



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ  
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΚΑΙ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ  
ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

**ΠΡΟΒΛΕΨΗ ΖΗΤΗΣΗΣ ΣΕ ΚΑΤΑΣΤΗΜΑΤΑ ΛΙΑΝΙΚΗΣ ΜΕ  
ΜΕΘΟΔΟΥΣ ΙΕΡΑΡΧΙΚΗΣ ΣΥΝΑΘΡΟΙΣΗΣ ΚΑΙ ΤΕΧΝΙΚΕΣ  
ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΚΩΝΣΤΑΝΤΙΝΟΣ ΣΕΓΡΕΔΟΣ

**Επιβλέπων:** Βασίλειος Ασημακόπουλος

Καθηγητής Ε.Μ.Π.

**Υπεύθυνος:** Ευάγγελος Σπηλιώτης

Διδάκτωρ Ε.Μ.Π





ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ

ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ ΜΗΧΑΝΙΚΩΝ ΗΛΕΚΤΡΟΝΙΚΩΝ  
ΥΠΟΛΟΓΙΣΤΩΝ

ΤΟΜΕΑΣ ΗΛΕΚΤΡΙΚΩΝ ΚΑΙ ΒΙΟΜΗΧΑΝΙΚΩΝ ΔΙΑΤΑΞΕΩΝ ΚΑΙ  
ΣΥΣΤΗΜΑΤΩΝ ΑΠΟΦΑΣΕΩΝ

**ΠΡΟΒΛΕΨΗ ΖΗΤΗΣΗΣ ΣΕ ΚΑΤΑΣΤΗΜΑΤΑ ΛΙΑΝΙΚΗΣ ΜΕ  
ΜΕΘΟΔΟΥΣ ΙΕΡΑΡΧΙΚΗΣ ΣΥΝΑΘΡΟΙΣΗΣ ΚΑΙ ΤΕΧΝΙΚΕΣ  
ΜΗΧΑΝΙΚΗΣ ΜΑΘΗΣΗΣ**

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

ΚΩΝΣΤΑΝΤΙΝΟΣ ΣΕΓΡΕΔΟΣ

**Επιβλέπων:** Βασίλειος Ασημακόπουλος

Καθηγητής Ε.Μ.Π.

**Υπεύθυνος:** Ευάγγελος Σπηλιώτης

Διδάκτωρ Ε.Μ.Π

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....

Βασίλειος Ασημακόπουλος

Καθηγητής Ε.Μ.Π.

.....

Ιωάννης Ψαρράς

Καθηγητής Ε.Μ.Π.

.....

Δημήτριος Ασκούνης

Καθηγητής Ε.Μ.Π.

.....

**ΣΕΓΡΕΔΟΣ ΚΩΝΣΤΑΝΤΙΝΟΣ**

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π

Copyright © Σεγρέδος Κωνσταντίνος, 2020

Με επιφύλαξη παντός δικαιώματος. All rights reserved.

Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τους συγγραφείς.

Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τους συγγραφείς και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν τις επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου.

## Πρόλογος

Η διπλωματική αυτή εργασία εκπονήθηκε στα πλαίσια των ερευνητικών δραστηριοτήτων της Μονάδας Προβλέψεων και Στρατηγικής κατά το ακαδημαϊκό έτος 2019-2020. Η μονάδα υπάγεται στον Τομέα Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων της Σχολής Ηλεκτρολόγων Μηχανικών και Μηχανικών Η/Υ, του Εθνικού Μετσόβιου Πολυτεχνείου.

Αρχικά, θα ήθελα να ευχαριστήσω τον Καθηγητή κ. Βασίλειο Ασημακόπουλο για την ευκαιρία που μου έδωσε να εκπονήσω την συγκεκριμένη διπλωματική εργασία στον τομέα των προβλέψεων. Ακόμα θα ήθελα να ευχαριστήσω τον Καθηγητή κ. Ιωάννη Ψαρρά και τον Καθηγητή κ. Δημήτριο Ασκούνη για τη συμμετοχή τους στην επιτροπή εξέτασης της εργασίας.

Τέλος θα ήθελα να ευχαριστήσω ιδιαίτερα τον Ευάγγελο Σπηλιώτη για την συνεχή βοήθεια του και την καθοδήγηση του σε όλη τη διάρκεια εκπόνησης της εργασίας, καθώς και τους φίλους και συγγενείς μου για την υποστήριξη τους.

Κωνσταντίνος Σεργέδος

## Περίληψη

Η παρούσα διπλωματική εργασία εξετάζει την πρόβλεψη των πωλήσεων σε καταστήματα λιανικής με την χρήση μεθόδων ιεραρχικής συνάθροισης και τεχνικές μηχανικής μάθησης. Σκοπός είναι να εξακριβωθεί πόσο καλά ανταποκρίνονται οι μέθοδοι μηχανικής μάθησης σε ιεραρχικά δεδομένα σε σχέση με τις κλασικές μεθόδους πρόβλεψης.

Η μελέτη γίνεται σε ένα σύνολο μιας αλυσίδας καταστημάτων διαφορετικών γεωγραφικών περιοχών, οπότε κρίνεται αναγκαία η συμφιλίωση των επιμέρους προβλέψεων που θα παραχθούν ώστε οι προβλέψεις των χαμηλότερων επιπέδων να αθροίζουν σε αυτές των υψηλότερων. Για αυτό το λόγο γίνεται χρήση μεθόδων ιεραρχικής πρόβλεψης(hierarchical forecasting).

Εισαγωγικά παρουσιάζονται κάποια βασικά χαρακτηριστικά των προβλέψεων, των μοντέλων χρονοσειρών καθώς και η ανάλυση που γίνεται πριν τη διαδικασία της πρόβλεψης. Αναλύονται στη συνέχεια οι μέθοδοι πρόβλεψης που θα χρησιμοποιηθούν ως σημείο αναφοράς και οι μέθοδοι μηχανικής μάθησης καθώς και οι δείκτες σφάλματος για την σύγκριση των μεθόδων. Ύστερα περιγράφονται οι μέθοδοι ιεραρχικής συνάθροισης και οι βασικές τεχνικές παραγωγής ιεραρχικών προβλέψεων.

Τέλος περιγράφεται αναλυτικά η μεθοδολογία των προβλέψεων που ακολουθήθηκε και παρουσιάζονται τα αποτελέσματα των μεθόδων που ελέγχθηκαν.

Λέξεις Κλειδιά: Τεχνικές προβλέψεων, Ιεραρχική Συνάθροιση Χρονοσειρών, Τεχνικές Μηχανικής Μάθησης, Ιεραρχικές Προβλέψεις.

## **Abstract**

This paper examines the sales forecasts of retail stores via the use of hierarchical aggregation methods and machine learning techniques. The aim is to ascertain how well machine learning methods do, in hierarchical data, over the classic forecasting methods.

The study is done in a retail stores chain. The stores are in different geographical areas, so it is deemed necessary to reconcile the individual forecasts in order for the lower level forecasts to sum up to the higher level ones. For this reason hierarchical forecasting methods are used.

First of all, some basic characteristics of forecasts, time series models and data preprocess are introduced. Later, the author analyzes the methods that are used as benchmark as well as the machine learning methods used and the error indicators for the comparison of the methods. After that, hierarchical aggregation of data and the most notable hierarchical forecasting methods are described.

Finally, the methodology that was followed for the forecast methods is described and the results of the study are displayed.

**Key Words:** Forecasting Techniques, Hierarchical Aggregation, Machine Learning Techniques, Hierarchical Forecasting





# Πίνακας Περιεχομένων

Πρόλογος.....	5
Περίληψη.....	6
Abstract .....	7
Πίνακας Περιεχομένων .....	9
Ευρετήριο Εικόνων.....	11
Ευρετήριο Πινάκων .....	12
Κεφάλαιο 1 : Εισαγωγή .....	13
1.1 Αντικείμενο της εργασίας.....	13
1.2 Δομή της εργασίας .....	17
Κεφάλαιο 2 : Τεχνικές Προβλέψεων .....	19
2.1 Γενικά για τις προβλέψεις .....	19
2.2 Χαρακτηριστικά Χρονοσειρών .....	20
2.2.1 Γενικά για τις χρονοσειρές .....	20
2.2.2 Αναπαράσταση και Ποιοτικά Χαρακτηριστικά Χρονοσειρών.....	21
2.3 Κατηγορίες Μεθόδων Πρόβλεψης.....	22
2.3.1 Ποσοτικές Μέθοδοι.....	23
2.4 Μέθοδοι Πρόβλεψης .....	24
2.4.1 Naive.....	25
2.4.2 ETS .....	25
2.4.3 ARIMA (Autoregressive Integrated Moving Average) .....	28
2.4.4 Πολυεπίπεδο Αντίληπτρο(Multi-Layer Perceptrons).....	30
2.5 Δείκτες Σφάλματος.....	38
Κεφάλαιο 3 : Ιεραρχική Συνάθροιση και Ιεραρχικές Προβλέψεις.....	41
3.1 Ιεραρχική Συνάθροιση.....	41
3.2 Ιεραρχικές Χρονοσειρές .....	42
3.3 Bottom-up Μέθοδος.....	44
3.4 Top-down Μέθοδος .....	46
3.5 Optimal Reconciliation Μέθοδος .....	47
Κεφάλαιο 4 : Μεθοδολογική Προσέγγιση .....	49
4.1 Μεθοδολογία .....	49
4.2 Επεξεργασία Δεδομένων.....	49

4.3 Υλοποίηση Μοντέλων Πρόβλεψης .....	50
4.4 Αξιολόγηση Μοντέλων .....	51
Κεφάλαιο 5 : Case Study σε πωλήσεις αλυσίδας καταστημάτων λιανικής .....	53
5.1 Συλλογή και Επεξεργασία Δεδομένων .....	53
5.2 Μέθοδοι Αναφοράς(benchmark).....	57
5.3 Μέθοδοι Μηχανικής Μάθησης.....	61
5.4 Σύγκριση Μεθόδων .....	73
Κεφάλαιο 6 : Αποτελέσματα και Προεκτάσεις .....	75
6.1 Συνοπτικά Αποτελέσματα της Εργασίας .....	75
6.2 Προεκτάσεις .....	76
Βιβλιογραφία .....	79

## Ευρετήριο Εικόνων

Εικόνα 1 : Συνάρτηση ενεργοποίησης λογιστική σιγμοειδής .....	32
Εικόνα 2 : Συνάρτηση ενεργοποίησης υπερβολική εφαπτομένη .....	32
Εικόνα 3 : Συνάρτηση ενεργοποίησης βηματική .....	33
Εικόνα 4 : Συνάρτηση ενεργοποίησης γραμμική .....	34
Εικόνα 5 : Συνάρτηση ενεργοποίησης ReLu.....	35
Εικόνα 6 : Γραφική αναπαράσταση νευρωνικού δικτύου εμπρόσθιας τροφοδότησης.....	35
Εικόνα 7: Διάγραμμα ιεραρχίας δύο επιπέδων .....	42
Εικόνα 8 : Ιεραρχία δεδομένων case study .....	53
Εικόνα 9 : Χρονοσειρά συνολικής ζήτησης .....	54
Εικόνα 10 : Χρονοσειρά ζήτησης Αττικής.....	54
Εικόνα 11 : Χρονοσειρά ζήτησης Ελληνικών νησιών .....	55
Εικόνα 12 : Χρονοσειρά ζήτησης Βόρειας Ελλάδας .....	55
Εικόνα 13 : Χρονοσειρά ζήτησης Νότιας Ελλάδας.....	56
Εικόνα 14 : Χρονοσειρά ζήτησης Θεσσαλονίκης.....	56
Εικόνα 15 : Σφάλματα ανά επίπεδο, μεθόδου top-down με προβλέψεις naive,ets και arima .....	59
Εικόνα 16 : Σφάλματα ανά επίπεδο, μεθόδου bottom-up με προβλέψεις naive,ets και arima .....	59
Εικόνα 17 : Σφάλματα ανά επίπεδο, μεθόδου optimal με προβλέψεις naive,ets και arima .	60
Εικόνα 18 : Σύγκριση naive, ets και arima για κάθε ιεραρχική μέθοδο .....	60
Εικόνα 19 : Σύγκριση σφαλμάτων της top-down για 74 mlr με διαφορετικά input sets .....	64
Εικόνα 20: Σύγκριση σφαλμάτων της bottom-up για 74 mlr με διαφορετικά input sets.....	64
Εικόνα 21: Σύγκριση σφαλμάτων της optimal για 74 mlr με διαφορετικά input sets.....	65
Εικόνα 22 : Σύγκριση σφαλμάτων της top-down για ένα mlr,εκπαιδευμένο από όλα τα δεδομένα, με διαφορετικά input sets.....	67
Εικόνα 23: Σύγκριση σφαλμάτων της bottom-up για ένα mlr,εκπαιδευμένο από όλα τα δεδομένα, με διαφορετικά input sets.....	68
Εικόνα 24: Σύγκριση σφαλμάτων της optimal για ένα mlr,εκπαιδευμένο από όλα τα δεδομένα, με διαφορετικά input sets.....	68
Εικόνα 25 : Σύγκριση σφαλμάτων της top-down για ένα mlr,εκπαιδευμένο από τυχαία δεδομένα, με διαφορετικά input sets.....	70
Εικόνα 26 : Σύγκριση σφαλμάτων της bottom-up για ένα mlr,εκπαιδευμένο από τυχαία δεδομένα, με διαφορετικά input sets.....	71
Εικόνα 27: Σύγκριση σφαλμάτων της optimal για ένα mlr,εκπαιδευμένο από τυχαία δεδομένα, με διαφορετικά input sets.....	72

## Ευρετήριο Πινάκων

Πίνακας 1 : Σφάλματα μεθόδων αναφοράς .....	58
Πίνακας 2 : Σφάλματα για χρήση 74 mlr με tanh.....	62
Πίνακας 3 : Σφάλματα για χρήση 74 mlr με logistic.....	63
Πίνακας 4 : Σφάλματα για χρήση ενός mlr εκπαιδευμένου από όλα τα δεδομένα με tanh.	66
Πίνακας 5 : Σφάλματα για χρήση ενός mlr εκπαιδευμένου από όλα τα δεδομένα με logistic .....	66
Πίνακας 6 : Σφάλματα για χρήση ενός mlr εκπαιδευμένου από τυχαία δεδομένα με tanh.	69
Πίνακας 7 : Σφάλματα για χρήση ενός mlr εκπαιδευμένου από τυχαία δεδομένα με logistic .....	70
Πίνακας 8 : Σύγκριση καλύτερων μεθόδων κάθε προσέγγισης, ανά επίπεδο και μέσο όρο.	73

## **Κεφάλαιο 1 : Εισαγωγή**

### **1.1 Αντικείμενο της εργασίας**

Η παραγωγή προβλέψεων αποτελεί πολύ σημαντικό παράγοντα στις μέρες μας σε διάφορους τομείς. Ένας από αυτούς είναι ο τομέας των επιχειρήσεων και πιο συγκεκριμένα των πωλήσεων τους. Για την παραγωγή σωστών προβλέψεων είναι απαραίτητη η συλλογή και οργάνωση ιστορικών δεδομένων από την εκάστοτε επιχείρηση και η αξιοποίησή τους με τον καλύτερο δυνατό τρόπο.

Τα δεδομένα αυτά, πολλές φορές μπορούν να ενταχθούν σε κάποιου είδους ιεραρχία. Για παράδειγμα μπορούν να χωριστούν οι πωλήσεις μιας επιχείρησης με βάση τη γεωγραφική περιοχή που ανήκουν τα μαγαζιά, ή με βάση κάποιων χαρακτηριστικών των προϊόντων της. Για παράδειγμα ο συνολικός αριθμός των ποδηλάτων μιας εταιρείας μπορεί να χωριστεί σε ποδήλατα δρόμου, βουνού παιδικά και σε άλλες κατηγορίες.

Ας υποθέσουμε λοιπόν ότι μια εταιρεία με δέκα μαγαζιά στην Αττική χρειάζεται να προμηθευτεί προϊόντα για τον επόμενο μήνα. Το κάθε μαγαζί ξεχωριστά κάνει κάποια πρόβλεψη για την ζήτηση που θα έχει, όμως αντίστοιχα κάποια πρόβλεψη γίνεται και από τον υπεύθυνο τομέα για όλη την Αττική. Όπως είναι λογικό θα πρέπει οι προβλέψεις των δέκα αυτών μαγαζιών να αθροίζονται στην πρόβλεψη για την ζήτηση όλης της Αττικής, ώστε να υπάρχει συνέχεια στις προβλέψεις και να γίνει σωστός σχεδιασμός και διανομή των προϊόντων στα αντίστοιχα μαγαζιά.

Με βάση τα προηγούμενα γίνεται εμφανής η ανάγκη για χρήση των ιεραρχικών χρονοσειρών και ιεραρχικών προβλέψεων. Μια χρονοσειρά αποτελείται από διαχρονικές παρατηρήσεις μιας μεταβλητής και στην ουσία δείχνει την εξέλιξη αυτής της μεταβλητής σε ένα χρονικό διάστημα. Ένα παράδειγμα θα μπορούσαν να είναι οι ημερήσιες (ή εβδομαδιαίες, μηνιαίες, κλπ) πωλήσεις ενός μαγαζιού. Οι ιεραρχικές χρονοσειρές αποτελούν μια συλλογή από χρονοσειρές που ανήκουν σε

μια ιεραρχία. Αν υποθέσουμε λοιπόν την προηγούμενη εταιρεία με τα δέκα μαγαζιά στην Αττική, τότε οι ημερήσιες πωλήσεις κάθε μαγαζιού αποτελούν μια χρονοσειρά. Το ίδιο όμως ισχύει και για τις συνολικές ημερήσιες πωλήσεις όλης της Αττικής για την εταιρεία αυτή. Προκύπτουν άρα έντεκα χρονοσειρές που ανήκουν σε μια δομή ιεραρχίας ώστε οι χαμηλότερες χρονοσειρές να αθροίζουν στο υψηλότερο επίπεδο.

