



Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών
και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρικής Ισχύος

Εντοπισμός Ρευματοκλοπών με Μηχανική Μάθηση

Διπλωματική Εργασία

Μητσέλος Αθανάσιος

Περίληψη

Λέξεις Κλειδιά

Abstract

Keywords

Ευχαριστίες

Περιεχόμενα

1	Εισαγωγή	11
1.1	Κίνητρο και υπόβαθρο διπλωματικής	11
1.2	Δομή Διπλωματικής	11
2	Θεωρητικό υπόβαθρο	12
2.1	Έξυπνοι μετρητές	12
2.2	Μηχανική μάθηση	12
2.2.1	Επιβλεπόμενη μάθηση	12
2.2.2	Ημι-επιβλεπόμενη μάθηση	12
2.2.3	Μη επιβλεπόμενη μάθηση	12
2.3	Μετρικές μηχανικής μάθησης	12
3	Περιγραφή και οργάνωση δεδομένων	13
3.1	Περιγραφή δεδομένων	13
3.1.1	Ανάλυση καταναλώσεων	13
3.1.2	Μοντελοποίηση εποχιακών δεικτών	13
3.2	Προεπεξεργασία και καθάρισμα δεδομένων	22
3.3	Προσομοίωση απάτης	22
3.3.1	Τύποι απάτης	22
4	Αλγόριθμοι επιβλεπόμενης μάθησης	23
4.1	Θεωρία Λογιστικής Παλινδρόμησης	23
4.2	Δοκιμή ταξινόμησης με λογιστική παλινδρόμηση	23
4.3	Θεωρία Μηχανών Διανυσμάτων Υποστήριξης	23
4.4	Δοκιμή ταξινόμησης με Μηχανές Διανυσμάτων Υποστήριξης	23
4.4.1	Δοκιμή χρονοσειρών χωρίς πυρήνα	23
4.4.2	Δοκιμή χαρακτηριστικών με πυρήνα RBF	23
4.5	Σχόλια	23
5	Αλγόριθμοι μη επιβλεπόμενης μάθησης	24
5.1	Θεωρία αλγορίθμων συσταδοποίησης	24
5.2	Δοκιμή αλγορίθμου μη επιβλεπόμενης μάθησης	24
5.3	Θεωρία αλγορίθμου μείωσης διάστασης	24
5.4	Θεωρία αλγορίθμου εύρεσης ανωμαλιών	24
5.5	Δοκιμή αλγορίθμου ημι επιβλεπόμενης μάθησης	24
5.6	Σχόλια	24
6	Δυσκολίες και μελλονική κατεύθυνση	25
6.1	Τεχνικά εμπόδια	25
6.1.1	Δυσκολία γενίκευσης σε άλλες καταναλωτικές συνήθειες	25
6.1.2	Δυσκολία επιλογής μετρικών	25
6.1.3	Εύρεση αξιόπιστων δυαδικών χαρακτηρισμών	25

6.1.4	Ανατροφοδότηση ελέγχων	25
6.2	Ασφάλεια Καταναλωτών	25
6.2.1	Ασφάλεια Μετρητών	25
6.2.2	Απειλή ιδιωτικότητας	25
7	Συμπεράσματα	26
7.1	Σύγκριση αποτελεσμάτων	26
7.2	Συμπερασματικές σημειώσεις	26
A'	Συχνές ερωτήσεις	27

Κατάλογος σχημάτων

3.1	Εφαρμογή πολυωνύμου δευτέρου βαθμού	14
3.2	Εβδομαδιαία εποχιακότητα ομάδας 1	16
3.3	Εβδομαδιαία εποχιακότητα ομάδας 2	16
3.4	Εβδομαδιαία εποχιακότητα ομάδας 3	17
3.5	Εβδομαδιαία εποχιακότητα ομάδας 4	17
3.6	Μηνιαία εποχιακότητα	18
3.7	Κατανάλωση χωρίς εποχιακούς δείκτες ανά εβδομάδα	19
3.8	Κατανάλωση χωρίς εποχιακούς δείκτες ανά μήνα	20
3.9	Εκτίμηση ακανόνιστης συνιστώσας με εβδομαδιαία εποχιακότητα	21
3.10	Εκτίμηση ακανόνιστης συνιστώσας με μηνιαία εποχιακότητα	21

Κατάλογος πινάκων

3.1 Έλεγχος συσταδοποίησης Σαββάτου	22
---	----

Κεφάλαιο 1

Εισαγωγή

1.1 Κίνητρο και υπόβαθρο διπλωματικής

Ο Παγκόσμιος Ιστός αποτελεί χώρο διακίνησης τεράστιου όγκου πληροφοριών. Ωστόσο, η συντριπτική πλειοψηφία των πληροφοριών του Ιστού, είναι προσανατολισμένη προς τον άνθρωπο-χρήστη και δεν είναι κατανοητή από τις εφαρμογές. Για να αξιοποιηθεί λοιπόν η διαθέσιμη πληροφορία και να γίνει πιο εύκολη η ανταλλαγή και η επεξεργασία της, ο Παγκόσμιος Ιστός εξελίσσεται στο Σημασιολογικό Ιστό.

1.2 Δομή Διπλωματικής

Κεφάλαιο 2

Θεωρητικό υπόβαθρο

2.1 Έξυπνοι μετρητές

2.2 Μηχανική μάθηση

2.2.1 Επιβλεπόμενη μάθηση

2.2.2 Ημι-επιβλεπόμενη μάθηση

2.2.3 Μη επιβλεπόμενη μάθηση

2.3 Μετρικές μηχανικής μάθησης

Κεφάλαιο 3

Περιγραφή και οργάνωση δεδομένων

3.1 Περιγραφή δεδομένων

3.1.1 Ανάλυση καταναλώσεων

3.1.2 Μοντελοποίηση εποχιακών δεικτών

Για βαθύτερη κατανόηση των χρονοσειρών γίνεται εκτίμηση της εποχιακής και μη εποχιακής καταναλωτικής τάσης με τη χρήση παραμετρικών μοντέλων. Με αυτό τον τρόπο θα καταστεί δυνατή η παρατήρηση της επαναληψιμότητας και των μορφών των καταναλώσεων. Για να γίνει αυτό χρησιμοποιείται αρχικά ο αλγόριθμος K-Means για την ομαδοποίηση των καταναλωτών σε τέσσερις συστάδες βάση του ετήσιου μέσου όρου καθενός. Στη συνέχεια δημιουργείται ένα προφίλ κατανάλωσης για κάθε συστάδα βρίσκοντας το μέσο ημερήσιο όρο κατανάλωσης. Χρειάστηκαν 2000 καταναλωτές για αυτή την ανάλυση με περισσότερους 1800 να ομαδοποιούνται σε δύο ομάδες υποδεικνύοντας προφίλ οικιακών καταναλωτών.

