



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
& ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΤΙΤΛΟΣ ΕΡΓΑΣΙΑΣ :

ΕΛΕΓΧΟΣ ΧΡΟΝΟΣΕΙΡΩΝ ΤΗΣ ΒΑΣΗΣ ΔΕΔΟΜΕΝΩΝ
ΤΟΥ ΖΩΝΤΑΝΟΥ ΕΡΓΑΣΤΗΡΙΟΥ ΤΟΥ ΠΑΝΕΠΙΣΤΗΜΙΟΥ
ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της Ιωαννίδου Σοφίας

Επιβλέπων: Γεώργιος Χριστοφορίδης

Συνεπιβλέπων: Υποψήφιος Διδάκτωρ, Σταύρος Φιλιππίδης

ΚΟΖΑΝΗ/ΟΚΤΩΒΡΙΟΣ/2023



HELLENIC DEMOCRACY
UNIVERSITY OF WESTERN MACEDONIA
SCHOOL OF ENGINEERING
DEPARTMENT OF ELECTRICAL
& COMPUTER ENGINEERING

TITLE:

TIME SERIES ANALYSIS OF THE UNIVERCITY OF
WESTERN MACEDONIA'S LIVING LAB DATABASE

THESIS

Sofia Ioannidou

SUPERVISOR: Prof. Georgios Christoforidis

CO-SUPERVISOR: PhD Student Stavros Filippidis

KOZANI/OCTOBER/2023

Περίληψη:

Η παρούσα διπλωματική εργασία έχει ως σκοπό την αξιολόγηση της ποιότητας των δεδομένων χρονοσειρών του Ζωντανού εργαστηρίου του Πανεπιστημίου Δυτικής Μακεδονίας με τη χρήση της Python. Πρωταρχικός στόχος είναι να διασφαλιστεί η αξιοπιστία και η ακεραιότητα των δεδομένων που συλλέγονται από διαφορετικές πηγές. Το πλαίσιο χρησιμοποιεί διάφορες βιβλιοθήκες Python για την εκτέλεση μιας σειράς ελέγχων ποιότητας δεδομένων, που περιλαμβάνουν την ανίχνευση ελλιπών τιμών, τον εντοπισμό ακραίων τιμών, την εξομάλυνση δεδομένων, την οπτικοποίηση και την ανίχνευση ανωμαλιών.

Το πλαίσιο ξεκινά με τη φόρτωση δεδομένων χρονοσειρών από ένα αρχείο Excel σε ένα πλαίσιο δεδομένων Pandas DataFrame. Στη συνέχεια, προχωρά στον εντοπισμό και την ποσοτικοποίηση των ελλιπών τιμών, τον εντοπισμό διπλών χρονοσφραγίδων και την επικύρωση της χρονολογικής σειράς των χρονοσφραγίδων. Ο κώδικας περιλαμβάνει επίσης ανίχνευση ακραίων τιμών με βάση το Z-Score και επικύρωση εύρους δεδομένων για μετρικές στήλες. Η εξομάλυνση των δεδομένων επιτυγχάνεται μέσω του φίλτρου Savitzky-Golay και το πλαίσιο προσφέρει οπτικοποιήσεις για κάθε μετρική στήλη. Επιπλέον, εισάγει έναν μηχανισμό ανίχνευσης ανωμαλιών με βάση το προφίλ του πίνακα χρησιμοποιώντας το Stumpy.

Το πλαίσιο είναι ευπροσάρμοστο και προσαρμόζεται σε προσαρμοσμένους ελέγχους, επιτρέποντας τη συμπερίληψη κανόνων για συγκεκριμένους τομείς και πρόσθετων ελέγχων προσαρμοσμένων σε συγκεκριμένες απαιτήσεις. Εν κατακλείδι, η διατριβή παρέχει μια σύνοψη της ποιότητας των δεδομένων, συμπεριλαμβανομένου του αριθμού των γραμμών και των στηλών στο σύνολο δεδομένων, προσφέροντας μια σαφή επισκόπηση της ακεραιότητας των δεδομένων.

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ : Χρονοσειρές , Python, Ανάλυση , Ζωντανό Εργαστήριο , Πανεπιστήμιο Δυτικής Μακεδονίας , Αλγόριθμος .

Abstract:

This thesis aims at assessing the quality of time-series data of the Living Lab of the University of Western Macedonia using Python. The primary objective is to ensure the reliability and integrity of data collected from diverse sources. The framework utilizes various Python libraries to execute a range of data quality checks, encompassing missing value detection, outlier identification, data smoothing, visualization, and anomaly detection. The framework begins with loading time-series data from an Excel file into a Pandas DataFrame. It then proceeds to identify and quantify missing values, detect duplicate timestamps, and validate the chronological order of timestamps. The code also includes Z-Score-based outlier detection and data range validation for metric columns. Data smoothing is achieved through the Savitzky-Golay filter, and the framework offers visualizations for each metric column. Additionally, it introduces an anomaly detection mechanism based on the matrix profile using Stumpy. The framework is adaptable and accommodates custom checks, allowing for the inclusion of domain-specific rules and additional checks tailored to specific requirements. In conclusion, the thesis provides a data quality summary, including the count of rows and columns in the dataset, offering a clear overview of data integrity.

KEYWORDS: Time series , Python , Analysis , Living Lab ,
University of Western Macedonia , Algorithm .

Πίνακας Περιεχομένων

Περίληψη	<i>Error! Bookmark not defined.</i>
Abstract	<i>Error! Bookmark not defined.</i>
Κεφάλαιο 1: Εισαγωγή στην Python	<i>Error! Bookmark not defined.</i>
1.1 Εγκατάσταση της γλώσσας προγραμματισμού python	<i>Error! Bookmark not defined.</i>
Κεφάλαιο 2: Εισαγωγή στις χρονοσειρές	<i>Error! Bookmark not defined.</i>
2.1 Μοντελοποίηση χρονοσειρών	<i>Error! Bookmark not defined.</i>
2.2 Χρονοσειρές - Βαθμονόμηση παραμέτρων	<i>Error! Bookmark not defined.</i>
2.3 Εφαρμογές των χρονοσειρών	<i>Error! Bookmark not defined.</i>
Κεφάλαιο 3: Το ζωντανό εργαστήριο του Πανεπιστημίου Δυτικής Μακεδονίας	<i>Error! Bookmark not defined.</i>
3.1 Το ζωντανό εργαστήριο του Πανεπιστημίου Δυτικής Μακεδονίας (SMART).....	<i>Error! Bookmark not defined.</i>
3.2 Εξοπλισμός του ζωντανού εργαστηρίου	<i>Error! Bookmark not defined.</i>
Κεφάλαιο 4: Δημιουργία αλγορίθμου για τον έλεγχο χρονοσειρών από το ζωντανό εργαστήριο	<i>Error! Bookmark not defined.</i>
4.1 Παρουσίαση αρχείων χρονοσειρών και δημιουργία αρχείου αλγορίθμου.....	<i>Error! Bookmark not defined.</i>
4.2 Παρουσίαση της δομής του αλγορίθμου	<i>Error! Bookmark not defined.</i>
4.3 Παρουσίαση αποτελεσμάτων	<i>Error! Bookmark not defined.</i>
Παράρτημα κώδικα:	<i>Error! Bookmark not defined.</i>
Κώδικας έλεγχου χρονοσειρών.	<i>Error! Bookmark not defined.</i>
Κώδικας μορφοποίησης αποτελεσμάτων σε πίνακα	<i>Error! Bookmark not defined.</i>
Βιβλιογραφία	43

Πίνακας Εικόνων :

Figure 1 Εγκατάσταση της γλώσσας προγραμματισμού Python 3.7.2 - Επιλογή "Add Python 3.7 to Path.....	10
Figure 2 Εκτύπωση της έκδοσης της Python σε τερματικό.....	10
Figure 3 Εγκατάσταση βιβλιοθήκης matplotlib και βιβλιοθήκης pandas	11
Figure 4 Κάτοψη κτηρίου με τα φωτοβολταϊκά πάνελ και τη διάταξη των τμημάτων του κτηρίου της δημόσιας πιλοτικής εγκατάστασης στην Κοζάνη.....	21
Figure 5 Υβριδικός αντιστροφέας με τα βασικά χαρακτηριστικά του.....	22
Figure 6 Σύστημα μπαταριών με τα βασικά χαρακτηριστικά του.....	22
Figure 7 Μετρητής ισχύος με τα βασικά χαρακτηριστικά του.....	22
Figure 8 Επισκόπηση πιλοτικού συστήματος BESS (τοπολογία σύζευξης DC)	23
Figure 9 Φωτογραφίες εγκατάστασης (εσωτερική εγκατάσταση και φωτοβολταϊκά πάνελ στην οροφή)	24
Figure 10 Διάγραμμα κατανάλωσης πρώτης φάσης φωτισμού.....	34
Figure 11 Διάγραμμα πραγματικής ισχύος δεύτερης βάσης για το πρώτο υποσύστημα Inverter.....	34
Figure 12 Διάγραμμα παροχής πραγματικής ισχύος για τις φάσεις L1 - L2 για το δεύτερο υποσύστημα Inverter.....	35
Figure 13 Διάγραμμα παροχής πραγματικής ισχύος για την πρώτη φάση του τρίτου υποσυστήματος Inverter.....	35
Figure 14 Διάγραμμα παραγωγής ισχύος σε ένα από τα δυο string του πρώτου φωτοβολταϊκού συστήματος.....	36
Figure 15 Διάγραμμα κατανάλωσης πραγματικής ισχύος πρώτης φάσης στην κατανάλωση των πλυντηρίων του ζωντανού εργαστηρίου.....	36
Figure 16 Διάγραμμα κατανάλωσης πραγματικής ισχύος της δεύτερης φάσης για ένα τμήμα της πρώτης πτέρυγας.....	37

Κεφάλαιο 1: Εισαγωγή στην Python

Η Python είναι μια γλώσσα προγραμματισμού με απλό συντακτικό, εξαιρετική αναγνωσιμότητα, φορητότητα και μοντέρνα χαρακτηριστικά που την κάνουν κατάλληλη ως πρώτη γλώσσα προγραμματισμού. Ο κώδικας μιας τέτοιας γλώσσας πρέπει να μετατραπεί σε "γλώσσα μηχανής" ώστε να εκτελεστεί από τον Η/Υ. Η επεξεργασία αυτή γίνεται από διερμηνευτές και μεταγλωττιστές. Στην περίπτωση της python η επεξεργασία γίνεται από διερμηνευτή. Η γλώσσα αυτή γράφτηκε από τον Ολλανδό προγραμματιστή Guido van Rossum στα τέλη της δεκαετίας 1980-90. [1] Η βασική κατανόηση οποιασδήποτε γλώσσας προγραμματισμού είναι απαραίτητη για έναν χρήστη που εργάζεται ή αναπτύσσει προβλήματα μηχανικής μάθησης. Παρακάτω παρουσιάζονται συνοπτικά πληροφορίες για μερικές από τις γνωστές γλώσσες προγραμματισμού:

- **Python:** Είναι μια υψηλού επιπέδου διερμηνευμένη γλώσσα προγραμματισμού, γρήγορη και εύκολη στον κώδικα. Η Python μπορεί να ακολουθήσει είτε διαδικαστικά είτε αντικειμενοστραφή παραδείγματα προγραμματισμού. Η παρουσία μιας ποικιλίας βιβλιοθηκών καθιστά την υλοποίηση πολύπλοκων διαδικασιών απλούστερη.
- **R:** Παρόμοια με την Python, η R είναι μια διερμηνευμένη γλώσσα πολλαπλών παραδειγμάτων, η οποία υποστηρίζει στατιστικούς υπολογισμούς και γραφικά. Η ποικιλία των πακέτων διευκολύνει την υλοποίηση μοντελοποίησης μηχανικής μάθησης στην R.
- **Java:** Πρόκειται για μια ερμηνευόμενη αντικειμενοστραφής γλώσσα προγραμματισμού, η οποία είναι ευρέως γνωστή για το μεγάλο εύρος διαθεσιμότητας πακέτων και τις εξελιγμένες τεχνικές οπτικοποίησης δεδομένων.
- **C/C++:** Πρόκειται για μεταγλωττισμένες γλώσσες και δύο από τις παλαιότερες γλώσσες προγραμματισμού. Αυτές οι γλώσσες προτιμώνται συχνά για την ενσωμάτωση δυνατοτήτων ML στις ήδη

υπάρχουσες εφαρμογές, καθώς επιτρέπουν την εύκολη προσαρμογή της υλοποίησης των αλγορίθμων ML".

- **MATLAB:** Το MATrix LABoratory είναι μια γλώσσα πολλαπλών παραδειγμάτων που δίνει τη δυνατότητα εργασίας με πίνακες. Επιτρέπει μαθηματικές πράξεις για πολύπλοκα προβλήματα. Χρησιμοποιείται κυρίως για αριθμητικές πράξεις, αλλά ορισμένα πακέτα επιτρέπουν επίσης τη γραφική προσομοίωση πολλαπλών πεδίων και τον σχεδιασμό βάσει μοντέλου.

Άλλες προτιμώμενες γλώσσες προγραμματισμού για προβλήματα μηχανικής μάθησης είναι η JavaScript, η LISP, η Prolog, η SQL, η Scala, η Julia, η SAS κ.λπ.

Η **Python** χρησιμοποιείται από τους χρήστες λόγω της εύκολης στη συγγραφή και εύληπτης δομής του κώδικά της, καθώς και της μεγάλης ποικιλίας βιβλιοθηκών ανοικτού κώδικα.

Μια βιβλιοθήκη είναι μια συλλογή από ήδη γραμμένες διαδικασίες και συναρτήσεις που μπορούν να επαναχρησιμοποιηθούν από ένα πρόγραμμα. Μια βιβλιοθήκη μας απαλλάσσει από το έργο της γραφής διαδικασιών και συναρτήσεων από το μηδέν. Μερικές από αυτές τις βιβλιοθήκες ανοικτού κώδικα παρουσιάζονται παρακάτω:

- **NumPy:** Η Numerical Python είναι μια βιβλιοθήκη που χρησιμοποιείται για επιστημονικούς υπολογισμούς. Λειτουργεί σε ένα αντικείμενο N-διάστατου πίνακα και παρέχει βασικές μαθηματικές λειτουργίες όπως μέγεθος, σχήμα, μέση τιμή, τυπική απόκλιση, ελάχιστο, μέγιστο καθώς και ορισμένες πιο σύνθετες λειτουργίες όπως γραμμικές αλγεβρικές συναρτήσεις και μετασχηματισμό Fourier.
- **Pandas:** Αυτή η βιβλιοθήκη παρέχει εξαιρετικά αποδοτικές και εύχρηστες δομές δεδομένων, όπως σειρές, πλαίσια δεδομένων και πίνακες. Έχει βελτιώσει τη λειτουργικότητα της Python από την απλή συλλογή και προετοιμασία δεδομένων μέχρι την ανάλυση δεδομένων. Οι δύο βιβλιοθήκες, Pandas και NumPy, καθιστούν πολύ απλή οποιαδήποτε λειτουργία σε μικρά έως πολύ μεγάλα σύνολα δεδομένων.

