



**Πρόγραμμα Μεταπτυχιακών Σπουδών
στην Αναλυτική των Επιχειρήσεων και Επιστήμη των Δεδομένων**

Τμήμα Οργάνωσης και Διοίκησης Επιχειρήσεων

Διπλωματική Εργασία

«Χρήση του Copernicus API στην Επιχειρηματική Αναλυτική»

**του
Γεωργίου Παπαδόπουλου του Αναστασίου**

**Υποβλήθηκε ως απαιτούμενο για την απόκτηση του μεταπτυχιακού διπλώματος
στην Αναλυτική των Επιχειρήσεων και Επιστήμη των Δεδομένων**

Νοέμβριος 2023

Ευχαριστίες

Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή μου κ. Ταραμπάνη Κωνσταντίνο, καθηγητή του τμήματος Οργάνωσης και Διοίκησης Επιχειρήσεων του Πανεπιστημίου Μακεδονίας για το ερέθισμα που μου έδωσε ώστε να ασχοληθώ με αυτόν τον ιδιαίτερα ενδιαφέροντα κλάδο της αναλυτικής των επιχειρήσεων και επιστήμης των δεδομένων.

Ακόμη, θα ήθελα να ευχαριστήσω θερμά τον κ. Καλαμπόκη Ευάγγελο, επίκουρου καθηγητή του τμήματος Οργάνωσης και Διοίκησης Επιχειρήσεων του Πανεπιστημίου Μακεδονίας για τη βοήθεια που μου προσέφερε προκειμένου να υλοποιηθεί η παρούσα διπλωματική εργασία.

Τέλος, θα ήθελα να ευχαριστήσω την οικογένεια μου για τη στήριξή τους σε όλο αυτό το χρονικό διάστημα.

Περίληψη

Η παρούσα διπλωματική εργασία με θέμα τη «Χρήση του Copernicus API στην Επιχειρηματική Αναλυτική» διενεργήθηκε στο πλαίσιο της απόκτησης του μεταπτυχιακού διπλώματος από το Πανεπιστήμιο Μακεδονίας και συγκεκριμένα από το Π.Μ.Σ. στην Αναλυτική των Επιχειρήσεων και Επιστήμη των Δεδομένων.

Στην εργασία μελετήθηκαν δεδομένα που αντλήθηκαν από το ευρωπαϊκό πρόγραμμα Copernicus. Τα δεδομένα που αντλήθηκαν αφορούν ενεργειακούς δείκτες παραγωγής και ζήτησης ηλεκτρικής ενέργειας μέσω ανανεώσιμων πηγών ενέργειας καθώς και κλιματικά δεδομένα όπως ταχύτητα ανέμου, θερμοκρασία, επιφανειακή ηλιακή ακτινοβολία, πίεση, βροχόπτωση κ.α.

Οι ανανεώσιμες πηγές ενέργειας αναφέρονται στην ενέργεια που παράγεται από φυσικούς πόρους και είναι φιλικές προς το περιβάλλον. Η Ελλάδα έχει κάνει σημαντικά βήματα προόδου στην ανάπτυξη ανανεώσιμων πηγών ενέργειας τα τελευταία χρόνια μεταβαίνοντας σε καθαρότερες και πιο βιώσιμες ενεργειακές επιλογές.

Στόχος της εργασίας ήταν η πραγματοποίηση ενός μοντέλου πρόβλεψης της παραγόμενης ηλεκτρικής ενέργειας την επόμενη ώρα από ανανεώσιμες πηγές και πιο συγκεκριμένα από την ηλιακή ενέργεια που συλλέγεται μέσω φωτοβολταϊκών στην Ελλάδα. Ο κώδικας υλοποιήθηκε στο Jupyter Notebook σε Python και για το μοντέλο πρόβλεψης χρησιμοποιήθηκε ο αλγόριθμος XGBoost. Ο κώδικας της εργασίας υπάρχει στην ιστοσελίδα GitHub καθώς και αποσπασματικά στην παρούσα εργασία.

Τα αποτελέσματα της πρόβλεψης ήταν ικανοποιητικά και μπορούν να χρησιμοποιηθούν στην ασφαλή πρόβλεψη της παραγωγής ηλεκτρικής ενέργειας μέσω φωτοβολταϊκών συστημάτων. Το μοντέλο φιλοδοξεί να βοηθήσει στην κατανόηση και την περαιτέρω μελέτη των ανανεώσιμων πηγών ενέργειας και την παροχή φιλική προς το περιβάλλον ενέργεια μέσω της σωστής εκμετάλλευσης των ανανεώσιμων πηγών που κατέχει η χώρα. Τέλος, η έρευνα έχει δυνατότητες εξέλιξης μέσω υλοποίησης μοντέλων για περισσότερα μέσα παραγωγής ανανεώσιμης ενέργειας στην Ελλάδα καθώς και συγκριτική μελέτη αυτών.

Λέξεις κλειδιά: Ανανεώσιμες Πηγές Ενέργειας (ΑΠΕ), Φωτοβολταϊκά, Machine Learning (ML), XGBoost, SHAP

Abstract

The topic of thesis is the " Use of Copernicus API in Business Analytics ". Thesis carried out in the context of obtaining the master's degree from the University of Macedonia, specifically from 'Business Analytics and Data Science' master.

Studied data obtained from the European Copernicus project. The data collected concern energy indicators of electricity production and demand through renewable energy sources as well as climate data such as wind speed, temperature, surface solar radiation, pressure, precipitation, etc.

Renewable energy refers to energy produced from natural resources and is environmentally friendly. Greece has made significant strides in renewable energy development in recent years by switching to cleaner and more sustainable energy options.

The aim of the work was to create a model for predicting the electricity produced in the next hour from renewable sources and more specifically from the solar energy collected through photovoltaics in Greece. The code was implemented in Jupyter Notebook and XGBoost algorithm was used for the prediction. The code of the work is available on GitHub as well as excerpts in this work.

The prediction results were satisfactory and can be used in the safe prediction of electricity generation through photovoltaic systems. The model aspires to help in the understanding and further study of renewable energy sources and the provision of environment-friendly energy through proper exploitation of the renewable resources that the country possesses. Finally, the research has potential for development through the implementation of models for more means of renewable energy production in Greece as well as a comparative study of them.

Keywords: Renewable Energy Sources (RES), Photovoltaics (PV), Machine Learning (ML), XGBoost, SHAP

Περιεχόμενα

| | |
|--|-----|
| Ευχαριστίες | i |
| Περίληψη..... | ii |
| Abstract | iii |
| 1. Εισαγωγή..... | 1 |
| 1.1 Το πρόγραμμα Copernicus..... | 1 |
| 1.1.1 Δεδομένα του Copernicus | 1 |
| 1.1.2 Εφαρμογές από το Copernicus..... | 4 |
| 1.1.3 Τα οφέλη του προγράμματος Copernicus | 5 |
| 1.1.4 Copernicus API | 6 |
| 1.2 Ανανεώσιμες πηγές ενέργειας..... | 7 |
| 1.2.1 Μορφές Ανανεώσιμων Πηγών Ενέργειας | 9 |
| 1.2.1 Ανανεώσιμες Πηγές Ενέργειας ανά χώρα | 15 |
| 1.2.2 Ανανεώσιμες Πηγές στην Ελλάδα..... | 17 |
| 1.3 Υλοποίηση Κώδικα..... | 20 |
| 1.3.1 Python..... | 20 |
| 1.3.2 Jupyter Notebook | 21 |
| 1.3.3 Μηχανική Μάθηση..... | 21 |
| 1.3.4 Αλγόριθμος XGBoost | 23 |
| 2. Μελέτη Δεδομένων Μοντέλου Πρόβλεψης..... | 25 |
| 2.1 Ανάλυση δεδομένων και μεταβλητών | 25 |
| 2.1.1 Θερμοκρασία αέρα..... | 25 |
| 2.1.2 Πίεση στο επίπεδο της θάλασσας..... | 25 |
| 2.1.3 Επιφανειακή ακτινοβολία | 25 |
| 2.1.4 Συνολική βροχόπτωση | 26 |
| 2.1.5 Ταχύτητα ανέμου..... | 26 |
| 2.1.6 Παραγωγή ηλιακής ενέργειας | 26 |
| 2.1.7 Παραγωγή αιολικής ενέργειας | 26 |
| 2.1.8 Παραγωγή υδροηλεκτρικής ενέργειας | 26 |
| 2.1.9 Ζήτηση ηλεκτρικής ενέργειας..... | 27 |
| 2.2 Επεξεργασία δεδομένων | 27 |
| 2.2.1 Εισαγωγή Δεδομένων..... | 27 |
| 2.3 Στατιστική Ανάλυση Δεδομένων | 29 |

| | | |
|-----|--|----|
| 2.4 | Δημιουργία επιπλέον μεταβλητών..... | 35 |
| 3. | Κατασκευή Μοντέλου Μηχανικής Μάθησης | 40 |
| 4. | Αξιολόγηση Μοντέλου Μηχανικής Μάθησης..... | 50 |
| 4.1 | Mean Absolute Error - MAE | 50 |
| 4.2 | Median Absolute Error - MedAE | 51 |
| 4.3 | Root Mean Squared Error (RMSE) | 51 |
| 4.4 | R-squared (R^2) | 53 |
| 5. | Συμπεράσματα | 55 |
| 6. | Βιβλιογραφία..... | 56 |
| 7. | Παράρτημα..... | 62 |

1. Εισαγωγή

Στην παρούσα διπλωματική εργασία αντλήθηκαν δεδομένα από το ευρωπαϊκό πρόγραμμα Copernicus σε στόχο την μελέτη τους και την πραγματοποίηση ενός μοντέλου πρόβλεψης. Τα δεδομένα που αντλήθηκαν αφορούν ενεργειακούς δείκτες και παραγωγή ηλεκτρικής ενέργειας μέσω ανανεώσιμων πηγών ενέργειας καθώς και κλιματικά δεδομένα όπως ταχύτητα ανέμου, θερμοκρασία, επιφανειακή ηλιακή ακτινοβολία, πίεση και βροχόπτωση. Στόχος είναι η μελέτη των δεδομένων και η αξιοποίηση τους στην παραγωγή ενός μοντέλου πρόβλεψης της παραγωγής ηλεκτρικής ενέργειας μέσω φωτοβολταϊκών για την Ελλάδα. Τα δεδομένα είναι από το 1979 έως και τον Ιούλιο του 2023. Η οργάνωση των δεδομένων, η στατιστική μελέτη τους και το μοντέλο πρόβλεψης πραγματοποιήθηκαν με τη γλώσσα προγραμματισμού Python στο περιβάλλον του Jupyter Notebook.

1.1 Το πρόγραμμα Copernicus

Το πρόγραμμα Copernicus είναι το ευρωπαϊκό πρόγραμμα γεωσκόπησης και γεωμορφολογίας, το οποίο εξετάζει τα χαρακτηριστικά του περιβάλλοντος με προσπάθεια να εκμεταλλευτεί τα δεδομένα αυτά προς όφελος των ευρωπαίων πολιτών. Το πρόγραμμα Copernicus διαχειρίζεται η Ευρωπαϊκή Επιτροπή σε συνεργασία με τον Ευρωπαϊκό Οργανισμό Διαστήματος (ESA), τον Ευρωπαϊκό Οργανισμό Περιβάλλοντος (ΕΟΠ) και άλλους φορείς. (*Copernicus*, 2023).

1.1.1 Δεδομένα του προγράμματος Copernicus

Το πρόγραμμα Copernicus συνδυάζει δεδομένα που συλλέγονται από ένα σύνθετο σύνολο συστημάτων πηγών στο διάστημα μέσω δορυφόρων παρατήρησης της γης, στο έδαφος, στη θάλασσα και στον αέρα, μέσω επιτόπιων, εναέριων και θαλάσσιων αισθητήρων, για την παροχή των υπηρεσιών του. Τα δεδομένα που συλλέγονται υποβάλλονται σε επεξεργασία για την παροχή ενός συνόλου υπηρεσιών που βασίζονται σε αξιόπιστες και σχεδόν σε πραγματικό χρόνο πληροφορίες. Τα δεδομένα καθώς και οι πληροφορίες από αυτά παρέχονται πλήρη, ελεύθερα και ανοικτά σε πρόσβαση σε όλους τους χρήστες. Πιο αναλυτικά:

- Το διαστημικό τμήμα του Copernicus περιλαμβάνει τους ειδικούς δορυφόρους γεωσκόπησης Sentinels 1, 2, 3 και 6 καθώς και τους Sentinel 4 και 5 που είναι μετεωρολογικοί δορυφόροι επί του σκάφους του Ευρωπαϊκού Οργανισμού για την Εκμετάλλευση Μετεωρολογικών Δορυφόρων (EUMETSAT¹). Επιπρόσθετα υπάρχουν και οι δορυφόροι τρίτων που αναφέρονται ως συνεισφέρουσες αποστολές του Copernicus που τελούν αποστολές υπό τη διαχείριση εθνικών, ευρωπαϊκών ή διεθνών οργανισμών και παρέχουν πληθώρα δεδομένων για τις υπηρεσίες του Copernicus.
- Το επίγειο τμήμα περιλαμβάνει τον έλεγχο των αποστολών των δορυφόρων καθώς και τη διαχείριση και επεξεργασία των δεδομένων που λαμβάνονται από αυτούς. Επίσης συμπληρώνει τα δεδομένα του διαστημικού τμήματος με τη βοήθεια αισθητήρων τοποθετημένους στις όχθες ποταμών, μεταφερόμενους με μετεωρολογικά μπαλόνια ή αεροπλάνα, τραβηγμένους στη θάλασσα από πλοία ή παρασυρόμενους στον ωκεανό με πλωτήρες ή σημαντήρες. Αυτή η επιτόπια συνιστώσα είναι απαραίτητη για την παροχή υψηλής ποιότητας, αξιόπιστων και βιώσιμων υπηρεσιών πληροφόρησης. Το επίγειο τμήμα παρέχεται από διεθνείς οργανισμούς καθώς και από εθνικές δημόσιες και ιδιωτικές εγκαταστάσεις.

Από την έναρξη λειτουργίας του πρώτου δορυφόρου του Copernicus, του Sentinel-1, το 2014, ο αντίκτυπος του προγράμματος έγινε ευρέως αισθητός. Το πρόγραμμα αποφέρει οικονομικά, κοινωνικά, περιβαλλοντικά και στρατηγικά οφέλη στην Ευρωπαϊκή καθώς και στην ευρύτερη κοινότητα των χρηστών. Τα δεδομένα που συλλέγει, επεξεργάζεται και παρέχει χωρίζονται και αξιοποιούνται από διάφορες υπηρεσίες όπως:

- i. Την Υπηρεσία Παρακολούθησης της Ατμόσφαιρας του Κοπέρνικου, που παρέχει συνεχή δεδομένα και πληροφορίες σχετικά με τη σύνθεση της ατμόσφαιρας της Γης. Η υπηρεσία περιγράφει την τρέχουσα κατάσταση, προβλέπει την κατάσταση λίγες ημέρες πριν και αναλύει τα αρχεία δεδομένων για τα τελευταία χρόνια.

¹ EUMETSAT European Organisation for the Exploitation of Meteorological Satellites - Ευρωπαϊκός Οργανισμός για την Εκμετάλλευση Μετεωρολογικών Δορυφόρων

Υποστηρίζει πολλές εφαρμογές σε διάφορους τομείς, συμπεριλαμβανομένης της υγείας, της περιβαλλοντικής παρακολούθησης, των ανανεώσιμων πηγών ενέργειας, της μετεωρολογίας και της κλιματολογίας.

- ii. Την υπηρεσία παρακολούθησης του θαλάσσιου περιβάλλοντος του Copernicus που παρέχει τακτικές και συστηματικές πληροφορίες αναφοράς σχετικά με τη φυσική και βιο-γεωχημική κατάσταση, τη μεταβλητότητα και τη δυναμική των ωκεάνιων και θαλάσσιων οικοσυστημάτων για τους παγκόσμιους ωκεανούς και τις ευρωπαϊκές περιφερειακές θάλασσες. Οι παρατηρήσεις και οι προβλέψεις που παράγονται από την υπηρεσία υποστηρίζουν θαλάσσιες εφαρμογές που στοχεύουν στην ασφάλεια.
- iii. Την υπηρεσία παρακολούθησης της γης του Copernicus που παρέχει γεωγραφικές πληροφορίες σχετικά με την κάλυψη γης και τις αλλαγές της, τη χρήση γης, την κατάσταση της βλάστησης, τον κύκλο του νερού και τις ενεργειακές μεταβλητές της επιφάνειας της γης σε ευρύ φάσμα χρηστών στην Ευρώπη και σε ολόκληρο τον κόσμο. Υποστηρίζει εφαρμογές σε διάφορους τομείς όπως ο χωροταξικός και πολεοδομικός σχεδιασμός, η διαχείριση δασών και υδάτων, η γεωργία και η επισιτιστική ασφάλεια, η διατήρηση και αποκατάσταση της φύσης, η αγροτική ανάπτυξη και άλλοι.
- iv. Την Υπηρεσία Κλιματικής Αλλαγής του Κοπέρνικου που στηρίζει την κοινωνία παρέχοντας έγκυρες πληροφορίες σχετικά με το παρελθόν, το παρόν και το μέλλον του κλίματος στην Ευρώπη και τον υπόλοιπο κόσμο. Η αποστολή της είναι να υποστηρίξει τις πολιτικές προσαρμογής και μετριασμού της Ευρωπαϊκής Ένωσης παρέχοντας συνεπείς και έγκυρες πληροφορίες σχετικά με την κλιματική αλλαγή.
- v. Την υπηρεσία για εφαρμογές ασφάλειας που αποσκοπεί στη στήριξη των πολιτικών της Ευρωπαϊκής Ένωσης μέσω της παροχής πληροφοριών για την αντιμετώπιση των προκλήσεων ασφάλειας της Ευρώπης. Βελτιώνει την πρόληψη, την ετοιμότητα και την αντίδραση σε κρίσεις σε τρεις βασικούς τομείς: επιτήρηση των συνόρων, θαλάσσια επιτήρηση και στήριξη της εξωτερικής δράσης της ΕΕ.
- vi. Την υπηρεσία διαχείρισης καταστάσεων έκτακτης ανάγκης που παρέχει σε όλους τους φορείς που εμπλέκονται στη διαχείριση φυσικών καταστροφών, ανθρωπογενών καταστάσεων έκτακτης ανάγκης και ανθρωπιστικών κρίσεων έγκαιρες και ακριβείς γεωχωρικές πληροφορίες που προέρχονται από δορυφορική

τηλεπισκόπηση και συμπληρώνονται από διαθέσιμες επιτόπιες πηγές ή ανοικτές πηγές δεδομένων. Το EMS του Copernicus αποτελείται από μια συνιστώσα χαρτογράφησης και μια συνιστώσα έγκαιρης προειδοποίησης.

1.1.2 Εφαρμογές του προγράμματος Copernicus

Το Copernicus υποστηρίζει ποικίλες εφαρμογές σε διάφορους τομείς όπως αναλύθηκε μέσω των υπηρεσιών του. Το φάσμα των εφαρμογών του προγράμματος Copernicus είναι τεράστιο και καλύπτει τον δημόσιο και τον ιδιωτικό τομέα σε πολλά τμήματα της αγοράς. Τα δεδομένα που συλλέγονται βρίσκουν εφαρμογή στους τομείς της κλιματικής αλλαγής, της ασφάλειας και τις καταστάσεις έκτακτης ανάγκης, την προστασίας του περιβάλλοντος, τη διαχείριση των αστικών περιοχών, του περιφερειακού και τοπικού σχεδιασμού, της γεωργίας, της δασοκομίας, της αλιείας, της υγείας, των μεταφορών, τη βιώσιμη ανάπτυξης, την πολιτικής προστασίας και του τουρισμού. Πιο αναλυτικά ορισμένες βασικές εφαρμογές περιλαμβάνουν:

- i. Τη γεωργία, μέσω της αξιολόγησης της χρήσης και των τάσεων της γεωργικής γης, των συνθηκών των καλλιεργειών και των προβλέψεων απόδοσης.
- ii. Τα ύδατα, χάρη στην ικανότητά του να παρακολουθεί τα φυσικά και βιογεωχημικά χαρακτηριστικά των παγκόσμιων ωκεανών και των περιφερειακών θαλασσών, το Copernicus βοηθάει στην αξιολόγηση της ποιότητας των υδάτων, τη χαρτογράφηση των αλιευτικών ζωνών ή την παρακολούθηση επιβλαβών ανθίσεων φυκιών.
- iii. Την κλιματική αλλαγή και το περιβάλλον, παρέχοντας δείκτες σχετικά με τους παράγοντες της κλιματικής αλλαγής, όπως τα επίπεδα διοξειδίου του άνθρακα, και αξιολογώντας τις επιπτώσεις της στο περιβάλλον, όπως το λιώσιμο των παγετώνων και την άνοδο της στάθμης της θάλασσας.
- iv. Την ενέργεια και τους φυσικούς πόρους, με την πρόβλεψη της παραγωγής ηλιακής και αιολικής ενέργειας, την επιλογή τοποθεσιών ανανεώσιμων πηγών ενέργειας, την παρακολούθηση της βιομάζας και την στάθμη των υδάτων. Με αυτό τον τρόπο καθιστά δυνατή την αποδοτική εκμετάλλευση των ανανεώσιμων πηγών ενέργειας και συμβάλει στην κάλυψη των αυξανόμενων παγκόσμιων ενεργειακών απαιτήσεων χωρίς την αύξηση των εκπομπών CO².

- v. Τη δασοκομία, μέσω χαρτογράφησης δασών και δασικών αλλαγών καθώς και στην καταγραφή των εθνικών δασικών απογραφών.
- vi. Τις αρχές δημόσιας υγείας με την παρακολούθηση της ποιότητας του αέρα και τις υπεριώδεις ακτίνες. Το Copernicus μπορεί επίσης να συμβάλει στον εντοπισμό περιοχών επιρρεπών στην εμφάνιση και εξάπλωση επιδημιών, οι οποίες εξαρτώνται από περιβαλλοντικούς παράγοντες, όπως η ποιότητα του νερού, της αποχέτευσης, των τροφίμων ή του αέρα.
- vii. Τις δημόσιες αρχές στη διαχείριση καταστροφών, μέσω στήριξης των επιχειρήσεων πολιτικής προστασίας από την ετοιμότητα, την πρόληψη έως και την αντιμετώπιση καταστροφών με χάρτες καταστροφών, παρακολούθηση των μεσοπρόθεσμων και μακροπρόθεσμων επιπτώσεων στο περιβάλλον, την ανθρώπινη ασφάλεια και την οικονομία.
- viii. Την ασφάλεια και άμυνα της ΕΕ, με συνοριακή και θαλάσσια επιτήρηση έως και παρακολούθηση υποδομών ζωτικής σημασίας, όπως σταθμοί ηλεκτροπαραγωγής, υπεράκτιες εξέδρες και αγωγούς.
- ix. Τον τουρισμό, παρέχοντας υπηρεσίες όπως η παρακολούθηση της ποιότητας των παράκτιων υδάτων ή ο εντοπισμός και παρακολούθηση πληθυσμών μεδουσών ή ανθίσεων φυκιών.
- x. Στις μεταφορές, με πληροφορίες υπηρεσιών δρομολόγησης πλοίων και σχετικά με περιβαλλοντικούς κινδύνους, όπως ηφαιστειακές εκρήξεις ή παρουσία θαλάσσιου πάγου και ρευμάτων που μπορούν να διαταράξουν τις αεροπορικές και θαλάσσιες μεταφορές.
- xi. Στην πολεοδομία, βοηθώντας τους δήμους και τις περιφέρειες της Ευρώπης να αντιμετωπίσουν ποικίλες προκλήσεις σχεδιασμού, όπως η διαχείριση αποβλήτων και η εκμετάλλευση ανανεώσιμων πηγών ενέργειας, παρέχοντας πληροφορίες σχετικά με τη χρήση γης και την ταξινόμηση της κάλυψης γης, την αστική ανάπτυξη, τις αστικές περιοχές πρασίνου, τις αστικές θερμικές νησίδες, τη στεγανότητα κ.λπ.

1.1.3 Τα οφέλη του προγράμματος Copernicus

Το Copernicus αποτελεί μακροπρόθεσμη επένδυση στην ευημερία του πλανήτη. Με τακτική και συνεχή παρατήρηση της Γης, με σκοπό την καλύτερη προετοιμασία στην αντιμετώπιση των περιβαλλοντικών προκλήσεων και την κατανόηση και ανταπόκριση

της κλιματικής αλλαγής. Τα δεδομένα του Copernicus στηρίζουν διάφορους κλάδους, δημιουργούν νέες εξειδικευμένες θέσεις εργασίας και συμβάλλουν στην ανάπτυξη της ευρωπαϊκής οικονομίας.

Με το πρόγραμμα Copernicus, η Ευρώπη εξαρτάται λιγότερο από πηγές δεδομένων τρίτων χωρών και μπορεί να ενεργεί ως αυτόνομος διεθνής παράγοντας. Οι πληροφορίες του είναι ζωτικής σημασίας για την ασφάλεια της Ευρώπης και συμβάλλει στην οικοδόμηση ενός ασφαλέστερου μέλλοντος για όλους τους ευρωπαίους πολίτες. Τέλος, υπάρχει σημαντική εξοικονόμηση κόστους από την ύπαρξη ενός ενιαίου προγράμματος στην ΕΕ, το οποίο θα συντονίζει τις εθνικές πρωτοβουλίες.

