

ΠΑΝΕΠΙΣΤΗΜΙΟ ΙΩΑΝΝΙΝΩΝ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ



ΤΕΧΝΙΚΕΣ ΑΝΑΛΥΣΗΣ ΕΝΕΡΓΕΙΑΚΩΝ ΔΕΔΟΜΕΝΩΝ

ΠΤΥΧΙΑΚΗ ΕΡΓΑΣΙΑ

ΓΙΑΛΟΥΡΗ ΕΡΓΙΝΑ ΔΕΣΠΟΙΝΑ

A.M:1469

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ:

ΓΡΗΓΟΡΙΟΣ ΔΟΥΜΕΝΗΣ

ΑΡΤΑ, ΙΟΥΝΙΟΣ 2022

ENERGY DATA ANALYSIS METHODOLOGIES

Εγκρίθηκε από τριμελή εξεταστική επιτροπή
Άρτα, 22/07/2022

ΕΠΙΤΡΟΠΗ ΑΞΙΟΛΟΓΗΣΗΣ

1. Επιβλέπων καθηγητής
Γρηγόριος Δουμένης
Επίκουρος Καθηγητής
2. Μέλος επιτροπής
Νικόλαος Γιαννακέας
Επίκουρος Καθηγητής
3. Μέλος επιτροπής
Σπυριδούλα Μαργαρίτη
Μέλος ΕΔΠΠ

© Γιαλούρη, Εργίνα Δέσποινα, 2022.

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Δήλωση μη λογοκλοπής

Δηλώνω υπεύθυνα και γνωρίζοντας τις κυρώσεις του Ν. 2121/1993 περί Πνευματικής Ιδιοκτησίας, ότι η παρούσα μεταπτυχιακή εργασία είναι εξ ολοκλήρου αποτέλεσμα δικής μου ερευνητικής εργασίας, δεν αποτελεί προϊόν αντιγραφής ούτε προέρχεται από ανάθεση σε τρίτους. Όλες οι πηγές που χρησιμοποιήθηκαν (κάθε είδους, μορφής και προέλευσης) για τη συγγραφή της περιλαμβάνονται στη βιβλιογραφία.

Γιαλούρη Εργίνα Δέσποινα

Υπογραφή

Στην οικογένεια μου

ΠΕΡΙΛΗΨΗ

Στην παρούσα πτυχιακή εργασία θα γίνει αναφορά σε τεχνικές που αφορούν στην ανάλυση και μπορούν να εφαρμοστούν σε ενεργειακά δεδομένα τα οποία προέρχονται από ανανεώσιμες πηγές ενέργειας και πιο συγκεκριμένα από πηγές φωτοβολταϊκών στοιχείων. Αυτά τα δεδομένα αποτελούν χρονοσειρές που παρουσιάζονται με ιδιαίτερα χαρακτηριστικά και ιδιότητες ενώ συγχρόνως εμφανίζουν ιδιαίτερα χαρακτηριστικά ανάλογα με το χρονικό διάστημα που εξετάζεται. Κύριο στόχο αποτελεί η ανάλυση δεδομένων και τα χαρακτηριστικά από τα οποία αποτελούνται τα δεδομένα αυτά, ακόμη ένας σημαντικός στόχος είναι η παρουσίαση διαφόρων τεχνικών όπως της παλινδρόμησης και της μηχανικής μάθησης καθώς και η σωστή προεπεξεργασία δεδομένων με βάση τα χαρακτηριστικά που έχουν παρατηρηθεί στην ανάλυση δεδομένων. Τέλος, στις τεχνικές που θα αναλυθούν θα παρουσιαστούν κατά κύριο λόγο βιβλιογραφικά, μοντέλα που θα μπορούσαν να βοηθήσουν στις προβλέψεις αντίστοιχων χρονοσειρών.

ABSTRACT

In this thesis, reference will be made to techniques related to the analysis that can be applied to energy data that comes from renewable energy sources and more specifically from sources of photovoltaic elements. These data are time series that are presented with particular characteristics and properties while at the same time displaying particular characteristics depending on the period of time under consideration. The main objective is the analysis of data and the characteristics of which these data are composed, another important objective is the presentation of various techniques such as regression and machine learning as well as the correct preprocessing of data based on the characteristics observed in the data analysis. Finally, in the techniques that will be analyzed, mainly bibliographic models will be presented that could help in the predictions of respective time series.

Λίστα Εικόνων

Εικόνα 1 Ανανεώσιμες Πηγές Ενέργειας.....	3
Εικόνα 2 Κατηγορίες ΑΠΕ	4
Εικόνα 3 Κατηγορίες Μη Ανανεώσιμων Πηγών Ενέργειας.....	5
Εικόνα 4 Φωτοβολταϊκά Και Ανεμογεννήτριες.....	7
Εικόνα 5 Ενεργειακή Κατανάλωση Ευρωπαϊκών Χωρών.....	8
Εικόνα 6 Χαρακτηριστικά Δεδομένων.....	9
Εικόνα 7 Διαγράμματα Ανάλυσης Εννοιών.....	10
Εικόνα 8 Βήματα Προεπεξεργασίας Δεδομένων	11
Εικόνα 9 Απεικόνιση Διαφόρων Χρονοσειρών	13
Εικόνα 10 Γραφήματα με Στάσιμες και Μη Χρονοσειρές.....	14
Εικόνα 11 Γραφήματα με Χαρακτηριστικά Χρονοσειρών.....	15
Εικόνα 12 Κατηγορίες Χρονοσειρών	15
Εικόνα 13 Μηχανική Μάθηση.....	20
Εικόνα 14 Είδη Μηχανικής Μάθησης	21
Εικόνα 15 Decision Trees.....	22
Εικόνα 16 K Nearest Neighbors.....	23
Εικόνα 17 Ενισχυτική Μάθηση.....	24
Εικόνα 18 Στάδια Μηχανικής Μάθησης	25
Εικόνα 19 Κενές Τιμές ορισμένες με τον χαρακτήρα True	32
Εικόνα 20 Στήλες Δεδομένων που Μελετήθηκαν.....	32
Εικόνα 21 Κενές Τιμές που Εντοπίστηκαν στο Αρχείο Ανά Λεπτό	33
Εικόνα 22 Σύγκριση Τιμών για Κατανόηση του Αλγορίθμου	33
Εικόνα 23 Επεξήγηση Κώδικα	34
Εικόνα 24 Flowchart.....	34
Εικόνα 25 Δύο Τυχαίες Κενές Τιμές	35
Εικόνα 26 Συμπλήρωση Των Κενών Τιμών	35
Εικόνα 27 Γράφημα με Τιμές Πριν και Μετά	36
Εικόνα 28 Τιμές Πριν και Μετά τη Πρόβλεψη	36

Πίνακας Περιεχομένων

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ	1
ΚΕΦΑΛΑΙΟ 2: ΑΝΑΝΕΩΣΙΜΕΣ ΠΗΓΕΣ ΕΝΕΡΓΕΙΑΣ.....	2
2.1 ΟΡΙΣΜΟΣ ΕΝΕΡΓΕΙΑΣ	2
2.2 ΜΟΡΦΕΣ ΕΝΕΡΓΕΙΑΣ.....	2
2.3 ΑΝΑΝΕΩΣΙΜΕΣ ΠΗΓΕΣ ΕΝΕΡΓΕΙΑΣ	3
2.3.1 ΟΡΙΣΜΟΣ ΑΝΑΝΕΩΣΙΜΩΝ ΠΗΓΩΝ ΕΝΕΡΓΕΙΑΣ	3
2.3.2 ΚΑΤΗΓΟΡΙΕΣ ΑΝΑΝΕΩΣΙΜΩΝ ΠΗΓΩΝ ΕΝΕΡΓΕΙΑΣ	4
2.3.3 ΟΡΙΣΜΟΣ ΜΗ ΑΝΑΝΕΩΣΙΜΩΝ ΠΗΓΩΝ ΕΝΕΡΓΕΙΑΣ	4
2.3.4 ΚΑΤΗΓΟΡΙΕΣ ΜΗ ΑΝΑΝΕΩΣΙΜΩΝ ΠΗΓΩΝ ΕΝΕΡΓΕΙΑΣ.....	5
2.4 ΠΡΟΚΛΗΣΕΙΣ ΑΠΟ ΤΗΝ ΕΙΣΑΓΩΓΗ ΑΝΑΝΕΩΣΙΜΩΝ ΠΗΓΩΝ ΕΝΕΡΓΕΙΑΣ	5
ΚΕΦΑΛΑΙΟ 3: ΔΕΔΟΜΕΝΑ ΠΑΡΑΓΩΓΗΣ ΑΠΕ	7
3.1 ΠΗΓΕΣ ΔΕΔΟΜΕΝΩΝ	7
3.2 ΜΑΖΙΚΟΤΗΤΑ ΔΕΔΟΜΕΝΩΝ	7
3.3 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΔΕΔΟΜΕΝΩΝ	9
3.4 ΠΟΙΟΤΗΤΑ ΔΕΔΟΜΕΝΩΝ	11
ΚΕΦΑΛΑΙΟ 4: ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ (ΑΝΑΛΥΤΙΚΕΣ ΤΕΧΝΙΚΕΣ-REGRESSION)	13
4.1 ΟΡΙΣΜΟΣ ΧΡΟΝΟΣΕΙΡΑΣ.....	13
4.2 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΧΡΟΝΟΣΕΙΡΩΝ.....	13
4.3 ΚΑΤΗΓΟΡΙΕΣ	15
4.4 ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ	16
4.5 ΑΝΑΛΥΣΗ ΜΟΝΤΕΛΩΝ ΜΕ ΤΕΧΝΙΚΕΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ	16
4.5.1 ΜΟΝΤΕΛΟ ΑΥΤΟΠΑΛΙΝΔΡΟΜΗΣΗΣ (AutoRegression-AR).....	17
4.5.2 ΜΟΝΤΕΛΟ ΚΙΝΗΤΟΥ ΜΕΣΟΥ (MOVING AVERAGE - MA).....	17
4.5.3 ΜΟΝΤΕΛΟ ΑΥΤΟΠΑΛΙΝΔΡΟΜΟΥ ΚΙΝΗΤΟΥ ΜΕΣΟΥ (AutoRegression Moving Average - ARMA)	18
4.5.4 ΑΥΤΟΠΑΛΙΝΔΡΟΜΙΚΟΥ ΜΟΝΤΕΛΟ ΚΙΝΗΤΟΥ ΜΕΣΟΥ ΟΡΟΥ (ARIMA)	18
4.5.5 ΜΟΝΤΕΛΟ SARIMA (SEASONAL AYTOREGRESSION INTEGRATED MOVING AVERAGE)	19
ΚΕΦΑΛΑΙΟ 5: ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ(ΤΕΧΝΙΚΕΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ).....	20
5.1 ΟΡΙΣΜΟΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ.....	20
5.2 ΕΙΔΗ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ	21
5.2.1 ΜΑΘΗΣΗ ΜΕ ΕΠΙΒΛΕΨΗ (Supervised Learning)	21
5.2.2 ΜΑΘΗΣΗ ΧΩΡΙΣ ΕΠΙΒΛΕΨΗ (Unsupervised Learning).....	23
5.2.3 ΕΝΙΣΧΥΤΙΚΗ ΜΑΘΗΣΗ (Reinforcement Learning).....	24
5.3 ΣΤΑΔΙΑ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ	25

5.4 ΜΟΝΤΕΛΑ.....	26
ΚΕΦΑΛΑΙΟ 6: ΒΕΛΤΙΩΣΗ ΠΟΙΟΤΗΤΑΣ ΔΕΔΟΜΕΝΩΝ ΑΠΕ	28
6.1 ΒΕΛΤΙΩΣΗ ΠΟΙΟΤΗΤΑΣ ΔΕΔΟΜΕΝΩΝ	28
6.2 ΑΝΑΛΥΣΗ ΕΝΕΡΓΕΙΑΚΩΝ ΔΕΔΟΜΕΝΩΝ	29
6.2.1 ΕΝΕΡΓΗ ΙΣΧΥΣ.....	29
6.2.2 ΤΑΣΗ	29
6.3 ΧΑΜΕΝΕΣ ΤΙΜΕΣ.....	30
ΚΕΦΑΛΑΙΟ 7: ΣΥΜΠΕΡΑΣΜΑΤΑ.....	37
ΒΙΒΛΙΟΓΡΑΦΙΑ	38
ΠΗΓΕΣ ΕΙΚΟΝΩΝ	44

ΚΕΦΑΛΑΙΟ 1: ΕΙΣΑΓΩΓΗ

Στην παρούσα πτυχιακή εργασία, στο πρώτο κεφάλαιο θα συζητηθεί η ανάλυση των κεφαλαίων που πρόκειται να υπάρξουν.

Στο δεύτερο κεφάλαιο δίνεται ένας γενικός ορισμός της ενέργειας καθώς επίσης πρόκειται να δοθεί και ένας ορισμός που αφορά τις ανανεώσιμες πηγές ενέργειας. Επίσης θα γίνει αναφορά στις προκλήσεις που έχουν προκύψει από την εισαγωγή τους στον σύγχρονο κόσμο.

Στο τρίτο κεφάλαιο, θα γίνει αναφορά στις πηγές των δεδομένων, δηλαδή από που μπορούν να προέρχονται τα δεδομένα, ποια δεδομένα τίθενται προς ανάλυση, τα χαρακτηριστικά τους καθώς και τον παράγοντα της προεπεξεργασίας των δεδομένων.

Στο τέταρτο κεφάλαιο θα παρουσιαστούν βασικοί ορισμοί όπως είναι ο ορισμός της χρονοσειράς αλλά θα γίνει ακόμα σημαντική αναφορά και σε αρκετά μοντέλα τα οποία προκύπτουν από τις τεχνικές παλινδρόμησης που μπορούν να χρησιμοποιηθούν.

Στο πέμπτο κεφάλαιο θα γίνει εξίσου μία σημαντική αναφορά στον τομέα της μηχανικής μάθησης καθώς και στις τεχνικές της. Επίσης θα γίνει παρουσίαση διαφόρων μοντέλων που χρησιμοποιούνται με βάση την μηχανική μάθηση.

Στο έκτο κεφάλαιο, θα μιλήσουμε για την βελτίωση που επιδέχονται τα ενεργειακά δεδομένα των ανανεώσιμων πηγών ενέργειας που έχουμε επιλέξει τόσο στον τομέα της παλινδρόμησης όσο και στον τομέα της μηχανικής μάθησης.

Στο έβδομο και τελευταίο κεφάλαιο της παρούσας πτυχιακής εργασίας παρουσιάζονται τα συμπεράσματα τα οποία προκύπτουν ύστερα από την προεπεξεργασία και την ανάλυση που γίνεται στα δεδομένα χρησιμοποιώντας τεχνικές που αφορούν τον τομέα της παλινδρόμησης όσο και τον τομέα της μηχανικής μάθησης.

ΚΕΦΑΛΑΙΟ 2: ΑΝΑΝΕΩΣΙΜΕΣ ΠΗΓΕΣ ΕΝΕΡΓΕΙΑΣ

2.1 ΟΡΙΣΜΟΣ ΕΝΕΡΓΕΙΑΣ

Η έννοια της ενέργειας προέρχεται από την ικανότητα που ήδη υπάρχει για την παραγωγή έργου, δηλαδή για παραγωγή δύναμης σε μία προσδιορισμένη απόσταση. Η δύναμη αυτή μπορεί να προέρχεται τόσο από ανθρώπινο όσο και από κάποιο μηχανικό παράγοντα αναλόγως το παράδειγμα και το πρόβλημα που τίθεται προς επίλυση κάθε φορά. Όπως γίνεται αντιληπτό λοιπόν σε κάθε ανθρώπινη δραστηριότητα και όχι μόνο υπάρχει η έννοια της ενέργειας και έχει τη δυνατότητα να εναλλάσσεται σε πολλές διαφορετικές μορφές. (Νικόλαος, 2015) Ιδιαίτερα στις σύγχρονες κοινωνίες ο τομέας της ενέργειας αποτελεί βασικό αγαθό για τον άνθρωπο και γι' αυτό εμφανίζεται σε μεγάλα ποσά παραγωγής, κατανάλωσης αλλά και αποθήκευσης. Για λόγους που αφορούν τόσο οικονομικούς όσο και περιβαλλοντικούς παράγοντες υπάρχει ανάπτυξη στην παραγωγή ενέργειας μέσω φυσικών πόρων που ονομάζονται Ανανεώσιμες Πηγές Ενέργειας (ΑΠΕ). (Μαγίδου, 2016)

2.2 ΜΟΡΦΕΣ ΕΝΕΡΓΕΙΑΣ

Όσον αφορά τις μορφές ενέργειας των ΑΠΕ χαρακτηρίζονται από πολλές υποκατηγορίες. Μια εξ αυτών είναι σίγουρα η ηλεκτρική ενέργεια την οποία μπορούμε με ευκολία να την εντοπίσουμε σε συσσωρευτές λόγω του ηλεκτρικού φορτίου που έχουμε. Ακόμη έχουμε την χημική ενέργεια η οποία εμφανίζεται λόγω των ουσιών που υπάρχουν αλλά και λόγω της τροποποίησης που μπορεί να υποστεί η χημική τους δομή. Στη συνέχεια υποκατηγορία αποτελεί και η μηχανική ενέργεια που αποκτά εφαρμογή σε υλικά μέσα και είναι συνδυασμός της κινητικής και της δυναμικής ενέργειας. Επίσης έχουμε την θερμική ενέργεια που στηρίζεται στη παραγωγή θερμότητας. Εξίσου σημαντική μορφή ενέργειας μπορεί να θεωρηθεί η ηλεκτρομαγνητική ενέργεια που στηρίζεται στα κύματα τόσο του ηλεκτρικού όσο και του μαγνητικού πεδίου. Τελευταία και ιδιαίτερα σημαντική είναι η πυρηνική ενέργεια που βασίζεται στη διάσπαση του πυρήνα ενός ατόμου. (Μαγίδου, 2016)

Συνοψίζοντας, η ενέργεια διαθέτει την ιδιότητα που μπορεί να προέρχεται τόσο από ανανεώσιμες όσο και από μη ανανεώσιμες πηγές ενέργειας χωρίς να αλλοιώνεται ή να καταστρέφεται.

2.3 ΑΝΑΝΕΩΣΙΜΕΣ ΠΗΓΕΣ ΕΝΕΡΓΕΙΑΣ



Εικόνα 1 Ανανεώσιμες Πηγές Ενέργειας

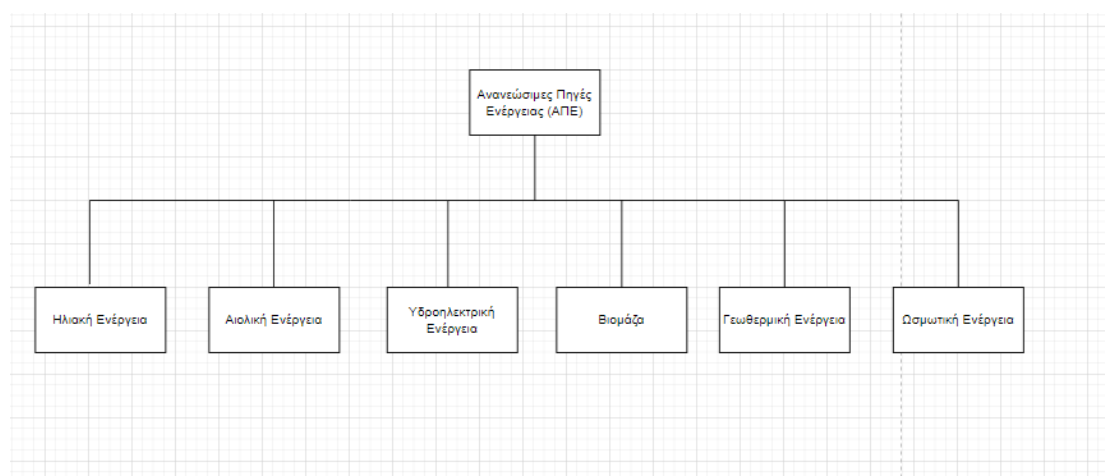
Οι πηγές ενέργειας είναι οι φυσικοί πόροι που γίνονται εκμεταλλεύσιμοι από τον ανθρώπινο παράγοντα ώστε να αποκτήσει ενέργεια και να την χρησιμοποιήσει τόσο στις πόλεις τις οποίες ζει όσο και στις βιομηχανίες στις οποίες αναπτύσσει σημαντική δραστηριότητα. Κατά αυτό τον τρόπο οι πηγές ενέργειας ταξινομούνται τόσο σε ανανεώσιμες όσο και σε μη ανανεώσιμες πηγές ενέργειας.

2.3.1 ΟΡΙΣΜΟΣ ΑΝΑΝΕΩΣΙΜΩΝ ΠΗΓΩΝ ΕΝΕΡΓΕΙΑΣ

Οι Ανανεώσιμες Πηγές Ενέργειας ή ΑΠΕ χαρακτηρίζονται ως τεχνολογία που προσφέρει λύσεις αφού η παραγωγή της γίνεται μέσω της φύσης. Οι ανανεώσιμες πηγές διαθέτουν την ικανότητα τόσο της γρήγορης όσο και της άμεσης ανανέωσης τους καθώς όπως προαναφέρθηκε η παραγωγή τους γίνεται μέσω της φύσης δηλαδή μέσω του ήλιου, του νερού και του αέρα κατά κύρια βάση (Ανοπ., n.d.). Εξάλλου γι' αυτό το λόγο χαρακτηρίζονται ως ανεξάντλητες και ταυτόχρονα παρουσιάζουν σημαντικά περιβαλλοντικά οφέλη και βοηθούν στην τεχνολογική ανάπτυξη πολλών χωρών.