Οι ιεραρχικές προβλέψεις αποτελούν την διαδικασία παραγωγής προβλέψεων οι οποίες υπακούνε στον προηγούμενο κανόνα της συνάθροισης ή την διαδικασία "συμφιλίωσης" των ασυνεχών προβλέψεων μιας ιεραρχίας. Στην ουσία επιτρέπουν την παραγωγή προβλέψεων σε κάθε χρονοσειρά ξεχωριστά χωρίς όμως να χάνεται η σχέση που έχουν λόγω της ιεραρχίας που ανήκουν. Δεν πρόκειται για μια μέθοδο πρόβλεψης όπως η *naive* ή τα μοντέλα *arima*, αλλά για ένα σύνολο από τεχνικές που εξασφαλίζουν ότι οι προβλέψεις των χρονοσειρών μιας ιεραρχίας θα αθροίζουν στα εκάστοτε υψηλότερα επίπεδα που συνδέονται.

Γίνεται επομένως προφανής η χρήση των ιεραρχικών προβλέψεων σε όλες τις επιχειρήσεις που εμφανίζουν κάποιου είδους ιεραρχία. Πολύ συνηθισμένο είναι να χρειάζεται οι προβλέψεις κάθε μαγαζιού μιας γεωγραφικής περιοχής να αθροίζουν στις προβλέψεις όλης της περιοχής. Οι κυριότερες μέθοδοι ιεραρχικής συνάθροισης δεδομένων που χρησιμοποιούνται σήμερα είναι οι *bottom-up*, *top-down* και *optimal reconciliation*.

Στην *bottom-up* οι προβλέψεις παράγονται για τις χρονοσειρές του χαμηλότερου επιπέδου, με τις προβλέψεις των επόμενων επιπέδων να προκύπτουν από το άθροισμα των προβλέψεων των χαμηλότερων επιπέδων με βάση τις σχέσεις μεταξύ αυτών όπως προκύπτουν από την ιεραρχία. Πλεονέκτημα είναι πως δεν χάνεται πληροφορία λόγω συνάθροισης των δεδομένων. Αντίθετα αγνοεί τις σχέσεις μεταξύ των χρονοσειρών λόγω ιεραρχίας και συνήθως δεν αποδίδει τόσο καλά στα υψηλότερα επίπεδα της ιεραρχίας. Επίσης είναι φανερό πως η παραγωγή προβλέψεων στο χαμηλότερο επίπεδο απαιτεί περισσότερο χρόνο και πόρους.

Η *top-down* ακολουθεί την ακριβώς αντίθετη διαδικασία από την *bottom-up*, δηλαδή παράγουμε τις προβλέψεις μόνο για την υψηλότερη χρονοσειρά της ιεραρχίας και μετά αυτές χωρίζονται στα χαμηλότερα επίπεδα με βάση κάποιες

αναλογίες(συνήθως προκύπτουν από τα ιστορικά δεδομένα). Αποτελεί την πιο απλή μέθοδο και στην ουσία χρειάζεται μόνο ένα μοντέλο πρόβλεψης αφού παράγονται προβλέψεις μόνο για μια χρονοσειρά. Στα υψηλότερα επίπεδα αποδίδει πολύ καλά όμως λόγω απώλειας πληροφορίας είναι λιγότερο αξιόπιστη όσο κατεβαίνουμε στην ιεραρχία.

Η optimal reconciliation χρησιμοποιεί ανεξάρτητες προβλέψεις όλων των χρονοσειρών, όλων των επιπέδων αξιοποιώντας όλη την πληροφορία και τις σχέσεις που προσφέρει η ιεραρχία. Μπορεί να δώσει πιο ακριβείς και αμερόληπτες προβλέψεις σε σχέση με τις προηγούμενες μεθόδους σε πολλές περιπτώσεις, όμως χρειάζεται τον περισσότερο χρόνο και πόρους που σημαίνει ότι δεν είναι η πιο χρήσιμη επιλογή για πολύ μεγάλο όγκο χρονοσειρών.

Η επιλογή της κατάλληλης μεθόδου ιεραρχικής συνάθροισης εξαρτάται συνήθως από τα δεδομένα μας και τον στόχο που θέλουμε να επιτύχουμε. Αν για παράδειγμα έχουμε μεγάλο όγκο χρονοσειρών αλλά θέλουμε να παράγουμε προβλέψεις μόνο για το υψηλότερο επίπεδο τότε η λογικότερη επιλογή θα ήταν η χρήση της top-down. Αντίστοιχα αν θέλουμε προβλέψεις για το χαμηλότερο επίπεδο πιθανότατα να αποδώσει καλύτερα η bottom-up.

Όσον αφορά την παραγωγή των προβλέψεων πριν από την "συμφιλίωση" τους ώστε να αθροίζουν σωστά στην ιεραρχία, αυτές προκύπτουν με την χρήση κλασσικών μεθόδων και μοντέλων πρόβλεψης. Μπορεί για παράδειγμα να προκύψουν από την μέθοδο naïve ή τα μοντέλα ets και arima. Είναι προφανές ότι όσο πιο αξιόπιστες είναι οι βασικές προβλέψεις, τόσο πιο ακριβείς θα είναι και οι συμφιλιωμένες προβλέψεις της ιεραρχίας. Προκειμένου να υλοποιήσουμε τα κλασσικά μοντέλα πρόβλεψης, συνήθως απαιτείται καλή γνώση της συμπεριφοράς τους, καθώς και γνώση των χαρακτηριστικών των χρονοσειρών, ώστε να γίνει η επιλογή του καλύτερου μοντέλου. Ως αποτέλεσμα η χρήση των μεθόδων αυτών μπορεί να είναι δύσκολη και χρονοβόρα πολλές φορές.

Ένας πιο σύγχρονος τρόπος παραγωγής προβλέψεων είναι με τη χρήση τεχνητών νευρωνικών δικτύων. Πρόκειται ουσιαστικά για μαθηματικά μοντέλα τα οποία λαμβάνοντας κάποιες εισόδους παράγουν κάποιες εξόδους και προσπαθούν να

προσομοιάσουν την λειτουργία των ανθρώπινων νευρώνων. Μεγάλο πλεονέκτημα τους είναι ότι δεν χρειάζεται να κατανοήσουμε πλήρως τη λειτουργία τους για να τα μοντελοποιήσουμε. Επίσης έχει αποδειχθεί ότι λειτουργούν πολύ καλά τόσο σε γραμμικά όσο και σε μη γραμμικά δεδομένα.

Είναι επομένως δυνατή η χρήση των νευρωνικών δικτύων για την παραγωγή των βασικών προβλέψεων και την μετέπειτα "συμφιλίωσή" τους με τις γνωστές ιεραρχικές μεθόδους. Η χρήση των νευρωνικών δικτύων για την παραγωγή των προβλέψεων στην συγκεκριμένη εργασία προκύπτει σαν ιδέα για δύο βασικούς λόγους. Πρώτον, η μεγάλη ικανότητα τους να αναγνωρίσουν μοτίβα και να παράγουν πολύ ακριβείς προβλέψεις σε συνδυασμό με την εύκολη υλοποίηση τους τα καθιστά ανώτερη και πιο αξιόπιστη επιλογή σε σχέση με τις περισσότερες κλασικές στατιστικές μεθόδους πρόβλεψης. Δεύτερον, η εκπαίδευση ενός νευρωνικού δικτύου με εισόδους από όλες τις χρονοσειρές της ιεραρχίας και η χρήση αυτού για την παραγωγή των προβλέψεων προσφέρει πολλά πλεονεκτήματα. Αρχικά εκπαιδεύοντας το δίκτυο από όλες τις χρονοσειρές, ενσωματώνουμε έμμεσα πληροφορία από όλη την ιεραρχία σε ένα ενιαίο μοντέλο πριν καν παράγουμε τις βασικές προβλέψεις, ενώ με τις κλασικές μεθόδους αυτό θα γινόταν μόνο μετά την ιεραρχική "συμφιλίωση" των προβλέψεων. Ακόμα μέθοδοι όπως η bottom-up και optimal που χρειάζονταν ένα μοντέλο πρόβλεψης για κάθε χρονοσειρά ξεχωριστά, με αυτή την προσέγγιση χρειάζονται μόνο ένα ενιαίο μοντέλο για όλες τις χρονοσειρές. Είναι φανερό επομένως, πως όχι μόνο βελτιώνεται η απόδοση των προβλέψεων και αξιοποιείται καλύτερα η σχέση των χρονοσειρών στην ιεραρχία, αλλά μειώνεται και δραστικά η χρήση πόρων και χρόνου για την υλοποίηση των μοντέλων πρόβλεψης και παραγωγής των προβλέψεων.

Στην παρούσα διπλωματική γίνεται μια προσπάθεια συνδυασμού των δύο μεθόδων σε δεδομένα πωλήσεων μιας αλυσίδας καταστημάτων λιανικής. Σκοπός είναι να δούμε πόσο καλά αποδίδει η παραγωγή των βασικών προβλέψεων των χρονοσειρών με χρήση νευρωνικών δικτύων και η συμφιλίωση τους με ιεραρχικές μεθόδους, σε σύγκριση με τις κλασικές τεχνικές όπου οι βασικές προβλέψεις προκύπτουν με στατιστικά μοντέλα. Πραγματοποιείται ένα case-study σε



πραγματικά δεδομένα και εξετάζονται τέσσερις διαφορετικές προσεγγίσεις για την παραγωγή προβλέψεων. Η πρώτη αφορά τις κλασικές μεθόδους που χρησιμοποιούνται και θα λειτουργήσει ως σημείο αναφοράς, ενώ οι άλλες τρεις αφορούν την χρήση των νευρωνικών δικτύων. Η μία εξετάζει την εκπαίδευση ενός νευρωνικού για κάθε χρονοσειρά, η οποία αντικαθιστά την αντίστοιχη χρήση ενός στατιστικού μοντέλου για κάθε μία, ενώ οι τελευταίες δύο προσεγγίσεις αφορούν τη χρήση ενός νευρωνικού δικτύου για όλες τις χρονοσειρές. Οι δύο τελευταίες προσεγγίσεις αυτές αποτελούν και την κυριότερη ιδέα της εργασίας.

## 1.2 Δομή της εργασίας

Στο πρώτο κεφάλαιο γίνεται μια εισαγωγή του θέματος της παρούσας διπλωματικής εργασίας και της βασικής δομής των κεφαλαίων της.

Στο δεύτερο κεφάλαιο περιγράφονται οι τεχνικές προβλέψεων και τα μοντέλα χρονοσειρών, καθώς και τα βασικά χαρακτηριστικά τους. Αναλύονται οι κύριες μέθοδοι πρόβλεψης που χρησιμοποιήθηκαν στην παρούσα διπλωματική (naive,ets, ARIMA,MLP), ενώ το τέλος του κεφαλαίου αφιερώνεται στα στατιστικά σφάλματα που χρησιμοποιούνται για την αξιολόγηση των μεθόδων πρόβλεψης.

Στο τρίτο κεφάλαιο περιγράφεται η ιεραρχική συνάθροιση δεδομένων και οι ιεραρχικές χρονοσειρές. Παρουσιάζονται αναλυτικά οι κύριες μέθοδοι παραγωγής ιεραρχικών προβλέψεων που χρησιμοποιούνται στην διπλωματική, δηλαδή οι top-down, bottom-up και optimal reconciliation.

Το τέταρτο κεφάλαιο παρουσιάζει την μεθοδολογία που ακολουθήθηκε στην διπλωματική. Γίνεται μια αναφορά στην επεξεργασία των δεδομένων, την υλοποίηση των μοντέλων πρόβλεψης και τη διαδικασία σύγκρισης των μεθόδων.

Στο πέμπτο κεφάλαιο παρουσιάζεται ένα case-study σε πραγματικά δεδομένα για τις πωλήσεις μιας αλυσίδας καταστημάτων λιανικής. Στόχος είναι η ανάδειξη της αποτελεσματικότητας των μεθόδων που εξετάστηκαν, μέσω της αξιολόγησης τους

από πραγματικά δεδομένα. Παρουσιάζεται μία προσέγγιση που ακολουθεί τις κλασικές ιεραρχικές προβλέψεις και χρησιμοποιείται ως σημείο αναφοράς, καθώς και τρεις προσεγγίσεις που αφορούν την χρήση πολυεπίπεδων αντίληπτρων για την παραγωγή των προβλέψεων.

Στο τελευταίο κεφάλαιο παρατίθενται τα κύρια συμπεράσματα που προέκυψαν με βάση τα αποτελέσματα των πειραμάτων και προτείνονται πιθανά θέματα για μελέτη και ανάλυση στο μέλλον.

## Κεφάλαιο 2 : Τεχνικές Προβλέψεων

### 2.1 Γενικά για τις προβλέψεις

Από τα αρχαία χρόνια η παραγωγή προβλέψεων βοηθούσε τον άνθρωπο στη λήψη αποφάσεων , τόσο σε απλά θέματα όπως η πρόβλεψη του καιρού όσο σε πιο περίπλοκα ,όπως το αν θα κάνει μία επένδυση. Μπορεί οι δύο περιπτώσεις να μην είναι το ίδιο σημαντικές όμως αυτό που μας ενδιαφέρει και στις δύο είναι να μειωθεί η απόκλιση της πρόβλεψης μας από το πραγματικό αποτέλεσμα όσο το δυνατόν περισσότερο. Οι προβλέψεις αυτές μπορούν είτε να είναι εμπειρικές ή να βασίζονται σε ένα σύνολο μοντέλων και μεθόδων. Έτσι λοιπόν προέκυψε η ανάπτυξη του τομέα των προβλέψεων τόσο σε προσωπικό επίπεδο, όσο και σε ακαδημαϊκό επίπεδο αλλά και επίπεδο επιχειρήσεων.

Η ανάγκη για προβλέψεις οφείλεται κυρίως στην αβεβαιότητα. Με τον όρο αβεβαιότητα εννοούμε τόσο τις μικρές αποκλίσεις από την συνηθισμένη πορεία των γεγονότων, όσο και εντελώς απρόβλεπτα γεγονότα, τα οποία επηρεάζουν σε υπερβολικό βαθμό την αναμενόμενη έκβαση των πραγμάτων. Όπως είναι φανερό η αβεβαιότητα αποτελεί τον χειρότερο εχθρό των προβλέψεων και έχει μελετηθεί από τον Σπύρο Μακρυδάκη στο βιβλίο <<Χορεύοντας με την Τύχη>> στο οποίο χρησιμοποιεί δύο διαφορετικές κατηγορίες, την <<αβεβαιότητα του μετρώ>> και την <<αβεβαιότητα της καρύδας>>. Η πρώτη αφορά τις μικρές αποκλίσεις στις προβλέψεις που οφείλονται σε κάποιο τυχαίο γεγονός όπως η καθυστέρηση του μετρώ γιατί άργησαν να κλείσουν οι πόρτες στην προηγούμενη στάση, ενώ η δεύτερη κατηγορία αφορά τα εντελώς απρόβλεπτα γεγονότα ,όπως το να περπατάει κάποιος σε μια εξωτική παραλία και να του πέσει μια καρύδα στο κεφάλι. Επομένως, σκοπός των επιστημονικών είναι να μειώσουν όσο το δυνατόν περισσότερο τις αποκλίσεις στα προβλήματα της πρώτης κατηγορίας, δηλαδή σαν αυτά του μετρώ.

Οι αβεβαιότητες αυτές και κυρίως της κατηγορίας της καρύδας δημιουργούν προβλήματα στα μοντέλα πρόβλεψης με αποτέλεσμα ο τομέας των προβλέψεων να αντιμετωπίζεται συχνά με κριτική και καχυποψία. Σε αντίθεση όμως, όσο πιο

απρόβλεπτο και περίπλοκο γίνεται το περιβάλλον τόσο περισσότερο δυσκολεύεται ο άνθρωπος να παράγει κάποια πρόβλεψη μόνος του και γίνεται επιτακτική η ανάγκη των συστηματικών μεθόδων πρόβλεψης. Στο πλαίσιο αυτής της σύγχρονης ανάγκης αναπτύχθηκε η παρούσα εργασία, η οποία ασχολείται με την παραγωγή προβλέψεων ζήτησης σε καταστήματα λιανικής

## **2.2 Χαρακτηριστικά Χρονοσειρών**

### **2.2.1 Γενικά για τις χρονοσειρές**

Για την παραγωγή προβλέψεων είναι πολύ σημαντική η συλλογή και οργάνωση ιστορικών δεδομένων που αφορούν το αντικείμενο που προσπαθούμε να προβλέψουμε. Είναι φανερό πως όσο πιο έγκυρα, ακριβή και πρόσφατα είναι αυτά τα δεδομένα τόσο καλύτερα αποτελέσματα θα έχει μια μέθοδος πρόβλεψης.

Τα ιστορικά αυτά δεδομένα μπορούν να έχουν πολλές μορφές, η κυριότερη των οποίων είναι οι χρονοσειρές τις οποίες θα χρησιμοποιήσουμε και στην συγκεκριμένη διπλωματική. Οι χρονοσειρές αφορούν διαχρονικές παρατηρήσεις μίας συγκεκριμένης μεταβλητής και στην ουσία δείχνουν την εξέλιξη αυτής της μεταβλητής σε ένα χρονικό διάστημα.

Οι χρονοσειρές μπορούν να χωριστούν σε δύο κατηγορίες, τις ντετερμινιστικές και τις στοχαστικές. Στην πρώτη κατηγορία οι παρατηρήσεις είναι μεταξύ τους εξαρτημένες και γνωρίζοντας την σχέση που τις συνδέει μπορούμε να υπολογίσουμε ακριβώς τις μελλοντικές τιμές. Αντιθέτως, στην δεύτερη κατηγορία οι παρατηρήσεις συνδέονται από μια στοχαστική διαδικασία και οι μελλοντικές τιμές τους δεν μπορούν να προσδιοριστούν πλήρως. Στην πραγματικότητα παρατηρούμε μόνο στοχαστικές χρονοσειρές αφού όλα τα φυσικά μεγέθη χαρακτηρίζονται από τον παράγοντα της τύχης και στην ουσία μόνο για αυτές έχει νόημα η παραγωγή προβλέψεων, αφού στις ντετερμινιστικές η πορεία τους είναι προκαθορισμένη. Σε όποια κατηγορία όμως και αν ανήκουν οι χρονοσειρές είναι χρήσιμο να κατανοήσουμε τους παράγοντες από τους οποίους εξαρτώνται αυτές, ώστε να παράγουμε προβλέψεις.