1.1.4 Copernicus API

Το Application Programming Interface (API), είναι η διεπαφή μέσω της οποίας τα δεδομένα τα οποία καταχωρούνται, σε μία εφαρμογή ή πλατφόρμα, αποστέλλονται σε έναν Server και στέλνονται ξανά πίσω στο χρήστη, με τις απαντήσεις και τα αποτελέσματα που επιθυμεί. Είναι ένα μεσάζον εργαλείο πληροφόρησης που επιτρέπει σε εφαρμογές να επικοινωνούν και να συνεργάζονται με άλλες εφαρμογές και χρήστες. Τα API είναι χρήσιμα για την πρόσβαση ή την ενσωμάτωση σε μια υπηρεσία.

Το πρόγραμμα Copernicus προσφέρει πρόσβαση στα δεδομένα και τις υπηρεσίες του μέσω διεπαφών προγραμματισμού εφαρμογών διευκολύνοντας με αυτές την ενσωμάτωση δεδομένων Copernicus σε μια ποικιλία εφαρμογών και πλατφόρμων. Επίσης, καθιστούν δυνατή την πρόσβαση στα δεδομένα μέσω προγραμματισμού και τη χρήση δορυφορικών δεδομένων και υπηρεσιών Copernicus από προγραμματιστές, ακαδημαϊκούς και οργανισμούς. Το Copernicus Data Space Ecosystem API προσφέρει πολλαπλές διεπαφές που κυμαίνονται από λίστα δεδομένων, λήψη αυτών ή/και οπτικοποίηση μέσω επεξεργασίας υπηρεσιών web όπως STAC, openEO και Sentinel Hub API. Μερικά τυπικά χαρακτηριστικά και API που παρέχονται από το Copernicus είναι:

1. Το Copernicus Open Access Hub API που παρέχει πλήρη και δωρεάν πρόσβαση σε προγραμματιστές, ερευνητές και οργανισμούς στα δορυφορικά δεδομένα και υπηρεσίες που προέρχονται από τους δορυφόρους Sentinel-1, Sentinel-2, Sentinel-3 και Sentinel-5P.

2. Τα συστήματα Copernicus DIAS, όπως το Sentinel Hub, που μέσω API παρέχουν πρόσβασης στην απόκτηση και επεξεργασία δεδομένων Copernicus για μια ποικιλία εφαρμογών και επιτρέπουν στους χρήστες να αναλύουν δεδομένα εξάγοντας τα δικά τους αποτελέσματα.
3. Το Copernicus Climate Data Store (CDS) API που παρέχει δεδομένα από την υπηρεσία Copernicus Climate Change Service (C3S) και την Υπηρεσία Παρακολούθησης Ατμόσφαιρας Copernicus (CAMS) που σχετίζονται με το κλίμα καθώς και εργαλεία για την ανάλυση των κλιματικών δεδομένων αυτών.

1.2 Ανανεώσιμες Πηγές Ενέργειας

Οι ανανεώσιμες πηγές ενέργειας είναι μορφές εκμεταλλεύσιμης ενέργειας που προέρχονται από την εκμετάλλευση του ανέμου, του ήλιου, της γεωθερμίας, της ροής του νερού κ.α. Οι ενεργειακές αυτές πηγές αναπληρώνονται με υψηλότερο ρυθμό από ό,τι καταναλώνονται και επομένως θεωρούνται βιώσιμες. Μερικά παραδείγματα ανανεώσιμων πηγών ενέργειας περιλαμβάνουν την ηλιακή ενέργεια, την αιολική ενέργεια, την υδροηλεκτρική ενέργεια, τη γεωθερμική ενέργεια και τα βιοκαύσιμα.

Οι ανανεώσιμες πηγές ενέργειας έχουν πολλά πλεονεκτήματα σε σχέση με τις μη ανανεώσιμες. Μερικά από τα οφέλη της χρήσης ανανεώσιμων πηγών ενέργειας περιλαμβάνουν την παραγωγή ενέργειας που δεν παράγει εκπομπές αερίων θερμοκηπίου από ορυκτά καύσιμα και μειώνει ορισμένους τύπους ατμοσφαιρικής ρύπανσης, διαφοροποιούν τον ενεργειακό εφοδιασμό και μειώνουν την εξάρτηση από εισαγόμενα καύσιμα, ενισχύοντας αξιοπιστία, την ασφάλεια και την ανθεκτικότητα του εθνικού δικτύου ηλεκτρικής ενέργειας με παράλληλη μείωση του κινδύνου διαρροών καυσίμων και ατυχημάτων που μπορούν να βλάψουν την ανθρώπινη υγεία και το περιβάλλον. Επιπλέον, δημιουργούνται νέες θέσεις εργασίας στην κατασκευή, την εγκατάσταση και την συντήρηση τους καθώς και νέοι επιστημονικοί κλάδοι για τη μελέτη και την αποδοτική τους αξιοποίηση.

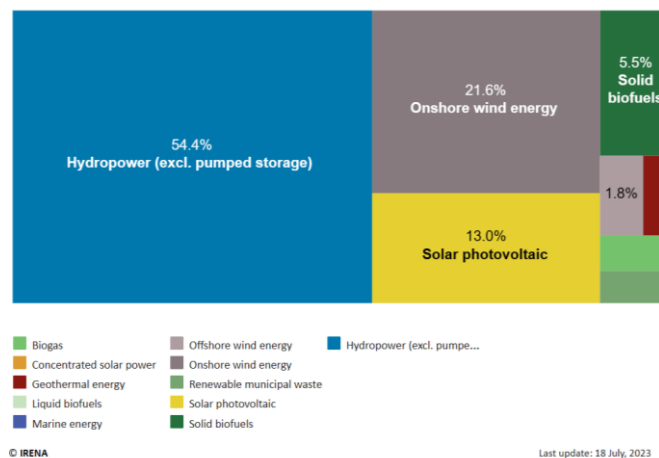
Οι ανανεώσιμες πηγές ενέργειας έχουν επίσης ορισμένα μειονεκτήματα. Μερικά από τα μειονεκτήματα περιλαμβάνουν το υψηλό αρχικό κόστος, τις περιορισμένες δυνατότητες αποθήκευσης καθώς και την πιθανή δύσκολη εγκατάσταση τους σε κάποιες γεωγραφικές περιοχές. Ωστόσο, η ανανεώσιμη ενέργεια εξακολουθεί να είναι πολύ καλύτερη από τη χρήση ορυκτών καυσίμων για την παραγωγή ηλεκτρικής ενέργειας,

καθώς οι πράσινες ή ανανεώσιμες πηγές ενέργειας είναι απεριόριστες και μπορούν να χρησιμοποιηθούν για την επανειλημμένη παραγωγή ενέργειας.

Πίνακας 1: Πλεονεκτήματα και μειονεκτήματα ανανεώσιμων πηγών ενέργειας (EnergySage, 2022).

| ΠΛΕΟΝΕΚΤΗΜΑΤΑ | ΜΕΙΟΝΕΚΤΗΜΑΤΑ |
|--|--|
| Οι ανανεώσιμες πηγές ενέργειας δεν θα εξαντληθούν | Οι ανανεώσιμες πηγές ενέργειας έχουν υψηλό αρχικό κόστος |
| Οι ανανεώσιμες πηγές ενέργειας έχουν χαμηλότερες απαιτήσεις συντήρησης | Οι ανανεώσιμες πηγές ενέργειας είναι διαλείπουσες |
| Οι ανανεώσιμες πηγές ενέργειας εξοικονομούν χρήματα | Οι ανανεώσιμες πηγές ενέργειας έχουν περιορισμένες δυνατότητες αποθήκευσης |
| Οι ανανεώσιμες πηγές ενέργειας έχουν πολλά περιβαλλοντικά οφέλη | Οι ανανεώσιμες πηγές ενέργειας έχουν γεωγραφικούς περιορισμούς |
| Οι ανανεώσιμες πηγές ενέργειας μειώνουν την εξάρτηση από ξένες πηγές ενέργειας | Οι ανανεώσιμες πηγές ενέργειας δεν είναι πάντα 100% απαλλαγμένες από άνθρακα |
| Οι ανανεώσιμες πηγές ενέργειας οδηγούν σε καθαρότερο νερό και αέρα | |
| Οι ανανεώσιμες πηγές ενέργειας δημιουργούν θέσεις εργασίας | |
| Οι ανανεώσιμες πηγές ενέργειας μπορούν να μειώσουν τα απόβλητα | |

Η αποδοτικότητα των ανανεώσιμων πηγών εξαρτάται από τα γεωγραφικά χαρακτηριστικά κάθε περιοχής και τη δυνατότητα εκμετάλλευσής τους. Η απόδοση τους μετράτε με τη βοήθεια του συντελεστή χωρητικότητας. Ο συντελεστής χωρητικότητας είναι ο λόγος της πραγματικής παραγωγής ενός σταθμού ηλεκτροπαραγωγής για μια χρονική περίοδο προς τη δυναμική του παραγωγή εάν λειτουργούσε με πλήρη ισχύ (IRENA, 2022).



Εικόνα 1: Ποσοστά ανανεώσιμων πηγών για την παραγωγή ηλεκτρικής ενέργειας (irena).

Τέλος, υπάρχουν διάφοροι τρόποι αποθήκευσης πράσινης ενέργειας συμπεριλαμβανομένων των μπαταριών, της αποθήκευσης αντλίας υδροηλεκτρικής ενέργειας, της αποθήκευσης ενέργειας πεπιεσμένου αέρα κ.α. Ανάλογα με το είδος της ανανεώσιμης ενέργειας που χρειάζεται να αποθηκευτεί υπάρχει και η βέλτιστη λύση. Η πλέον διαδεδομένη πηγή ανανεώσιμων πηγών για την παραγωγή ηλεκτρικής ενέργειας είναι η υδροηλεκτρική και ακολουθούν η αιολική και ηλιακή ενέργεια (IRENA, 2022).

1.2.1 Μορφές Ανανεώσιμων Πηγών Ενέργειας

Ηλιακή Ενέργεια

Η ηλιακή ενέργεια είναι το ακτινοβολούμενο φως που προέρχεται από τον ήλιο. Ο ήλιος δημιουργεί την ενέργειά του μέσω μιας θερμοπυρηνικής διαδικασίας που δημιουργεί θερμότητα και ηλεκτρομαγνητική ακτινοβολία (Householder, 1984). Η ηλιακή ενέργεια χωρίζεται σε τρία είδη τα θερμικά ηλιακά, τα παθητικά ηλιακά και τα φωτοβολταϊκά συστήματα.

- **Θερμικά Ηλιακά Συστήματα**

Τα θερμικά ηλιακά συστήματα π.χ. ηλιακοί θερμοσίφωνες, απορροφούν την ηλιακή ενέργεια και στη συνέχεια τη μεταφέρουν με τη μορφή θερμότητας σε κάποιο ρευστό π.χ. νερό. Η απορρόφηση της ηλιακής ενέργειας πραγματοποιείται μέσω ηλιακών συλλεκτών που είναι σε επαφή με νερό και του μεταδίδουν μέρος της θερμότητας που έλαβαν από την προσπίπτουσα ακτινοβολία.

- **Παθητικά Ηλιακά Συστήματα**

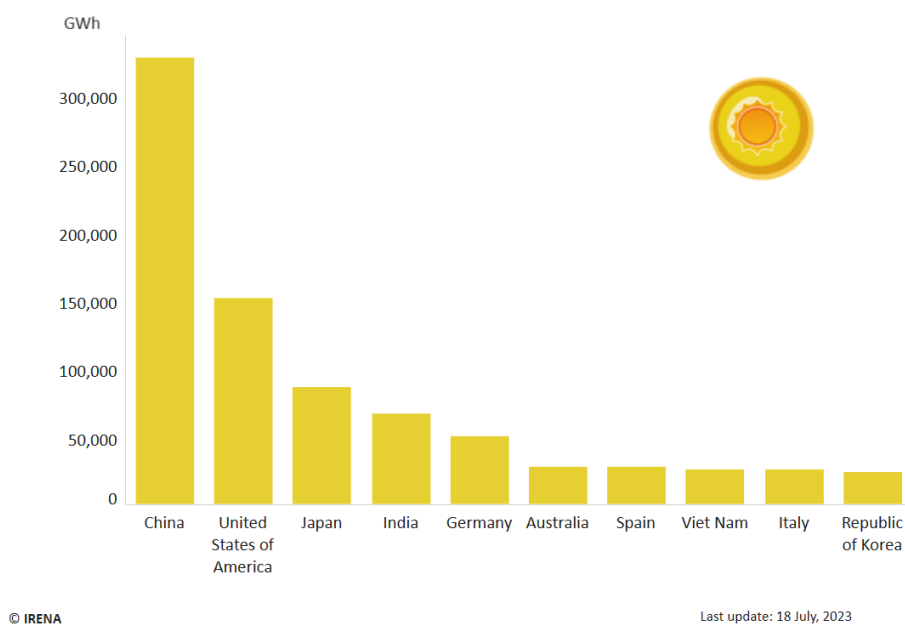
Τα παθητικά ηλιακά συστήματα κατάλληλα σχεδιασμένες διατάξεις που βοηθούν στην εκμετάλλευση της ηλιακής ενέργειας με σκοπό τον φυσικό φωτισμό των κτιρίων ή/και τη ρύθμιση της θερμοκρασίας σε αυτά.

- **Φωτοβολταϊκά Συστήματα**

Τα φωτοβολταϊκά συστήματα μετατρέπουν την ηλιακή ακτινοβολία σε ηλεκτρική ενέργεια. Χρησιμοποιούνται συνήθως για την ηλεκτροδότηση απομακρυσμένων από ηλεκτρικό δίκτυο συστημάτων και κτιρίων όπως δορυφόροι, φάροι και απομονωμένα σπίτια χωρίς όμως αυτό να είναι ο γενικός κανόνας. Ανάλογα με τη χρήση του

παραγόμενου ρεύματος τα φωτοβολταϊκά κατατάσσονται σε αυτόνομα συστήματα όταν η παραγόμενη ενέργεια καταναλώνεται επιτόπου και εξολοκλήρου και διασυνδεδεμένα συστήματα, όταν η παραγόμενη ενέργεια διοχετεύεται στο ηλεκτρικό δίκτυο για να μεταφερθεί και να καταναλωθεί (ΥΠΠΕΝ) .

Η ηλεκτροπαραγωγή από φωτοβολταϊκά έχει το πλεονέκτημα πως αποδίδει την μέγιστη ισχύ της κατά τη διάρκεια της ημέρας όπου παρουσιάζεται και η μέγιστη ζήτηση, δεν παράγει ατμοσφαιρικούς ρύπους ή διοξείδιο του άνθρακα και τέλος τα φωτοβολταϊκά μπορούν να τοποθετηθούν ακόμα και σε οροφές κτιρίων. Βέβαια η ηλιακή ενέργεια έχει ορισμένα μειονεκτήματα με το μεγαλύτερο να είναι η έλλειψη συνεχούς ηλιακού φωτός. Η «ποσότητα» του ήλιου ποικίλλει ανάλογα με την ώρα της ημέρας, τις εποχές του έτους, τον καιρό και την τοποθεσία. Η ηλιακή ενέργεια είναι μια πολλά υποσχόμενη εναλλακτική πηγή ενέργειας που μπορεί να χρησιμοποιηθεί για διάφορους σκοπούς, συμπεριλαμβανομένης της παραγωγής ηλεκτρικής ενέργειας αλλά και ως εφεδρική πηγή ενέργειας (Santhiarsa, 2005; Teknik, 2013).



Εικόνα 2: Ηλιακή Ενέργεια ανά χώρα (irena.org, 2023)

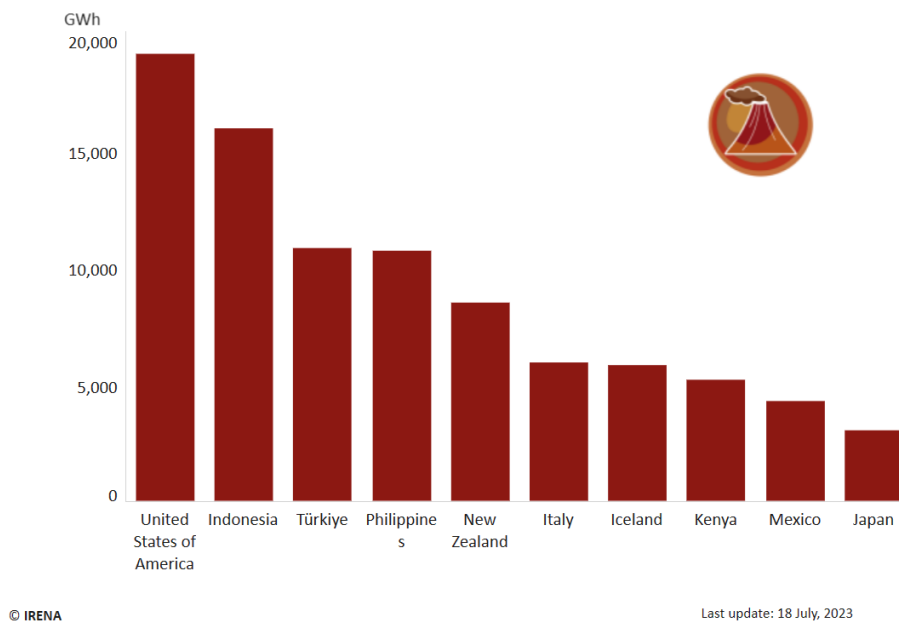
Γεωθερμική Ενέργεια

Η γεωθερμική ενέργεια είναι μια ήπια και ανανεώσιμη μορφή ενέργειας που παράγεται από την εκμετάλλευση της θερμότητας από τον φλοιό της Γης και έχει ελάχιστο έως μηδενικό περιβαλλοντικό αποτύπωμα. Μπορεί να χρησιμοποιηθεί άμεσα

ή έμμεσα για την παραγωγή ηλεκτρικής και θερμικής ενέργειας. Η εκμετάλλευση του υπόγειου γεωθερμικού δυναμικού προϋποθέτει η θερμοκρασία του υπογείου θερμού ρευστού να υπερβαίνει τους 30°C (ΥΠΠΕΝ) .

Ένα τμήμα της θερμικής αυτής ενέργειας προέρχεται από τον πυρήνα της Γης, ο οποίος σχηματίστηκε από την τριβή και τη βαρυτική έλξη κατά τη δημιουργία της Γης πριν από περισσότερα από 4 δισεκατομμύρια χρόνια ενώ η συντριπτική πλειονότητα της θερμότητας της Γης παράγεται συνεχώς από τη διάσπαση ραδιενεργών ισοτόπων (Lund, J. W. 2023). Οι γεωθερμικοί πόροι είναι διαθέσιμοι συνέχεια (24/7) και έχουν υψηλότερο συντελεστή χωρητικότητας από πολλές άλλες πηγές ενέργειας (Barbier, E. 2002).

Οι βασικές χρήσεις της γεωθερμικής ενέργειας κυμαίνονται από θέρμανση θερμοκηπίων και υδατοκαλλιεργειών, ξήρανση γεωργικών προϊόντων, αφαλάτωση νερού για την κάλυψη ύδρευσης έως και ήπιες βιομηχανικές χρήσεις και τηλεθέρμανση κτιρίων, οικισμών, χωριών ή και πόλεων.



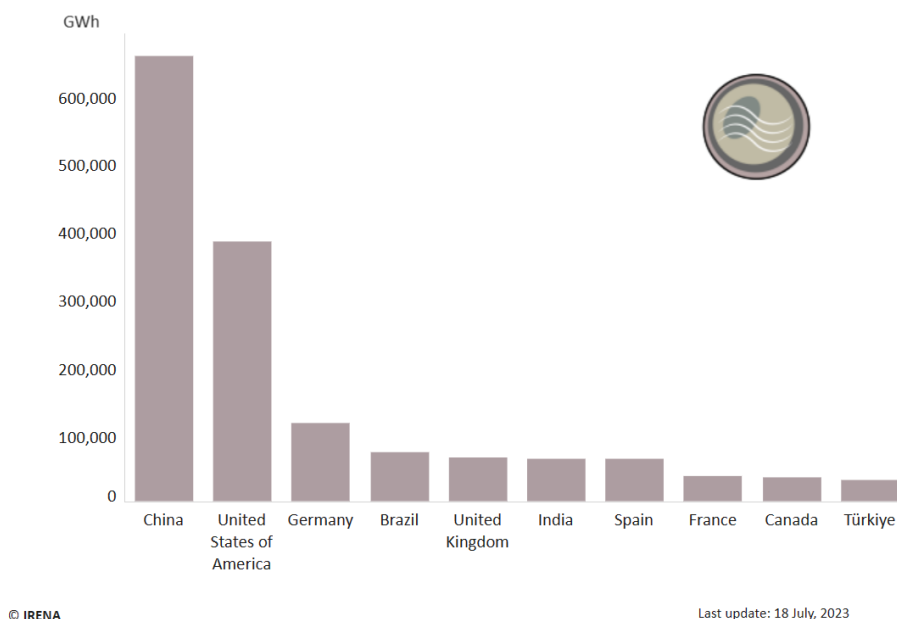
Εικόνα 3: Γεωθερμική Ενέργεια ανά χώρα (irena.org, 2023)

Αιολική Ενέργεια

Η αιολική ενέργεια είναι ο ρυθμός της ροής της κινητικής ενέργειας που μεταφέρεται από τον κινούμενο αέρα και αξιοποιείται με τη βοήθεια ανεμογεννητριών (Α/Γ). Η ανεμογεννήτρια διοχετεύει την ισχύ του ανέμου ο οποίος περιστρέφει τα πτερύγια της και στη συνέχεια μια γεννήτρια μετατρέπει την μηχανική αυτή ενέργειας σε ηλεκτρική ενέργεια (Kalmikov, 2017; Selin, 2023). Η αιολική ενέργεια μπορεί να χρησιμοποιηθεί για την παραγωγή ηλεκτρικής ενέργειας μέσω ανεμογεννητριών τοποθετημένες στην οροφή κτιρίων ή σε βουνά (Gordon 1986).

Οι ανεμογεννήτριες χρησιμοποιούνται για την πλήρη κάλυψη ή και τη συμπλήρωση των ενεργειακών αναγκών. Το παραγόμενο από τις ανεμογεννήτριες ηλεκτρικό ρεύμα είτε καταναλώνεται επιτόπου είτε διοχετεύεται στο ηλεκτρικό δίκτυο. Η παραγόμενη ηλεκτρική ενέργεια από τις ανεμογεννήτριες μπορεί να αποθηκευτεί με σκοπό τη χρησιμοποίησή της αργότερα. Η αποθήκευση της πραγματοποιείται είτε με ηλεκτρικούς συσσωρευτές (μπαταρίες) για μικρής κλίμακας παραγωγικές μη διασυνδεδεμένες στο κεντρικό δίκτυο μονάδες είτε με άντληση ύδατος σε τεχνητές λίμνες κατασκευασμένες σε υψόμετρο το οποίο είναι ικανό να τροφοδοτήσει υδροηλεκτρικούς σταθμούς όταν η παραγόμενη ηλεκτρική ενέργεια είναι μεγάλη (ΥΠΕΝ).

Η αιολική ενέργεια είναι πλέον μια σημαντική και αναπτυσσόμενη πηγή ανανεώσιμης ενέργειας με χερσαία και υπεράκτια αιολικά πάρκα συνδεδεμένα απευθείας με δίκτυα μεταφοράς ενέργειας. Επιπλέον, προηγμένες τεχνολογίες προσπαθούν να βελτιώσουν τα χαρακτηριστικά των ανεμογεννητριών με στόχο την πλήρη ενσωμάτωσή τους στο ηλεκτρικό δίκτυο (Blaabjerg 2017). Υπάρχουν πολλά και διαφορετικά μοντέλα που χρησιμοποιούνται για την αξιολόγηση των αιολικών πόρων και την επιλογή τοποθεσίας, καθώς και τα διάφορα στοιχεία των ανεμογεννητριών και τα οικονομικά των συστημάτων μετατροπής αιολικής ενέργειας τα οποία προσπαθούν να δώσουν βιώσιμη λύση στην ενέργεια. Η αιολική ενέργεια θεωρείται βιώσιμη ανανεώσιμη πηγή ενέργειας και έχει πολύ μικρότερο αντίκτυπο στο περιβάλλον σε σύγκριση με την καύση ορυκτών καυσίμων (Herbert 2007).

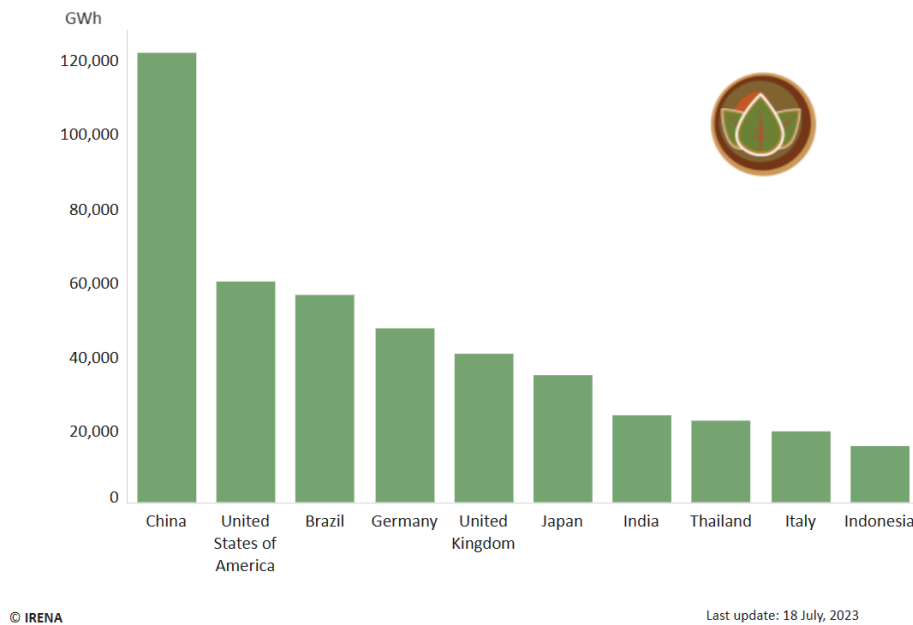


Εικόνα 4: Αιολική Ενέργεια ανά χώρα (irena.org, 2023).