2.3.2 ΚΑΤΗΓΟΡΙΕΣ ΑΝΑΝΕΩΣΙΜΩΝ ΠΗΓΩΝ ΕΝΕΡΓΕΙΑΣ

Οι Ανανεώσιμες Πηγές Ενέργειας χωρίζονται σε κατηγορίες ανάλογα με την ενέργεια που χρησιμοποιείται από τη φύση. Η ηλιακή ενέργεια χρησιμοποιείται για την



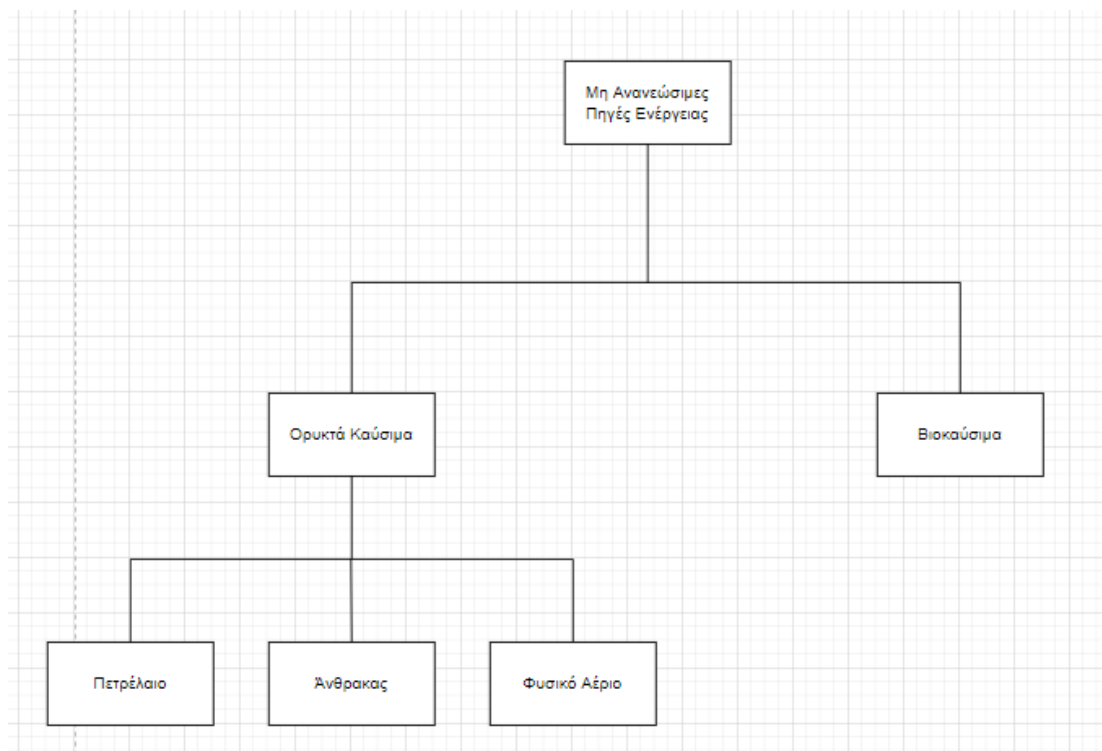
Εικόνα 2 Κατηγορίες ΑΠΕ

παραγωγή ηλεκτρικής ενέργειας και αυτό επιτυγχάνεται μέσω του φωτός του ηλίου. Η αιολική ενέργεια που προέρχεται από την ενέργεια που προκαλεί ο άνεμος και παράγει ηλεκτρική ενέργεια λοιπόν μέσω ανεμογεννητριών ή και ανεμόμυλων. Εν συνεχεία, υπάρχει η υδροηλεκτρική ενέργεια η οποία μέσω της βοήθειας του νερού δηλαδή μέσω κάποιου φράγματος ή κάποιου ειδικά σχεδιασμένου υδροηλεκτρικού σταθμού που μπορεί να παράγει την ανάλογη ενέργεια. Επίσης, κατηγορία αποτελεί και η βιομάζα η οποία μέσω της κατάλληλης επεξεργασίας παράγει ηλεκτρική ενέργεια (Αnon., n.d.). Ακόμη έχουμε τη γεωθερμική ενέργεια που οφείλεται στην αποσύνθεση των πετρωμάτων. Τέλος, υπάρχει η ενέργεια που προέρχεται από τη θάλασσα, μέσω της παλίρροιας και των κυμάτων καθώς και η ωσμωτική ενέργεια που σχετίζεται με την παραγόμενη ενέργεια από τη μίξη γλυκού και αλμυρού νερού. (Αnon., n.d.)

2.3.3 ΟΡΙΣΜΟΣ ΜΗ ΑΝΑΝΕΩΣΙΜΩΝ ΠΗΓΩΝ ΕΝΕΡΓΕΙΑΣ

Οι μη ανανεώσιμες πηγές ενέργειας εν αντιθέσει με όσα έχουν αναφερθεί μέχρι στιγμής δεν είναι ανεξάντλητες και οι πόροι τους μπορούν να εξαντληθούν εφόσον χρειάζονται αρκετά χρόνια ώστε να αναγεννηθούν. Γίνεται αντιληπτό λοιπόν ότι η επαναδημιουργία τους γίνεται με ιδιαίτερα αργούς ρυθμούς και ότι αυτός είναι ο λόγος για τον οποίο θεωρούνται και ιδιαίτερα ευάλωτες οι συγκεκριμένες πηγές ενέργειας.

2.3.4 ΚΑΤΗΓΟΡΙΕΣ ΜΗ ΑΝΑΝΕΩΣΙΜΩΝ ΠΗΓΩΝ ΕΝΕΡΓΕΙΑΣ



Εικόνα 3 Κατηγορίες Μη Ανανεώσιμων Πηγών Ενέργειας

Οι μη ανανεώσιμες πηγές ενέργειας μπορούν να χωριστούν σε δύο βασικές κατηγορίες τη συμβατική που έχει σαν βασικές υποκατηγορίες της τα ορυκτά καύσιμα όπως είναι το πετρέλαιο, ο άνθρακας και το φυσικό αέριο, και την μη συμβατική που σαν κύρια υποκατηγορία της έχει τα βιοκαύσιμα. Αυτές οι πηγές ενέργειας έχουν την τάση τα επόμενα χρόνια να εξαλειφθούν καθώς χρησιμοποιούνται σε αλόγιστο βαθμό. Για παράδειγμα τα ορυκτά καύσιμα και πιο συγκεκριμένα ο άνθρακας, το πετρέλαιο και το φυσικό αέριο στις μέρες μας έχουν ιδιαίτερα υψηλή κατανάλωση και για να αναδημιουργηθούν θα περάσει ένα σημαντικό μεγάλο χρονικό διάστημα αρκετών ετών. Από την άλλη πλευρά τα βιοκαύσιμα αποτελούν μείγμα από οργανικές ουσίες και προέρχεται είτε από τη γεωργία είτε από διάφορα είδη δασών (Αnon., n.d.). Ακόμη μία μη ανανεώσιμη πηγή ενέργειας θεωρείται η πυρηνική ενέργεια. Παρότι δεν εκπέμπει αέρια που μπορούν να επηρεάσουν το φαινόμενο του θερμοκηπίου, υπάρχει περίπτωση να προκύψει κάποιο ατύχημα. Τότε θα υπάρξουν σοβαρά μειονεκτήματα που αφορούν τόσο το περιβάλλον όσο και τον ανθρώπινο παράγοντα (Μαγίδου, 2016).

2.4 ΠΡΟΚΛΗΣΕΙΣ ΑΠΟ ΤΗΝ ΕΙΣΑΓΩΓΗ ΑΝΑΝΕΩΣΙΜΩΝ ΠΗΓΩΝ ΕΝΕΡΓΕΙΑΣ

Σαφέστατα προκύπτουν διάφορες προκλήσεις οι οποίες προέρχονται από τις ΑΠΕ όπως είναι η μαζικότητα που αφορά τις μονάδες παραγωγής. Οι μονάδες παραγωγής μπορούν είτε να αποτελούνται από πολλά φωτοβολταϊκά στοιχεία παραδείγματος χάριν, είτε από λιγότερα. Αυτό εμφανώς επηρεάζει και το μέγεθος των ενεργειακών δεδομένων. Στις μονάδες αυτές ανεξαρτήτως μεγέθους πάντα έχει σημασία το κόστος το οποίο είναι αυξημένο αρχικά διότι είναι απαραίτητη η εγκατάσταση των

ανάλογων στοιχείων. Ακόμη μία πρόκληση σχετίζεται με την τοποθέτηση καθώς αναλόγως με τη μορφή ενέργειας που πρόκειται να αξιοποιηθεί επιλέγεται και η τοποθεσία εγκατάστασης. Επομένως ανάλογα με την εγκατάσταση και την τοποθεσία παράγονται και τα ανάλογα δεδομένα. Τα ενεργειακά δεδομένα που αφορούν ανανεώσιμες πηγές ενέργειας προέρχονται τόσο από μετρήσεις όσο και από μοντελοποίηση των εκάστοτε δεδομένων. Έχουν την ιδιότητα να εμφανίζονται συνήθως ανά τακτά χρονικά διαστήματα και να αποτελούν μία χρονοσειρά ενώ συνήθως σε αυτές τις χρονοσειρές τα δεδομένα έχουν χαρακτηριστικά που αλλάζουν βάση των δεδομένων που βρίσκονται κάθε φορά προς ανάλυση. Όλα αυτά όμως θα συζητηθούν στο επόμενο κεφάλαιο που αφορά τα δεδομένα παραγωγής.

ΚΕΦΑΛΑΙΟ 3: ΔΕΔΟΜΕΝΑ ΠΑΡΑΓΩΓΗΣ ΑΠΕ

3.1 ΠΗΓΕΣ ΔΕΔΟΜΕΝΩΝ



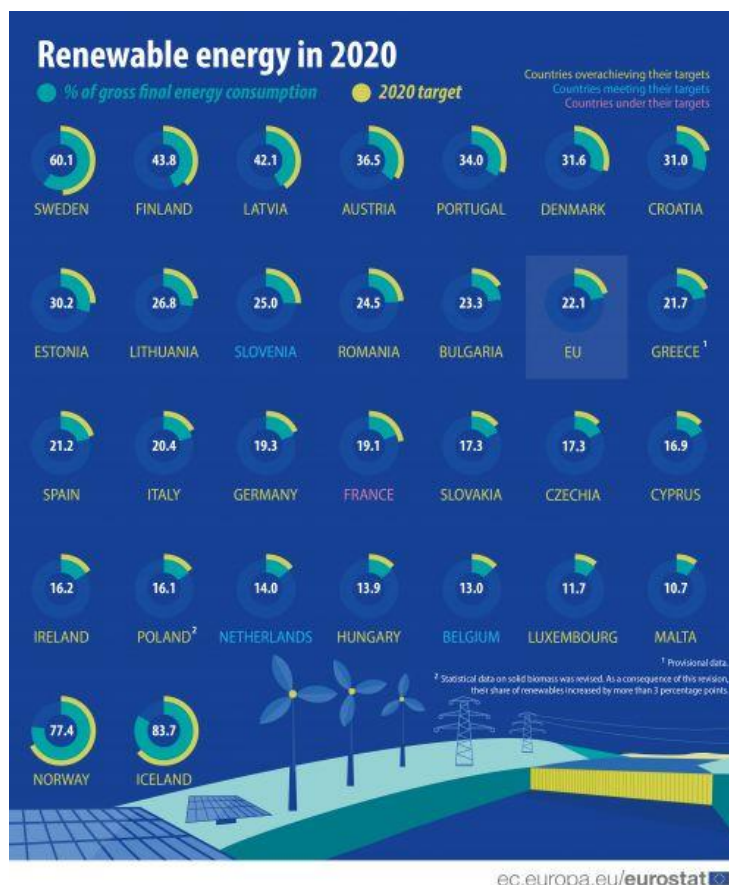
Εικόνα 4 Φωτοβολταϊκά Και Ανεμογεννήτριες

Οι πηγές από τις οποίες μπορούμε να αντλήσουμε δεδομένα για τις ανανεώσιμες πηγές ενέργειας σχετίζονται σαφέστατα και με τις μορφές ενέργειας που υπάρχουν και κατατάσσονται σε αυτή την κατηγορία. Η ηλιακή ενέργεια αποτελεί λοιπόν μία σημαντική μορφή ανανεώσιμης πηγής ενέργειας καθώς πλέον χρησιμοποιείται για την παραγωγή ηλεκτρικής ενέργειας μέσω διαφόρων τεχνολογιών. Μία από αυτές είναι τα φωτοβολταϊκά συστήματα που έχουν την ικανότητα να μετατρέπουν την ηλιακή ενέργεια την οποία δέχονται σε ηλεκτρική. Οι φωτοβολταϊκοί σταθμοί οι οποίοι σχεδιάζονται για την παραγωγή της ηλεκτρικής ενέργειας αναπτύσσονται σε διαφορετικές κλίμακες ανάλογα με την συνολική ισχύ που πρόκειται να παράγουν. Με αυτό τον τρόπο χωρίζονται λοιπόν σε μικρής, μεσαίας και μεγάλης κλίμακας (Αnon., n.d.). Αντίστοιχα με την παραγωγή και την κατανάλωση ισχύος που γίνεται στα φωτοβολταϊκά πάρκα έχουμε και την ανάλογη συλλογή από δεδομένα. Μία ακόμη τεχνολογία από την οποία μπορούμε να αντλήσουμε αντίστοιχα δεδομένα είναι οι ανεμογεννήτριες που στηρίζονται στην αιολική ενέργεια. Οι ανεμογεννήτριες έχουν και αυτές σαν χαρακτηριστικό ότι μετατρέπουν την ενέργεια του αέρα εξίσου σε ηλεκτρική. Επομένως σε ένα αιολικό πάρκο μπορούν να εντοπιστούν συστοιχίες από ανεμογεννήτριες που κυμαίνονται από κάποια KW μέχρι και MW και αντίστοιχα η άντληση των δεδομένων μπορεί να γίνει με σαφή ρυθμό (Καραϊσάς, 2014). Ακόμα άντληση δεδομένων μπορεί να υπάρξει και από άλλες ανανεώσιμες πηγές ενέργειας όπως από γεωθερμικές εγκαταστάσεις και από την υδροηλεκτρική ενέργεια.

3.2 ΜΑΖΙΚΟΤΗΤΑ ΔΕΔΟΜΕΝΩΝ

Τα δεδομένα είναι ιδιαίτερος σημαντικά καθώς από την ανάλυση και την επεξεργασία τους μπορούν να προκύψουν σημαντικά οφέλη. Με την πάροδο των χρόνων η συλλογή δεδομένων έχει αυξηθεί σημαντικά καθώς η παραγωγή τους ολοένα αυξάνεται με ταχύ ρυθμό. Εξίσου αυξάνεται και η παραγωγή που αφορά τις

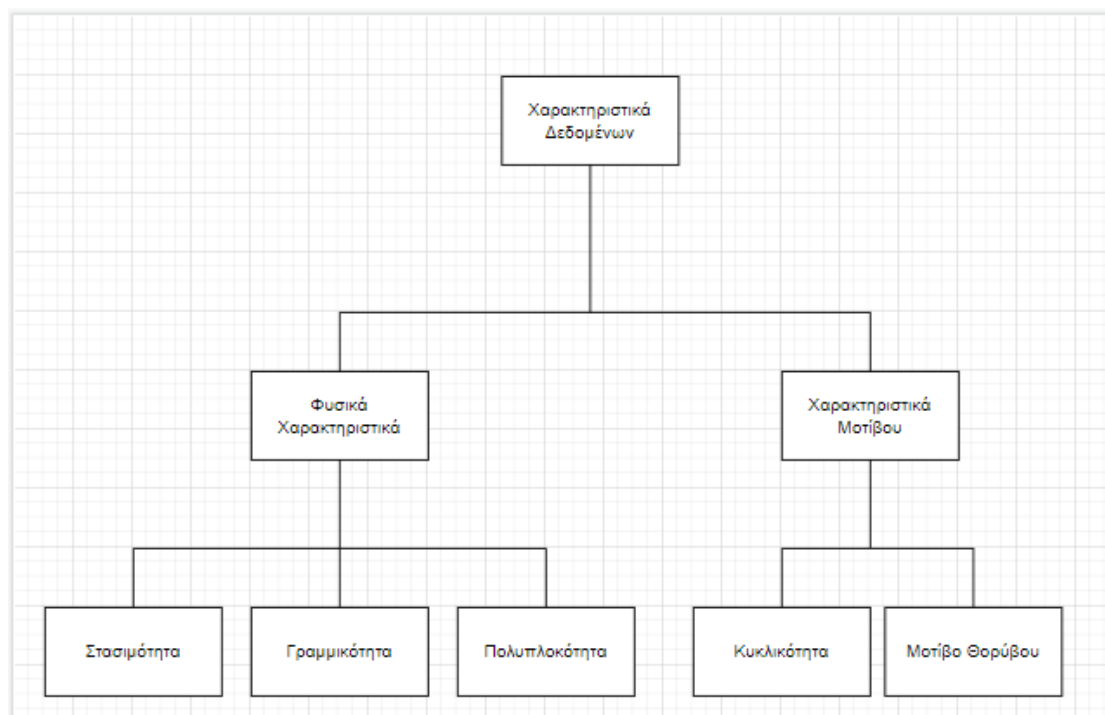
ανανεώσιμες πηγές ενέργειας (ΑΠΕ) τόσο σε παγκόσμιο επίπεδο όσο και σε ευρωπαϊκό. Αναλόγως με την ανανεώσιμη πηγή αλλά και τη χώρα που επιλέγουμε μπορούμε να δούμε ότι υπάρχουν και διαφορετικά ποσοστά που σχετίζονται με την παραγωγή δεδομένων και την κατανάλωση ενέργειας (Eurostat, 2022).



Εικόνα 5 Ενεργειακή Κατανάλωση Ευρωπαϊκών Χωρών

Φυσικά, η παραγωγή δεδομένων διαφοροποιείται ανάλογα με την τοπολογία εγκατάστασης των φωτοβολταϊκών, με αυτό εννοούμε ότι υπάρχουν διαφοροποιήσεις ανάλογα με το ότι παίρνουμε δεδομένα από μία εγκατάσταση που έχει γίνει σε ένα σπίτι ή σε ένα μικρής κλίμακας φωτοβολταϊκό σταθμό και διαφορετικό όταν μιλάμε για φωτοβολταϊκά πάρκα μεγάλης κλίμακας. Συνεπώς πρόκειται να δούμε δεδομένα που σχετίζονται από κάποια KW μέχρι και κάποια GW (Wikipedia, n.d.). Συνοψίζοντας λοιπόν η παραγωγή δεδομένων ιδίως τα τελευταία χρόνια είναι εμφανής όσον αφορά όμως τις ανανεώσιμες πηγές ενέργειας τα δεδομένα συνεχίζουν να είναι ποσοτικά σε ένα μεγάλο βαθμό όμως παρουσιάζουν διαφορές ανάλογα με το ποσοστό ισχύος που επιλέγουμε να μελετήσουμε κάθε φορά.

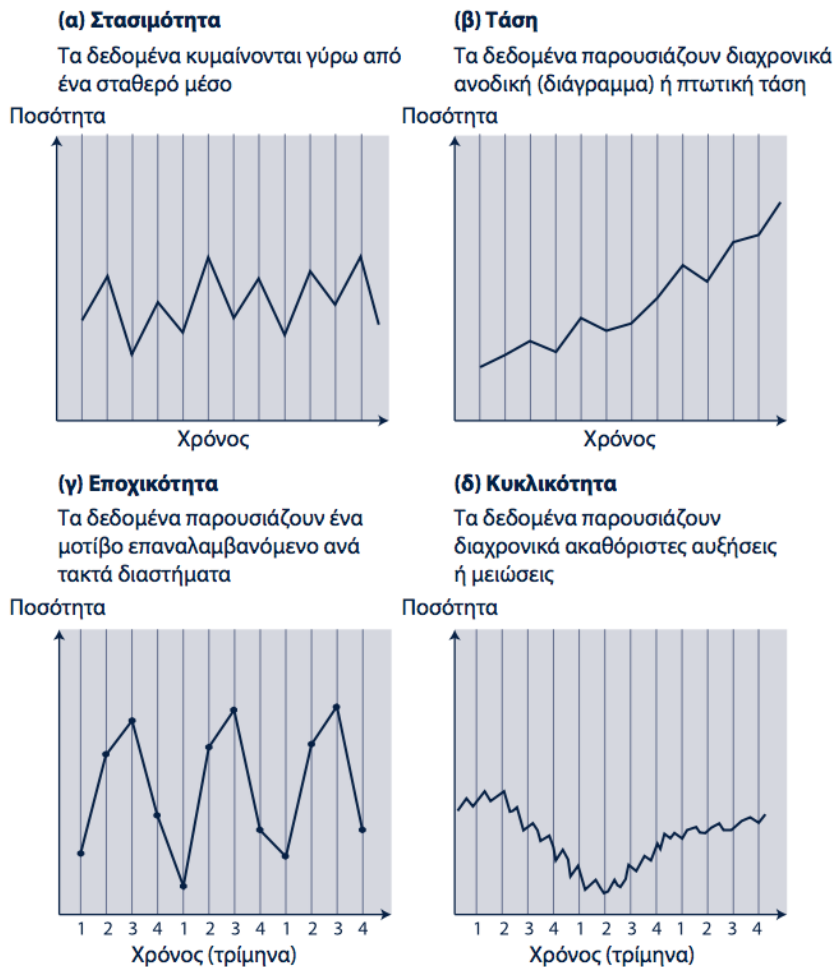
3.3 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΔΕΔΟΜΕΝΩΝ



Εικόνα 6 Χαρακτηριστικά Δεδομένων

Αρχικά μπορούμε να έχουμε δεδομένα από πολλές διαφορετικές πηγές. Μερικές εκ αυτών μπορούν να είναι ψηφιακές συσκευές αλλά και κοινωνικά δίκτυα. Γενικότερα τα δεδομένα μπορούν να έχουν πολλές μορφές όπως είναι οι δομημένοι πίνακες, οι χρονοσειρές, τα γραφήματα και πολλά ακόμα (Κατσαντώνης, 2020). Όσον αφορά όμως τα δεδομένα στο σύνολο τους αποτελούνται από χαρακτηριστικά που διακρίνονται αρχικά σε δύο κατηγορίες τα φυσικά χαρακτηριστικά που αποτελούν γενική κατηγορία για τη στασιμότητα των δεδομένων ή την μη, την γραμμικότητα ή την μη και την πολυπλοκότητα. Εν συνεχεία η επόμενη κατηγορία είναι τα χαρακτηριστικά μοτίβου σε αυτήν ανήκει η κυκλικότητα και το μοτίβο του θορύβου (Ling, 2013).