- **SciPy:** Η Science Python είναι μια βιβλιοθήκη που χρησιμοποιείται για επιστημονικούς και τεχνικούς υπολογισμούς. Παρέχει λειτουργίες για βελτιστοποίηση, επεξεργασία σήματος και εικόνας, ολοκλήρωση, παρεμβολή και γραμμική άλγεβρα. Αυτή η βιβλιοθήκη είναι χρήσιμη κατά την εκτέλεση μηχανικής μάθησης.
- **Scikit Learn:** Αυτή η βιβλιοθήκη είναι μια εργαλειοθήκη SciPy Toolkit που χρησιμοποιείται ευρέως για στατιστική μοντελοποίηση, μηχανική μάθηση και βαθιά μάθηση, καθώς περιέχει διάφορα προσαρμόσιμα μοντέλα παλινδρόμησης, ταξινόμησης και ομαδοποίησης. Συνεργάζεται καλά με τις βιβλιοθήκες Numpy, Pandas και άλλες βιβλιοθήκες, γεγονός που διευκολύνει τη χρήση της.
- **Statsmodels:** Όπως το Scikit Learn, αυτή η βιβλιοθήκη χρησιμοποιείται για την εξερεύνηση στατιστικών δεδομένων και τη στατιστική μοντελοποίηση. Λειτουργεί επίσης καλά με άλλες βιβλιοθήκες της Python.
- **Matplotlib:** Αυτή η βιβλιοθήκη χρησιμοποιείται για την οπτικοποίηση δεδομένων σε διάφορες μορφές, όπως γραμμικό διάγραμμα, ραβδόγραμμα, χάρτες θερμότητας, διαγράμματα διασποράς, ιστόγραμμα κ.λπ. Περιέχει όλες τις λειτουργικότητες που σχετίζονται με γραφήματα που απαιτούνται από την απεικόνιση έως την επισήμανση.

Μια άλλη βιβλιοθήκη, ιδιαίτερα σημαντική, εκτός από αυτές που συζητήθηκαν παραπάνω, για την αντιμετώπιση χρονοσειρών είναι η Datetime.

Αυτή η βιβλιοθήκη, με τις δύο ενότητες της datetime και calendar, παρέχει όλες τις απαραίτητες λειτουργίες datetime για την ανάγνωση, τη μορφοποίηση και τον χειρισμό του χρόνου.

1.1 Εγκατάσταση της γλώσσας προγραμματισμού python

Σε αυτή την παράγραφο παρουσιάζεται ο τρόπος εγκατάστασης της γλώσσας προγραμματισμού python. Η Python μπορεί να ληφθεί μέσω της ιστοσελίδας: <https://www.python.org/downloads/>.

Εάν η εγκατάσταση θα γίνει σε υπολογιστή με Windows, είναι σημαντικό η επιλογή "add python to path" να επιλεγεί, ώστε οι εφαρμογές που χρησιμοποιούν τη γλώσσα να ανιχνεύουν αυτόματα τα αρχεία της.

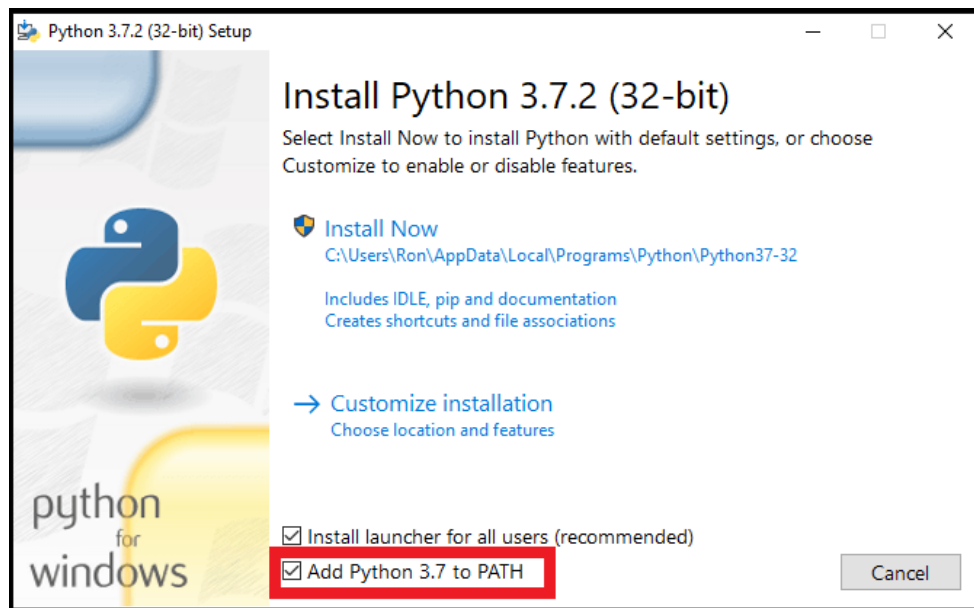


Figure 1 Εγκατάσταση της γλώσσας προγραμματισμού Python 3.7.2 - Επιλογή "Add Python 3.7 to Path".

Μετά την εγκατάσταση, η εγκατάσταση μπορεί να επαληθευτεί με την εκκίνηση ενός διερμηνέα γραμμής εντολών, όπως το CMD στα Windows, το terminal σε MacOS ή linux, και πληκτρολογώντας την εντολή: «python3 -V». Εάν η python έχει εγκατασταθεί επιτυχώς, το τερματικό θα εμφανίσει την εγκατεστημένη της έκδοση:

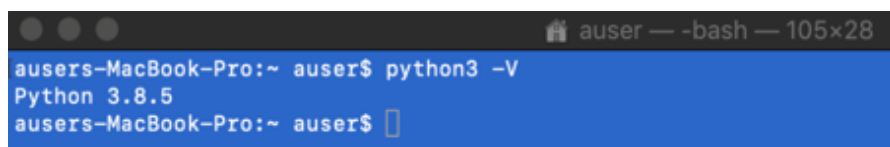


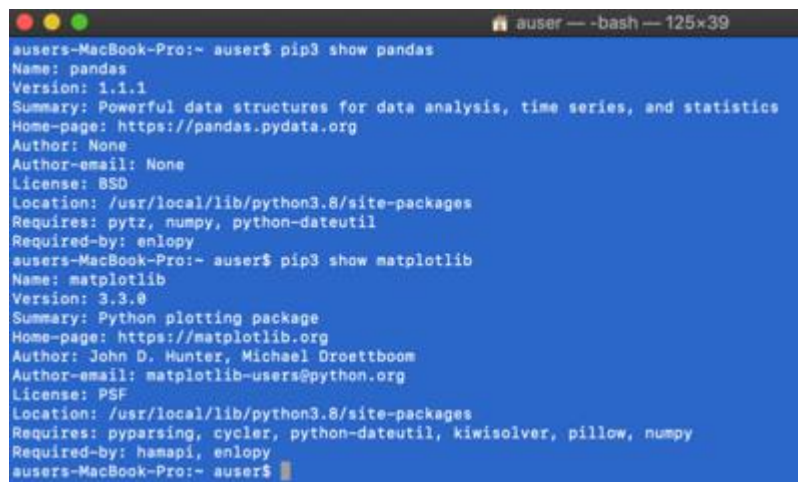
Figure 2 Εκτύπωση της έκδοσης της Python σε τερματικό

Ο υπολογιστής πλέον μπορεί να εκτελεί αρχεία της γλώσσας προγραμματισμού python με κατάλιξη «.py». Υπάρχουν πολλαπλοί τρόποι επεξεργασίας και εκτέλεσης αρχείων δέσμης ενεργειών. Ένας απλός τρόπος είναι να χρησιμοποιήσετε τον ενσωματωμένο επεξεργαστή κειμένου του υπολογιστή (Notepad, Textedit) και να εκτελέσετε το αρχείο .py μεταβαίνοντας στη διαδρομή του και πληκτρολογώντας: "python3 yourfilename.py". Μια πιο τυπική προσέγγιση είναι η χρήση ενός ειδικού λογισμικού κωδικοποίησης, όπως το Visual Studio Code της Microsoft (VSC) (<https://code.visualstudio.com>), το Atom ή το

GNOME builder. Στην παρούσα πτυχιακή εργασία χρησιμοποιείται τον Visual Studio Code της Microsoft.

Πριν από την εκκίνηση του VSC, πρέπει να εγκατασταθούν ορισμένες βιβλιοθήκες της python. Η Python διαθέτει τον δικό της διαχειριστή πακέτων, ο οποίος μπορεί εύκολα να χρησιμοποιηθεί χρησιμοποιώντας έναν διερμηνέα γραμμής εντολών. Εκκινήστε ένα παράθυρο τερματικού και εγκαταστήστε τις βιβλιοθήκες matplotlib και pandas χρησιμοποιώντας:

- pip3 install pandas
- pip3 install matplotlib



```
ausers-MacBook-Pro:~ auser$ pip3 show pandas
Name: pandas
Version: 1.1.1
Summary: Powerful data structures for data analysis, time series, and statistics
Home-page: https://pandas.pydata.org
Author: None
Author-email: None
License: BSD
Location: /usr/local/lib/python3.8/site-packages
Requires: pytz, numpy, python-dateutil
Required-by: enlopy
ausers-MacBook-Pro:~ auser$ pip3 show matplotlib
Name: matplotlib
Version: 3.3.0
Summary: Python plotting package
Home-page: https://matplotlib.org
Author: John D. Hunter, Michael Droettboom
Author-email: matplotlib-users@python.org
License: PSF
Location: /usr/local/lib/python3.8/site-packages
Requires: pyparsing, cycler, python-dateutil, kiwisolver, pillow, numpy
Required-by: hamapi, enlopy
ausers-MacBook-Pro:~ auser$
```

Figure 3 Εγκατάσταση βιβλιοθήκης matplotlib και βιβλιοθήκης pandas

Στην παρούσα πτυχιακή οι παραπάνω δύο βιβλιοθήκες "pandas" και "matplotlib" χρησιμοποιούνται για τον έλεγχο χρονοσειρών. [3]

Κεφάλαιο 2: Εισαγωγή στις χρονοσειρές

Μια χρονοσειρά είναι μια ακολουθία παρατηρήσεων κατά τη διάρκεια μιας ορισμένης περιόδου. Μια μονομεταβλητή χρονοσειρά αποτελείται από τις τιμές που λαμβάνει μια μεταβλητή σε περιοδικές χρονικές στιγμές κατά τη διάρκεια μιας περιόδου, ενώ μια πολυμεταβλητή χρονοσειρά αποτελείται από τις τιμές που λαμβάνουν πολλές μεταβλητές στις ίδιες περιοδικές χρονικές στιγμές κατά τη διάρκεια μιας περιόδου. Το απλούστερο κατά τη διάρκεια της ημέρας ή της εβδομάδας ή του μήνα ή του έτους. Η ανάλυση των χρονικών δεδομένων είναι ικανή να μας δώσει χρήσιμες πληροφορίες για το πώς μια μεταβλητή μεταβάλλεται με την πάροδο του χρόνου ή πώς εξαρτάται από τη

παράδειγμα χρονοσειράς που συναντάμε σε καθημερινή βάση είναι η μεταβολή της θερμοκρασίας μεταβολή των τιμών μιας ή περισσότερων άλλων μεταβλητών.

Τα τελευταία χρόνια παρατηρήθηκε η ευρεία εφαρμογή της στατιστικής και της μηχανικής μάθησης για την εξαγωγή αξιοποιήσιμων πληροφοριών και επιχειρηματικής αξίας από τα δεδομένα σε όλους σχεδόν τομείς. Συνεπώς, γίνεται επιτακτική ανάγκη για τους αναλυτές επιχειρήσεων και τους επαγγελματίες λογισμικού να είναι σε θέση να αντιμετωπίζουν διαφορετικούς τύπους συνόλων δεδομένων. Συχνά, τα δεδομένα είναι μια χρονοσειρά με τη μορφή μιας ακολουθίας ποσοτικών παρατηρήσεων σχετικά με ένα σύστημα ή μια διαδικασία και γίνονται σε διαδοχικά χρονικά σημεία. Συνήθως, τα χρονικά σημεία απέχουν μεταξύ τους ίσες αποστάσεις. [4]

Η ανάλυση χρονοσειρών αποτελείται από δύο βασικά βήματα:

- Οπτικοποίηση δεδομένων για τον έλεγχο της παρουσίας τάσεων, εποχικότητας και κυκλικών προτύπων.
- Προσαρμογή της τάσης και της εποχικότητας για τη δημιουργία στάσιμων σειρών

Η δημιουργία στάσιμων δεδομένων είναι σημαντική για την ενίσχυση του μοντέλου πρόβλεψης χρονοσειρών. Η αφαίρεση της τάσης, της εποχικότητας και των κυκλικών συνιστωσών θα μας άφηνε με ακανόνιστες διακυμάνσεις οι οποίες δεν μπορούν να μοντελοποιηθούν χρησιμοποιώντας μόνο τον δείκτη χρόνου ως επεξηγηματική μεταβλητή. Επομένως, για να βελτιωθεί περαιτέρω η πρόβλεψη, οι ακανόνιστες διακυμάνσεις θεωρούνται ανεξάρτητες και ταυτόσημα κατανομημένες παρατηρήσεις και μοντελοποιούνται με γραμμική παλινδρόμηση σε μεταβλητές εκτός του δείκτη χρόνου.

Η στασιμοποίηση μιας χρονοσειράς, είναι ένα σημαντικό βήμα για την ανάπτυξη του μοντέλου γραμμικής παλινδρόμησης σε εξωγενείς μεταβλητές. Αυτό οφείλεται στο γεγονός ότι υπάρχουν καθιερωμένες στατιστικές μέθοδοι, για παράδειγμα το κεντρικό οριακό θεώρημα, η μέθοδος των ελαχίστων τετραγώνων κ.ο.κ..

Παραδείγματα χρονοσειρών δεδομένων περιλαμβάνουν το ακαθάριστο εγχώριο προϊόν, τον όγκο πωλήσεων, τις τιμές των μετοχών, τα χαρακτηριστικά του καιρού όταν καταγράφονται σε μια χρονική διασπορά

αρκετών ετών, μηνών, ημερών, ωρών κ.ο.κ. Η συχνότητα παρατήρησης εξαρτάται από τη φύση της μεταβλητής και τις εφαρμογές της. Ο όγκος των πωλήσεων δημοσιεύεται κάθε μήνα, τρίμηνο ή διετία, αν και τα στοιχεία για μεγαλύτερη χρονική διάρκεια μπορεί να έχουν παραχθεί με τη συγκέντρωση πιο λεπτομερών δεδομένων, όπως οι ημερήσιες ή εβδομαδιαίες πωλήσεις. Στο άλλο άκρο, υπάρχουν αρκετές φυσικές διεργασίες που παράγουν δεδομένα χρονοσειρών σε κλάσματα του δευτερολέπτου.

