

# Μεθοδολογία Έρευνας- Ανακάλυψη γνώσης από δεδομένα

**Δρ. Δ. Παπακυριακόπουλος,**  
Επιστημονικός Συνεργάτης  
Τμήμα Διοίκησης Επιχειρήσεων,  
ΤΕΙ Αθήνας

**Δρ. Σ. Μπινιώρης**  
Καθηγητής  
Τμήμα Διοίκησης Επιχειρήσεων,  
ΤΕΙ Αθήνας



# Το πρόβλημα

- **ΣΤΟΧΟΣ**

- Δημιουργία (διοικητικού) μηχανισμού για τον εντοπισμό των ελλείψεων στα ράφια των Supermarket.

- **ΕΡΕΥΝΗΤΙΚΑ ΚΙΝΗΤΡΟ**

- Πρόβλημα εφοδιαστικής αλυσίδας
  - **Αυξημένη** πολυπλοκότητα. Συνδέει
    - Δυνάμεις προσφοράς (παραδόσεις) στο κατάστημα
    - Ικανότητα αναπλήρωσης καταστήματος
    - Ζήτηση καταναλωτή
  - Συνδέεται με **Bullwhip effect** (Lee et al, 1997)
  - Σχετικές εργασίες αναφέρονται στην **αντίδραση του καταναλωτή** (Fitzimons et al, , Anderson et al.,)
- Τοποθετείται ψηλά στα προβλήματα του λιανεμπορίου
  - Ο ρυθμός ελλείψεων είναι άνω του 8% (Gruen et al, 2007)
  - Η απώλεια εσόδου είναι



# Εναλλακτικές προσεγγίσεις

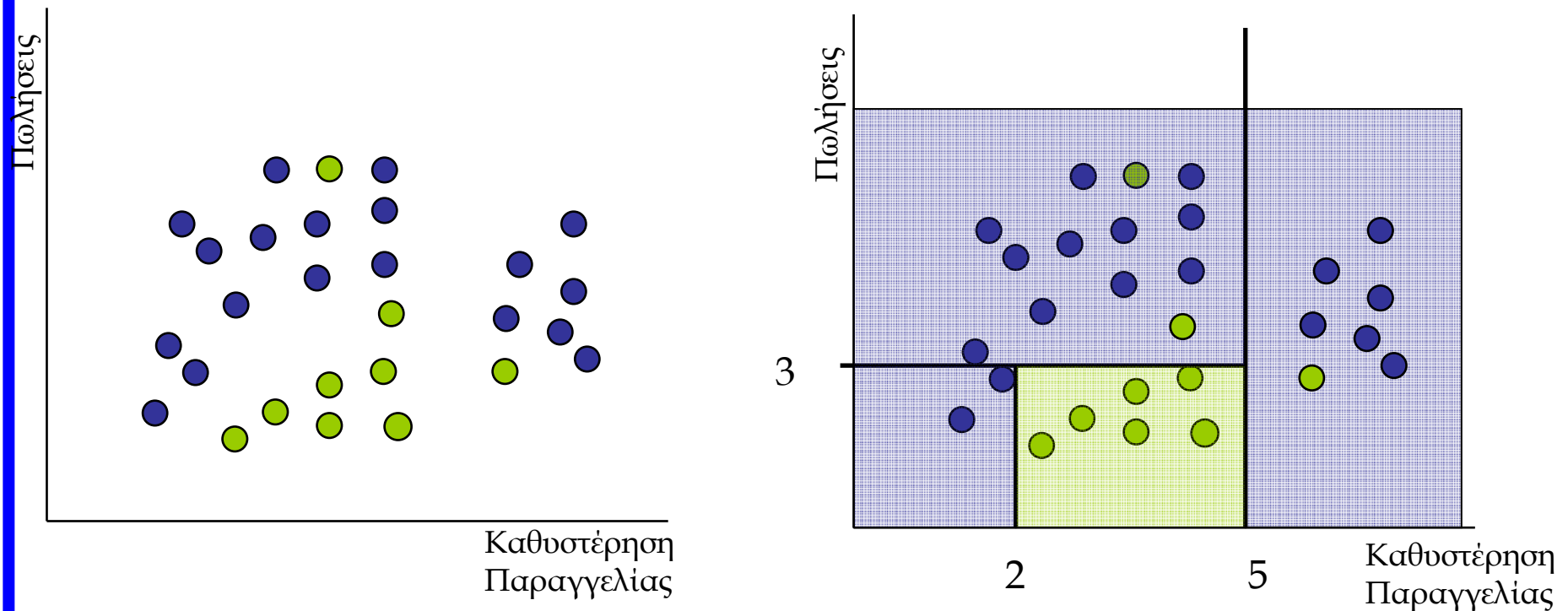
## 1. Αναλυτικές προσεγγίσεις – Top Down