## 2.2.2 Αναπαράσταση και Ποιοτικά Χαρακτηριστικά Χρονοσειρών

Οι χρονοσειρές ανάλογα τη γραφική τους αναπαράσταση μπορούν να χωριστούν σε τρεις κατηγορίες. Η πρώτη κατηγορία είναι τα διαγράμματα χρόνου, τα οποία απεικονίζουν τις τιμές μιας μεταβλητής στην πάροδο του χρόνου. Είναι εύκολα στην κατανόηση των δεδομένων και τα πιο διαδεδομένα ειδικά στον χώρο των επιχειρήσεων. Στη δεύτερη κατηγορία ανήκουν τα εποχιακά διαγράμματα τα οποία χρησιμοποιούνται για χρονοσειρές που εμφανίζουν εποχικότητα. Ένα παράδειγμα είναι οι πωλήσεις μιας εταιρείας καλοκαιρινών ειδών καθ' όλη τη διάρκεια του χρόνου. Τέλος, υπάρχουν τα διαγράμματα διασποράς, τα οποία χρησιμεύουν στην συσχέτιση δύο διαφορετικών μεγεθών ή χαρακτηριστικών που αφορούν τα δεδομένα μας. Ένα παράδειγμα είναι η θερμοκρασία με την πώληση σκούφων, ή η σύγκριση δύο προϊόντων της ίδιας εταιρείας.

Όλες οι χρονοσειρές, ανεξαρτήτως τύπου, διαθέτουν ορισμένα βασικά χαρακτηριστικά που μας βοηθάνε στην κατανόηση και επεξεργασία τους. Αυτά είναι η τάση, η εποχικότητα, η κυκλικότητα και η τυχαιότητα.

Με τον όρο τάση αναφερόμαστε στην μακροπρόθεσμη μεταβολή του μέσου επιπέδου των τιμών μιας χρονοσειράς. Η τάση μπορεί να είναι ανοδική, καθοδική ή και σταθερή και συνήθως εμφανίζεται σαν μια μορφή καμπύλης ή σπανιότερα ευθείας. Το πότε η μεταβολή αυτή θα χαρακτηριστεί ως μακροπρόθεσμη εξαρτάται κάθε φορά από τα δεδομένα, αλλά το σίγουρο είναι πως πρέπει πάντα να υπάρχει ικανοποιητικός αριθμός δεδομένων ώστε να βγάλουμε σωστά συμπεράσματα.

Η εποχικότητα εκφράζει μια περιοδική διακύμανση σταθερού και μικρότερου ή ίσου μήκους από ένα έτος. Είναι εύκολα αναγνωρίσιμη σαν ένα επαναλαμβανόμενο μοτίβο στην χρονοσειρά και μπορεί να αντιμετωπιστεί υπολογίζοντας τους δείκτες εποχικότητας για τα αντίστοιχα χρονικά και τη διαίρεση τους με τα πραγματικά δεδομένα, κάτι που θα οδηγήσει σε μια νέα χρονοσειρά που ονομάζεται αποεποχικοποιημένη.

Η κυκλικότητα είναι η μεταβολή που οφείλεται σε εξωγενείς παράγοντες κατά μεγάλες χρονικές περιόδους. Το μήκος των περιόδων που εμφανίζεται η κυκλικότητα δεν είναι πάντα σταθερό, όμως είναι πάντα μεγαλύτερο του ενός έτους. Παρατηρείται κυρίως σε οικονομικές χρονοσειρές εξαιτίας των ανόδων και των υφέσεων που παρουσιάζουν παγκοσμίως οι οικονομίες.

Το τελευταίο χαρακτηριστικό, δηλαδή η τυχαιότητα είναι κάτι στοχαστικό και ουσιαστικά αποτελεί την διαφορά ανάμεσα στην συνδυασμένη επίδραση των τριών πρώτων χαρακτηριστικών και των πραγματικών δεδομένων.

Πέρα από αυτά τα χαρακτηριστικά οι χρονοσειρές πολλές φορές εμφανίζουν απότομες αλλαγές, η πρόβλεψη των οποίων είναι αδύνατη μόνο με τη χρήση ιστορικών δεδομένων. Αυτές οι αλλαγές ονομάζονται ασυνέχειες και χωρίζονται σε special events(ή outliers) και σε level shifts.

Τα special events επηρεάζουν παροδικά τις χρονοσειρές και συνήθως οφείλονται σε απρόβλεπτα γεγονότα για αυτό και είναι δύσκολο να αναγνωριστούν και να αφαιρεθούν από τα εκάστοτε ιστορικά δεδομένα.

Αντιθέτως τα level shifts επηρεάζουν μόνιμα την πορεία των δεδομένων και παρόλο που αυτά συνεχίζουν να εξελίσσονται στο ίδιο μοτίβο, αλλάζει το επίπεδο στο οποίο βρισκόντουσαν. Ένα παράδειγμα θα μπορούσε να είναι η αύξηση των πωλήσεων λόγω του κλεισίματος μιας ανταγωνιστικής επιχείρησης.

## 2.3 Κατηγορίες Μεθόδων Πρόβλεψης

Οι μέθοδοι πρόβλεψης μπορούν να χωριστούν σε τρεις μεγάλες κατηγορίες, τις ποσοτικές(quantitative), τις κριτικές(judgmental) και τις τεχνολογικές(technological).

Οι ποσοτικές μέθοδοι, που είναι και αυτές που θα χρησιμοποιήσουμε στην διπλωματική αυτή, χρησιμοποιούν κυρίως την στατιστική και τα μαθηματικά μοντέλα. Τα μαθηματικά μοντέλα όμως δεν μπορούν να προβλέψουν σωστά όταν στις χρονοσειρές εμφανίζονται special events ή απότομες αλλαγές και εδώ είναι που χρειάζονται οι κριτικές προβέψεις οι οποίες βασίζονται στις γνώσεις, την εμπειρία και την κριτική ικανότητα αυτού που προβλέπει. Τέλος οι τεχνολογικές χρησιμοποιούνται για μακροπρόθεσμες προβλέψεις σε οικονομικά, τεχνολογικά,

πολιτικά και κοινωνικά θέματα. Χωρίζονται σε διερευνητικές, που ξεκινούν από το παρελθόν ή το παρόν και εξετάζουν όλα τα πιθανά σενάρια για να οδηγηθούν στο μέλλον και σε κανονιστικές, που πρώτα προσδιορίζουν όλους τους μελλοντικούς στόχους και μετά ελέγχουν την δυνατότητα επίτευξής τους με βάση τα δεδομένα.

### **2.3.1 Ποσοτικές Μέθοδοι**

Οι ποσοτικές μέθοδοι μπορούν να χωριστούν σε δύο μεγάλες κατηγορίες, το μοντέλο χρονοσειρών, που είναι και αυτό που μας ενδιαφέρει και το αιτιοκρατικό μοντέλο. Για το μοντέλο χρονοσειρών, τα δεδομένα πρέπει να έχουν στοιχεία ανά ίδια χρονικά διαστήματα και να ακολουθούν κάποιο σταθερό και επαναλαμβανόμενο πρότυπο στο οποίο θα βασιστεί η πρόβλεψη. Οι μέθοδοι χρονοσειρών χωρίζονται περαιτέρω σε μεθόδους εξομάλυνσης, αποσύνθεσης και αυτοπαλινδρομικές. Όσον αφορά το αιτιοκρατικό μοντέλο, υπάρχει μια σταθερή σχέση μεταξύ του μεγέθους που εξετάζουμε και κάποιων παραμέτρων, στις οποίες αν αντικαταστήσουμε τις νέες τιμές τους παράγουμε πρόβλεψη για το ζητούμενο μέγεθος.

#### ***2.3.1.1 Μέθοδοι Εξομάλυνσης***

Βασικός στόχος μιας μεθόδου εξομάλυνσης είναι να εξομαλύνει το πρότυπο που βασίζονται τα ιστορικά μας δεδομένα και έπειτα με βάση αυτό το μοντέλο να παράγουμε προβλέψεις, κυρίως βραχυπρόθεσμες. Η μέθοδος αυτή μπορεί να πραγματοποιηθεί είτε με κινητούς μέσους όρους, όπου όλες οι τιμές των δεδομένων συμμετέχουν ισάξια στην εξομάλυνση, είτε με κάποια εκθετική μέθοδο με την οποία δίνεται η ανάλογη βαρύτητα στις πιο παλιές και πιο πρόσφατες παρατηρήσεις.

### **2.3.1.2 Μέθοδοι Αποσύνθεσης**

Στόχος των μεθόδων αποσύνθεσης είναι να απομονωθούν τα βασικά χαρακτηριστικά των χρονοσειρών, δηλαδή η τάση, η εποχικότητα, η κυκλικότητα και η τυχαιότητα. Οι κυριότερες μέθοδοι αποσύνθεσης είναι οι εξής:

- \*Σταθερή Προσθετική Μέθοδος(Fixed Additive Method)
- \*Σταθερή Πολλαπλασιαστική Μέθοδος(Fixed Multiplicative Method)
- \*Κινητή Προσθετική Μέθοδος(Moving Additive Method)
- \*Κινητή Πολλαπλασιαστική Μέθοδος(Moving Multiplicative Method)

### **2.3.1.3 Αυτοπαλινδρομικές Μέθοδοι Κινητού Μέσου Όρου**

Οι μέθοδοι αυτές βασίζονται στην αλληλεξάρτηση που παρουσιάζουν οι τιμές των χρονοσειρών σε διάφορες χρονικές στιγμές και πρόκειται ουσιαστικά για στοχαστικά μοντέλα, τα οποία περιγράφουν την εξέλιξη στο χρόνο του μεγέθους που εξετάζουμε. Τα μοντέλα αυτά συνδυάζουν γραμμικά την τυχαιότητα, τις ιστορικές τιμές των δεδομένων και άλλους στοχαστικούς παράγοντες.

## **2.4 Μέθοδοι Πρόβλεψης**

Υπάρχουν πάρα πολλές μέθοδοι πρόβλεψης στην γνωστή βιβλιογραφία, όμως εμείς θα αναλύσουμε αυτές που χρησιμοποιήθηκαν στην παρούσα διπλωματική. Αυτές είναι οι εξής :

- 1.Naive (Απλοϊκή μέθοδος)
- 2.ETS
- 3.ARIMA
- 4.Πολυεπίπεδο Αντίληπτρο-Multi-layer Perceptrons (MLP)

Προφανώς θα μπορούσαν να συμμετέχουν και άλλες μέθοδοι πρόβλεψης στην εργασία όμως ο σκοπός των τριών πρώτων ήταν να λειτουργήσουν σαν σημείο



αναφοράς. Αντίστοιχα θα μπορούσαμε να χρησιμοποιήσουμε διαφορετικές μεθόδους μηχανικής μάθησης όμως δεν ψάχνουμε να βρούμε ποια δουλεύει καλύτερα, αλλά να εξετάσουμε τον συνδυασμό τους με τις ιεραρχικές μεθόδους πρόβλεψης. Από εκεί και πέρα κάποιος μπορεί να εξετάσει τη χρήση άλλων μεθόδων για να βελτιώσει τα αποτελέσματα του.

### 2.4.1 Naive

Πρόκειται για την πιο απλή μέθοδο πρόβλεψης και λόγω της απλότητας της χρησιμοποιείται πιο πολύ σαν σημείο αναφοράς και όχι σαν κύρια μέθοδος πρόβλεψης. Η πρόβλεψη που δίνει για την επόμενη χρονική περίοδο είναι η τιμή που είχε το μέγεθος την ακριβώς προηγούμενη περίοδο, δηλαδή :

$$F_t = Y_{t-1}$$

Αν θέλουμε να παράγουμε προβλέψεις για περισσότερες χρονικές περιόδους τότε όλες οι μελλοντικές τιμές ισούνται με την τελευταία διαθέσιμη ιστορική παρατήρηση. Οι προβλέψεις της συνήθως είναι ανακριβείς, εκτός από πολύ συγκεκριμένες περιπτώσεις, και για αυτό χρησιμοποιείται ως μέθοδος αναφοράς(benchmark).

### 2.4.2 ETS

Η ETS είναι ένα State Space Model και ουσιαστικά πρόκειται για πολλά μοντέλα, το καθένα από τα οποία αποτελείται από μια εξίσωση που περιγράφει τα παρατηρούμενα δεδομένα και κάποιες εξισώσεις κατάστασης που περιγράφουν χαρακτηριστικά που αλλάζουν με τον χρόνο (level,trend,seasonal). Το όνομα ETS είναι αρχικά για τα Error, Trend, Seasonal τα οποία είναι τα στοιχεία που χρησιμοποιεί η μέθοδος για την παραγωγή προβλέψεων. Για κάθε μέθοδο υπάρχουν δύο μοντέλα, ένα με προσθετικά σφάλματα και ένα με πολλαπλασιαστικά. Αντίστοιχα η εποχιακότητα μπορεί να είναι προσθετική,

πολλαπλασιαστική ή να μην λαμβάνεται καθόλου υπόψη. Τέλος όσον αφορά την τάση μπορεί να μην λαμβάνεται καθόλου υπόψη (None), μπορεί να είναι προσθετική (Additive), προσθετική φθίνουσα (Additive Damped), πολλαπλασιαστική (Multiplicative) και πολλαπλασιαστική φθίνουσα (Multiplicative Damped). Οι αντίστοιχοι συμβολισμοί για τα σφάλματα είναι  $Error=\{A,M\}$ , για την εποχιακότητα  $Seasonal=\{N,A,M\}$  και για την τάση  $Trend=\{N,A,A_d,M,M_d\}$ . Η πραγματική τιμή μπορεί να γραφτεί ως η τιμή του προηγούμενου επιπέδου συν ένα σφάλμα, οπότε προκύπτει η εξίσωση:

$$Y_t = S_{t-1} + e_t$$

Για ένα state space model αρκεί να προσδιοριστεί η κατανομή πιθανοτήτων για  $e_t$ . Αν στο μοντέλο με προσθετικά σφάλματα υποτεθεί ότι τα σφάλματα ενός βήματος κατανέμονται κανονικά με μέση τιμή 0 και διακύμανση  $\sigma^2$  μπορούμε να χρησιμοποιήσουμε την συντομογραφία  $e_t = \varepsilon_t \sim NID(0, \sigma^2)$ , όπου NID (Normally and Independently Distributed) κατανεμημένα κανονικά και ανεξάρτητα. Προκύπτουν επομένως οι εξισώσεις:

$$Y_t = S_{t-1} + \varepsilon_t \quad (1)$$

$$S_t = S_{t-1} + \alpha \cdot \varepsilon_t \quad (2)$$

Η εξίσωση (1) ονομάζεται εξίσωση μέτρησης ενώ η (2) εξίσωση κατάστασης ή μετάβασης. Με αυτές τις εξισώσεις και την κατανομή των σφαλμάτων καθορίζεται πλήρως ένα μοντέλο. Η εξίσωση μέτρησης εκφράζει τη σχέση μεταξύ παρατηρήσεων και μη παρατηρούμενων καταστάσεων. Στην περίπτωση αυτή, η παρατήρηση  $Y_t$  είναι γραμμική συνάρτηση του επιπέδου  $S_t$  και το τυχαίο σφάλμα  $\varepsilon_t$ , το απρόβλεπτο μέρος του  $Y_t$ . Για άλλα μοντέλα αυτή η σχέση μπορεί να είναι μη γραμμική. Η εξίσωση κατάστασης δείχνει την εξέλιξη της κατάστασης στον χρόνο. Όσο υψηλότερη είναι η τιμή του  $\alpha$  τόσο γρηγορότερες θα είναι και οι μεταβολές του επιπέδου. Όμοια μπορούμε να φτιάξουμε το μοντέλο με πολλαπλασιαστικά σφάλματα, γράφοντας τα τυχαία σφάλματα σαν σχετικά σφάλματα, οπότε έχουμε:

$$\varepsilon_t = Y_t - F_t \quad \text{όπου } \varepsilon_t \sim NID(0, \sigma^2)$$

και αντικαθιστώντας  $F_t = S_{t-1}$  παίρνουμε:

$$Y_t = S_{t-1} + S_{t-1} \cdot \varepsilon_t \quad (1)$$

$$e_t = Y_t - F_t = S_{t-1} \cdot \varepsilon_t \quad (2)$$

και μπορούμε να γράψουμε την πολλαπλασιαστική μορφή του μοντέλου ως:

$$Y_t = S_{t-1} \cdot (1 + \varepsilon_t) S_t = S_{t-1} \cdot (1 + \alpha \cdot \varepsilon_t)$$

### Προσδιορισμός Παραμέτρων

Προκειμένου να υπολογίσουμε τις παραμέτρους του μοντέλου, μπορούμε εκτός από την ελαχιστοποίηση του αθροίσματος των τετραγώνων των σφαλμάτων, να μεγιστοποιήσουμε την "πιθανότητα". Με τον όρο αυτό εννοούμε την πιθανότητα των δεδομένων να προκύψουν από το καθορισμένο μοντέλο. Επομένως μεγάλη πιθανότητα συνεπάγεται και καλύτερο μοντέλο. Οι παράμετροι εξομάλυνσης και οι αρχικές καταστάσεις υπολογίζονται με βάση τη μεγιστοποίηση αυτής της πιθανότητας. Ένα πλεονέκτημα τη ETS είναι ότι μπορούν να χρησιμοποιηθούν ορισμένα κριτήρια πληροφορίας για τον καθορισμό των καλύτερων μοντέλων για μια συγκεκριμένη χρονοσειρά. Συγκεκριμένα τα κριτήρια αυτά είναι :

-Το κριτήριο πληροφορίας του Akaike ή αλλιώς Akaike's Information Criteria (AIC) το οποίο ορίζεται ως

$$AIC = -2\log(L) + 2k$$

όπου L είναι η πιθανότητα του μοντέλου και k ο συνολικός αριθμός των παραμέτρων και αρχικών καταστάσεων που έχουν εκτιμηθεί.