Η επιτυχής αξιοποίηση των δεδομένων χρονοσειρών θα οδηγήσει στην παρακολούθηση της υγείας του συστήματος με την πάροδο του χρόνου. Για παράδειγμα, η απόδοση μιας εταιρείας παρακολουθείται από τα τριμηνιαία περιθώρια κέρδους της. Η ανάλυση χρονοσειρών αποσκοπεί στην αξιοποίηση τέτοιων δεδομένων για διάφορους σκοπούς που μπορούν να κατηγοριοποιηθούν σε γενικές γραμμές ως εξής:

- Για την κατανόηση και ερμηνεία των υποκείμενων δυνάμεων που παράγουν την παρατηρούμενη κατάσταση ενός συστήματος ή μιας διαδικασίας με την πάροδο του χρόνου .
- Για την πρόβλεψη της μελλοντικής κατάστασης του συστήματος ή της διαδικασίας από την άποψη των παρατηρήσιμων χαρακτηριστικών .

Η ανάλυση χρονολογικών σειρών εφαρμόζει διάφορες στατιστικές μεθόδους για τη διερεύνηση και τη μοντελοποίηση των εσωτερικών δομών των δεδομένων χρονολογικών σειρών, για την πραγματοποίηση των προαναφερθέντων στόχων, όπως τάσεις, εποχιακές διακυμάνσεις, κυκλική συμπεριφορά και ακανόνιστες μεταβολές. Για τον αποτελεσματικό σχεδιασμό προγραμμάτων υπολογιστών που μπορούν να διερευνήσουν, να απεικονίσουν και να μοντελοποιήσουν πρότυπα σε δεδομένα χρονοσειρών, υπάρχουν διάφορες μαθηματικές τεχνικές και εργαλεία προγραμματισμού. [5]

Οι επιχειρηματικοί αναλυτές και οι επιστήμονες δεδομένων συναντούν πολλούς διαφορετικούς τύπους δεδομένων στα έργα ανάλυσης. Τα περισσότερα δεδομένα που συναντώνται σε ακαδημαϊκά και βιομηχανικά έργα μπορούν να ταξινομηθούν σε γενικές γραμμές στις ακόλουθες κατηγορίες:

- Συγχρονικά δεδομένα
- Δεδομένα χρονολογικών σειρών
- Βάση δεδομένων

Η κατανόηση του είδους των δεδομένων που απαιτούνται για την επίλυση ενός προβλήματος και του είδους των δεδομένων που μπορούν να ληφθούν από τις διαθέσιμες πηγές είναι σημαντική για τη διατύπωση του προβλήματος και την επιλογή της σωστής μεθοδολογίας για την ανάλυση. Τα συγχρονικά δεδομένα ή η διατομή ενός πληθυσμού λαμβάνονται με τη λήψη παρατηρήσεων από πολλά άτομα την ίδια χρονική στιγμή. Τα συγχρονικά δεδομένα μπορεί να περιλαμβάνουν παρατηρήσεις που λαμβάνονται σε διαφορετικές χρονικές στιγμές. Στην ουσία, τα συγχρονικά δεδομένα αντιπροσωπεύουν ένα στιγμιότυπο σε μια δεδομένη χρονική στιγμή. Αλλά οι σκοποί ανάλυσης δεν καθορίζονται από το χρόνο.

2.1 Μοντελοποίηση χρονοσειρών

Μια χρονοσειρά έχει 4 στοιχεία που αναφέρονται παρακάτω:

- **Στάθμη:** Είναι η μέση τιμή γύρω από την οποία μεταβάλλεται η σειρά.
- **Τάση:** Είναι η αυξανόμενη ή μειούμενη συμπεριφορά μιας μεταβλητής με το χρόνο.
- **Εποχικότητα:** Είναι η κυκλική συμπεριφορά των χρονολογικών σειρών.
- **Θόρυβος:** Είναι το σφάλμα στις παρατηρήσεις που προστίθεται λόγω περιβαλλοντικών παραγόντων.

Υπάρχουν διάφορες δημοφιλείς τεχνικές μοντελοποίησης χρονοσειρών για την αποτύπωση αυτών των στοιχείων. Κάποιες από αυτές είναι οι εξής:

- **Naïve μέθοδοι:** Πρόκειται για απλές τεχνικές εκτίμησης, όπως ότι η προβλεπόμενη τιμή λαμβάνει την τιμή που ισούται με τη μέση τιμή των προηγούμενων τιμών της χρονικά εξαρτημένης μεταβλητής ή την προηγούμενη πραγματική τιμή. Αυτές χρησιμοποιούνται για σύγκριση με εξελιγμένες τεχνικές μοντελοποίησης.
- **Αυτόματη παλινδρόμηση:** Η αυτόματη παλινδρόμηση προβλέπει τις τιμές των μελλοντικών χρονικών περιόδων ως συνάρτηση των τιμών

σε προηγούμενες χρονικές περιόδους. Οι προβλέψεις της αυτόματης παλινδρόμησης μπορεί να ταιριάζουν καλύτερα με τα δεδομένα από ό,τι οι αφελείς μέθοδοι, αλλά μπορεί να μην είναι σε θέση να λάβει υπόψη την εποχικότητα.

- **Μοντέλο ARIMA:** Ένα αυτοπαλινδρομικό ολοκληρωμένο κινητό μέσο όρο μοντελοποιεί την τιμή μιας μεταβλητής ως γραμμική συνάρτηση των προηγούμενων τιμών και των υπολειμματικών σφαλμάτων σε προηγούμενα χρονικά βήματα μιας στάσιμης χρονοσειράς. Ωστόσο, τα δεδομένα του πραγματικού κόσμου μπορεί να είναι μη στάσιμα και να έχουν εποχικότητα, οπότε αναπτύχθηκαν το Seasonal-ARIMA και το Fractional-ARIMA. Η ARIMA λειτουργεί σε μονομεταβλητές χρονοσειρές, ενώ για τον χειρισμό πολλαπλών μεταβλητών εισήχθη η VARIMA.
- **Εκθετική εξομάλυνση:** Μοντελοποιεί την τιμή μιας μεταβλητής ως εκθετική για να ληφθούν υπόψη οι μακροπρόθεσμες εξαρτήσεις. Μπορεί να εκπαιδευτεί με μεγάλο όγκο δεδομένων για να συλλάβει τις τάσεις σε χρονοσειρές πολλαπλών μεταβλητών.

2.2 Χρονοσειρές - Βαθμονόμηση παραμέτρων

Κάθε στατιστικό μοντέλο ή μοντέλο μηχανικής μάθησης έχει κάποιες παραμέτρους που επηρεάζουν σε μεγάλο βαθμό τον τρόπο με τον οποίο μοντελοποιούνται τα δεδομένα. Για παράδειγμα, το ARIMA έχει τις τιμές p , d , q . Αυτές οι παράμετροι πρέπει να αποφασιστούν έτσι ώστε το σφάλμα μεταξύ των πραγματικών τιμών και των μοντελοποιημένων τιμών να είναι ελάχιστο. Η βαθμονόμηση των παραμέτρων λέγεται ότι είναι η πιο κρίσιμη και χρονοβόρα εργασία της προσαρμογής μοντέλων. Για το λόγο αυτό, είναι πολύ σημαντικό για εμάς να επιλέξουμε τις βέλτιστες παραμέτρους.

Μέθοδοι για τη βαθμονόμηση των παραμέτρων:

- **Hit-and-try:** Ένας συνηθισμένος τρόπος βαθμονόμησης των μοντέλων είναι η βαθμονόμηση με το χέρι, όπου ξεκινάτε με την απεικόνιση της χρονοσειράς και δοκιμάζετε διαισθητικά κάποιες τιμές παραμέτρων και τις αλλάζετε ξανά και ξανά μέχρι να επιτύχετε μια αρκετά καλή προσαρμογή. Απαιτεί μια καλή κατανόηση του

μοντέλου που δοκιμάζουμε. Για το μοντέλο ARIMA, η βαθμονόμηση με το χέρι γίνεται με τη βοήθεια του διαγράμματος αυτοσυσχέτισης για την παράμετρο "p", του διαγράμματος μερικής αυτοσυσχέτισης για την παράμετρο "q" και του ADF-test για την επιβεβαίωση της "στασιμότητας των χρονοσειρών και τη ρύθμιση της παραμέτρου "d".

- Αναζήτηση πλέγματος: Ένας άλλος τρόπος βαθμονόμησης των μοντέλων είναι η αναζήτηση πλέγματος, που ουσιαστικά σημαίνει ότι προσπαθείτε να δημιουργήσετε ένα μοντέλο για όλους τους πιθανούς συνδυασμούς παραμέτρων και επιλέγετε εκείνο με το ελάχιστο σφάλμα. Αυτό είναι χρονοβόρο και ως εκ τούτου είναι χρήσιμο όταν ο αριθμός των παραμέτρων που πρέπει να βαθμονομηθούν και το εύρος των τιμών που παίρνουν είναι λιγότερες, καθώς περιλαμβάνει πολλαπλούς εμφωλευμένους βρόχους for.
- Γενετικός αλγόριθμος: Ο γενετικός αλγόριθμος λειτουργεί με βάση τη βιολογική αρχή ότι μια καλή λύση θα εξελιχθεί τελικά στην πιο "βέλτιστη" λύση. Χρησιμοποιεί τις βιολογικές λειτουργίες της μετάλλαξης, της διασταύρωσης και της επιλογής για να φτάσει τελικά σε μια βέλτιστη λύση.

2.3 Εφαρμογές των χρονοσειρών

Τα μοντέλα χρονοσειρών αναγνωρίζουν πρώτα την τάση και την εποχικότητα από τις υπάρχουσες παρατηρήσεις και στη συνέχεια προβλέπουν μια τιμή με βάση αυτή την τάση και την εποχικότητα. Μια τέτοια ανάλυση είναι χρήσιμη σε διάφορους τομείς όπως:

- Χρηματοοικονομική ανάλυση: Περιλαμβάνει την πρόβλεψη πωλήσεων, την ανάλυση αποθεμάτων, την ανάλυση της χρηματιστηριακής αγοράς, την εκτίμηση τιμών.
- Ανάλυση καιρού: Περιλαμβάνει την εκτίμηση της θερμοκρασίας, την αλλαγή του κλίματος, την αναγνώριση εποχιακών μετατοπίσεων, την πρόβλεψη του καιρού.
- Ανάλυση δεδομένων δικτύου: Περιλαμβάνει πρόβλεψη χρήσης δικτύου, ανίχνευση ανωμαλιών ή εισβολών, προγνωστική συντήρηση.

- Ανάλυση υγειονομικής περίθαλψης: Περιλαμβάνει πρόβλεψη απογραφής, πρόβλεψη ασφαλιστικών παροχών, παρακολούθηση ασθενών".

Επίσης τα μοντέλα χρονοσειρών χρησιμοποιούνται στη μηχανική μάθηση (ML) ως εξής:

- Δεδομένα χρονοσειρών: Πρόκειται για τα δεδομένα που αλλάζουν ανάλογα με το χρόνο και, ως εκ τούτου, ο χρόνος παίζει καθοριστικό ρόλο σε αυτά, τα οποία συζητήσαμε σε μεγάλο βαθμό σε αυτό το σεμινάριο.
- Δεδομένα μη χρονοσειρών: Είναι τα δεδομένα που είναι ανεξάρτητα από το χρόνο και ένα μεγάλο ποσοστό των προβλημάτων ML αφορά δεδομένα μη χρονοσειρών. Για λόγους απλότητας, θα τα κατηγοριοποιήσουμε περαιτέρω ως εξής
- Αριθμητικά δεδομένα: Οι υπολογιστές, σε αντίθεση με τους ανθρώπους, κατανοούν μόνο αριθμούς, οπότε όλα τα είδη δεδομένων μετατρέπονται τελικά σε αριθμητικά δεδομένα για τη μηχανική μάθηση, για παράδειγμα, τα δεδομένα εικόνας μετατρέπονται σε τιμές (r,b,g), οι χαρακτήρες μετατρέπονται σε κωδικούς ASCII ή οι λέξεις ευρετηριάζονται σε αριθμούς, τα δεδομένα ομιλίας μετατρέπονται σε αρχεία mfcc που περιέχουν αριθμητικά δεδομένα.
- Δεδομένα εικόνας: Η υπολογιστική όραση έχει φέρει επανάσταση στον κόσμο των υπολογιστών, έχει διάφορες εφαρμογές στον τομέα της ιατρικής, της δορυφορικής απεικόνισης κ.λπ.
- Δεδομένα κειμένου: Η επεξεργασία φυσικής γλώσσας (NLP) χρησιμοποιείται για την ταξινόμηση κειμένων, την ανίχνευση παραφράσεων. [6]

Κεφάλαιο 3: Το ζωντανό εργαστήριο του

Πανεπιστημίου Δυτικής Μακεδονίας:

Ένα ζωντανό εργαστήριο είναι ένα οικοσύστημα ανοικτής καινοτομίας με επίκεντρο τον χρήστη, το οποίο ενσωματώνει ταυτόχρονες διαδικασίες έρευνας και καινοτομίας στο πλαίσιο μιας σύμπραξης δημόσιου-ιδιωτικού τομέα και ανθρώπων.

Η έννοια βασίζεται σε μια συστηματική προσέγγιση συν-δημιουργίας των χρηστών που ενσωματώνει τις διαδικασίες έρευνας και καινοτομίας. Αυτές ενσωματώνονται μέσω της συνδημιουργίας, διερεύνησης, πειραματισμού και αξιολόγησης καινοτόμων ιδεών, σεναρίων, εννοιών και σχετικών τεχνολογικών αντικειμένων σε πραγματικές περιπτώσεις χρήσης. Σε αυτές τις περιπτώσεις χρήσης εμπλέκονται κοινότητες χρηστών, όχι μόνο ως παρατηρούμενα υποκείμενα αλλά και ως πηγή δημιουργίας. Η προσέγγιση αυτή επιτρέπει σε όλους τους εμπλεκόμενους φορείς να εξετάζουν ταυτόχρονα τόσο τη συνολική απόδοση ενός προϊόντος ή μιας υπηρεσίας όσο και την πιθανή υιοθέτησή του από τους χρήστες. Η εξέταση αυτή μπορεί να γίνει στο προηγούμενο στάδιο της έρευνας και της ανάπτυξης και σε όλα τα στοιχεία του κύκλου ζωής του προϊόντος, από το σχεδιασμό μέχρι την ανακύκλωση.

Μέσα σε μια μεγάλη ποικιλία διαφορετικών ζωντανών εργαστηρίων, όλα τα ζωντανά εργαστήρια χρησιμοποιούν έξι ίδια δομικά στοιχεία.