Βιομάζα

Η βιομάζα είναι ανανεώσιμη ενέργεια που έχει βιολογική (οργανική) προέλευση και προέρχεται από τα φυτά, τα ζώα ή τον άνθρωπο. Πρακτικά περιλαμβάνεται σε αυτήν οποιοδήποτε υλικό προέρχεται άμεσα ή έμμεσα από τον φυτικό ή ζωικό κόσμο καθώς και τα αστικά απορρίμματα. Οι πηγές ενέργειας από βιομάζα μπορεί να περιλαμβάνουν καυσόξυλα, σφαιρίδια ξύλου, καλαμπόκι, σόγια, ζωική κοπριά και άλλα (ΥΠΠΕΝ) .

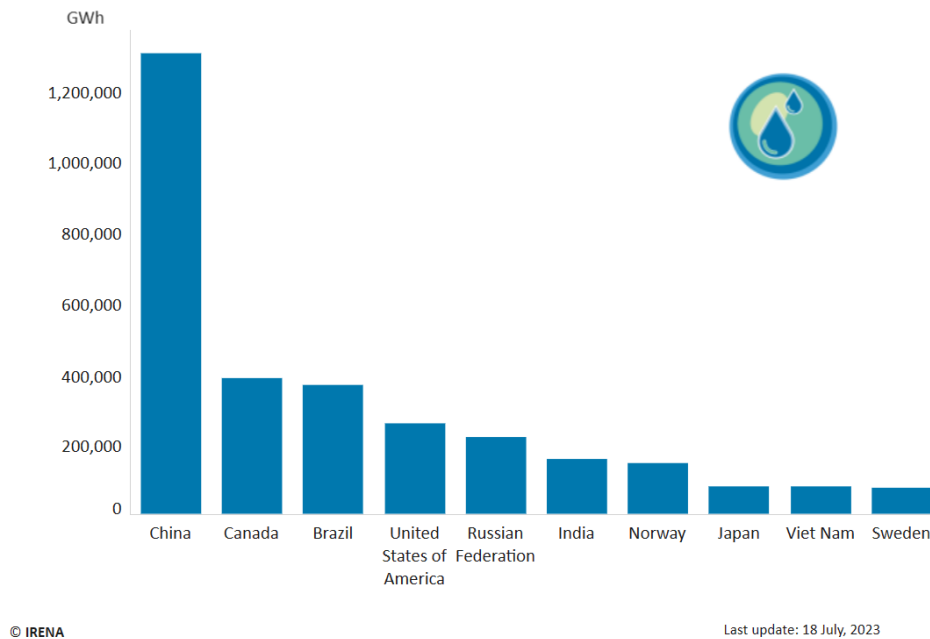
Η βιομάζα χρησιμοποιείται κυρίως για την παραγωγή θερμικής και ηλεκτρικής ενέργειας ακόμα για την παραγωγή υγρών βιοκαυσίμων. Τα βιοκαύσιμα είναι ουσίες που παράγονται από μείγμα οργανικής ύλης (βιομάζα) που μπορούν να χρησιμοποιηθούν από κινητήρες, συστήματα εσωτερικής καύσης ή να μετατραπεί σε ηλεκτρική ενέργεια. Τα πλεονεκτήματα της γρήγορης πυρόλυσης της βιομάζας, η οποία προσφέρει ένα υγρό προϊόν, το βιοέλαιο μπορεί εύκολα να αποθηκευτεί και να μεταφερθεί (Czernik, 2004). Η ενέργεια από βιομάζα είναι μια ανανεώσιμη πηγή ενέργειας που έχει χρησιμοποιηθεί ιστορικά για μαγείρεμα, θέρμανση και μικρές βιομηχανίες. Η ενέργεια από βιομάζα παρέχει πολλαπλά ενεργειακά και περιβαλλοντικά οφέλη, που κυμαίνονται από ηλεκτρική ενέργεια, θερμότητα έως καύσιμα μεταφοράς (Rosillo-calle, 2016).



Εικόνα 5: Βιομάζα ανά χώρα (irena.org, 2023).

Υδροηλεκτρική ενέργεια

Η υδροηλεκτρική ενέργεια είναι η ηλεκτρική ενέργεια που παράγεται από γεννήτριες που κινούνται από στρόβιλους που μετατρέπουν την κινητική ενέργεια από τη ροή του νερού σε μηχανική ενέργεια και μετέπειτα σε ηλεκτρική. Ήταν μία από τις πρώτες ανανεώσιμες πηγές που χρησιμοποιήθηκαν για την παραγωγή ηλεκτρικής ενέργειας ενώ στις αρχές του 21^{ου} αιώνα η υδροηλεκτρική ενέργεια ήταν η ευρέως χρησιμοποιούμενη μορφή ανανεώσιμης ενέργειας (Britannica, 2023). Είναι μια καθαρή και ανανεώσιμη πηγή ενέργειας που δεν καταναλώνει ούτε μολύνει το νερό που χρησιμοποιεί. Οι υδροηλεκτρικοί σταθμοί παρέχουν τουλάχιστον το 50% της συνολικής παροχής ηλεκτρικής ενέργειας σε περισσότερες από 35 χώρες και παράγουν περίπου το 15,8% της παγκόσμιας παραγωγής ηλεκτρικής ενέργειας (Kaygusuz 2016). Ωστόσο, η ανάπτυξη της υδροηλεκτρικής ενέργειας συνοδεύεται και από περιβαλλοντικές και κοινωνικές προκλήσεις, όπως η υποβάθμιση του περιβάλλοντος και η κλιματική αλλαγή, η οποία μπορεί να επηρεάσει αρνητικά την παραγωγή υδροηλεκτρικής ενέργειας (Kaunda 2012). Παρά τις προκλήσεις αυτές υπάρχουν ευκαιρίες για περαιτέρω παραγωγή υδροηλεκτρικής ενέργειας μέσω ανακαίνισης και βελτιώσεων της απόδοσης σε υπάρχουσες εγκαταστάσεις και συνεχούς ανάπτυξης υδροηλεκτρικών σταθμών μικρής κλίμακας που συνδέονται με το δίκτυο (Bahadori 2013).

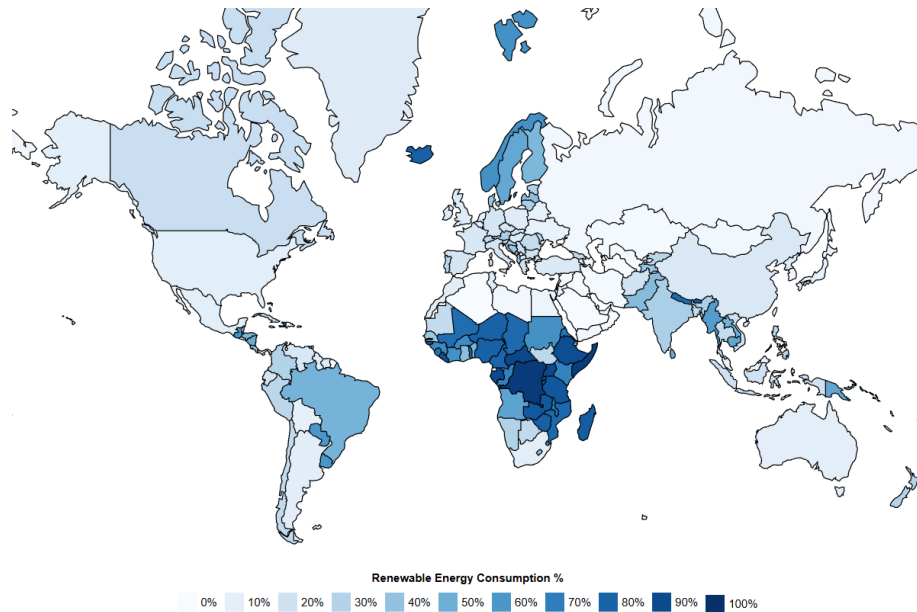


Εικόνα 6: Υδροηλεκτρική ενέργεια ανά χώρα (irena.org, 2023)

1.2.1 Ανανεώσιμες Πηγές Ενέργειας ανά χώρα

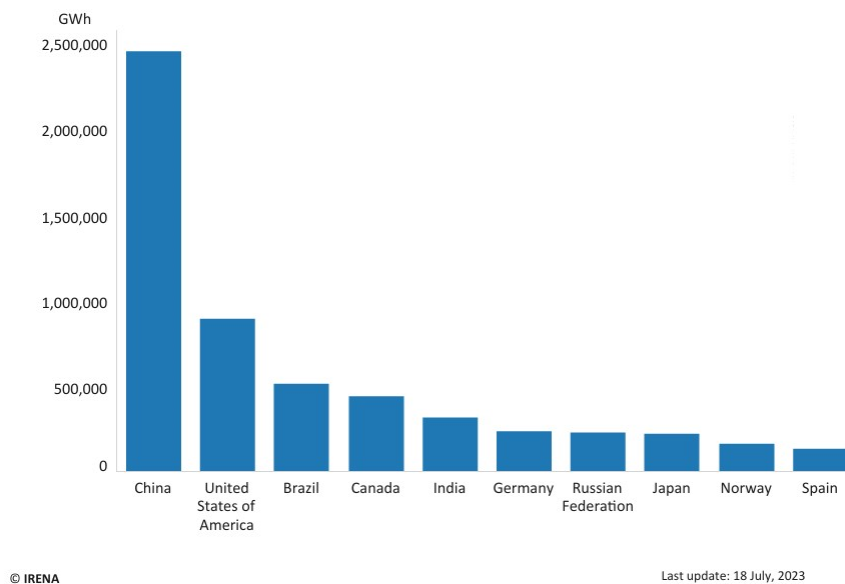
Σύμφωνα με τον Διεθνή Οργανισμό Ενέργειας (IEA), οι ανανεώσιμες πηγές ενέργειας αντιπροσώπευαν το 29% της παγκόσμιας παραγωγής ηλεκτρικής ενέργειας το 2020 (irena, 2022). Αντίστοιχα το 20% της ακαθάριστης τελικής κατανάλωσης ενέργειας της Ευρωπαϊκής Ένωσης (ourworldindata, 2020) προέρχονταν από ανανεώσιμες πηγές.

Σύμφωνα με την ‘World Population Review’, οι χώρες που με τα μεγαλύτερα ποσοστά στην κατανάλωση ενέργειας που προέρχεται από ανανεώσιμες είναι η Λ.Δ. του Κονγκό, η Σομαλία, η Κεντροαφρικανική Δημοκρατία, η Ουγκάντα, η Γκαμπόν, η Αιθιοπία, η Λιβερία, η Γουινέα-Μπισάου, η Τανζανία και το Μπουρούντι. Η Ελλάδα παράγει μόλις το 20% της ενέργειας που καταναλώνει από ανανεώσιμες πηγές ενέργειας (Renewable Energy by Country, 2023).



Εικόνα 7: Συνολική κατανάλωση ενέργειας που προέρχεται ανανεώσιμες πηγές ενέργειας ανά χώρα, 2023 (worldpopulationreview.com)

Σύμφωνα με το ‘Statista’ και το ‘IRENA’ οι χώρες με την υψηλότερη παραγωγή ηλεκτρικής ενέργειας από ανανεώσιμες πηγές είναι η Κίνα, οι ΗΠΑ και η Βραζιλία (Statista 2022, IRENA 2001).



Εικόνα 8: Παραγωγή ηλεκτρικής ενέργειας ανά Χώρα, 2023 (irena, 2023)

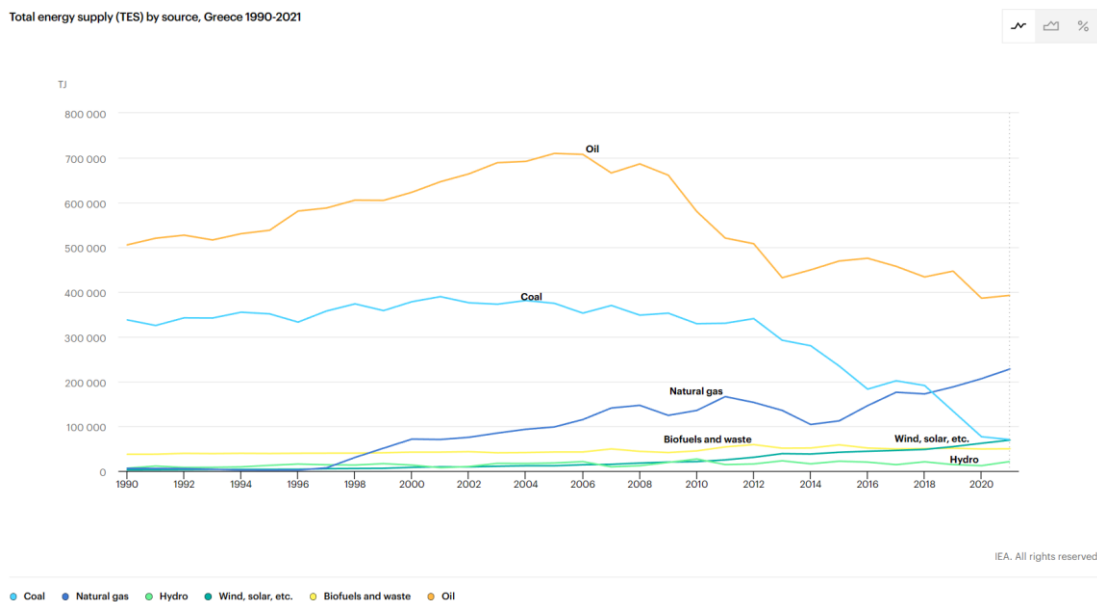
1.2.2 Ανανεώσιμες Πηγές στην Ελλάδα

Η Ελλάδα έχει σημειώσει τεράστια πρόοδο στον τομέα των ανανεώσιμων πηγών ενέργειας προκειμένου να ελαχιστοποιηθούν οι εκπομπές αερίων του θερμοκηπίου και να βελτιωθεί η ενεργειακή ασφάλεια. Για να ενθαρρύνει την ανάπτυξη των ανανεώσιμων πηγών ενέργειας, η ελληνική κυβέρνηση έχει θέσει σε εφαρμογή μια σειρά από πολιτικές και κίνητρα όπως έργα ανανεώσιμων πηγών ενέργειας, τιμολόγια τροφοδοσίας, επιδοτήσεις, φορολογικές ελαφρύνσεις κ.α.

Οι πιο κοινές ανανεώσιμες πηγές ενέργειας στην Ελλάδα είναι η αιολική και η ηλιακή ενέργεια. Το ισχυρό αιολικό δυναμικό της Ελλάδας, ειδικότερα στα νησιά και τις παράκτιες περιοχές της, καθιστά τα αιολικά πάρκα τόσο στην ξηρά όσο και στη θάλασσα, κυρίαρχα στον τομέα των ανανεώσιμων πηγών ενέργειας για τη χώρα. Επίσης, τα ηλιακά θερμικά συστήματα για τη θέρμανση του νερού και τα ηλιακά πάνελ για την παραγωγή ενέργειας χρησιμοποιούνται πλέον ευρέως στην Ελλάδα.

Επιπρόσθετα, οι μικρού και μεσαίου μεγέθους υδροηλεκτρικές μονάδες σε λειτουργία, η καύση βιομάζας και η παραγωγή βιοαερίου, καθώς και η γεωθερμική ενέργεια που χρησιμοποιείται για την παραγωγή ηλεκτρικής ενέργειας και τη θέρμανση του αέρα προσθέτουν ποσοστά στη χωρητικότητα της χώρας για ανανεώσιμες πηγές ενέργειας.

Από το 2011 έως το 2021, η συνεισφορά των ανανεώσιμων πηγών ενέργειας στη συνολική παραγωγή και κατανάλωση ενέργειας της Ελλάδας αυξήθηκε από 11% σε περίπου 20%, κυρίως λόγω της αυξανόμενης παραγωγής ηλεκτρικής ενέργειας από αιολικά και ηλιακά συστήματα ανανεώσιμων πηγών (IEA, irena). Στη συνέχεια παρουσιάζονται και οι ενεργειακές μονάδες που παράγονται με ανανεώσιμες και η πηγές στην Ελλάδα από το 1990 έως και το 2022.



Εικόνα 9: Συνολική παροχή ενέργειας (TES) ανά πηγή, Ελλάδα 1990-2021 (IEA)

Παρατηρείται πως το 2020 το μεγαλύτερο ποσοστό ενέργειας που προέρχεται από ανανεώσιμες πηγές ήταν κυρίως από αιολικά πάρκα. Το συνολικό ποσοστό παραγωγής ηλεκτρισμού από ανανεώσιμες πηγές το 2020 ήταν 36% δηλαδή 17.554 GWh ενέργειας ([irena](#)).

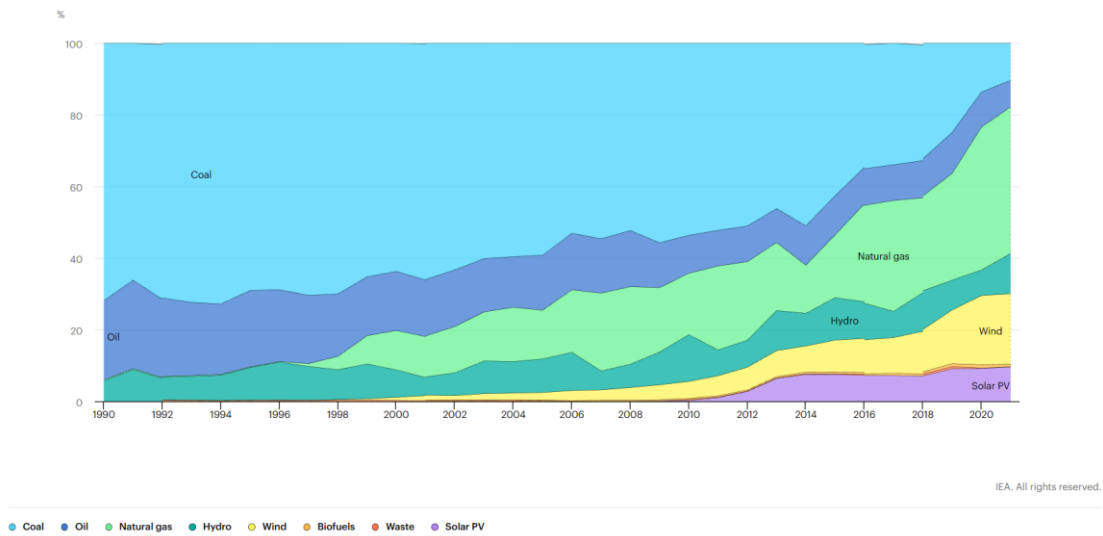
Πίνακας 2: Παραγωγή ηλεκτρικής ενέργειας Ελλάδα 2020 ([irena](#)).

| Generation in 2020 | GWh | % |
|----------------------|---------------|------------|
| Non-renewable | 30 698 | 64 |
| Renewable | 17 554 | 36 |
| Hydro and marine | 3 344 | 7 |
| Solar | 4 447 | 9 |
| Wind | 9 310 | 19 |
| Bioenergy | 454 | 1 |
| Geothermal | 0 | 0 |
| Total | 48 252 | 100 |

Το 2021, οι ανανεώσιμες πηγές ενέργειας αντιπροσώπευαν το 22% της ακαθάριστης τελικής κατανάλωσης ενέργειας στην Ελλάδα. Τα ποσοστά για την ηλεκτροπαραγωγής ήταν στο 36%, ζήτησης θέρμανσης όμοια στο 36% και της ψύξης στο 4,3% ([IEA](#), [Rae](#)).

Αντίστοιχα, το 2022, το 41% της ηλεκτρικής ενέργειας που παρήχθη στην Ελλάδα προήλθε από Ανανεώσιμες Πηγές Ενέργειας (ΑΠΕ) σύμφωνα με τον Διεθνή Οργανισμό Ενέργειας ([IEA](#)). Φαίνεται δηλαδή πως το ποσοστό της ενέργειας που προέρχεται από ανανεώσιμες πηγές συνεχώς αυξάνεται.

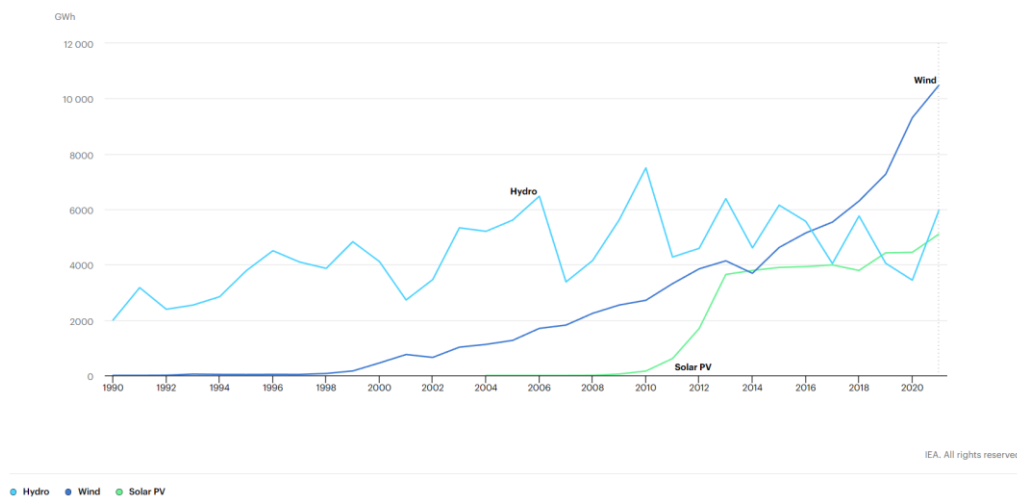
Electricity generation by source, Greece 1990-2021



Εικόνα 10: Συνολική παροχή ενέργειας (TES) ανά πηγή, Ελλάδα 1990-2021 (IEA)

Τέλος, στο επόμενο σχήμα παρατηρείται η μεγάλη άνοδος της παραγωγής ηλεκτρικής ενέργειας από φωτοβολταϊκά μετά το 2010 η οποία βέβαια μέχρι και το 2020 δεν καταφέρνει να ξεπεράσει την αιολική σε απόδοση.

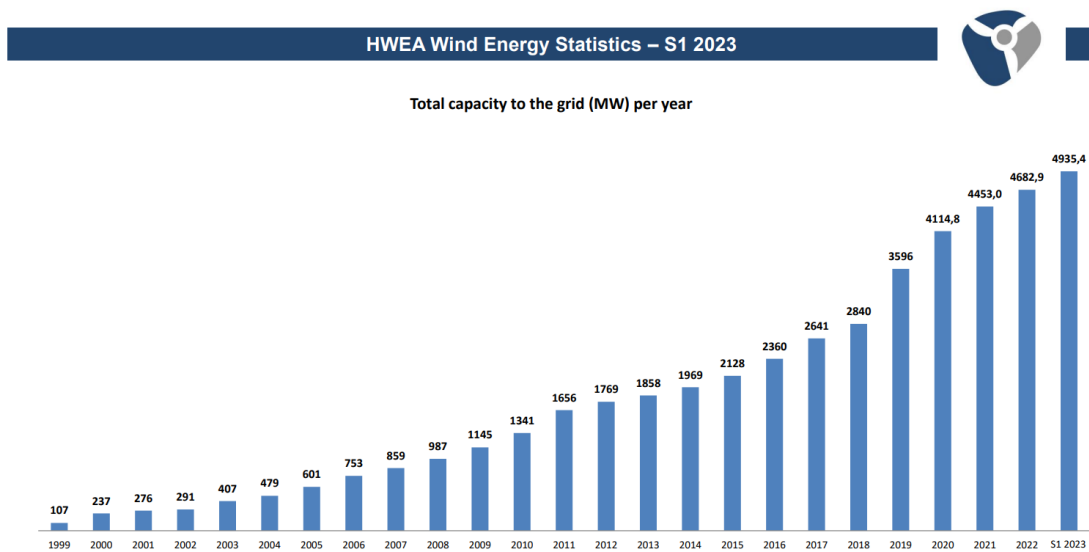
Renewable electricity generation by source (non-combustible), Greece 1990-2021



Εικόνα 11: Παραγωγή ηλεκτρικής ενέργειας από ανανεώσιμες πηγές, Ελλάδα 1990-2021 (IEA).

Σύμφωνα με την Ελληνική Επιστημονική Ένωση Αιολικής Ενέργειας υπάρχει ραγδαία αύξηση της συνολικής αιολικής ενέργειας από το 1999 έως και το 2023. Το 2022, 21,5% της συνολικής ηλεκτροπαραγωγής προήλθε από αιολικά πάρκα, δηλαδή 10,7 TWh ηλεκτρική ενέργεια. Ενώ για το 2023 τα στοιχεία δείχνουν πως η ημερήσια παραγωγή

αιολικής ενέργειας κυμαίνεται στις 20.6GWh με το 19.5% του ηλεκτρισμού να προέρχεται από αιολική ενέργεια (ΕΛΕΤΑΕΝ).



Εικόνα 12: Συνολική αιολική ισχύς (MW) ανά έτος (ΕΛΕΤΑΕΝ)

1.3 Υλοποίηση Κώδικα

Η ανάλυση των δεδομένων που λήφθηκαν από το ευρωπαϊκό πρόγραμμα Copernicus πραγματοποιήθηκε στο Jupyter Notebook με τη γλώσσα προγραμματισμού Python.

