



# ΑΝΙΧΝΕΥΣΗ ΜΗ ΤΕΧΝΙΚΩΝ ΑΠΩΛΕΙΩΝ ΜΕ ΣΥΣΤΗΜΑΤΑ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

23 ΟΚΤΩΒΡΙΟΥ – ΜΗΤΣΕΛΟΣ ΑΘΑΝΑΣΙΟΣ



## ΡΕΥΜΑΤΟΚΛΟΠΕΣ ΣΤΗΝ ΕΛΛΑΔΑ

Διαφυγόντα έσοδα βάσει ΡΑΕ

Εταιρίες	Εκατομμύρια €
ΔΕΗ	120-125
ΥΚΩ	21
ΕΤΜΕΑΡ	32
ΑΔΜΗΕ	7,3
ΔΕΔΔΗΕ	26,5
Σύνολο	206,8-211,8

Τρόποι εξαπάτησης παρόχου

- Επέμβαση στο μετρητή
- Απευθείας σύνδεση
- Άτυπες ρυθμίσεις χρέωσης
- Απλήρωτοι λογαριασμοί

Πηγή: ΔΕΗ, Δελτίο τύπου 5/5/2017

# ΔΙΑΔΙΚΑΣΙΑ ΑΞΙΟΠΟΙΗΣΗΣ ΧΡΟΝΟΣΕΙΡΩΝ



An Coimisiún  
um Rialáil Fóntas

**Commission for  
Regulation of Utilities**

- 5.000 οικιακοί καταναλωτές
- 600 μικρομεσαίες επιχειρήσεις
- Διάρκεια συλλογής: 2009-2011

- Προεπεξεργασία δεδομένων έξυπνων μετρητών
  - Καθάρισμα δεδομένων
  - Επιλογή καταναλωτών
  - Επιλογή χρονικού πεδίου
- Προσομοίωση απάτης
  - Καθορισμός διαφορετικών τύπων απάτης
  - Επιλογή έντασης απάτης
  - Επιλογή ποσοστού προσομοίωσης καταναλωτών
- Ταξινόμηση καταναλωτών με μηχανική μάθηση
  - Επιλογή/Αξιολόγηση αλγορίθμων
  - Δημιουργία συστημάτων

# ΠΡΟΕΠΕΞΕΡΓΑΣΙΑ ΔΕΔΟΜΕΝΩΝ ΕΞΥΠΝΩΝ ΜΕΤΡΗΤΩΝ

Στιγμιότυπο αρχείου έξυπνων μετρητών

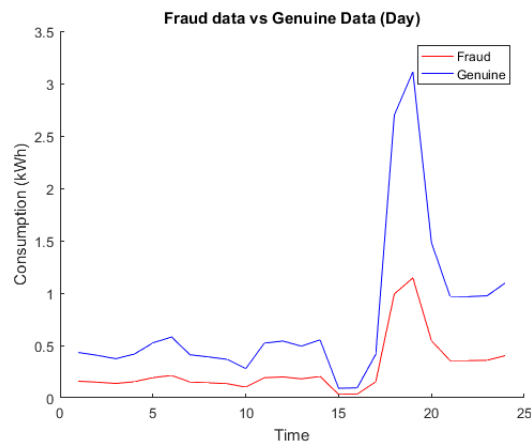
ID	ημέρα ώρα	kWh
1392	19503	0.140
1392	19504	0.138
...	...	...
1187	22028	1.367
1187	22029	1.425
1392	19940	0.234

Τα δεδομένα οργανώθηκαν ως εξής:

- 4.500 οικιακοί καταναλωτές και μικρομεσαίες επιχειρήσεις.
- Μετρήσεις για 1 έτος με μισάωρες μετρήσεις.

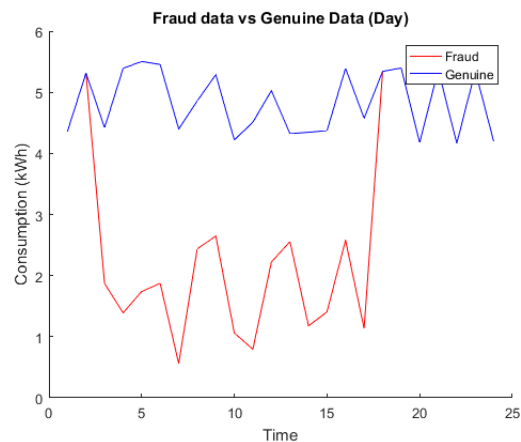
# ΠΡΟΣΟΜΟΙΩΣΗ ΑΠΑΤΗΣ

## Τύπου 1



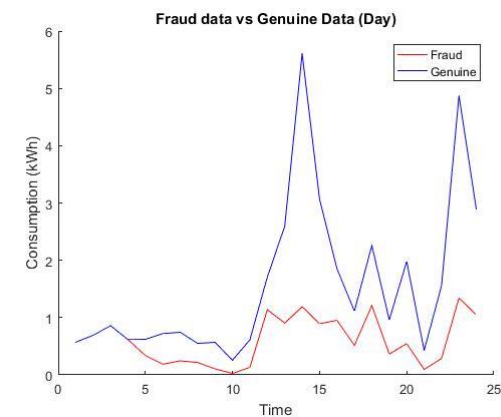
- Σταθερή ένταση σε όλο το έτος
- Συνεχόμενη απάτη