- Προβλέψεις Ζήτησης
  - **Συμπέρασμα:** Το 2% των προϊόντων παρουσιάζουν χαρακτηριστικά εργοδικής ζήτησης
- Διαχείριση Αποθεμάτων
  - **Συμπέρασμα:** Το 54% των προϊόντων παρουσιάζουν αξιόπιστες εγγραφές αποθέματος. Πρόβλημα μείωσης του αποθέματος (Inventory Shrinkage)

## 2. Ευρετικές προσεγγίσεις – Bottom Up

- Ανακάλυψη Γνώσης από τα Δεδομένα
  - η εξεύρεση μιας (ενδιαφέρουσας, αυτονόητης, μη προφανής και πιθανόν χρήσιμης) πληροφορίας ή προτύπων από μεγάλες βάσεις δεδομένων με χρήση αλγορίθμων
  - Ο κλάδος διαθέτει σημαντικές και αναξιόποιητες πηγές δεδομένων

# Αντίληψη του προβλήματος



●: OOS  
●: EXISΩΣ

ΕΑΝ “Καθυστέρηση Παραγγελίας” > 5 ΤΟΤΕ EXISΩΣ  
ΕΑΝ “Πωλήσεις” > 3 ΤΟΤΕ EXISΩΣ  
ΕΑΝ  $X > 2$  ΤΟΤΕ OOS  
ΓΙΑ ΤΑ ΥΠΟΛΟΙΠΑ EXISΩΣ

# Ερευνητική Μεθοδολογία

Ποια καταστήματα;  
Ποια προϊόντα;  
Για πόσο διάστημα;

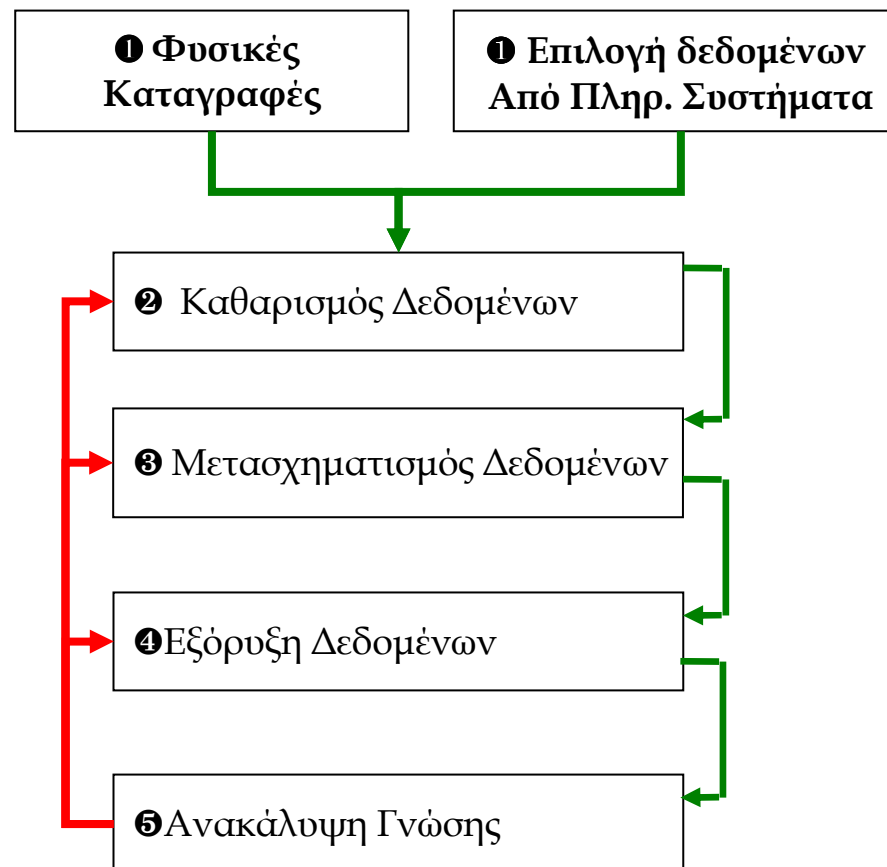
Όλα τα δεδομένα;  
Τα διαθέτει ο κλάδος;  
Είναι ορθά;