1.3.1 Python

Η Python είναι μια υψηλού επιπέδου γλώσσα προγραμματισμού που χρησιμοποιείται ευρέως λόγω της ευελιξίας, της απλότητας της και των πολλαπλών βιβλιοθηκών της. Δημιουργήθηκε από τον Guido van Rossum και κυκλοφόρησε για πρώτη φορά το 1991 ενώ παραμένει γλώσσα ανοιχτού κώδικα με μεγάλη κοινότητα προγραμματιστών. Επιπρόσθετα, είναι διαθέσιμη σε διάφορες πλατφόρμες όπως Windows, macOS και διαφορετικών διανομών Linux και αυτό την καθιστά ευέλικτη. Αυτή τη στιγμή η προτεινόμενη και νεότερη έκδοση της είναι η Python 3.

Η python είναι μια γλώσσα προγραμματισμού γενικής χρήσης, δηλαδή μπορεί να χρησιμοποιηθεί σε ένα ευρύ φάσμα εφαρμογών, συμπεριλαμβανομένης της ανάπτυξης ιστού, των επιστημονικών υπολογιστών, της ανάλυσης δεδομένων, της τεχνητής νοημοσύνης, του αυτοματισμού και πολλά άλλα. Η σύνταξη της έχει σχεδιαστεί ώστε να παρέχει εύκολη ανάγνωση και εγγραφή, και αναγνωσιμότητα του κώδικα. Είναι

ερμηνευμένη γλώσσα, δηλαδή δεν χρειάζεται μεταγλώττιση του κώδικα πριν την εκτέλεση του.

Τέλος, διαθέτει μια τεράστια τυπική βιβλιοθήκη που παρέχει πακέτα για μια μεγάλη ποικιλία διεργασιών καθώς και βιβλιοθήκες τρίτων που επεκτείνουν τις δυνατότητές της όπως για παράδειγμα, το NumPy και το Panda που χρησιμοποιούνται για χειρισμό και ανάλυση δεδομένων, το Django και το Flask για ανάπτυξη ιστού, το TensorFlow και το PyTorch για μηχανική εκμάθηση και πολλά άλλα. Τέλος, κατέχει ισχυρή ενεργή κοινότητα προγραμματιστών, η οποία συμβάλλει στην ανάπτυξη και τη βελτίωσή της.

1.3.2 Jupyter Notebook

Το Jupyter Notebook είναι ένα υπολογιστικό περιβάλλον ανοιχτού κώδικα βασισμένο στον ιστό που επιτρέπει τη δημιουργία και το διαμοιρασμό εγγράφων κώδικα. Χρησιμοποιείται ευρέως στην ανάλυση δεδομένων, την έρευνα και τη διδασκαλία καθώς παρέχει εύκολη και αποδοτική απεικόνιση του κώδικα. Το Jupyter Notebook υποστηρίζει διάφορες γλώσσες προγραμματισμού μεταξύ αυτών και την Python που είναι μια από τις πιο δημοφιλείς επιλογές.

1.3.3 Μηχανική Μάθηση

Η μηχανική μάθηση είναι μια τεχνική ανάλυσης δεδομένων που βοηθάει τον κώδικα να «μαθαίνει» μέσω των δεδομένων του προβλήματος που καλείται να βρει λύση. Με άλλα λόγια, η μηχανική μάθηση είναι η διαδικασία εκπαίδευσης ενός συστήματος υπολογιστή ώστε να κάνει προβλέψεις ή αποφάσεις με βάση ορισμένα δεδομένα εισόδου. Το σύστημα μαθαίνει από τα δεδομένα και βελτιώνει την ακρίβειά του με την πάροδο του χρόνου (Bisong, 2019). Βρίσκεται στον πυρήνα της τεχνητής νοημοσύνης και της επιστήμης δεδομένων και έχει υιοθετηθεί σε διάφορους τομείς, όπως η υγειονομική περίθαλψη, η κατασκευή, η εκπαίδευση, η οικονομική μοντελοποίηση, η αστυνόμευση και το μάρκετινγκ (Jordan, 2015). Οι δυνατότητες της μηχανικής μάθησης για τη βελτιστοποίηση και την αυτοματοποίηση πολύπλοκων διαδικασιών, όπως η διασφάλιση ποιότητας φυσικής ακτινοβολίας, ο σχεδιασμός περιγράμματος και θεραπείας, η ακτινοθεραπεία με καθοδήγηση εικόνας, η διαχείριση αναπνευστικής κίνησης, η μοντελοποίηση απόκρισης στη θεραπεία και η πρόβλεψη αποτελεσμάτων συνεχώς εξελίσσεται (Naqa 2015).

Υπάρχουν πολλές προσεγγίσεις μηχανικής εκμάθησης, συμπεριλαμβανομένης της εποπτευόμενης, χωρίς επίβλεψη και της ενισχυτικής μάθησης, καθώς και ένα πλήθος γλωσσών προγραμματισμού που χρησιμοποιούνται για αυτή (Shaveta 2023). Η κύρια διαφορά μεταξύ εποπτευόμενης και μη εποπτευόμενης μάθησης είναι ότι η εποπτευόμενη μάθηση είναι ένας τύπος μηχανικής μάθησης όπου ο αλγόριθμος μαθαίνει να προβλέπει μεταβλητές εξόδου από μεταβλητές εισόδου με βάση παραδείγματα με ετικέτα, ενώ η μη εποπτευόμενη μάθηση είναι ένας τύπος μηχανικής μάθησης όπου ο αλγόριθμος μαθαίνει να αναγνωρίζει μοτίβα σε δεδομένα χωρίς να του δίνονται ρητές ετικέτες εξόδου (Love, 2002).

Εποπτευόμενων αλγορίθμων μηχανικής εκμάθησης

Η εποπτευόμενη μηχανική μάθηση είναι μια υποκατηγορία της μηχανικής μάθησης και της τεχνητής νοημοσύνης. Ορίζεται από τη χρήση δεδομένων που ορίζονται για την εκπαίδευση των αλγορίθμων και ταξινομούν δεδομένα ή προβλέπουν τα αποτελέσματα με ακρίβεια. Καθώς τα δεδομένα εισόδου τροφοδοτούνται στο μοντέλο, προσαρμόζει τα βάρη του έως ότου το μοντέλο τοποθετηθεί κατάλληλα, κάτι που συμβαίνει ως μέρος της διαδικασίας διασταυρούμενης επικύρωσης. Η εποπτευόμενη μάθηση βοηθά τους στην επίλυση μιας ποικιλίας πραγματικών προβλημάτων όπως η ταξινόμηση των ανεπιθύμητων μηνυμάτων σε ξεχωριστό φάκελο από τα εισερχόμενά (Gahlan, 2020). Μερικά δημοφιλή παραδείγματα εποπτευόμενων αλγορίθμων μηχανικής εκμάθησης είναι:

- Linear regression
- Random forest
- Support vector machines (SVM)

Μη εποπτευόμενη μηχανική εκμάθηση

Η μη εποπτευόμενη μηχανική μάθηση είναι ένας τύπος μηχανικής μάθησης που χρησιμοποιεί αλγόριθμους μηχανικής μάθησης για την ανάλυση και τη ομαδοποίηση συνόλων δεδομένων χωρίς ομαδοποίηση. Αυτοί οι αλγόριθμοι ανακαλύπτουν κρυφά μοτίβα ή ομαδοποιήσεις δεδομένων χωρίς την ανάγκη ανθρώπινης παρέμβασης. Τα μοντέλα μάθησης χωρίς επίβλεψη χρησιμοποιούνται για τρεις κύριες εργασίες:

- Ομαδοποίηση

- συσχέτιση
- μείωση διαστάσεων συνόλου με βάση τις ομοιότητες ή τις διαφορές τους.

Η μη εποπτευόμενη μηχανική εκμάθηση χρησιμοποιείται κυρίως για να ομαδοποιήσει τα χαρακτηριστικά που φαίνονται να είναι παρόμοια μεταξύ τους (Singh, 2019). Η μάθηση χωρίς επίβλεψη βρίσκει πολλές πραγματικές εφαρμογές όπως την εξερεύνηση δεδομένων, την τμηματοποίηση πελατών, των συστημάτων συστάσεων, των εκστρατειών μάρκετινγκ στόχων και της προετοιμασίας και οπτικοποίησης δεδομένων. Μερικά παραδείγματα είναι:

- k-means clustering
- k-nearest neighbors (KNN)
- hierarchical clustering
- Density-Based Spatial Clustering of Applications with Noise (DBSCAN)
- Gaussian Mixtures Model (GMM)

Η μερικώς εποπτευόμενη μάθηση

Η μερικώς εποπτευόμενη μάθηση είναι ένας τύπος μηχανικής μάθησης που συνδυάζει μια μικρή ποσότητα δεδομένων με ετικέτες (labels) και μια μεγάλη ποσότητα δεδομένων χωρίς ετικέτες κατά τη διάρκεια της «εκπαίδευσης» του αλγορίθμου. Αυτή η μέθοδος χρησιμοποιείται όταν δεν υπάρχουν αρκετά επισημασμένα δεδομένα διαθέσιμα για εποπτευόμενη μάθηση, αλλά υπάρχουν αρκετά διαθέσιμα δεδομένα για τον εντοπισμό μοτίβων στα δεδομένα (Zhou, 2013).

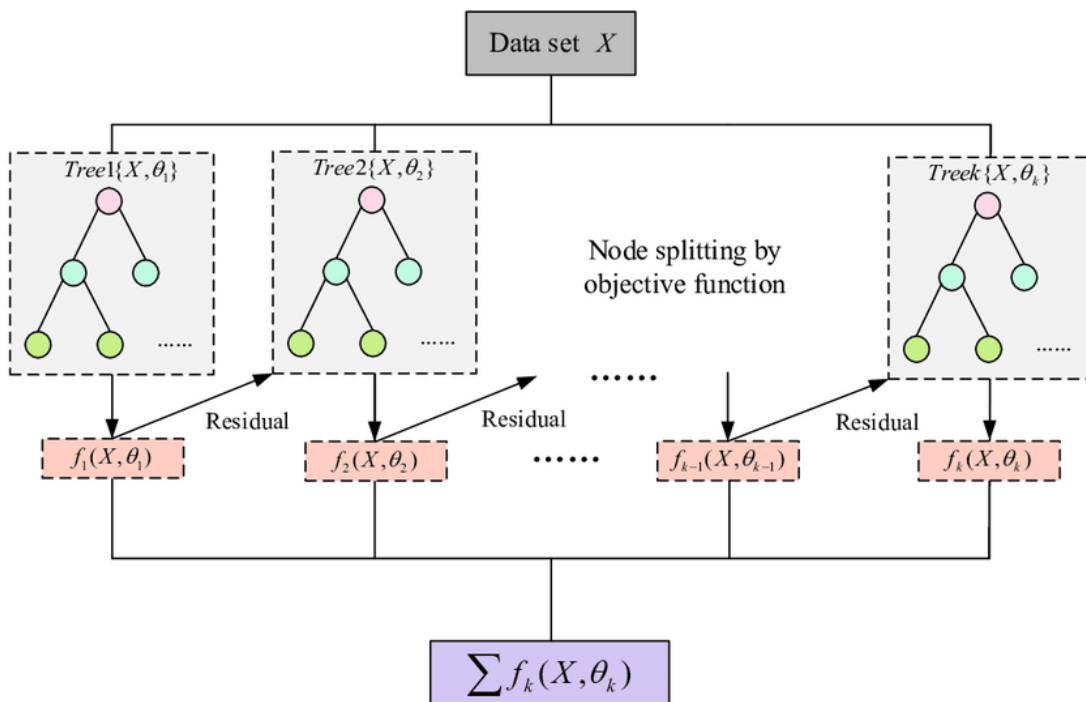
1.3.4 Αλγόριθμος XGBoost

Ο αλγόριθμος «Gradient Boosting» υπάρχει από 1999 και είναι μια μέθοδος εκμάθησης συνόλων κατά την οποία βελτιστοποιείται η απόδοση και γίνεται πιο αποτελεσματικό κάθε φορά το μοντέλο. Με τη μέθοδο Boosting επιτυγχάνεται το γεγονός πως παρατηρήσεις που ταξινομήθηκαν εσφαλμένα σε προηγούμενο μοντέλο δεν αυξάνουν την πιθανότητα να επιλεχθούν ξανά και ελαχιστοποιεί το συνολικό σφάλμα της πρόβλεψης (Natekin & Knoll, 2013).

Ο XGBoost (eXtreme Gradient Boosting) είναι αλγόριθμος εποπτευόμενης μηχανικής μάθησης που αναπτύχθηκε από τον Tianqi Chen και τους συνεργάτες του και είναι μια

βελτιστοποιημένη έκδοση του Gradient Boosting η οποία εστιάζει στην υπολογιστική ταχύτητα και την αποτελεσματικότητα του μοντέλου (Chen, 2016). Χρησιμοποιεί δέντρα αποφάσεων που ενισχύουν τη διαβάθμιση και τη κλίση με σκοπό τη διαχείριση μεγάλων συνόλων δεδομένων. Χαρακτηρίζεται από γρήγορο υπολογισμό και καλή απόδοση του μοντέλου, είναι επεκτάσιμος και μπορεί να χειριστεί δισεκατομμύρια δεδομένα χρησιμοποιώντας πολύ λίγους υπολογιστικούς πόρους (Chen, 2015; Chen, 2016).

Ο XGBoost αυξάνεται σε δημοτικότητα και χρησιμοποιείται από πολλούς επιστήμονες δεδομένων παγκοσμίως για την επίλυση προβλημάτων παλινδρόμησης, ταξινόμησης, κατάταξης και προκλήσεων πρόβλεψης που καθορίζονται από τον χρήστη ([educba](#)). Ο κώδικας του αλγορίθμου υπάρχει διαθέσιμος στην ιστοσελίδα [GitHub](#).



Εικόνα 13: Διάγραμμα ροής του XGBoost (Guo et.al. 2020)

2. Μελέτη Δεδομένων Μοντέλου Πρόβλεψης

2.1 Ανάλυση δεδομένων και μεταβλητών

Τα δεδομένα που παρέχονται από το πρόγραμμα Copernicus περιέχουν κλιματικούς και ενεργειακούς δείκτες. Οι δείκτες που σχετίζονται με το κλίμα είναι η θερμοκρασία του αέρα, οι βροχοπτώσεις, η εισερχόμενη ηλιακή ακτινοβολία, η ταχύτητα του ανέμου στα 10 m και στα 100 m και η μέση πίεση του αέρα στο επίπεδο της θάλασσας. Οι ενεργειακοί δείκτες είναι η ζήτηση ηλεκτρικής ενέργειας και η παραγωγή ηλεκτρικής ενέργειας από διάφορες πηγές όπως αιολική (τόσο χερσαία όσο και υπεράκτια), ηλιακή και υδροηλεκτρική ενέργεια από ποτάμια και λίμνες.

Ανάλογα με τον δείκτη τα δεδομένα είναι διαθέσιμα σε εθνικό ή περιφερειακό επίπεδο για τις περισσότερες ευρωπαϊκές χώρες. Οι μεταβλητές παρουσιάζονται αναλυτικά παρακάτω.

2.1.1 Θερμοκρασία αέρα

Η θερμοκρασία του αέρα περιβάλλοντος παρέχεται κοντά στην επιφάνεια, συνήθως σε ύψος δυο μέτρα (2m). Οι τιμές δεδομένων είναι στιγμιαίες μετρήσεις σε μονάδες Kelvin (K).

2.1.2 Πίεση στο επίπεδο της θάλασσας

Τα δεδομένα αντιπροσωπεύουν τον μέσο όρο σε μονάδες Πασκάλ (Pa) σε επίπεδο χώρας και υποχώρας. Οι τιμές των δεδομένων είναι στιγμιαίες μετρήσεις.

2.1.3 Επιφανειακή ακτινοβολία

Η ποσότητα της ηλιακής ακτινοβολίας σε μονάδες W/m^2 που φτάνει σε ένα οριζόντιο επίπεδο στην επιφάνεια της Γης. Αυτή η παράμετρος περιλαμβάνει τόσο την άμεση όσο και τη διάχυτη ηλιακή ακτινοβολία. Οι τιμές προκύπτουν από την επιφανειακή ακτινοβολία καθώς και τη διόρθωση τους με τη χρήση νεφοκάλυψης της Μονάδας Κλιματικής Έρευνας.

2.1.4 Συνολική βροχόπτωση

Είναι το βάθος βρόχινου νερού που συσσωρεύεται σε επίπεδη, οριζόντια και αδιαπέραστη επιφάνεια ανά μονάδα επιφάνειας κατά τη διάρκεια δεδομένης χρονικής περιόδου. Τα δεδομένα αντιπροσωπεύουν τον μέσο όρο έκτασης σε επίπεδο χώρας και υποχώρας. Οι τιμές δεδομένων είναι συσσωρευμένες μετρήσεις σε μονάδες μέτρων (m).

2.1.5 Ταχύτητα ανέμου

Είναι το μέγεθος της δυσδιάστατης οριζόντιας ταχύτητας αέρα σε ύψη 10 μέτρων και 100 μέτρων. Τα δεδομένα αντιπροσωπεύουν τον μέσο όρο σε επίπεδο χώρας, υποχώρας και θαλάσσιες περιφέρειες. Οι τιμές δεδομένων είναι στιγμιαίες μετρήσεις σε μονάδες μέτρα ανά δευτερόλεπτο (m/s).

2.1.6 Παραγωγή ηλιακής ενέργειας

Παραγωγή ηλιακής ενέργειας εκφραζόμενη ως συντελεστής δυναμικότητας (λόγος πραγματικής παραγωγής προς εγκατεστημένη ισχύ), ενέργειας (MWh ή GWh) και μέσης ισχύς (MW ή GW) δηλαδή ενέργεια στη μονάδα του χρόνου. Ο μέσος όρος των δεδομένων δίνεται σε επίπεδο χώρας και υποχώρας με τις τιμές των δεδομένων να είναι οι στιγμιαίες μετρήσεις.

2.1.7 Παραγωγή αιολικής ενέργειας

Η υπεράκτια παραγωγή αιολικής ενέργειας εκφράζεται ως συντελεστής δυναμικότητας (λόγος πραγματικής παραγωγής προς εγκατεστημένη ισχύ), ενέργειας (MWh ή GWh) και σε μέση ισχύς (MW ή GW) δηλαδή ενέργεια ανά μονάδα χρόνου για μη χερσαίες περιοχές. Οι τιμές δεδομένων είναι στιγμιαίες μετρήσεις σε μονάδες

Η χερσαία παραγωγή αιολικής ενέργειας εκφράζεται ως συντελεστής δυναμικότητας (λόγος πραγματικής παραγωγής προς εγκατεστημένη ισχύ), ενέργειας (MWh ή GWh) και σε μέση ισχύς (MW ή GW) δηλαδή ενέργεια ανά μονάδα χρόνου για χερσαίες περιοχές. Οι τιμές δεδομένων είναι στιγμιαίες μετρήσεις.

2.1.8 Παραγωγή υδροηλεκτρικής ενέργειας

Η παραγωγή υδροηλεκτρικής ενέργειας από τεχνητές λίμνες παρέχεται εκφραζόμενη ως συντελεστής δυναμικότητας (λόγος πραγματικής παραγωγής προς εγκατεστημένη

ισχύ), ως ενέργεια σε μονάδες ενέργειας (MWh ή GWh) είτε ως μέση ισχύς (MW ή GW) δηλαδή ενέργεια ανά μονάδα χρόνου, και τα στοιχεία δίνονται σε επίπεδο χώρας για τις χώρες όπου υπάρχει παραγωγή.

Όμοια η υδροηλεκτρική παραγωγή από ποτάμια παρέχεται και αυτή εκφραζόμενη ως συντελεστής δυναμικότητας (λόγος πραγματικής παραγωγής προς εγκατεστημένη δυναμικότητα), σε ενεργειακές μονάδες (MWh ή GWh) και σε μέση ισχύς (MW ή GW) δηλαδή ενέργεια ανά μονάδα χρόνου σε επίπεδο χώρας για τις χώρες όπου υπάρχει παραγωγή.

2.1.9 Ζήτηση ηλεκτρικής ενέργειας

Η ζήτηση ηλεκτρικής ενέργειας είναι η κατανάλωση ηλεκτρικής ενέργειας εκφρασμένη σε μονάδες ενέργειας (MWh ή GWh) ή ως μέση ισχύς (MW ή GW) δηλαδή ενέργεια ανά μονάδα χρόνου, και τα στοιχεία παρέχονται σε επίπεδο χώρας.

2.2 Επεξεργασία Δεδομένων

2.2.1 Εισαγωγή Δεδομένων

Τα δεδομένα παρέχονται σε ξεχωριστά αρχεία (.csv) ανά κλιματική και ενεργειακή μεταβλητή για όλες τις χώρες. Αρχικά, αναπτύχθηκε ένας αλγόριθμος στον οποίο επιλέγοντας τη χώρα ή την περιοχή επιθυμίας δεσμεύει τα δεδομένα από όλα τα αρχεία, επτά στο σύνολο, και τα αποθηκεύει σε ξεχωριστό αρχείο (.csv) με το όνομα της χώρας και τα στοιχεία των κλιματικών και ενεργειακών μεταβλητών ανά ημερομηνία ή χρονικό διάστημα.

Για την παρούσα εργασία επιλέχθηκε η Ελλάδα ('EL'). Στη συνέχεια με τη βοήθεια της εντολής επανάληψης λαμβάνονται από όλα τα ξεχωριστά αρχεία δεδομένων που λήφθηκαν από το Copernicus API μόνο τα στοιχεία για τη χώρα ενδιαφέρονται και αντιγράφονται σε νέο αρχείο ίδιου τύπου (.csv - 'Comma Separated Values').

Παρουσιάζεται το κομμάτι του κώδικα:

```
cwd = os.path.abspath(r'C:\Forecasting Solar PV Power Generation\data')
files = os.listdir(cwd)

data = pd.DataFrame()
data['Date'] = ""

for file in files:
    df_temp = pd.DataFrame()
    df = pd.DataFrame()
    y = file

    if file.endswith('.csv'):
        df_temp = df_temp.append(pd.read_csv(file), ignore_index=True)
        df[["Date", y]] = df_temp[["Date", name]]
        data = data.merge(df, on = 'Date', how = 'outer', suffixes=('data', 'df'))
```

Εικόνα 14: Κώδικας για την εισαγωγή δεδομένων.

Έπειτα το νέο αρχείο επεξεργάζεται ώστε η ημερομηνία να έχει τη σωστή για την ανάλυση δομή και ελέγχεται για τυχόν ανωμαλίες. Επιπρόσθετα, πραγματοποιείται μετονομασία των στηλών των δεδομένων με τα κατάλληλα ονόματα και τις μονάδες μέτρησης των μεταβλητών ώστε να υπάρχει ίδια βάση σύγκρισης. Μετά την επεξεργασία το τελικό αρχείο είναι έτοιμο για την στατιστική ανάλυση των δεδομένων και την ανάπτυξη του μοντέλου. Οι ακαθόριστες τιμές (“NaN”) που υπάρχουν στα δεδομένα λήφθηκαν κατά τις στιγμές που δεν υπήρχε ηλιακή ενέργεια και ακτινοβολία, π.χ. βραδινές ώρες. Το τελικό αποτέλεσμα του πίνακα δεδομένων φαίνεται παρακάτω.

Πίνακας 3: Παρουσίαση δεδομένων προβλήματος.

| | Date | Solar PV Power | Air Temperature | Global Horizontal Irradiance | Mean Sea Level Pressure | Total Precipitation | Wind Speed 10m | Wind Speed 100m |
|--------|---------------------|----------------|-----------------|------------------------------|-------------------------|---------------------|----------------|-----------------|
| 0 | 1979-01-01 00:00:00 | NaN | 284.90286 | NaN | 100365.470 | NaN | 3.524685 | 6.391961 |
| 1 | 1979-01-01 01:00:00 | NaN | 284.95416 | NaN | 100294.310 | NaN | 3.385256 | 6.170972 |
| 2 | 1979-01-01 02:00:00 | NaN | 285.61070 | NaN | 100214.970 | NaN | 3.327355 | 6.026396 |
| 3 | 1979-01-01 03:00:00 | NaN | 285.44745 | NaN | 100149.040 | NaN | 3.410202 | 6.118879 |
| 4 | 1979-01-01 04:00:00 | NaN | 285.35626 | NaN | 100078.020 | NaN | 3.421252 | 6.124903 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 413008 | 2023-07-31 18:00:00 | 11.919418 | 299.32834 | 1.548849e+01 | 100839.734 | 3.029865e-07 | 2.036868 | 3.357661 |
| 413009 | 2023-07-31 19:00:00 | NaN | 298.88220 | 6.467518e-14 | 100940.990 | 4.022958e-07 | 1.722882 | 3.008210 |
| 413010 | 2023-07-31 20:00:00 | NaN | 297.24496 | 6.467518e-14 | 100986.950 | 1.343638e-07 | 1.588435 | 2.707174 |
| 413011 | 2023-07-31 21:00:00 | NaN | 296.87473 | 6.467518e-14 | 101005.640 | 6.926348e-06 | 1.507571 | 2.452604 |
| 413012 | 2023-07-31 22:00:00 | NaN | 295.26300 | 6.467518e-14 | 101013.400 | 6.448036e-06 | 1.426133 | 2.342447 |

413013 rows × 8 columns

Σε δεύτερη φάση, το νέο αρχείο εισάγεται σε νέο κομμάτι κώδικα με σκοπό την ανάλυση και επεξεργασία του. Ο αλγόριθμος μπορεί να παρέχει άμεσα πληροφορίες για τις κλιματικές και ενεργειακές μεταβλητές κάθε διαθέσιμης χώρας ή περιοχής με την

κατάλληλη εντολή εισόδου (συντομογραφία ονόματος χώρας) με τις διαθέσιμες για τη χώρα αυτή πληροφορίες από το δεδομένα του Copernicus. Για την Ελλάδα τα δεδομένα που παρέχονται φαίνονται στον παραπάνω και προέρχονται από το εννοποιημένο αρχείο που δημιουργήθηκε στην προηγούμενη φάση.