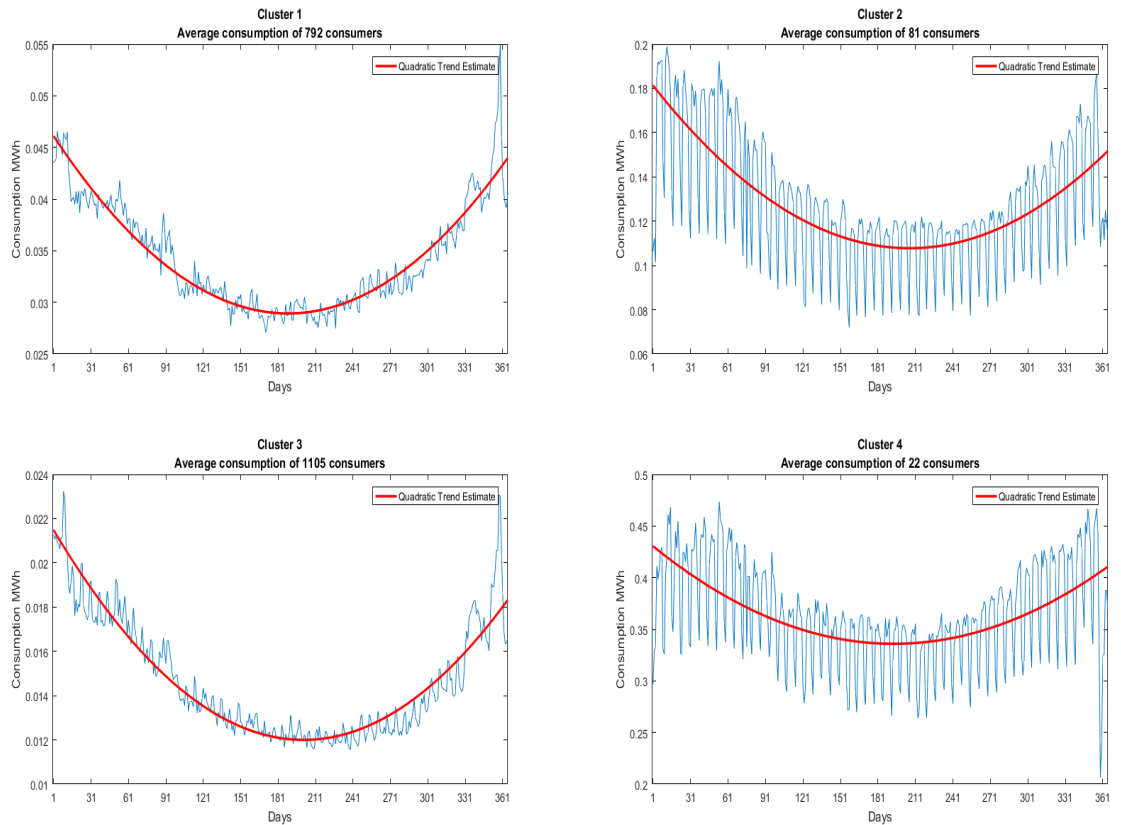
Ανάλυση Παλινδρόμησης

Σκοπός, λοιπόν αυτού του μέρους είναι να γίνει στατιστική μελέτη του πολυωνυμικού μοντέλου στα δεδομένα μας και να δούμε αν οι χρονοσειρές κάθε συστάδας μπορούν να περιγραφούν με πολυώνυμο δευτέρου βαθμού. [3]

$$T_t = \beta_0 + \beta_1 t + \beta_2 t^2$$

Όπως φαίνεται στο Σχήμα 3.1 οι συστάδες μπορούν να χαρακτηριστούν από μια παραβολική καμπύλη με θετικό συντελεστή μεγιστοβάθμιου όρου.

- Η συστάδα 1 αποτελείται από 792 καταναλωτές και έχει η παραβολική καμπύλη τάσης λαμβάνει ελάχιστη τιμή την 189η μέρα του έτους.
- Η συστάδα 2 αποτελείται από 81 καταναλωτές και έχει η παραβολική καμπύλη τάσης λαμβάνει ελάχιστη τιμή την 206η μέρα του έτους.
- Η συστάδα 3 αποτελείται από 81 καταναλωτές και έχει η παραβολική καμπύλη τάσης λαμβάνει ελάχιστη τιμή την 201η μέρα του έτους.
- Η συστάδα 4 αποτελείται από 81 καταναλωτές και έχει η παραβολική καμπύλη τάσης λαμβάνει ελάχιστη τιμή την 194η μέρα του έτους.



Σχήμα 3.1: Εφαρμογή πολυωνύμου δευτέρου βαθμού

Εύκολα, λοιπόν, βγάνει το συμπέρασμα πως οι οικιακοί καταναλωτές έχουν την τάση να έχουν πιο ομοιόμορφα κατανεμημένα την παραβολική καμπύλη, ενώ οι επιχειρήσεις έχουν μεγαλύτερο βαθμό τυχαιότητας και λιγότερο συμμετρική καμπύλη ως προς το ελάχιστο σημείο της.