Ποιες είναι οι ασυνέπειες των δεδομένων;  
Τι κάνω με μη-έγκυρες καταχωρήσεις;  
Ποια είναι η ποιότητα των δεδομένων;

Ποιες είναι οι επεξηγηματικές μεταβλητές;  
Τι κάνω με τις "ά-τιμες" μεταβλητές;  
Έμφαση σε συνεχή ή διακριτές τιμές;

Ποιο είναι το καλύτερο μοντέλο;  
Πως η ασθενής κλάση μπορεί να ενδυναμωθεί;  
Υπάρχει μεροληψία ή διακύμανση;

Γενικεύονται τα αποτελέσματα;  
Εσωτερική ή εξωτερική εγκυρότητα;  
Μπορούν να αναπαρασταθούν;



# Εφαρμογή Μεθοδολογίας

## Προφίλ

- Έρευνα πεδίου (δεν ελέγχουμε όλες τις μεταβλητές)
- Τέσσερις αλυσίδες λιανικής
  - Τρία αντιπροσωπευτικά καταστήματα από κάθε αλυσίδα
- Επιλογή 12 κύριων προϊόντικών κατηγοριών
  - Στρωματοποιημένη δειγματοληψία για 10 προϊόντα από κάθε κατηγορία (ταχύτητα κίνησης και Διακύμανση ζήτησης) + 15 προϊόντα που θα είναι σε προώθηση κατά την διάρκεια των Φυσικών απογραφών
  - Τα προϊόντα παραδίδονται από κεντρικής αποθήκης λιανέμπορου
- Διάρκεια 4 εβδομάδες με 1 επίσκεψη την ημέρα (πρωί-απόγευμα)

**Σύνολο μάθησης 19.400 περίπου εγγραφών**

# Επιλογή δεδομένων

- **Φάση I**

- Σύνολα δεδομένων για την επιλογή προϊόντων και καταστημάτων που θα εξετασθούν
- Ορίσθηκαν οι βασικές προδιαγραφές των πειραμάτων

- **Φάση II**

- Ιστορικά στοιχεία πωλήσεων και παραγγελιών καταστήματος (6 μήνες έως 1 χρόνος ανάλογα με την αλυσίδα)

# Καθαρισμός δεδομένων

- Οι αλυσίδες έχουν διαφορετικά προβλήματα με τα δεδομένα που χρησιμοποιούν
  - Μείγμα καταστήματος
  - Backordered παραγγελίες
  - Διαφορές Παραγγελίας - Παράδοση
  - Προβλήματα στην κατηγοριοποίηση
- Κατά την διαδικασία της φυσικής απογραφής έπρεπε να επιβεβαιωθεί το καρτελάκι στο ράφι
  - Προϊόντα σε διαφορετικές θέσεις (προωθητικά)
  - Μη-επικαιροποιημένο καρτελάκι στο ράφι

**Το 60% της ερευνητικής προσπάθειας**

**Συνεπές σύνολο δεδομένων**



# Μετασχηματισμός Δεδομένων

- **Όγκος Πωλήσεων**
  - Ημερήσιος μέσος όρος και τυπική απόκλιση
  - Συχνότητα όγκου πώλησης και τυπική απόκλιση
  - Εύρος περιόδων χωρίς πώληση
  - Δείκτης ταχυκινήσις [0..1]
  - Μερίδιο κατηγορίας
- **Απόθεμα**
  - Τελευταία παραγγελία
  - Τελευταία παράδοση
  - Εκτίμηση επιπέδου αποθέματος
- **Τάσεις ζήτησης**
  - Προωθητικό προϊόν
  - Εποχικότητα
- **Κατάστημα**
  - Μέγεθος καταστήματος
  - Μέγεθος αποθήκης
  - Αριθμός εργαζομένων στη βάρδια
  - Ημέρα
  - Ώρα της ημέρας