Με τον όρο της στασιμότητας στα ενεργειακά δεδομένα εννοείται ότι οι διακυμάνσεις που μπορεί να υπάρχουν στις τιμές της χρονοσειράς (όρος που θα αναφερθεί σε επόμενο κεφάλαιο αναλυτικότερα) δεν εμφανίζουν διαφορές με τον χρόνο. Όταν τα δεδομένα μπορούν να θεωρηθούν ως μη στατικά τότε πρέπει τα δεδομένα να εμφανίζουν έντονα κάποια χαρακτηριστικά όπως είναι η εποχικότητα και η τάση. Η στασιμότητα είναι ευδιάκριτη τις περισσότερες φορές στα ενεργειακά δεδομένα καθώς εντοπίζεται έντονα είτε τάση είτε εποχικότητα (Κουγιουμτζής, 2019-2020). Ο πιο απλός τρόπος είναι να εμφανίσουμε τα δεδομένα σε ένα γράφημα και να προσπαθήσουμε να τα παρατηρήσουμε για να βγάλουμε και τα ανάλογα συμπεράσματα που θα έχουν σημαντικό ρόλο στη συνέχεια για την προσαρμογή του μοντέλου. Ακόμη ένας τρόπος είναι μέσω στατιστικής μεθόδου και πιο συγκεκριμένα μέσω του πολύ γνωστού τεστ που είναι το Augmented Dickey-Fuller test.



Εικόνα 7 Διαγράμματα Ανάλυσης Εννοιών

Με τον όρο της γραμμικότητας αυτομάτως συμπεραίνουμε ότι οι μεταβλητές του συστήματος έχουν μία αλληλεπίδραση που γίνεται με γραμμικό τρόπο. Δηλαδή όλοι οι όροι του συστήματος είναι γραμμικοί ως προς τις μεταβλητές (Κουγιουμτζής, 2019-2020). Για να μπορέσει να εξετάσει κάποιος αν υπάρχει γραμμικότητα ή όχι μπορεί να το ελέγξει είτε με γραφικές μεθόδους είτε με άλλες μεθόδους που μπορούν να εντοπίσουν σχετικά την γραμμικότητα ή όχι.

Η πολυπλοκότητα που σχετίζεται με τα ενεργειακά δεδομένα μπορεί αντίστοιχα να αντιμετωπιστεί μέσω διαφόρων μεθόδων που εν μέρει μπορεί να είναι απαραίτητες ώστε να γίνει σωστή εφαρμογή του μοντέλου που ακολουθείται μετέπειτα σαν διαδικασία.

Σχετικά με την κυκλικότητα αφορά μοτίβα που επαναλαμβάνονται ανά περιόδους αλλά με αυτά τα μοτίβα ή αλλιώς διακυμάνσεις να αφορούν χρονικό διάστημα μεγαλύτερο του ενός έτους. Το πιο συνηθισμένο βέβαια είναι να τέτοιου είδους κυκλικότητας να εμφανίζονται σε οικονομικούς κύκλους. Επίσης όταν υπάρχει κυκλικότητα τότε στα δεδομένα παρουσιάζονται διάφορες αυξομειώσεις.

Το κομμάτι του θορύβου που αφορά δεδομένα και πόσο μάλλον ενεργειακά δεδομένα σχετίζεται με πιθανά σφάλματα που μπορούν να εντοπιστούν σε μία

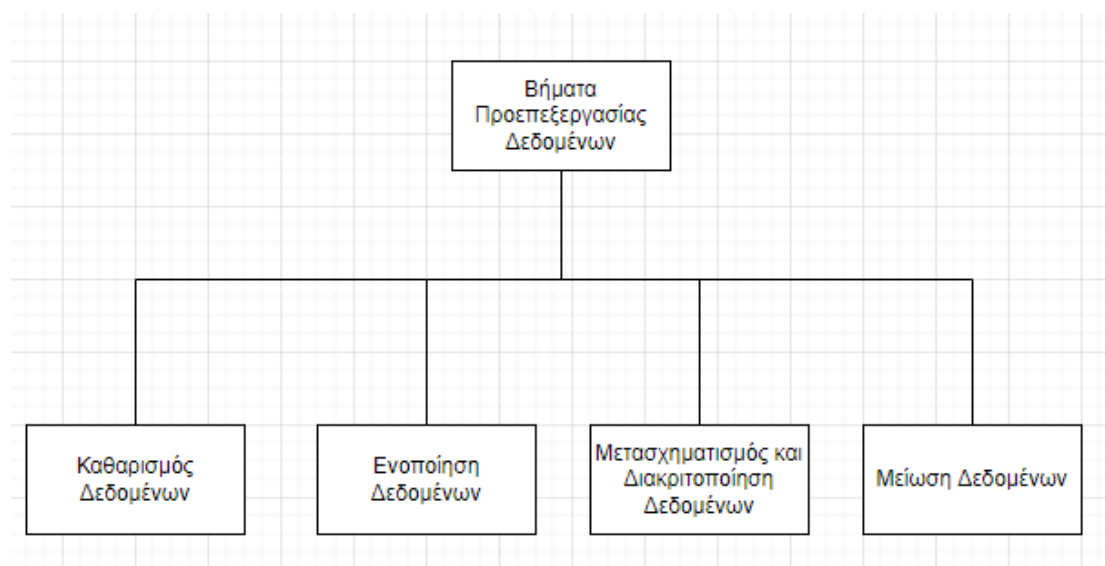
χρονοσειρά. Γι' αυτό το λόγο μπορεί να ευθύνεται είτε ο τρόπος δειγματοληψίας είτε μπορούν να ευθύνονται οι μετρήσεις. (Ling, 2013)

3.4 ΠΟΙΟΤΗΤΑ ΔΕΔΟΜΕΝΩΝ

Η ποιότητα που πρόκειται να έχουν τα δεδομένα μελλοντικά εξαρτάται από την ανάλογη προεπεξεργασία δεδομένων που θα προηγηθεί. Μέσο του όρου προεπεξεργασία δεδομένων θεωρούνται όλα τα δεδομένα που υπόκεινται σε ανάλογες διαδικασίες ώστε να υπάρχει η διαδικασία της προετοιμασίας πριν την ανάλογη παραγωγή της γνώσης (Κύρκος, 2015).

Τα δεδομένα όπως είναι φυσικό πρέπει να υποστούν αυτή τη προεπεξεργασία καθώς από τη πρώτη στιγμή που έγινε η συλλογή τους μπορεί να εμφανιστούν πολλά προβλήματα τα οποία η προεπεξεργασία δεδομένων έχει σαν κύριο στόχο την επίλυση τους. Επομένως είναι ιδιαίτερως σημαντικά να έχουμε ποιοτικά δεδομένα που δεν θα προκαλούν προβλήματα.

Πριν από την προεπεξεργασία λοιπόν αυτά τα δεδομένα λέγονται «ακάθαρτα» καθώς μπορεί να υπάρχουν κενές ή χαμένες τιμές, να υπάρχει θόρυβος μέσα στα δεδομένα, διπλοτυπίες αλλά ακόμη και ασυνέπειες που αφορούν την κωδικοποίηση. Αντίστοιχα βέβαια η επεξεργασία αυτών των δεδομένων έχει την ονομασία «καθαρισμός δεδομένων» ή αλλιώς data cleaning (Κύρκος, 2015).



Εικόνα 8 Βήματα Προεπεξεργασίας Δεδομένων

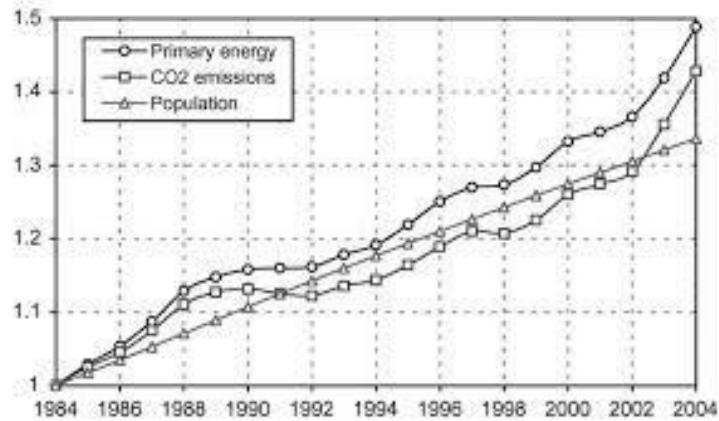
Τα βασικά βήματα λοιπόν που απαρτίζουν τον τομέα της προεπεξεργασίας δεδομένων αρχικά είναι ο καθαρισμός των δεδομένων όπου γίνεται προσπάθεια για αντικατάσταση χαμένων τιμών ή για απομάκρυνση του θορύβου. Στη συνέχεια υπάρχει η ενοποίηση δεδομένων (data integration) που σχετίζεται με την απαλοιφή πλεονασμών αλλά και με τη ενοποίηση σε βάσεις δεδομένων. Ακόμη έχουμε το μετασχηματισμό των δεδομένων και τη Διακριτοποίηση (data transformation and data discretization) που σχετίζεται με την Κανονικοποίηση. Τελευταίο η μείωση των

δεδομένων (data reduction) που σχετίζεται με τη συμπίεση που γίνεται στα δεδομένα. (Σκούρα, 2020-2021)

Εύκολα λοιπόν μπορούμε να καταλήξουμε στο συμπέρασμα ότι ο καθαρισμός των δεδομένων είναι από τις σημαντικότερες διαδικασίες που αφορούν την ανάλυση των δεδομένων.

ΚΕΦΑΛΑΙΟ 4: ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ (ΑΝΑΛΥΤΙΚΕΣ ΤΕΧΝΙΚΕΣ-REGRESSION)

4.1 ΟΡΙΣΜΟΣ ΧΡΟΝΟΣΕΙΡΑΣ



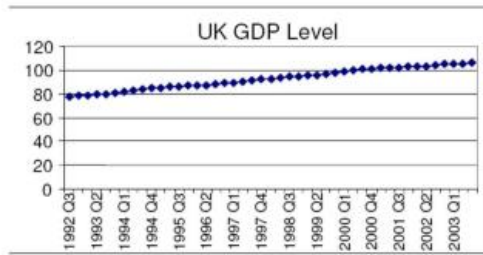
Εικόνα 9 Απεικόνιση Διαφόρων Χρονοσειρών

Με τον όρο χρονοσειρά ή χρονική σειρά (time series) εννοούμε την ακολουθία ανά τακτά χρονικά διαστήματα του μεγέθους που έχουμε επιλέξει να θέσουμε ως προς παρατήρηση με βάση το σύστημα. Τα χρονικά διαστήματα των παρατηρήσεων εξαρτώνται άμεσα από το χρόνο δειγματοληψίας ή αλλιώς sampling time και μπορούν να εμφανίζονται ανά δευτερόλεπτα, ανά δεκαπέντε λεπτά, ωριαία, ημερήσια ακόμα και μηνιαία και ούτω καθεξής. Όσον αφορά το σύστημα των χρονοσειρών είναι ιδιαίτερος σημαντικό να γίνει σωστή εκτίμηση ανάλογα με το σύστημα που μελετάμε κάθε φορά, καθώς πρέπει για αρχή να μην υπάρχει λευκός θόρυβος δηλαδή οι τιμές που λαμβάνονται να μην στηρίζονται στην τυχαιότητα και εν συνεχεία οι μεταβλητές να είναι κατά κύριο λόγο ανεξάρτητες. Τέλος είναι ιδιαίτερος σημαντικό να τονιστεί ότι εκτός από τις μονοδιάστατες χρονοσειρές υπάρχουν και οι πολυδιάστατες χρονοσειρές (multivariate time series), όπου για το ήδη υπάρχων σύστημα εμφανίζονται ταυτόχρονες παρατηρήσεις για πολλά μεγέθη. Ένα χαρακτηριστικό παράδειγμα θα μπορούσε να είναι η καταγραφή τόσο της πίεσης όσο και της θερμοκρασίας (Κουγιουμτζής, 2019-2020).

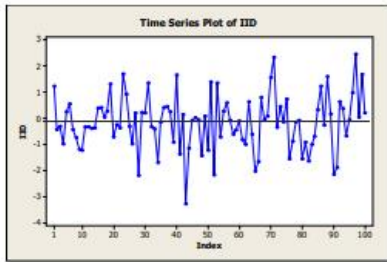
4.2 ΧΑΡΑΚΤΗΡΙΣΤΙΚΑ ΧΡΟΝΟΣΕΙΡΩΝ

Με βάση τα παραπάνω οι χρονοσειρές αποτελούνται από παρατηρήσεις οι οποίες είναι διαδοχικές με βάση κάποιο χρονικό διάστημα που μπορεί να είναι από ετήσιο μέχρι και ανά δευτερόλεπτο ανάλογα πάντα με τις συνθήκες παρατήρησης που υπάρχουν. Ακόμη οι χρονοσειρές αποτελούνται από κάποια βασικά ποιοτικά χαρακτηριστικά. Πρώτο και πάρα πολύ βασικό είναι η στασιμότητα (stationarity), η οποία ουσιαστικά βασίζεται στο χρόνο και στις τιμές της χρονοσειράς που δεν μεταβάλλονται ως προς αυτόν, επομένως και εμφανίζεται σταθερότητα ως προς τη μέση τιμή (Νικολάου, 2007), ενώ μία μη στάσιμη (non-stationary) χρονοσειρά εμφανίζει αλλαγές στις παρατηρήσεις της με τις εναλλαγές του χρόνου (Κουγιουμτζής,

2019-2020). Η στασιμότητα μπορεί αντίστοιχα να κατηγοριοποιηθεί σε δύο



(α) μη στάσιμη
χρονολογική σειρά

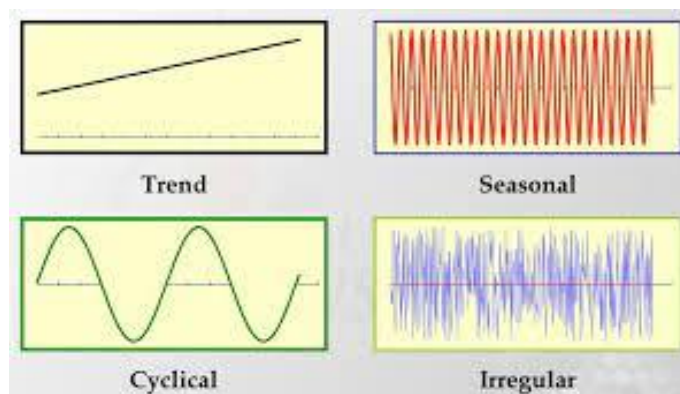


(β) στάσιμη
χρονολογική σειρά

Εικόνα 10 Γραφήματα με Στάσιμες και Μη Χρονοσειρές

κατηγορίες, η μία αφορά την ασθενή στασιμότητα ή στασιμότητα δεύτερης τάξης (weak stationarity or second order stationarity) και η δεύτερη αφορά την αυστηρή στασιμότητα (strict stationarity) (Τσιμπάνος, 2018). Η ασθενής στασιμότητα στηρίζεται στη σταθερότητα που υπάρχει τόσο στο μέσο όρο της όσο και στην πεπερασμένη διακύμανση ανάλογα με το χρονικό διάστημα που έχουμε ορίσει για την χρονοσειρά που μελετάμε ενώ εξίσου σημαντικό είναι ότι υπάρχει εξάρτηση για ξένες χρονικές στιγμές μόνο από την απόσταση που διαθέτουν μεταξύ τους δύο αυτοσυνδιακυμάνσεις (Νταβίντ, 2018). Από την άλλη πλευρά η αυστηρή στασιμότητα αποτελεί μέρος της ασθενής στασιμότητας και πρόκειται για την διαδικασία η οποία δεν επηρεάζεται από τυχαίες αλλαγές που μπορεί να συμβαίνουν στο χρόνο (Νταβίντ, 2018). Συμπερασματικά συνήθως σαν κύρια επιλογή βρίσκεται αυτή της ασθενής στασιμότητας καθώς είναι υψηλής δυσκολίας να μπορέσει να επιβεβαιωθεί η ύπαρξη αυστηρής στασιμότητας σε μία χρονοσειρά (Τσιμπάνος, 2018). Δεύτερο εξίσου σημαντικό χαρακτηριστικό που μπορεί να εμφανίσει μία χρονοσειρά είναι η τάση ή trend που βασίζεται είτε στην αύξηση είτε στη μείωση που μπορεί να παρουσιάζουν οι τιμές ανά επίπεδα (Νικολάου Ι. Ευθύμιος, 2007, σελ. 26), πιο συγκεκριμένα υπάρχει η πιθανότητα δηλαδή να παρατηρήσουμε αλλαγή στη μέση τιμή της χρονοσειράς που μελετάτε είτε με αυξητική είτε με φθίνουσα τάση, η οποία θα αποτελείται όμως από μία συμπεριφορά που θα είναι μακροχρόνια (Κουγιουμτζής, 2019-2020). Τρίτο και πολύ σημαντικό ποιοτικό χαρακτηριστικό των χρονοσειρών αποτελεί η περιοδικότητα (periodicity), η οποία ανάλογα με το έτος μπορεί να χωριστεί σε περιόδους με βάση το μήνα, το τρίμηνο και ούτω καθεξής (Κουγιουμτζής, 2019-2020). Η περιοδικότητα χωρίζεται σε δύο υποκατηγορίες, στην εποχικότητα ή seasonality και στη κυκλική διακύμανση ή cycle. Η εποχικότητα (seasonality) έχει σαν κύριο χαρακτηριστικό της ότι έχει την ιδιότητα της επανάληψης σε τακτά και ίσα χρονικά διαστήματα μέσα στην χρονοσειρά (Τσαούσης, 2018). Αναλόγως με τη διαδικασία που μελετάτε μπορεί η συχνότητα των περιοδικών

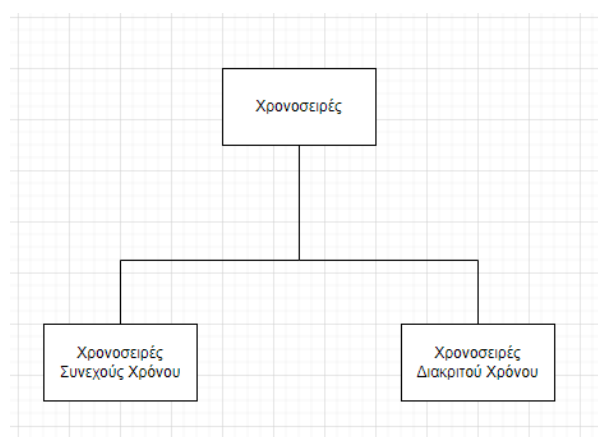
ταλαντώσεων να εμφανίζονται ανά συγκεκριμένες ώρες, ημέρες, εβδομάδες ακόμα και μήνες (Κουγιουμτζής, 2019-2020). Από την άλλη πλευρά η κυκλική διακύμανση



Εικόνα 11 Γραφήματα με Χαρακτηριστικά Χρονοσειρών

(cycle) στη μακροχρόνια σχέση και συμπεριφορά συνήθως μεγαλύτερη του ενός έτους (Χασιρτζόγλου, 2020), αποτελείται επίσης από διακυμάνσεις οι οποίες είναι κυματοειδής και παράλληλα δεν παρουσιάζει σταθερότητα ως προς τον χρόνο και τη διάρκεια του (Τσαούσης, 2018). Τέταρτο χαρακτηριστικό αποτελούν οι ασυνέχειες (discontinuity) που κατά κύριο λόγο μπορούν να προέλθουν από ασυνήθιστες τιμές (outliers) αλλά μπορεί να οφείλονται ακόμα και σε αλλαγή του επιπέδου (level shift) (Νταβίντ, 2018). Τελευταίο αλλά εξίσου σημαντικό χαρακτηριστικό αποτελεί η τυχαιότητα (randomization) και πρόκειται για διακυμάνσεις που δεν έχουν προβλεφθεί δηλαδή προκύπτουν από άγνωστους παράγοντες που δεν εμφανίζονται με συγκεκριμένο ρυθμό ούτε θεωρούνται ως κανονικοί (Νταβίντ, 2018).