Import Data

```
#Select Country
# AL,AT,BA,BE,BG,CH,CY,CZ,DE,DK,EE,EL,ES,FI,FR,HR,HU,IE,IS,IT,LI,LT,LU,LV,ME,MK,MT,NL,NO,PL,PT,RO,RS,SE,SI,SK,TR,UK
name = 'EL'

data = pd.read_csv(name + '.csv')

data_safe = data.copy()
data.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 413013 entries, 0 to 413012
Data columns (total 8 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Date                   413013 non-null object
1   Solar PV Power         227570 non-null float64
2   Air Temperature        413013 non-null float64
3   Global Horizontal Irradiance  413006 non-null float64
4   Mean Sea Level Pressure  413011 non-null float64
5   Total Precipitation     413006 non-null float64
6   Wind Speed 10m         413011 non-null float64
7   Wind Speed 100m       413011 non-null float64
dtypes: float64(7), object(1)
memory usage: 25.2+ MB
```

Εικόνα 15: Εισαγωγή δεδομένων για τη χώρα προτίμησης.

2.3 Στατιστική Ανάλυση Δεδομένων

Αρχικά θα πραγματοποιηθεί έλεγχος ώστε να διαπιστωθεί η κατανομή που ακολουθούν τα δεδομένα. Όπως φαίνεται τα δεδομένα του προβλήματος δεν ακολουθούν κανονική κατανομή με βάση το “Kolmogorov-Smirnov ktest”.

Kolmogorov-Smirnov k test

```
#Kolmogorov-Smirnov k test

def k_test(data):

    k_test_list = list(kstest(data, 'norm'))

    pval = k_test_list[1]
    alpha = 0.05

    print("H0: Data was drawn from a Normal Distribution")
    if (pval<alpha):
        print("pval {} is lower than significance level: {}, therefore null hypothesis is rejected".format(pval, alpha))
    else:
        print("pval {} is higher than significance level: {}, therefore null hypothesis cannot be rejected".format(pval, alpha))

print('Kolmogorov-Smirnov test for Solar PV Power')
print('')
k_test(SP)

print('\n',kstest(SP, 'norm'))
```

```
Kolmogorov-Smirnov test for Solar PV Power

H0: Data was drawn from a Normal Distribution
pval 0.0 is lower than significance level: 0.05, therefore null hypothesis is rejected

KstestResult(statistic=0.5144313110293043, pvalue=0.0, statistic_location=3.0477145692, statistic_sign=-1)
```

Εικόνα 16: Κώδικας και αποτελέσματα Kolmogorov-Smirnov ktest.

Ο μέσος όρος των δεδομένων για την παροχή ηλιακής ενέργεια μέσω φωτοβολταϊκών είναι ανά μονάδα χρόνου είναι 400,67 (MW) με τυπική απόκλιση 552,53 (MW) και βαθμό συγκέντρωσης γύρω από αυτή (κύρτωση) να είναι 2,71. Η διάμεσος παρουσιάζεται στα 10.64 (MW) ενώ η ασυμμετρία του δείγματος στο 1,09.

```
mean = round(np.mean(SP.values),2)
std = round(np.std(SP.values),2)
median = round(np.median(SP.values),2)
skew = round(skew(SP.values),2)
ex_kurt = round(kurtosis(SP),2)

print('\nΜέσος Όρος:', mean, '(MW)', '\nΤυπική απόκλιση:', std, '(MW)', '\nΔιάμεσος:', median, '(MW)')
print("Ασυμμετρία: {} \nΚύρτωση :{}".format(skew, ex_kurt+3))
```

Μέσος Όρος: 400.67 (MW)
 Τυπική απόκλιση: 552.53 (MW)
 Διάμεσος: 10.64 (MW)
 Ασυμμετρία: 1.09
 Κύρτωση :2.71

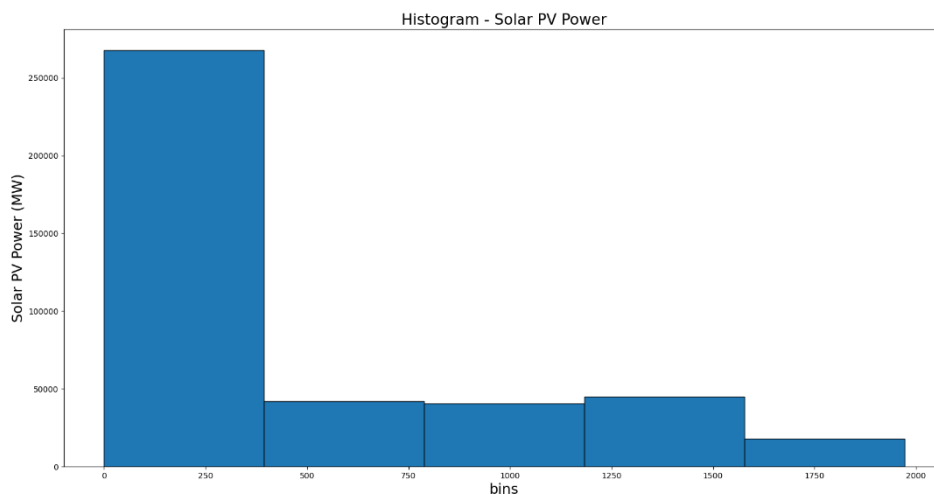
Εικόνα 17: Κώδικας και αποτελέσματα στατιστικής ανάλυσης.

Έπειτα παρουσιάζεται το ιστόγραμμα στο οποίο φαίνεται γραφικά η κατανομή του δείγματος.

Histogram

```
# histogram
fig1=plt.figure(figsize=(20,10))
plt.hist(SP, edgecolor='black', bins=5)
plt.title('Histogram - Solar PV Power',fontsize=19)
plt.xlabel('bins',fontsize=18)
plt.ylabel('Solar PV Power (MW)',fontsize=18)
fig1.savefig('histogram Solar PV Power.png')
```

Εικόνα 18: Κώδικας υλοποίησης ιστογράμματος.



Σχήμα 1: Ιστόγραμμα για την Ηλιακή Ενέργεια από φωτοβολταϊκά (MW).

Στη συνέχεια κατασκευάστηκαν και χρονικές μεταβλητές για την παροχή και αξιοποίηση των δεδομένων και την κατανομή του δείγματα ανά ώρα, μέρα, βδομάδα, μήνα και χρόνο ώστε να κατασκευαστούν τα αντίστοιχα γραφήματα.

```
data['Date'] = pd.to_datetime(data['Date'])
data_time = data.copy()

#hour, day, month, year

data_time['Date'] = pd.to_datetime(data_time['Date'])
data_time['hour'] = data_time['Date'].dt.hour
data_time['day'] = data_time['Date'].dt.day
data_time['weekday'] = data_time['Date'].dt.weekday
data_time['month'] = data_time['Date'].dt.month
data_time['year'] = data_time['Date'].dt.year

data_time.info()
```

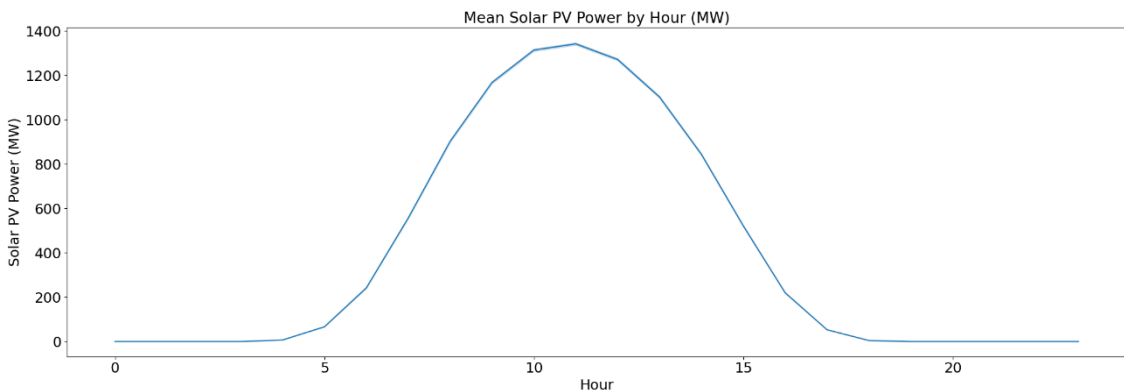
Εικόνα 19: Κώδικας για την κατασκευή των χρονικών μεταβλητών.

Η παροχή ηλιακής ενέργειας από τα φωτοβολταϊκά ανά ώρα παρουσιάζει ένα μέγιστο στις πρωινές προς μεσημβρινές ώρες. Το διάγραμμα δείχνει τη μέση τιμή των δεδομένων από ώρα από το 1979 έως τον Ιούλιο του 2023.

Hourly

```
fig=plt.figure(figsize=(20,7))
sns.lineplot(data=data_time,x='hour',y='Solar PV Power')
plt.xticks(fontsize=18)
plt.yticks(fontsize=18)
plt.title('Mean Solar PV Power by Hour (MW)',fontsize=19)
plt.xlabel('Hour',fontsize=18)
plt.ylabel('Solar PV Power (MW)',fontsize=18)
plt.tight_layout()
plt.savefig('Mean Solar PV Power by Hour (MW).png')
plt.show()
```

Εικόνα 20: Κώδικας για την κατασκευή του διαγράμματος της μέσης ηλιακής ενέργειας ανά ώρα.



Σχήμα 2: Μέση ηλιακή ενέργεια ανά ώρα.

Επιπρόσθετα, παρουσιάζεται και το θηκόγραμμα (box plot) δηλαδή η οπτική αναπαράσταση που δείχνει την κατανομή, την κεντρική τάση και την εξάπλωση του συνόλου δεδομένων. Ανάλογα την εποχή και το μήνα παρατηρείται πως η μέση ηλιακή

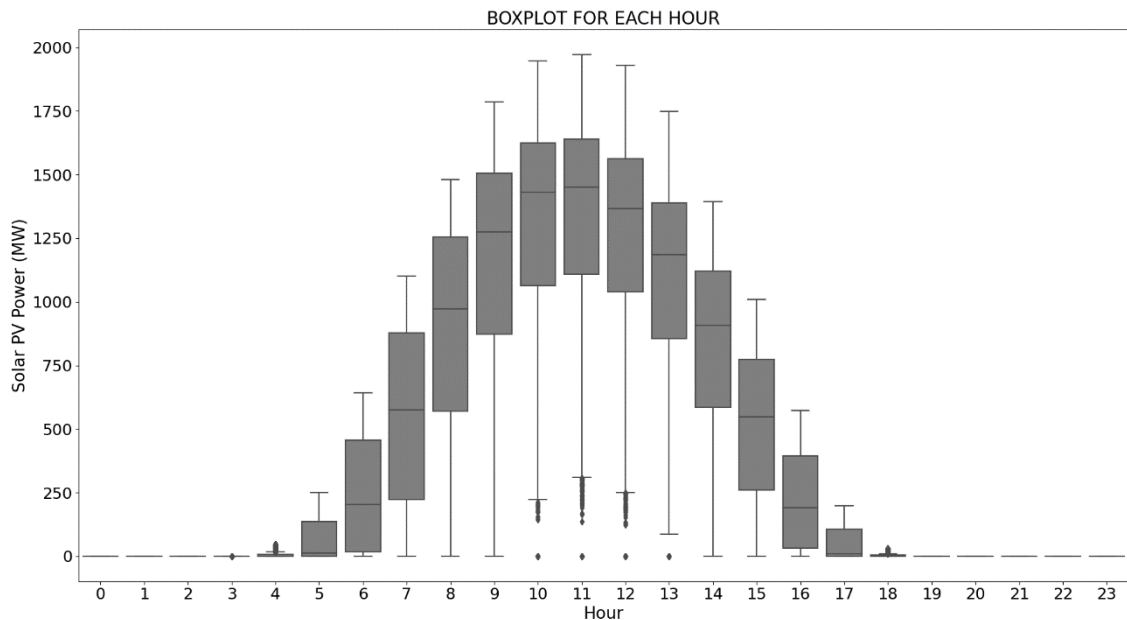
ενέργεια έχει μεγάλες διακυμάνσεις με αυξημένη τάση στις μεσημεριανές ώρες. Οι αποκλίσεις που φαίνονται οφείλονται στους μήνες κατά τους οποίους παρουσιάστηκε η μέτρηση καθώς και στην χρονολογία των δεδομένων.

```
#BOXPLOT FOR EACH HOUR

fig=plt.figure(figsize=(18,10))
sns.boxplot(data=data_time,x='hour',y='Solar PV Power',color='grey')
plt.title('BOXPLOT FOR EACH HOUR',fontsize=20)
plt.xlabel('Hour',fontsize=19)
plt.xticks(fontsize=18)
plt.yticks(fontsize=18)
plt.ylabel('Solar PV Power (MW)',fontsize=19)
plt.tight_layout()
plt.show()

fig.savefig('BOXPLOT FOR EACH HOUR.png')
```

Εικόνα 21: Κώδικας για την κατασκευή του boxplot της μέσης ηλιακής ενέργειας ανά ώρα.



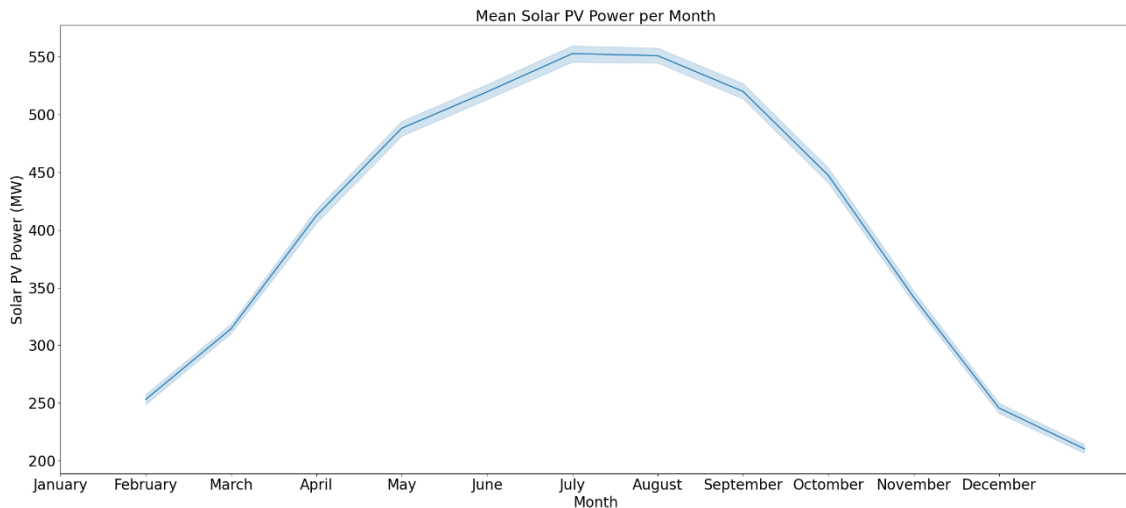
Σχήμα 3: Boxplot της μέσης ηλιακής ενέργειας ανά ώρα.

Η αντίστοιχη παροχή ενέργειας ανά μήνα φανερώνει πως το μέγιστο βρίσκεται στους θερινούς μήνες, ιδιαίτερα τον Αύγουστο, ενώ στους χειμερινούς παρουσιάζει απότομη πτώση. Το διάγραμμα δείχνει τη μέση τιμή των δεδομένων ανά μήνα από το 1979 έως τον Ιούλιο του 2023.

Monthly

```
fig = plt.figure(figsize=(22,10))
sns.lineplot(data=data_time,x='month',y='Solar PV Power')
labels=['January','February','March','April','May','June','July','August','September','October','November','December']
plt.xticks(np.arange(12),labels,fontsize=20)
plt.yticks(fontsize=20)
plt.title('Mean Solar PV Power per Month',fontsize=21)
plt.xlabel('Month',fontsize=20)
plt.ylabel('Solar PV Power (MW)',fontsize=20)
plt.tight_layout()
plt.savefig('Mean Solar PV Power per Month (MW).png')
plt.show()
```

Εικόνα 22: Κώδικας για την κατασκευή του διαγράμματος της μέσης ηλιακής ενέργειας ανά μήνα.

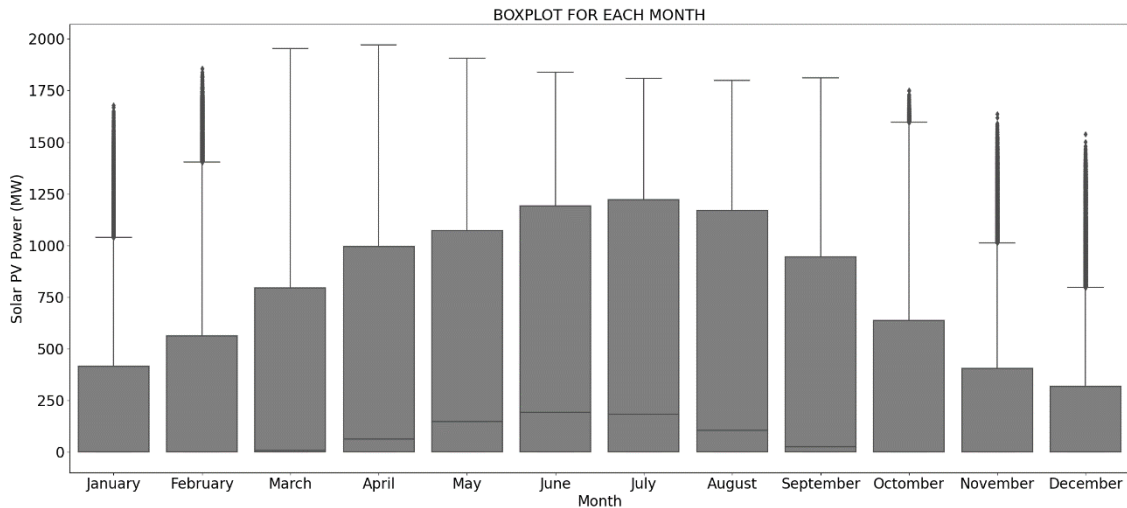


Σχήμα 4: Διάγραμμα μέσης ηλιακής ενέργειας ανά μήνα.

Επιπρόσθετα παρουσιάζεται και το θηκόγραμμα (box plot) στο οποίο παρουσιάζεται αυξημένη τάση στους θερινούς μήνες και μεγάλες και πάλι αποκλίσεις το μέσο όρο σε όλους τους μήνες.

```
#BOXPLOT FOR EACH MONTH
fig = plt.figure(figsize=(22,10))
sns.boxplot(data=data_time,x='month',y='Solar PV Power',color='grey')
labels=['January','February','March','April','May','June','July','August','September','October','November','December']
plt.xticks(np.arange(12),labels,fontsize=20)
plt.yticks(fontsize=20)
plt.title('BOXPLOT FOR EACH MONTH',fontsize=21)
plt.xlabel('Month',fontsize=20)
plt.ylabel('Solar PV Power (MW)',fontsize=20)
plt.tight_layout()
plt.show()
fig.savefig('BOXPLOT FOR EACH MONTH.png')
```

Εικόνα 23: Κώδικας για την κατασκευή του διαγράμματος της μέσης ηλιακής ενέργειας ανά μήνα.

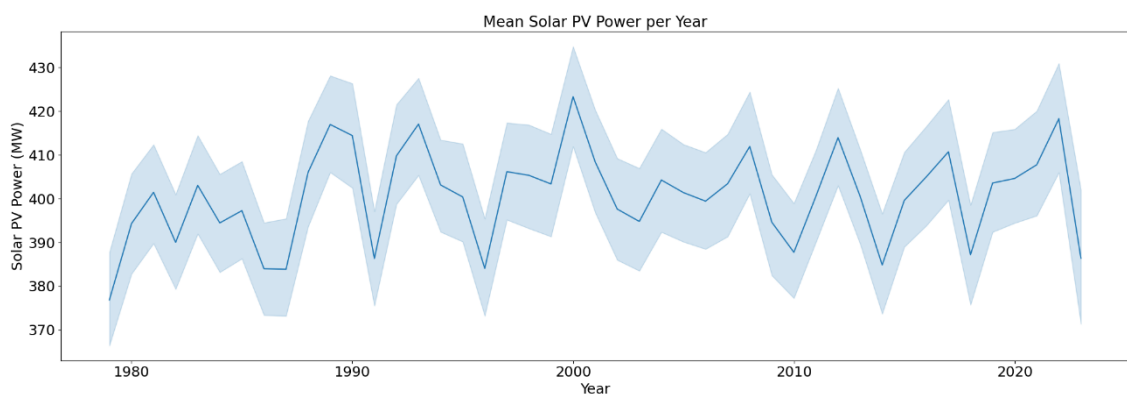


Σχήμα 5: Boxplot της μέσης ηλιακής ενέργειας ανά μήνα.

Τέλος, η μέση και η συνολική παροχή ηλιακής ενέργειας ανά χρονιά για όλο το σύνολο δεδομένων, από το 1979 έως τον Ιούνιο του 2022, παρουσιάζει μεγάλες αυξομειώσεις. Το 2023 δεν παρουσιάζεται καθώς δεν υπάρχουν τα δεδομένα όλου του έτους.

```
# Solar PV Power (mean) until 2022
fig=plt.figure(figsize=(20,7))
sns.lineplot(data=data_time,x='year',y='Solar PV Power')
plt.xticks(fontsize=18)
plt.yticks(fontsize=18)
plt.title('Mean Solar PV Power per Year',fontsize=19)
plt.xlabel('Year',fontsize=18)
plt.ylabel('Solar PV Power (MW)',fontsize=18)
plt.tight_layout()
plt.savefig('Mean Solar PV Power per Year (MW).png')
plt.show()
```

Εικόνα 24: Κώδικας για την κατασκευή του διαγράμματος της μέσης ηλιακής ενέργειας ανά χρόνο.

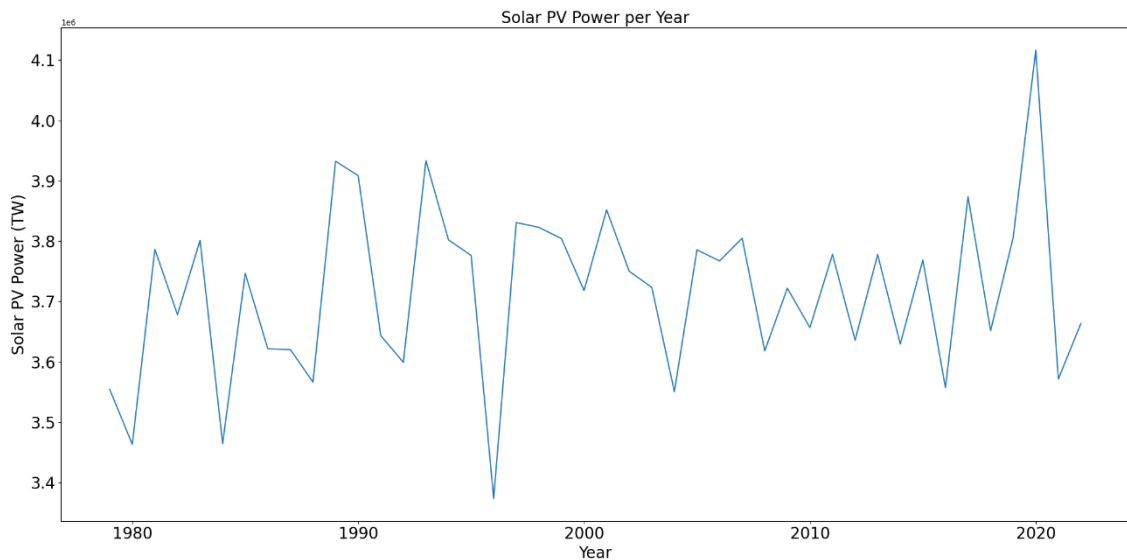


Σχήμα 6: Διάγραμμα της μέσης ηλιακής ενέργειας ανά χρόνο.

Στη συνολική ενέργεια που παράχθηκε ανά χρονιά από φωτοβολταϊκά ακολουθεί το μέσο όρο παρουσιάζοντας ξαφνική μεγάλη το 2020 η οποία βέβαια πέφτει κοντά στο μέσο όρο την επόμενη κιόλας χρονιά.

```
# Solar PV Power (sum) until 2022
fig=plt.figure(figsize=(20,10))
sns.lineplot(data=data_time_2022,x='year',y='SP_SumByYear')
plt.xticks(x='year', fontsize=20)
plt.yticks(fontsize=20)
plt.title('Solar PV Power per Year',fontsize=20)
plt.xlabel('Year',fontsize=20)
plt.ylabel('Solar PV Power (TW)',fontsize=20)
plt.tight_layout()
plt.savefig('Solar PV Power per Year (TW).png')
plt.show()
```

Εικόνα 25: Κώδικας για την κατασκευή του διαγράμματος της συνολικής ηλιακής ενέργειας ανά χρόνο.



Σχήμα 7: Διάγραμμα της συνολικής ηλιακής ενέργειας ανά χρόνο

2.4 Δημιουργία επιπλέον μεταβλητών

Για την ανάπτυξη του μοντέλου πρόβλεψης δημιουργήθηκαν δυο επιπλέον στήλες με τις ώρες και τους μήνες της μέτρησης με σκοπό το συσχετισμό της παροχής ενέργειας με βάση την ώρα και τον μήνα καθώς όπως προέκυψε από τη στατιστική μελέτη δίνουν μεγάλο βάρος στα αποτελέσματα. Οι απροσδιόριστες τιμές ('NaN') παρουσιάζονται στις περιπτώσεις που δεν υπάρχει ηλιοφάνεια, π.χ. βραδινές ώρες και αντικαταστάθηκαν με την τιμή μηδέν ώστε να μην τις αποκλείσει το μοντέλο καθώς τα δεδομένα δίνονται σε όλη τη διάρκεια της ημέρας και όχι μόνο στις ώρες ηλιοφάνειας.