Εφαρμογή κριτηρίων Info Gain και  $\chi^2$

Ημερίδα Μεθοδολογία της Έρευνας

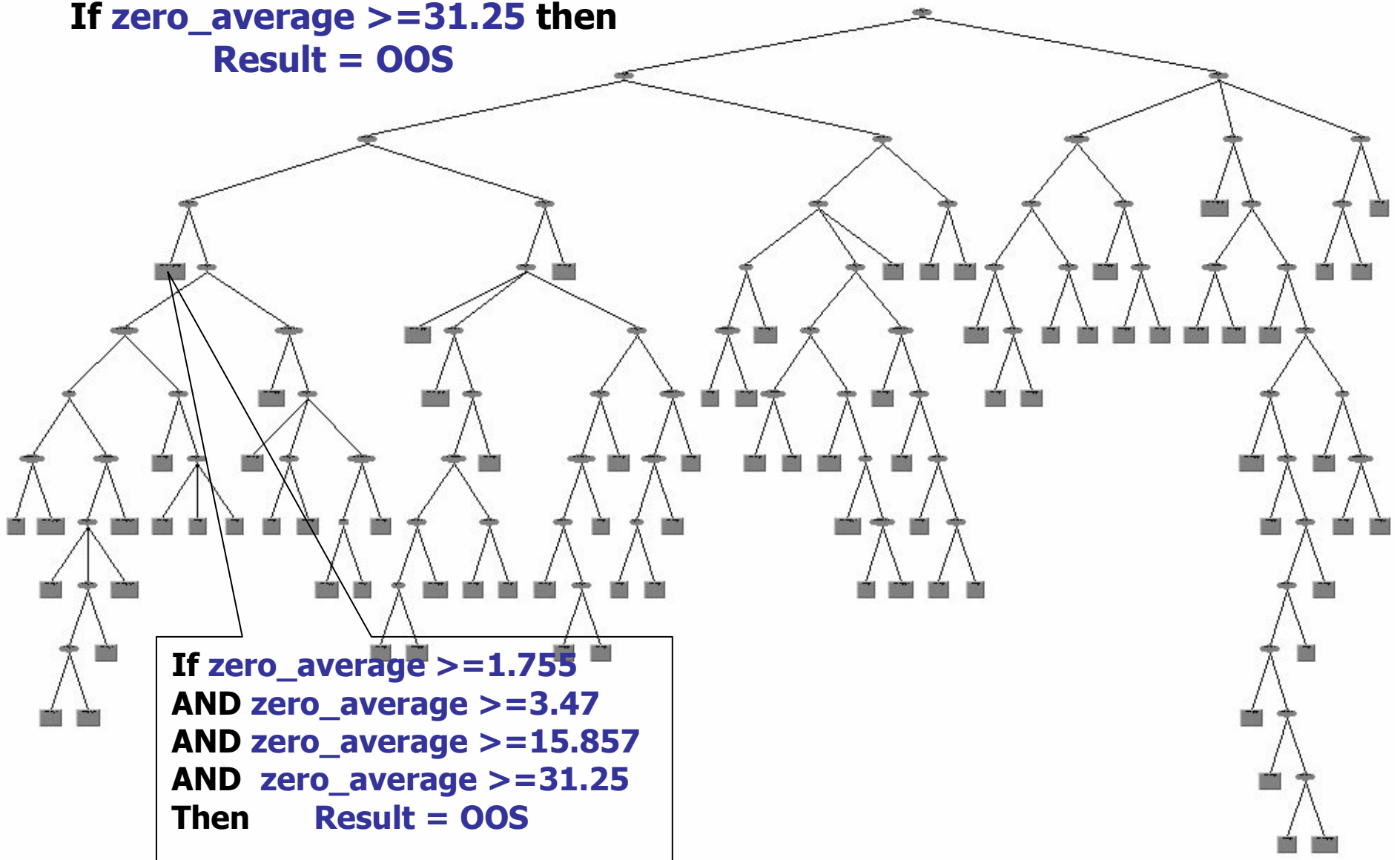
# Εξόρυξη Δεδομένων (1/2)

- Επιλογή τεχνικών εξόρυξης διαφορετικής προσέγγισης
  - Στατιστικές τεχνικές (πχ Naïve Bayes)
  - Δέντρα Απόφασης (πχ C4.5)
  - Νευρωνικά Δίκτυα (πχ Multilayer Perceptron)
- Αξιοποίηση Συνόλου Μάθησης με δυο τρόπους
  - Εκπαίδευση (training)
  - Έλεγχος (testing)
- Εξετάσθηκαν 21 διαφορετικές τεχνικές εξόρυξης
  - Κάθε τεχνική έχει διαφορετικές παραμέτρους
    - Υπάρχει η ανάγκη καθορισμού των βέλτιστων τιμών
- Μελέτη προβλεπτικής ικανότητας αλγορίθμων.
  - Χρήση 10-Fold Cross Validation
    - Analysis of Variance