4.3 ΚΑΤΗΓΟΡΙΕΣ



Εικόνα 12 Κατηγορίες Χρονοσειρών

Οι χρονοσειρές (time series) μπορούν να κατηγοριοποιηθούν ανάλογα με τα δεδομένα που διαθέτουν και το χρόνο κατά τον οποίο λήφθηκαν σε συνεχείς ή χρονοσειρές συνεχούς χρόνου (continuous time series) και σε διακριτές ή χρονοσειρές διακριτού χρόνου (discrete time series). Όσον αφορά τις χρονοσειρές

συνεχούς χρόνου είναι οι χρονοσειρές οι οποίες για οποιαδήποτε χρονική στιγμή μπορούν να έχουν μέτρηση της παρατήρησης. Ένα πολύ ιδιαίτερο παράδειγμα αποτελεί η καταγραφή των παρατηρήσεων της θερμοκρασίας που η τιμή της έχει τη δυνατότητα καταγραφής οποιαδήποτε επιθυμητή στιγμή στη διάρκεια του εικοσιτετράωρου. Σχετικά με τις διακριτές χρονοσειρές οι τιμές που καταγράφονται γίνονται σε συγκεκριμένα χρονικά διαστήματα όπως είναι η ημέρα, ο μήνας, το τρίμηνο και το έτος. Οι πιο γνωστές χρονοσειρές που ανήκουν σε αυτή τη κατηγορία προέρχονται κατά κύριο λόγο από τον οικονομικό τομέα. Όμως οι διακριτές χρονοσειρές αποτελούνται ακόμα από δύο υποκατηγορίες, τις στιγμιαίες καταγραμμένες χρονοσειρές (*instantaneously recorded time series*) και τις συσσωρευμένες χρονοσειρές (*accumulated time series*). Οι χρονοσειρές που είναι στιγμιαία καταγραμμένες αφορούν τις παρατηρήσεις που έχουν καταγραφεί σε συγκεκριμένα χρονικά διαστήματα, ενώ μπορούν να καταγραφούν οποιαδήποτε στιγμή μέσα στην ημέρα. Ενώ οι συσσωρευμένες χρονοσειρές είναι εκείνες που η καταγραφή της παρατήρησης δεν μπορεί να είναι άμεση οπότε παρουσιάζονται συνολικά οι παρατηρήσεις σε συγκεκριμένο χρονικό διάστημα (Τσιμπάνος, 2018). Πρακτικά λοιπόν μπορούμε να καταλάβουμε ότι οι χρονοσειρές είναι κατά κύριο λόγο διακριτού χρόνου. Φυσικά οι χρονοσειρές θα μπορούσαν να κατηγοριοποιηθούν και με βάση τις επόμενες τους τιμές σε ντετερμινιστικές και σε στοχαστικές. Οι ντετερμινιστικές στις οποίες για την κατοχύρωση των μελλοντικών τιμών θα ήταν απαραίτητη και η ύπαρξη των προηγούμενων τιμών και οι στοχαστικές στις οποίες οι μελλοντικές τιμές δεν μπορούν να περιγράψουν πλήρως τις παρελθοντικές τιμές (Χασιριτζόγλου, 2020).

4.4 ΑΠΛΗ ΓΡΑΜΜΙΚΗ ΠΑΛΙΝΔΡΟΜΗΣΗ

Η παλινδρόμηση (*regression*) ή ακόμα πιο συγκεκριμένα η ανάλυση παλινδρόμησης ή *regression analysis* αποτελεί σημαντικό τομέα της Στατιστικής επιστήμης και ασχολείται με τη σχέση που υπάρχει μεταξύ των μεταβλητών του κάθε προβλήματος. Οι μεταβλητές αυτές μπορούν να ξεκινούν από δύο και να είναι και περισσότερες. Όπως μπορεί να γίνει αντιληπτό η πιο απλή περίπτωση που θα μπορούσε να τεθεί υπό εξέταση είναι αυτή των δύο μεταβλητών δηλαδή της απλής γραμμικής παλινδρόμησης (*simple linear regression*). Όσον αφορά τις δύο μεταβλητές, η μία μεταβλητή θα είναι η εξαρτημένη ή *dependent variable*, η οποία θα απαρτίζεται από τιμές οι οποίες θα εμφανίζονται συνεχόμενα. Η άλλη μεταβλητή θα είναι η ανεξάρτητη μεταβλητή ή *independent variable* που δεν εξαρτάται απαραίτητα από τον τύπο δεδομένων που θα έχει. Επομένως, αντιλαμβανόμαστε ότι το μοντέλο ή εξίσωση παλινδρόμησης προκύπτει από την σχέση που έχουν μεταξύ τους οι μεταβλητές, ενώ το απλό γραμμικό μοντέλο παλινδρόμησης, παίρνει τον ορισμό της γραμμικότητας διότι φαίνεται η αντίστοιχη σχέση μεταξύ της εξαρτημένης και της ανεξάρτητης μεταβλητής. Θεωρείται ως απλό εφόσον πρόκειται για συσχέτιση μεταξύ δύο μεταβλητών και όχι περισσότερων (Αδαμόπουλος, 1999-2000).

4.5 ΑΝΑΛΥΣΗ ΜΟΝΤΕΛΩΝ ΜΕ ΤΕΧΝΙΚΕΣ ΠΑΛΙΝΔΡΟΜΗΣΗΣ

Οι τεχνικές παλινδρόμησης χρησιμοποιούνται ώστε μέσω των σχέσεων που υπάρχουν μεταξύ των μεταβλητών να προκύψουν αντίστοιχα αποτελέσματα και

προβλέψεις βάσει των δεδομένων. Ορισμένα από τα μοντέλα που πρόκειται να αναλυθούν είναι το μοντέλο αυτοπαλινδρόμησης, το μοντέλο κινητού μέσου, το μοντέλο αυτοπαλινδρόμησης κινητού μέσου, το ARIMA και το SARIMA.

4.5.1 ΜΟΝΤΕΛΟ ΑΥΤΟΠΑΛΙΝΔΡΟΜΗΣΗΣ (AutoRegression-AR)

Τα μοντέλα αυτοπαλινδρόμησης κατά βάση στηρίζονται στο γεγονός ότι οι τιμές που εμφανίζονται στις εκάστοτε χρονοσειρές ανά χρονικές στιγμές και διαστήματα αποτελούνται από ένα είδος εξάρτησης σχετικά με τις παρελθοντικές τους τιμές που υπάρχουν σε κάθε χρονοσειρά που μελετάται. Επομένως πρόκειται για μία εξάρτηση που αφορά συγκεκριμένες περιπτώσεις ανάλογα με το μοντέλο πρόβλεψης και παράλληλα αναπτύσσεται γραμμική σχέση μεταξύ των πρόσφατων παρατηρήσεων και των παρελθοντικών τιμών τους στη χρονοσειρά (Χατζηευλογημένος, 2019).

Υποθετικά μιλώντας λοιπόν, αν υπάρξει μία παρατήρηση στη χρονοσειρά κατά ένα χρονικό διάστημα t , τότε αυτή η παρατήρηση εξαρτάται από παρελθοντικές τιμές. Κατά αυτό τον τρόπο δημιουργείται το μοντέλο αυτοπαλινδρόμησης, το οποίο αλγεβρικά μπορεί να αναπαρασταθεί με τον εξής τρόπο:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t \quad (1)$$

όπου, το c είναι μία σταθερά, το e_t είναι λευκός θόρυβος που προέρχεται από κάποια κατάλοιπα που μπορεί να υπάρξουν στην χρονοσειρά και τα $\phi_1, \phi_2, \dots, \phi_p$ είναι συντελεστές αυτοσυσχέτισης του μοντέλου αυτοπαλινδρόμησης.

Με βάση τον παραπάνω τύπο της σχέσης (1) αντιλαμβανόμαστε ότι η τιμή y_t της χρονοσειράς εξαρτάται άμεσα από τις προηγούμενες παρατηρήσεις δηλαδή από την $\phi_1, \phi_2, \dots, \phi_p$ άρα από την προηγούμενη, την προ-προηγούμενη και την κατά p στιγμές πίσω παρατηρήσεις. Ενώ σε αυτό το γραμμικό υπολογισμό προαιρετικά μπορεί να προστεθεί και η σταθερά c (Δαγούμας, 2020).

Γενικά είναι γνωστό ότι οι πιο απλές περιπτώσεις αυτού του μοντέλου είναι της πρώτης και της δεύτερης τάξης, αντίστοιχα με $p = 1$ και $p = 2$. Εξίσου σημαντικό είναι να γνωρίζουμε ότι η τάξη ενός τέτοιου μοντέλου ορίζεται από τη συνάρτηση μερικής αυτοσυσχέτισης (Partial Autocorrelation Function - PACF).

Συνοπτικά, για ένα μοντέλο αυτοπαλινδρόμησης με τάξη p ισχύει ότι η συνάρτηση αυτοσυσχέτισης ή ACF τείνει κατά φθίνοντα ρυθμό προς τη μηδενική τιμή ενώ παράλληλα η μέση τιμή ορίζεται από τον τύπο: $\mu = \phi_0 / (1 - \phi_1 - \phi_2 - \dots - \phi_p)$ (Τσαούσης, 2018).

4.5.2 ΜΟΝΤΕΛΟ ΚΙΝΗΤΟΥ ΜΕΣΟΥ (MOVING AVERAGE - MA)

Το μοντέλο κινητού μέσου αναπτύσσει και αυτό γραμμικότητα σχέσεων αυτή τη φορά όμως ανάμεσα στην τιμή της χρονοσειράς εκείνη τη χρονική στιγμή και στα σφάλματα που προέρχονται από προηγούμενες παρατηρήσεις. Άρα, μία παρατήρηση μπορεί να επηρεαστεί από άλλα γεγονότα ή πληροφορίες και λόγω του μοντέλου κινητού μέσου αυτό διαδίδεται και στις επόμενες παρατηρήσεις. Αλγεβρικά μπορούμε να έχουμε την εξής αναπαράσταση:

$$y_t = c - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \dots - \theta_q e_{t-q} + e_t \quad (2)$$

όπου, το c είναι μία σταθερά, το $\theta_1, \theta_2, \theta_q$ είναι οι συντελεστές μερικής αυτοσυσχέτισης για το μοντέλο κινητού μέσου (MA) και το e_t αποτελεί λευκό θόρυβο (Χατζηευλογημένος, 2019).

Με βάση τον τύπο (2) η τιμή y_t της χρονοσειράς βασίζεται από τους παράγοντες $\theta_1, \theta_2, \theta_q$ όπου προκύπτουν από σφάλματα που έγιναν με βάση το μοντέλο σε προηγούμενη, προ-προηγούμενη και κατά q περιόδους πίσω. Ενώ σε αυτό το συνδυασμό που είναι γραμμικός επιτρέπεται και η προσθήκη μίας σταθεράς c χωρίς να είναι υποχρεωτική όμως (Δαγούμας, 2020).

Αντίστοιχα, και στο μοντέλο του κινούμενου μέσου οι πιο απλές περιπτώσεις είναι αυτές της πρώτης και της δεύτερης τάξης όπου $q=1$ ή $q=2$. Σε αντίθεση όμως με παραπάνω για την τάξη q και τον προσδιορισμό της είναι η συνάρτηση αυτοσυσχέτισης (ACF).

Συνοψίζοντας, το μοντέλο κινητού μέσου (MA) έχει τη συνάρτηση μερικής αυτοσυσχέτισης (PACF) που τείνει προς το μηδέν με εκθετικό ρυθμό, έχει τη συνάρτηση αυτοσυσχέτισης (ACF) που μηδενίζεται ύστερα από q χρονικές στιγμές και έχει και τη μηδενική μέση τιμή (Τσαούσης, 2018).

4.5.3 ΜΟΝΤΕΛΟ ΑΥΤΟΠΑΛΙΝΔΡΟΜΟΥ ΚΙΝΗΤΟΥ ΜΕΣΟΥ (AutoRegression Moving Average - ARMA)

Το μοντέλο αυτοπαλίνδρομου κινητού μέσου ή Autoregressive Moving Average αποτελεί ένα μεικτό μοντέλο το οποίο απαρτίζεται από το αυτοπαλίνδρομο μοντέλο (AR) και το μοντέλο κινητού μέσου (MA) εκ των οποίων το πρώτο είναι τάξης p ενώ το δεύτερο τάξης q (Χατζηευλογημένος, 2019). Στόχος αυτού του μοντέλου, όπως είναι αναμενόμενο είναι κατά κύριο λόγο η στασιμότητα της χρονοσειράς και σε δεύτερο χρόνο η διαδικασία της αντιστρεψιμότητας (Τσαούσης, 2018). Αλγεβρικά λοιπόν το μεικτό μοντέλο αποτυπώνεται ως εξής:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p} + e_t - \theta_1 e_{t-1} - \theta_2 e_{t-2} - \theta_q e_{t-q}$$

Διαπιστώνουμε, ότι σε αυτό το μοντέλο που είναι μεικτού τύπου μπορούμε να έχουμε διάφορους συνδυασμούς ανάλογα με τις τιμές των αντίστοιχων τάξεων p και q . Σημαντική παρατήρηση βέβαια είναι πως αν κάποιος πρόκειται για μεικτό μοντέλο αλλά για ένα απλό μοντέλο είτε αυτοπαλίνδρομο είτε κινητού μέσου.

Συνοψίζοντας, το μοντέλο ARMA παίρνει μηδενική τιμή μέσω της συνάρτησης αυτοσυσχέτισης (ACF) ύστερα από χρόνο q και παίρνει μηδενική τιμή ως προς τη συνάρτηση μερικής αυτοσυσχέτισης (PACF) ύστερα από χρόνο p (Τσαούσης, 2018).

4.5.4 ΑΥΤΟΠΑΛΙΝΔΡΟΜΙΚΟ ΜΟΝΤΕΛΟ ΚΙΝΗΤΟΥ ΜΕΣΟΥ ΟΡΟΥ (ARIMA)

Το μοντέλο ARIMA ή Autoregressive Integrated Moving Average αποτελείται από τρεις συντελεστές τους (p, d, q) όπου ο καθένας έχει και διαφορετική ιδιότητα καθώς αντιστοιχεί και σε διαφορετικό μοντέλο και αντίστοιχα και μέρος του μοντέλου ARIMA. Ο συντελεστής p υποδεικνύει την τάξη της παλινδρόμησης, ο συντελεστής d την σταθερότητα που είναι απαραίτητη για την ύπαρξη ενός μοντέλου και τέλος τον

συντελεστή q για να μπορέσει να υποδείξει την τάξη που αφορά το μοντέλο του κινητού μέσου όρου (MA) (Χατζηευλογημένος, 2019).

Όταν πρόκειται να εξετάσουμε μία χρονοσειρά είναι απαραίτητο να γνωρίζουμε ότι πρέπει να αποτελείται από στασιμότητα καθώς με αυτό τον τρόπο η χρονοσειρά μπορεί να τεθεί σε σοβαρή εξέταση και να ξεκινήσει αντίστοιχα η εφαρμογή των μοντέλων που είναι προς μελέτη, στην συγκεκριμένη περίπτωση θέτουμε υπό εξέταση την ικανότητα λειτουργίας του μοντέλου ARIMA καθώς αναφερόμαστε σε αυτό. Επομένως σε μία αρκετά απλή μορφή είναι σχεδόν απαραίτητο δύο συντελεστές ο ϕ_i και ο θ_i να πληρούν συγκεκριμένους περιορισμούς για να γίνεται εμφανής η στασιμότητα που έχει η κάθε χρονοσειρά και αν δεν διαθέτει με ανάλογες κατάλληλες μεθόδους να την κατευθύνουμε ώστε να αποκτήσει (Δαγούμας, 2020).

Για παράδειγμα ένα μοντέλο ARIMA το οποίο γνωρίζουμε ότι απαρτίζεται από σταθερότητα και επομένως ο συντελεστής d είναι μηδενικός η συνάρτηση αυτοσυσχέτισης (ACF) θα έχει τάση προς το μηδέν σε περιόδους $q-p$ ενώ η συνάρτηση μερικής συσχέτισης (PACF) θα έχει τάση πάλι προς το μηδέν αλλά σε διαφορετικές περιόδους αυτή τη φορά δηλαδή ως προς $p-q$ (Χατζηευλογημένος, 2019).

4.5.5 ΜΟΝΤΕΛΟ SARIMA (SEASONAL AUTOREGRESSION INTEGRATED MOVING AVERAGE)

Το μοντέλο SARIMA αποτελεί μία γενική συγχώνευση του μοντέλου ARIMA όμως ταυτόχρονα προσθέτει και το χαρακτηριστικό της εποχικότητας. Επομένως επειδή είναι ιδιαίτερος πιθανό μία χρονοσειρά η οποία μελετάτε να εμφανίζει είτε εντόνως είτε ηπιότερα το φαινόμενο της εποχικότητας για να μπορέσουμε να το αντιμετωπίσουμε είναι απαραίτητο να χρησιμοποιήσουμε ένα μοντέλο που διαθέτει περισσότερη εξειδίκευση και είναι καταλληλότερο και αυτό είναι το μοντέλο SARIMA. Αλγεβρικά το μοντέλο απεικονίζεται ως εξής:

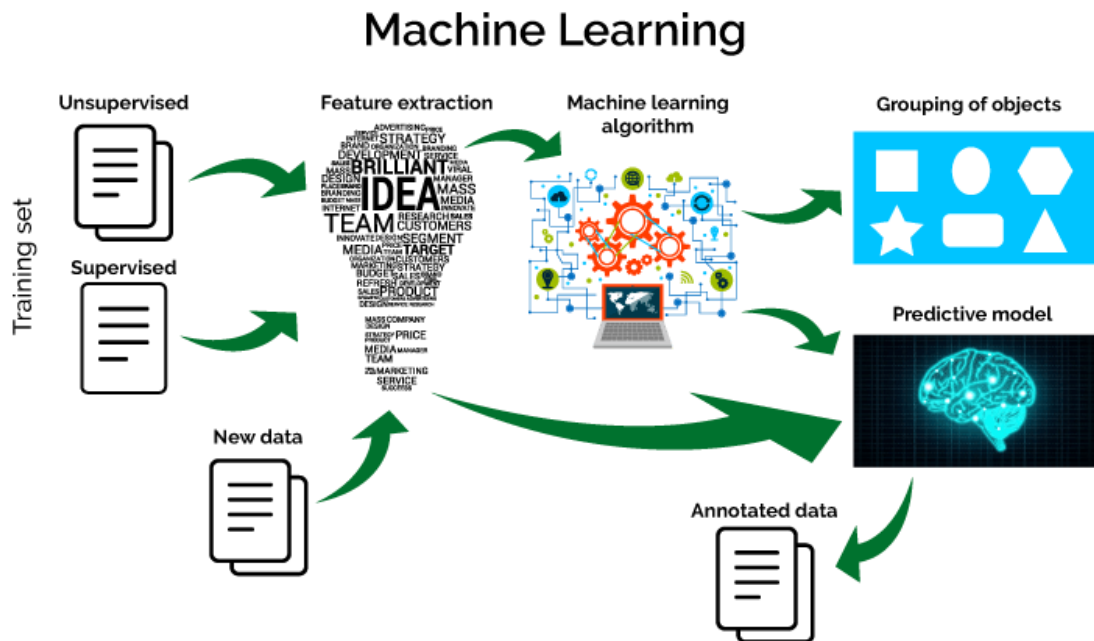
$$\varphi(B)(1 - B)^d y_t = \theta(B)Z_t$$

$$\Phi(B^s)(1 - B^s)^D y_t = \Theta(B^s)Z_t$$

Όπου το $\varphi(B)$ και το $\theta(B)$ είναι παράγοντες που σχετίζονται με την μελέτη και τον υπολογισμό της τάσης (trend) στη χρονοσειρά, ενώ οι παράγοντες $\Phi(B)$ και $\Theta(B)$ σχετίζονται με την εποχικότητα που εμφανίζεται στη χρονοσειρά και αντίστοιχα επηρεάζει και το μοντέλο. Επομένως όπως μπορεί να γίνει αντιληπτό και οι όροι d, D λειτουργούν κατά τον ίδιο τρόπο για την τάση και την εποχικότητα αντίστοιχα (Παναγοπούλου, 2020).