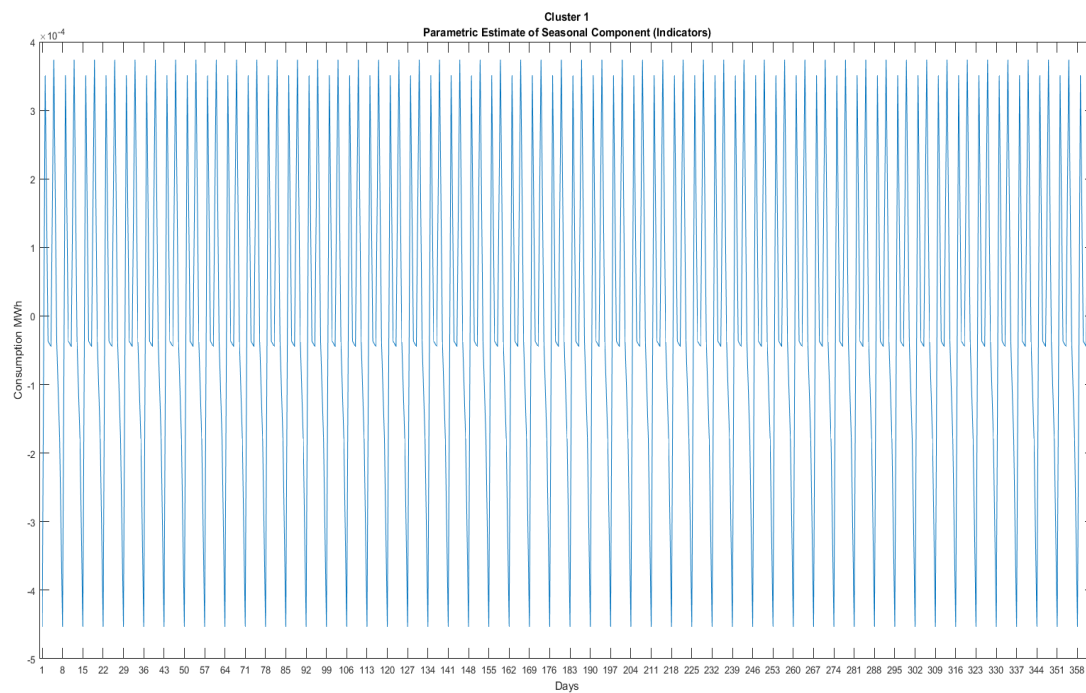
Εκτίμηση εποχιακών δεικτών

Αρχικά για την εκτίμηση των εποχιακών δεικτών απαιτείται η αφαίρεση του πολυωνύμου δευτέρου βαθμού από τις χρονοσειρές των ομάδων.[2] Δεδομένης της μικρής διάρκειας των καταναλώσεων (1 έτος) καθίσταται αδύνατη η εξαγωγή εποχιακών δεικτών ανά μήνα έτους ή ανά εποχή έτους. Για αυτό το λόγο οι εποχιακοί δείκτες μεταφέρθηκαν ανά ημέρα της εβδομάδας ή ανά ημέρα του μήνα. Για την πρώτη περίπτωση οι δείκτες αναφέρονται στις ημέρες κάθε εβδομάδας, ενώ για την δεύτερη αναφέρονται στις ημέρες κάθε μήνα δημιουργώντας 7 ή 30 δείκτες αντίστοιχα. Για την εβδομαδιαία εποχιακότητα έχω τις παρακάτω καμπύλες για κάθε ομάδα.

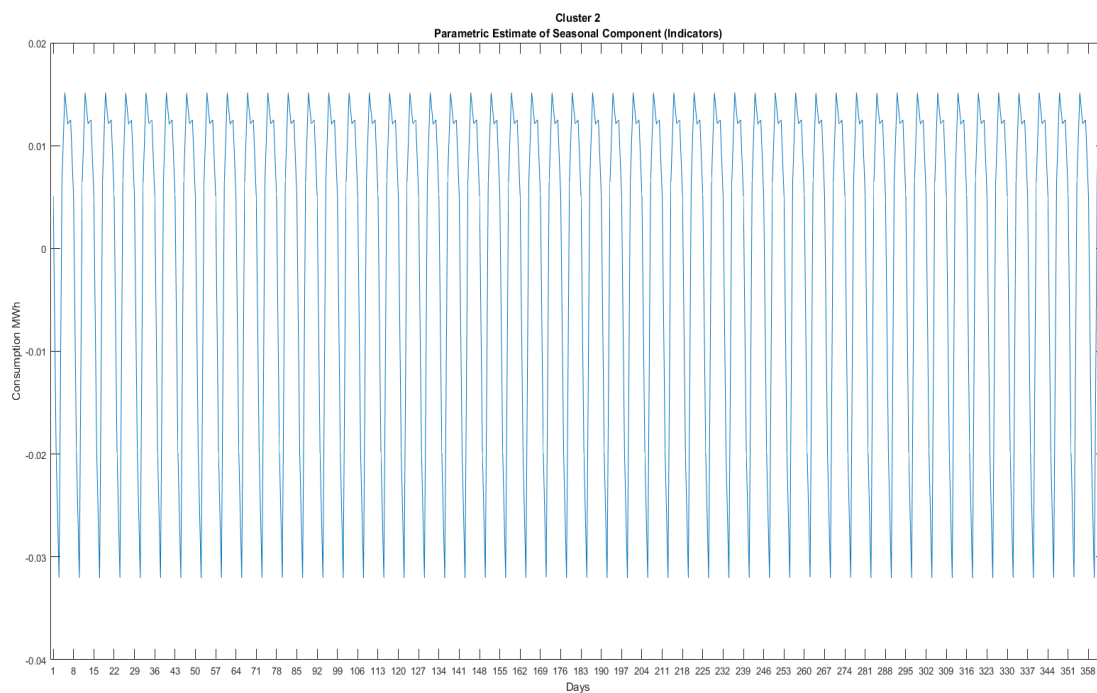
Εκτίμηση με διαστήματα ημέρας ανά εβδομάδα