**Τα δέντρα απόφασης είναι η αποτελεσματικότερη προσέγγιση**

Tree View

**If zero\_average  $\geq 31.25$  then  
Result = OOS**



**If zero\_average  $\geq 1.755$   
AND zero\_average  $\geq 3.47$   
AND zero\_average  $\geq 15.857$   
AND zero\_average  $\geq 31.25$   
Then Result = OOS**

# Ανακάλυψη Γνώσης

	Class	Length	Supp.	Conf.
0	exists	0	100.0%	84.9%
1	oos	3	3.7%	91.3%
2	exists	1	22.3%	84.9%
3	oos	3	1.4%	89.4%
4	oos	1	13.7%	64.5%
5	exists	4	0.6%	66.7%
6	oos	4	0.1%	100.0%
7	exists	3	1.9%	96.4%
8	oos	2	3.8%	55.8%
9	exists	1	3.2%	96.3%
10	exists	1	21.6%	97.7%
11	exists	2	9.4%	99.1%
12	exists	2	16.3%	96.5%
13	exists	1	29.0%	96.0%
14	exists	1	25.3%	72.3%
15	exists	3	0.5%	93.8%
16	exists	2	2.7%	81.5%
17	exists	3	3.6%	99.1%
18	oos	5	0.1%	100.0%
19	oos	2	5.9%	82.6%
20	exists	2	17.7%	89.9%