ΚΕΦΑΛΑΙΟ 5: ΑΝΑΛΥΣΗ ΔΕΔΟΜΕΝΩΝ(ΤΕΧΝΙΚΕΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ)

5.1 ΟΡΙΣΜΟΣ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ



Εικόνα 13 Μηχανική Μάθηση

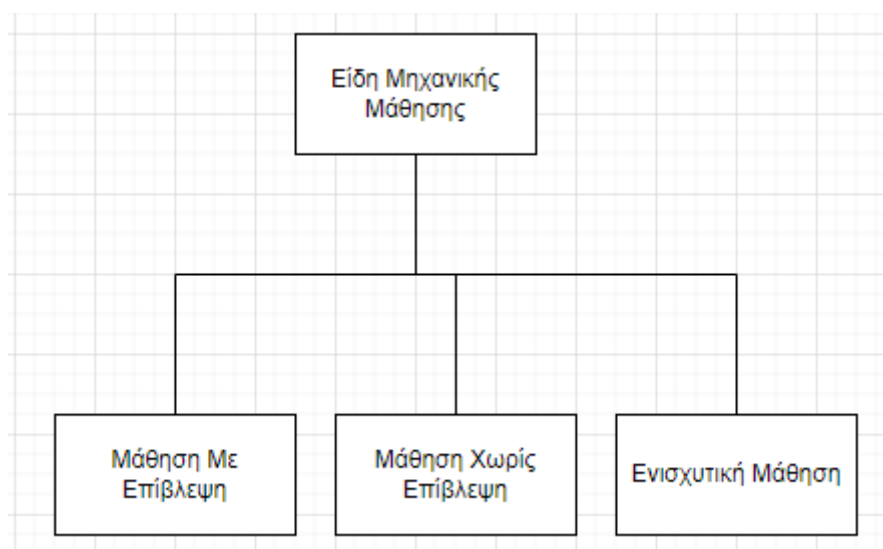
Ο άνθρωπος ως ον είναι γνωστό ότι διαθέτει ευφυή χαρακτηριστικά και ξεχωρίζει για την αντίληψη και την νοημοσύνη που έχει. Γι' αυτό το λόγο υπήρξε και ακόμα υπάρχει ενδιαφέρον ώστε όλα αυτά τα νοητικά χαρακτηριστικά να μπορέσουν με κάποιο τρόπο να αποδοθούν σε μηχανές, οι οποίες θα έχουν τη δυνατότητα της προσφοράς υπηρεσιών και όχι μόνο απλά σαν μηχανές αποθήκευσης δεδομένων και στοιχείων. Με αυτή τη προοπτική δημιουργήθηκε ο κλάδος της Τεχνητής Νοημοσύνης (Artificial Intelligence - AI). Η Τεχνητή Νοημοσύνη κατά κύριο λόγο αποτελείται από δύο προσεγγίσεις τη συμβολική και την μη συμβολική ή υπολογιστική. Στην πρώτη προσέγγιση δηλαδή αυτή της συμβολικής θεωρείται ότι μπορεί να γίνει η προσέγγιση της ανθρώπινης λειτουργίας και νοημοσύνης μέσω αλγορίθμων, ενώ στη δεύτερη προσέγγιση δηλαδή αυτή της υπολογιστικής βασίζεται στην αντιγραφή διαφόρων διεργασιών που έχουν οι άνθρωποι, ένα σαφές παράδειγμα θα μπορούσαν να αποτελέσουν τα τεχνητά νευρωνικά δίκτυα. Ένα ιδιαίτερα σημαντικό λοιπόν πεδίο της Τεχνητής Νοημοσύνης είναι η Μηχανική Μάθηση όπου έχει σαν στόχο την εκμάθηση γνώσης της μηχανής (Αναστασίου, 2019).

Γενικότερα με τον όρο Μηχανική Μάθηση (Machine Learning - ML) θεωρείται η αξιοποίηση ενός συνόλου δεδομένων για την δημιουργία είτε μοντέλων δηλαδή τη δημιουργία μίας εκδοχής για την αντίληψη του χώρου είτε προτύπων δηλαδή νέα γνώση που στηρίζεται σε παλαιότερες εμπειρίες από ένα υπολογιστικό σύστημα για την βελτίωση της απόδοσης τους. Επομένως μέσω της μηχανικής μάθησης είναι σημαντική η απόκτηση καινούριας γνώσης έχοντας σαν βοήθεια γνώση που έχει

αποκτηθεί σε παρελθοντικό χρόνο, με σκοπό τη βελτίωση του υπολογιστικού συστήματος που μελετάται. (Βλαχάβας, 2011)

Εξάλλου με τον όγκο των δεδομένων ο οποίος είναι ολοένα αυξανόμενος δίνεται λύση μέσω της τεχνητής νοημοσύνης αλλά και της μηχανικής μάθησης για τη σωστή διαχείριση τους και την εφαρμογή των κατάλληλων τεχνικών. Επιπροσθέτως ο κλάδος της μηχανικής μάθησης μπορεί να εφαρμοστεί σε πολλούς διαφορετικούς κλάδους, ακόμη και διαφορετικών πεδίων. Μερικοί από τους κλάδους που εφαρμόζεται είναι της οικονομίας, της βιοϊατρικής, της βιομηχανίας αλλά και της ρομποτικής (Αναστασίου, 2019).

5.2 ΕΙΔΗ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ



Εικόνα 14 Είδη Μηχανικής Μάθησης

Ο κλάδος της Μηχανικής Μάθησης χωρίζεται σε υποπεδία ανάλογα με το είδος που αντιπροσωπεύει το κάθε ένα. Τα τρία κυριότερα είδη μηχανικής μάθησης είναι η μάθηση με επίβλεψη, η μάθηση χωρίς επίβλεψη και η ενισχυτική μάθηση, τα οποία θα τα δούμε αναλυτικά παρακάτω.

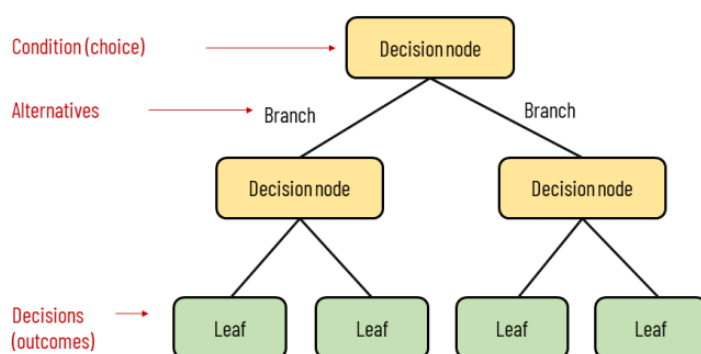
5.2.1 ΜΑΘΗΣΗ ΜΕ ΕΠΙΒΛΕΨΗ (Supervised Learning)

Η μάθηση με επίβλεψη ή supervised learning θεωρείται ως μία διαδικασία όπου δημιουργείται μία συνάρτηση που δέχεται εισόδους οι οποίες είναι γνωστές μέσω των δεδομένων ώστε η συνάρτηση να μπορέσει να γενικευτεί και να υπάρξει ώθηση σε μία έξοδο η οποία όμως είναι άγνωστη καθώς είναι και η πρόβλεψη του μοντέλου (Αναστασίου, 2019). Επομένως, οι εισοδοί που δέχεται η συνάρτηση είναι τα δεδομένα που έχουν εκπαιδευτεί δηλαδή τα training data, τα οποία εντοπίζονται με την εξής μορφή ζευγαριών x_i, y_i όπου i είναι ένας φυσικός αριθμός. Η τιμή του x_i θεωρείται ως η γνωστή τιμή εισόδου και η y_i είναι η μεταβλητή απόφασης. Επομένως, ο σκοπός είναι να γίνει σωστή πρόβλεψη της εξόδου που θεωρείται ως η μεταβλητή y με βάση δεδομένα που στηρίζονται στο μέλλον και είναι η μεταβλητή x (Κωστόπουλος, 2018).

Στην επιβλεπόμενη μάθηση διακρίνονται δύο ειδών προβλήματα, η μία κατηγορία αφορά τα προβλήματα κατηγοριοποίησης (ταξινόμησης), ενώ η άλλη κατηγορία αφορά προβλήματα παλινδρόμησης (παρεμβολή). Όπως γίνεται κατανοητό τα προβλήματα ταξινόμησης σχετίζονται με διακριτές τιμές και σχετίζονται με την αντίστοιχη δημιουργία μοντέλων, ενώ τα προβλήματα παρεμβολής έχουν συνεχείς τιμές. Ένα παράδειγμα που αφορά τα προβλήματα κατηγοριοποίησης είναι σίγουρα η απορία του αν θα έχει ήλιο ή αν θα βρέξει την επόμενη ημέρα, ενώ για το πρόβλημα της παλινδρόμησης ένα παράδειγμα θα μπορούσε να είναι η πρόβλεψη κάποιας τιμής (Αναστασιάδου, 2019).

Ακόμη είναι απαραίτητο να έχουμε ένα δείγμα ελέγχου δηλαδή ένα test set που είναι ένα σύνολο δεδομένων το οποίο όμως είναι διαφορετικό από το σύνολο δεδομένων της εκπαίδευσης. Το test set χρησιμοποιείται ώστε να μπορέσει να φανεί η ακρίβεια που κατέχει το μοντέλο και χρησιμοποιείται σε κάθε κατάσταση (Κωστόπουλος, 2018).

Elements of a decision tree



Εικόνα 15 Decision Trees

Ορισμένοι από τους πιο γνωστούς αλγόριθμους που χρησιμοποιεί η επιβλεπόμενη μάθηση είναι οι κοντινότεροι γείτονες (Nearest Neighbors), τα δέντρα απόφασης (decision trees), απλά δίκτυα Bayes, η γραμμική παρεμβολή (linear regression) και πολλοί ακόμη (Αναστασιάδου, 2019).

Ένα παράδειγμα εφαρμογή που θα μπορούσε να δοθεί για αυτό το είδος μηχανικής μάθησης είναι ότι σαν είσοδο έχουμε δεδομένα από μετρήσεις φωτοβολταϊκών στοιχείων που αφορούν ενεργειακά δεδομένα και επομένως θα μπορούσαν να είναι από μετρήσεις παραγωγής ενέργειας, ενώ σαν έξοδο να υπάρχουν εξίσου ενεργειακές μετρήσεις που αφορούν όμως μετρήσεις πρόβλεψης και τη συνέχεια της παραγωγής ενέργειας απλά με τρόπο ελεγχόμενο ώστε να μην μπορέσουν να έχουν ξαφνική απόκλιση οι μετρήσεις.

Συνοψίζοντας, στην μάθηση με επίβλεψη ορίζονται ορισμένες σημαντικές διαδικασίες. Πρώτη από όλες είναι ο σημαντικότερος καθορισμός του προβλήματος ώστε να γίνει αντιληπτό το πρόβλημα το οποίο τίθεται προς ανάλυση. Εν συνέχεια,

είναι απαραίτητο να γνωρίζουμε τα δεδομένα που έχουμε στο πρόβλημα και να ξεκινήσουμε τη διαδικασία της προ-επεξεργασίας που είναι ιδιαίτερα σημαντική για τα δεδομένα και έχει αναλυθεί σε παραπάνω κεφάλαιο. Ακόμη ένα σημαντικό βήμα είναι η δημιουργία του συνόλου εκπαίδευσης ή αλλιώς training set, ώστε να μπορέσουμε να διαλέξουμε τον σωστό ταξινομητή (classifier) και να δημιουργηθεί το κομμάτι της εκπαίδευσης πάνω σε αυτόν για να μπορέσει να γίνει σωστή επεξεργασία από το μοντέλο και να επιφέρει τα ανάλογα αποτελέσματα. Τέλος, όσον αφορά το σύνολο ελέγχου (test set) να γίνει σωστά η αξιολόγηση που αφορά τα αποτελέσματα (Κωστόπουλος, 2018) .

5.2.2 ΜΑΘΗΣΗ ΧΩΡΙΣ ΕΠΙΒΛΕΨΗ (Unsupervised Learning)

Η μάθηση χωρίς επίβλεψη ή unsupervised learning αποτελεί ένα σύστημα εκπαίδευσης όπου δεν του είναι γνωστή η έξοδος και στηρίζεται μονάχα στις εισόδους που δέχεται ως παρατηρήσεις. Αυτό που καταλαβαίνουμε λοιπόν είναι ότι τα δεδομένα που διατίθενται κάθε φορά, τα οποία αποτελούν και τα στοιχεία εισόδου δεν είναι πάντοτε γνωστά αλλά γίνεται προσπάθεια διαχείρισης τους με βάση ομοιότητες ή μη που μπορεί να εμφανίζονται.

example of K nearest neighbor.

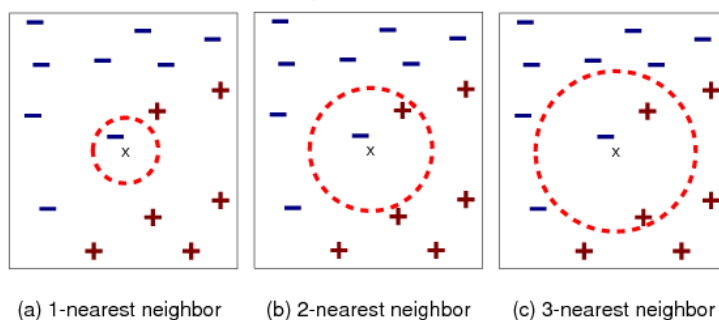


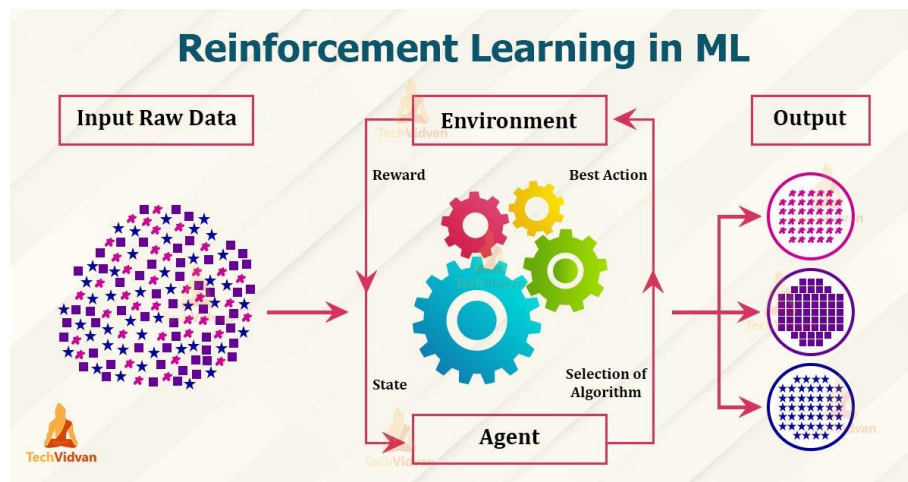
Figure 1: K nearest neighbor

Εικόνα 16 K Nearest Neighbors

Ακόμη σοβαρά προβλήματα που τίθενται προς επίλυση μπορούν να εμφανιστούν και στην μη επιβλεπόμενη μάθηση. Ένα από αυτά είναι η συσταδοποίηση (clustering), όπου ο κύριος σκοπός είναι τα δεδομένα που υπάρχουν κάθε φορά να μπορούν να χωριστούν σε συστάδες δηλαδή κατά κάποιο τρόπο σε ομάδες. Μέσω της διαδικασίας της ομαδοποίησης των δεδομένων είναι ιδιαίτερα σημαντικό να επιτευχθεί ο κύριος στόχος, που είναι η ύπαρξη σημαντικού ποσοστού ομοιότητας μέσα στη συστάδα. Ακόμα αξίζει να αναφέρουμε ότι από τους πιο γνωστούς αλγόριθμους που ανήκουν στην κατηγορία της συσταδοποίησης είναι ο αλγόριθμος k-means, που είναι επαναληπτικός και στηρίζεται στο διαχωρισμό των δεδομένων (Αναστασιάδου, 2019). Ακόμη ένα πρόβλημα της μη-επιβλεπόμενης μάθησης είναι η ελάττωση των διαστάσεων (dimensionality reduction) και ο εντοπισμός καινοτομιών (novelty detection). Στην ελάττωση των διαστάσεων τα δεδομένα ενώ διατηρούν τα χαρακτηριστικά και τις ιδιότητες τους πρέπει να μπορέσουν να μειώσουν ως προς το πλήθος τα χαρακτηριστικά, ενώ στον εντοπισμό των καινοτομιών γίνεται αναγνώριση περιπτώσεων που δεν ανήκουν στην πλειοψηφία των υπολοίπων (Κωστόπουλος, 2018).

Ένα πολύ σημαντικό παράδειγμα που θα μπορούσε να κάνει πιο κατανοητό το πως λειτουργεί γενικά η μάθηση χωρίς επίβλεψη είναι ο διαχωρισμός των πελατών ανά ομάδες με βάση κοινά χαρακτηριστικά που μπορεί να εμφανίζουν. Ξεκινώντας λοιπόν θεωρούμε ως δεδομένο ότι διαθέτουμε ένα αρχείο με δεδομένα το οποίο έχει πελάτες και διάφορα χαρακτηριστικά που τους αφορούν. Επομένως, γίνεται η προσπάθεια ένταξης τους σε ομάδες ανάλογα με τον αλγόριθμο που έχει επιλεγεί. Γίνεται λοιπόν αντιληπτό πως επειδή ο αρχικός αριθμός των κατηγοριών των χαρακτηριστικών μας είναι άγνωστος, το ίδιο άγνωστα θα μας είναι και τα αποτελέσματα κατά την έξοδο (Αναστασιάδου, 2019).

5.2.3 ΕΝΙΣΧΥΤΙΚΗ ΜΑΘΗΣΗ (Reinforcement Learning)



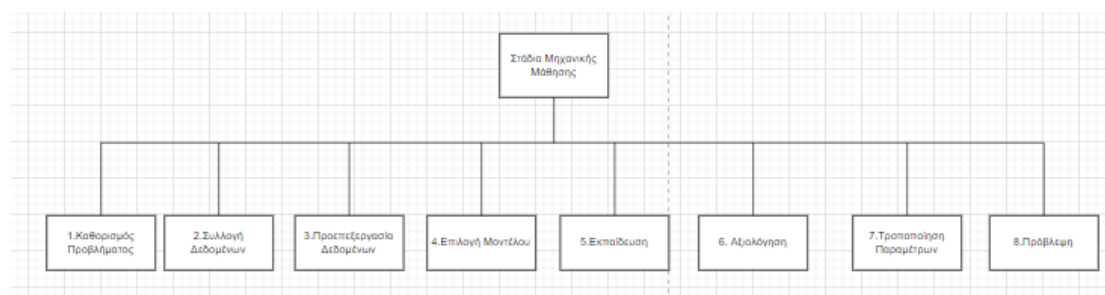
Εικόνα 17 Ενισχυτική Μάθηση

Η ενισχυτική μάθηση ή reinforcement learning χαρακτηρίζεται ως η διαδικασία όπου το σύστημα μέσω διαφόρων ενεργειών προσπαθεί να αλληλοεπιδράσει με βάση το εκάστοτε περιβάλλον. Δηλαδή προσπαθεί να καταλάβει τις καταστάσεις που υπάρχουν στο περιβάλλον που εξετάζετε και να τις μετατρέψει σε ενέργειες που πρέπει απαραίτητως να γίνουν όλο αυτή η διαδικασία μέσω της αλληλεπίδρασης. Επομένως, η μάθηση με ενίσχυση προσπαθεί να αντλήσει το κομμάτι της μάθησης μέσω της αλληλεπίδρασης με το περιβάλλον και συνάμα στηρίζεται στη μέθοδο της δοκιμής και του σφάλματος αλλά και στη μάθηση τόσο μέσω της επιβράβευσης όσο και της τιμωρίας, ενώ η επιβλεπόμενη και η μη επιβλεπόμενη προσπαθούν να αποκτήσουν γνώση μέσα από τα ίδια τα δεδομένα.

Η ενισχυτική μάθηση ουσιαστικά αποτελείται από έναν πράκτορα (agent) ο οποίος αναφέρετε στο σύστημα και προσπαθεί να κάνει την καλύτερη δυνατή επιλογή με βάση το περιβάλλον και την κατάσταση στην οποία βρίσκεται. Επομένως βρίσκει τις ανάλογες ενέργειες και προσπαθεί κάθε φορά να κάνει τη σωστότερη επιλογή αξιοποιώντας παράλληλα το κομμάτι της ανταμοιβής που σχετίζεται με τη βελτίωση της συμπεριφοράς του ίδιου του πράκτορα. Κατά αυτό τον τρόπο αφού μακροπρόθεσμα πρόκειται να βελτιωθεί η απόδοση γίνεται αντιληπτό ότι

ενδιαφέρεται τόσο για την ανταμοιβή (reward) όσο και για τις συνέπειες που θα αποφέρουν οι ενέργειες μελλοντικά (Αναστασιάδου, 2019).