Από την εβδομαδιαία εποχιακότητα λοιπόν εύκολα κάποιος αντιλαμβάνεται πως ανάλογα με τον τύπο των καταναλωτών οι μέρες που έχουμε μέγιστη και ελάχιστη κατανάλωση διαφέρουν ριζικά. Η πρώτη μέρα του έτους για το έτος που μελετάμε είναι Πέμπτη. Ειδικότερα:

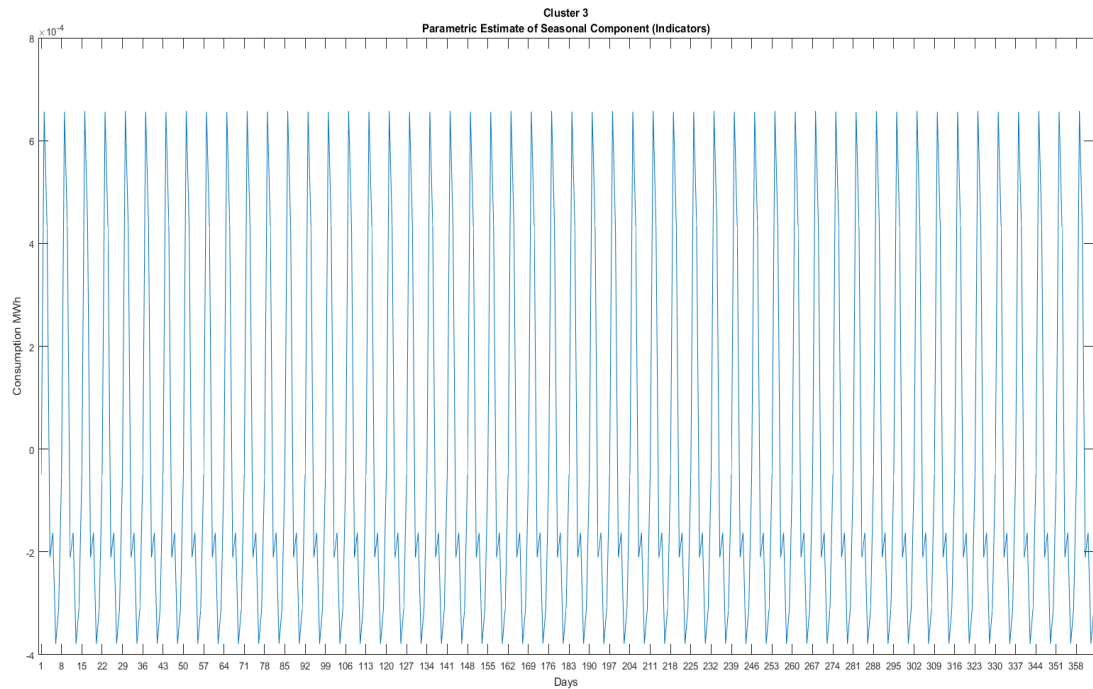
- Για τους καταναλωτές συστάδας 1 (οικιακοί καταναλωτές) έχουμε ελάχιστες καταναλώσεις τις Πέμπτες.
- Για τους καταναλωτές συστάδας 2 (επιχειρήσεις) έχουμε ελάχιστες καταναλώσεις τα Σάββατα.
- Για τους καταναλωτές συστάδας 3 (οικιακοί καταναλωτές) έχουμε ελάχιστες καταναλώσεις τις Τρίτες.
- Για τους καταναλωτές συστάδας 4 (επιχειρήσεις) έχουμε ελάχιστες καταναλώσεις τα Σάββατα.



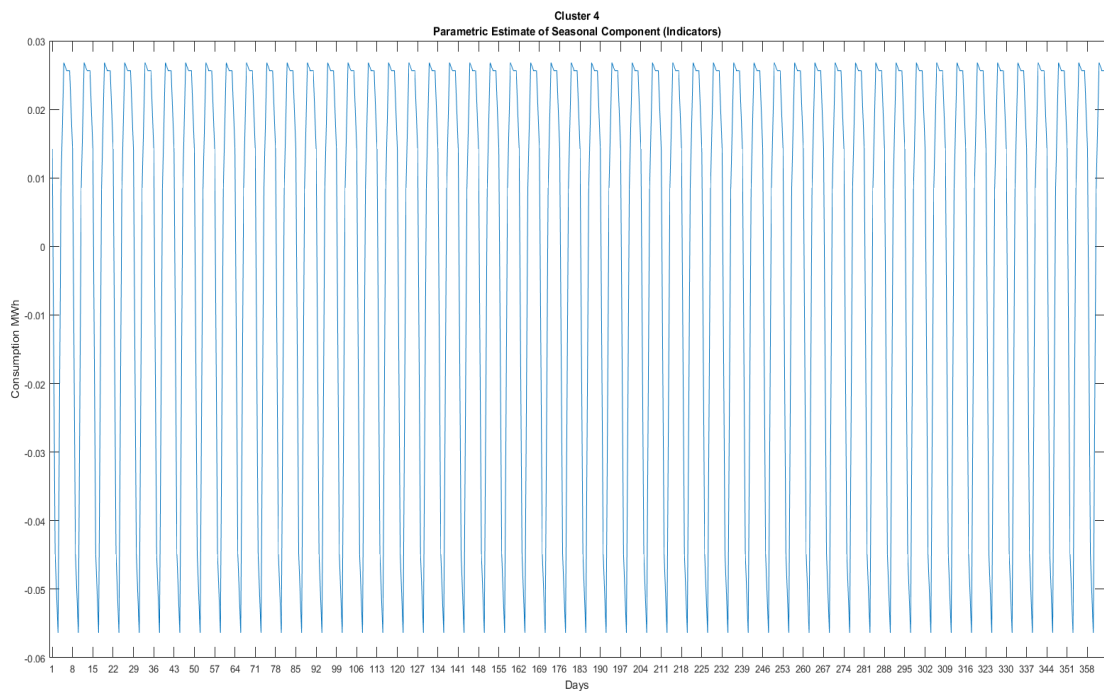
Σχήμα 3.2: Εβδομαδιαία εποχιακότητα ομάδας 1