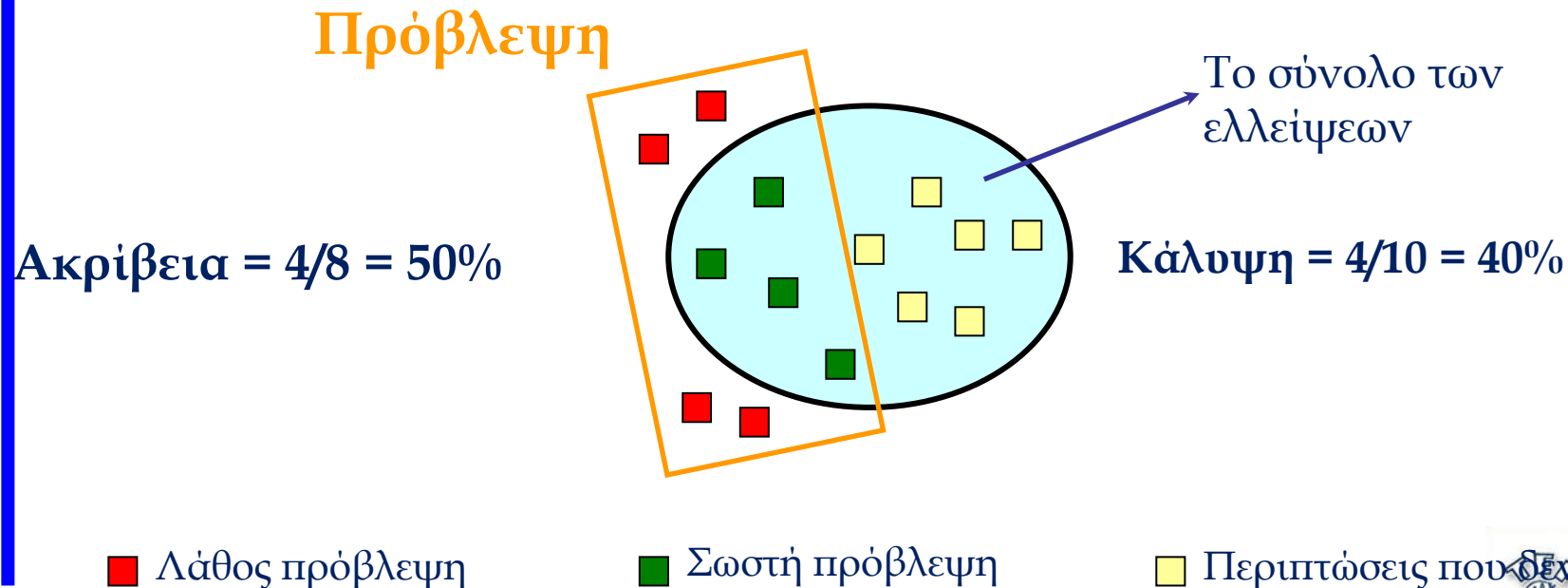
- 1 IF store\_type = 1 AND zero\_avg >= 31.25 AND zero\_seq < 4 ΩHEN RES = oos
- 2 IF store\_type = 2 ΩHEN RES = exists
- 3 IF store\_type = 3 AND zero\_avg >= 31.25 AND zero\_seq < 7 ΩHEN RES = oos
- 4 IF zero\_avg >= 31.25 ΩHEN RES = oos
- 5 IF store\_type = 1 AND zero\_avg >= 31.25 AND zero\_seq < 7 AND zero\_seq >= 4 ΩHEN RES = exists
- 6 IF store\_type = 2 AND zero\_avg >= 31.25 AND zero\_seq < 23 AND zero\_seq >= 14 ΩHEN RES = oos
- 7 IF store\_type = 3 AND zero\_seq < 12 AND zero\_seq >= 7 ΩHEN RES = exists
- 8 IF zero\_avg >= 63 AND zero\_seq < 3 ΩHEN RES = oos
- 9 IF zero\_avg < 1.230769 ΩHEN RES = exists
- 10 IF zero\_avg < 1.583333 ΩHEN RES = exists
- 11 IF store\_type = 3 AND zero\_avg < 1.775 ΩHEN RES = exists
- 12 IF store\_type = 1 AND zero\_avg < 31.25 ΩHEN RES = exists
- 13 IF avg >= 1.384615 ΩHEN RES = exists
- 14 IF zero\_seq >= 4 ΩHEN RES = exists
- 15 IF zero\_avg < 31.25 AND zero\_avg >= 21 AND zero\_seq >= 42 ΩHEN RES = exists
- 16 IF zero\_avg < 31.25 AND zero\_avg >= 15.85714 ΩHEN RES = exists
- 17 IF store\_type = 4 AND zero\_avg < 5.823529 AND zero\_avg >= 3.470588 ΩHEN RES = exists
- 18 IF store\_type = 4 AND zero\_avg < 31.25 AND zero\_avg < 6.052631 AND zero\_avg >= 5.823529 AND zero\_seq >= 4 ΩHEN RES = oos
- 19 IF store\_type = 1 AND zero\_avg >= 31.25 ΩHEN RES = oos
- 20 IF store\_type = 2 AND zero\_avg < 51.5 ΩHEN RES = exists