5.3 ΣΤΑΔΙΑ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ



Εικόνα 18 Στάδια Μηχανικής Μάθησης

Τα στάδια της μηχανικής μάθησης είναι ιδιαίτερος σημαντικά καθώς μπορούν εύκολα να αναλύσουν ποια βήματα είναι καλό να ακολουθηθούν ώστε να γίνει σωστή διαχείριση των ανάλογων δεδομένων μέσω της μηχανικής μάθησης. Αρχικά, καθορίζεται το πρόβλημα, το οποίο είναι και από τις σημαντικότερες διαδικασίες μαζί με την σωστή συλλογή δεδομένων, διότι είναι απαραίτητο στην αρχή ανάθεσης του προβλήματος να θέτονται καίρια και βάσιμα ερωτήματα καθώς και να θέτονται εξ αρχής οι στόχοι. Στην συνέχεια, έρχεται στο προσκήνιο η συλλογή των δεδομένων που είναι εξίσου σημαντική, καθώς έχουν σημασία για την τελική διαμόρφωση των αποτελεσμάτων. Είναι ιδιαίτερα σημαντικό να υπάρχουν δεδομένα αρκετών χρόνων, τα οποία όμως να είναι και ποιοτικά, δηλαδή να έχουν συνεχόμενες τιμές ανά τακτά χρονικά διαστήματα, να μην εμφανίζουν έντονα κενά και ακραίες τιμές αφού θα επηρεαστεί το μοντέλο μελλοντικά. Το επόμενο στάδιο αφορά την προεπεξεργασία των δεδομένων που έχει αναλυθεί και σε προηγούμενο κεφάλαιο και σαν διαδικασία θεωρείται απαραίτητη αφού διορθώνονται τυχόν λάθη και διαχειρίζονται οι τιμές που είναι κενές. Η προεπεξεργασία των δεδομένων θεωρείται ως μία αρκετά σημαντική διαδικασία καθώς πρόκειται αυτά τα δεδομένα να χρησιμοποιηθούν στο μοντέλο είτε για την εκπαίδευση του (training set) είτε για τον ορθό έλεγχο του (test set). Εν συνεχεία γίνεται η επιλογή του μοντέλου με τρόπο προσεκτικό ανάλογα με τα χαρακτηριστικά που διαθέτουν τα δεδομένα και το τελικό επιθυμητό αποτέλεσμα. Με βάση αυτό τον τρόπο υπάρχουν μοντέλα που σχετίζονται με μουσική, εικόνες, κείμενο αλλά υπάρχουν και μοντέλα που ασχολούνται με αριθμητικά δεδομένα. Γι' αυτό το λόγο αφού γίνει η ανάλογη επιλογή του μοντέλου ξεκινάει και η εκπαίδευση του. Πρόκειται για ένα κομμάτι ιδιαίτερα δύσκολο, αφού στόχος είναι η σωστή πρόβλεψη, η οποία θα γίνει με βάση τη σωστή ρύθμιση των αντίστοιχών παραμέτρων του μοντέλου. Αφού λοιπόν ολοκληρωθεί η εκπαίδευση, έρχεται η σειρά της αξιολόγησης που τεστάρει αν το μοντέλο είναι ορθό με βάση τις απαιτήσεις. Στην αξιολόγηση χρησιμοποιούνται τα δεδομένα του ελέγχου τα οποία δεν έχουν χρησιμοποιηθεί όμως για εκπαίδευση επομένως η αξιολόγηση που πρόκειται να γίνει είναι διαυγής. Ακόμη ένα στάδιο είναι η τροποποίηση των παραμέτρων του ίδιου του μοντέλου. Αυτό μπορεί να γίνει εάν θέλουμε να δούμε την εξελικτικότητα που μπορεί να έχει το κατ' εκπαιδευόμενο μοντέλο. Τελευταίο αλλά εξίσου σημαντικό στάδιο

είναι αυτό της πρόβλεψης, όπου αποτελεί το στάδιο που φαίνονται έντονα η δουλειά των προηγούμενων σταδίων και ειδικότερα στα στάδια της εκπαίδευσης και της αξιολόγησης του μοντέλου (Yufeng, 2017).

5.4 MONTEΛΑ

Στη μηχανική μάθηση, η μάθηση στηρίζεται σε μια συνάρτηση f όπου οι μεταβλητές εισόδου x αντιστοιχίζονται στις μεταβλητές εξόδου y , δηλαδή $y = f(x)$. Με αυτό τον τρόπο μπορούμε να πούμε ότι τα μοντέλα που στηρίζονται στη μηχανική μάθηση χωρίζονται κατά κύριο λόγο σε δύο κατηγορίες στα παραμετρικά και στα μη-παραμετρικά μοντέλα (Χατζηλυγερούδης, 2020).

Ένα παραμετρικό μοντέλο στηρίζεται στο γεγονός ότι αποτελείται από σταθερού μεγέθους παραμέτρους χωρίς να λαμβάνετε υπόψη κάθε φορά ο αριθμός από τα δεδομένα εκπαίδευσης. Επομένως, όσα δεδομένα και αν παρουσιαστούν σε ένα παραμετρικό μοντέλο αυτό θα συνεχίζει να χρειάζεται τον ίδιο αριθμό παραμέτρων όπως και στην αρχή. Άρα μπορεί να γίνει αντιληπτό ότι αν οι παράμετροι είναι γνωστοί τότε τα δεδομένα μπορούν μέχρι και να διαγραφούν, ενώ παράλληλα οι προβλέψεις δεν εξαρτώνται από τα δεδομένα.

Στην κατηγορία των παραμετρικών μοντέλων μπορούμε να θέσουμε τα γραμμικά μοντέλα, όπως τη γραμμική παλινδρόμηση, τη λογιστική παλινδρόμηση και πολλά ακόμη. Ακόμη είναι τα απλά νευρωνικά δίκτυα (simple neural networks) και οι γραμμικές μηχανές διανυσμάτων υποστήριξης (linear Support Vector Machines - SVM).

Συνοψίζοντας, τα πλεονεκτήματα των παραμετρικών μοντέλων είναι η απλότητα τους καθώς είναι ιδιαίτερα εύκολα στον τομέα της κατανόησης, η μικρότερη απαίτηση τους ως προς τα δεδομένα εκπαίδευσης και η ταχύτητα των αλγόριθμων τους ως προς την εκπαίδευση. Φυσικά υπάρχουν και τα μειονεκτήματα που σχετίζονται με την πολυπλοκότητα αφού εφαρμόζονται ευκολότερα σε προβλήματα που εμφανίζουν χαμηλή πολυπλοκότητα. Ακόμη ένα μειονέκτημα δεν υπάρχει και η καλύτερη δυνατή προσαρμογή όσον αφορά τα δεδομένα.

Τα μη-παραμετρικά μοντέλα από την άλλη πλευρά δεν έχουν την δυνατότητα να δημιουργήσουν υποθέσεις με βάση τα δεδομένα εκπαίδευσης (training set) που διαθέτουν. Ωστόσο ο αριθμός των παραμέτρων μπορεί να είναι αυξητικός κάθε φορά, ενώ το μοντέλο μπορεί να δώσει πολύ περισσότερες πληροφορίες όσο αυτό τροφοδοτείται συνεχώς με δεδομένα.

Στην κατηγορία των μη-παραμετρικών μοντέλων σίγουρα είναι ο αλγόριθμος των k -πλησιέστερων γειτόνων (k - Nearest Neighbors), τα μπαεσιανά μοντέλα (Bayesian Models), και τις γκαουσιανές διαδικασίες (Gaussian Processes).

Συνοψίζοντας, για τα μη-παραμετρικά μοντέλα τα πλεονεκτήματα αφορούν τις πιο ακριβείς προβλέψεις λόγω της καλύτερης προσαρμογής που διαθέτουν, αφορούν την ευκαμψία που έχουν αφού μπορούν να προσαρμοστούν σε πολλές διαφορετικές μορφές της αρχικής συνάρτησης και τελευταίο πλεονέκτημα είναι πως αυτά τα

μοντέλα μαθαίνουν κυρίως μέσω των δεδομένων. Βέβαια αυτό μπορεί να μετατραπεί και σαν βασικό μειονέκτημα όσον αφορά την ταχύτητα που επηρεάζεται λόγω των παραμέτρων που είναι άπειρες σε αυτή την περίπτωση και εξίσου τα δεδομένα εκπαίδευσης έχουν αυξηθεί. Ακόμα ένα πολύ σημαντικό μειονέκτημα βέβαια είναι η υπερπροσαρμογή (overfitting), η οποία προέρχεται από την πλήρη αποτύπωση που υπάρχει στις ιδιότητες των δεδομένων εκπαίδευσης του μοντέλου (Shruthi, 2020).

ΚΕΦΑΛΑΙΟ 6: ΒΕΛΤΙΩΣΗ ΠΟΙΟΤΗΤΑΣ ΔΕΔΟΜΕΝΩΝ ΑΠΕ

6.1 ΒΕΛΤΙΩΣΗ ΠΟΙΟΤΗΤΑΣ ΔΕΔΟΜΕΝΩΝ

Τα δεδομένα και η ποιότητα τους όπως αναφέραμε και σε προηγούμενο κεφάλαιο είναι ιδιαίτερα σημαντική καθώς με βάση αυτά θα γίνει η ανάλυση και η ένταξη τους σε κατηγορίες για την επίλυση του εκάστοτε προβλήματος που μελετάται. Με τον όρο της ποιότητας δεδομένων εννοούμε ότι δίνεται σαν μέτρο για την αξιολόγηση και τα χαρακτηριστικά τους. Ορισμένα από τα χαρακτηριστικά που αφορούν την ποιότητα των δεδομένων αφορούν τόσο την ακρίβεια (accuracy) και την πληρότητα όσο και την συνέπεια και την αξιοπιστία (Vaughan, 2019). Η ακρίβεια των δεδομένων σχετίζεται με δεδομένα που έχουν σωστή μορφή και σωστή εγγραφή και επομένως μπορούν να χρησιμοποιηθούν με μεγάλη αξιοπιστία (Farah, 2020). Επομένως με τον όρο ακρίβεια εννοούμε τη σωστή καταχώρηση ή εγγραφή των δεδομένων. Σχετικά με την πληρότητα (completeness) θέλουμε ανάμεσα στα δεδομένα να μην υπάρχουν ούτε κενές τιμές ούτε πληροφορίες που θα έπρεπε να έχουν καταγραφεί και δεν έχει γίνει (Anon., 2021). Εν συνεχεία με τον όρο της συνέπειας των δεδομένων (consistency of data) εννοούμε ότι όλα τα δεδομένα μπορούν να διαχειριστούν με τον ίδιο τρόπο. Ένα ιδιαίτερο σημαντικό και απλό παράδειγμα σχετίζεται με καιρικά δεδομένα όπου είναι απαραίτητο να μην υπάρχουν ασυνέπειες που αφορούν τις εγγραφές στην ημέρα και την ώρα, καθώς μπορεί με αυτό τον τρόπο ένα πολύ χρήσιμο αρχείο με δεδομένα να γίνει απολύτως ανούσιο και αχρείαστο (Abhishek, 2020). Τελευταίο και ιδιαίτερος σημαντικό χαρακτηριστικό της ποιότητας δεδομένων είναι η αξιοπιστία τους. Η αξιοπιστία των δεδομένων στηρίζεται στην πληρότητα και την ακρίβεια που διαθέτουν και είναι ιδιαίτερο σημαντικό χαρακτηριστικό ιδίως για εταιρίες που μπορεί να θέλουν να πάρουν και να επεξεργαστούν δεδομένα (Anon., n.d.).

Για την βελτίωση λοιπόν της ποιότητας των δεδομένων και σαν συνέπεια και των δεδομένων που μπορεί να έχουμε κάθε φορά στα χέρια μας είναι απαραίτητο να εντοπίσουμε κατά κύριο λόγο τις χαμένες τιμές και την ύπαρξη θορύβου.

Όταν τα δεδομένα και επομένως και τα σήματα τα οποία δημιουργούνται στις χρονοσειρές αποτελούνται από θόρυβο για να είναι δυνατή η βελτίωση της ποιότητας των δεδομένων χρησιμοποιείται μία διαδικασία που αποκαλείται αποθορυβοποίηση και έχει σαν στόχο της να μην υπάρξει θόρυβος στα δεδομένα, επομένως να τον απομακρύνει. Ως θόρυβος μπορεί να θεωρηθεί οποιοσδήποτε ανεπιθύμητος ήχος μπορεί να παρεμβάλλει σε μία ηχογράφιση για παράδειγμα. Αντίστοιχα σε δεδομένα που μπορεί να αφορούν χαρακτηριστικά που προέρχονται από ενεργειακές μετρήσεις, θόρυβος μπορεί να προκληθεί όταν στο σήμα παρεμβάλλονται μετρήσεις που είναι ακραίες και δεν έχουν σχέση με τις μέχρι εκείνης της στιγμής μετρήσεις αλλά ακόμα μπορεί να προκληθεί από διάφορους αστάθμητους παράγοντες που επηρεάζουν τη μέτρηση εκείνη τη δεδομένη χρονική στιγμή. Εξάλλου ο θόρυβος σαν έννοια μπορεί να χαρακτηριστεί και ως αμφίβολη, διότι πολλές φορές εξαρτάται από τον ανθρώπινο παράγοντα τι ορίζεται σαν θόρυβος και τι σαν σήμα που πρέπει να συμπεριληφθεί στα δεδομένα. Για την

αποθρομβοποίηση όπως είναι αναμενόμενο χρησιμοποιούνται ποικίλες τεχνικές όπου μία από αυτές είναι η αποθρομβοποίηση με φασματική πύλη(spectral gating). Ακόμη αποθρομβοποίηση είναι δυνατό να συμβεί και μέσω νευρωνικών δικτύων όπου γίνεται μέσω των αυτοκωδικοποιητών, οι ίδιοι θέλουν να αντιγράψουν την είσοδο του σήματος ή της τιμής στην έξοδο και αυτό το καταφέρνουν μέσω του κωδικοποιητή και του αποκωδικοποιητή όπου ο τελευταίος μπορεί και δημιουργεί ξανά το κρυφό πεδίο που σχετίζεται με την είσοδο και την κωδικοποίηση της (Κοντογιάννης, 2022).

Ακόμη ένας πολύ σημαντικός και ιδιαίτερος τρόπος έτσι ώστε να μπορέσει να υπάρξει βελτίωση των δεδομένων και πιο συγκεκριμένα βελτίωση ενεργειακών δεδομένων είναι η έννοια των χαμένων τιμών που είναι όρος ιδιαίτερος όσον αφορά τις χρονοσειρές και τα ενεργειακά δεδομένα. Με τον όρο των χαμένων τιμών γίνεται αντιληπτό ότι πρόκειται για τιμές που δεν έχουν καταχωρηθεί στα δεδομένα και είναι άγνωστες οι τιμές που μπορεί να υπήρχαν στην θέση τους. Η αναλυτική έννοια των χαμένων τιμών θα αναφερθεί σε επόμενο επιμέρους κεφάλαιο.

6.2 ΑΝΑΛΥΣΗ ΕΝΕΡΓΕΙΑΚΩΝ ΔΕΔΟΜΕΝΩΝ

Στην ανάλυση των ενεργειακών δεδομένων θα δοθεί μία αναφορά σχετικά με τις κατηγορίες που χρησιμοποιήθηκαν και με τις στήλες που επιλέχθηκαν από το σύνολο των δεδομένων. Εξάλλου είναι γνωστό ότι η κάθε κατηγορία έχει και διαφορετική σημασία για τα δεδομένα και την ανάλυση τους. Βεβαίως σε αυτά θα αναφερθούν και οι καταστάσεις που μπορεί να συνέβαλλαν για τις χαμένες τιμές που υπάρχουν μέσα στο dataset.

6.2.1 ΕΝΕΡΓΗ ΙΣΧΥΣ

Η ισχύς όπως είναι γνωστό αποτελεί το φυσικό μέγεθος που συνδέει την παραγόμενη ενέργεια με το χρονικό διάστημα (Αντωνίου, n.d.). Η ισχύς μπορεί να διαχωριστεί κατά κύριο λόγο σε τρεις κατηγορίες, την στιγμιαία ισχύ, την μέση ισχύ και την μέγιστη ισχύ. Στην στιγμιαία ισχύ υπολογίζεται κατά κύριο λόγο μέσω του γινομένου της τάσης επί το ρεύμα και η στιγμιαία ισχύς έχει σαν χαρακτηριστικό ότι πρόκειται για τη ισχύ που εμφανίζεται όμως σε μία δεδομένη χρονική στιγμή. Η μέση ισχύς από την άλλη μεριά αποτελεί την ισχύ που όμως προσμετράτε για μία αρκετά μεγάλη χρονική περίοδο και μπορούμε να βγάλουμε τη μέση ισχύ όταν μετρήσουμε με βάση τον τύπο μέσα σε ένα χρονικό διάστημα πολύ συγκεκριμένο (Αnon., n.d.).

Η ενεργή ισχύς (active power) από την άλλη μεριά η οποία βρίσκεται και σαν στήλη στο σύνολο των δεδομένων μας αποτελεί την ισχύ η οποία εμπίπτει στην κατανάλωση, επομένως σχετίζεται με το πόση κατανάλωση γίνεται πραγματικά ή αντίστοιχα μπορεί να χρησιμοποιηθεί σε ένα κύκλωμα. Ακόμα είναι σημαντικό να αναφερθεί ότι η ενεργή ισχύς αποκαλείται και ως μέση ισχύς και κατά κύριο λόγο αποτελεί το μέτρο που σχετίζεται με την κατανάλωση της ενέργειας και τον ρυθμό με τον οποίο επιτυγχάνεται η κατανάλωση αυτή. (Αnon., n.d.)

6.2.2 ΤΑΣΗ

Η τάση έχει σαν κύριο χαρακτηριστικό της ότι είναι η διαφορά δυναμικού που προκύπτει ανάμεσα σε δύο σημεία και ακόμα σαν ορισμός υφίσταται ο όρος της

ενέργειας ως προς τη μονάδα του φορτίου ή διαφορετικά $V = W/Q$, όπου V είναι ο συμβολισμός της τάσης, W είναι ο συμβολισμός για το έργο ή αλλιώς ενέργεια και Q είναι το φορτίο που σχετίζεται με τη μεταφορά της ενέργειας αυτής (Πουλάκης, 2015).

Η τάση μπορεί να εντοπιστεί τόσο σε κυκλώματα με συνεχές όσο και με εναλλασσόμενο ρεύμα όμως εμφανίζουν κάποιες μικρές διαφορές. Η τάση σε ένα κύκλωμα με εναλλασσόμενο ρεύμα είναι λογικό ότι παρότι έχει ομοιόμορφες κυματομορφές μπορεί και αλλάζει τη κατεύθυνση ανά παρόμοιες χρονικές περιόδους και συνήθως το εναλλασσόμενο ρεύμα είναι αυτό που παρέχεται στους καταναλωτές. Από την άλλη πλευρά το συνεχές ρεύμα έχει μόνο μία κατεύθυνση και βρίσκεται σε ευθεία πάντοτε. Ακόμη κατά κύριο λόγο βρίσκεται σε πηγές που μπορούν να αποθηκεύσουν ενέργεια όπως μπορεί να είναι οι μπαταρίες (Αnon., n.d.). Γενικότερα από τις σημαντικότερες πηγές που διαθέτει η τάση είναι ως πρώτη η κατηγορία των συσσωρευτών ή αλλιώς των μπαταριών που αποτελούνται από διάφορα ηλεκτροχημικά στοιχεία. Εν συνεχεία είναι τα ηλιακά ή φωτοβολταϊκά κύτταρα (solar cells) όπου με την αρωγή των φωτοβολταϊκών μπορούν να μετατρέψουν την ηλιακή σε ηλεκτρική ενέργεια. Ακόμη μία πηγή θεωρείται η γεννήτρια και τροφοδοτικό. Η γεννήτρια κάνει την απαραίτητη μετατροπή ενέργειας μέσω της αρχής της ηλεκτρομαγνητικής επαγωγής ενώ το τροφοδοτικό κάνει μετατροπή του εναλλασσόμενου σε συνεχές ρεύμα (Πουλάκης, 2015).

6.3 ΧΑΜΕΝΕΣ ΤΙΜΕΣ

Αρχικά είναι σημαντικό να υπενθυμίσουμε ότι τα δεδομένα μπορούν πολλές φορές να είναι είτε κακής ποιότητας είτε να χαρακτηρίζονται ως «ακάθαρτα». Αμφότερα και στις δύο περιπτώσεις μπορεί να γίνει αντιληπτό πως τα δεδομένα που βρίσκονται στο σύνολο δεδομένων μπορεί να μην είναι ολοκληρωμένα, να έχουν θόρυβο και να πέφτουν σε ασυμφωνίες. Οι παραπάνω λόγοι μπορεί να προέρχονται από ελλιπείς τιμές και έλλειψη σημαντικών χαρακτηριστικών. Επομένως γίνεται εμφανές ότι οι χαμένες τιμές που υπάρχουν τις περισσότερες φορές στα σύνολα δεδομένων μπορεί να οφείλονται σε βλάβες που προέρχονται από τον εξοπλισμό που καταγράφει τις τιμές, λόγω μη συμβατών μεταξύ τους δεδομένων ακόμα και δεδομένα που δεν έγινε ποτέ η συμπλήρωση τους λόγω του ανθρώπινου παράγοντα. Σε ενεργειακά δεδομένα που σχετίζονται με την ηλιοφάνεια πέραν των παραπάνω λόγων που είναι πολύ πιθανό να βρεθούν κενές τιμές, πρέπει να θέσουμε σαν παράγοντα και την παράμετρο της συννεφιάς όπου σε αυτή την περίπτωση είναι σχεδόν αναμενόμενο ότι περιμένουμε να δούμε είτε δεδομένα με τιμές διαφορετικές από τις προηγούμενες καθώς θα παρουσιάζονται μειωμένες.