-Το διορθωμένο κριτήριο του Akaike για μικρή προκατάληψη ή αλλιώς AIC<sub>c</sub> ορίζεται

ως εξής :

$$AIC_c = AIC + \frac{2k(k+1)}{T-k-1}$$

-Το μπεϋζιανό κριτήριο πληροφορίας ή BIC

$$BIC = AIC + k[\log(T) - 2]$$

Τρεις συνδυασμοί των (ERROR, TREND, SEASONAL) μπορούν να οδηγήσουν σε υπολογιστικές δυσκολίες λόγω διαίρεσης με τιμές που είναι πολύ κοντά στο μηδέν στις εξισώσεις κατάστασης. Συνήθως αποφεύγουμε αυτούς τους συνδυασμούς ,όταν επιλέγουμε ένα μοντέλο, οι οποίοι είναι οι ETS(A,N,M), ETS(A,A,M) και ETS(A,A<sub>d</sub>,M).

Τα μοντέλα με πολλαπλασιαστικά σφάλματα είναι χρήσιμα όταν τα δεδομένα είναι αυστηρά θετικά, αλλά όχι αριθμητικά σταθερά όταν τα δεδομένα περιέχουν

μηδενικές ή αρνητικές τιμές. Επομένως αν τα δεδομένα δεν είναι αυστηρά θετικά δεν λαμβάνονται υπόψη τα μοντέλα πολλαπλασιαστικών σφαλμάτων και χρησιμοποιούνται μόνο τα 6 μοντέλα προσθετικών σφαλμάτων.

### 2.4.3 ARIMA (Autoregressive Integrated Moving Average)

Μία ακόμα μέθοδος πρόβλεψης είναι τα ολοκληρωμένα αυτοπαλινδρομικά μοντέλα κινητού μέσου όρου. Εφαρμόζονται σε περιπτώσεις που τα δεδομένα παρουσιάζουν ενδείξεις μη στασιμότητας η οποία μπορεί να απομακρυνθεί με ένα πρώτο βήμα διαφόρισης(η οποία αναφέρεται στο "ολοκληρωμένο" μέρος του μοντέλου). Τα μοντέλα αυτά συμβολίζονται συνήθως ως ARIMA(p,d,q) όπου οι παράμετροι p,d,q είναι μη αρνητικοί ακέραιοι και αναφέρονται αντίστοιχα στο αυτοπαλινδρομικό μέρος, ολοκληρωμένο μέρος και μέρος κινητού μέσου όρου του μοντέλου. Όταν κάποιες παράμετροι είναι μηδενικές συνηθίζεται να παραλείπονται τα αντίστοιχα γράμματα από το ακρωνύμιο ARIMA οπότε για παράδειγμα ισχύει ARIMA(1,0,0) -->AR(1),ARIMA(0,1,0)-->I(1),ARIMA(0,0,1)-->MA(1). Τα μοντέλα ARIMA είναι στοχαστικά μοντέλα και βασίζονται στον υπολογισμό της πιθανότητας για την οποία η τιμή του μεγέθους βρίσκεται σε κάποιο διάστημα.

Προκειμένου να εφαρμοστούν τα μοντέλα ARIMA είναι αναγκαία μια επεξεργασία των δεδομένων ώστε η χρονοσειρά προς μελέτη να είναι στάσιμη, δηλαδή να έχει σταθερή μέση τιμή, διακύμανση και συνάρτηση αυτοσυσχέτισης στην πάροδο του χρόνου. Για να συμβούν αυτά χρειάζεται :

- Μετασχηματισμός των δεδομένων. Γίνεται για να περιοριστεί η τυχαιότητα της χρονοσειράς και των τυχόν ακραίων τιμών που υπάρχουν, δημιουργώντας έτσι μια νέα χρονοσειρά μικρότερης και πιο σταθερής διακύμανσης. Ως αποτέλεσμα το επίπεδο της χρονοσειράς δεν μεταβάλλεται τόσο απότομα και μπορεί να υποτεθεί στασιμότητα. Ο πιο απλός και διαδεδομένος μετασχηματισμός είναι η λογαρίθμηση της χρονοσειράς,η οποία δίνει αρκετά καλά αποτελέσματα με μικρό υπολογιστικό κόστος.

- Διαφόριση(differencing) της χρονοσειράς. Χρησιμοποιείται για να περιορίσει τις διακυμάνσεις επιπέδου αφαιρώντας τάση και εποχιακότητα. Προκύπτει και εδώ μια νέα χρονοσειρά με σταθερότερο επίπεδο και διακύμανση. Όταν διαφορίζεται μια χρονοσειρά η παρατηρήσεων τότε η νέα χρονοσειρά θα έχει στοιχεία τις διαφορές των παρατηρήσεων της πρώτης. Άρα :

$$1\eta \text{ τάξη: } Y'_t = Y_t - Y_{t-1}, 2\eta \text{ τάξη: } Y''_t = Y'_t - Y'_{t-1} = Y_t - 2Y_{t-1} - Y_{t-2} \text{ κ.ο.κ.}$$

- Εποχιακή διαφορίση σε χρονοσειρές με έντονη εποχιακότητα. Στην προκειμένη περίπτωση τα στοιχεία της νέας χρονοσειράς προκύπτουν απο τις διαφορές μεταξύ των παρατηρήσεων της αρχικής χρονοσειράς και των προηγούμενων αντίστοιχων εποχιακών περιόδων. Άρα:

$$1\eta \text{ τάξη: } Y'_t = Y_t - Y_{t-m}, 2\eta \text{ τάξη: } Y''_t = Y'_t - Y'_{t-m} = Y_t - 2Y_{t-m} - Y_{t-2m} \text{ κ.ο.κ. με } m \text{ την περίοδο εποχιακότητας.}$$

Έπειτα γίνεται η εκτίμηση των μοντέλων:

- Μοντέλα αυτοπαλινδρόμησης AR(p). Τα μοντέλα αυτά θεωρούν γραμμικές σχέσεις ανάμεσα στις παρατηρήσεις της χρονοσειράς και τις χρησιμοποιούν για την περιγραφή της και την παραγωγή προβλέψεων. Η παράμετρος p είναι η τάξη του μοντέλου, δηλαδή το πλήθος των παρελθοντικών παρατηρήσεων που χρησιμοποιούνται για την πρόβλεψη. Αλγεβρικά προκύπτει:

$$y_t = c + \phi_1 y_{t-1} + \phi_2 y_{t-2} + \dots + \phi_p y_{t-p}, \text{ όπου } \phi_i \text{ οι συντελεστές αυτοσυσχέτισης του μοντέλου για υστέρηση } i \text{ και } c \text{ μία σταθερά με } c = \mu(1 - \phi_1 - \dots - \phi_p).$$

- Μοντέλα κινητού μέσου όρου MA(q). Τα μοντέλα αυτά θεωρούν γραμμικές σχέσεις ανάμεσα στην παρατήρηση της χρονοσειράς που εξετάζεται και στα σφάλματα που εμφάνισε το μοντέλο MA σε προηγούμενες περιόδους. Παριστάνεται αλγεβρικά ως εξής :

$$y_t = c + \theta_1 e_{t-1} + \theta_2 e_{t-2} + \dots + \theta_q e_{t-q},$$

όπου  $\theta_i$  οι συντελεστές μερικής αυτοσυσχέτισης του μοντέλου MA για υστέρηση  $i$  και  $c$  μια σταθερά.

Τα μοντέλα AR και MA είναι δυνατόν να συνδυαστούν για την ανάλυση και πρόβλεψη στάσιμων χρονοσειρών. Χρησιμοποιώντας επομένως και τα μοντέλα

διαφόρισης  $I$  για να εξασφαλιστεί η στασιμότητα, προκύπτουν τα μοντέλα ARIMA( $p,d,q$ ) με  $p,d,q$ , όπως αναφέρθηκε παραπάνω, την τάξη του αντίστοιχου μοντέλου. Με τη χρήση του τελεστή ολίσθησης  $B$  το μοντέλο παριστάνεται ως:

$$(1 - \phi_1 B - \phi_2 B^2 - \dots - \phi_p B^p)(1 - B)^d y_t = c + (1 + \theta_1 B + \dots + \theta_q B^q) e_t .$$

Ο πρώτος όρος του πρώτου μέλους της εξίσωσης αναπαριστά το μοντέλο AR( $p$ ), ο δεύτερος την διαφόριση  $I(d)$  και το δεύτερο μέλος της εξίσωσης αντιπροσωπεύει το μοντέλο MA( $q$ ).

#### **2.4.4 Πολυεπίπεδο Αντίληπτρο(Multi-Layer Perceptrons)**

Τα μοντέλα αυτά ανήκουν στις μεθόδους υπολογιστικής ευφυΐας και είναι μια ειδική κατηγορία των τεχνητών νευρωνικών δικτύων, τα οποία χρησιμοποιούνται όλο και περισσότερο στις μέρες μας σε διάφορες εφαρμογές μία από τις οποίες είναι η πρόβλεψη χρονοσειρών. Υπάρχουν πολλές κατηγορίες νευρωνικών δικτύων όμως στην παρούσα διπλωματική χρησιμοποιήθηκε μόνο το πολυεπίπεδο αντίληπτρο, το οποίο είναι και αυτό που θα αναλύσουμε παρακάτω. Θα αναφερόμαστε από εδώ και πέρα σε αυτό ως MLP.

Σε αντίθεση με τις στατιστικές μεθόδους πρόβλεψης, βασικό πλεονέκτημα των νευρωνικών δικτύων είναι ότι μπορούν να δώσουν ικανοποιητικά αποτελέσματα και χωρίς να έχουν πλήρη δεδομένα. Ακόμα τα νευρωνικά δίκτυα δεν προγραμματίζονται αλλά μαθαίνουν. Είναι φανερό επομένως, ότι δεν χρειάζεται να κατανοήσουμε πλήρως το πρόβλημα για να το επιλύσουμε με νευρωνικά δίκτυα. Αξίζει να σημειωθεί πως τα νευρωνικά δίκτυα δεν δίνουν στο πως κατέληξαν σε ένα αποτέλεσμα, άρα δεν είναι κατάλληλα για εξαγωγή συμπερασμάτων (extrapolation), αλλά μόνο για παρεμβολή(interpolation).

Το πολυεπίπεδο αντίληπτρο(MLP) είναι μια κατηγορία τεχνητού νευρωνικού δικτύου εμπρόσθιας τροφοδότησης, επιβλεπόμενης μάθησης(supervised learning). Ένα νευρωνικό επιβλεπόμενης μάθησης τροφοδοτείται με παραδείγματα εισόδων και τις αντίστοιχες επιθυμητές εξόδους. Στη διαδικασία της μάθησης τα βάρη που συνδέουν τους νευρώνες προσαρμόζονται ώστε να ελαχιστοποιείται η

διαφορά ανάμεσα στην επιθυμητή έξοδο και στην έξοδο του δικτύου. Ένα δίκτυο εμπρόσθιας τροφοδότησης είναι υποκατηγορία ενός δικτύου επιβλεπόμενης μάθησης στην οποία τα δεδομένα εισόδου τροφοδοτούνται στο δίκτυο από τους νευρώνες εισόδου και μεταφέρονται προς τα επόμενα στρώματα του δικτύου. Προκύπτει μια απεικόνιση των εισόδων στις εξόδους, η οποία γίνεται μέσω κάποιων παραμέτρων. Οι παράμετροι αυτοί είναι τα βάρη που συνδέουν τους νευρώνες.

Τα πολυεπίπεδα αντίληπτρα είναι πολύ δημοφιλή νευρωνικά δίκτυα. Ουσιαστικά ο όρος "πολυεπίπεδο αντίληπτρο" δεν αναφέρεται σε ένα αντίληπτρο με πολλά επίπεδα, αλλά περιέχει πολλά αντίληπτρα οργανωμένα σε επίπεδα. Ένα αντίληπτρο παίρνει τον γραμμικό συνδυασμό των εισόδων του και μέσω μια συνάρτησης ενεργοποίησης(activation function) παράγει την έξοδο. Η διαδικασία αυτή αναπαριστάται μαθηματικά ως εξής :

$$o = f(\sum_{i=1}^n w_i x_i - b)$$

όπου:

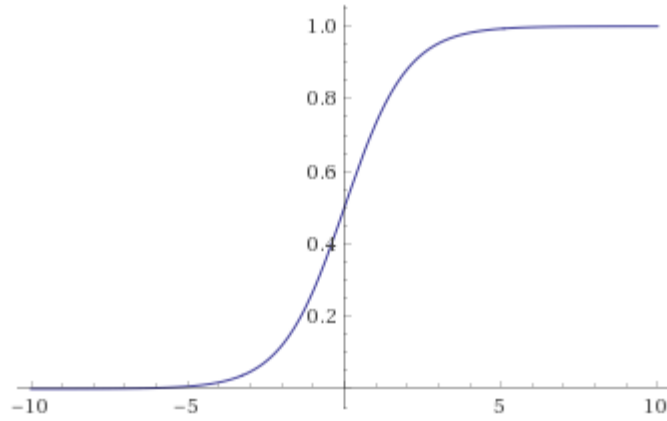
1.  $o$  η έξοδος
2.  $x_i$  οι είσοδοι
3.  $w_i$  τα βάρη
4.  $b$  ο συντελεστής πόλωσης
5.  $f$  η συνάρτηση ενεργοποίησης

Η συνάρτηση ενεργοποίησης είναι πολύ σημαντική και η επιλογή της αλλάζει ανάλογα το πρόβλημα. Η μορφή της συνήθως είναι βηματική, γραμμική, λογιστική σιγμοειδής ή υπερβολική εφαπτομένη.

Οι πιο συνηθισμένες συναρτήσεις ενεργοποίησης είναι η λογιστική σιγμοειδής και η υπερβολική εφαπτομένη η οποία ουσιαστικά είναι μια μετατοπισμένη και κλιμακωτή εκδοχή της σιγμοειδούς.

Λογιστική σιγμοειδής:

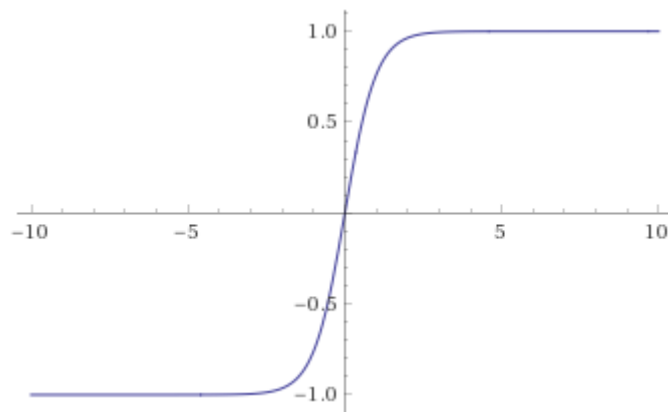
$$f(x) = \frac{1}{1+e^{-x}}$$



Εικόνα 1 : Συνάρτηση ενεργοποίησης λογιστική σιγμοειδής

Υπερβολική εφαπτομένη:

$$f(x) = \tanh(x) = \frac{e^x - e^{-x}}{e^x + e^{-x}}$$



Εικόνα 2 : Συνάρτηση ενεργοποίησης υπερβολική εφαπτομένη

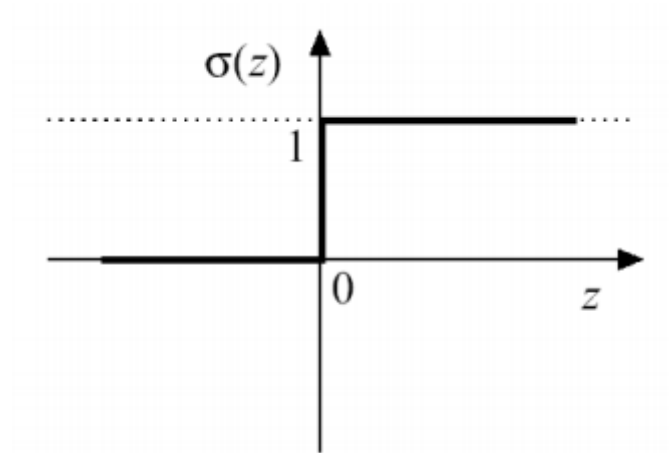


Σε περιπτώσεις που θέλουμε να κάνουμε διαχωρισμό δεδομένων εισόδου συνήθως χρησιμοποιείται η βηματική συνάρτηση με τιμές 0 ή 1 στη μία μορφή της ενώ στην άλλη τιμές -1 ή 1.