## Τύπου 2



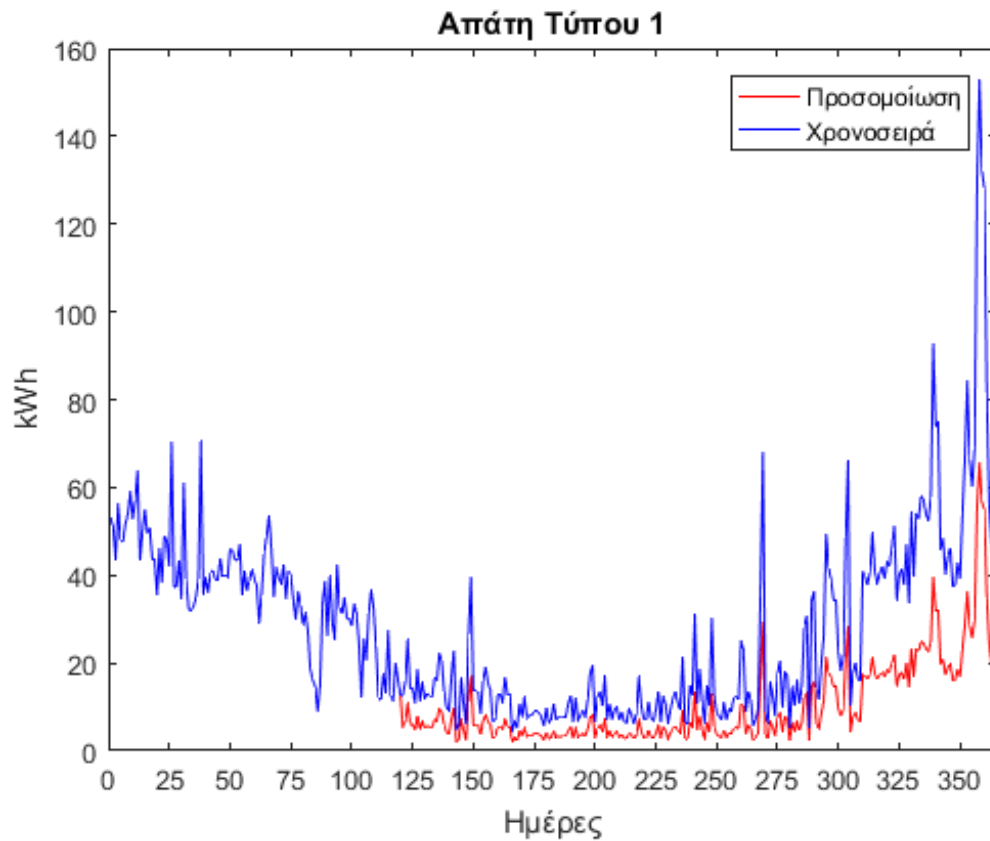
- Κάθε μέρα διαφορετική ένταση
- Διακοπτόμενη απάτη

## Τύπου 3

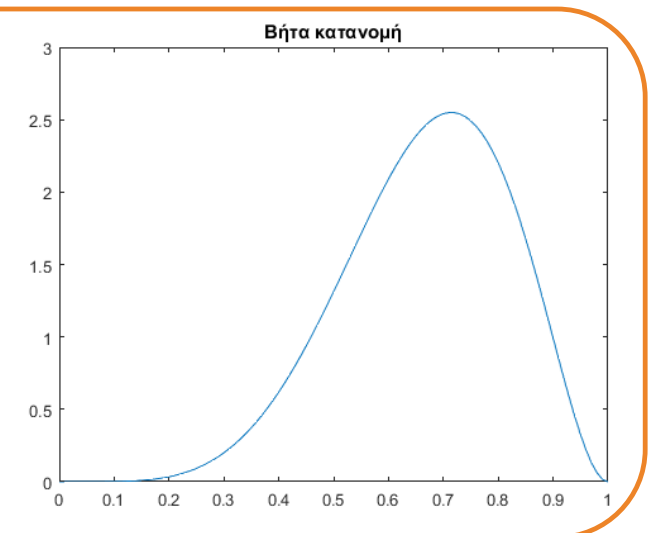
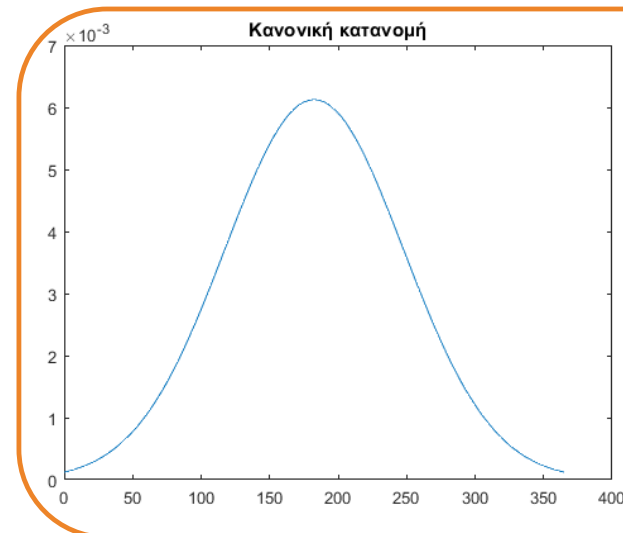