Γι' αυτό το λόγο είναι απαραίτητο να μπορέσουμε να χειριστούμε τα δεδομένα που λείπουν από το αντίστοιχο σύνολο. Πρώτος τρόπος ο οποίος όμως δεν είναι και ο προτιμότερος ειδικά όταν δεν θέλουμε να χάσουμε σημαντικές πληροφορίες από τα δεδομένα μας είναι η διαγραφή μίας ολόκληρης κενής γραμμής. Αυτός ο τρόπος δεν συνιστάται και χρησιμοποιείται συνήθως μόνο όταν η γραμμή έχει πολλές κενές τιμές. Ακόμη ένας τρόπος ο οποίος μπορεί να θεωρηθεί ως και ο βέλτιστος είναι η

προσπάθεια για την εύρεση της πραγματικής τιμής, όπου όπως είναι σχεδόν αναμενόμενο αποτελεί μία διαδικασία εξαιρετικά χρονοβόρα και ταυτόχρονα υπάρχει έντονη η πιθανότητα να μην μπορέσει ποτέ να βρεθεί αυτή η πραγματική τιμή. Εν συνεχεία είναι πιο συνήθης η αντικατάσταση των τιμών που λείπουν αυτό μπορεί να γίνει είτε με μία συγκεκριμένη τιμή είτε με τη πιο συνηθισμένη και τη μέση τιμή αλλά μπορεί να υπάρξει αντικατάσταση και με οποιαδήποτε τιμή. Τελευταία αλλά αρκετά ιδιαίτερη και σημαντική επιλογή θα μπορούσε να είναι η πρόβλεψη της χαμένης τιμής. Σε αυτή τη περίπτωση μπορεί να αναπτυχθεί ένας αλγόριθμος ή ακόμη και ένα μοντέλο ώστε να βρίσκει τις τιμές που αφορούν την κάθε στήλη και να βάζει την κοντινότερη κατά εκτίμηση πάντα τιμή (Δεμερτζής, 2020).

Αντίστοιχα λοιπόν όταν ξεκινήσαμε με το σύνολο δεδομένων στα χέρια μας με έναν απλό έλεγχο διαπιστώσαμε ότι υπήρξαν αρκετά κενά με τιμές που είχαν εντελώς κενές τιμές. Σε αυτό το σημείο έπρεπε να πάρουμε κάποια απόφαση σχετικά με το πώς θα γίνει η επεξεργασία των τιμών αυτών καθώς έπρεπε να εξεταστούν σωστά τα βήματα της προεπεξεργασίας και να συμπληρωθούν οι αντίστοιχες τιμές ώστε να μην υπάρξει πρόβλημα σε περίπτωση που το αντίστοιχο σύνολο δεδομένων χρησιμοποιηθεί σε κάποιο μοντέλο πρόβλεψης μετέπειτα. Για αρχή ήταν απαραίτητο ο εντοπισμός των τιμών που ήταν κενές. Με την έννοια του εντοπισμού εννοούμε ότι βρίσκουμε τις θέσεις μέσα στο dataset όπου εντοπίζονται οι τιμές αυτές. Στη συνέχεια αποφασίζουμε να τις απομονώσουμε σε ένα csv αρχείο για να μπορέσουμε να συμπληρώσουμε και τις κενές τιμές.

Κατά βήμα λοιπόν μπορούμε να πούμε ότι ξεκινήσαμε έχοντας τα δεδομένα τα όποια προέρχονται από το διαδίκτυο σε ένα txt αρχείο το οποίο έγινε η μετατροπή του σε csv για μεγαλύτερη ευκολία. Εν συνεχεία αποφασίσαμε ότι είναι σωστό να δούμε σε ένα αρχείο ποιες σειρές είναι κενές και ποιες σειρές έχουν μέσα τιμές. Αυτό συνέβη με το να εμφανίζεται σε ένα αρχείο οι κενές σειρές με την τιμή True ενώ στις σειρές που υπάρχουν όντως τιμές εμφανίζεται με την τιμή False. Αξίζει να αναφέρουμε ότι στο αρχείο που δημιουργήθηκε η ώρα εμφανίζεται με τη συχνότητα του ενός λεπτού και πρόκειται για αρχείο δεδομένων που έχει καταγραφές τεσσάρων χρόνων.

6835	2006-12-21 11:17:00,False,False
6836	2006-12-21 11:18:00,False,False
6837	2006-12-21 11:19:00,False,False
6838	2006-12-21 11:20:00,False,False
6839	2006-12-21 11:21:00,False,False
6840	2006-12-21 11:22:00,False,False
6841	2006-12-21 11:23:00,True,True
6842	2006-12-21 11:24:00,True,True
6843	2006-12-21 11:25:00,False,False
6844	2006-12-21 11:26:00,False,False
6845	2006-12-21 11:27:00,False,False
6846	2006-12-21 11:28:00,False,False
6847	2006-12-21 11:29:00,False,False

Εικόνα 19 Κενές Τιμές ορισμένες με τον χαρακτήρα True

Τα αρχεία μας γενικά αποτελούνται από τρεις βασικές στήλες. Η πρώτη έχει να κάνει με την καταγραφή της ημερομηνίας και της ώρας, η δεύτερη έχει να κάνει με την ενεργή ισχύ(active power) και η τρίτη έχει να κάνει με την τάση (voltage). Σε αυτό το

	A	B	C
1	datetime	Global_active_power	Voltage
2	16/12/2006 17:24	FALSE	FALSE
3	16/12/2006 17:25	FALSE	FALSE
4	16/12/2006 17:26	FALSE	FALSE
5	16/12/2006 17:27	FALSE	FALSE
6	16/12/2006 17:28	FALSE	FALSE

Εικόνα 20 Στήλες Δεδομένων που Μελετήθηκαν

σημείο θα γίνει επεξήγηση της διαδικασίας του κώδικα σε γλώσσα προγραμματισμού Python σε περιβάλλον εργασίας Spyder.

Συνεχίζοντας θεωρήσαμε σωστότερο να μπορέσουμε να δούμε τις σειρές και τις καταγραφές που εμφάνιζαν μόνο κενές τιμές ανά λεπτό.

	A	B	C
1	datetime	Global_active_power	Voltage
2	21/12/2006 11:23	TRUE	TRUE
3	21/12/2006 11:24	TRUE	TRUE
4	30/12/2006 10:08	TRUE	TRUE
5	30/12/2006 10:09	TRUE	TRUE
6	14/1/2007 18:36	TRUE	TRUE
7	28/1/2007 17:13	TRUE	TRUE
8	22/2/2007 22:58	TRUE	TRUE
9	22/2/2007 22:59	TRUE	TRUE
10	25/3/2007 17:52	TRUE	TRUE
11	28/4/2007 0:21	TRUE	TRUE
12	28/4/2007 0:22	TRUE	TRUE
13	28/4/2007 0:23	TRUE	TRUE
14	28/4/2007 0:24	TRUE	TRUE
15	28/4/2007 0:25	TRUE	TRUE
16	28/4/2007 0:26	TRUE	TRUE
17	28/4/2007 0:27	TRUE	TRUE
18	28/4/2007 0:28	TRUE	TRUE
19	28/4/2007 0:29	TRUE	TRUE

Εικόνα 21 Κενές Τιμές που Εντοπίστηκαν στο Αρχείο Ανά Λεπτό

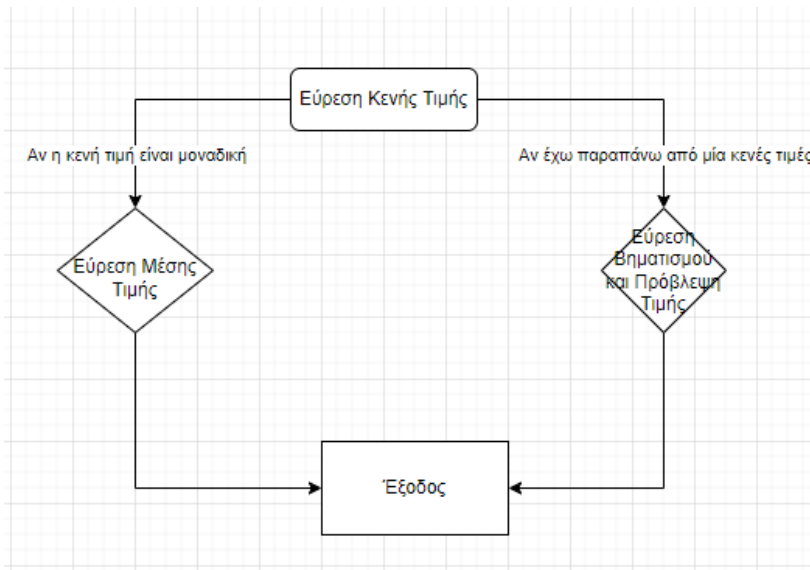
Ακόμα προχωρώντας θέσαμε την ώρα ανά δεκαπέντε λεπτά αντί για την συχνότητα του λεπτού και βρήκαμε τον μέσο όρο των τιμών του δεκαπεντάλεπτου μόνο αν οι κενές τιμές ήταν λιγότερες από δέκα αν υπήρχαν κενές τιμές παραπάνω από δέκα

11	2007-04-28 00:21:00,True,True				
12	2007-04-28 00:22:00,True,True	12697	12697,2007-04-27 23:24:00,1.0254000000000000,237.92500000000000		
13	2007-04-28 00:23:00,True,True	12698	12696,2007-04-27 23:39:00,1.3237333333333334,237.93933333333334		
14	2007-04-28 00:24:00,True,True	12699	12697,2007-04-27 23:54:00,1.5051999999999999,237.39333333333335		
15	2007-04-28 00:25:00,True,True	12700	12698,2007-04-28 00:09:00,1.3669333333333333,233.12133333333335		
16	2007-04-28 00:26:00,True,True	12701	12699,2007-04-28 00:24:00,,		
17	2007-04-28 00:27:00,True,True	12702	12700,2007-04-28 00:39:00,,		
18	2007-04-28 00:28:00,True,True	12703	12701,2007-04-28 00:54:00,,		
19	2007-04-28 00:29:00,True,True	12704	12702,2007-04-28 01:09:00,,		
20	2007-04-28 00:30:00,True,True	12705	12703,2007-04-28 01:24:00,,		
21	2007-04-28 00:31:00,True,True	12706	12704,2007-04-28 01:39:00,,		
22	2007-04-28 00:32:00,True,True	12707	12705,2007-04-28 01:54:00,,		
23	2007-04-28 00:33:00,True,True	12708	12706,2007-04-28 02:09:00,,		
24	2007-04-28 00:34:00,True,True	12709	12707,2007-04-28 02:24:00,,		
25	2007-04-28 00:35:00,True,True	12710	12708,2007-04-28 02:39:00,,		
26	2007-04-28 00:36:00,True,True	12711	12709,2007-04-28 02:54:00,,		
27	2007-04-28 00:37:00,True,True				
28	2007-04-28 00:38:00,True,True				
29	2007-04-28 00:39:00,True,True				

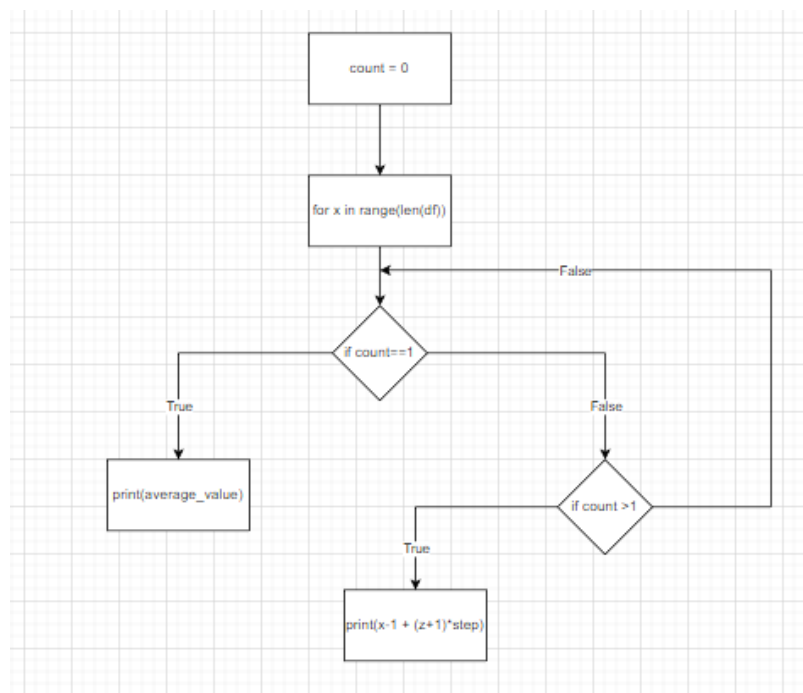
Εικόνα 22 Σύγκριση Τιμών για Κατανόηση του Αλγορίθμου

τότε οι σειρές των στηλών παράμεναν κενές και αντίστοιχα δημιουργήσαμε και το αντίστοιχο αρχείο.

Στη συνέχεια βρήκαμε το σύνολο των κενών γραμμών που υπάρχουν σε όλο το σύνολο δεδομένων και στη συνέχεια βρήκαμε και όλες τις κενές τιμές αναλυτικά ανά γραμμή. Αυτό έγινε μέσω του Project Jupyter, όπου πρόκειται για διαδραστική πλατφόρμα. Βέβαια το σημαντικότερο από όλα είναι που μέσω ενός αλγορίθμου καταφέραμε να γεμίσουμε τις κενές αυτές τιμές που υπάρχουν. Ο αλγόριθμος εφαρμόστηκε ανάλογα με το αν υπήρχε κενή τιμή.



Εικόνα 23 Επεξήγηση Κώδικα



Εικόνα 24 Flowchart

Παραδείγματος χάριν αν έχω δύο κενές τιμές όπως συμβαίνει με τις γραμμές που απεικονίζονται στις παρακάτω εικόνες τότε για την πρώτη στήλη θα έχω $0.3401333333333334 + (0+1) * 0.1435555555555552 = 0.4836888888888895$

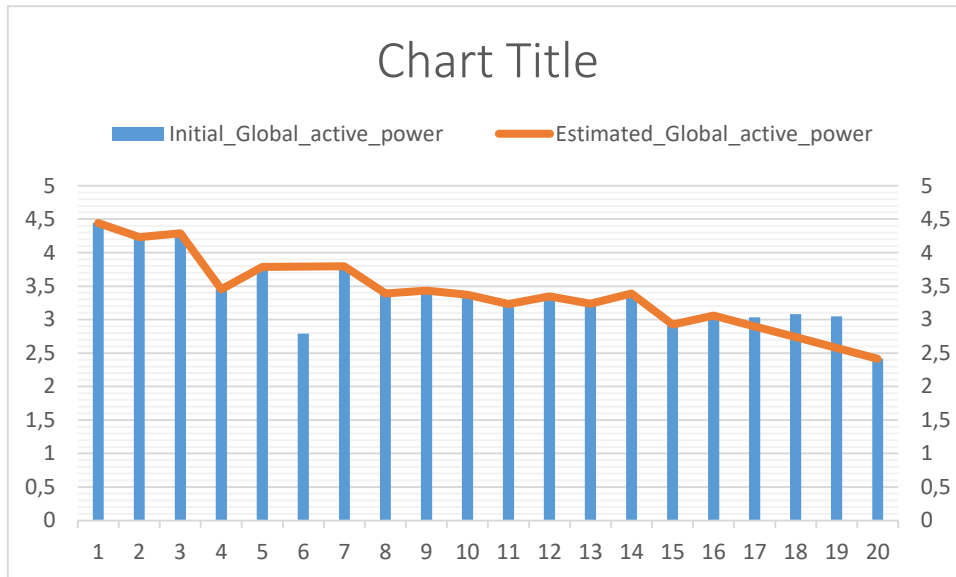
453	451,2006-12-21 10:24:00,3.195333333333333,237.115333333333333
454	452,2006-12-21 10:39:00,1.4573333333333334,240.181333333333333
455	453,2006-12-21 10:54:00,1.0906666666666667,241.624
456	454,2006-12-21 11:09:00,0.3401333333333334,240.378
457	455,2006-12-21 11:24:00,,
458	456,2006-12-21 11:39:00,,
459	457,2006-12-21 11:54:00,0.7707999999999999,242.631333333333332
460	458,2006-12-21 12:09:00,0.5206666666666667,243.980000000000002
461	459,2006-12-21 12:24:00,0.3845333333333334,243.680666666666667

Εικόνα 25 Δύο Τυχαίες Κενές Τιμές

454	452,2006-12-21 10:39:00,1.4573333333333334,240.181333333333333
455	453,2006-12-21 10:54:00,1.0906666666666667,241.624
456	454,2006-12-21 11:09:00,0.3401333333333334,240.378
457	455,2006-12-21 11:24:00,0.4836888888888895,240.521555555555555
458	456,2006-12-21 11:39:00,0.6272444444444445,240.665111111111111
459	457,2006-12-21 11:54:00,0.7707999999999999,242.631333333333333
460	458,2006-12-21 12:09:00,0.5206666666666667,243.98
461	459,2006-12-21 12:24:00,0.3845333333333333,243.680666666666667

Εικόνα 26 Συμπλήρωση Των Κενών Τιμών

Για να μπορέσουμε να διασταυρώσουμε τα αποτελέσματα που συμπληρώνονται ως προς τις κενές τιμές αποφασίσαμε να τεστάρουμε το σύνολο δεδομένων διαγράφοντας κάποιες τιμές που ξέρουμε ότι υπήρχαν στο σύνολο δεδομένων και μας είναι γνωστές. Τις διαγράφουμε λοιπόν με την προοπτική ότι θέλουμε να δούμε πως θα γίνει το γέμισμα τους και πόσο κοντινές θα είναι οι τιμές με αυτές που προϋπήρχαν ήδη. Οι τιμές και οι διαφορές τους φαίνονται στο παρακάτω διάγραμμα.



Εικόνα 27 Γράφημα με Τιμές Πριν και Μετά

	A	B	C
1	datetime	Initial_Global_active_power	Estimated_Global_active_power
2	16/12/2006 17:39	4,4441	4,4441
3	16/12/2006 17:54	4,2333	4,2333
4	16/12/2006 18:09	4,2867	4,2867
5	16/12/2006 18:24	3,4501	3,4501
6	16/12/2006 18:39	3,7854	3,7854
7	16/12/2006 18:54	2,7889	3,7908
8	16/12/2006 19:09	3,7961	3,7961
9	16/12/2006 19:24	3,3924	3,3924
10	16/12/2006 19:39	3,4311	3,4311
11	16/12/2006 19:54	3,3727	3,3727
12	16/12/2006 20:09	3,2314	3,2314
13	16/12/2006 20:24	3,3479	3,3479
14	16/12/2006 20:39	3,2364	3,2364
15	16/12/2006 20:54	3,3899	3,3899
16	16/12/2006 21:09	2,925	2,925
17	16/12/2006 21:24	3,0618	3,0618
18	16/12/2006 21:39	3,0328	2,9002
19	16/12/2006 21:54	3,0835	2,7386
20	16/12/2006 22:09	3,0494	2,5769
21	16/12/2006 22:24	2,4153	2,4153

Εικόνα 28 Τιμές Πριν και Μετά τη Πρόβλεψη

ΚΕΦΑΛΑΙΟ 7: ΣΥΜΠΕΡΑΣΜΑΤΑ

Συνοψίζοντας λοιπόν αντιλαμβανόμαστε ότι οι ανανεώσιμες πηγές ενέργειας κατακλύζουν την καθημερινότητα μας και γίνονται στον καθένα μας ολοένα και πιο απαραίτητες. Πέραν αυτού όμως τα δεδομένα που τίθενται προς ανάλυση πρέπει να υπόκεινται σε επεξεργασία για να μπορέσουν να αποδώσουν με τον καλύτερο δυνατό τρόπο τα μοντέλα τα οποία θα χρησιμοποιούν τα αντίστοιχα σύνολα δεδομένων. Είναι εξαιρετικά σημαντικό να διαθέτουμε ποιοτικά δεδομένα με τα λιγότερα δυνατά λάθη. Εξάλλου στις μέρες μας οι χρονοσειρές και ειδικά σε ενεργειακά δεδομένα γίνονται όλο και πιο γνωστές για την πρόβλεψη τιμών και όχι μόνο. Τα μοντέλα παλινδρόμησης τα οποία εξετάστηκαν έστω και σε θεωρητικό υπόβαθρο έχουν αρκετά καλή απόδοση με βάση κάποιες παραμέτρους και είναι ιδιαίτερα σημαντικό πριν να γίνει η προσπάθεια για να χρησιμοποιηθούν ο ενδιαφερόμενος να γνωρίζει εξαιρετικά τα δεδομένα που διαθέτει και τα χαρακτηριστικά τους, διότι διαφορετικά θα υπάρξει σημαντικό πρόβλημα και στην εφαρμογή των παραμέτρων των μοντέλων. Τέλος από μεριάς μας έγινε προσπάθεια για αναγνώριση σφαλμάτων που υπήρξαν στα δεδομένα και αντικατάσταση τους με τον καλύτερο δυνατό τρόπο ώστε να μην χαθούν τιμές που μπορεί να αποδειχτούν ιδιαίτερα πολύτιμες στο μέλλον.

ΒΙΒΛΙΟΓΡΑΦΙΑ

Abhishek, 2020. *What is Data Consistency?*. [Online]

Available at: <https://thecustomizewindows.com/2020/01/what-is-data-consistency/>
[Accessed 10 07 2022].

Anon., 2021. *What Is Data Completeness? Why One Missing Piece Matters*. [Online]

Available at: <https://www.bdex.com/blog/why-is-data-completeness-important/#:~:text=Data%20completeness%20refers%20to%20the,costly%20mistakes%20and%20false%20conclusions.>

[Accessed 10 07 2022].

Anon., n.d. [Online]

Available at: <https://www.renovablesverdes.com/el/que-son-las-fuentes-de-energia/>
[Accessed 21 03 2022].