Πίνακας 4: Παρουσίαση των επιπλέον μεταβλητών.

| | Date | Solar PV Power | Air Temperature | Global Horizontal Irradiance | Mean Sea Level Pressure | Total Precipitation | Wind Speed 10m | Wind Speed 100m | hour | month |
|--------|---------------------|----------------|-----------------|------------------------------|-------------------------|---------------------|----------------|-----------------|------|-------|
| 0 | 1979-01-01 00:00:00 | 0.000000 | 284.90286 | NaN | 100365.470 | NaN | 3.524685 | 6.391961 | 0 | 1 |
| 1 | 1979-01-01 01:00:00 | 0.000000 | 284.95416 | NaN | 100294.310 | NaN | 3.385256 | 6.170972 | 1 | 1 |
| 2 | 1979-01-01 02:00:00 | 0.000000 | 285.61070 | NaN | 100214.970 | NaN | 3.327355 | 6.026396 | 2 | 1 |
| 3 | 1979-01-01 03:00:00 | 0.000000 | 285.44745 | NaN | 100149.040 | NaN | 3.410202 | 6.118879 | 3 | 1 |
| 4 | 1979-01-01 04:00:00 | 0.000000 | 285.35626 | NaN | 100078.020 | NaN | 3.421252 | 6.124903 | 4 | 1 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 413008 | 2023-07-31 18:00:00 | 11.919418 | 299.32834 | 1.548849e+01 | 100839.734 | 3.029865e-07 | 2.036868 | 3.357661 | 18 | 7 |
| 413009 | 2023-07-31 19:00:00 | 0.000000 | 298.88220 | 6.467518e-14 | 100940.990 | 4.022958e-07 | 1.722882 | 3.008210 | 19 | 7 |
| 413010 | 2023-07-31 20:00:00 | 0.000000 | 297.24496 | 6.467518e-14 | 100986.950 | 1.343638e-07 | 1.588435 | 2.707174 | 20 | 7 |
| 413011 | 2023-07-31 21:00:00 | 0.000000 | 296.87473 | 6.467518e-14 | 101005.640 | 6.926348e-06 | 1.507571 | 2.452604 | 21 | 7 |
| 413012 | 2023-07-31 22:00:00 | 0.000000 | 295.26300 | 6.467518e-14 | 101013.400 | 6.448036e-06 | 1.426133 | 2.342447 | 22 | 7 |

413013 rows × 10 columns

Επιπλέον, κατασκευάστηκαν νέες μεταβλητές με την παροχή ενέργειας την προηγούμενη ώρα ('prev_hour'), τη μέση τιμή της ενέργεια το πρωί, απόγευμα και βράδυ ('morning_afternoon_evening') αλλά και τη μέση ενέργεια από την αυγή έως και τη δύση ('awn_dusk') με σκοπό την ορθή εκμάθηση του μοντέλου την μηδενική ενέργεια παροχή τις ώρες που δεν υπάρχει ηλιοφάνεια. Οι μεταβλητές κρίθηκαν απαραίτητες καθώς πριν την εισαγωγή τους η προβλέψεις στις ώρες που δεν υπήρχε ο ήλιος έδιναν μεγάλες τιμές που πολλές φορές κυμαίνονταν κοντά στον μέσο όρο.

```
def create_features(data, dep_name=None):
    data['index'] = data.index

    # previous hour
    def prev_hour_energy():
        data['prev_hour_energy'] = [data['Solar PV Power'][i] if i == 0 else data['Solar PV Power'][i-1]
                                   for i in range(len(data))]

    prev_hour_energy()

    #mean morning afternoon evening
    def morning_afternoon_evening():
        data['morning_afternoon_evening'] = data['Solar PV Power'].rolling(8).mean()

    morning_afternoon_evening()

    #mean dawn dusk
    # https://www.worlddata.info/europe/greece/sunset.php
    def dawn_dusk():
        data['dawn_dusk'] = data['Solar PV Power'].rolling(12).mean()

    dawn_dusk()

    data = data.drop(['index'], axis=1)
    return data
```

Εικόνα 26: Κώδικας δημιουργίας των επιπλέον μεταβλητών.

Μέρος του νέου πίνακα μεταβλητών φαίνεται παρακάτω:

Πίνακας 5: Παρουσίαση των επιπλέον μεταβλητών.

| | Date | Solar PV Power | Air Temperature | Global Horizontal Irradiance | Mean Sea Level Pressure | Total Precipitation | Wind Speed 10m | Wind Speed 100m | hour | month | prev_hour_energy | morning_afternoon_evening | dawn_dusk |
|--------|---------------------|----------------|-----------------|------------------------------|-------------------------|---------------------|----------------|-----------------|------|-------|------------------|---------------------------|-------------|
| 0 | 1979-01-01 00:00:00 | 0.000000 | 284.90286 | 0.000000e+00 | 100365.470 | 0.000000e+00 | 3.524685 | 6.391961 | 0 | 1 | 0.000000 | NaN | NaN |
| 1 | 1979-01-01 01:00:00 | 0.000000 | 284.95416 | 0.000000e+00 | 100294.310 | 0.000000e+00 | 3.385256 | 6.170972 | 1 | 1 | 0.000000 | NaN | NaN |
| 2 | 1979-01-01 02:00:00 | 0.000000 | 285.61070 | 0.000000e+00 | 100214.970 | 0.000000e+00 | 3.327355 | 6.026396 | 2 | 1 | 0.000000 | NaN | NaN |
| 3 | 1979-01-01 03:00:00 | 0.000000 | 285.44745 | 0.000000e+00 | 100149.040 | 0.000000e+00 | 3.410202 | 6.118879 | 3 | 1 | 0.000000 | NaN | NaN |
| 4 | 1979-01-01 04:00:00 | 0.000000 | 285.35626 | 0.000000e+00 | 100078.020 | 0.000000e+00 | 3.421252 | 6.124903 | 4 | 1 | 0.000000 | NaN | NaN |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 413008 | 2023-07-31 18:00:00 | 11.919418 | 299.32834 | 1.548849e+01 | 100839.734 | 3.029865e-07 | 2.038688 | 3.357661 | 18 | 7 | 151.358048 | 866.955674 | 1014.837886 |
| 413009 | 2023-07-31 19:00:00 | 0.000000 | 298.88220 | 6.467518e-14 | 100940.990 | 4.022958e-07 | 1.722882 | 3.008210 | 19 | 7 | 11.919418 | 666.644596 | 939.374512 |
| 413010 | 2023-07-31 20:00:00 | 0.000000 | 297.24496 | 6.467518e-14 | 100986.950 | 1.343638e-07 | 1.588435 | 2.707174 | 20 | 7 | 0.000000 | 481.989237 | 833.475497 |
| 413011 | 2023-07-31 21:00:00 | 0.000000 | 296.87473 | 6.467518e-14 | 101005.640 | 6.926348e-06 | 1.507571 | 2.452604 | 21 | 7 | 0.000000 | 313.039527 | 709.911467 |
| 413012 | 2023-07-31 22:00:00 | 0.000000 | 295.26300 | 6.467518e-14 | 101013.400 | 6.448036e-06 | 1.426133 | 2.342447 | 22 | 7 | 0.000000 | 176.003056 | 577.970450 |

413013 rows x 13 columns

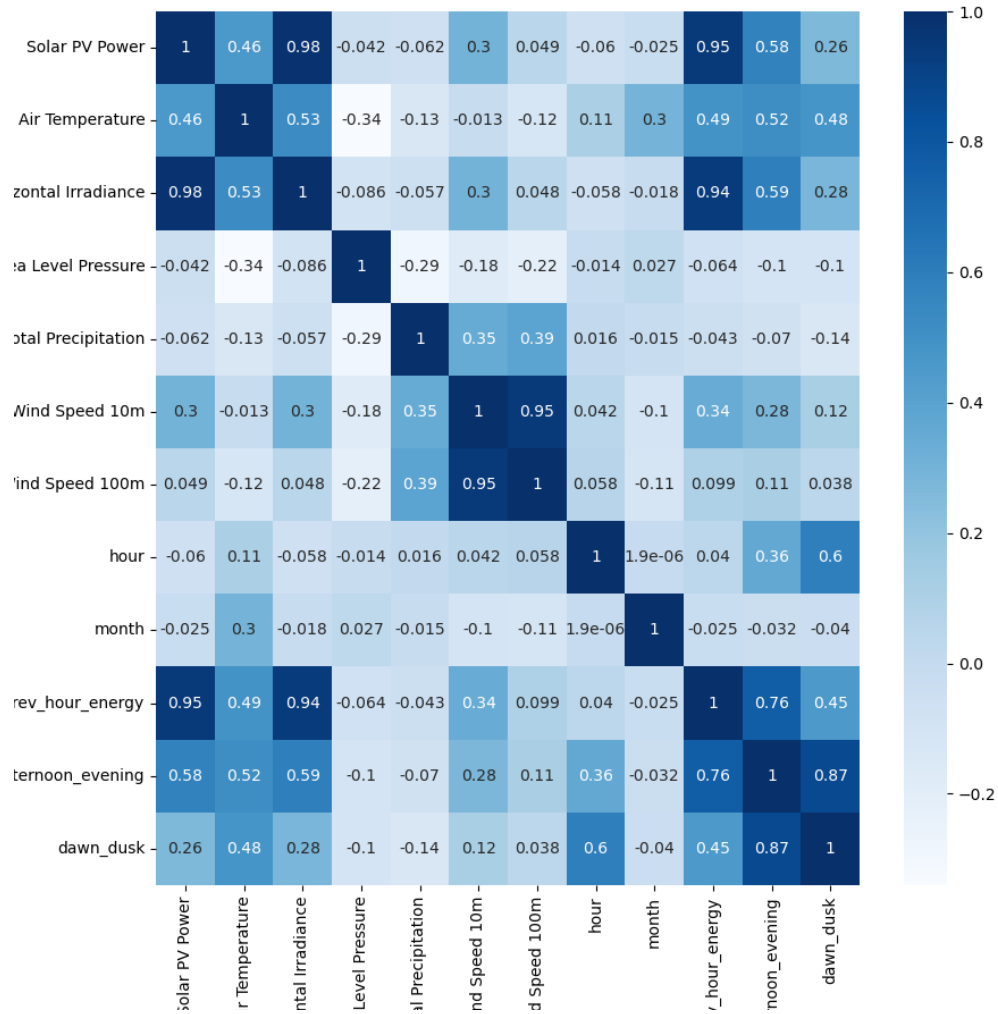
Στη συνέχεια πραγματοποιήθηκε ο υπολογισμός της χάρτη συσχέτισης (correlation map) ώστε να εντοπισθούν οι συσχετιζόμενες μεταβλητές οι οποίες οδηγούν σε υπερφόρτωση του μοντέλου και του όγκου των υπο επεξεργασία δεδομένων. Ο κώδικας που βοήθησε στην υλοποίηση της χάρτη συσχέτισης παρουσιάζεται παρακάτω:

Correlation Map

```
#correlation map
plt.figure(figsize=(10,10))
cor = data_pred.corr()
sns.heatmap(cor, annot=True, cmap=plt.cm.Blues)
plt.savefig('corr.png')
```

Εικόνα 27: Κώδικας υλοποίησης του χάρτη συσχέτισης.

Τα αποτελέσματα του κώδικα στα οποία φαίνονται όλες οι αρχικές μεταβλητές και η μεταξύ τους συσχέτιση δείχνει πως κάποιες από αυτές συσχετίζονται σε μεγάλο βαθμό (έντονο μπλε χρώμα).



Σχήμα 8: Χάρτης συσχέτισης (Correlation map).

Αφαιρώντας τις μεταβλητές που συσχετίζονται σε έντονο βαθμό το μοντέλο πραγματοποιεί με μεγαλύτερη ταχύτητα και ευστοχία τους κατάλληλους υπολογισμούς δίνοντας πιο αξιόπιστη πρόβλεψη. Στη συγκεκριμένη εργασία τέθηκε ως όριο συσχέτισης το 95% (0,95). Παρουσιάζεται ο κώδικας και τα αποτελέσματα σε νέο χάρτη συσχέτισης:

```
# Drop highly correlated feature
threshold = 0.95

columns = np.full((cor.shape[0],), True, dtype=bool)
for i in range(cor.shape[0]):
    for j in range(i+1, cor.shape[0]):
        if cor.iloc[i,j] >= threshold:
            if columns[j]:
                columns[j] = False
selected_columns = cor.columns[columns]
selected_columns

data_pred = data_pred[selected_columns]
```

Εικόνα 28: Κώδικας διαγραφής των σχετιζόμενων μεταβλητών.

```
# Drop highly correlated feature
threshold = 0.95

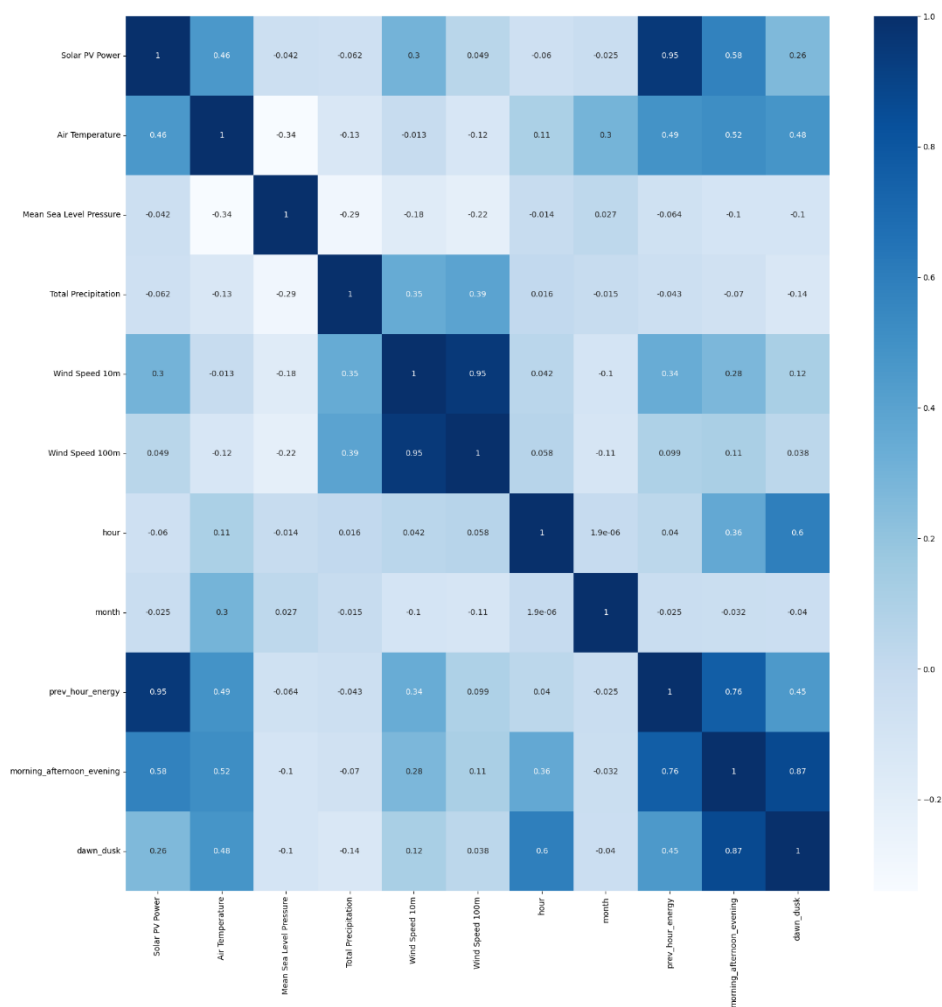
columns = np.full((cor.shape[0],), True, dtype=bool)
for i in range(cor.shape[0]):
    for j in range(i+1, cor.shape[0]):
        if cor.iloc[i,j] >= threshold:
            if columns[j]:
                columns[j] = False
selected_columns = cor.columns[columns]
selected_columns

data_pred = data_pred[selected_columns]
```

```
#correlation map

plt.figure(figsize=(20,20))
cor2 = data_pred.corr()
sns.heatmap(cor2, annot=True, cmap=plt.cm.Blues)
plt.savefig('not_corr.png')
```

Εικόνα 29: Κώδικας υλοποίησης του χάρτη συσχέτισης.



Σχήμα 9: Χάρτης συσχέτισης (Correlation map).

3. Κατασκευή Μοντέλου Μηχανικής Μάθησης

Τα δεδομένα που εισαγάγαμε στο μοντέλο είναι ενέργεια από φωτοβολταϊκά ('Solar PV Power'), η θερμοκρασία του αέρα ('Air Temperature'), η μέση πίεση στο επίπεδο της θάλασσας ('Mean Sea Level Pressure'), η συνολική βροχόπτωση ('Total Precipitation'), η ταχύτητα του ανέμου στο 10 και 100 μέτρα υψόμετρο ('Wind Speed 10m', 'Wind Speed 100m'), η ώρα που έγινε η μέτρηση ('hour'), η μέτρηση την προηγούμενη ώρα, ο μήνας που πραγματοποιήθηκε η μέτρηση ('month'), αν ήταν μεσημέρι, απόγευμα ή βράδυ ('morning_afternoon_evening') καθώς και βρισκόμαστε μετά την αυγή ή τη δύση του ήλιου ('dawn_dusk'). Στο σύνολο τους δηλαδή 413.013 γραμμές επί 11 στήλες, 4.543.143 δεδομένα για την ανάλυση.

```
data_xgb=data_pred.copy()
data_xgb = data_xgb.set_index(data_xgb['Date'], drop=False, append=False, inplace=False, verify_integrity=False).drop('Date', 1)
data_xgb.head()
```

Εικόνα 30: Κώδικας προετοιμασίας δεδομένων για την εισαγωγή στον αλγόριθμο.

Στον επόμενο πίνακα παρουσιάζονται και τα δεδομένα που εισάγουμε στον αλγόριθμο.

Πίνακας 6: Παρουσίαση των δεδομένων για την εισαγωγή στον αλγόριθμο.

| Date | Solar PV Power | Air Temperature | Mean Sea Level Pressure | Total Precipitation | Wind Speed 10m | Wind Speed 100m | hour | month | prev_hour_energy | morning_afternoon_evening | dawn_dusk |
|---------------------|----------------|-----------------|-------------------------|---------------------|----------------|-----------------|------|-------|------------------|---------------------------|-------------|
| 1979-01-01 00:00:00 | 0.000000 | 284.90286 | 100365.470 | 0.000000e+00 | 3.524685 | 6.391961 | 0 | 1 | 0.000000 | 0.000000 | 0.000000 |
| 1979-01-01 01:00:00 | 0.000000 | 284.95416 | 100294.310 | 0.000000e+00 | 3.385256 | 6.170972 | 1 | 1 | 0.000000 | 0.000000 | 0.000000 |
| 1979-01-01 02:00:00 | 0.000000 | 285.61070 | 100214.970 | 0.000000e+00 | 3.327355 | 6.026396 | 2 | 1 | 0.000000 | 0.000000 | 0.000000 |
| 1979-01-01 03:00:00 | 0.000000 | 285.44745 | 100149.040 | 0.000000e+00 | 3.410202 | 6.118879 | 3 | 1 | 0.000000 | 0.000000 | 0.000000 |
| 1979-01-01 04:00:00 | 0.000000 | 285.35626 | 100078.020 | 0.000000e+00 | 3.421252 | 6.124903 | 4 | 1 | 0.000000 | 0.000000 | 0.000000 |
| ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... | ... |
| 2023-07-31 18:00:00 | 11.919418 | 299.32834 | 100839.734 | 3.029865e-07 | 2.036868 | 3.357661 | 18 | 7 | 151.358048 | 866.955674 | 1014.837886 |
| 2023-07-31 19:00:00 | 0.000000 | 298.88220 | 100940.990 | 4.022958e-07 | 1.722882 | 3.008210 | 19 | 7 | 11.919418 | 666.644596 | 939.374512 |
| 2023-07-31 20:00:00 | 0.000000 | 297.24496 | 100986.950 | 1.343638e-07 | 1.588435 | 2.707174 | 20 | 7 | 0.000000 | 481.989237 | 833.475497 |
| 2023-07-31 21:00:00 | 0.000000 | 296.87473 | 101005.640 | 6.926348e-06 | 1.507571 | 2.452604 | 21 | 7 | 0.000000 | 313.039527 | 709.911467 |
| 2023-07-31 22:00:00 | 0.000000 | 295.26300 | 101013.400 | 6.448036e-06 | 1.426133 | 2.342447 | 22 | 7 | 0.000000 | 176.003056 | 577.970450 |

413013 rows x 11 columns

Η μεταβλητή για την οποία θα πραγματοποιηθεί μελέτη για την πρόβλεψη της είναι η ενέργεια από φωτοβολταϊκά ('Solar PV Power') μέσω του αλγορίθμου παλινδρόμησης 'XGBRegressor' από τη βιβλιοθήκη XGBoost. Το μοντέλο του XGBoost παρουσιάζεται παρακάτω.

Το δείγμα χωρίστηκε κατά 70% σε δεδομένα εκπαίδευσης (train data) και 30% σε δεδομένα προς έλεγχο (test data). Οι παράμετροι του μοντέλου ορίστηκαν όπως φαίνονται παρακάτω μετά από έλεγχο διαφόρων τιμών με τη μέθοδο «gridsearch» και score nMAE:-58.89 (MW).

- n_estimators : 200
- learning_rate: 0.1
- max_depth: 1
- min_child_weight: 1
- subsample: 1
- colsample_bytree: 0.70
- gamma: 0

```
#XGBoost Model

train_size=round(data_xgb.shape[0]*0.70)
df_train = data_xgb.iloc[:train_size].copy()
df_test = data_xgb.iloc[train_size:].copy()

X_train=df_train.iloc[:,1:] #features (col 2 - 9)
y_train=df_train.iloc[:,0] #target (col 1) Solar PV Power

X_test=df_test.iloc[:,1:] #features (col 2 - 9)
y_test=df_test.iloc[:,0] #target (col 1) Solar PV Power

paramGrid = {'n_estimators':[100, 150, 200],
             'learning_rate':[0.1, 0.05],
             'max_depth':[1],
             'min_child_weight':[1],
             'subsample':[1],
             'colsample_bytree':[0.70],
             'gamma':[0]
            }

tscv = TimeSeriesSplit(n_splits=2)
model=xgb.XGBRegressor()

gridsearch=GridSearchCV(model,paramGrid, cv=tscv, scoring='neg_mean_absolute_error')

fit = gridsearch.fit(X_train, y_train)

print('Score:', fit.best_score_)

print('\nThe best params:', fit.best_params_)

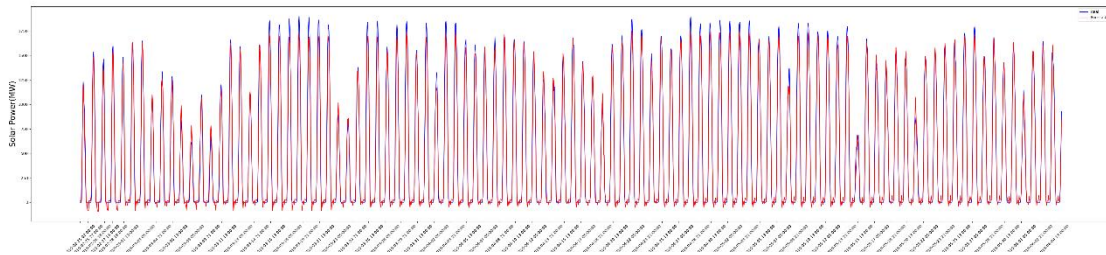
Score: -58.89379594816839

The best params: {'colsample_bytree': 0.7, 'gamma': 0, 'learning_rate': 0.1, 'max_depth': 1, 'min_child_weight': 1, 'n_estimators': 200, 'subsample': 1}
```

Εικόνα 31: Κώδικας υλοποίησης μοντέλου μηχανικής μάθησης (XGBoost).

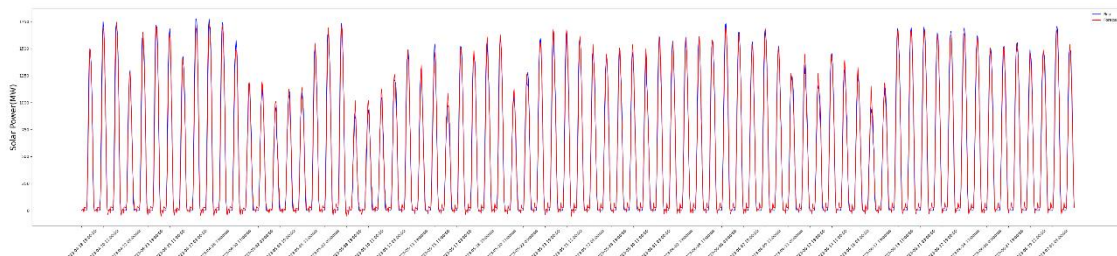
Τα αποτελέσματα του αλγορίθμου παρουσιάζονται στις επόμενες γραφικές παραστάσεις. Η πρώτη γραφική παράσταση δείχνει τα πρώτα 3000 δεδομένα της μεταβλητής που μελετάται, «Solar PV Power» σε σχέση με το χρόνο.