Βηματική:

$$f(x) = 0, x < 0$$

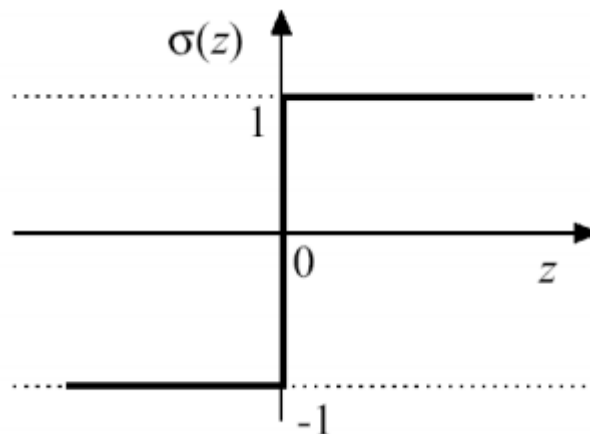
$$1, x \geq 0$$



ή

$$f(x) = -1, x < 0$$

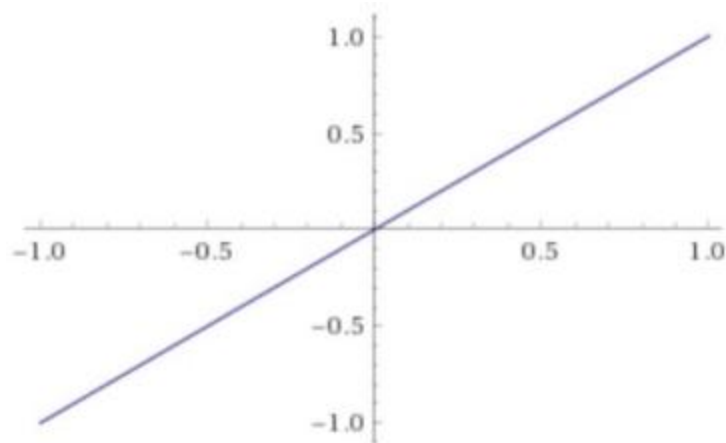
$$1, x \geq 0$$



Εικόνα 3 : Συνάρτηση ενεργοποίησης βηματική

Για χρονοσειρές με τάση αποδίδει αρκετά καλά η γραμμική συνάρτηση ενεργοποίησης:

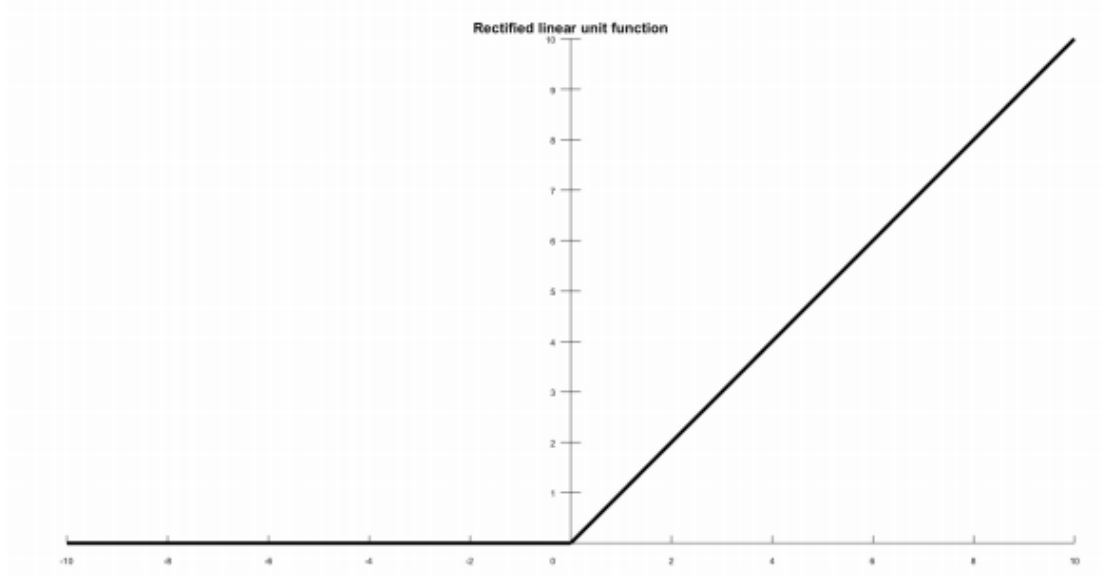
Γραμμική :  $f(x) = x$



Εικόνα 4 : Συνάρτηση ενεργοποίησης γραμμική

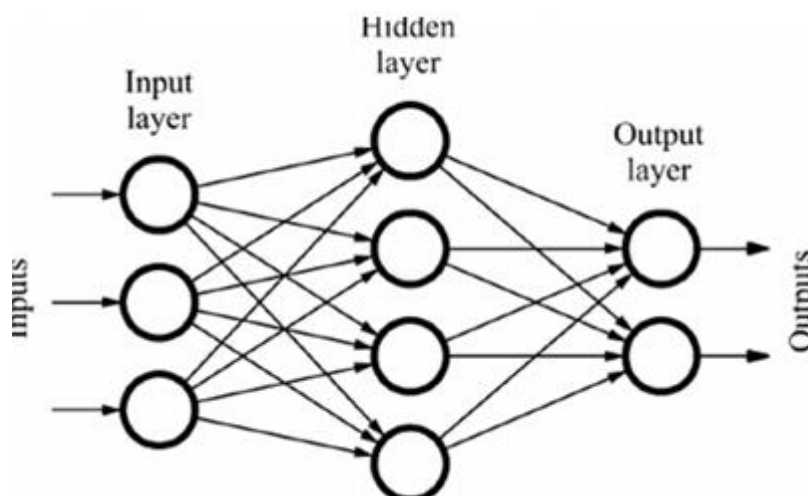
Μια παραλλαγή της γραμμικής, η οποία χρησιμοποιείται πάρα πολύ τα τελευταία χρόνια είναι η αποκομμένη γραμμική ή ReLu(Rectified Linear Unit).

ReLU :  $f(x) = 0, x < 0$   
 $x, x \geq 0$



Εικόνα 5 : Συνάρτηση ενεργοποίησης ReLu

Ένα νευρωνικό δίκτυο εμπρόσθιας τροφοδότησης έχει την εξής μορφή :



Εικόνα 6 : Γραφική αναπαράσταση νευρωνικού δικτύου εμπρόσθιας τροφοδότησης

Το νευρωνικό δίκτυο αποτελείται από στρώματα νευρώνων, με τον κάθε νευρώνα να ενώνεται με όλους τους νευρώνες του επόμενου στρώματος μέσω συνάψεων. Στα νευρωνικά εμπρόσθιας τροφοδότησης δεν υπάρχουν αναδράσεις και επομένως ούτε στο πολυεπίπεδο αντίληπτρο. Το κλασικό πολυεπίπεδο αντίληπτρο αποτελείται από τρία στρώματα, το στρώμα εισόδου το κρυφό στρώμα και το στρώμα εξόδου. Η είσοδος είναι ένα διάνυσμα η διαστάσεων και η έξοδος ένα

διάνυσμα  $m$  διαστάσεων. Άρα μπορεί το δίκτυο να θεωρηθεί σαν μια συνάρτηση από το χώρο  $n$ -διαστάσεων στο χώρο  $m$ -διαστάσεων.

Τα παραπάνω περιγράφουν τη βασική δομή ενός νευρωνικού δικτύου εμπρόσθιας τροφοδότησης. Μένει τώρα να δούμε πως γίνεται η διαδικασία μάθησης του δικτύου, ώστε να βρεθεί ο τελικός πίνακας με τα βάρη των συνάψεων που ενώνουν τους νευρώνες. Στη διαδικασία της μάθησης το δίκτυο τροφοδοτείται συνέχεια ζευγάρια εισόδων-εξόδων. Κάθε πλήρης προβολή όλων των εισόδων-εξόδων εκπαίδευσης ονομάζεται εποχή(epoch). Η εκπαίδευση του δικτύου γίνεται για αρκετές εποχές έως ότου σταθεροποιηθούν τα βάρη και αρχίζει να συγκλίνει το άθροισμα των τετραγωνικών σφαλμάτων. Η εκπαίδευση ενός νευρωνικού μπορεί να γίνει με διάφορους αλγορίθμους, ο πιο δημοφιλής όμως είναι ο αλγόριθμος της προς τα πίσω διάδοσης σφάλματος(backpropagation). Ο τυπικός αλγόριθμος backpropagation είναι μια μέθοδος πτώσης κλίσης(gradient descent). Για την έναρξη του αλγορίθμου τα συναπτικά βάρη αρχικοποιούνται σε τυχαίες τιμές. Μετά εισάγονται οι είσοδοι των δεδομένων, οι οποίες με τα τυχαία βάρη παράγουν κάποια έξοδο. Έπειτα έχοντας ως στόχο την ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος ανάμεσα στις εξόδους του δικτύου και στις πραγματικές εξόδους των δεδομένων προσαρμόζονται τα νέα βάρη των συνάψεων. Η διαδικασία αυτή επαναλαμβάνεται για μεγάλο αριθμό εποχών έως ότου το σφάλμα γίνει πολύ μικρό ή όταν ολοκληρωθεί ο αριθμός των εποχών που έχουμε ορίσει.

Ο ρυθμός μάθησης ενός δικτύου είναι μια παράμετρος που χρησιμοποιείται για να ελέγχει τον ρυθμό με τον οποίο αλλάζουν τα βάρη στο τέλος κάθε εποχής. Πολύ μεγάλος ρυθμός μάθησης μπορεί να οδηγήσει σε ταλαντώσεις και πιο αργή σύγκλιση από μια απότομη κάθοδο(direct descent). Αντίθετα μικρός ρυθμός μάθησης συνεπάγεται μικρά βήματα καθόδου, άρα και αργή σύγκλιση. Επομένως, προκύπτει κάθε φορά διαφορετικός βέλτιστος ρυθμός μάθησης ανάλογα τα δεδομένα.

Ο αλγόριθμος backpropagation παρουσιάζει και ορισμένα μειονεκτήματα. Το κυριότερο είναι η μεγάλη και χρονοβόρα διαδικασία εκπαίδευσης, ειδικά όταν έχει γίνει λανθασμένη επιλογή ρυθμού αύξησης. Ακόμα είναι πιθανή η παράλυση του δικτύου η οποία μπορεί να συμβεί κατά την εκπαίδευση αν τα βάρη πάρουν πολύ μεγάλες τιμές. Η είσοδος ενός νευρώνα επομένως θα είναι μεγάλη και η έξοδος θα είναι συνήθως ή 0 ή 1 λόγω της σιγμοειδούς συνάρτησης που χρησιμοποιείται κατά βάση. Αυτό θα έχει ως αποτέλεσμα οι προσαρμογές που γίνονται στα βάρη να μην προσφέρουν ιδιαίτερη βελτίωση και η εκπαίδευση να μένει στάσιμη. Τέλος, δεν είναι σίγουρο ότι ο αλγόριθμος θα συγκλίνει στο ελάχιστο της αντικειμενικής συνάρτησης αλλά σε κάποιο τοπικό ελάχιστο.

Στην αντιμετώπιση αυτών των μειονεκτημάτων παίζει πολύ σημαντικό ρόλο η προεπεξεργασία των δεδομένων. Υπάρχουν δύο κύριες μέθοδοι προεπεξεργασίας που χρησιμοποιούνται. Η πρώτη είναι ο μετασχηματισμός(transformation) και πρόκειται για επεξεργασία των δεδομένων ώστε να προκύψει άλλος τύπος εισόδου στο δίκτυο. Η δεύτερη μέθοδος είναι η κανονικοποίηση(normalization) και αποσκοπεί στην προσαρμογή των δεδομένων σε μια κλίμακα. Η κανονικοποίηση παίζει πολύ σημαντικό ρόλο στην διαδικασία της μάθησης του δικτύου και βοηθάει σημαντικά στην γρηγορότερη σύγκλιση, όμως ο έλεγχος της απόδοσης του θα πρέπει να γίνεται με δεδομένα της αρχικής μορφής και όχι κανονικοποιημένα προκειμένου να εξάγουμε χρήσιμα αποτελέσματα.

Για να θεωρηθεί ένα νευρωνικό δίκτυο επιτυχημένο, εκτός από την ελαχιστοποίηση του σφάλματος ανάμεσα στις πραγματικές και τις επιθυμητές εξόδους θα πρέπει να ανταποκρίνεται και σε δεδομένα εισόδου εκτός του συνόλου μάθησης. Ένα πρόβλημα που προκύπτει συνήθως κατά την εκπαίδευση ενός νευρωνικού δικτύου είναι η υπερπροσαρμογή(overfitting). Με το όρο αυτό αναφερόμαστε στην απομνημόνευση του συνόλου μάθησης κάτι που οδηγεί σε αδυναμία του δικτύου να ανταποκριθεί σε νέα δεδομένα σωστά. Σκοπός του δικτύου δεν πρέπει να είναι μια απλή αντιστοίχιση των δεδομένων εισόδου στα δεδομένα εξόδου μέσω μια απομνημόνευσης, αλλά η αναγνώριση μοτίβων και συσχετίσεων ανάμεσα στις εισόδους και τις εξόδους.

Παραπάνω περιγράφηκαν τα βασικά χαρακτηριστικά ενός νευρωνικού δικτύου εμπρόσθιας τροφοδότησης και ορισμένα σημεία που θέλουν ιδιαίτερη προσοχή κατά τη μοντελοποίηση αυτού. Με την προϋπόθεση ότι ακολουθούμε ορισμένους βασικούς κανόνες κατά την μοντελοποίηση ενός MLP και μερικές δοκιμές για να βρούμε τις κατάλληλες παραμέτρους, τότε το μοντέλο αυτό μπορεί να αποδώσει πολύ καλά στην πρόβλεψη χρονοσειρών.

## 2.5 Δείκτες Σφάλματος

Προκειμένου να επιλεγεί η καλύτερη μέθοδος πρόβλεψης χρειαζόμαστε κάποιον τρόπο να μετρήσουμε την ακρίβεια των μεθόδων που εξετάζονται. Το ρόλο αυτό αναλαμβάνουν διάφοροι στατιστικοί δείκτες σφαλμάτων. Ως σφάλμα ορίζεται η διαφορά της πραγματικής τιμής μιας παρατήρησης από την αντίστοιχη πρόβλεψη που έχει παραχθεί. Η αξιολόγηση των σφαλμάτων μπορεί να γίνει είτε σε ένα in-sample δείγμα, δηλαδή σε παρατηρήσεις που ήδη έχουμε, είτε σε ένα out-of sample δείγμα, δηλαδή σε παρατηρήσεις που θα μάθουμε αργότερα αφού κάνουμε τις αντίστοιχες προβλέψεις. Περισσότερο χρησιμοποιείται η δεύτερη προσέγγιση διότι σημασία έχει η απόδοση του μοντέλου μας σε νέα δεδομένα. Παρ' όλα αυτά αν οι in-sample δείκτες είναι καλοί συνήθως θα είναι και out-of sample. Οι πιο βασικοί δείκτες σφάλματος είναι οι εξής:

### 1. Μέσο Σφάλμα (Mean Error-ME)

$$ME = \frac{\sum_{i=1}^n (Y_i - F_i)}{n}$$

Όταν ο δείκτης αυτός παίρνει τιμές κοντά στο μηδέν τα σφάλματα θεωρούνται τυχαία και όχι συστηματικά. Θετικές τιμές του δείκτη υποδηλώνουν απαισιοδοξία στις προβλέψεις ενώ αντίθετα αρνητικές τιμές υποδηλώνουν αισιοδοξία.

### 2. Μέσο Απόλυτο Σφάλμα (Mean Absolute Error-MAE)

$$MAE = \frac{\sum_{i=1}^n |Y_i - F_i|}{n}$$

Ο δείκτης αυτός δεν δείχνει την προκατάληψη της μεθόδου, αλλά μόνο της ακρίβειας της. Όσο μεγαλύτερος είναι ο δείκτης τόσο χειρότερη είναι η ακρίβεια του μοντέλου.

### 3. Μέσο Τετραγωνικό Σφάλμα (Mean Squared Error-MSE)

$$MSE = \frac{\sum_{i=1}^n (Y_i - F_i)^2}{n}$$

Παρόμοια με τον προηγούμενο δείκτη δείχνει την ακρίβεια των προβλέψεων, με τη διαφορά ότι δίνει περισσότερη βάση στα μεγάλα σφάλματα και λιγότερη στα μικρά λόγω του τετραγωνισμού τους. Χρησιμεύει κυρίως στον υπολογισμό των καλύτερων συντελεστών εξομάλυνσης.

### 4. Ρίζα Μέσου Τετραγωνικού Σφάλματος (Root Mean Squared Error-RMSE)

$$RMSE = \frac{\sum_{i=1}^n \sqrt{(Y_i - F_i)^2}}{n}$$

Έχει ίδιες ιδιότητες με τον MSE όμως εκφράζεται στις ίδιες μονάδες με την χρονοσειρά.

### 5. Μέσο Απόλυτο Ποσοστιαίο Σφάλμα (Mean Absolute Percentage Error-MAPE)

$$MAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - F_i|}{Y_i}$$

Παίρνει τιμές μόνο μεγαλύτερες ή ίσες του μηδενός και προφανώς όσο πιο κοντά στο μηδέν τόσο καλύτερη ακρίβεια έχει η μέθοδος πρόβλεψης. Δεν συνίσταται η χρήση του σε χρονοσειρές διακοπτόμενης ζήτησης γιατί προκαλεί απροσδιοριστία. Είναι πολύ χρήσιμος όταν οι πραγματικές τιμές είναι πολύ υψηλές και όταν θέλουμε να συγκρίνουμε τα σφάλματα μιας μεθόδου πρόβλεψης που έχει εφαρμοστεί σε χρονοσειρές με διαφορετικό επίπεδο μέσης τιμής.

6.Συμμετρικό Μέσο Απόλυτο Ποσοστιαίο Σφάλμα(Symmetric Mean Absolute Percentage Error-sMAPE)

$$sMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - F_i|}{(Y_i + F_i)/2}$$

Πρόκειται για παρόμοιο δείκτη με τον MAPE, με τη διαφορά ότι ο δείκτης αυτός έχει ανώτατο όριο και παίρνει τιμές από 0% έως 200%.

7.Μέσο Απόλυτο Κανονικοποιημένο Σφάλμα(Mean Absolute Scaled Error-MAsE)

$$MASE = \frac{\frac{1}{n} \sum_{i=1}^n |Y_i - F_i|}{\frac{1}{n-1} \sum_{i=2}^n |Y_i - Y_{i-1}|}$$

Πρόκειται για ένα δείκτης που χρησιμοποιεί το σφάλμα της Naive ως σημείο αναφοράς προκειμένου να συγκριθεί η αποδοτικότητα διαφόρων μεθόδων πρόβλεψης.