Το ENoLL, το «Ευρωπαϊκό Δίκτυο Ζωντανών Εργαστηρίων», τα περιγράφει ως εξής:

- **Ενορχήστρωση:** το ζωντανό εργαστήριο λειτουργεί ως ενορχηστρωτής μέσα στο οικοσύστημα για να συνδεθεί και να συνεργαστεί με σχετικούς ενδιαφερόμενους φορείς
- **Πολυμερής συμμετοχή:** υιοθέτηση μιας ολιστικής άποψης για την κοινωνία, με τη συμμετοχή ενδιαφερομένων από το μοντέλο της τετραπλής έλικας (κυβέρνηση, ακαδημαϊκός χώρος, ιδιωτικός τομέας και πολίτες)
- **Ενεργή συμμετοχή των χρηστών:** ένα ζωντανό εργαστήριο εμπλέκει τους σχετικούς ενδιαφερόμενους «ενεργά» σε όλες τις σχετικές δραστηριότητες, διασφαλίζοντας ότι τα σχόλιά τους καταγράφονται και εφαρμόζονται καθ 'όλη τη διάρκεια του κύκλου ζωής της καινοτομίας
- **Συνδημιουργία:** σε ένα ζωντανό εργαστήριο, οι αξίες συνδημιουργούνται από κάτω προς τα πάνω όχι μόνο για αλλά και από όλους τους σχετικούς ενδιαφερόμενους, εξασφαλίζοντας υψηλότερη υιοθέτηση στο τέλος

- **Πραγματικές συνθήκες:** ένα ζωντανό εργαστήριο λειτουργεί στο πραγματικό περιβάλλον των τελικών χρηστών, εισάγοντας καινοτομίες στην πραγματική τους ζωή αντί να μετακινεί τους χρήστες σε ιστότοπους δοκιμών για να εξερευνήσουν τις καινοτομίες
- **Προσέγγιση πολλαπλών μεθόδων:** κάθε δραστηριότητα ζωντανού εργαστηρίου καθοδηγείται από προβλήματα. Ως εκ τούτου, η μεθοδολογική προσέγγιση για κάθε μεμονωμένη δραστηριότητα θα επιλεγεί με βάση τα αναμενόμενα αποτελέσματα της δραστηριότητας και τα ενδιαφερόμενα μέρη που πρέπει να συμμετάσχουν.

3.1 Το ζωντανό εργαστήριο του Πανεπιστημίου Δυτικής Μακεδονίας (SMART)

Το ζωντανό εργαστήριο του Πανεπιστημίου Δυτικής Μακεδονίας (SMART) (<https://livinglab.ece.uowm.gr/>) έχει συνολική δομημένη επιφάνεια 4212 m² και επιφάνεια οροφής / δαπέδου 2430 m². Στο μεγαλύτερο μέρος του κτιρίου υπάρχουν δύο όροφοι και η επιφάνεια των ανοιγμάτων είναι 800 m². Αποτελείται από 190 μονόκλινα δωμάτια, τρία γραφεία πανεπιστημιακού προσωπικού, 9 αποθήκες, 9 σοφίτες, 3 πλυντήρια και ένα εξειδικευμένο δωμάτιο ελέγχου κατασκευασμένο στο πλαίσιο του ζωντανού εργαστηρίου. Έχει αναπτυχθεί μέσω χρηματοδότησης μέσω τεσσάρων ερευνητικών προγραμμάτων που έχουν συνεισφέρει σταδιακά εξοπλισμό και εργασία, ξεκινώντας από το 2017. Ο στόχος του ζωντανού εργαστηρίου είναι να δημιουργήσει ένα «φιλικό προς το χρήστη» περιβάλλον για τη δοκιμή βιομηχανικών προϊόντων και για ερευνητικούς σκοπούς.

Το SMART είναι ένα Ζωντανό Εργαστήριο σχεδιασμένο να λειτουργεί ως ένα ανοικτό οικοσύστημα καινοτομίας με επίκεντρο τον χρήστη, συνδεδεμένο με το δυναμικό αξίας των νέων αγορών. Οι εγκαταστάσεις του Πανεπιστημίου Δυτικής Μακεδονίας και οι καταναλωτές εντός της περιφέρειας Κεντρικής και Δυτικής Μακεδονίας στην Ελλάδα συνθέτουν το πραγματικό περιβάλλον δοκιμών. Στόχος του SMART LL είναι να εξισορροπήσει εκ νέου τη δημόσια δράση σε μια λογική λαμβάνοντας υπόψη τη γενικότητα της ανησυχίας των

πολιτών μέσω της ενεργού συμμετοχής των χρηστών. Οι διαφορετικές κλιματικές συνθήκες (μεγάλο υψόμετρο, παραλία) και γεωγραφικές ιδιαιτερότητες (λιγνιτικοί σταθμοί, υψηλή διείδυση ΑΠΕ) των εγκατεστημένων χρηστών επιτρέπουν ένα ευρύ φάσμα δοκιμών σε πραγματικές συνθήκες, συμπεριλαμβανομένων παραγόντων με μικρό οικονομικό βάρος. Η άρτια εκπαιδευμένη ομάδα του SMART, αποτελούμενη από Καθηγητές, μεταδιδακτορικούς ερευνητές, υποψήφιους διδάκτορες και μηχανικούς, προσπαθεί συνεχώς να δημιουργήσει ισχυρούς δεσμούς μεταξύ της ακαδημαϊκής κοινότητας και της τοπικής επιχειρηματικής υποδομής με ιδιαίτερη έμφαση στις Μικρομεσαίες Επιχειρήσεις. Το SMART βασίζεται στους πυλώνες της έρευνας, της εκπαίδευσης και των βιομηχανικών δοκιμών. Μέσω της εκπαιδευτικής διάστασης, οι μαθητές μπορούν να έχουν πρόσβαση σε αρχεία και περιουσιακά στοιχεία πραγματικών χρηστών. Τα ερευνητικά ενδιαφέροντα περιλαμβάνουν έξυπνα δίκτυα, βοηθητικές υπηρεσίες σε διαχειριστές δικτύων, τεχνικές διαχείρισης ζήτησης, βελτιστοποίηση λειτουργίας φωτοβολταϊκών και συστημάτων αποθήκευσης ενέργειας

3.2 Εξοπλισμός του ζωντανού εργαστηρίου

Οι χρονοσειρές που εξετάζονται στην παρούσα εργασία προέρχονται από τον εξής εξοπλισμό του ζωντανού εργαστηρίου:

1. 35,4 kWp φωτοβολταϊκών, χωρισμένα σε τρία υποσυστήματα, δηλ. τρεις υβριδικούς μετατροπείς με διπλές σειρές το καθένα.
2. 31,4 kWh συστημάτων αποθήκευσης ενέργειας με μπαταρίες, χωρισμένα σε τρία υποσυστήματα, ένα για κάθε φωτοβολταϊκό σύστημα.
3. Αναλυτές ποιότητας ισχύος που παρακολουθούν όλες τις κύριες γραμμές της εγκατάστασης

Το Φωτοβολταϊκό σύστημα αποθήκευσης αποτελείται 84 πολυκρυσταλλικά πλαίσια ισχύος 270 Wp το καθένα. Η εικόνα 4 δείχνει την κάτοψη του κτιρίου με τη διάταξη των δύο σειρών φωτοβολταϊκών πλαϊσίων μαζί με τη διάταξη του τμήματος του κτιρίου.

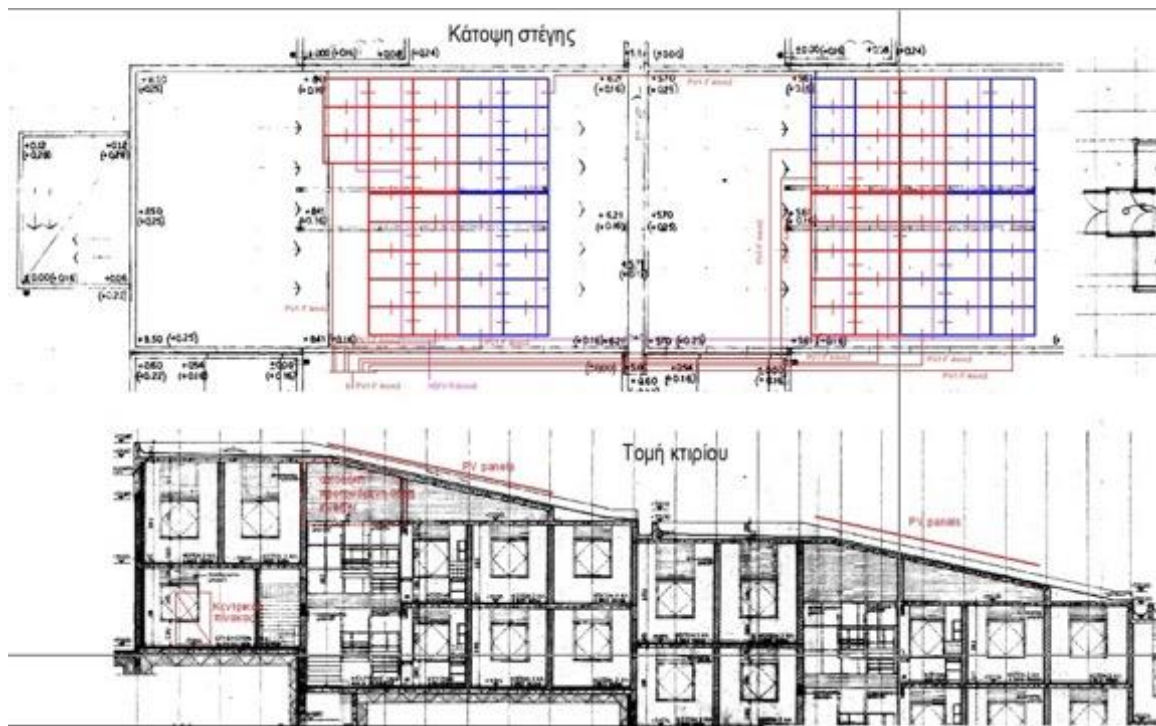


Figure 4 Κάτοψη κτηρίου με τα φωτοβολταϊκά πάνελ και τη διάταξη των τμημάτων του κτηρίου της δημόσιας πιλοτικής εγκατάστασης στην Κοζάνη

Η σύζευξη μεταξύ των φωτοβολταϊκών και του BESS είναι στην πλευρά DC, λόγω της χρήσης ενός υβριδικού μετατροπέα. Δύο αντιστροφείς 10 kW SMA KOSTAL PLENTICORE Plus 10, όπως φαίνεται στην εικόνα 5, είναι υπεύθυνοι για τη φόρτιση και την αποφόρτιση των συσσωρευτών και έχουν συνδυαστεί με έναν υπερσύγχρονο συσσωρευτή μέσω τριφασικής σύνδεσης. Για το σύστημα αποθήκευσης, έχουν εγκατασταθεί δύο μονάδες μπαταριών (Lithium-Iron Phosphate), όπως φαίνεται στο σχήμα 6, με ενεργειακή ισχύ 10,24 kWh η καθεμία και κάθε μία από αυτές τις μονάδες συνδέεται με έναν από τους δύο υβριδικούς μετατροπείς.

Επιπλέον, έχουν εγκατασταθεί δύο μετρητές ισχύος (το ειδικό μοντέλο που χρησιμοποιείται είναι το EM300 LR), όπως φαίνεται στην εικόνα 7, προκειμένου να παρακολουθείται η κατανάλωση ενέργειας εντός της εγκατάστασης. Ο ρόλος αυτών των μετρητικών συστημάτων είναι να παρέχουν τις μετρήσεις της κατανάλωσης φορτίου στον μετατροπέα, προκειμένου να ελέγχουν είτε τη φόρτιση είτε την εκφόρτιση του συστήματος συσσωρευτή. Μετρούν την πραγματική και άεργο ισχύ, τη φαινόμενη ισχύ, την πραγματική, άεργη και φαινόμενη ενέργεια, το ρεύμα, την τάση και τον συντελεστή ισχύος για τις τρεις φάσεις του εναλλασσόμενου ρεύματος. Επιπλέον, το σύστημα αποθήκευσης είναι

εξοπλισμένο με BMS που μπορεί να παρέχει μετρήσεις όπως η κατάσταση φόρτισης (SoC), η ισχύς της μπαταρίας DC κ.λπ.

Inverter	Hybrid
	<ul style="list-style-type: none">• DC-Coupled System• 2X10 kW rated power• Round-trip efficiency 96.5 %• Indoor-Outdoor (IP65)• Communication with battery- RS485

image taken from <https://www.kostal-solar-electric.com/en-gr/products/hybrid-inverters/premcore-plus>

Figure 5 Υβριδικός αντιστροφέας με τα βασικά χαρακτηριστικά του


Battery	LiFePO4
	<ul style="list-style-type: none">• High Voltage (570 VDC)• 20.48 kWh Usable Capacity• Deep Cycle (95% DoD)• Indoor (IP55)• Temperature: -10 °C to 50 °C

image taken from www.mg-solar-shop.com

Figure 6 Σύστημα μπαταριών με τα βασικά χαρακτηριστικά του


Meters	Load Meter
	<ul style="list-style-type: none">• Battery System Monitoring• Load monitoring• Energy consumption of installation• Energy exchange with the grid• Optimized battery control

image taken from www.b-control.com

Figure 7 Μετρητής ισχύος με τα βασικά χαρακτηριστικά του

Όπως φαίνεται στο σχήμα 8, η συσκευή κεντρικού ελεγκτή και ο υπόλοιπος εξοπλισμός συνδέονται πίσω από τον μετρητή ενέργειας της

υπηρεσίας κοινής ωφέλειας, με αποτέλεσμα ένα BtM BESS. Η ηλεκτρική διάταξη της εγκατάστασης παρουσιάζεται στην εικόνα 8, όπου απεικονίζονται οι συνδέσεις μεταξύ των εξαρτημάτων του πιλοτικού συστήματος. Αυτή η διάταξη αποτυπώνει τη σύζευξη DC της εγκατάστασης, καθώς δείχνει ότι τόσο τα φωτοβολταϊκά πάνελ όσο και τα συστήματα συσσωρευτών συνδυάζονται με τη χρήση ενός μόνο υβριδικού αντιστροφέα. Επιπλέον, παρουσιάζεται όλος ο εξοπλισμός προστασίας μαζί με τις πρόσθετες συσκευές μέτρησης και έτσι απεικονίζεται η βασική ιδέα του συστήματος ελέγχου φόρτισης/εκφόρτισης του συστήματος μπαταρίας. Σε αυτή την ηλεκτρική διάταξη περιλαμβάνονται όλα τα κυκλώματα που τροφοδοτούν το φορτίο εντός της εγκατάστασης και έτσι εμφανίζεται και η παρακολούθηση φορτίου των πρόσθετων εγκατεστημένων μετρητικών συσκευών. Η εικόνα 9 δείχνει την εσωτερική εγκατάσταση (αριστερή φωτογραφία) και την όψη της οροφής του κτιρίου (δεξιά φωτογραφία). Το πρώτο απεικονίζει και τα δύο συστήματα συσσωρευτών με τους δύο υβριδικούς μετατροπείς, ενώ το δεύτερο δείχνει ένα μέρος ενός από τα δύο φωτοβολταϊκά πάνελ .

