., Watson, R.T., “Analysing the past to prepare for the future: Writing a literature review”. *MIS Quarterly* Vol. 26 No. 2, pp. xiii-xxiii/June 2002.
- Wisesa, O., Adriansyah, A. and Khalaf, O. I., "Prediction Analysis Sales for Corporate Services Telecommunications Company using Gradient Boost Algorithm," 2020 2nd International Conference on Broadband Communications, Wireless Sensors and Powering (BCWSP), 2020, pp. 101-106.
- Xia, Z., Xue, S., Wu, L. et al. “ForeXGBoost: passenger car sales prediction based on XGBoost”. *Distrib Parallel Databases* 38, 713–738 (2020).
- Xie, H., Liu, L. “Adaptive GRU With Regularization Item for Book Sales Prediction”. 2021 IEEE 23rd International Conference on High Performance Computing and Communications, 7th International Conference on Data Science and Systems, 19th International Conference on Smart City and 7th International Conference on Dependability in Sensor, Cloud and Big Data Systems and Applications, HPCC-DSS-SmartCity-DependSys 2021, Haikou - Hainan, pp. 245-252, 2022.
- Xin, S., Ester, M., Bu, J., Yao, C., Li, Z., Zhou, X., Ye, Y., and Wang, C. “Multi-task based Sales Predictions for Online Promotions”. *CIKM '19: Proceedings of the 28th ACM International Conference on Information and Knowledge Management*, November 2019. Pages 2823–2831.
- Yellamma, P., Abhinav, B., Jaya Vaishnavi, B., Ushaswini, G., Srinivas, M. “Forecasting techniques for sales prediction”. *International Journal of Advanced Science and Technology* 29(6), pp. 3042-3049 (2020).
- You, L., Kou, J., Wang, S. “Online Retail Sales Prediction with Integrated Framework of K-mean and Neural Network”. *ICEME 2019: Proceedings of the 2019 10th International Conference on E business, Management and Economics*, July 2019. Pages 115–118.