Σχήμα 3.3: Εβδομαδιαία εποχιακότητα ομάδας 2



Σχήμα 3.4: Εβδομαδιαία εποχιακότητα ομάδας 3



Σχήμα 3.5: Εβδομαδιαία εποχιακότητα ομάδας 4

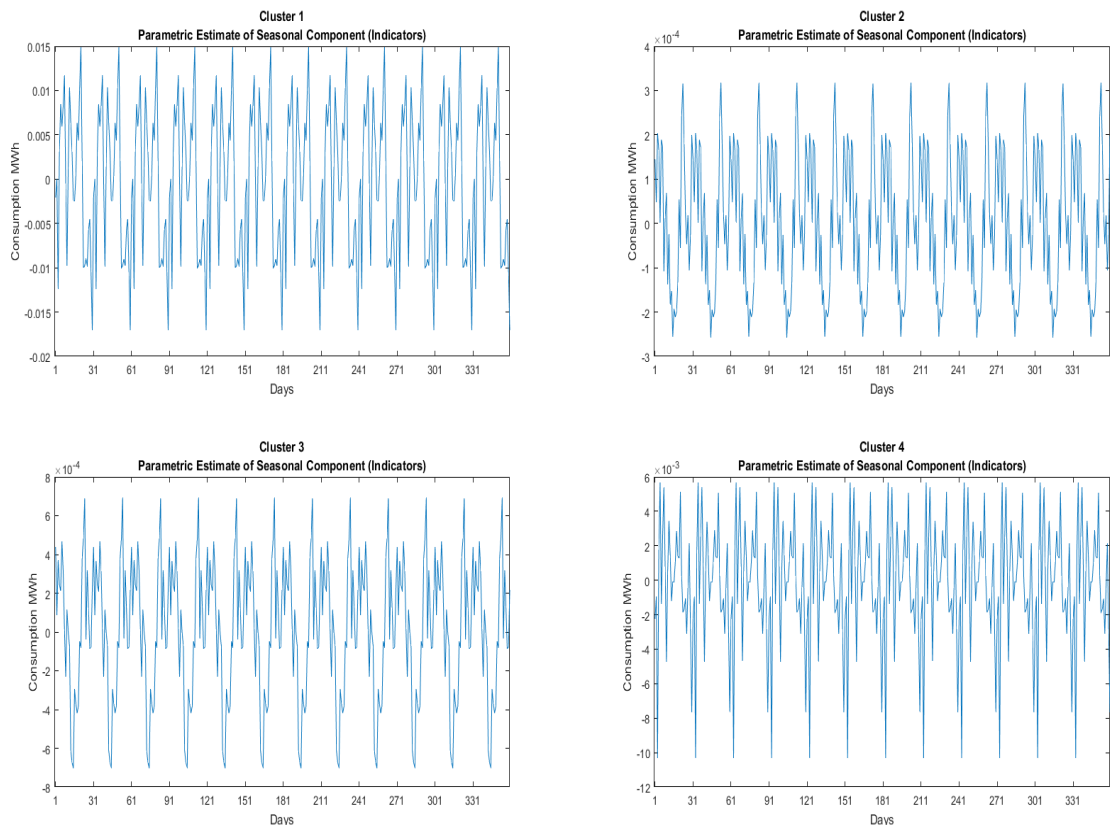
Εκτίμηση σε διαστήματα ημέρας ανά μήνα

Το διάστημα ενός μήνα αφήνει μεγαλύτερα περιθώριο εποπτείας της χρονοσειράς, ενώ ταυτόχρονα δημιουργεί αποτελέσματα με μεγαλύτερη συνοχή. Από την άλλη

πλευρά οι 12 μήνες του έτους δεν μπορούν να εξάγουν πολύ ασφαλή δεδομένα αν συγκριθούν με τις 52 εβδομάδες.

Από την μηνιαία εποχιακότητα γίνεται εύκολα αντιληπτό πως ανάλογα με τον τύπο των καταναλωτών οι μέρες που έχουμε μέγιστη και ελάχιστη κατανάλωση διαφέρουν ριζικά. Ειδικότερα:

- Για τους καταναλωτές συστάδας 1 (επιχειρήσεις) έχουμε ελάχιστες καταναλώσεις στις 30 του μηνός.
- Για τους καταναλωτές συστάδας 2 (οικιακοί καταναλωτές) έχουμε ελάχιστες καταναλώσεις στις 15 του μηνός.
- Για τους καταναλωτές συστάδας 3 (οικιακοί καταναλωτές) έχουμε ελάχιστες καταναλώσεις στις 15 του μηνός.
- Για τους καταναλωτές συστάδας 4 (επιχειρήσεις) έχουμε ελάχιστες καταναλώσεις στις 3 του μηνός.