Anon., n.d. [Online]

Available at:

https://el.wikipedia.org/wiki/%CE%91%CE%BD%CE%B1%CE%BD%CE%B5%CF%8E%CF%83%CE%B9%CE%BC%CE%B5%CF%82_%CF%80%CE%B7%CE%B3%CE%AD%CF%82_%CE%B5%CE%BD%CE%AD%CF%81%CE%B3%CE%B5%CE%B9%CE%B1%CF%82#%CE%95%CE%AF%CE%B4%CE%B7_%CE%AE%CF%80%CE%B9%CF%89%CE%BD_%CE%BC%CE%BF%CF%

[Accessed 21 03 2022].

Anon., n.d. *Electricity Fundamentals - Active Power*. [Online]

Available at: <https://www.electricityforum.com/active-power#:~:text=Active%20power%20is%20the%20power,the%20electric%20circuits%20or%20load.>

[Accessed 01 07 2022].

Anon., n.d. *What is data reliability?*. [Online]

Available at: <https://www.talend.com/resources/what-is-data-reliability/#:~:text=Data%20reliability%20means%20that%20data,data%20quality%2C%20and%20regulatory%20compliance.>

[Accessed 10 07 2022].

Anon., n.d. *What is power?*. [Online]

Available at: <https://www.khanacademy.org/science/physics/work-and-energy/work-and-energy-tutorial/a/what-is-power>

[Accessed 01 07 2022].

Anon., n.d. *What is Voltage?*. [Online]

Available at: <https://www.fluke.com/en/learn/blog/electrical/what-is-voltage>
[Accessed 02 07 2022].

Anon., n.d. *Ανάπτυξη φωτοβολταϊκών σταθμών*. [Online]

Available at: <https://mgd-energy.gr/paragogi-energeias/>
[Accessed 2022].

Eurostat, 2022. *Renewable energy statistics*. [Online]

Available at: <https://ec.europa.eu/eurostat/statistics->

[explained/index.php?title=Renewable energy statistics#Share of renewable energy more than doubled between 2004 and 2020](#)
[Accessed 12 04 2022].

Farah, 2020. *What is Data Accuracy, Why it Matters and How Companies Can Ensure They Have Accurate Data..* [Online]
Available at: <https://dataladder.com/what-is-data-accuracy/>
[Accessed 10 07 2022].

Ling, 2013. *Energy Time Series Data Analysis based on a Novel Integrated Data Characteristic Testing Approach.* [Online]
Available at:
https://www.researchgate.net/publication/257719585_Energy_Time_Series_Data_Analysis_based_on_a_Novel_Integrated_Data_Characteristic_Testing_Approach
[Accessed 15 04 2022].

Shruthi, 2020. *Parametric and Non-parametric Models in Machine Learning.* [Online]
Available at: <https://medium.com/analytics-vidhya/parametric-and-nonparametric-models-in-machine-learning-a9f63999e233#:~:text=A%20parametric%20model%20can%20predict,they%20make%20about%20the%20data.>
[Accessed 27 05 2022].

Vaughan, 2019. *TechTarget.* [Online]
Available at: <https://www.techtarget.com/searchdatamanagement/definition/data-quality>
[Accessed 29 06 2022].

Wikipedia, n.d. [Online]
Available at: https://en.wikipedia.org/wiki/List_of_photovoltaic_power_stations
[Accessed 12 04 2022].

Yufeng, 2017. *The 7 Steps of Machine Learning.* [Online]
Available at: <https://towardsdatascience.com/the-7-steps-of-machine-learning-2877d7e5548e>
[Accessed 26 05 2022].

Αδαμόπουλος, 1999-2000. *Πιθανότητες και Στοιχεία Στατιστικής (Γ' Γενικού Λυκείου) ΙΤΥΕ Διόφαντος.* [Online]
Available at: http://ebooks.edu.gr/ebooks/v/html/8547/4704/Mathimatika-kai-Stoicheia-Statistikis_G-EPAL_html-apli/index2_4.html
[Accessed 08 05 2022].

Αναστασιάδου, 2019. *Ερμηνεύσιμη Μηχανική Μάθηση.* [Online]
Available at:
<http://ikee.lib.auth.gr/record/309395/files/%CE%91%CE%BB%CE%B5%CE%BE%CE%AC%CE%BD%CE%B4%CF%81%CE%B1%CE%A7%CF%81%CE%B9%CF%83%CF%84%CE%AF%CE%BD%CE%B1%CE%91%CE%BD%CE%B1%CF%83%CF%84%CE%B1%CF%83%CE%B9%CE%AC%CE%B4%CE%BF%CF%85.pdf>
[Accessed 25 05 2022].

Αναστασίου, 2019. *Αλγόριθμοι Μηχανικής Μάθησης και Εφαρμογές σε Ιατροβιολογικά Προβλήματα.* [Online]

Available at:
https://nemertes.library.upatras.gr/jspui/bitstream/10889/13208/1/eytychia_anastasiou_di_plomatiki.pdf
[Accessed 20 05 2022].

Αντωνίου, n.d. *5.8 Ισχύς*. [Online]
Available at: http://ebooks.edu.gr/ebooks/v/html/8547/2204/Fysiki_B-Gymnasiou_html-empl/index.html
[Accessed 01 07 2022].

Βλαχάβας, 2011. *Τεχνητή Νοημοσύνη- Β' Έκδοση, Κεφάλαιο 18*. [Online]
Available at: <http://aibook.csd.auth.gr/include/slides/Chap18.pdf>
[Accessed 20 05 2022].

Δαγούμας, 2020. *Αυτοπαλίνδρομα Μοντέλα Κινητού Μέσου Όρου (ARIMA) (Σημειώσεις για το μάθημα «Τεχνικές Πρόβλεψεις»)*. [Online]
Available at:
<https://eclass.unipi.gr/modules/document/file.php/DES103/%CE%94%CE%B1%CE%B3%CE%BF%CF%8D%CE%BC%CE%B1%CF%82/ARIMA/%CE%91%CF%85%CF%84%CE%BF%CF%80%CE%B1%CE%BB%CE%B9%CE%BD%CE%B4%CF%81%CE%BF%CE%BC%CE%B9%CE%BA%CE%AC%20%CE%9C%CE%BF%CE%BD%CF%84%CE%AD%CE%BB%CE%B1%2>
[Accessed 10 05 2022].

Δεμερτζής, 2020. *Πρόγραμμα Σπουδών Ηλεκτρολόγων Μηχανικών - Νευρωνικά Δίκτυα*. [Online]
Available at:
<https://eclass.emt.ihu.gr/modules/document/file.php/ED167/2020/3.%CE%9C%CE%B7%CF%87%CE%B1%CE%BD%CE%B9%CE%BA%CE%AE%20%CE%9C%CE%AC%CE%B8%CE%B7%CF%83%CE%B7.pdf>
[Accessed 03 07 2022].

Καραϊσάς, 2014. *Φωτοβολταϊκά Συστήματα Και Ανανεώσιμες Πηγές Ενέργειας*. [Online]
Available at: https://imegsevee.gr/wp-content/uploads/2018/01/fotovoltaika_systimata.pdf
[Accessed 25 03 2022].

Κατσαντώνης, 2020. *Ανάλυση Δεδομένων και Κατηγοριοποίηση*. [Online]
Available at:
https://dione.lib.unipi.gr/xmlui/bitstream/handle/unipi/12983/Katsantonis_1608.pdf?sequence=1&isAllowed=y
[Accessed 12 04 2022].

Κοντογιάννης, 2022. *Αποθορυβοποίηση Χρονοσειρών με Χρήση Νευρωνικών Δικτύων Βαθέων Αρχιτεκτονικών*. [Online]
Available at:
https://nemertes.library.upatras.gr/jspui/bitstream/10889/16246/1/Thesis_up1059526.pdf
[Accessed 29 06 2022].

Κουγιουμτζής, 2019-2020. *Ανάλυση Δεδομένων 2019-2020, Κεφάλαιο 6 – Χρονοσειρές*. [Online]
Available at: <https://users.auth.gr/dkugiu/Teach/DataAnalysis/Chp6.pdf>
[Accessed 15 04 2022].

Κύρκος, 2015. 7. Προεπεξεργασία Δεδομένων. [Online]
Available at: <https://repository.kallipos.gr/handle/11419/1234>
[Accessed 15 04 2022].

Κωστόπουλος, 2018. Πρόβλεψη Χρεοκοπίας Επιχειρήσεων με Χρήση Μεθόδων Μηχανικής Μάθησης. [Online]
Available at:
<http://repository.teiwest.gr/xmlui/bitstream/handle/123456789/7089/DE%20%CE%9A%CE%A9%CE%A3%CE%A4%CE%9F%CE%A0%CE%9F%CE%A5%CE%9B%CE%9F%CE%A3%20%CE%93%CE%95%CE%A9%CE%A1%CE%93%CE%99%CE%9F%CE%A3.pdf?sequence=1&isAllowed=y>
[Accessed 25 05 2022].

Μαγίδου, 2016. Ανάπτυξη κώδικα Matlab για την διενέργεια μακροχρόνιων ενεργειακών προβλέψεων στην Κύπρο. [Online]
Available at:
<https://ktisis.cut.ac.cy/bitstream/10488/8717/1/%CE%9C%CE%B1%CF%81%CE%AF%CE%BD%CE%B1%20%CE%9C%CE%B1%CE%B3%CE%AF%CE%B4%CE%BF%CF%85.pdf#page=21&zoo m=100,101,162>
[Accessed 20 03 2022].

Νικόλαος, Α., 2015. 1. Ενέργεια: Ορισμοί, Μορφές και χρήσεις. [Online]
Available at:
http://archive.eclass.uth.gr/eclass/modules/document/file.php/MHXB122/Enotita_1.pdf
[Accessed 20 03 2022].

Νικολάου, 2007. Συγκριτική ανάλυση και εφαρμογή γραμμικών, μη-γραμμικών και νευρο-ασαφών μεθόδων, για τη βραχυπρόθεσμη πρόβλεψη παραγωγής ενέργειας από αιολικά πάρκα. [Online]
Available at: <http://artemis.library.tuc.gr/DT2008-0113/DT2008-0113.pdf>
[Accessed 06 05 2022].

Νταβίντ, 2018. Συλλογή Μακροχρόνιων Χρονοσειρών για Παραγωγή και Αξιολόγηση Μακροπρόθεσμων Προβλέψεων. [Online]
Available at:
<https://dspace.lib.ntua.gr/xmlui/bitstream/handle/123456789/46919/%CE%A3%CF%85%CE%BB%CE%BB%CE%BF%CE%B3%CE%AE%20%CE%9C%CE%B1%CE%BA%CF%81%CE%BF%CF%87%CF%81%CF%8C%CE%BD%CE%B9%CF%89%CE%BD%20%CF%87%CF%81%CE%BF%CE%BD%CE%BF%CF%83%CE%B5%CE%B9%CF%81%CF%8E%CE%BD%20>
[Accessed 06 05 2022].

Παναγοπούλου, 2020. Πρόβλεψη Χρονοσειρών με Έντονη Εποχικότητα. [Online]
Available at:
https://nemertes.library.upatras.gr/jspui/bitstream/10889/13878/1/DT_M_%ce%a0%ce%b1%ce%bd%ce%b1%ce%b3%ce%bf%cf%80%ce%bf%cf%8d%ce%bb%ce%bf%cf%85_%ce%95%cf%85%cf%83%cf%84%ce%b1%ce%b8%ce%af%ce%b1_247004.pdf
[Accessed 11 05 2022].

Πουλάκης, 2015. ΚΕΦΑΛΑΙΟ 2: Τάση, Ρεύμα και Αντίσταση. [Online]
Available at:
<https://eclass.teiwm.gr/modules/document/file.php/EE108/%CE%98%CE%B5%CF%89%CF%>

[81%CE%AF%CE%B1/%CE%A3%CE%B7%CE%BC%CE%B5%CE%B9%CF%8E%CF%83%CE%B5%CE%B9%CF%82%20%CE%B4%CE%B9%CE%B1%CE%BB%CE%AD%CE%BE%CE%B5%CF%89%CE%BD%20%CE%B1%CE%BA%CE%B1%CE%B4.%20%CE%AD%CF%84%CE](https://www.ceid.upatras.gr/webpages/courses/cplusplus/dm/1_Data_preprocessing.pdf)
[Accessed 02 07 2022].

Σκούρα, 2020-2021. *Εξόρυξη Δεδομένων και Αλγόριθμοι Μάθησης- Προεπεξεργασία Δεδομένων*. [Online]

Available at:

https://www.ceid.upatras.gr/webpages/courses/cplusplus/dm/1_Data_preprocessing.pdf
[Accessed 15 04 2022].

Τσαούσης, 2018. [Online]

[Accessed 05 2022].

Τσαούσης, 2018. *Ανάλυση Χρηματοοικονομικών Χρονοσειρών και Πρόβλεψη με Γραμμικά Μοντέλα*. [Online]

Available at: <http://ikee.lib.auth.gr/record/297900/files/GRI-2018-21573.pdf>

[Accessed 07 05 2022].

Τσιμπάνος, 2018. *Κεφάλαιο 1- Βασικές Έννοιες Χρονοσειρών*. [Online]

Available at: <https://eclass.aegean.gr/modules/document/file.php/SAS142/KEF1.pdf>

[Accessed 06 05 2022].

Χασιρτζόγλου, 2020. *Ανάλυση και Πρόβλεψη Χρονοσειρών με Μοντέλα ARIMA και Εφαρμογές*. [Online]

Available at:

<https://dspace.lib.ntua.gr/xmlui/bitstream/handle/123456789/52222/%CE%94%CE%B9%CF%80%CE%BB%CF%89%CE%BC%CE%B1%CF%84%CE%B9%CE%BA%CE%AE%20%CE%B5%CF%81%CE%B3%CE%B1%CF%83%CE%AF%CE%B1%20%CE%A7%CE%B1%CF%83%CE%B9%CF%81%CF%84%CE%B6%CF%8C%CE%B3%CE%BB%CE%BF%CF%85%20>

[Accessed 07 05 2022].

Χατζηευλογημένος, 2019. *Εφαρμογές Στοχαστικών Διαδικασιών στα Παιχνίδια Τυχαίων Αριθμών, το Παράδειγμα του Παιχνιδιού Τζόκερ*. [Online]

Available at:

<https://hellanicus.lib.aegean.gr/bitstream/handle/11610/19647/%CE%A0%CF%84%CF%85%CF%87%CE%B9%CE%B1%CE%BA%CE%AE%20%CE%B5%CF%81%CE%B3%CE%B1%CF%83%CE%AF%CE%B1%20%CE%A7%CE%B1%CF%84%CE%B6%CE%B7%CE%B5%CF%85%CE%BB%CE%BF%CE%B3%CE%B7%CE%BC%CE%AD%CE%BD%CE%BF%CF%82%>

[Accessed 10 05 2022].

Χατζηλυγερούδης, 2020. *Τεχνητή Νοημοσύνη- 6.Εισαγωγή στη Μηχανική Μάθηση και Βελτιστοποίηση*. [Online]

Available at:

<https://eclass.upatras.gr/modules/document/file.php/CEID1104/%CE%94%CE%99%CE%91%CE%9B%CE%95%CE%9E%CE%95%CE%99%CE%A3/%CE%94%CE%99%CE%91%CE%9B%CE%95%CE%9E%CE%95%CE%99%CE%A3%202020-21/6%20-%20%CE%95%CE%99%CE%A3%CE%91%CE%93%CE%A9%CE%93%CE%97%20%CE%A3%CE%A4%CE>

[Accessed 27 05 2022].

ΠΗΓΕΣ ΕΙΚΟΝΩΝ

Εικόνα 1: <https://www.nea-acropoli-athens.gr/arhra/oikologia/758-enallaktikes-morfes-energeias>

Εικόνα 4: <https://ypodomes.com/to-2022-provlepetai-na-mpoyn-sto-diktyo-1-900-mw-apo-ape/>

Εικόνα 5: https://ec.europa.eu/eurostat/statistics-explained/index.php?title=Renewable_energy_statistics#Share_of_renewable_energy_more_than_doubled_between_2004_and_2020

Εικόνα 7: <https://kritiki.gr/microsites/dioikisi-epixeirisiakwn-leitourgiwn/chapter/ch7/graphics/img7-1>

Εικόνα 9: <http://artemis.cslab.ece.ntua.gr:8080/jspui/bitstream/123456789/13056/1/DT2016-0036.pdf>

Εικόνα 10: <https://slideplayer.gr/slide/1987676/>

Εικόνα 11: <http://ikee.lib.auth.gr/record/297900/files/GRI-2018-21573.pdf>

Εικόνα 13: <https://www.csc.com.gr/machine-learning-%CE%BC%CE%B7%CF%87%CE%B1%CE%BD%CE%B9%CE%BA%CE%AE-%CE%BC%CE%AC%CE%B8%CE%B7%CF%83%CE%B7-%CF%84%CE%B9-%CE%B5%CE%AF%CE%BD%CE%B1%CE%B9/>

Εικόνας 15: <https://why-change.com/2021/11/13/how-to-create-decision-trees-for-business-rules-analysis/>

Εικόνα 16: <https://www.semanticscholar.org/paper/K-Nearest-Neighbor-for-Uncertain-Data-Agrawal/dbef97e1b4d834c69050f552523c2aac0aa31162>

Εικόνα 17: <https://techvidvan.com/tutorials/reinforcement-learning/>

ΠΑΡΑΡΤΗΜΑΤΑ

Spyder:

```
from pandas import read_csv

import numpy as np

import pandas as pd

import matplotlib.pyplot as plt

dataset = read_csv('household_power_consumption.csv',
infer_datetime_format=True, parse_dates=['datetime'], index_col=['datetime'],
usecols=['datetime','Global_active_power','Global_reactive_power','Voltage','Global
_intensity','Sub_metering_3'])

df_miss = read_csv('household_power_consumption.csv',
infer_datetime_format=True, parse_dates=['datetime'], index_col=['datetime'],
usecols=['datetime','Global_active_power','Voltage'])

daily_groups = dataset.resample('D')

daily_data = daily_groups.sum()

miss_value = df_miss.isnull()

miss_value.to_csv('power_consumption_find_nan_values.csv')

rows_miss = read_csv('power_consumption_find_nan_values.csv',
infer_datetime_format=True, parse_dates=['datetime'], index_col=['datetime'])

a = rows_miss.loc[rows_miss.Global_active_power == True]

a.to_csv('power_consumption_choice.csv')

average_value = read_csv('household_power_consumption.csv',
infer_datetime_format=True, parse_dates=['datetime'],
usecols=['datetime','Global_active_power','Voltage'])

power = average_value['Global_active_power'].to_list()

voltage = average_value['Voltage'].to_list()

times = pd.to_datetime(average_value['datetime']).tolist()

def my_function(arg1):

    e = []

    d = pd.date_range('2006-12-16 17:39:00','2010-11-26 21:02:00',freq='15min')

    for x in d:

        e.append(x)
```



```

    return e

def avg(data_1,step,thresh):
    datasum = cnt = valid_count = 0

    c = []

    for num in data_1:
        #for num in data:
            cnt+=1

            if num != None:
                datasum+=num

                valid_count+=1

            if cnt == step:
                #start,end = '2006-12-16 17:39:00', '2010-11-26 21:02:00'

                if valid_count < thresh:
                    c.append(None)

                else:
                    c.append(datasum / valid_count)

                datasum = cnt = valid_count = 0

    return c

lists =
{'datetime':list(my_function(times)), 'Global_active_power':list(avg(power,15,10)), 'V
oltage':list(avg(voltage,15,10))}

df = pd.DataFrame(lists)
df.to_csv('average.csv',index=True)

w = read_csv('average.csv')
plt.subplot(211)
plot_1 = plt.plot(lists['datetime'],w['Global_active_power'])
plt.subplot(212)

```

```
plot_2 = plt.plot(lists['datetime'],w['Voltage'])
```

Jupyter:

```
import pandas as pd
```

```
import numpy as np
```

```
df = pd.read_csv("average.csv",index_col=0,parse_dates=True)
```

```
df.index.freq = '15T'
```

```
np.where(df['Voltage'].isnull() == True)[0]
```

```
df.isna().sum()
```

```
count = 0
```

```
for x in range(len(df)):
```

```
    #print(x)
```

```
    #print(pd.isnull(df.iloc[x,:]))
```

```
    if pd.isnull(df.iloc[x,1]) == True:
```

```
        print(x)
```

```
        for y in range(len(df)-x):
```

```
            count += 1
```

```
            #print(count)
```

```
            if pd.isnull(df.iloc[x+count,1]) == False:
```

```
                print(count)
```

```
                break
```

```
    if count==1:
```

```
        #print(x)
```

```
        #print(df.iloc[x,0])
```

```
        df.iloc[x,1] = (df.iloc[x-1,1] + df.iloc[x+1,1])/2
```

```
        df.iloc[x,2] = (df.iloc[x-1,2] + df.iloc[x+1,2])/2
```

```
        print(df.iloc[x,1])
```

```
        print(df.iloc[x,2])
```

```
    elif count>1:
```

```
step = (df.iloc[x+count,1] - df.iloc[x-1,1])/(count+1)
#print(df.iloc[x-1,1])
#print(df.iloc[x+count,1])
print(step)
for z in range(count):
    df.iloc[x+z,1] = df.iloc[x-1,1] + ((z+1) * step)
    df.iloc[x+z,2] = df.iloc[x-1,2] + ((z+1) * step)
    #print(z)
    #print(df.iloc[x+z,1])
    #print(df.iloc[x+z,2])

count=0
df.to_csv('fulfill_nan_value',index=True)
```