- Κάθε ώρα διαφορετική ένταση
- Διακοπτόμενη απάτη

# ΕΠΙΛΟΓΗ ΠΡΟΣΟΜΟΙΩΣΗΣ

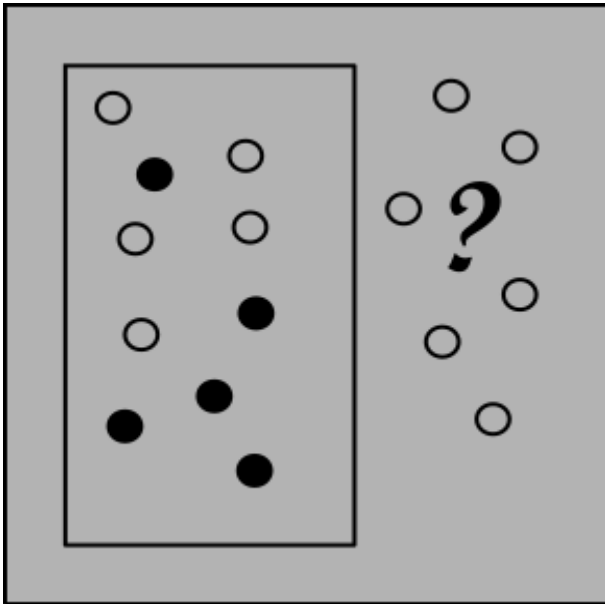


- Επιλογή ημέρας εκκίνησης κλοπής από κανονική κατανομή με  $\mu=182,5$  και  $\sigma=56,15$ .
- Επιλογή έντασης απάτης από βήτα κατανομή με  $\alpha=6$ ,  $\beta=3$ .
- Επιλογή ποσοστού καταναλωτών με μη τεχνικές απώλειες (10%).
- Τυχαία επιλογή καταναλωτών



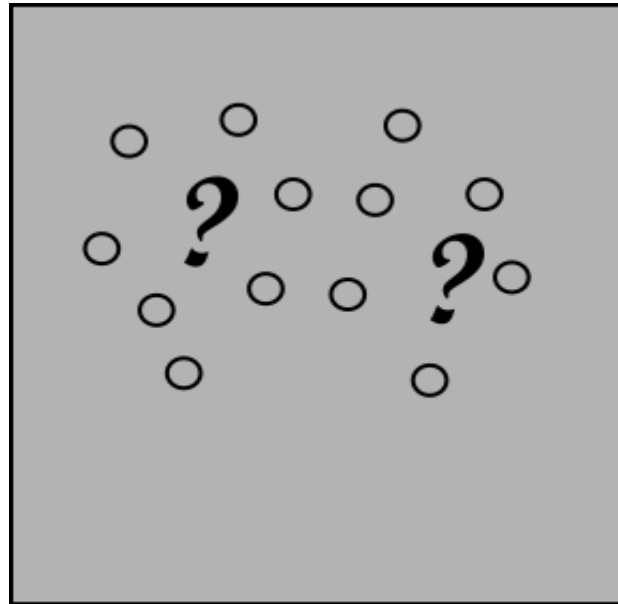
# ΤΑΞΙΝΟΜΗΣΗ ΚΑΤΑΝΑΛΩΤΩΝ ΜΕ ΑΛΓΟΡΙΘΜΟΥΣ/ΣΥΣΤΗΜΑΤΑ ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ

## Επιβλεπόμενη μάθηση



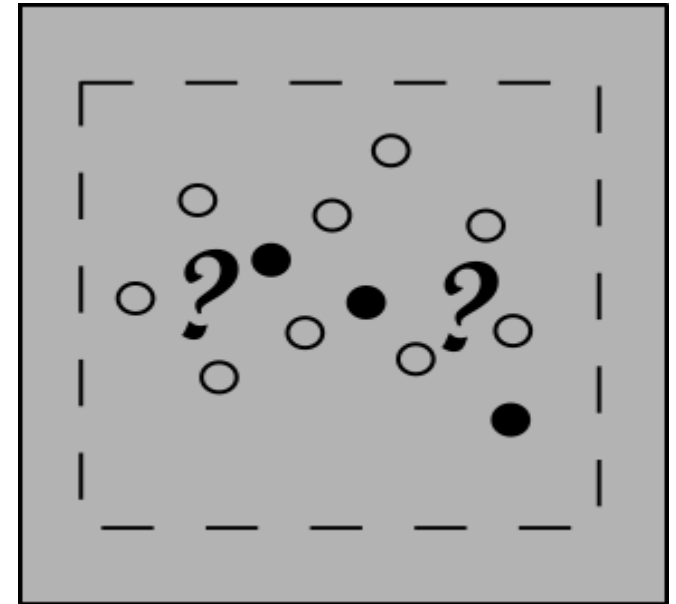
- Εκπαίδευση με 70% των labels και πρόβλεψη στο 30% των δεδομένων.

## Μη επιβλεπόμενη μάθηση



- Πρόβλεψη μέσω κανόνων χωρίς τη χρήση labels.

## Ημι-επιβλεπόμενη μάθηση



- Βελτιστοποίηση ταξινόμησης με 30% των labels πρόβλεψη στο 70% των δεδομένων.

# ΜΕΤΡΙΚΕΣ ΑΞΙΟΛΟΓΗΣΗΣ

		Predicted class	
		P	N
Actual Class	P	True Positives (TP)	False Negatives (FN)
	N	False Positives (FP)	True Negatives (TN)

- $DR = \frac{TP}{TP+FN}$

- $FPR = \frac{FP}{FP+TN}$

- $Accuracy = \frac{TP+TN}{TP+FP+FN+TN}$

- $F1\ score = \frac{2 \cdot Precision \cdot DR}{Precision+DR}$

- $BDR = \frac{P(I) \cdot DR}{P(I) \cdot DR + P(I') \cdot FPR}$

- $Precision = \frac{TP}{TP+FP}$

- $P(I)$ =πιθανότητα απάτης

- TP=πλήθος των σωστών προβλέψεων στο θετικό αποτέλεσμα
- TN=πλήθος των σωστών προβλέψεων στο αρνητικό αποτέλεσμα
- FN=πλήθος των λανθασμένων προβλέψεων στο θετικό αποτέλεσμα (αρνητική πρόβλεψη)
- FP=πλήθος των λανθασμένων προβλέψεων στο αρνητικό αποτέλεσμα (θετική πρόβλεψη)