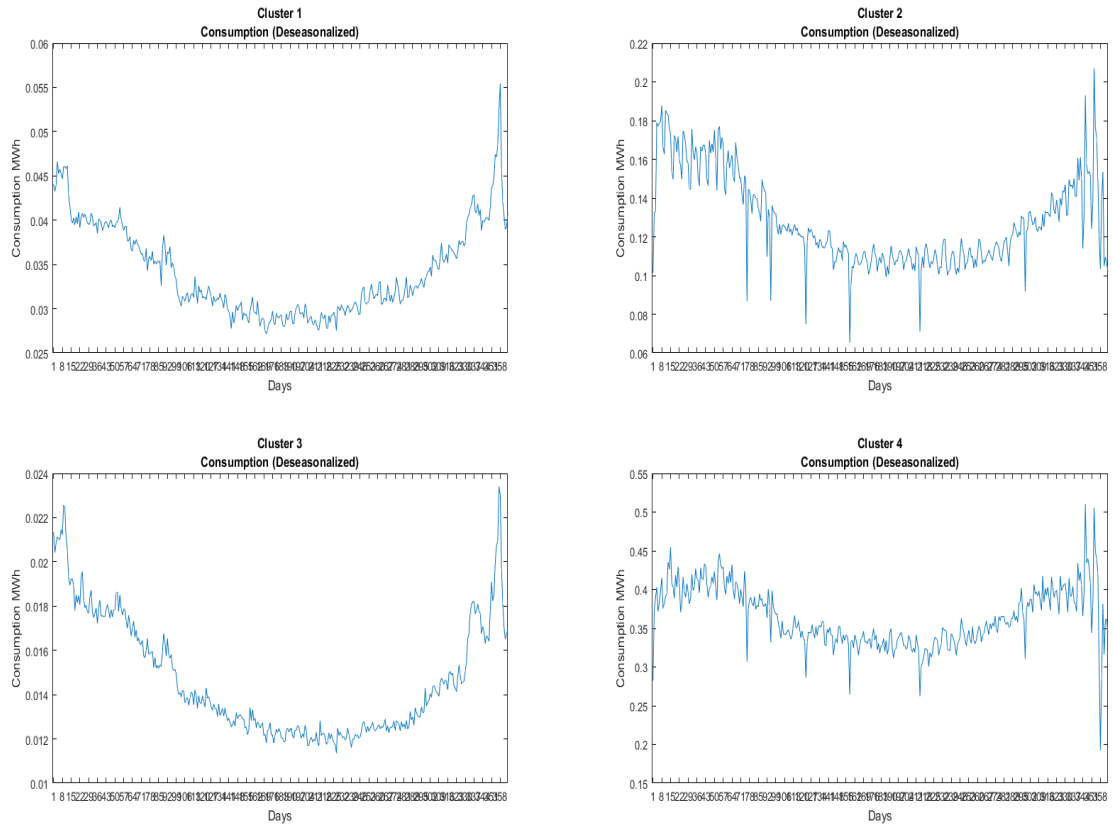


Σχήμα 3.6: Μηνιαία εποχιακότητα

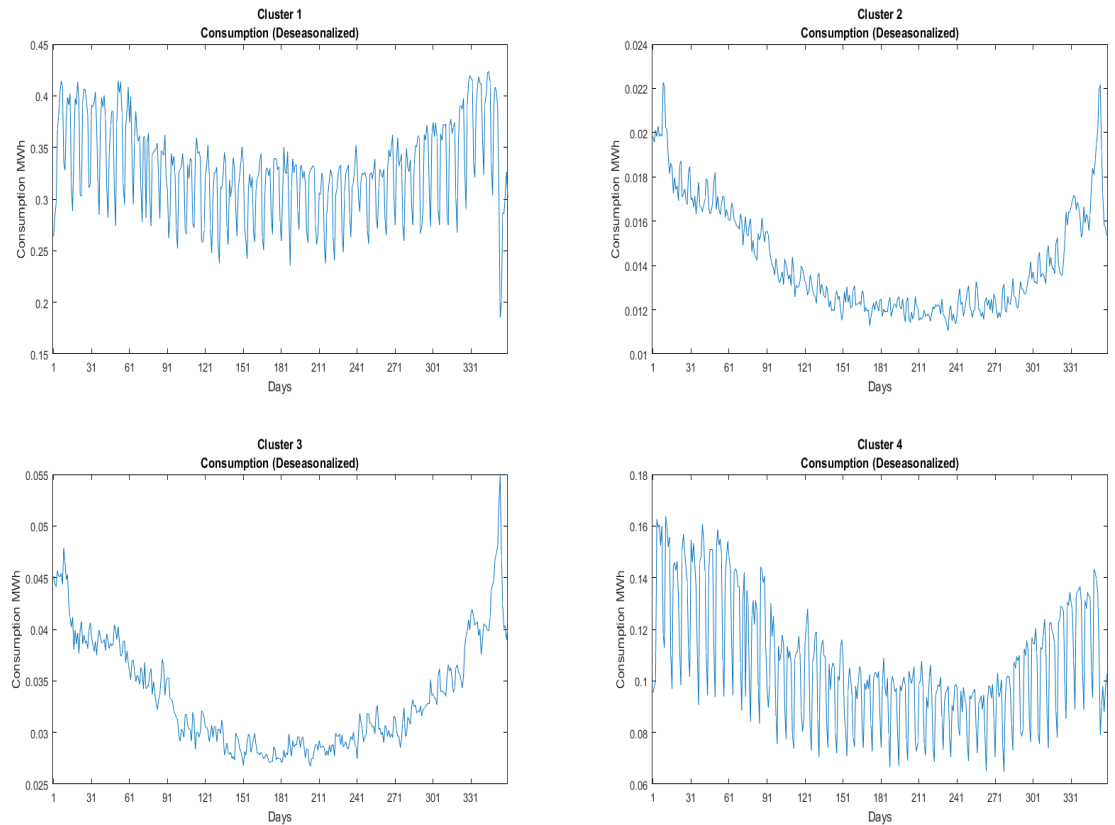
Αφαίρεση εποχιακών δεικτών

Σε αυτό το σημείο είναι σημαντικό να παρατηρηθεί η κατανάλωση χωρίς τους εποχιακούς δείκτες. Με αυτό τον τρόπο καθίσταται ευκολότερη η θεώρηση της μορφής των κυματομορφών και η σύγκρισή τους με τις αρχικές καταναλώσεις του πρώτου μέρους. Αφαιρώντας τα εποχιακά χαρακτηριστικά οι καμπύλες πλησιάζουν περισσότερο

στην παραβολική συνάρτηση. Έτσι η καταναλωτική τους τάση χωρίς τους εποχιακούς δείκτες γίνεται πιο έντονη και ευδιάκριτη.