Τα πραγματικά δεδομένα παρουσιάζονται με μπλε χρώμα, ενώ τα δεδομένα του προέκυψαν από το μοντέλο πρόβλεψης με κόκκινο χρώμα. Παρατηρείται ικανοποιητική προσαρμογή των δεδομένων πρόβλεψης πάνω στα πραγματικά με απόκλιση 4% (Μέσος Όρος) για όλο το δείγμα.



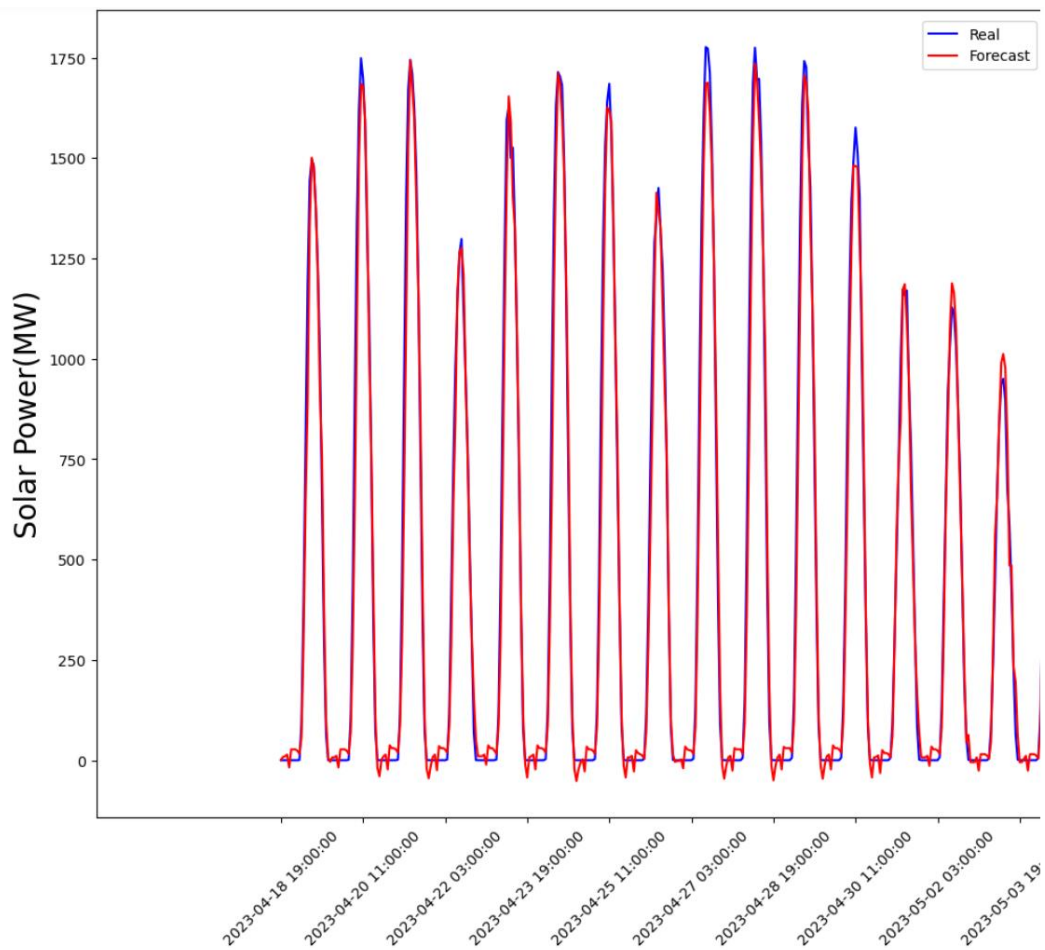
Σχήμα 10: Διάγραμμα παρουσίασης των προβλέψεων και των πραγματικών τιμών ανά χρόνο (πρώτες χρονολογίες).

Στη συνέχεια φαίνονται τα τελευταία 3000 δεδομένα. Πάλι με μπλε χρώμα παρουσιάζονται τα πραγματικά ενώ με κόκκινο χρώμα τα δεδομένα που προέκυψαν από το μοντέλο.



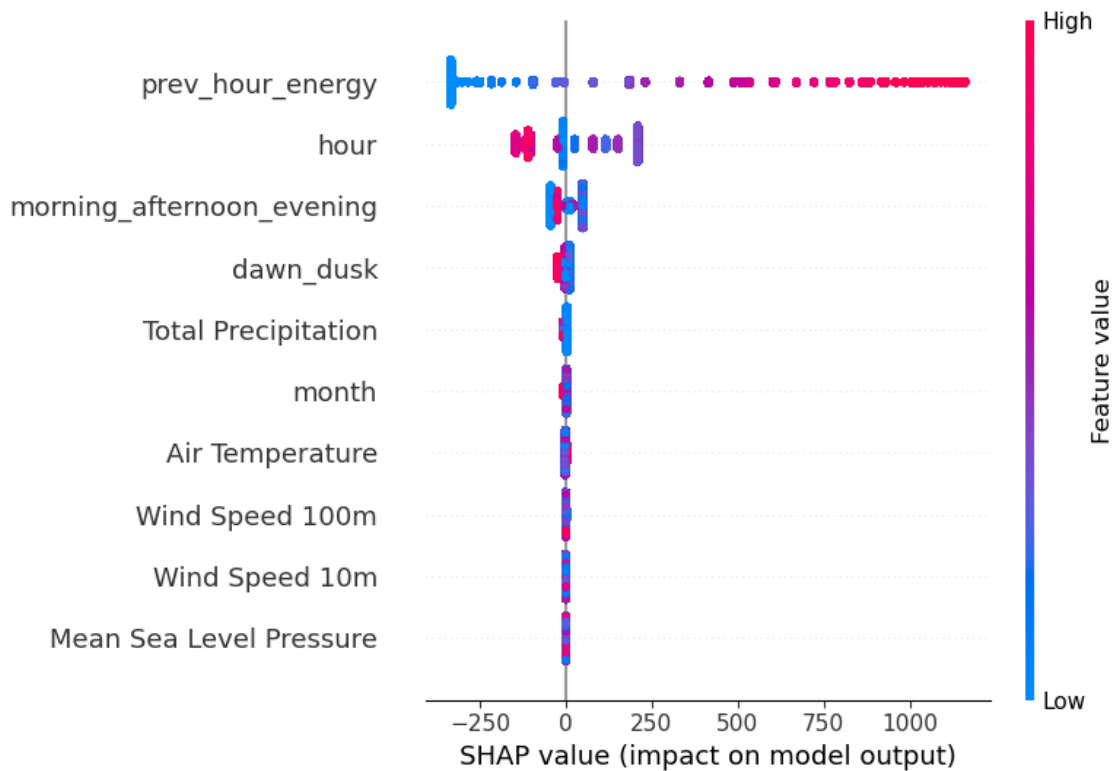
Σχήμα 11: Διάγραμμα παρουσίασης των προβλέψεων και των πραγματικών τιμών ανά χρόνο (τελευταίες χρονολογίες).

Τέλος, σε μεγέθυνση φαίνεται ένα μέρος του δείγματος. Στον y-κάθετο άξονα υπάρχει η ηλιακή ενέργεια σε MW ενώ στο x-οριζόντιο άξονα ο χρόνος ανά περιόδους.



Σχήμα 12: Διάγραμμα παρουσίασης των προβλέψεων και των πραγματικών τιμών ανά χρόνο.

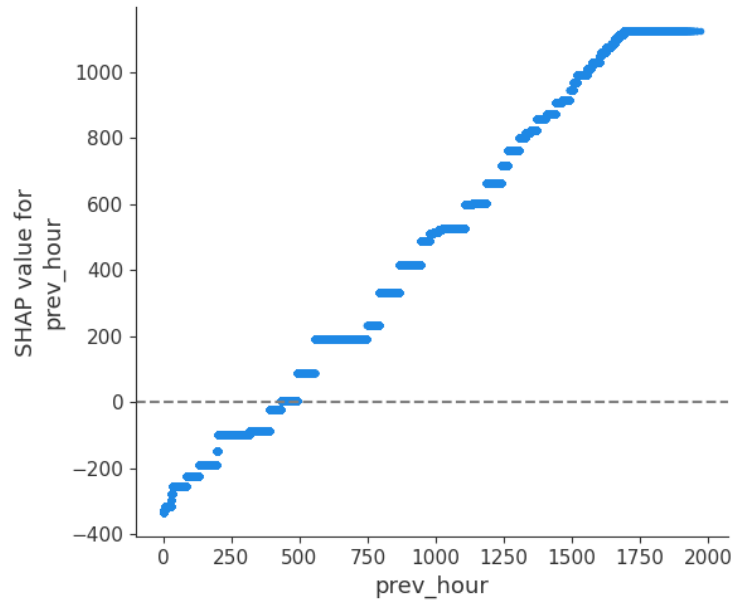
Στη συνέχεια, για την ερμηνεία των προβλέψεων του μοντέλου μηχανικής μάθησης χρησιμοποιήθηκε μέρος της βιβλιοθήκης SHAP (SHapley Additive Explanations) και συγκεκριμένα το «shap.TreeExplainer». Με τη βοήθεια του «shap.TreeExplainer» λήφθηκαν οι πληροφορίες σχετικά με την επιρροή που άσκησαν οι μεταβλητές του μοντέλου, εκτός της μεταβλητής στόχου, στην πρόβλεψη του αλγορίθμου. Τόσο το μοντέλο όσο και η γραφική επεξήγηση του δείχνουν ικανοποιητικά αποτελέσματα καθώς όχι μόνο προβλέπει ικανή βάση τις τιμές της ενέργειας αλλά δείχνει πως κατανοεί σε βάθος πότε υπάρχει παραγωγή ηλιακής ενέργειας.



Σχήμα 13: Διάγραμμα παρουσίασης της ερμηνείας των προβλέψεων του μοντέλου μηχανικής μάθησης.

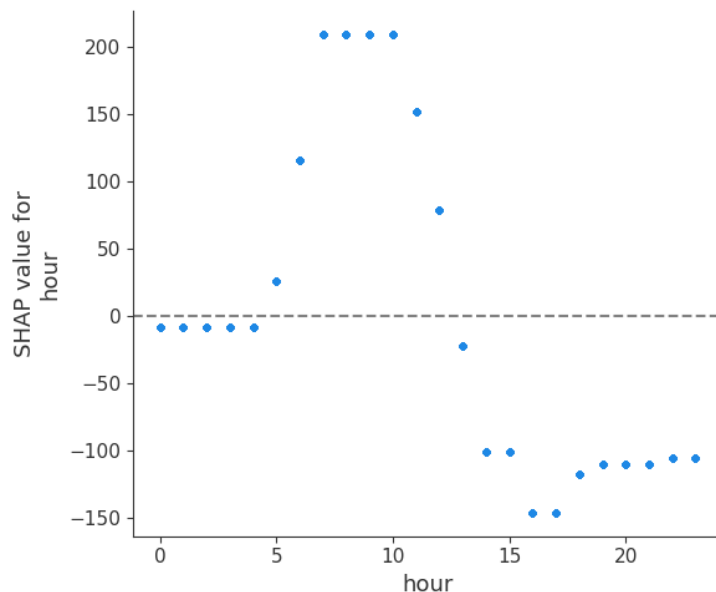
Αναλυτικότερα, το ποσοστό επηρεασμού της παραγωγής ηλεκτρικής ενέργειας από φωτοβολταϊκά (μεταβλητή στόχος) παρουσιάζεται σε σχέση με καθεμιά από τις υπόλοιπες μεταβλητές στα παρακάτω διαγράμματα. Τα διαγράμματα υλοποιήθηκαν με το «shap.dependence_plot» της βιβλιοθήκης SHAP (SHapley Additive Explanations).

Παρατηρείται πως η τιμή ηλιακής ενέργειας (ισχύς επί ώρα) που καταγράφηκε την προηγούμενη ώρα είναι η σημαντικότερη μεταβλητή στο μοντέλο. Όταν η τιμή της ενέργειας που καταγράφηκε είναι υψηλή τότε υπάρχει μεγάλη πιθανότητα να υπάρξει παραγωγή ηλιακής ενέργειας από φωτοβολταϊκά. Αντίστοιχα, μια χαμηλή τιμή καταγεγραμμένης ενέργειας (ισχύς επί ώρα) την προηγούμενη ώρα δίνει σχετικά μεγάλη πιθανότητα να μην υπάρξει καταγραφή ή να υπάρχουν χαμηλά επίπεδα παραγωγής ηλιακής ενέργειας.



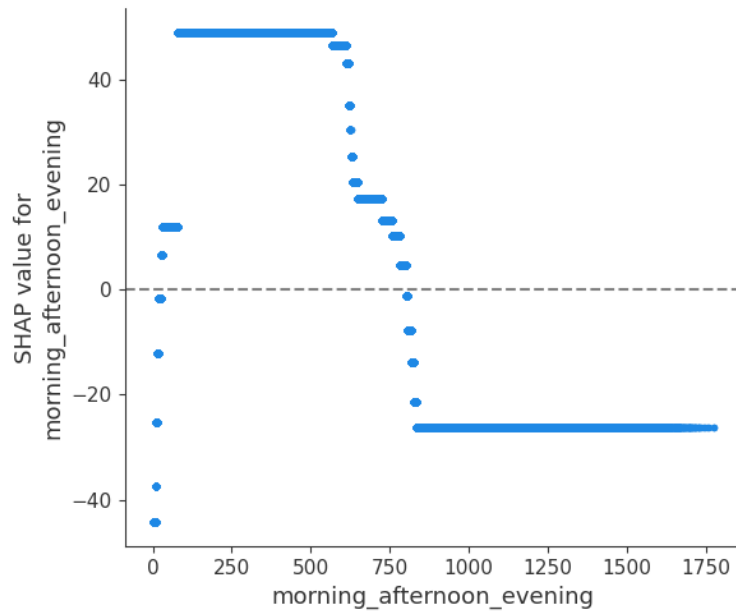
Σχήμα 14: Διάγραμμα παρουσίασης της ερμηνείας των προβλέψεων για τη μεταβλητή “prev_hour”.

Επιπρόσθετα η ώρα συσχετίζεται άμεσα με την παροχή ισχύος από ηλιακή ενέργεια στα φωτοβολταϊκά καθώς οι απογευματινές και βραδινές ώρες (π.χ. 20:00) δείχνουν πιθανότητα μη ύπαρξης ηλιακής ενέργειας, ώρες κοντά στο μηδέν (π.χ. 02:00) δείχνουν μηδενική παραγωγή ενέργειας ενώ οι ώρες που κυμαίνονται σε πρωί και μεσημέρι μεγάλη παραγωγή ενέργειας (π.χ. 10:00).

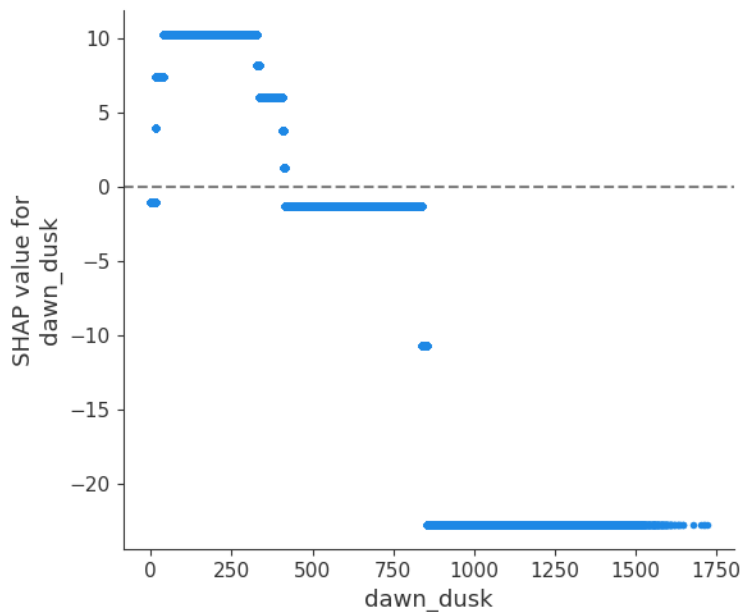


Σχήμα 15: Διάγραμμα παρουσίασης της ερμηνείας των προβλέψεων για την μεταβλητή “hour”.

Όμοια οι μεταβλητές που ορίζουν αν είναι μέρα, απόγευμα ή βράδυ δείχνει ότι επιδρά με τον ίδιο τρόπο στο μοντέλο.

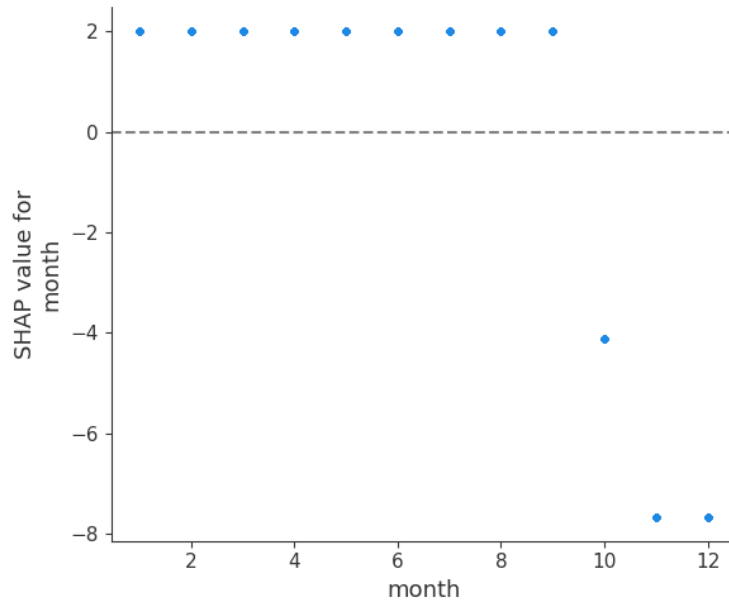


Σχήμα 16: Διάγραμμα παρουσίασης της ερμηνείας των προβλέψεων για τη μεταβλητή “morning_afternoon_evening”.



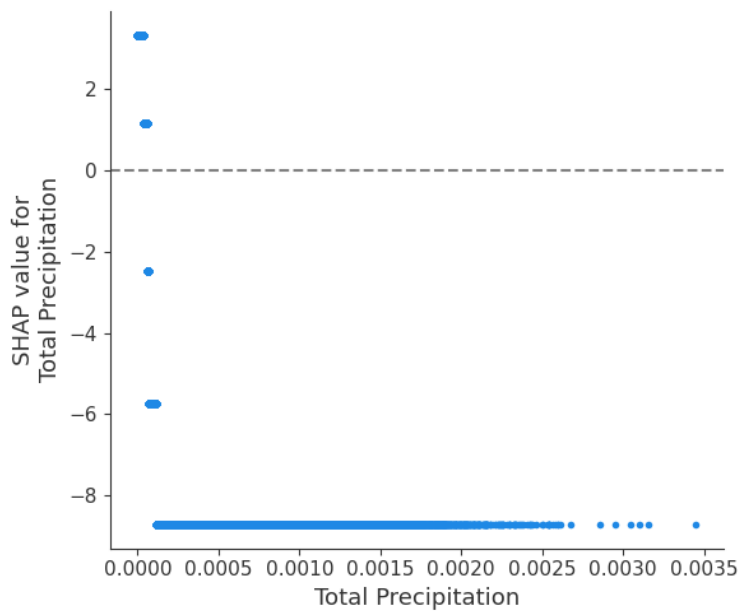
Σχήμα 17: Διάγραμμα παρουσίασης της ερμηνείας των προβλέψεων για τη μεταβλητή “dawn_dusk”.

Επίσης, βλέπουμε πως ο μήνας του έτους έχει πολύ μικρή επίδραση στην παραγωγή ενέργειας από την ηλιακή ακτινοβολία.

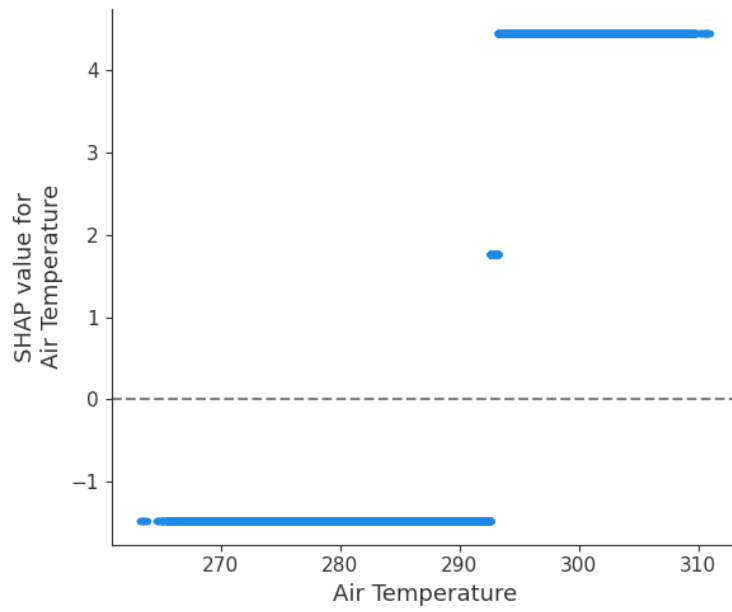


Σχήμα 18: Διάγραμμα παρουσίασης της ερμηνείας των προβλέψεων για τη μεταβλητή “month”.

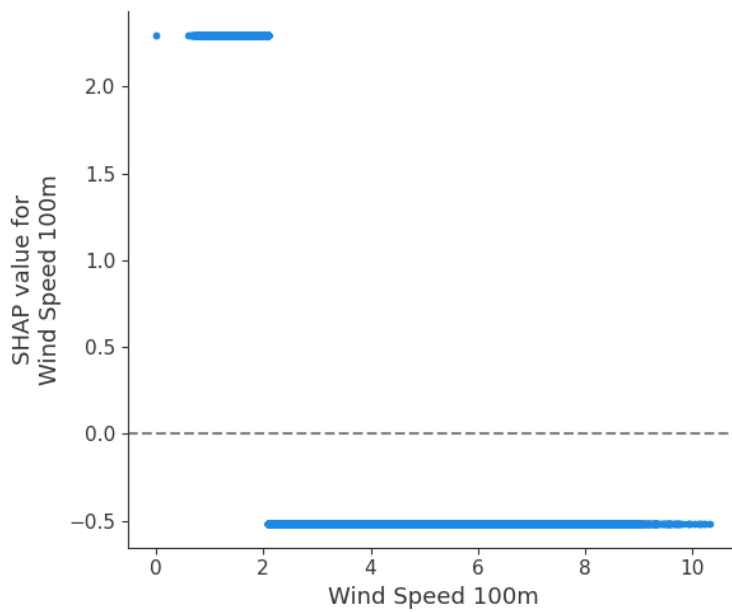
Τέλος, παρατηρείται πως τα κλιματικά δεδομένα, εκτός της μεταβλητής ηλιακής ακτινοβολίας που ήταν άμεσα συνδεδεμένη με την κύρια μεταβλητή και αφαιρέθηκε μετά τον έλεγχο συσχέτισης, δείχνουν πως δεν επηρεάζουν την παραγωγή ηλιακής ενέργειας από φωτοβολταϊκά.



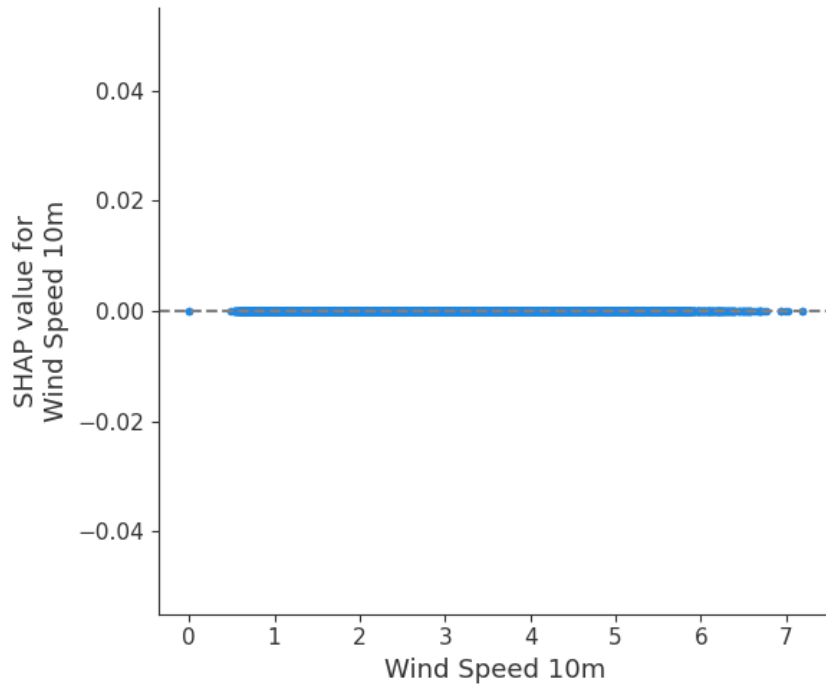
Σχήμα 19: Διάγραμμα παρουσίασης της ερμηνείας των προβλέψεων για τη μεταβλητή “Total Precipitation”.