# ΑΛΓΟΡΙΘΜΟΙ ΕΠΙΒΛΕΠΟΜΕΝΗΣ ΜΑΘΗΣΗΣ

## Δοκιμή επιβλεπόμενων ταξινομητών

- Αλγόριθμοι LIBLINEAR
  - Λογιστική Παλινδρόμηση
  - Γραμμικά SVM
  - SVM Crammer και Singer
- Αλγόριθμοι LIBSVM
  - SVM χωρίς πυρήνα
  - SVM με RBF πυρήνα

Λεπτομέρειες δοκιμής:

- 4.500 χρονοσειρές με ωριαίες μετρήσεις
- 10% καταναλωτών με μη τεχνικές απώλειες
- 70% εκπαίδευση 30% πρόβλεψη

## Επιβλεπόμενοι ταξινομητές (απάτες τύπου 1)

Αλγόριθμοι	DR	FPR	Acc.	FI	BDR
Λογιστική Παλινδρόμηση	75,65	1,38	96,67	79,45	86
Linear SVM (LIBLINEAR)	80	1,46	96,96	81,78	86
Crammer Singer SVM	81,74	1,94	96,67	80,69	82
Linear SVM (LIBSVM)	81,43	1,24	96,96	84,76	88

## Γραμμικό SVM σε όλες τις απάτες

Τύπος	DR	FPR	Acc.	FI	BDR
1	81,43	1,24	96,96	84,76	88
2	22,63	7,25	85,63	24,22	26
3	23,78	10,36	82,67	22,52	20
Μικτός	27,13	7,37	86,37	27,56	29

# ΕΞΑΓΩΓΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ

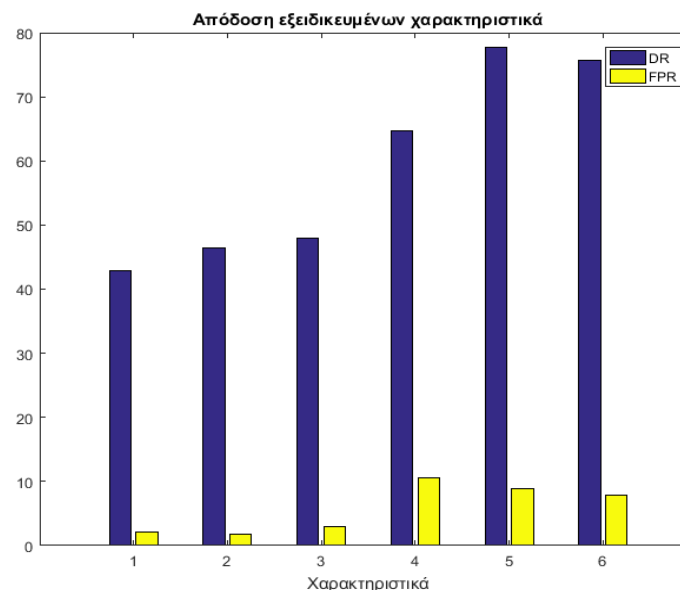
## Γενικά χαρακτηριστικά

1. Ετήσια μέση τιμή ωριαίας κατανάλωσης
2. Ετήσια τυπική απόκλιση ωριαίας κατανάλωσης

## Εξειδικευμένα χαρακτηριστικά\*

1. Κινούμενος μέσος όρος μηνιαίου μέσου όρου
2. Κινούμενος μέσος όρος μηνιαίας τυπικής απόκλισης
3. Συμμετρική διαφορά καταναλώσεων
4. Συμμετρική διαφορά τυπικής απόκλισης
5. Τμηματική διαφορά κατανάλωσης με όμοιους καταναλωτές
6. Τμηματική διαφορά τυπικής απόκλισης με όμοιους καταναλωτές

\*Αν το χαρακτηριστικό δεν έχει μεγάλη τιμή τίθεται μηδέν.