Στην παρούσα διπλωματική χρησιμοποιείται κυρίως το συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα(sMAPE), καθώς εφαρμόζονται διάφορες μέθοδοι προβλέψεων σε πολλές χρονοσειρές και δεν μας ενδιαφέρει τόσο η προκατάληψη των μεθόδων όσο η ακρίβεια τους. Επίσης η έκφραση του δείκτη σε ποσοστά καθιστά εύκολη τη σύγκριση μεταξύ των διάφορων μοντέλων που αναπτύχθηκαν.



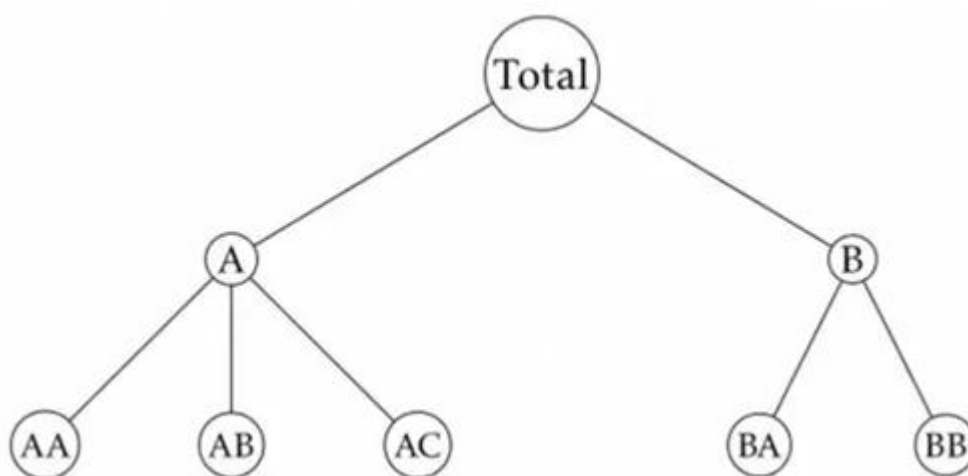
## Κεφάλαιο 3 : Ιεραρχική Συνάθροιση και Ιεραρχικές Προβλέψεις

### 3.1 Ιεραρχική Συνάθροιση

Οι χρονοσειρές μπορούν πολλές φορές να αποσυντεθούν σε επιμέρους χρονοσειρές με βάση κάποιο χαρακτηριστικό τους. Για παράδειγμα ο συνολικός αριθμός πωλήσεων μιας επιχείρησης μπορεί να χωριστεί σε πωλήσεις από το κάθε προϊόν της, ή ο αριθμός των τουριστών σε μια χώρα να χωριστεί με γεωγραφικά κριτήρια σε επιμέρους περιοχές. Κάθε μία από αυτές τις νέες χρονοσειρές μπορεί να χωριστεί σε ακόμα περισσότερες χρονοσειρές είτε βάση το ίδιο χαρακτηριστικό με πριν είτε με κάποιο άλλο χαρακτηριστικό. Έτσι δημιουργείται μια ιεραρχική δομή από χρονοσειρές στις οποίες αναφερόμαστε ως "hierarchical time series". Με την τεχνική της ιεραρχικής συνάθροισης μπορούμε να παράγουμε προβλέψεις για κάθε χρονοσειρά ξεχωριστά και στην συνέχεια να τις συνδυάσουμε για την παραγωγή μιας τελικής πρόβλεψης. Οι δύο κυριότερες διαδικασίες ιεραρχικής πρόβλεψης είναι η top-down και η bottom-up. Κατά καιρούς έχουν γίνει αρκετές μελέτες προκειμένου να βρεθεί ποια μέθοδος είναι καλύτερη, όμως τα αποτελέσματα και οι γνώμες δίστανται. Πολλές έρευνες έχουν δείξει ότι η top-down δίνει καλύτερα αποτελέσματα, αλλά πολλοί είναι και εκείνοι που έχουν επιχειρηματολογήσει υπέρ της bottom-up. Υπάρχουν βέβαια και οι έρευνες που έχουν δείξει ότι το πόσο καλά θα ανταπεξέλθει μια από τις δύο μεθόδους εξαρτάται από τα δεδομένα που χρησιμοποιούμε και τα αποτελέσματα που επιθυμούμε να πετύχουμε. Πλεονέκτημα της bottom-up είναι ότι παράγει καλύτερες προβλέψεις σε δεδομένα που ποικίλουν και έχουν μεγάλες διαφορές μεταξύ τους, ενώ πλεονέκτημα της top-down είναι ότι δίνει μεγαλύτερη βάση στην τάση που χαρακτηρίζει τα δεδομένα. Μία ακόμα μέθοδος την οποία πρότειναν το 2011 οι R.J.Hyndman και R.A.Ahmed, η οποία φαίνεται να είναι πιο αποτελεσματική από τις δύο προηγούμενες είναι η optimal. Στην παρούσα διπλωματική χρησιμοποιήθηκαν και οι τρεις μέθοδοι, τις οποίες και θα αναλύσουμε παρακάτω.

## 3.2 Ιεραρχικές Χρονοσειρές

Έστω ότι έχουμε την ιεραρχική δομή του σχήματος 3.1 , $K=2$  επιπέδων. Στην κορυφή της ιεραρχίας βρίσκεται το "Total" , το οποίο ονομάζεται και επίπεδο μηδέν(level 0) και πρόκειται για το συνολικό άθροισμα όλων των χρονοσειρών των δεδομένων μας. Η παρατήρηση  $t$  της Total χρονοσειράς συμβολίζεται ως  $y_t$  για  $t=1,2,3,\dots,T$ . Η Total χωρίζεται σε δύο χρονοσειρές στο επίπεδο 1, τις A και B, οι οποίες με τη σειρά τους χωρίζονται σε 3 και 2 αντίστοιχα, οι οποίες είναι οι AA, ABC και BA, BB στο επίπεδο 2, το οποίο ονομάζεται bottom-level. Για την  $t$  παρατήρηση στα υπόλοιπα επίπεδα χρησιμοποιούμε τον συμβολισμό  $y_{j,t}$  με το  $j$  να αναφέρεται στον αντίστοιχο κόμβο. Για παράδειγμα η  $y_{B,t}$  συμβολίζει την παρατήρηση  $t$  της χρονοσειράς B, ενώ η  $y_{AB,t}$  την παρατήρηση  $t$  της χρονοσειράς AB.



Εικόνα 7: Διάγραμμα ιεραρχίας δύο επιπέδων

Στο συγκεκριμένο παράδειγμα ο συνολικός αριθμός των χρονοσειρών είναι  $n=8$ , ενώ ο αριθμός των χρονοσειρών του κατώτερου επιπέδου είναι  $m=5$ . Αναγκαστικά σε όλες τις ιεραρχίες θα ισχύει  $n>m$ .

Για κάθε στιγμή  $t$  οι παρατηρήσεις ενός επιπέδου θα αθροίζουν στην παρατήρηση  $t$  του από πάνω επιπέδου. Για παράδειγμα:

$$y_t = y_{A,t} + y_{B,t} = (y_{AA,t} + y_{AB,t} + y_{AC,t}) + (y_{BA,t} + y_{BB,t})$$

Οι εξισώσεις αυτές μπορούν να περιγραφούν καλύτερα χρησιμοποιώντας πίνακες. Συγκεκριμένα χρησιμοποιούμε έναν πίνακα  $S$   $n \times m$  διαστάσεων τον οποίο ονομάζουμε πίνακα αθροίσματος (Summing matrix) ο οποίος καθορίζει τον τρόπο που αθροίζονται οι bottom-level χρονοσειρές.

Στην προκειμένη περίπτωση έχουμε :

$$\begin{bmatrix} y_t \\ y_{A,t} \\ y_{B,t} \\ y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} y_{AA,t} \\ y_{AB,t} \\ y_{AC,t} \\ y_{BA,t} \\ y_{BB,t} \end{bmatrix}$$

ή σε άλλη μορφή :

$$y_t = S b_t$$

όπου  $y_t$  , ένα διάνυσμα  $n$ -διάστασης με όλες τις παρατηρήσεις της ιεραρχίας την χρονική στιγμή  $t$  ,  $S$  ο πίνακας αθροίσματος που αναφέρθηκε παραπάνω και  $b_t$  ένα διάνυσμα  $m$  διάστασης όλων των παρατηρήσεων των bottom-level χρονοσειρών την χρονική στιγμή  $t$ . Η πρώτη γραμμή του πίνακα αθροίσματος ουσιαστικά αθροίζει όλες τις χρονοσειρές, οι δύο επόμενες γραμμές αθροίζουν τις bottom-level χρονοσειρές που ανήκουν στους κόμβους του πρώτου επιπέδου όπως δείχνει το σχήμα 3.1 και οι υπόλοιπες γραμμές είναι ένας μοναδιαίος πίνακας που δίνει την κάθε bottom-level χρονοσειρά μόνη της.

Γενικά στις ιεραρχικές μεθόδους μας ενδιαφέρει περισσότερο να δουλεύουμε με τις προβλέψεις παρά με τις κανονικές παρατηρήσεις των χρονοσειρών. Έστω λοιπόν ότι έχουμε παράγει τις προβλέψεις για ορίζοντα  $h$  κάθε μίας χρονοσειράς ξεχωριστά, σε όλα τα επίπεδα. Τότε αν έχουμε δεδομένα  $n$ -παρατηρήσεων θα έχουμε προβλέψεις μέχρι και τη στιγμή  $n+h$ . Όλες οι βασικές προβλέψεις για ορίζοντα  $h$  του επιπέδου της ιεραρχίας  $i$  θα συμβολίζονται ως  $\hat{Y}_{i,n}(h)$  , ενώ οι βασικές προβλέψεις όλης της ιεραρχίας θα συμβολίζονται ως  $\hat{Y}_n(h)$  , το οποίο θα είναι ένα διάνυσμα με στοιχεία στην ίδια σειρά με το  $y_t$ . Χρησιμοποιώντας τους συμβολισμούς αυτού όλες οι υπάρχουσες ιεραρχικές μέθοδοι μπορούν να περιγραφούν με την εξής σχέση :

$$\tilde{Y}_n(h) = SP \hat{Y}_n(h) \quad (3.1)$$

Αν τώρα συμβολίσουμε με  $m_i$  τον αριθμό των χρονοσειρών του επιπέδου  $i$ , ο συνολικός αριθμός όλων των χρονοσειρών της ιεραρχίας επιπέδων  $K$  θα είναι  $m = m_0 + m_1 + \dots + m_K$ . Ο πίνακας  $S$  θα είναι ο πίνακας αθροίσματος που αναφέρθηκε πιο πάνω διαστάσεων  $m \times m_K$ , ενώ ο πίνακας  $P$  θα έχει διαστάσεις  $m_K \times m$  και ο ρόλος του θα διαφέρει ανάλογα με τη μέθοδο πρόβλεψης που εφαρμόζουμε κάθε φορά. Προκύπτει από τη σχέση (3.1) ότι οι τελικές αναθεωρημένες προβλέψεις  $\tilde{Y}_n(h)$  είναι το αποτέλεσμα του γραμμικού συνδυασμού των ανεξάρτητων βασικών προβλέψεων  $\hat{Y}_n(h)$ .

### 3.3 Bottom-up Μέθοδος.

Με τη μέθοδο αυτή παράγουμε προβλέψεις αρχικά για όλες τις bottom-level χρονοσειρές και στην συνέχεια τις προσθέτουμε για να δημιουργήσουμε προβλέψεις για τις υπόλοιπες χρονοσειρές της ιεραρχίας. Για παράδειγμα στην προηγούμενη ιεραρχία έστω ότι παράγουμε προβλέψεις για οριζόντια πρόβλεψη  $h$  για κάθε bottom-level χρονοσειρά. Αυτές θα συμβολίζονται ως εξής :

$$\hat{Y}_{AA,h}, \hat{Y}_{AB,h}, \hat{Y}_{AC,h}, \hat{Y}_{BA,h}, \hat{Y}_{BB,h}$$

Προσθέτοντας τες παίρνουμε τις προβλέψεις για τις χρονοσειρές των υπόλοιπων επιπέδων:

$$\tilde{Y}_h = \hat{Y}_{AA,h} + \hat{Y}_{AB,h} + \hat{Y}_{AC,h} + \hat{Y}_{BA,h} + \hat{Y}_{BB,h}$$

$$\tilde{Y}_{A,h} = \hat{Y}_{AA,h} + \hat{Y}_{AB,h} + \hat{Y}_{AC,h}$$

$$\tilde{Y}_{B,h} = \hat{Y}_{BA,h} + \hat{Y}_{BB,h}$$

Χρησιμοποιώντας εδώ τον πίνακα αθροίσματος  $S$  προκύπτει η εξίσωση

$$\begin{bmatrix} \tilde{Y}_h \\ \tilde{Y}_{A,h} \\ \tilde{Y}_{B,h} \\ \tilde{Y}_{AA,h} \\ \tilde{Y}_{AB,h} \\ \tilde{Y}_{AC,h} \\ \tilde{Y}_{BA,h} \\ \tilde{Y}_{BB,h} \end{bmatrix} = \begin{bmatrix} 1 & 1 & 1 & 1 & 1 \\ 1 & 1 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 1 \\ 1 & 0 & 0 & 0 & 0 \\ 0 & 1 & 0 & 0 & 0 \\ 0 & 0 & 1 & 0 & 0 \\ 0 & 0 & 0 & 1 & 0 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix} \begin{bmatrix} \hat{Y}_{AA,h} \\ \hat{Y}_{AB,h} \\ \hat{Y}_{AC,h} \\ \hat{Y}_{BA,h} \\ \hat{Y}_{BB,h} \end{bmatrix}$$

ή σε πιο συμπαγή μορφή :  $\tilde{Y}_h = S \hat{b}_h$

όπου  $\tilde{Y}_h$  είναι ένα διάνυσμα όλων των προβλέψεων ώστε οι προβλέψεις των ψηλότερων επιπέδων να προκύπτουν από το άθροισμα των αντίστοιχων bottom-level προβλέψεων με τον τρόπο που δείχνει η ιεραρχία και  $\hat{b}_h$  ένα διάνυσμα με τις βασικές προβλέψεις για τις χρονοσειρές του κατώτατου επιπέδου, με ορίζοντα πρόβλεψης  $h$ .

Αν χρησιμοποιήσουμε την εξίσωση (3.1) τότε ο  $P$  θα έχει τη μορφή :

$$P = [O_{(m_K) \times (m-m_K)} | I_{m_K}]$$

όπου  $O_{ij}$  είναι ο μηδενικός πίνακας διαστάσεων  $i \times j$ . Ο ρόλος του  $P$  εδώ είναι να εξάγει τις bottom-level προβλέψεις οι οποίες μετά αθροίζονται από τον  $S$  και προκύπτουν οι αναθεωρημένες προβλέψεις για όλη την ιεραρχία.

Βασικό πλεονέκτημα της μεθόδου αυτής αποτελεί ότι παράγουμε προβλέψεις μόνο στο κατώτατο επίπεδο της ιεραρχίας με αποτέλεσμα να μην χάνεται πληροφορία λόγω του αθροίσματος πολλών χρονοσειρών. Αρνητικό δε, είναι το γεγονός ότι αυτές οι χρονοσειρές μπορεί να έχουν αρκετό "θόρυβο"(για παράδειγμα special events) ο οποίος κάνει πιο προκλητική την μοντελοποίηση τους και την παραγωγή πρόβλεψης.

### 3.4 Top-down Μέθοδος

Αντίθετα με την bottom-up μέθοδο, στην top-down παράγουμε προβλέψεις μόνο για την Total χρονοσειρά, δηλαδή αυτή που βρίσκεται στο ψηλότερο επίπεδο της ιεραρχίας και μετά τις αποσυνθέτουμε στα χαμηλότερα επίπεδα. Το μοίρασμα των προβλέψεων της Total χρονοσειράς στα χαμηλότερα επίπεδα μπορεί να γίνει με διάφορους τρόπους, όμως ο πιο δημοφιλής, τον οποίο χρησιμοποιήσαμε και στην παρούσα διπλωματική, είναι με τις ιστορικές αναλογίες των δεδομένων. Ο πίνακας P τώρα θα πάρει τη μορφή :

$$P = [p | O_{(m_K) \times (m-1)}]$$

όπου  $p = [p_1, p_2, \dots, p_{m_K}]'$  είναι οι αναλογίες των bottom-level χρονοσειρών. Ο ρόλος του P εδώ πέρα είναι να μοιράσει τις προβλέψεις του ανώτατου επιπέδου σε προβλέψεις των χρονοσειρών του κατώτερου επιπέδου. Όσον αφορά τις αναλογίες χρησιμοποιούμε τις παρακάτω, οι οποίες απέδωσαν καλά στην έρευνα των Gross και Sohl το 1990.

Η πρώτη προσέγγιση χρησιμοποιεί τις αναλογίες οι οποίες δίνονται από τον τύπο:

$$p_j = \sum_{t=1}^n \frac{Y_{j,t}}{Y_t} / n$$

για  $j = 1, 2, \dots, m_K$ . Ο κάθε συντελεστής εδώ συμβολίζει το μέσο όρο των ιστορικών αναλογιών των bottom-level χρονοσειρών  $\{Y_{j,t}\}$  για περίοδο  $t = 1, 2, \dots, n$  σε σχέση με το συνολικό άθροισμα  $\{Y_t\}$ .

Η δεύτερη προσέγγιση για τους συντελεστές  $p$  είναι :

$$p_j = \sum_{t=1}^n \frac{Y_{j,t}}{n} / \sum_{t=1}^n \frac{Y_t}{n}$$

για  $j = 1, 2, \dots, m_K$ . Κάθε αναλογία  $p_j$  εδώ συμβολίζει τη μέση ιστορική τιμή των bottom-level χρονοσειρών  $\{Y_{j,t}\}$  σε σχέση με τη μέση τιμή του συνολικού αθροίσματος  $\{Y_t\}$ .