# Εξωτερική εγκυρότητα

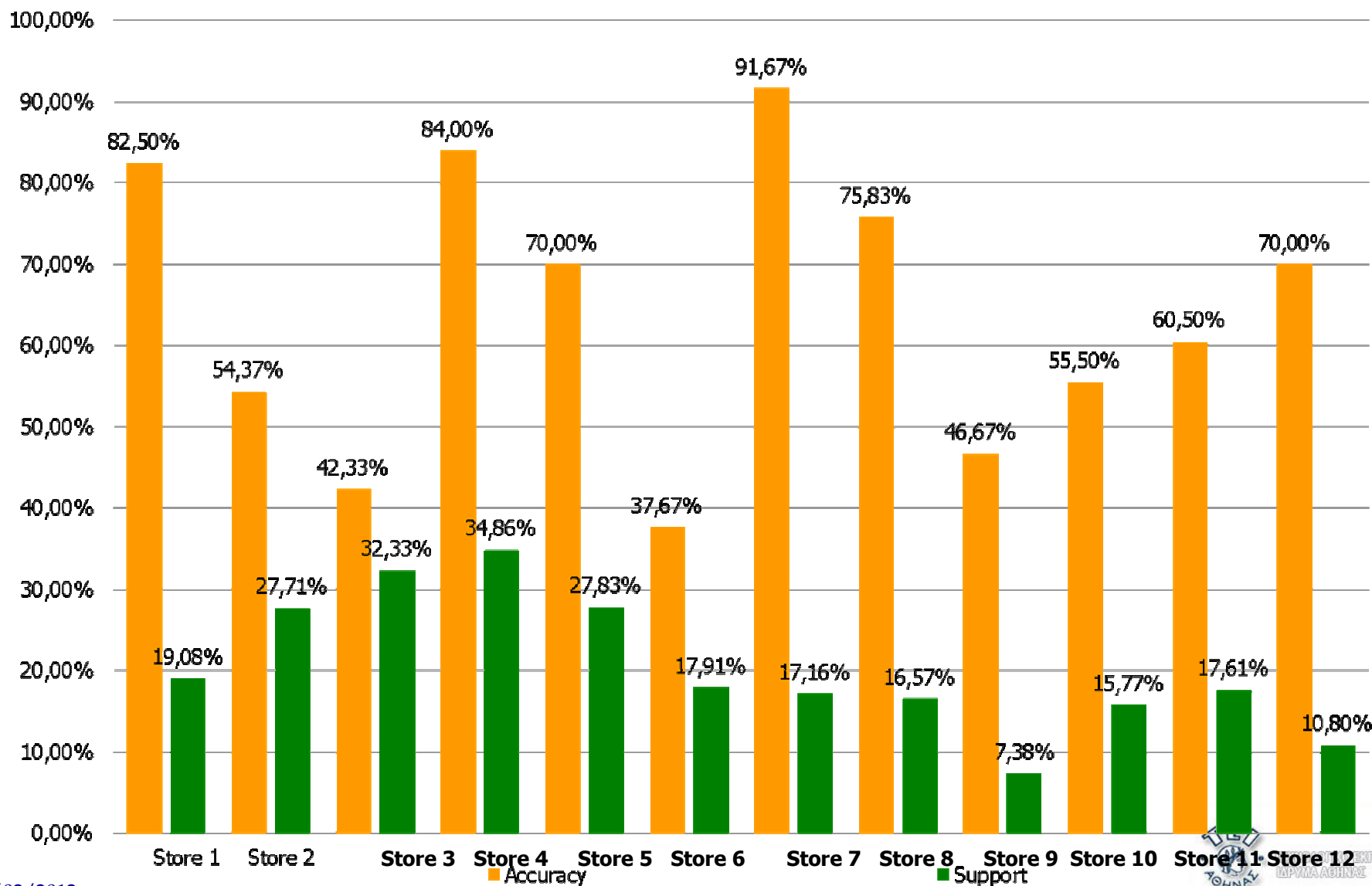
- Εφαρμογή των κανόνων σε νέο και άγνωστο σύνολο δεδομένων.
- Καθημερινή και αυτοματοποιημένη διασύνδεση με τις αλυσίδες λιανικής
- Η ταυτόχρονη εφαρμογή των κανόνων δημιουργεί αναφορές
  - Καθημερινή λίστα προϊόντων σε έλλειψη
  - Καθημερινή αναφορά σε λιανέμπορους και προμηθευτές

# Μέτρα Αξιολόγησης

- **Ακρίβεια:**  $\text{OOS True Predicted} / \text{Σύνολο OOS που εντοπίστηκαν}$
- **Κάλυψη:**  $\text{OOS True Predicted} / \text{Σύνολο των ελλείψεων στο κατάστημα}$



# Αποτελέσματα ανά κατάστημα



21/02/2012

Ημερίδα Μεθοδολογία της Έρευνας



# Συμπεράσματα

- Η χρήση της διαδικασίας ανακάλυψης γνώσης από τα δεδομένα είναι εφικτή.
- Σε κάθε βήμα της διαδικασίας υπάρχει πλήθος ερευνητικών υποθέσεων και αποφάσεων που πρέπει να ληφθούν
- Η γνώση που δημιουργείται αποτυπώνεται με την μορφή κανόνων
  - Κάθε αλυσίδα λιανικής έχει διαφορετικούς κανόνες
  - Κάθε κατάσταση έχει διαφορετικούς κανόνες
  - Σε κάθε προϊοντική κατηγορία είναι δυνατό να υπάρχουν διαφορετικοί κανόνες
- Πολυπλοκότητα vs Απλότητα της ευρετικής λύσης
- Η πρόταση απαιτεί συστηματική επένδυση από τη πλευρά της επιχείρησης
- Παράγοντες που επιδρούν στις μετρήσεις
  - Ανθρώπινη συμπεριφορά
  - Μετρήσεις ελλείψεων γίνονται ταυτόχρονα με την αναπλήρωση των προϊόντων