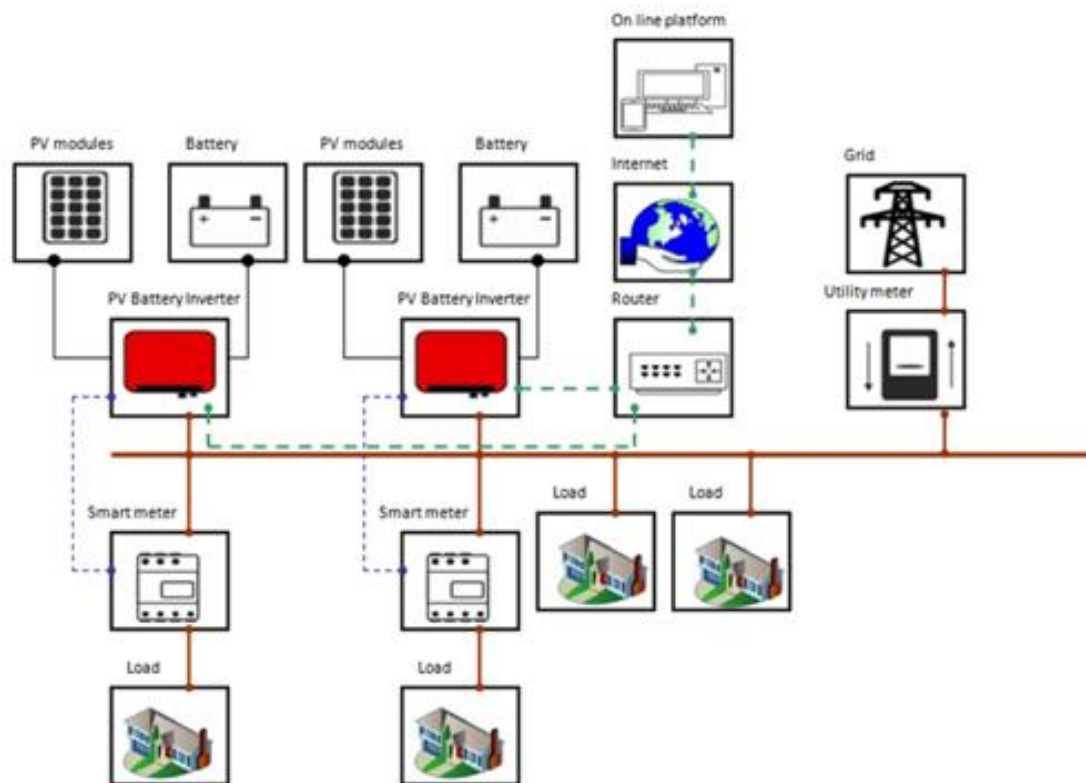


Figure 8 Επισκόπηση πιλοτικού συστήματος BESS (τοπολογία σύζευξης DC)



Figure 9 Φωτογραφίες εγκατάστασης (εσωτερική εγκατάσταση και φωτοβολταϊκά πάνελ στην οροφή)

Στον Πίνακα 1 παρουσιάζονται τα βασικά τεχνικά χαρακτηριστικά που αφορούν τα Φ/Β πάνελ, τους υβριδικούς αντιστροφείς και τα συστήματα συσσωρευτών. Λεπτομερέστερες πληροφορίες παρέχονται στο προσάρτημα..

Πίνακας 1: Τεχνικά χαρακτηριστικά του δημόσιου εξοπλισμού BESS

PV Modules (x84)	CSUN 270 - 60p
Technology	Polycrystalline
Maximum Power	270W
Module Efficiency	16.63%
Temperature Range	-40°C to 85°C
cells	60
Junction box	IP67, 3 bypass diodes
Frame	Anodized aluminium
Application class	Class A
Safety class	Class II
PV-Battery hybrid inverter (x2)	KOSTAL PLENTICORE Plus 10
Rated/peak power	10/10 kW
Round-trip efficiency	96.5%
Relative humidity range	4-100%
Communication with battery	RS485
IP rate	IP65
DC input voltage range/rated	120-1000/570 V

Installation	Indoor/outdoor
Battery unit (x2)	BYD B-box HV 10.2
Technology	Lithium-Iron Phosphate (LiFePO ₄)
Energy capacity (nominal/usable)	10.24/10.24 kWh
Maximum charge/discharge power	10.24 kW
Relative humidity range	----
Depth of Discharge (DoD)	95%
Communication with the inverter	RS485
IP rate	IP55
Operating Temperature Range	-10 to 50 °C
Installation	Indoor

Κεφάλαιο 4: Δημιουργία αλγορίθμου για τον έλεγχο χρονοσειρών από το ζωντανό εργαστήριο

Σκοπός της παρούσας εργασίας είναι η δημιουργία ενός αλγορίθμου για τον έλεγχο χρονοσειρών δεδομένων ηλεκτρικής ενέργειας που συλλέχθηκαν από τις συσκευές που παρουσιάστηκαν στο τρίτο κεφάλαιο. Σε αυτό το κεφάλαιο παρουσιάζεται η λειτουργία του αλγορίθμου, τα βήματα που ακολουθήθηκαν ώστε να λειτουργήσει, καθώς και τα αποτελέσματα της εκτέλεσής του.

4.1 Παρουσίαση αρχείων χρονοσειρών και δημιουργία αρχείου αλγορίθμου

Αρχικά δημιουργήθηκε ένας φάκελος ο οποίος χρησιμοποιήθηκε ώστε να περιέχει τα αρχεία των χρονοσειρών αλλά και τα αρχεία του κώδικα που θα χρησιμοποιηθούν για τον έλεγχο των χρονοσειρών. Για την συγγραφή

και εκτέλεση του κώδικα στην παρούσα πτυχιακή χρησιμοποιείται το λογισμικό Visual Studio Code. Το αρχείο με τα δεδομένα χρονοσειρών προέρχεται από τη βάση δεδομένων του ζωντανού εργαστηρίου: <https://smart.ece.uowm.gr>.

Αυτό το αρχείο περιέχει ένα ολοκληρωμένο σύνολο δεδομένων χρονοσειρών που καταγράφει ένα ευρύ φάσμα μετρήσεων και μετρήσεων ηλεκτρικής ισχύος που σχετίζονται με την ενέργεια. Το σύνολο δεδομένων είναι οργανωμένο σε πολλαπλές στήλες, καθεμία από τις οποίες αντιπροσωπεύει μια συγκεκριμένη παράμετρο. Τα δεδομένα καταγράφονται με την πάροδο του χρόνου, συνήθως σε τακτά χρονικά διαστήματα, και είναι πολύτιμα για την ανάλυση και την κατανόηση διαφόρων πτυχών της κατανάλωσης και της παραγωγής ενέργειας σε ένα συγκεκριμένο σύστημα ή εγκατάσταση. Ακολουθεί ανάλυση μερικών στηλών που υπάρχουν στα δεδομένα χρονοσειρών:

1. Χρονοσφραγίδα: Η ημερομηνία και η ώρα κατά την οποία καταγράφηκε κάθε σημείο δεδομένων.
2. Ισχύς συνεχούς ρεύματος φόρτισης της μπαταρίας I: Η ισχύς συνεχούς ρεύματος (DC) που χρησιμοποιείται για τη φόρτιση της μπαταρίας I.
3. Ισχύς συνεχούς ρεύματος εκφόρτισης της μπαταρίας I: Η ισχύς συνεχούς ρεύματος που χρησιμοποιείται όταν η μπαταρία I εκφορτίζει την αποθηκευμένη ενέργειά της.
4. Battery II-Charging DC Power: Παρόμοια με την Battery I, αυτή η στήλη αντιπροσωπεύει την ισχύ συνεχούς ρεύματος που χρησιμοποιείται για τη φόρτιση της Battery II.
5. Μπαταρία II-Εκφόρτιση DC Power: Η ισχύς DC που χρησιμοποιείται όταν η μπαταρία II εκφορτίζεται.
6. Κοινόχρηστα φώτα-ενεργή ισχύς L1/L2/L3: Η κατανάλωση ενεργού ισχύος (σε Watt) για τα κοινόχρηστα φώτα σε διαφορετικές ηλεκτρικές φάσεις (L1, L2 και L3). Τα δεδομένα αυτά χωρίζονται σε ξεχωριστές στήλες για κάθε φάση.
7. Κοινόχρηστα φώτα-ενεργός ισχύς Άθροισμα L1-L3: Το άθροισμα της κατανάλωσης ενεργού ισχύος για τα κοινόχρηστα φώτα και στις τρεις φάσεις (L1, L2 και L3).

8. Άθροισμα ενεργού ισχύος εξόδου των αντιστροφών I/II/III L1-L3: Η ενεργός ισχύς που παράγεται ή εξάγεται από τους αντιστροφείς I, II και III, αντίστοιχα, σε διαφορετικές ηλεκτρικές φάσεις (L1, L2 και L3). Αυτές οι στήλες αντιπροσωπεύουν την παραγωγή ισχύος.
9. Inverter III Output-Janitza-Active Power Sum L1-L3: Ειδικά για τον Inverter III, η στήλη αυτή παρέχει δεδομένα παραγωγής ενεργού ισχύος σε διαφορετικές ηλεκτρικές φάσεις (L1, L2 και L3).
10. PCC με το δίκτυο-άθροισμα ενεργού ισχύος L1-L3: Η ενεργός ισχύς που ανταλλάσσεται μεταξύ του σημείου κοινής ζεύξης (PCC) και του ηλεκτρικού δικτύου, κατανεμημένη ανά διαφορετικές ηλεκτρικές φάσεις (L1, L2 και L3).
11. Ισχύς παραγωγής PV III: Η ισχύς που παράγεται από το φωτοβολταϊκό (PV) σύστημα III, που συνήθως συνδέεται με την παραγωγή ηλιακής ενέργειας.
12. Πλυντήρια ρούχων-ενεργή ισχύς L1/L2/L3: Η κατανάλωση ενεργού ισχύος των πλυντηρίων ρούχων σε διαφορετικές ηλεκτρικές φάσεις (L1, L2 και L3). Τα δεδομένα αυτά διαχωρίζονται ανά φάση.
13. Πλυντήρια ρούχων-ενεργή ισχύς Άθροισμα L1-L3: Η συνολική κατανάλωση ενεργού ισχύος των πλυντηρίων ρούχων και στις τρεις φάσεις (L1, L2 και L3).
14. Πτέρυγα 1-1/1-2/2-1/2-2/2-3/3-1/3-2/3-3-Active Power Sum L1-L3: Οι στήλες αυτές αντιπροσωπεύουν το άθροισμα της κατανάλωσης ενεργού ισχύος σε διάφορες πτέρυγες και τμήματα μιας εγκατάστασης, με τα δεδομένα να διαχωρίζονται ανά ηλεκτρική φάση (L1, L2 και L3). Ωστόσο, ο αριθμός των μεγεθών που θα ελεγχθούν μέσω του κώδικα είναι περίπου 270 μετρικά.

4.2 Παρουσίαση της δομής του αλγορίθμου

Παρακάτω παρουσιάζονται τα βήματα του κώδικα Python για την αξιολόγηση της ποιότητας των δεδομένων και την ανίχνευση ανωμαλιών στις χρονοσειρές.

Βήμα 1: Ο κώδικας ξεκινά με την εισαγωγή των απαραίτητων βιβλιοθηκών, συμπεριλαμβανομένων των ``os``, ``numpy``, ``matplotlib``, ``datetime``, ``pandas``, ``scipy``, ``openpyxl`` και ``stumpy``. Στη συνέχεια, ορίζει τον κατάλογο εργασίας στη θέση του αρχείου του σεναρίου.

Βήμα 2: Καθορίζει τη διαδρομή προς ένα αρχείο Excel με όνομα 'testing.xlsx' και διαβάζει τα δεδομένα από αυτό το αρχείο σε ένα πλαίσιο δεδομένων Pandas DataFrame. Ο κώδικας αντικαθιστά επίσης τυχόν συμβολοσειρές "NaN" στο σύνολο δεδομένων με πραγματικές τιμές NaN.

Βήμα 3: Το σενάριο ελέγχει για ελλείπουσες τιμές στο DataFrame και μετρά τον αριθμό των τιμών NaN σε κάθε μετρική στήλη. Τα αποτελέσματα εκτυπώνονται στην κονσόλα.

Βήμα 4: Ελέγχει για διπλές χρονοσφραγίδες στη στήλη 'timestamp' του DataFrame και εκτυπώνει τον αριθμό των διπλών χρονοσφραγίδων.

Βήμα 5: Ο κώδικας μετατρέπει τη στήλη 'timestamp' σε μορφή datetime, ταξινομεί το DataFrame με βάση τη στήλη 'timestamp' και στη συνέχεια ελέγχει αν οι χρονοσφραγίδες είναι σε χρονολογική σειρά. Εκτυπώνει αν οι χρονοσφραγίδες είναι σε χρονολογική σειρά ή όχι.

Βήμα 6: Ορίζει μια συνάρτηση 'detect_outliers_zscore' για την ανίχνευση ακραίων τιμών σε αριθμητικές στήλες με τη χρήση Z-scores. Στη συνέχεια, ο κώδικας εφαρμόζει αυτή τη συνάρτηση ανίχνευσης ακραίων τιμών σε καθεμία από τις στήλες αριθμητικών μετρήσεων στο DataFrame και εκτυπώνει τα αποτελέσματα.

Βήμα 7: Αυτό το βήμα εκτελεί επικύρωση εύρους δεδομένων, εξομάλυνση δεδομένων και οπτικοποίηση για κάθε μετρική στήλη. Επαναλαμβάνει τις μετρικές στήλες, υπολογίζει το εύρος τιμών και εκτυπώνει την ελάχιστη και τη μέγιστη τιμή. Εφαρμόζει επίσης εξομάλυνση δεδομένων χρησιμοποιώντας το φίλτρο Savitzky-Golay για να δημιουργήσει μια εξομαλυμένη έκδοση των δεδομένων.

Για κάθε μετρική στήλη, εκτελεί μια διαδικασία ανίχνευσης ανωμαλιών χρησιμοποιώντας το προφίλ του πίνακα που υπολογίζεται με τη βιβλιοθήκη 'stumpy'. Οι ανωμαλίες επισημαίνονται και οπτικοποιούνται σε ένα διάγραμμα διασποράς.

Βήμα 8: Υπάρχει ένα σχόλιο τοποθέτησης που προτείνει την προσθήκη προσαρμοσμένων ελέγχων με βάση κανόνες και απαιτήσεις συγκεκριμένου τομέα. Αυτό το βήμα αφήνεται κενό για τυχόν πρόσθετους ελέγχους ποιότητας δεδομένων.