Μεγάλο πλεονέκτημα της μεθόδου αυτής αποτελεί η απλότητα της εφαρμογής της αφού χρειάζεται να μοντελοποιήσουμε και να παράγουμε προβλέψεις μόνο για την ανώτατη χρονοσειρά "Total" της ιεραρχίας. Η προσέγγιση αυτή συνήθως δίνει

αρκετά καλά αποτελέσματα για τα πάνω επίπεδα της ιεραρχίας όμως χάνεται πληροφορία λόγω του αθροίσματος όλων των χρονοσειρών. Έτσι δεν είναι δυνατό να εκμεταλλευτούμε τα special events ή άλλα χαρακτηριστικά των bottom-level χρονοσειρών για τις προβλέψεις μας. Τέλος οι προβλέψεις που βασίζονται στις ιστορικές αναλογίες τείνουν να χάνουν την τάση που πιθανόν εμφανίζουν τα δεδομένα μας.

### 3.5 Optimal Reconciliation Μέθοδος

Η μέθοδος αυτή συνδυάζει με τον βέλτιστο τρόπο τις βασικές προβλέψεις προκειμένου να παράγουμε ένα σύνολο από αναθεωρημένες προβλέψεις, οι οποίες είναι όσο πιο κοντά γίνεται στις προβλέψεις που εξαρτώνται από μία μεταβλητή, αλλά και να πληρούν την προϋπόθεση οι προβλέψεις στα ψηλότερα επίπεδα της ιεραρχίας να είναι ίσες με το άθροισμα των αντίστοιχων προβλέψεων των χαμηλότερων επιπέδων. Σε αντίθεση με τις άλλες μεθόδους, η συγκεκριμένη χρησιμοποιεί όλη την πληροφορία μια ιεραρχίας και αν οι βασικές προβλέψεις είναι αμερόληπτες, δίνει αμερόληπτες προβλέψεις σε όλη την ιεραρχία.

Η βασική ιδέα προέρχεται από την αναπαράσταση των βασικών προβλέψεων μιας ιεραρχίας για ορίζοντα πρόβλεψης  $h$  με το μοντέλο της γραμμικής παλινδρόμησης :

$$\hat{Y}_n(h) = S\beta_h + \varepsilon_h$$

όπου  $\beta_h = E[\hat{Y}_{k,n}(h) | Y_1, \dots, Y_n]$  είναι η άγνωστη μέση τιμή των βασικών προβλέψεων των bottom-level χρονοσειρών και το  $\varepsilon_h$  έχει μηδενική μέση τιμή και πίνακα συνδιακύμανσης  $V[\varepsilon_h] = \Sigma_h$ . Ο όρος  $\varepsilon_h$  συμβολίζει το σφάλμα της παραπάνω παλινδρόμησης και δεν πρέπει να συγχέεται με το σφάλμα της πρόβλεψης. Γνωρίζοντας τώρα το  $\Sigma_h$  με μέθοδο ελαχίστων τετραγώνων μπορούμε να εκτιμήσουμε την ελάχιστη αμερόληπτη διακύμανση του  $\beta_h$ . Σύμφωνα με τους Hyndman et al.(2007), με την υπόθεση ότι  $\varepsilon_h \approx S\varepsilon_{k,h}$ , όπου  $\varepsilon_{k,h}$  τα σφάλματα

προβλέψεων στο κατώτατο επίπεδο, ο καλύτερος γραμμικός αμερόληπτος εκτιμητής για το  $\beta_h$  θα είναι  $\hat{\beta}_h = (S'S)^{-1}S' \hat{Y}_n(h)$ . Οι αναθεωρημένες προβλέψεις θα προκύπτουν από τον τύπο  $\check{Y}_n(h) = S \hat{\beta}_h$ , άρα ο πίνακας P της γενικής εξίσωσης (3.1) θα είναι :

$$P = (S'S)^{-1}S'$$



## Κεφάλαιο 4 : Μεθοδολογική Προσέγγιση

### 4.1 Μεθοδολογία

Στα προηγούμενα κεφάλαια αναλύσαμε τις βασικές μεθόδους πρόβλεψης καθώς και τον τρόπο πρόβλεψης με πολυεπίπεδα αντίληπτρα. Είδαμε επίσης τις βασικές μεθόδους ιεραρχικής συνάθροισης, η οποία συμβάλλει στη συμφιλίωση των προβλέψεων και αυξάνει την απόδοση τους. Υπάρχουν αρκετές μέθοδοι ιεραρχικής συνάθροισης, όμως εξετάσαμε τις τρεις πιο βασικές, την top-down ,την bottom-up και την optimal χρησιμοποιώντας κάθε φορά διαφορετική μέθοδο για την παραγωγή των βασικών προβλέψεων. Υλοποιήθηκαν αρχικά τα μοντέλα naive, ets και arima εφαρμόστηκαν και οι τρεις ιεραρχικές μέθοδοι για το καθένα. Αυτά τα αποτελέσματα χρησιμοποιήθηκαν ως σημείο αναφοράς(benchmark). Στη συνέχεια εξετάσαμε την εφαρμογή των ιεραρχικών συναθροίσεων σε προβλέψεις οι οποίες προέκυψαν με χρήση πολυεπίπεδων αντίληπτρων προκειμένου να εξετάσουμε την απόδοση τους σε σχέση με τις κλασικές μεθόδους. Για την αξιολόγηση των μεθόδων χρησιμοποιήσαμε τον δείκτη συμμετρικού μέσου απόλυτου ποσοστιαίου σφάλματος(sMAPE). Παρακάτω φαίνεται αναλυτικά η διαδικασία που ακολουθήσαμε για την ανάπτυξη και τον έλεγχο των μοντέλων.

### 4.2 Επεξεργασία Δεδομένων

Φαινομενικά , το πρώτο βήμα για την παραγωγή των προβλέψεων είναι η εισαγωγή των δεδομένων μας στα μοντέλα που έχουμε δημιουργήσει. Η αλήθεια είναι όμως ότι τα δεδομένα συνήθως απαιτούν κάποια επεξεργασία προκειμένου να τα φέρουμε στη μορφή που θέλουμε. Αν για παράδειγμα διαχειριζόμαστε ημερήσια δεδομένα πωλήσεων πολλών μαγαζιών, είναι πιθανό οι χρονοσειρές μας να μην έχουν το ίδιο μήκος λόγω απώλειας τιμών. Οι τιμές αυτές μπορεί να λείπουν είτε γιατί κάποια μαγαζιά ήταν κλειστά ορισμένες μέρες ή γιατί όλα τα μαγαζιά δεν δούλευαν μια μέρα, όπως για παράδειγμα στις αργίες. Επομένως, πρέπει αρχικά κάπως να "γεμίσουμε" αυτές τις κενές τιμές ώστε οι χρονοσειρές μας να έχουν το ίδιο μήκος και την ίδια συχνότητα. Στην παρούσα διπλωματική οι κενές τιμές συμπληρώθηκαν με το μέσο όρο των τιμών της προηγούμενης και της επόμενης

εβδομάδας της αντίστοιχης μέρας. Αυτό έγινε γιατί οι ημερήσιες πωλήσεις παρουσιάζουν εποχιακότητα, αφού εξαρτώνται από τη μέρα. Τα σαββατοκύριακα υπάρχει μεγαλύτερη κίνηση στα μαγαζιά άρα και συνήθως περισσότερες πωλήσεις. Επίσης χρησιμοποιώντας τις δύο πιο κοντινές τιμές λαμβάνουμε έμμεσα υπόψη και την εποχή ώστε να μην υπάρχει μεγάλη απόκλιση από την πραγματικότητα. Αν για παράδειγμα μας έλειπε η τιμή μιας καλοκαιρινής Τετάρτης σε πωλήσεις παγωτών και χρησιμοποιούσαμε δύο χειμωνιάτικες Τετάρτες η τιμή που θα προέκυπτε δεν θα ήταν αντιπροσωπευτική.

Εφόσον οι χρονοσειρές μας αποκτήσουν ίδιο μήκος και συχνότητα, τότε μπορούμε να τις διαχειριστούμε για το επόμενο βήμα, το οποίο είναι να τις εντάξουμε σε μια ιεραρχία. Η ιεραρχία σε αυτήν την διπλωματική χωρίστηκε σε δύο επίπεδα με γεωγραφικά κριτήρια, ώστε οι bottom-level χρονοσειρές να αθροίζουν στο επόμενο επίπεδο ανάλογα με την περιοχή που βρίσκονται τα αντίστοιχα μαγαζιά και έπειτα να αθροίζουν όλες οι περιοχές στο "Total". Δημιουργούνται έτσι κάποιες νέες χρονοσειρές οι οποίες είναι οι αθροισμένες χρονοσειρές των χαμηλότερων επιπέδων και χρησιμεύουν στις ιεραρχικές μεθόδους.

### 4.3 Υλοποίηση Μοντέλων Πρόβλεψης

Αφού δημιουργήσουμε τις ιεραρχικές χρονοσειρές σειρά έχει η υλοποίηση των μοντέλων πρόβλεψης και η παραγωγή προβλέψεων. Αρχικά φτιάχνουμε τους πίνακες  $S$  και  $P$ , που αναλύθηκαν στο κεφάλαιο της ιεραρχικής συνάθροισης, για κάθε μία από τις μεθόδους top-down, bottom-up και optimal. Στη συνέχεια χρησιμοποιούμε τους πίνακες ώστε να φτιάξουμε μια συνάρτηση για κάθε μέθοδο, ο ρόλος της οποίας θα είναι να κάνει την συμφιλίωση των βασικών προβλέψεων με βάση τη γνωστή θεωρία. Εφόσον τώρα έχουμε έτοιμες τις συναρτήσεις αυτές μένει να φτιάξουμε τον πίνακα των βασικών προβλέψεων που θα χρησιμοποιήσουν οι συναρτήσεις ως είσοδο για την παραγωγή των ιεραρχικών προβλέψεων. Υλοποιούμε κάθε μέθοδο βασικής πρόβλεψης ξεχωριστά και στις βασικές προβλέψεις που προκύπτουν εφαρμόζουμε τις προηγούμενες συναρτήσεις ώστε να παράγουμε τις τελικές συμφιλιωμένες προβλέψεις. Οι πρώτες μέθοδοι που υλοποιήθηκαν ήταν οι naive, ets και arima, ενώ στην συνέχεια εξετάστηκε η παραγωγή βασικών προβλέψεων με αρκετές διαφορετικές προσεγγίσεις χρησιμοποιώντας τα πολυεπίπεδα αντίληπτα.

## 4.4 Αξιολόγηση Μοντέλων

Προκειμένου να συγκρίνουμε όλα τα μοντέλα μεταξύ τους εξετάσαμε την απόδοση τους με το συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα (sMAPE). Αρχικά χωρίζουμε τα δεδομένα μας κρατώντας κάποιες από τις τελευταίες τιμές τους, ώστε να τις συγκρίνουμε με τις αντίστοιχες προβλέψεις των μοντέλων μας και να βρούμε το in-sample σφάλμα. Εξετάζουμε την απόδοση των μεθόδων σε κάθε επίπεδο ξεχωριστά αλλά και συνολικά με το μέσο όρο τους. Το σφάλμα κάθε επιπέδου της ιεραρχίας προκύπτει από το μέσο όρο των σφαλμάτων όλων των χρονοσειρών που ανήκουν στο επίπεδο αυτό. Αντίστοιχα το σφάλμα κάθε χρονοσειράς προκύπτει από το μέσο όρο των σφαλμάτων πέντε περιόδων. Ο λόγος που το σφάλμα κάθε χρονοσειράς προκύπτει από πέντε περιόδους και όχι από μία είναι για να περιορίσουμε το ενδεχόμενο κάποια μέθοδος να δούλευε καλύτερα σε κάποια δεδομένα κατά τύχη, αλλά σε διαφορετικές περιόδους περιόδους να μην ανταποκρινόταν το ίδιο καλά.

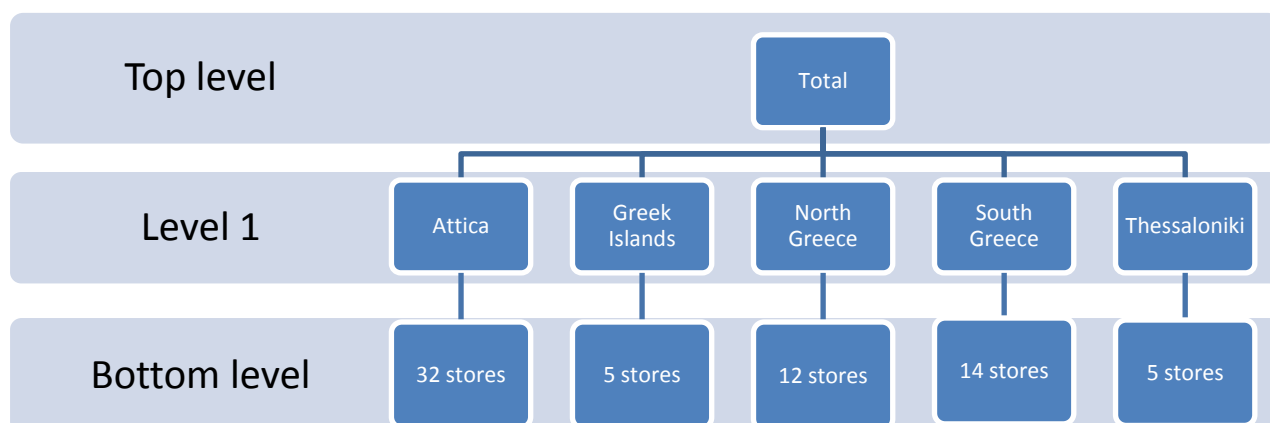


## Κεφάλαιο 5 : Case Study σε πωλήσεις αλυσίδας καταστημάτων λιανικής

### 5.1 Συλλογή και Επεξεργασία Δεδομένων

Στην παρούσα διπλωματική εξετάστηκαν οι μεθοδολογίες που αναφέρθηκαν παραπάνω, σε δεδομένα από μια αλυσίδα καταστημάτων. Συγκεκριμένα είχαμε τις πωλήσεις από 68 μαγαζιά για την περίοδο από 1/9/2018 έως 25/5/2019 εκτός των Κυριακών. Τα μαγαζιά αυτά ήταν κατανεμημένα σε πέντε γενικές περιοχές, την Αττική, τη Θεσσαλονίκη, τη Βόρεια Ελλάδα, τη Νότια Ελλάδα και τα Ελληνικά Νησιά. Με βάση αυτές τις γεωγραφικές περιοχές ορίστηκε και η ιεραρχία, ώστε να προκύψουν οι ιεραρχικές χρονοσειρές. Όπως αναφέρθηκε και παραπάνω, προκειμένου οι χρονοσειρές να έχουν όλες το ίδιο μήκος χρειάστηκε να συμπληρωθούν ορισμένες κενές τιμές. Αν υποθέσουμε την κενή τιμή  $n$  και εφόσον τα δεδομένα έχουν συχνότητα  $f=6$  (δεν έχουμε δεδομένα για Κυριακές), τότε οι τιμές αυτές προέκυψαν ως  $n = (n+6) \times (n-6) / 2$ .

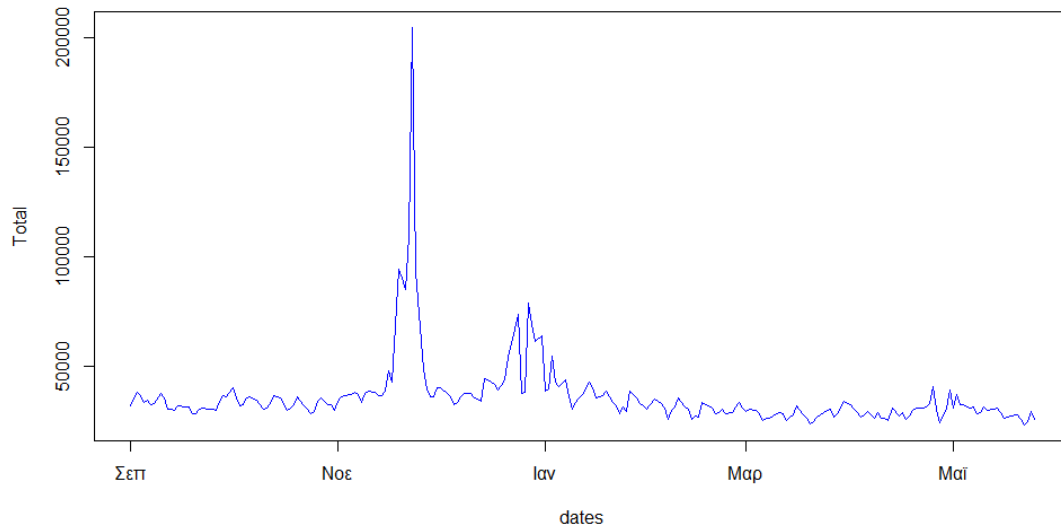
Η ιεραρχία που προέκυψε με αυτά τα δεδομένα είχε την παρακάτω μορφή :