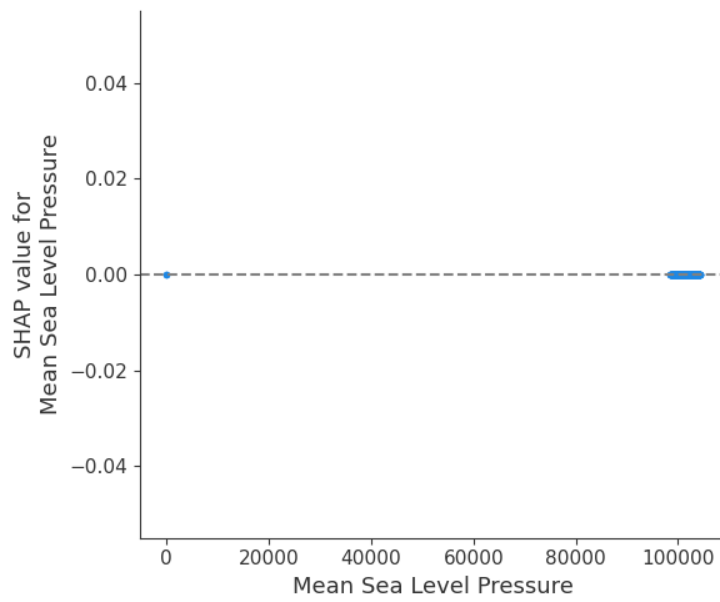
Σχήμα 20: Διάγραμμα παρουσίασης της ερμηνείας των προβλέψεων για τη μεταβλητή “Air Temperature”.



Σχήμα 21: Διάγραμμα παρουσίασης της ερμηνείας των προβλέψεων για τη μεταβλητή “Wind Speed 100m”.



Σχήμα 22: Διάγραμμα παρουσίασης της ερμηνείας των προβλέψεων για τη μεταβλητή “Wind Speed 10m”



Σχήμα 23: Διάγραμμα παρουσίασης της ερμηνείας των προβλέψεων για τη μεταβλητή “Mean Sea Level Pressure”

4. Αξιολόγηση Μοντέλου Μηχανικής Μάθησης

Για την αξιολόγηση των αποτελεσμάτων του μοντέλου πρόβλεψης που υλοποιήθηκε με τον αλγόριθμο πρόβλεψης XGBoost υπολογίστηκαν μια σειρά από στατιστικά μεγέθη τα οποία παρουσιάζονται σε αυτή την ενότητα.

4.1 Mean Absolute Error - MAE

Το μέσο απόλυτο σφάλμα (Mean Absolute Error, MAE) έχει στόχο την αξιολόγηση της απόδοσης του προγνωστικού μοντέλου. Το «Μέσο Απόλυτο Σφάλμα» δείχνει το μέσο μέγεθος των σφαλμάτων μεταξύ των προβλεπόμενων και των πραγματικών τιμών στο σύνολο των δεδομένων. Το MAE παρουσιάζει κατά πόσο οι προβλέψεις που πραγματοποιούνται από το μοντέλο προσεγγίζουν τις πραγματικές τιμές του συνόλου δεδομένων. Όσο χαμηλότερη είναι η τιμή του MAE τόσο πιο ακριβές είναι το μοντέλο. Η μαθηματική έκφραση για το Μέσο Απόλυτο Σφάλμα είναι:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i|$$

- MAE is the Mean Absolute Error.
- n is the number of data points in the test dataset.
- y_i represents the actual target value for the i -th data point.
- \hat{y}_i represents the predicted value for the i -th data point.

Ο κώδικας για τον υπολογισμό του μέσου απόλυτου σφάλματος καθώς και τα αποτελέσματα του παρουσιάζεται παρακάτω. Παρατηρείται πως το Μέσο Απόλυτο Σφάλμα στα δεδομένα ελέγχου (Test Data) είναι 58,79 MW ενώ στα δεδομένα εκπαίδευσης (Train Data) 58,43 MW, δηλαδή τουλάχιστον μία τάξη μεγέθους κάτω από τη μέση τιμή των δεδομένων.

```
#MAE (Mean absolute error)
#difference between the original and predicted values extracted by averaged the absolute difference over the data set.

#test
y_pred=model.predict(X_test).round()
mae_test=mean_absolute_error(y_test,y_pred)

#train
y_pred_train=model.predict(X_train).round()
mae_train=mean_absolute_error(y_train,y_pred_train)

print('Test Data ', '\nMAE (Mean absolute error): ',round(mae_test,2),'MW' )
print('\nTrain Data ', '\nMAE (Mean absolute error): ',round(mae_train,2),'MW' )

Test Data
MAE (Mean absolute error): 58.79 MW

Train Data
MAE (Mean absolute error): 58.43 MW
```

Εικόνα 32: Κώδικας υλοποίησης και αποτελεσμάτων του στατιστικού μεγέθους MAE.

4.2 Median Absolute Error - MedAE

Το διάμεσο απόλυτο σφάλμα (Median Absolute Error, MedAE) είναι παρόμοιο στατιστικό με το Μέσο Απόλυτο Σφάλμα (MAE) με μόνη διαφορά πως αντί του μέσου όρου μετρά το διάμεσο των σφαλμάτων μεταξύ των προβλεπόμενων και των πραγματικών τιμών στο σύνολο δεδομένων. Το Διάμεσο Απόλυτο Σφάλμα επηρεάζεται λιγότερο από τις ακραίες τιμές των δεδομένων και είναι πιο αξιόπιστο μέτρο στα σύνολα δεδομένων που μπορεί να περιέχουν ακραίες τιμές. Η μαθηματική έκφραση για το Διάμεσο Απόλυτο Σφάλμα είναι:

$$MedAE = \text{median}(|y_i - \hat{y}_i|)$$

- *MedAE* is the Median Absolute Error.
- y_i represents the actual target value for the i -th data point.
- \hat{y}_i represents the predicted value for the i -th data point.

Ο κώδικας για τον υπολογισμό του διάμεσου απολύτου σφάλματος καθώς και τα αποτελέσματα του παρουσιάζεται παρακάτω. Παρατηρείται πως το Διάμεσο Απόλυτο Σφάλμα στα δεδομένα ελέγχου (Test Data) είναι 34 MW ενώ στα δεδομένα εκπαίδευσης (Train Data) 34 MW, δηλαδή τουλάχιστον μία τάξη μεγέθους κάτω από τη μέση τιμή των δεδομένων.

```
#MedAE (Median Absolute Error)

#Test
mdae_test = median_absolute_error(y_test, y_pred).round()

#Train
mdae_train = median_absolute_error(y_train, y_pred_train).round()

print('Test Data', '\nMedAE (Median Absolute Error):', mdae_test, 'MW')
print('Train Data', '\nMedAE (Median Absolute Error):', mdae_train, 'MW')

Test Data
MedAE (Median Absolute Error): 34.0 MW
Train Data
MedAE (Median Absolute Error): 34.0 MW
```

Εικόνα 33: Κώδικας υλοποίησης και αποτελεσμάτων του στατιστικού μεγέθους MedAE.

4.3 Root Mean Squared Error (RMSE)

Η τετραγωνική ρίζα στο μέσο τετράγωνο του σφάλματος (RMSE) χρησιμοποιείται για την αξιολόγηση της απόδοσης του μοντέλου πρόβλεψης. Το RMSE είναι μέτρο τυπικού σφάλματος για τις προβλέψεις του μοντέλου και έχει και αυτό, όπως τα προηγούμενα μεγέθη που αναφέρθηκαν, την ίδια μονάδα μέτρησης με τη μεταβλητές πρόβλεψης. Το

RMSE προσδίδει πιο ερμηνεύσιμες και άμεσα συγκρίσιμες τιμές σε σχέση με τις αρχικές μονάδες της μεταβλητής στόχου. Η μαθηματική έκφραση για το RMSE είναι:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}$$

- *RMSE* is the Root Mean Squared Error.
- *n* is the number of data points in the test dataset.
- y_i represents the actual target value for the *i*-th data point.
- \hat{y}_i represents the predicted value for the *i*-th data point.

Επιπλέον, μπορεί να πραγματοποιηθεί κανονικοποίηση του Root Mean Squared Error (RMSE) με στόχο την έκφραση του στατιστικού μεγέθους σε σχέση με το εύρος ή την εξάπλωση της μεταβλητής που στοχεύει η ανάλυση. Η κανονικοποίηση του RMSE βοηθάει στην αξιολόγηση του μοντέλου σε μονάδες ανεξάρτητες της μεταβλητής που στοχεύει το μοντέλο. Με αυτό τον τρόπο γίνεται εύκολη η σύγκριση της απόδοσης διαφορετικών μοντέλων, και παρέχεται μια πιο τυποποιημένη μέτρηση αξιολόγησης. Η μαθηματική έκφραση για το κανονικοποιημένο RMSE είναι:

$$PRMSE = \frac{RMSE}{\max(y) - \min(y)} \times 100$$

- *PRMSE* is the Percentage RMSE.
- *RMSE* is the Root Mean Squared Error.
- *y* is the vector of actual target values.

Ο κώδικας για τον υπολογισμό του RMSE και του PRMSE καθώς και των αποτελεσμάτων τους παρουσιάζονται παρακάτω. Παρατηρείται πως το RMSE στα δεδομένα ελέγχου (Test Data) είναι 84,22 MW ενώ στα δεδομένα εκπαίδευσης (Train Data) 83,63 MW, δηλαδή τουλάχιστον μία τάξη μεγέθους κάτω από τη μέση τιμή των δεδομένων. Το κανονικοποιημένο σφάλμα πρόβλεψης (PRMSE) και στις δυο περιπτώσεις ήταν στο 0.04 (4%).

```

# RMSE (Root Mean Squared Error)

#test
rmse_test=np.sqrt(mean_squared_error(y_test,y_pred))

#Normalize RMSE
test_max = y_test.max()
test_min = y_test.min()
Normalized_rmse_test = rmse_test / (test_max - test_min)
Normalized_rmse_test

#train
rmse_train=np.sqrt(mean_squared_error(y_train,y_pred_train))

#Normalize RMSE
train_max = y_train.max()
train_min = y_train.min()
Normalized_rmse_train = rmse_train / (train_max - train_min)
Normalized_rmse_train

print('Test Data', '\nRMSE (Root Mean Squared Error) :', round(rmse_test,2), ' MW', '\nPRMSE (Normalized RMSE): ',round(Normalized_rmse_test,2), ' MW')
print('\nTrain Data', '\nRMSE (Root Mean Squared Error) :', round(rmse_train,2), ' MW', '\nPRMSE (Normalized RMSE): ',round(Normalized_rmse_train,2), ' MW')

```

Test Data
RMSE (Root Mean Squared Error) : 84.22 MW
PRMSE (Normalized RMSE): 0.04

Train Data
RMSE (Root Mean Squared Error) : 83.63 MW
PRMSE (Normalized RMSE): 0.04

Εικόνα 34: Κώδικας υλοποίησης και αποτελεσμάτων των στατιστικών μεγεθών RMSE και NRMSE.

4.4 R-squared (R^2)

Ο συντελεστής προσδιορισμού (R^2) είναι η στατιστική μέτρηση που χρησιμοποιείται για την αξιολόγηση της καλής προσαρμογής (fitting) ενός μοντέλου. Η τιμή R^2 κυμαίνεται συνήθως μεταξύ 0 και 1 και μετρά το ποσοστό της διακύμανσης στην εξαρτημένη μεταβλητή. Η τιμή 0 υποδεικνύει ότι το μοντέλο δεν εξηγεί καμία από τις διακυμάνσεις στην εξαρτημένη μεταβλητή, ενώ η τιμή 1 υποδεικνύει ότι το μοντέλο εξηγεί τέλεια όλη τη διακύμανση στην εξαρτημένη μεταβλητή. Η μαθηματική έκφραση για το R^2 είναι:

$$R^2 = 1 - \frac{SSR}{SST}$$

- R^2 is the R-squared value, which measures the goodness of fit of the regression model.
- SSR is the sum of squared residuals, which represents the sum of the squared differences between the predicted values and the actual values of the dependent variable. It quantifies the unexplained variation in the dependent variable by the model.
- SST is the total sum of squares, which represents the sum of the squared differences between the actual values of the dependent variable and their mean. It quantifies the total variation in the dependent variable.

Ο κώδικας για τον υπολογισμό του διάμεσου απολύτου σφάλματος καθώς και τα αποτελέσματα του παρουσιάζεται παρακάτω. Παρατηρείται πως το R^2 στα δεδομένα ελέγχου (Test Data) είναι 0.98 (98%) ενώ στα δεδομένα εκπαίδευσης (Train Data) είναι και πάλι 0.98 (98%). Δηλαδή το 98% της εξαρτημένης μεταβλητής εξηγείται από το μοντέλο.

```
#Scikit-Learn's R-squared
from sklearn.metrics import r2_score

#test
R2_test = r2_score(y_test, y_pred)

#train
R2_train = r2_score(y_train, y_pred_train)

print('Test Data', '\nR-squared (Coefficient of determination): ', round(R2_test, 2))
print('\nTrain Data', '\nR-squared (Coefficient of determination): ', round(R2_train, 2))

Test Data
R-squared (Coefficient of determination): 0.98

Train Data
R-squared (Coefficient of determination): 0.98
```

Εικόνα 35: Κώδικας υλοποίησης και αποτελεσμάτων του στατιστικού R^2 .

Οι τιμές των στατιστικών μεγεθών που παρουσιάστηκαν δείχνουν την αποτελεσματικότητα του μοντέλου πρόβλεψης με στόχο την μελέτη της παραγωγής ηλεκτρικής ενέργειας από φωτοβολταϊκά την επόμενη ώρα.

5. Συμπεράσματα

Τα αποτελέσματα δείχνουν πως το μοντέλο πρόβλεψης λειτουργεί ορθά καθώς εντοπίζει με ακρίβεια τις εναλλαγές στην ώρα και εντοπίζει ως κύρια την λογική και αποδεδειγμένη συνθήκη την ώρα ως κύριας μεταβλητή.

Οι χαμηλότερες τιμές για τα στατιστικά μεγέθη MAE, MSE και RMSE υποδεικνύουν ικανοποιητική απόδοση, ενώ η υψηλή τιμή του R^2 επιβεβαιώνει τα αποτελέσματα. Τέλος, η γραφική απεικόνιση φανερώνει αυτό που δείχνουν τα μαθηματικά πως τα αποτελέσματα διαγράφουν σωστή και ικανοποιητική προσέγγιση πάνω στα πραγματικά δεδομένα.

Αξίζει να σημειωθεί πως υπάρχει εθνικό σχέδιο για την ενέργεια στη χώρα που στοχεύει στην αύξηση του συνολικού μεριδίου των ανανεώσιμων πηγών ενέργειας (ΑΠΕ) στην ακαθάριστη τελική κατανάλωση ενέργειας σε 35% έως το 2030 (europa.eu).

Ενδιαφέρον θα παρουσίαζε αντίστοιχη μελέτη θέτοντας ως μεταβλητές στόχους την παραγωγή ενέργειας από τον άνεμο (Wind power generation) και το νερό (Hydro power generation reservoirs and rivers) καθώς και μερική ή ολική σύγκριση τους με την ζήτηση ηλεκτρικής ενέργειας ανά ώρα της χώρας όμως δυστυχώς τα δεδομένα που παρέχει το ευρωπαϊκό πρόγραμμα Copernicus για την Ελλάδα περιορίζονται στην παραγωγή ενέργειας από φωτοβολταϊκά (ηλιακή ενέργεια ανά ώρα). Για τις υπόλοιπες ανανεώσιμες πηγές ενέργειας που υπάρχουν στη χώρα δίνεται μόνο ο συντελεστής χωρητικότητας (Capacity factor ratio) ανά ώρα.

6. Βιβλιογραφία

Ηλεκτρονικές Πηγές

1. Benefits of Renewable Energy Use, Union of Concerned Scientists, (2017), <https://www.ucsusa.org/resources/benefits-renewable-energy-use>
2. Britannica, T. Editors of Encyclopaedia (2023, June 6). hydroelectric power. Encyclopedia Britannica. <https://www.britannica.com/science/hydroelectric-power>
3. Copernicus API – Copernicus, <https://cds.climate.copernicus.eu/api-how-to>
4. Country Rankings 2023, IRENA, [Country Rankings \(irena.org\)](https://www.irena.org/en/country-rankings)
5. EIA. “Annual Energy Outlook 2022.” Washington, D.C.: U.S. Energy Information Administration, March 2022. <https://www.eia.gov/outlooks/aeo/>
6. Energy Efficiency & Renewable Energy, [Office of Energy Efficiency & Renewable Energy | Department of Energy](https://www.energy.gov/eere)
7. EU Space Programme, what is Copernicus? (2023), [What is Copernicus? | EU Agency for the Space Programme \(europa.eu\)](https://www.europa.eu/what-is-copernicus/)
8. Generation of electricity, BBC, <https://www.bbc.co.uk/bitesize/guides/zbsdmp3/revision/4>
9. GES, <https://www.greenenergystorage.eu/>
10. Greensolver “Country Overview, Greece Renewable Energy 2020-2030”, [PowerPoint Presentation \(greendealflow.com\)](https://www.greensolver.com/powerpoint-presentation)
11. <https://www.iea.org/energy-system/renewables/solar-pv>
12. International Energy Agency (IEA), World Energy Statistics and Balances, Greece, [Greece - Countries & Regions - IEA](https://www.iea.org/countries-and-regions/greece)
13. IRENA - International Renewable Energy Agency, [IRENA – International Renewable Energy Agency](https://www.irena.org/)
14. JRC Photovoltaic Geographical Information System (PVGIS) - European Commission ([europa.eu](https://pvgis.jrc.ec.europa.eu/))
15. Jupyter Notebook, [Project Jupyter | Home](https://projectjupyter.com/)
16. Leading countries in installed renewable energy capacity worldwide in 2022, Statista, <https://www.statista.com/statistics/267233/renewable-energy-capacity-worldwide-by-country/>

17. Local Renewable Energy Benefits and Resources, EPA USA,
<https://www.epa.gov/statelocalenergy/local-renewable-energy-benefits-and-resources>
18. Lund, J. W. (2023, July 13). geothermal energy. Encyclopedia Britannica.
<https://www.britannica.com/science/geothermal-energy>
19. National Geographic Society, [Education | National Geographic Society](#)
20. Open Access Hub – Copernicus, <https://scihub.copernicus.eu/>
21. Our World in Data: <https://ourworldindata.org/renewable-energy>
22. Renewable Energy by Country 2023, World Population Review, [Renewable Energy by Country 2023 \(worldpopulationreview.com\)](#)
23. Renewable Energy, Office of Energy Efficiency & Renewable Energy,
<https://www.energy.gov/eere/renewable-energy>
24. Renewable Energy, Online Tutorials.
https://www.tutorialspoint.com/renewable_energy/renewable_energy_introduction.htm
25. Renewable Resources, National Geographic Society,
<https://www.nationalgeographic.org/encyclopedia/renewable-resources/>
26. Selin, N. Eckley (2023, July 20). wind power. Encyclopedia Britannica.
<https://www.britannica.com/science/wind-power>
27. Selin, N. Eckley (2023, July 21). *renewable energy*. Encyclopedia Britannica.
<https://www.britannica.com/science/renewable-energy>
28. Supervised and Unsupervised Machine Learning Algorithms,
<https://machinelearningmastery.com/supervised-and-unsupervised-machine-learning-algorithms/>
29. [Welcome to Python.org](#)
30. What are the pros and cons of renewable energy?, EnergySage, (2022),
<https://news.energysage.com/advantages-and-disadvantages-of-renewable-energy/>
31. What is renewable energy? , United Nations,
<https://www.un.org/en/climatechange/what-is-renewable-energy>.
32. What is Supervised Learning?, IBM. <https://www.ibm.com/topics/supervised-learning>.

33. What is Unsupervised Learning?, IBM.
<https://www.ibm.com/topics/unsupervised-learning>
34. XGBoost Documentation, [XGBoost Documentation — xgboost 2.0.0-dev documentation](#)
35. XGBoost: A Scalable Tree Boosting System, [arXiv:1603.02754](#)
36. [Ανεξάρτητος Διαχειριστής Μεταφοράς Ηλεκτρικής Ενέργειας | ΑΔΜΗΕ \(admie.gr\)](#)
37. [Ελλάδα | Καθαρή ενέργεια για τα νησιά της ΕΕ \(europa.eu\)](#)
38. Ελληνική Επιστημονική Ένωση Αιολικής Ενέργειας, [Οι Αιολικές Εγκαταστάσεις στην Ελλάδα | HWEA Wind Statistics – ΕΛΕΤΑΕΝ \(eletaen.gr\)](#)
39. Ρυθμιστική Αρχή Ενέργειας Ελλάδας (PAEE), [Rae Website](#)
40. Υπουργείο Περιβάλλοντος και Ενέργειας, Ηλιακή/ Φωτοβολταϊκά, [Ηλιακή/ Φωτοβολταϊκά - \(ypen.gov.gr\)](#)

Διεθνής Βιβλιογραφία

A. Natekin and A. Knoll, “Gradient Boosting Machines, A Tutorial,” *Frontiers in neurorobotics*, vol. 7, p. 21, Jun. 2013, doi: 10.3389/fnbot.2013.00021.

Andronie, M., Simion, V.E., Gurgu, E., Dijmǎrescu, A., Dijmǎrescu, I., & Bucharest, R.P. (2019). Social Responsibility of Firms and the Impact of Bio-Economy in Intelligent Use of Renewable Energy Source. www.amfiteatrueconomic.ro.

Bahadori, A., Zahedi, G., & Zendehboudi, S. (2013). An overview of Australia's hydropower energy: Status and future prospects. *Renewable & Sustainable Energy Reviews*, 20, 565-569.

Barbier, E. (2002). Geothermal energy technology and current status: an overview. *Renewable & Sustainable Energy Reviews*, 6, 3-65.

Bisong, E. (2019). What Is Machine Learning? *Building Machine Learning and Deep Learning Models on Google Cloud Platform*.

Blaabjerg, F., & Ma, K. (2017). Wind Energy Systems. *Proceedings of the IEEE*, 105, 2116-2131.

Chen, T., & Guestrin, C. (2015). XGBoost : Reliable Large-scale Tree Boosting System.

Chen, T., & Guestrin, C. (2016). XGBoost: A Scalable Tree Boosting System. *Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*.

Christopher, K., & Dimitrios, R. (2012). A review on exergy comparison of hydrogen production methods from renewable energy sources. *Energy and Environmental Science*, 5, 6640-6651.

Czernik, A., & Bridgwater, A.V. (2004). Overview of Applications of Biomass Fast Pyrolysis Oil. *Energy & Fuels*, 18, 590-598.

Gahlan, N. (2020). Assessment of Supervised Machine Learning Algorithms.

Gordon, D.H. (1986). Renewable Energy Resources.

Guo, Rui & Zhao, Zhiqian & Wang, Tao & Liu, Guangheng & Zhao, Jingyi & Gao, Dianrong. (2020). Degradation state recognition of piston pump based on ICEEMDAN and XGBoost. *Applied Sciences*. 10. 6593. 10.3390/app10186593.

Herbert, G.M., Iniyan, S., Sreevalsan, E., & Rajapandian, S. (2007). A review of wind energy technologies. *Renewable & Sustainable Energy Reviews*, 11, 1117-1145.

Householder, D.L., & Marshall, J.A. (1984). An introduction to solar energy. *Industrial Education*, 73, 16-18.

IRENA (2022), Renewable Energy Statistics 2022, The International Renewable Energy Agency, Abu Dhabi

Jordan, M.I., & Mitchell, T. (2015). Machine learning: Trends, perspectives, and prospects. *Science*, 349, 255 - 260.

Kalmikov, A.G. (2017). Wind Power Fundamentals.

Kaunda, C.S., Kimambo, C.Z., & Nielsen, T.K. (2012). Hydropower in the Context of Sustainable Energy Supply: A Review of Technologies and Challenges. *International Scholarly Research Notices*, 2012, 1-15.

Kaygusuz, K. (2016). Hydropower as clean and renewable energy source for electricity generation. *The Journal of Engineering Research*, 5, 359-369.

Liu, B., Lee, W.S. (2011). Partially Supervised Learning. In: Web Data Mining. Data-Centric Systems and Applications. Springer, Berlin, Heidelberg. https://doi.org/10.1007/978-3-642-19460-3_5

Love, B.C. (2002). Comparing supervised and unsupervised category learning. *Psychonomic Bulletin & Review*, 9, 829-835.

Naqa, I.M., & Murphy, M.J. (2017). Machine Learning. *Industry, Innovation and Infrastructure*.

Newman, R. (2014). Promotion of the use of energy from renewable sources.

Rosillo-calle, F. (2016). A review of biomass energy – shortcomings and concerns. *Journal of Chemical Technology & Biotechnology*, 91, 1933-1945.

Santhiarsa, I.N., & Kusuma, I.G. (2005). Kajian energi surya untuk pembangkit tenaga listrik.

Shaveta (2023). A review on machine learning. *International Journal of Science and Research Archive*.

Singh, P. (2019). Unsupervised Machine Learning. Learn PySpark.

Solís, J.M., Franquelo, L.G., Bialasiewicz, J.T., Galván, E., Guisado, R.P., Prats, M.Á., Galván, J.I., & Moreno-Alfonso, N. (2006). Power-Electronic Systems for the Grid Integration of Renewable Energy Sources: A Survey. *IEEE Transactions on Industrial Electronics*, 53, 1002-1016.

Teknik, D.F. (2013). Pemanfaatan energi surya sebagai suplai cadangan pada laboratorium elektro dasar di institut teknologi padang.

Wang, Yuanchao & Pan, Z. & Zheng, J. & Qian, L. & Mingtao, Li. (2019). A hybrid ensemble method for pulsar candidate classification. *Astrophysics and Space Science*. 364. 10.1007/s10509-019-3602-4.

Wilson, S. (2021). Solarities or Solarculture. *South Atlantic Quarterly*, 120, 137-150.

Zhang, R., Li, B., & Jiao, B. (2019). Application of XGboost Algorithm in Bearing Fault Diagnosis. *IOP Conference Series: Materials Science and Engineering*, 490.

Zhou, Z., & Schwenker, F. (2013). Partially Supervised Learning. *Lecture Notes in Computer Science*.

7. Παράρτημα

Ο κώδικας που αναπτύχθηκε βρίσκεται στον ιστότοπο GitHub:

<https://github.com/Grapadoa/UOM-Thesis-2023.git>