Βήμα 9: Τέλος, το σενάριο εκτυπώνει μια σύνοψη ποιότητας δεδομένων, συμπεριλαμβανομένου του αριθμού των γραμμών και των στηλών στο DataFrame.

Συνολικά, αυτός ο κώδικας είναι ένα ολοκληρωμένο σενάριο αξιολόγησης της ποιότητας των δεδομένων και ανίχνευσης ανωμαλιών για δεδομένα χρονοσειρών, συμπεριλαμβανομένης της προεπεξεργασίας δεδομένων, της ανίχνευσης ακραίων τιμών και της οπτικοποίησης. Τα σχόλια εντός του κώδικα παρέχουν επεξηγήσεις για κάθε βήμα και για το τι επιδιώκει να επιτύχει.

4.3 Παρουσίαση αποτελεσμάτων

Στην παρούσα εργασία ελέγχθηκαν 271 μεγέθη για διάστημα ενός έτους. Τα αποτελέσματα των ελέγχων μετά την εκτέλεση του κώδικα παρουσιάζονται παρακάτω:

Πίνακας 2: Αποτελέσματα ελέγχων μετά την εκτέλεση του κώδικα

Όνομα μετρικού	Εύρος	Αριθμός «NaN»	Αριθμός μη φυσιολογικών τιμών
Active power sum L1-L3 P-Grid	-9161.632 - 12759.988	24113	24123
Battery I-Charging DC power	-5720.613 - 5410.209	5405	5439
Battery I-Discharging DC power	-5720.613 - 5410.209	5405	5439
Battery II-Charging DC power	-5682.927 - 5433.2	5543	5577
Battery II-Discharging DC power	-5682.927 - 5433.2	5543	5577
Communal lights-Active power L1	-4.671 - 2543.615	2087	2100
Communal lights-Active power L2	0.0 - 908.123	2087	2100
Communal lights-Active power L3	-4.877 - 1881.028	2087	2100
Communal lights-Active power sum L1-L3	-3.396 - 5056.165	2087	2100
Communal lights-Reactive power L1	-46.641 - 4390.547	2087	2100

Communal lights- Reactive power L2	-56.287 - 1333.405	2087	2100
Communal lights- Reactive power L3	-44.666 - 2836.981	2087	2100
Communal lights- Reactive power sum L1- L3	-143.062 -	2087	2100
Inverter I output- Active power L1	-1276.06 -	129	143
	3094.653		
Inverter I output- Active power L2	-916.192 -	129	143
	3341.867		
Inverter I output- Active power L3	-923.085 -	129	143
	3425.24		
Inverter I output- Active power sum L1-L3	-3010.608 -	129	143
	9831.739		
Inverter II output- Active power L1	-927.266 -	129	143
	3350.046		
Inverter II output- Active power L2	-952.163 -	129	143
	3355.15		
Inverter II output- Active power L3	-957.641 -	129	143
	3439.904		
Inverter II output- Active power sum L1-L3	-2837.101 -	129	143
	10127.759		
Inverter II supplied loads-Active power sum L1-L3	0.0 - 18563.338	5543	5577
Inverter II supplied loads-Reactive power sum L1-L3	-384.2 - 1248.902	5543	5577
Inverter III output- Active power L1	- 10665.969	1507	1543
	- 0.0		
Inverter III output- Active power L2	-9397.582 -	1507	1543
	- 0.0		
Inverter III output- Active power L3	-8691.14 -	1507	1543
	- 0.0		
Inverter III output- Active power sum L1-L3	- 23284.236	1507	1543
	- 0.0		
Inverter III output- Janitza-Active power L1	-177.611 -	67	80
	3352.05		
Inverter III output- Janitza-Active power L2	-182.792 -	67	80
	3375.426		

Inverter III output- Janitza-Active power L3	-190.168 - 3437.056	67	80
Inverter III output- Janitza-Active power sum L1-L3	-548.906 - 10148.962	67	80
PCC with the grid- Active power L1	-8397.662 - 20220.725	125	139
PCC with the grid- Active power L2	-9291.002 - 23647.301	125	139
PCC with the grid- Active power L3	-9276.789 - 23733.881	125	139
PCC with the grid- Active power sum L1-L3	- 26598.727 - 55935.516	125	139
PCC with the grid- Current L1	1.968 - 88.631	125	139
PCC with the grid- Current L2	0.0 - 101.211	125	139
PCC with the grid- Current L3	0.848 - 101.351	125	139
PCC with the grid- Reactive power L1	-800.696 - 4448.278	125	139
PCC with the grid- Reactive power L2	-714.544 - 2523.198	125	139
PCC with the grid- Reactive power L3	-350.525 - 4034.217	125	139
PCC with the grid- Reactive power sum L1- L3	-1310.613 - 9996.861	125	139
PV I-Generation DC power string I	-0.195 - 5946.654	5405	5439
PV I-Generation DC power string II	-0.227 - 5982.993	5405	5439
PV I-Self-consumed power	0.0 - 8197.409	5405	5439
PV II-Generation DC power string I	-0.001 - 5592.348	5543	5577
PV II-Generation DC power string II	0.0 - 5486.959	5543	5577
PV II-Self-consumed power	0.0 - 7942.431	5543	5577
PV III-Generation power	0.0 - 10326.2	1507	1543

Washing machines-Active power L1	0.0 - 5540.981	38	49
Washing machines-Active power L2	0.0 - 7302.845	38	49
Washing machines-Active power L3	0.0 - 4564.402	38	49
Washing machines-Active power sum L1-L3	0.0 - 14236.046	38	49
Washing machines-Reactive power sum L1-L3	-331.678 -	38	49
	1543.959		
Wing 1-1-Active power L1	0.0 - 3349.163	2086	2099
Wing 1-1-Active power L2	0.0 - 4355.887	2086	2099
Wing 1-1-Active power L3	0.0 - 6989.689	2086	2099
Wing 1-1-Active power sum L1-L3	0.0 - 8187.585	2086	2099
Wing 1-2-Active power L1	0.0 - 6077.688	2089	2102
Wing 1-2-Active power L2	0.0 - 4611.573	2089	2102
Wing 1-2-Active power L3	4.129 - 6418.906	2089	2102
Wing 1-2-Active power sum L1-L3	25.138 - 11684.929	2089	2102
Wing 1-3-Active power L1	0.0 - 4310.989	38	49
Wing 1-3-Active power L2	3.059 - 6015.442	38	49
Wing 1-3-Active power L3	-3.947 - 5814.555	38	49
Wing 1-3-Active power sum L1-L3	3.059 - 11026.772	38	49
Wing 2-1-Active power L1	0.0 - 1783.859	38	49
Wing 2-1-Active power L2	0.0 - 4099.357	38	49
Wing 2-1-Active power L3	0.0 - 3408.488	38	49
Wing 2-1-Active power sum L1-L3	0.0 - 6868.951	38	49
Wing 2-2-Active power L1	0.0 - 2244.778	210	226
Wing 2-2-Active power L2	0.0 - 5112.256	210	226
Wing 2-2-Active power L3	0.0 - 4601.401	210	226
Wing 2-2-Active power sum L1-L3	0.0 - 8958.302	210	226

Wing 2-3-Active power L1	0.0 - 5990.083	2091	2104
Wing 2-3-Active power L2	0.0 - 967.28	2091	2104
Wing 2-3-Active power L3	0.0 - 5693.057	2091	2104
Wing 2-3-Active power sum L1-L3	0.0 - 9786.695	2091	2104
Wing 3-1-Active power L1	0.0 - 3832.455	38	49
Wing 3-1-Active power L2	23.629 - 4774.305	38	49
Wing 3-1-Active power L3	20.556 - 2709.193	38	49
Wing 3-1-Active power sum L1-L3	94.399 - 8873.503	38	49
Wing 3-2-Active power L1	-2.88 - 2168.028	38	49
Wing 3-2-Active power L2	0.0 - 5037.852	38	49
Wing 3-2-Active power L3	0.0 - 4259.315	38	49
Wing 3-2-Active power sum L1-L3	0.0 - 7865.876	38	49
Wing 3-3-Active power L1	-2.788 - 3139.731	125	137
Wing 3-3-Active power L2	0.0 - 7089.327	125	137
Wing 3-3-Active power L3	0.0 - 3122.763	125	137
Wing 3-3-Active power sum L1-L3	0.0 - 10219.822	125	137

Κατά τη διάρκεια εκτέλεσης του κώδικα παράχθηκαν συνολικά 271 διαγράμματα, τα οποία για λόγους συντομίας δεν παρατίθενται όλα σε αυτό το κεφάλαιο. Στις επόμενες 7 εικόνες παρουσιάζονται ενδεικτικά 7 μεγέθη. Η γαλάζια γραμμή του διαγράμματος παρουσιάζει τα πρωτογενή δεδομένα, ενώ η κόκκινη γραμμή τα δεδομένα όπου έχει εφαρμοστεί το φίλτρο Savitzky-Golay που χρησιμοποιείται στην επεξεργασία σήματος για την εξάλειψη του θορύβου σε ένα σήμα και τη βελτίωση της ομαλότητας της τάσης ενός σήματος. Με κόκκινο «X» παρουσιάζονται τα σημεία όπου ο αλγόριθμος της βιβλιοθήκης stumpy ανίχνευσε ανωμαλίες.

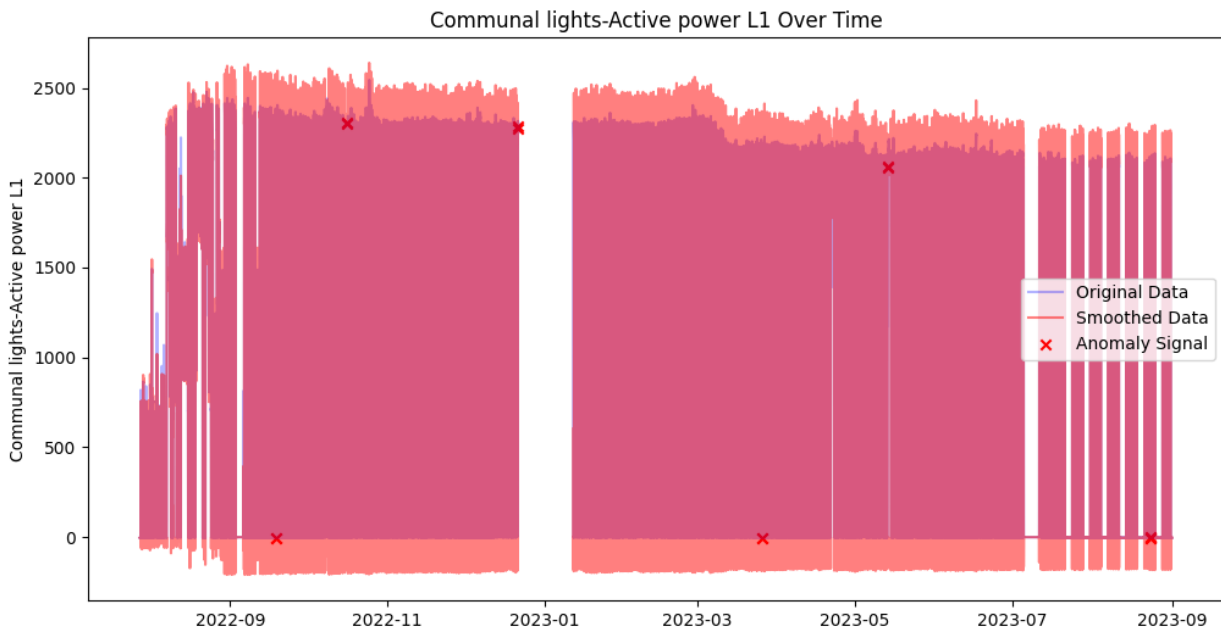


Figure 10 Διάγραμμα κατανάλωσης πρώτης φάσης φωτισμού

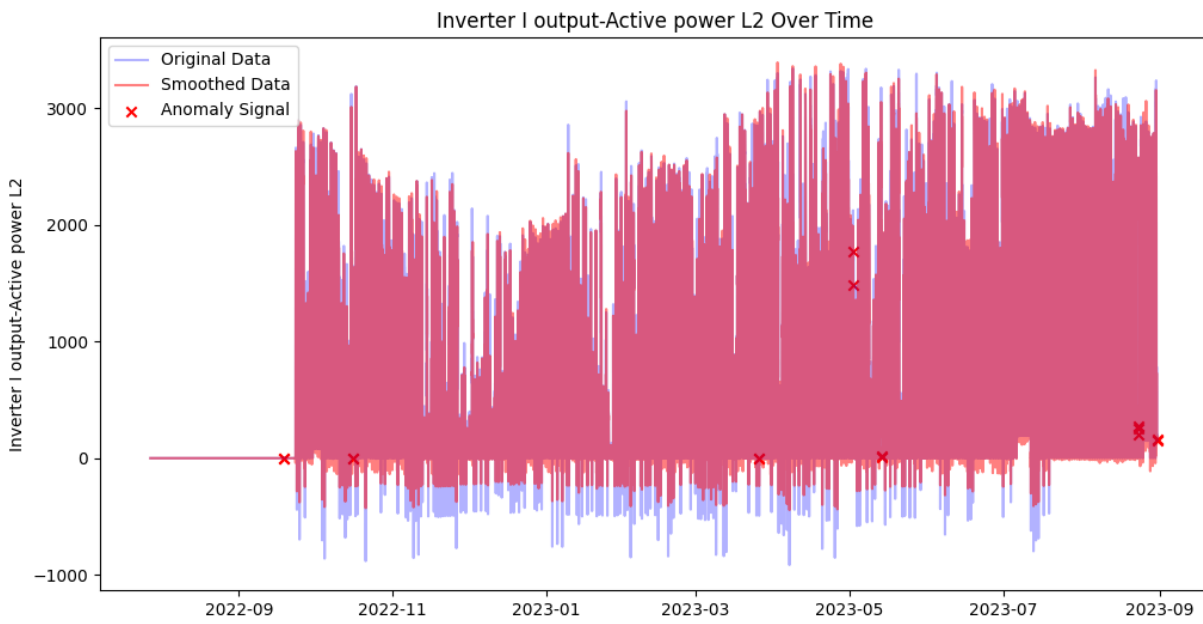


Figure 11 Διάγραμμα πραγματικής ισχύος δεύτερης βάσης για το πρώτο υποσύστημα Inverter