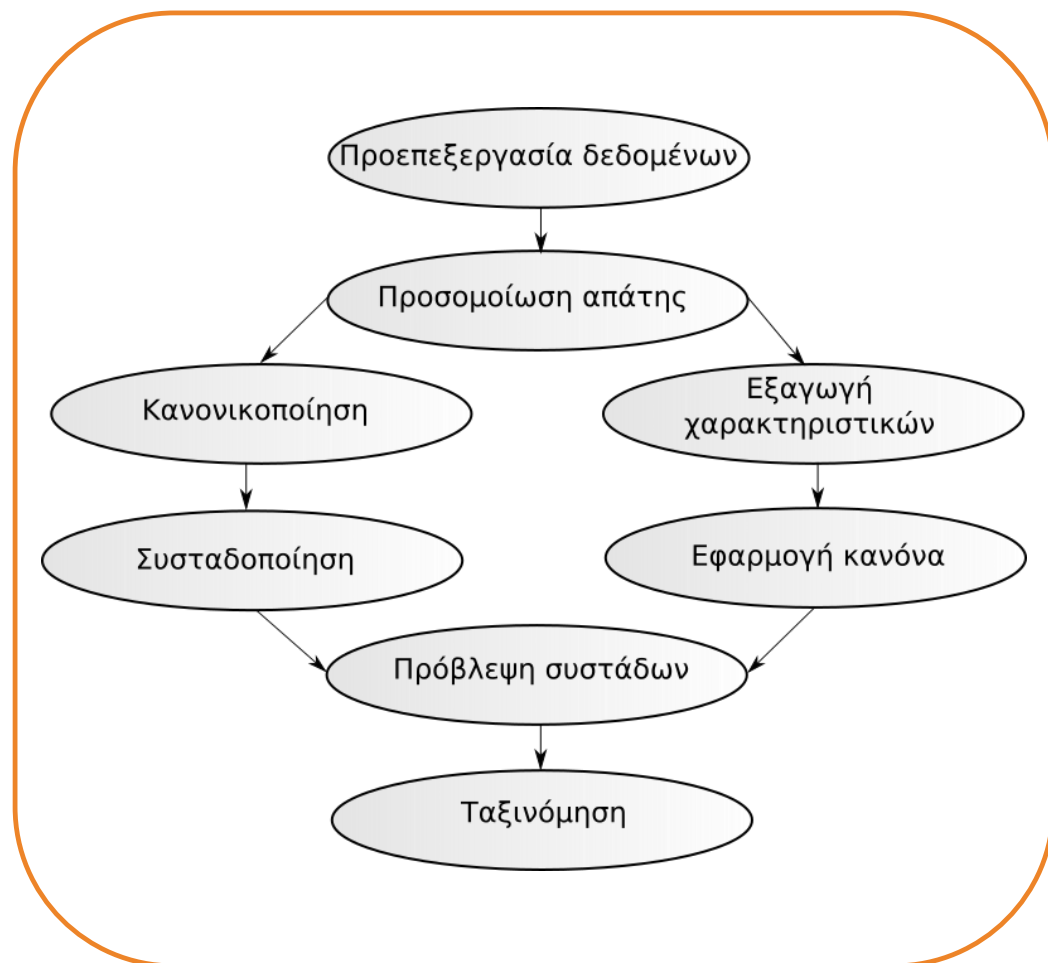
Λεπτομέρειες δοκιμής

- 2.000 καταναλωτές
- 50% καταναλωτών με μη τεχνικές απώλειες

Τι προσφέρουν τα χαρακτηριστικά;

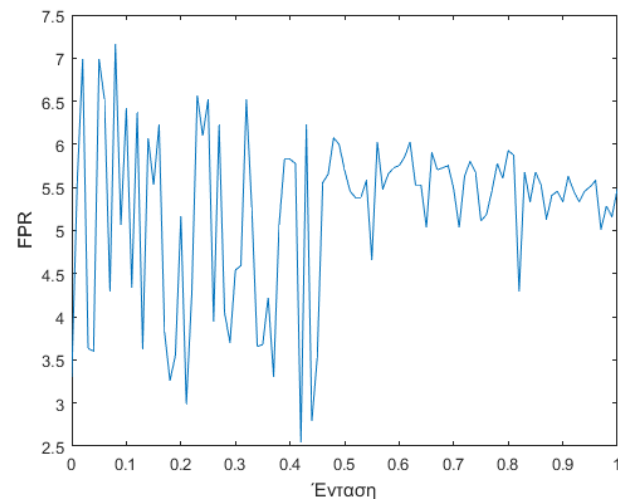
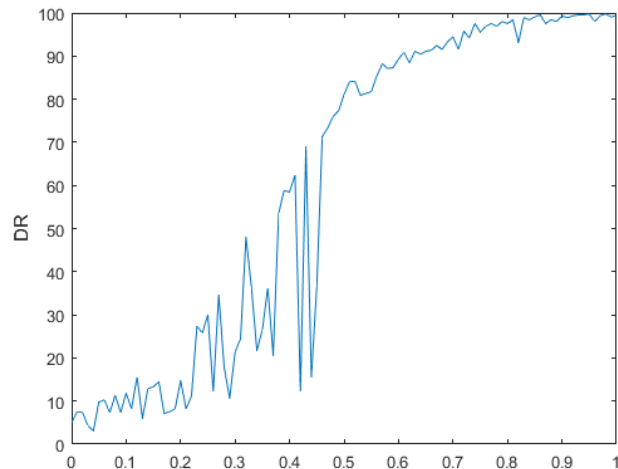
- Ανωνυμοποίηση καταναλωτών
- Σημαντική μείωση όγκου δεδομένων
- Εξειδίκευση των δεδομένων στο ζητούμενο

# ΣΥΣΤΗΜΑΤΑ ΜΗ ΕΠΙΒΛΕΠΟΜΕΝΗΣ ΜΑΘΗΣΗΣ



- Κανονικοποίηση χρονοσειρών
  - Κανονικοποίηση ετήσιου διανύσματος
  - Εύρος κανονικοποίησης  $[-1,1]$
- Συσταδοποίηση σε 2 συστάδες
  - 1 συστάδα με φυσιολογικές καταναλώσεις
  - 1 συστάδα με ιδιόμορφες καταναλώσεις
- Εξαγωγή 8 χαρακτηριστικών
  - 2 γενικά χαρακτηριστικά
  - 6 εξειδικευμένα χαρακτηριστικά
- Εφαρμογή κανόνα
  - Απενοχοποίηση όταν οι καταναλωτές έχουν λιγότερα από 3 εξειδικευμένα χαρακτηριστικά