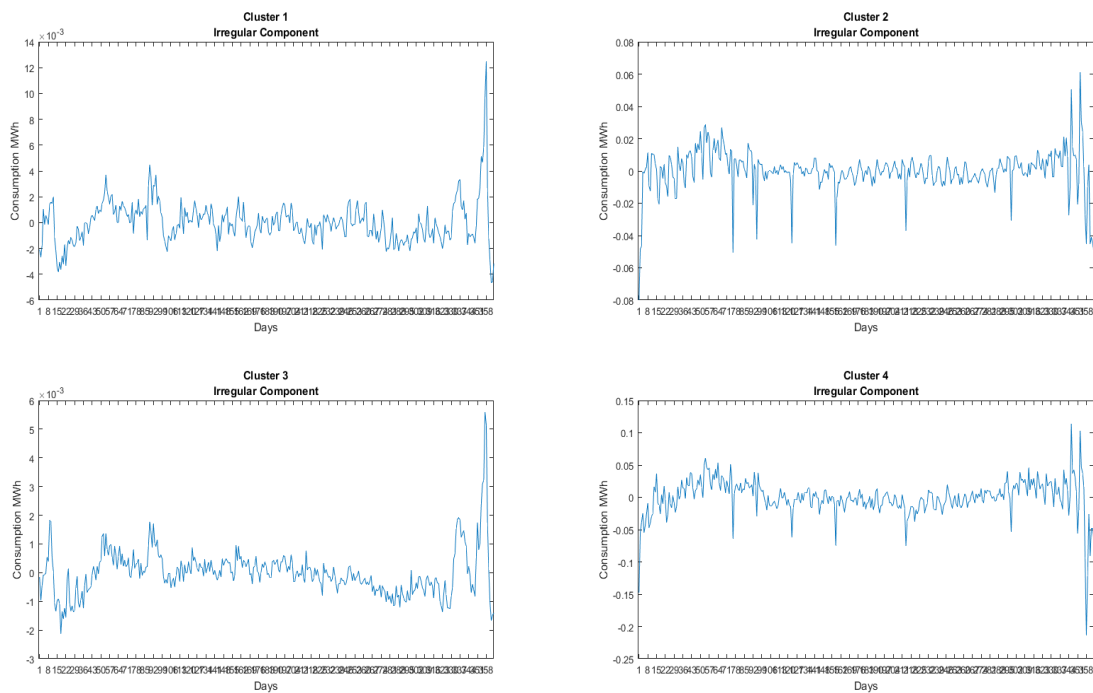
Σχήμα 3.7: Κατανάλωση χωρίς εποχιακούς δείκτες ανά εβδομάδα



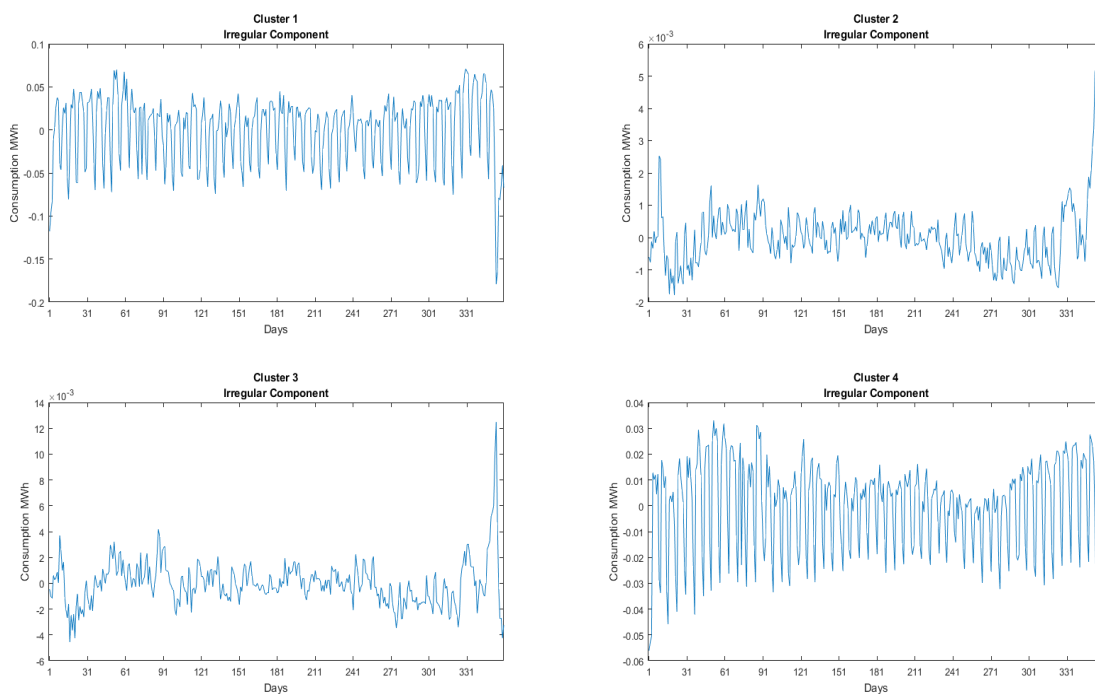
Σχήμα 3.8: Κατανάλωση χωρίς εποχιακούς δείκτες ανά μήνα

Εκτίμηση ακανόνιστης συνιστώσας

Τέλος έχει ενδιαφέρον να δούμε το βαθμό της τυχαιότητας που έχουμε στις καταναλώσεις των συστάδων που δημιουργήθηκαν. Αυτό επιτυγχάνεται αφαιρώντας την εποχιακή χρονοσειρά και την καταναλωτική τάση της αρχικής χρονοσειράς. Με αυτό τον τρόπο γίνεται σαφές ότι παρόλο την εποχιακότητα και την τάση οι χρονοσειρές έχουν αισθητό τυχαίο παράγοντα. Η αφαίρεση δημιουργεί αλλαγές στο επίπεδο της χρονοσειράς, σταθεροποιώντας έτσι το μέσο όρο της. Γίνεται αντιληπτό πως έχουν μη προβλέψιμα πρότυπα τουλάχιστον με δεδομένα διάρκειας ενός έτους. Τέτοιου τύπου δεδομένα λέγονται στατικές χρονοσειρές.[1]