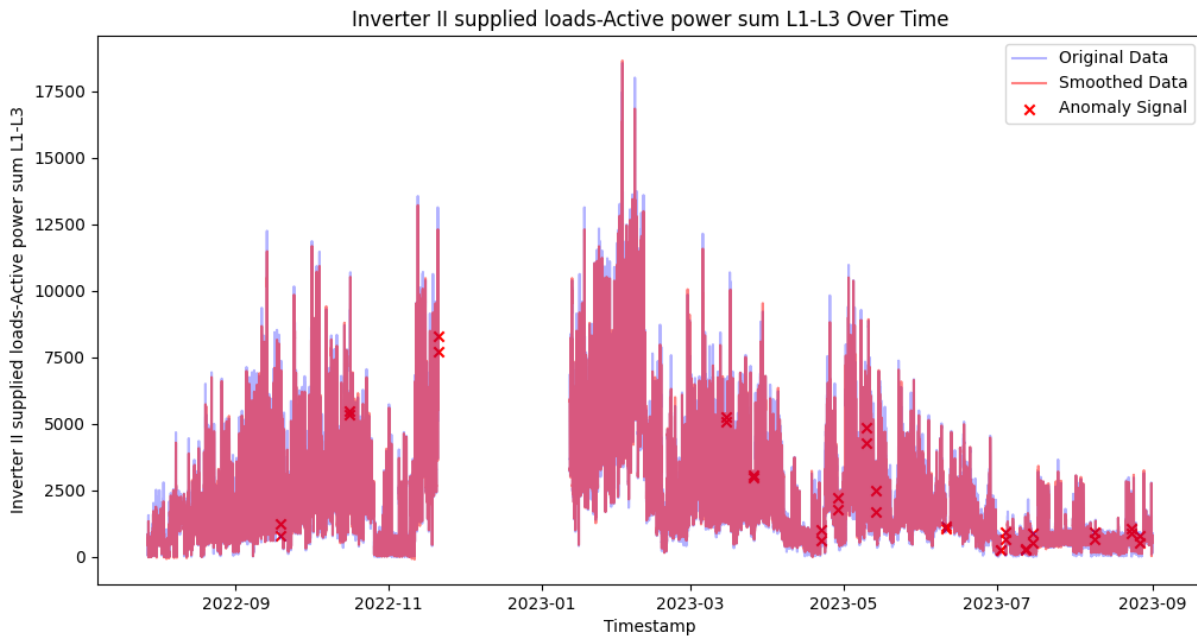


Figure 12 Διάγραμμα παροχής πραγματικής ισχύος για τις φάσεις L1 - L2 για το δεύτερο υποσύστημα Inverter

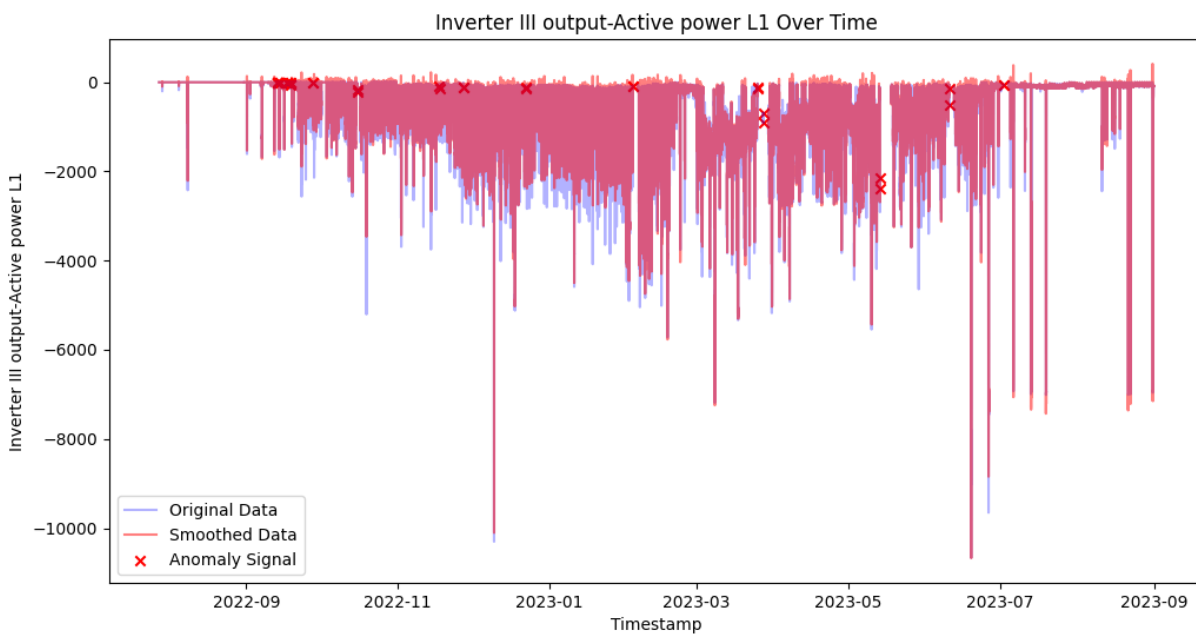


Figure 13 Διάγραμμα παροχής πραγματικής ισχύος για την πρώτη φάση του τρίτου υποσυστήματος Inverter

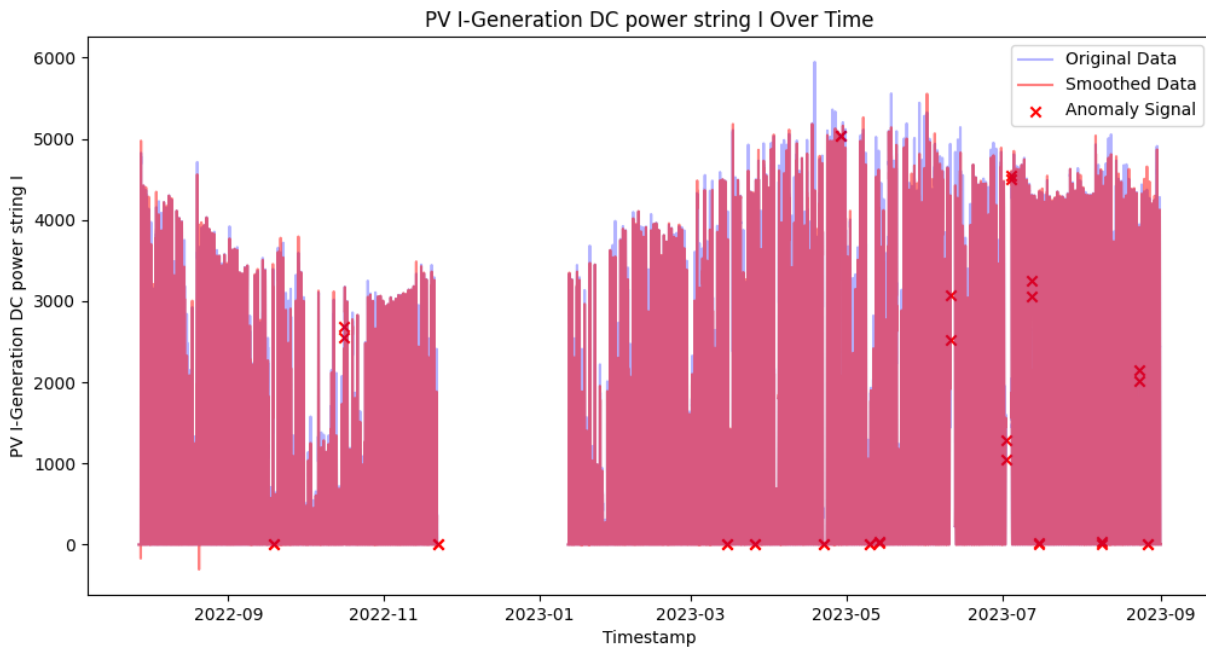


Figure 14 Διάγραμμα παραγωγής ισχύος σε ένα από τα δυο string του πρώτου φωτοβολταϊκού συστήματος

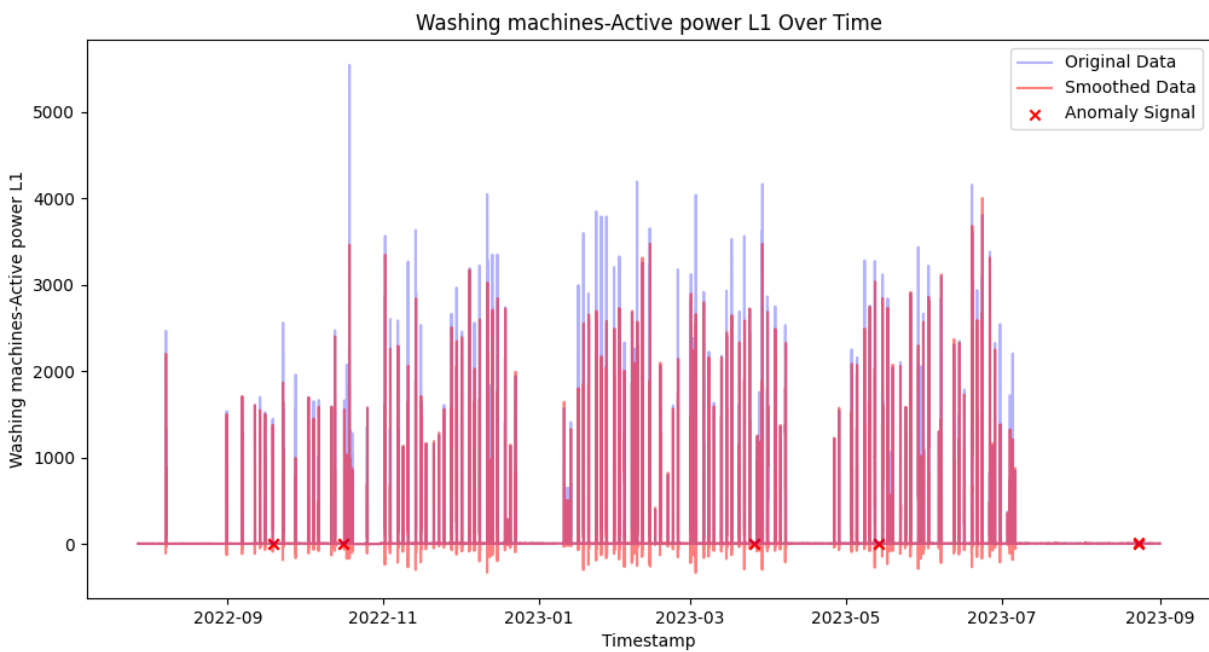


Figure 15 Διάγραμμα κατανάλωσης πραγματικής ισχύος πρώτης φάσης στην κατανάλωση των πλυντηρίων του ζωντανού εργαστηρίου

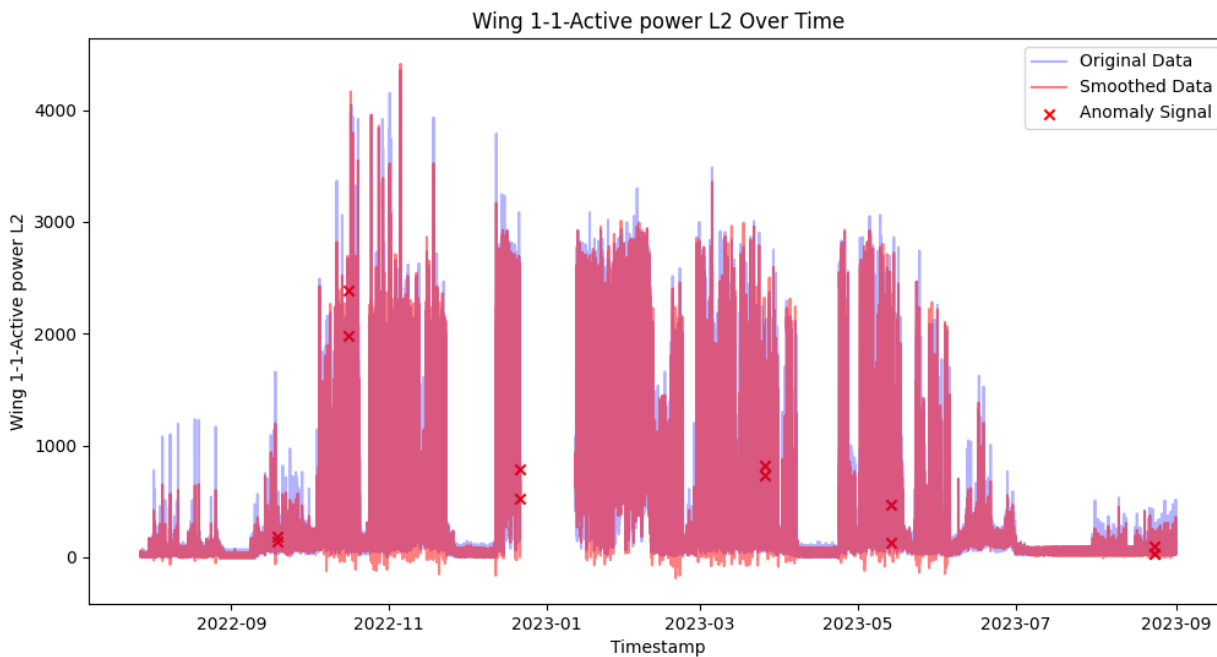


Figure 16 Διάγραμμα κατανάλωσης πραγματικής ισχύος της δεύτερης φάσης για ένα τμήμα της πρώτης πτέρυγας

Ολοκληρώνοντας την αξιολόγηση της ποιότητας των δεδομένων, είναι προφανές ότι τα δεδομένα από το Ζωντανό Εργαστήριο του Πανεπιστημίου Δυτικής Μακεδονίας έχουν ικανοποιητική αξιοπιστία. Στα δεδομένα εμφανίζεται ένα κενό περίπου 1 μήνα το διάστημα Δεκέμβριος 2022 – Ιανουάριος 2023, το οποίο οφείλεται σε γνωστή βλάβη. Μέσω της χρήσης της Python και ενός συστηματικού πλαισίου, αξιολογήθηκαν διάφορες πτυχές της ποιότητας των δεδομένων. Οι έλεγχοι των δεδομένων περιελάμβαναν τον εντοπισμό και την ποσοτικοποίηση των ελλειπών τιμών, την ανίχνευση διπλών χρονοσφραγίδων και την επικύρωση της χρονολογικής σειράς των χρονοσφραγίδων. Οι ακραίες τιμές εντοπίστηκαν αποτελεσματικά χρησιμοποιώντας τεχνικές με βάση το Z-Score και οι μετρικές στήλες επικυρώθηκαν για τη συνέπεια του εύρους δεδομένων. Η εξομάλυνση των δεδομένων επιτεύχθηκε επιτυχώς με τη χρήση του φίλτρου Savitzky-Golay, βελτιώνοντας τη συνολική ποιότητα των δεδομένων. Οι οπτικοποιήσεις για κάθε μετρική στήλη παρείχαν πληροφορίες και οπτική κατανόηση των προτύπων και της συμπεριφοράς των δεδομένων.

Η εισαγωγή του μηχανισμού ανίχνευσης ανωμαλιών με βάση το προφίλ που χρησιμοποιεί το Stumpy αποδεικνύει περαιτέρω την ενδελεχή αξιολόγηση των δεδομένων και τη δέσμευση για τη διασφάλιση της αξιοπιστίας του συνόλου δεδομένων.