# ΑΠΟΤΕΛΕΣΜΑΤΑ ΜΗ ΕΠΙΒΛΕΠΟΜΕΝΟΥ ΑΛΓΟΡΙΘΜΟΥ ΩΣ ΠΡΟΣ ΤΗΝ ΜΕΤΑΒΟΛΗ ΤΗΣ ΕΝΤΑΣΗΣ



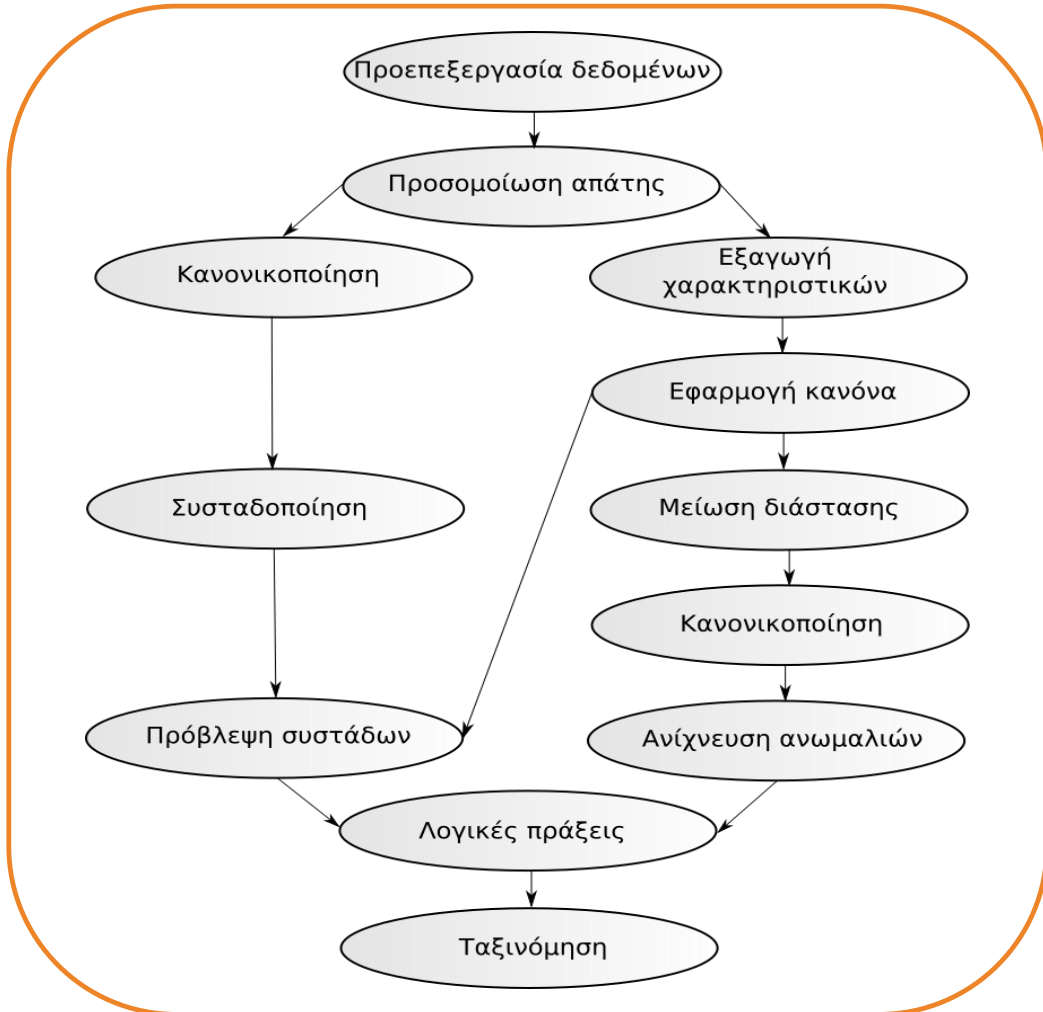
## Παρατηρήσεις

- Έντονες μεταβολές μέχρι 50% έντασης κλοπής
- Ομαλή αύξηση DR μετά το 50%
- Σταθεροποίηση FPR μετά το 50%
- Προσανατολισμός συστήματος σε απάτες υψηλής έντασης
- Υψηλό DR
- FPR με περιθώρια βελτίωσης

## Λεπτομέρειες δοκιμής

- 4.500 καταναλωτές
- 8 χαρακτηριστικά κάθε καταναλωτής
- 10% καταναλωτών με μη τεχνικές απώλειες
- Απάτη τύπου 1