Σχήμα 3.9: Εκτίμηση ακανόνιστης συνιστώσας με εβδομαδιαία εποχιακότητα



Σχήμα 3.10: Εκτίμηση ακανόνιστης συνιστώσας με μηνιαία εποχιακότητα

Εξερεύνηση ημερών με χαμηλές καταναλώσεις

Για να αντληθούν περαιτέρω χαρακτηριστικά των χρονοσειρών χρειάστηκε η υλοποίηση αλγορίθμου με διπλή συσταδοποίηση. Σύμφωνα με τον αλγόριθμο πρώτα συσταδοποιούνται οι καταναλωτές με βάση την ημερήσια κατανάλωση, εν συνεχεία για κάθε συστάδα δημιουργείται νέα ομαδοποίηση με βάση την ομοιότητα κάθε ημερήσιας κατανάλωσης. Με αυτό τον τρόπο μπορεί να παρατηρηθεί ποιες μέρες όμοιων καταναλωτών έχουν παρόμοιες καταναλώσεις. Καθίσταται έτσι εφικτό, να φιλτράρουμε από τα δεδομένα μας μέρες με χαμηλή κατανάλωση που γνωρίζουμε πως θα δυσκόλευαν το πρόβλημα της ταξινόμησης σε αληθή και αλλοιωμένα δεδομένα.

Τα αποτελέσματα του αλγορίθμου έδειξαν πως μόνο τα Σάββατα μιας συστάδας εμφανίζουν έντονη ομοιότητα οικιακών καταναλώσεων. Οι Κυριακές κατά κύριο λόγο συσταδοποιούνται με την υπόλοιπη εβδομάδα δημιουργώντας την εβδομαδιαία τάση, γεγονός που δείχνει πως για τους περισσότερους καταναλωτές η Κυριακή είναι εργάσιμη ημέρα. Παράλληλα, παρατηρείται πως ανά περιόδους οι καταναλώσεις δημιουργούν νέες συστάδες αφήνοντας μόνο τα Σάββατα να σπάνε την συνεχόμενη συσταδοποίηση. Στον Πίνακα 3.1 φαίνεται πως ακόμη και στα Σάββατα δεν έχουμε απολύτως γεμάτες συστάδες.

Συστάδες Καταναλωτών				
Συστάδες Σαββάτου	Συστάδα 1	Συστάδα 2	Συστάδα 3	Συστάδα 4
Συστάδα 1	0	24	30	19
Συστάδα 2	9	11	0	15
Συστάδα 3	0	9	0	0
Συστάδα 4	42	0	0	0
Συστάδα 5	0	2	0	0
Συστάδα 6	0	4	0	7
Συστάδα 7	0	1	21	10

Πίνακας 3.1: Έλεγχος συσταδοποίησης Σαββάτου

Παρατηρήσεις

Τα εμφανή χαρακτηριστικά εποχιακότητας και η εφαρμογή πολυωνύμου δευτέρου βαθμού στις χρονοσειρές θέτει καλό υποψήφιο τα μοντέλα πρόβλεψης χρονοσειρών. Με ένα τέτοιο σύστημα θα δημιουργείται μια πρόβλεψη κατανάλωσης από έμπιστους καταναλωτές για κάποιο χρονικό διάστημα. Εν συνεχεία θα αλλοιώνονται τα χαρακτηριστικά κάποιου μέρους των καταναλωτών και θα ελέγχεται αν ο αλγόριθμος μπορεί να διαχωρίσει τις αλλοιωμένες τιμές από αυτές που προέβλεψε.

3.2 Προεπεξεργασία και καθάρισμα δεδομένων

3.3 Προσομοίωση απάτης

3.3.1 Τύποι απάτης

Κεφάλαιο 4

Αλγόριθμοι επιβλεπόμενης μάθησης

4.1 Θεωρία Λογιστικής Παλινδρόμησης

4.2 Δοκιμή ταξινόμησης με λογιστική παλινδρόμηση

4.3 Θεωρία Μηχανών Διανυσμάτων Υποστήριξης

4.4 Δοκιμή ταξινόμησης με Μηχανές Διανυσμάτων Υποστήριξης

4.4.1 Δοκιμή χρονοσειρών χωρίς πυρήνα

4.4.2 Δοκιμή χαρακτηριστικών με πυρήνα RBF

4.5 Σχόλια

Κεφάλαιο 5

Αλγόριθμοι μη επιβλεπόμενης μάθησης

- 5.1 Θεωρία αλγορίθμων συσταδοποίησης
- 5.2 Δοκιμή αλγορίθμου μη επιβλεπόμενης μάθησης
- 5.3 Θεωρία αλγορίθμου μείωσης διάστασης
- 5.4 Θεωρία αλγορίθμου εύρεσης ανωμαλιών
- 5.5 Δοκιμή αλγορίθμου ημι επιβλεπόμενης μάθησης
- 5.6 Σχόλια

Κεφάλαιο 6

Δυσκολίες και μελλονική κατεύθυνση

6.1 Τεχνικά εμπόδια

6.1.1 Δυσκολία γενίκευσης σε άλλες καταναλωτικές συνήθειες

6.1.2 Δυσκολία επιλογής μετρικών

6.1.3 Εύρεση αξιόπιστων δυαδικών χαρακτηρισμών

6.1.4 Ανατροφοδότηση ελέγχων

6.2 Ασφάλεια Καταναλωτών

6.2.1 Ασφάλεια Μετρητών

6.2.2 Απειλή ιδιωτικότητας

Κεφάλαιο 7

Συμπεράσματα

7.1 Σύγκριση αποτελεσμάτων

7.2 Συμπερασματικές σημειώσεις

Παράρτημα Α΄

Συχνές ερωτήσεις

Bibliography

- [1] Rob J. Hyndman and George Athanasopoulos. *Forecasting: principles and practice*. 2012. URL: <https://www.otexts.org/fpp>.
- [2] Gregory C. Reinsel George E. P. Box Gwilym M. Jenkins and Greta M. Ljung. *Time Series Analysis: Forecasting and Control*. New Jersey: Wiley, 2016.
- [3] Mathworks. *Parametric Trend Estimation*. 2017. URL: <https://www.mathworks.com/help/econ/parametric-trend-estimation.html>.