Παράρτημα κώδικα:

Κώδικας έλεγχου χρονοσειρών

```
import os
import sys
import numpy as np
import matplotlib.pyplot as plt
import pandas as pd
import stumpy
from datetime import datetime, date
from scipy.stats import zscore
from scipy.signal import savgol_filter

abspath = os.path.abspath(__file__)
dname = os.path.dirname(abspath)
os.chdir(dname)

output_file = open('output22.txt', 'w')
sys.stdout = output_file
# Step 1: Load the Excel file into a DataFrame
file_path = 'esties-UOWM_15m_all.xlsx'

# Read the data and replace "NaN" strings with actual NaN values during reading
df = pd.read_excel(file_path)
df = df.applymap(lambda x: None if x == "NaN" else x) #apply pandas NaN values

# Step 2: Check for missing values and count NaN values in metric columns
missing_values = df.isnull().sum()
print("Missing Values:")
print(missing_values)

metric_columns = df.columns[1:] #exclude timestamp

# Step 4: Check for duplicate timestamps
duplicate_timestamps = df['timestamp'].duplicated().sum()
print("\nDuplicate Timestamps:")
print(duplicate_timestamps)

# Step 5: Check timestamp order
df['timestamp'] = pd.to_datetime(df['timestamp'], format='%d/%m/%y %H:%M')
df_sorted = df.sort_values(by='timestamp')
if df.equals(df_sorted):
    print("\nTimestamps are in chronological order.")
else:
    print("\nTimestamps are not in chronological order.")
```

```

# Step 6: Outliers Detection
def detect_outliers_zscore(data, threshold=3):
    z_scores = zscore(data)
    outliers = (np.abs(z_scores) > threshold)
    return outliers

# Apply outliers detection to numeric columns
outliers = df[metric_columns].apply(detect_outliers_zscore)
print("\nOutliers Detection (Z-Score):")
print(outliers)

# Step 7: Data Range Validation, Data Smoothing, and Visualization (Example)
for col in metric_columns:
    # Data Range Validation
    min_val = df[col].min()
    max_val = df[col].max()
    print(f"\n{col} Range: {min_val} - {max_val}")
    print("Column:", col, "has", df[col].isna().sum(axis = 0), "NaN Values")

    # Data Smoothing (Savitzky-Golay filter)
    smoothed_data = savgol_filter(df[col], window_length=5, polyorder=2)
    # Abnormal detection's
    time_series = df[col].values

    # Compute the matrix profile using stumpy
    matrix_profile = stumpy.stump(time_series, m=3)
    matrix_profile_values = matrix_profile[:, 0]
    df_anom= pd.DataFrame()
    col_anomaly=f'{col}_anom'
    col_anomaly_signal=f'{col}_anom_sing'
    df_anom[col_anomaly] = np.nan
    df_anom[col_anomaly_signal] = np.nan
    df_anom[col_anomaly] = matrix_profile_values
    threshold = df[col].std()    ### Get volatility price of column values
    df_anom[col_anomaly_signal] = df_anom[col_anomaly] > threshold
    df_anom['timestamp'] = df['timestamp']
    df = pd.merge(df, df_anom, on='timestamp', how='left')

    df[col_anomaly_signal].fillna(False, inplace=True)

    # Mark anomalies with a scatter plot
    anomaly_indices = df[df[col_anomaly_signal]].index
    print(f'Abnormal signals in {col} - {df[col_anomaly_signal].sum()}')
    plt.figure(figsize=(12, 6))
    plt.plot(df['timestamp'], df[col], label='Original Data',
color='blue', alpha=0.3)
    plt.plot(df['timestamp'], smoothed_data, label='Smoothed Data',
color='red', alpha=0.5)
    plt.scatter(df.loc[anomaly_indices, 'timestamp'], df.loc[anomaly_indices,
col], c='red', marker='x', label='Anomaly Signal', alpha=1.0)
    plt.xlabel('Timestamp')

```

```

plt.ylabel(col)
plt.title(f'{col} Over Time')
plt.legend()
#plt.show()
# Step 8: Additional checks based on domain-specific rules and requirements
# Add custom checks here if needed
# Step 9: Print data quality summary
print("\nData Quality Summary:")
print("Number of Rows:", len(df))
print("Number of Columns:", len(df.columns))
output_file.close()

```

Κώδικας μορφοποίησης αποτελεσμάτων σε πίνακα

```

import pandas as pd
from tabulate import tabulate
import os

abspath = os.path.abspath(__file__)
dname = os.path.dirname(abspath)
os.chdir(dname)
# Your data as a list of lists
data = [
    ["Active power sum L1-L3 P-Grid", "Range: -9161.632 - 12759.988"],
    ["Column: Active power sum L1-L3 P-Grid has", "24113 NaN Values"],
    ["Abnormal signals in Active power sum L1-L3 P-Grid", "- 24123"],
    ["Battery I-Charging DC power", "Range: -5720.613 - 5410.209"],
    ["Column: Battery I-Charging DC power has", "5405 NaN Values"],
    ["Abnormal signals in Battery I-Charging DC power", "- 5439"],
    ["Battery I-Discharging DC power", "Range: -5720.613 - 5410.209"],
    ["Column: Battery I-Discharging DC power has", "5405 NaN Values"],
    ["Abnormal signals in Battery I-Discharging DC power", "- 5439"],
    ["Battery II-Charging DC power", "Range: -5682.927 - 5433.2"],
    ["Column: Battery II-Charging DC power has", "5543 NaN Values"],
    ["Abnormal signals in Battery II-Charging DC power", "- 5577"],
    ["Battery II-Discharging DC power", "Range: -5682.927 - 5433.2"],
    ["Column: Battery II-Discharging DC power has", "5543 NaN Values"],
    ["Abnormal signals in Battery II-Discharging DC power", "- 5577"],
    ["Communal lights-Active power L1", "Range: -4.671 - 2543.615"],
    ["Column: Communal lights-Active power L1 has", "2087 NaN Values"],
    ["Abnormal signals in Communal lights-Active power L1", "- 2100"],
    ["Communal lights-Active power L2", "Range: 0.0 - 908.123"],
    ["Column: Communal lights-Active power L2 has", "2087 NaN Values"],
    ["Abnormal signals in Communal lights-Active power L2", "- 2100"],
    ["Communal lights-Active power L3", "Range: -4.877 - 1881.028"],
    ["Column: Communal lights-Active power L3 has", "2087 NaN Values"],
    ["Abnormal signals in Communal lights-Active power L3", "- 2100"],
    ["Communal lights-Active power sum L1-L3", "Range: -3.396 - 5056.165"],
    ["Column: Communal lights-Active power sum L1-L3 has", "2087 NaN Values"],
    ["Abnormal signals in Communal lights-Active power sum L1-L3", "- 2100"],
    ["Communal lights-Reactive power L1", "Range: -46.641 - 4390.547"],
    ["Column: Communal lights-Reactive power L1 has", "2087 NaN Values"],
    ["Abnormal signals in Communal lights-Reactive power L1", "- 2100"],

```



```

["Communal lights-Reactive power L2", "Range: -56.287 - 1333.405"],
["Column: Communal lights-Reactive power L2 has", "2087 NaN Values"],
["Abnormal signals in Communal lights-Reactive power L2", "- 2100"],
["Communal lights-Reactive power L3", "Range: -44.666 - 2836.981"],
["Column: Communal lights-Reactive power L3 has", "2087 NaN Values"],
["Abnormal signals in Communal lights-Reactive power L3", "- 2100"],
["Communal lights-Reactive power sum L1-L3", "Range: -143.062 - 8002.834"],
["Column: Communal lights-Reactive power sum L1-L3 has", "2087 NaN Values"],
["Abnormal signals in Communal lights-Reactive power sum L1-L3", "- 2100"],
["Inverter I output-Active power L1", "Range: -1276.06 - 3094.653"],
["Column: Inverter I output-Active power L1 has", "129 NaN Values"],
["Abnormal signals in Inverter I output-Active power L1", "- 143"],
["Inverter I output-Active power L2", "Range: -916.192 - 3341.867"],
["Column: Inverter I output-Active power L2 has", "129 NaN Values"],
["Abnormal signals in Inverter I output-Active power L2", "- 143"],
["Inverter I output-Active power L3", "Range: -923.085 - 3425.24"],
["Column: Inverter I output-Active power L3 has", "129 NaN Values"],
["Abnormal signals in Inverter I output-Active power L3", "- 143"],
["Inverter I output-Active power sum L1-L3", "Range: -3010.608 -
9831.739"],
["Column: Inverter I output-Active power sum L1-L3 has", "129 NaN Values"],
["Abnormal signals in Inverter I output-Active power sum L1-L3", "- 143"],
["Inverter II output-Active power L1", "Range: -927.266 - 3350.046"],
["Column: Inverter II output-Active power L1 has", "129 NaN Values"],
["Abnormal signals in Inverter II output-Active power L1", "- 143"],
["Inverter II output-Active power L2", "Range: -952.163 - 3355.15"],
["Column: Inverter II output-Active power L2 has", "129 NaN Values"],
["Abnormal signals in Inverter II output-Active power L2", "- 143"],
["Inverter II output-Active power L3", "Range: -957.641 - 3439.904"],
["Column: Inverter II output-Active power L3 has", "129 NaN Values"],
["Abnormal signals in Inverter II output-Active power L3", "- 143"],
["Inverter II output-Active power sum L1-L3", "Range: -2837.101 -
10127.759"],
["Column: Inverter II output-Active power sum L1-L3 has", "129 NaN Values"],
["Abnormal signals in Inverter II output-Active power sum L1-L3", "- 143"],
["Inverter II supplied loads-Active power sum L1-L3", "Range: 0.0 -
18563.338"],
["Column: Inverter II supplied loads-Active power sum L1-L3 has", "5543 NaN
Values"],
["Abnormal signals in Inverter II supplied loads-Active power sum L1-L3", "-
5577"],
["Inverter II supplied loads-Reactive power sum L1-L3", "Range: -384.2 -
1248.902"],
["Column: Inverter II supplied loads-Reactive power sum L1-L3 has", "5543
NaN Values"],
["Abnormal signals in Inverter II supplied loads-Reactive power sum L1-L3",
"- 5577"],
["Inverter III output-Active power L1", "Range: -10665.969 - 0.0"],
["Column: Inverter III output-Active power L1 has", "1507 NaN Values"],
["Abnormal signals in Inverter III output-Active power L1", "- 1543"],
["Inverter III output-Active power L2", "Range: -9397.582 - 0.0"],
["Column: Inverter III output-Active power L2 has", "1507 NaN Values"],

```

```

["Abnormal signals in Inverter III output-Active power L2", "- 1543"],
["Inverter III output-Active power L3", "Range: -8691.14 - 0.0"],
["Column: Inverter III output-Active power L3 has", "1507 NaN Values"],
["Abnormal signals in Inverter III output-Active power L3", "- 1543"],
["Inverter III output-Active power sum L1-L3", "Range: -23284.236 - 0.0"],
["Column: Inverter III output-Active power sum L1-L3 has", "1507 NaN
Values"],
["Abnormal signals in Inverter III output-Active power sum L1-L3", "-
1543"],
["Inverter III output-Janitza-Active power L1", "Range: -177.611 -
3352.05"],
["Column: Inverter III output-Janitza-Active power L1 has", "67 NaN
Values"],
["Abnormal signals in Inverter III output-Janitza-Active power L1", "- 80"],
["Inverter III output-Janitza-Active power L2", "Range: -182.792 -
3375.426"],
["Column: Inverter III output-Janitza-Active power L2 has", "67 NaN
Values"],
["Abnormal signals in Inverter III output-Janitza-Active power L2", "- 80"],
["Inverter III output-Janitza-Active power L3", "Range: -190.168 -
3437.056"],
["Column: Inverter III output-Janitza-Active power L3 has", "67 NaN
Values"],
["Abnormal signals in Inverter III output-Janitza-Active power L3", "- 80"],
["Inverter III output-Janitza-Active power sum L1-L3", "Range: -548.906 -
10148.962"],
["Column: Inverter III output-Janitza-Active power sum L1-L3 has", "67 NaN
Values"],
["Abnormal signals in Inverter III output-Janitza-Active power sum L1-L3",
"- 80"],
["PCC with the grid-Active power L1", "Range: -8397.662 - 20220.725"],
["Column: PCC with the grid-Active power L1 has", "125 NaN Values"],
["Abnormal signals in PCC with the grid-Active power L1", "- 139"],
["PCC with the grid-Active power L2", "Range: -9291.002 - 23647.301"],
["PCC with the grid-Active power L2", "Range: -9291.002 - 23647.301"],
["Column: PCC with the grid-Active power L2 has", "125 NaN Values"],
["Abnormal signals in PCC with the grid-Active power L2", "- 139"],
["PCC with the grid-Active power L3", "Range: -8687.854 - 23980.049"],
["Column: PCC with the grid-Active power L3 has", "125 NaN Values"],
["Abnormal signals in PCC with the grid-Active power L3", "- 139"],
["PCC with the grid-Active power sum L1-L3", "Range: -25685.002 -
67847.036"],
["Column: PCC with the grid-Active power sum L1-L3 has", "125 NaN Values"],
["Abnormal signals in PCC with the grid-Active power sum L1-L3", "- 139"],
["PCC with the grid-Reactive power L1", "Range: -1711.723 - 869.201"],
["Column: PCC with the grid-Reactive power L1 has", "125 NaN Values"],
["Abnormal signals in PCC with the grid-Reactive power L1", "- 139"],
["PCC with the grid-Reactive power L2", "Range: -1625.108 - 642.753"],
["Column: PCC with the grid-Reactive power L2 has", "125 NaN Values"],
["Abnormal signals in PCC with the grid-Reactive power L2", "- 139"],
["PCC with the grid-Reactive power L3", "Range: -1804.283 - 767.625"],
["Column: PCC with the grid-Reactive power L3 has", "125 NaN Values"],

```

```

["Abnormal signals in PCC with the grid-Reactive power L3", "- 139"],
["PCC with the grid-Reactive power sum L1-L3", "Range: -5100.0 - 2288.788"],
["Column: PCC with the grid-Reactive power sum L1-L3 has", "125 NaN
Values"],
["Abnormal signals in PCC with the grid-Reactive power sum L1-L3", "- 139"]
]

# Create a DataFrame from the table data
df = pd.DataFrame(data, columns=["Measurement", "Value"])

# Save the DataFrame to an Excel file
excel_file = "measurement_data.xlsx"
df.to_excel(excel_file, index=False, engine='openpyxl')

print(f>Data has been saved to {excel_file}")

```

Βιβλιογραφία:

- [1] Εισαγωγή στον προγραμματισμό με αρωγό τη γλώσσα python , Γεώργιος Μανής , ΣΕΕΑΒ, Καλλιπος , 2015
- [2] Introduction to programming with python, Vivian Bailey, Copyright © 2015 by Software development training
- [3] Deep time series forecasting with python , Dr. N.D Lewis, Copyright © 2016 by N.D. Lewis
- [4][5][6] ractical time series analysis , dr. avishek pal & dr. pks prakash copyright © 2017 packt publishing