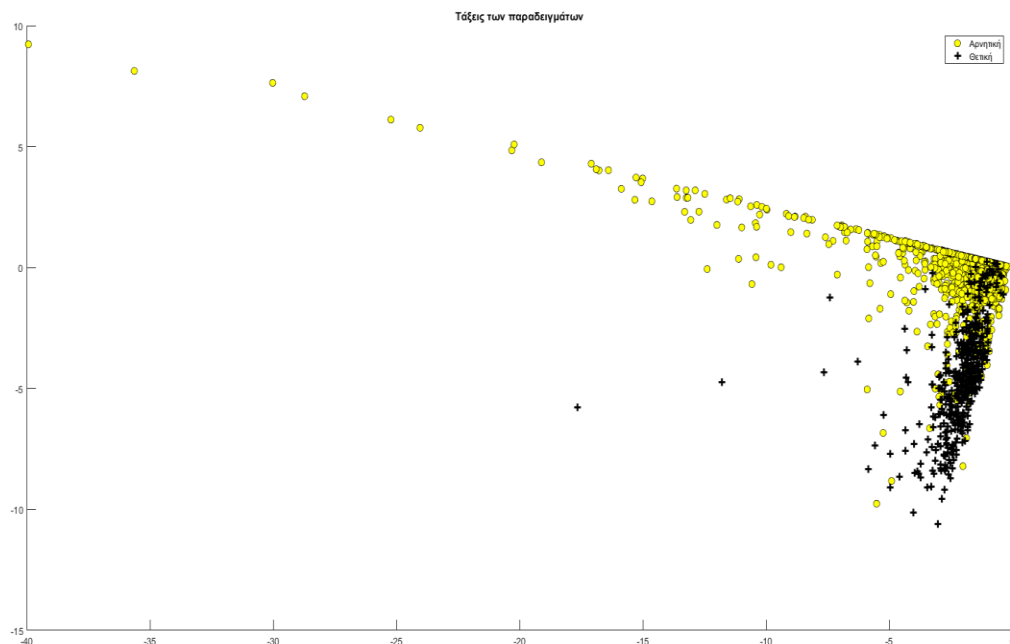
# ΣΥΣΤΗΜΑΤΑ ΗΜΙ-ΕΠΙΒΛΕΠΟΜΕΝΗΣ ΜΑΘΗΣΗΣ



- Κανονικοποίηση χρονοσειρών
- Συσταδοποίηση σε 2 συστάδες
- Εξαγωγή 8 χαρακτηριστικών
- Εφαρμογή κανόνα
- Μείωση διάστασης
  - Principal Component Analysis
  - Οπτικοποίηση χαρακτηριστικών σε 2 διαστάσεις
- Κανονικοποίηση χαρακτηριστικών
  - Εύρος κανονικοποίησης [0,1]
- Ανίχνευση ανωμαλιών
  - Πολυμεταβλητή Γκαουσιανή κατανομή
- Λογικές Πράξεις μεταξύ ταξινομητών
  - AND/OR

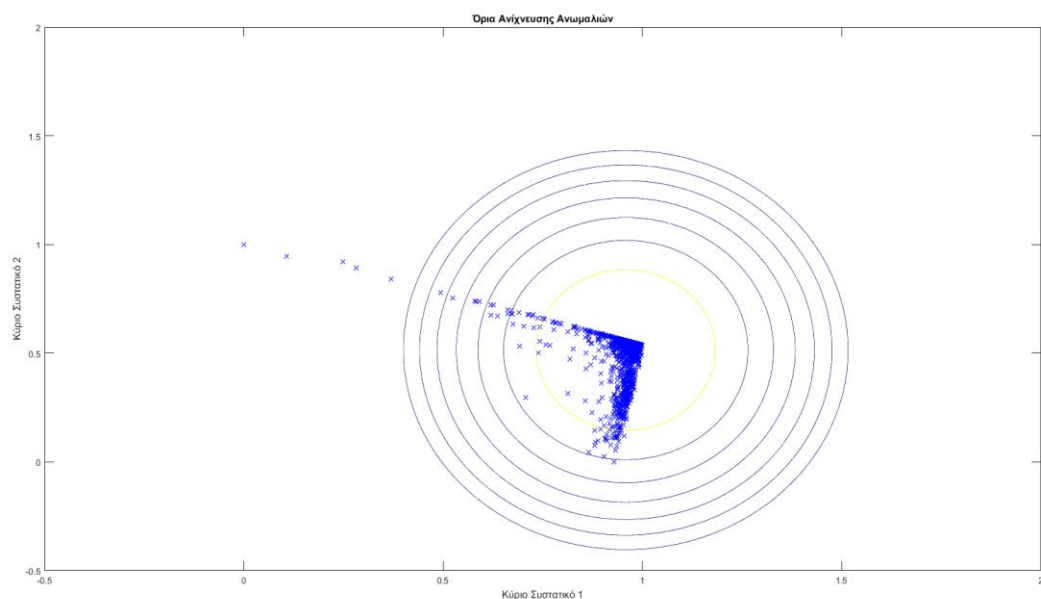
# ΟΠΤΙΚΟΠΟΙΗΣΗ ΧΑΡΑΚΤΗΡΙΣΤΙΚΩΝ ΜΕ ΜΕΙΩΣΗ ΔΙΑΣΤΑΣΗΣ

Διαστάσεις	DR	FPR	Acc.	FI	BDR
8	71,09	2	95,49	74,64	80
2	79,01	2,51	95,59	78,65	78



- Οι περισσότεροι καταναλωτές βρίσκονται στο κέντρο των αξόνων
- Πάνω αριστερά βρίσκονται οι μικρομεσαίες επιχειρήσεις
- Κάτω δεξιά βρίσκονται οι καταναλωτές με ιδιαίτερα χαμηλές-ύποπτες καταναλώσεις

# ΠΟΛΥΜΕΤΑΒΛΗΤΗ ΓΚΑΟΥΣΙΑΝΗ ΚΑΤΑΝΟΜΗ

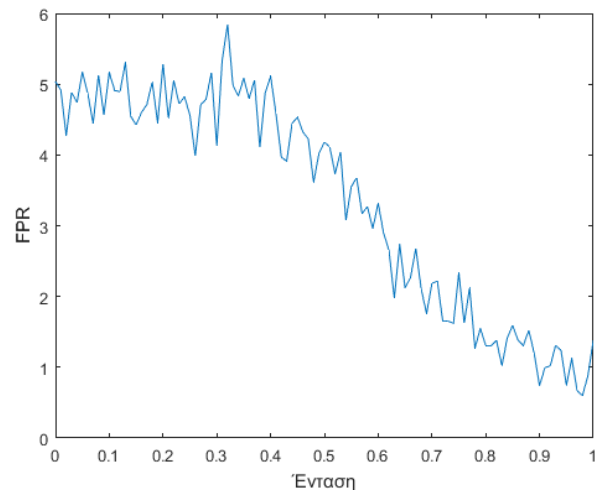
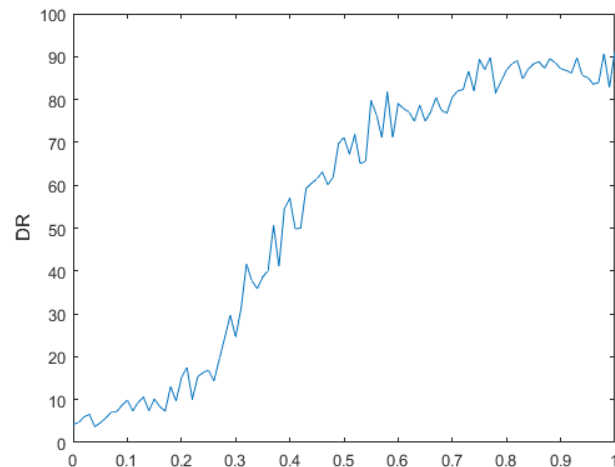


Βήματα ταξινόμησης με ανίχνευση ανωμαλιών:

1. Με 70% των καταναλωτών εξάγονται διανύσματα μέσου όρου και διακύμανσης.
2. Με 30% των καταναλωτών βελτιστοποιείται το όριο διαχωρισμού βάσει βέλτιστου FI score.
3. Ταξινόμηση του 70% των καταναλωτών βάσει του βέλτιστου ορίου διαχωρισμού.

- Πλατύς κύκλος
- Διαχωρισμός των τυπικών καταναλωτών από τους καταναλωτές με ιδιαίτερες ενεργειακές ανάγκες

# ΑΠΟΤΕΛΕΣΜΑΤΑ ΗΜΙ-ΕΠΙΒΕΛΟΜΕΝΟΥ ΣΥΣΤΗΜΑΤΟΣ ΩΣ ΠΡΟΣ ΤΗΝ ΜΕΤΑΒΟΛΗ ΤΗΣ ΕΝΤΑΣΗΣ



## Παρατηρήσεις

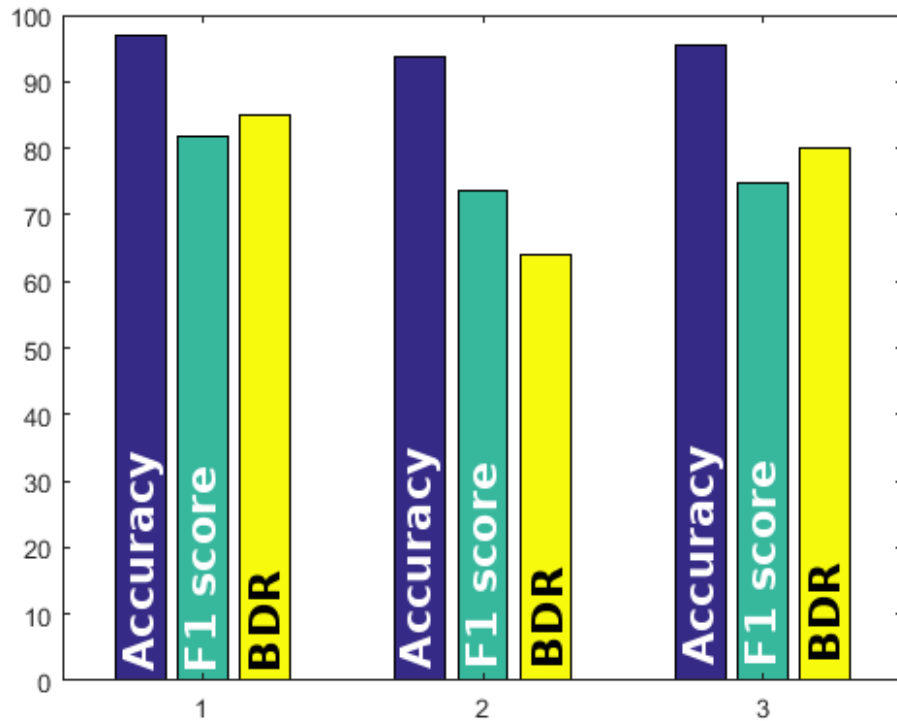
- Ομαλές μεταβολές του DR και FPR
- Σταθερά χαμηλό FPR
- Σχετικά χαμηλό DR
- Ικανοποιητικά αποτελέσματα για πάνω από 60% ένταση κλοπής

## Λεπτομέρειες δοκιμής

- 4.500 καταναλωτές
- 8 χαρακτηριστικά κάθε καταναλωτής
- 10% καταναλωτών με μη τεχνικές απώλειες
- Απάτη τύπου 1



## ΣΥΓΚΡΙΣΗ ΑΠΟΤΕΛΕΣΜΑΤΩΝ



Αλγ.	DR	FPR	Acc.	FI	BDR
Επιβλ.	81,43	1,24	96,96	84,76	88
Μη επιβλ.	85,78	4,99	94,09	74,37	66
Ημι-επιβλ.	71,09	2	95,49	74,64	80

1. Γραμμικό SVM
2. Μη επιβλεπόμενο σύστημα με κανόνες
3. Ανίχνευση ανωμαλιών με πολυμεταβλητή Γκαουσιανή κατανομή

# ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΗ ΕΠΕΚΤΑΣΗ

## Αξιόπιστες τεχνικές για τον εντοπισμό ρευματοκλοπών

- Γραμμική ταξινόμηση χρονοσειρών
- Κανονικοποίηση καταναλώσεων και συσταδοποίηση χρονοσειρών
- Εξαγωγή χαρακτηριστικών και κανονικοποίηση
- Ανίχνευση ανωμαλιών στα χαρακτηριστικά
- Συνδυασμός αλγορίθμων και σύνθεση συστημάτων.

## Μελλοντική κατεύθυνση αυτής της διπλωματικής

- Εφαρμογή ταξινόμησης σε περισσότερες από δύο κλάσεις
- Εξερεύνηση τεχνικών ανίχνευσης ανωμαλιών
- Ταξινόμηση βάσει πρόβλεψης μελλοντικής χρονοσειράς

# ΕΡΩΤΗΣΕΙΣ