Εικόνα 8 : Ιεραρχία δεδομένων case study

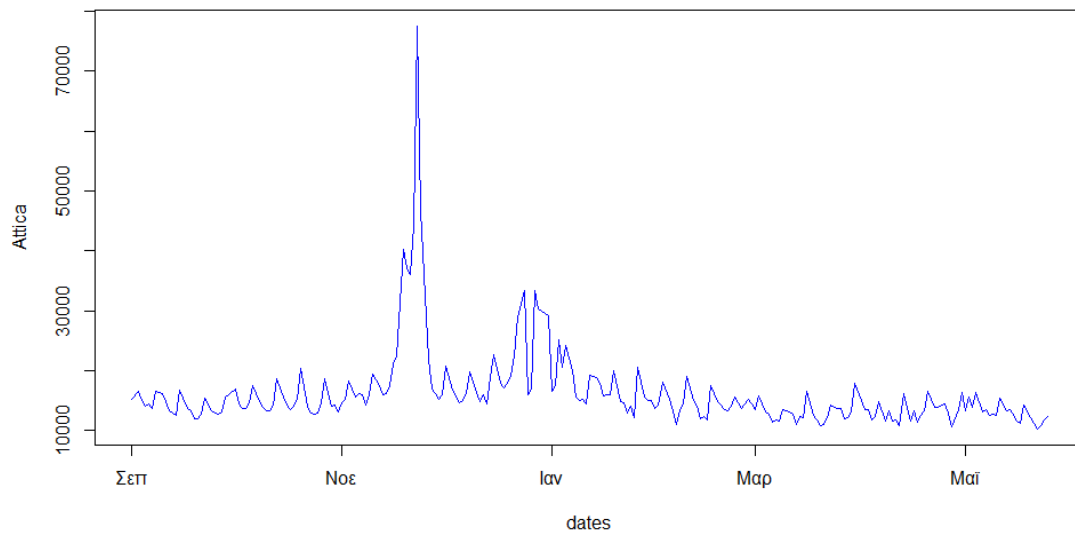
Στο επίπεδο  $K = 0$  έχουμε την τις συνολικές πωλήσεις και των 68 καταστημάτων των 5 περιοχών, επομένως έχουμε μία χρονοσειρά, την total. Στο επίπεδο  $K = 1$  έχουμε τις συνολικές πωλήσεις κάθε γεωγραφικής περιοχής ξεχωριστά, δηλαδή η Αττική θα είναι μια χρονοσειρά με τις συνολικές πωλήσεις όλων των μαγαζιών που βρίσκονται σε αυτή. Αντίστοιχα, προκύπτουν και οι χρονοσειρές των υπόλοιπων περιοχών, με

το επίπεδο 1 να έχει 5 διαφορετικές χρονοσειρές. Τέλος, στο επίπεδο  $K = 2$  έχουμε τις πωλήσεις των επιμέρους καταστημάτων, με το κάθε μαγαζί να μας δίνει μία χρονοσειρά. Προκύπτουν επομένως από αυτές τις συναθροίσεις των δεδομένων, 6 νέες χρονοσειρές και δουλεύουμε στις προβλέψεις μας με ένα σύνολο 74 χρονοσειρών. Ακολουθεί το διάγραμμα με τις συνολικές πωλήσεις όλων των καταστημάτων :

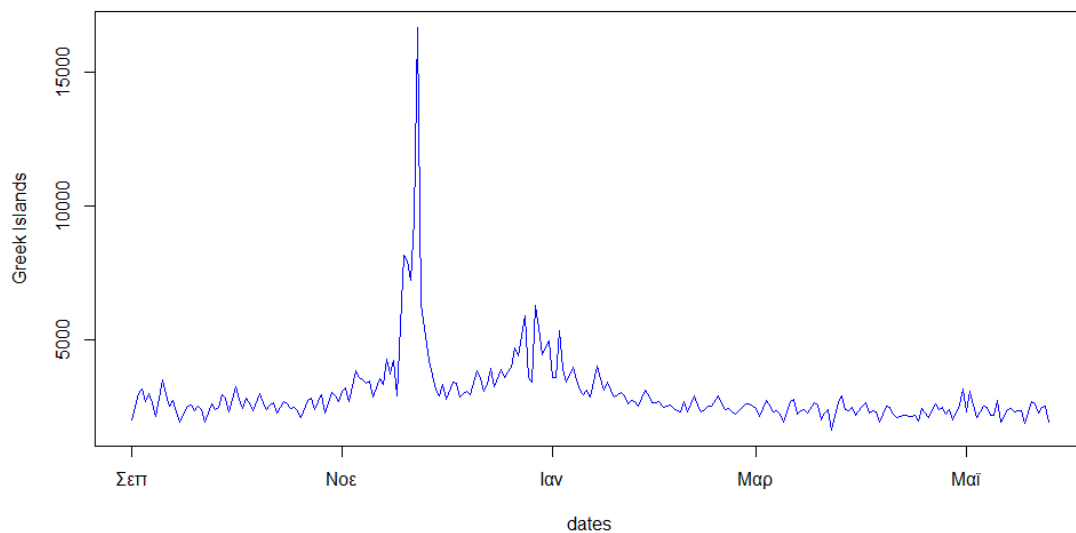


Εικόνα 9 : Χρονοσειρά συνολικής ζήτησης

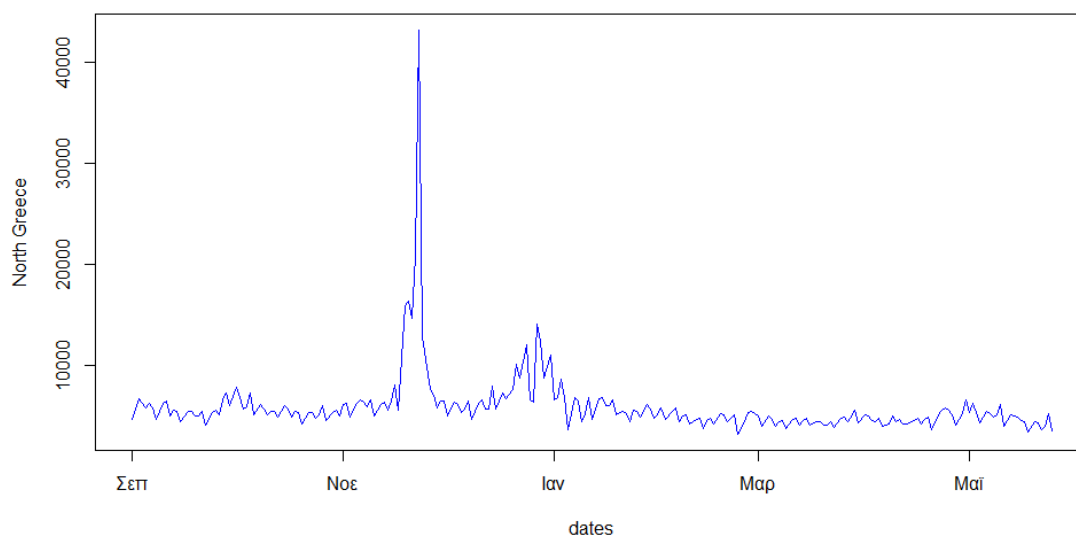
Καθώς και τα διαγράμματα των χρονοσειρών κάθε γεωγραφικής περιοχής:



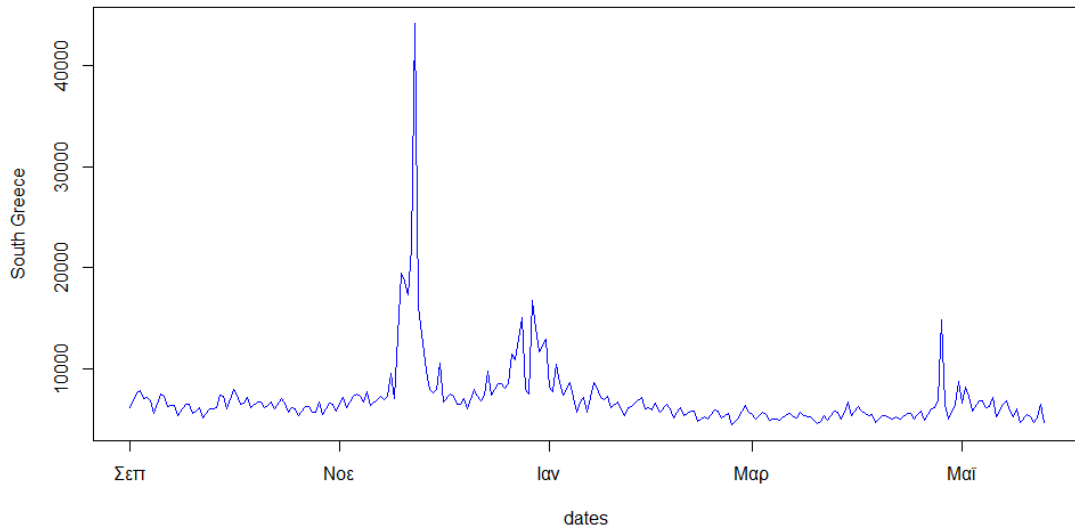
Εικόνα 10 : Χρονοσειρά ζήτησης Αττικής



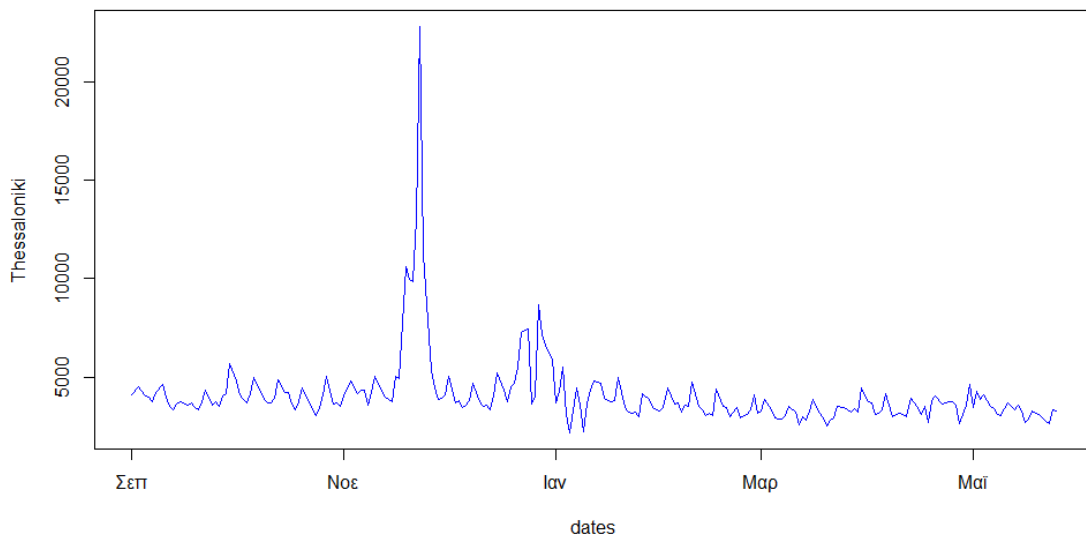
**Εικόνα 11 : Χρονοσειρά ζήτησης Ελληνικών νησιών**



**Εικόνα 12 : Χρονοσειρά ζήτησης Βόρειας Ελλάδας**



Εικόνα 13 : Χρονοσειρά ζήτησης Νότιας Ελλάδας



Εικόνα 14 : Χρονοσειρά ζήτησης Θεσσαλονίκης

Από τις παραπάνω χρονοσειρές ανά γεωγραφική περιοχή παρατηρούμε εύκολα ότι εμφανίζουν παρόμοια χαρακτηριστικά. Η κύρια διαφορά που έχουν οι χρονοσειρές κάθε περιοχής βρίσκεται στο επίπεδο τους. Αυτό είναι αναμενόμενο εφόσον οι περιοχές έχουν διαφορετικό αριθμό μαγαζιών και διαφορετικό αριθμό πελατών. Για παράδειγμα η Αττική είναι λογικό να έχει μεγαλύτερη ζήτηση λόγω του όγκου των καταστημάτων και των πελατών σε σχέση με τις υπόλοιπες περιοχές. Οι χρονοσειρές ανά περιοχή παρουσιάζουν παρόμοια μοτίβα με συνεχόμενες αυξομειώσεις της ζήτησης οι οποίες οφείλονται στην μέρα της εβδομάδας. Αξίζει να σημειωθεί ότι η χρονοσειρά της Νότιας Ελλάδας εμφανίζει στα τέλη Απριλίου μια



ακόμα αύξηση η οποία δεν εμφανίζεται στις άλλες περιοχές. Αυτό πιθανότατα να οφείλεται σε κάποιο γεγονός που οδήγησε στην αύξηση της ζήτησης μόνο σε αυτή την γεωγραφική περιοχή, ενώ οι υπόλοιπες δεν επηρεάστηκαν.

Τα διαγράμματα των bottom-level χρονοσειρών δεν παρουσιάζονται λόγω μεγάλου όγκου, όμως κατά κύριο λόγο ακολουθούν το μοτίβο της αντίστοιχης γεωγραφικής περιοχής που εντάσσονται. Είναι φανερό από τα διαγράμματα ότι γενικά οι χρονοσειρές δεν εμφανίζουν κάποια ανοδική ή καθοδική τάση. Ακόμα εκτός από κάποιες περιόδους που είναι αναμενόμενο να αυξηθούν οι πωλήσεις, όπως τα Χριστούγεννα, φαίνεται ότι στα τέλη Νοέμβρη υπάρχει πιθανότατα κάποιο special event καθώς εμφανίζεται μια πολύ μεγάλη αύξηση. Η αύξηση αυτή είναι πιθανό να επηρεάσει και την απόδοση των μεθόδων σε κάποιο βαθμό. Με βάση τα δεδομένα των 74 χρονοσειρών και την ιεραρχία που αναλύθηκε, αναπτύχθηκαν και εξετάστηκαν τα μοντέλα που θα δούμε παρακάτω.

## 5.2 Μέθοδοι Αναφοράς(benchmark)

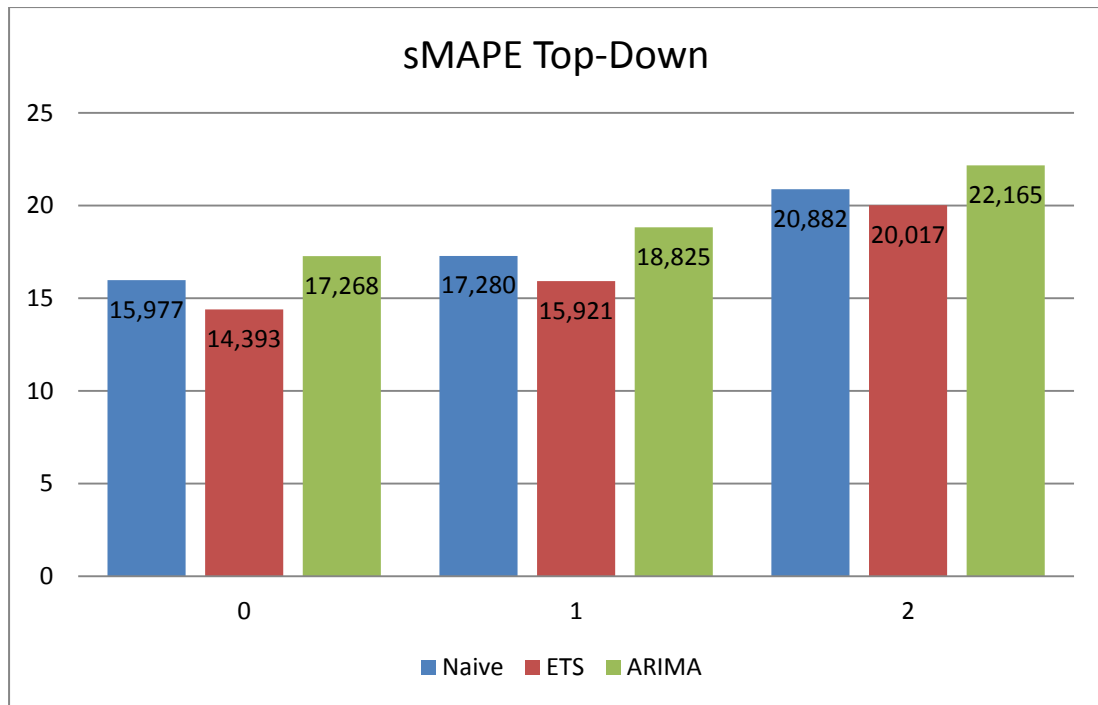
Προκειμένου να έχουμε ένα σημείο αναφοράς ,ώστε να συγκρίνουμε με αυτό τις μεθόδους μηχανικής μάθησης, παραγάγαμε προβλέψεις με τρεις διαφορετικές μεθόδους. Με την naive, με τα state space μοντέλα-ets και με τα μοντέλα arima. Οι προβλέψεις κάθε μοντέλου με τη σειρά τους συμφιλώθηκαν με τρεις διαφορετικές βασικές μεθόδους ιεραρχικής συνάθροισης, τις top-down, bottom-up και optimal reconciliation. Προκειμένου να συγκρίνουμε τα μοντέλα μας χρησιμοποιήθηκε ο δείκτης sMAPE. Για λόγους μεγαλύτερης ακρίβειας χρησιμοποιήθηκαν για την εξαγωγή των σφαλμάτων οι τελευταίες 60 τιμές των δεδομένων χωρισμένες σε πέντε δωδεκάδες, αφού ο ορίζοντας πρόβλεψης ήταν 12,με τα τελικά σφάλματα να προκύπτουν από το μέσο όρο των 5 αυτών σφαλμάτων κάθε φορά. Με αυτόν τον τρόπο ελέγχουμε τα σφάλματα σε μεγαλύτερο εύρος χρόνου και μειώνουμε την πιθανότητα κάποια μέθοδος να εμφανίζει καλύτερα αποτελέσματα κατά τύχη. Παρακάτω παρουσιάζονται τα σφάλματα sMAPE των μεθόδων:

Σφάλμα sMAPE			
Top-Down			
	Naive	ETS	ARIMA
<b>Level 0</b>	15,97699	<b>14,39327</b>	17,26768
<b>Level 1</b>	17,27959	<b>15,92052</b>	18,82489
<b>Level 2</b>	20,88178	<b>20,01706</b>	22,16543
Average	18,04612	<b>16,77695</b>	19,41933
Bottom-Up			
	Naive	ETS	ARIMA
<b>Level 0</b>	15,97699	<b>11,75367</b>	13,63812
<b>Level 1</b>	18,47013	<b>11,68944</b>	14,56445
<b>Level 2</b>	24,52556	<b>16,04994</b>	19,32666
Average	19,65756	<b>13,16435</b>	15,84308
Optimal			
	Naive	ETS	ARIMA
<b>Level 0</b>	15,97699	<b>14,00027</b>	17,31568
<b>Level 1</b>	18,47013	<b>14,34153</b>	19,16199
<b>Level 2</b>	24,52556	<b>18,54572</b>	25,80937
Average	19,65756	<b>15,62917</b>	20,76235

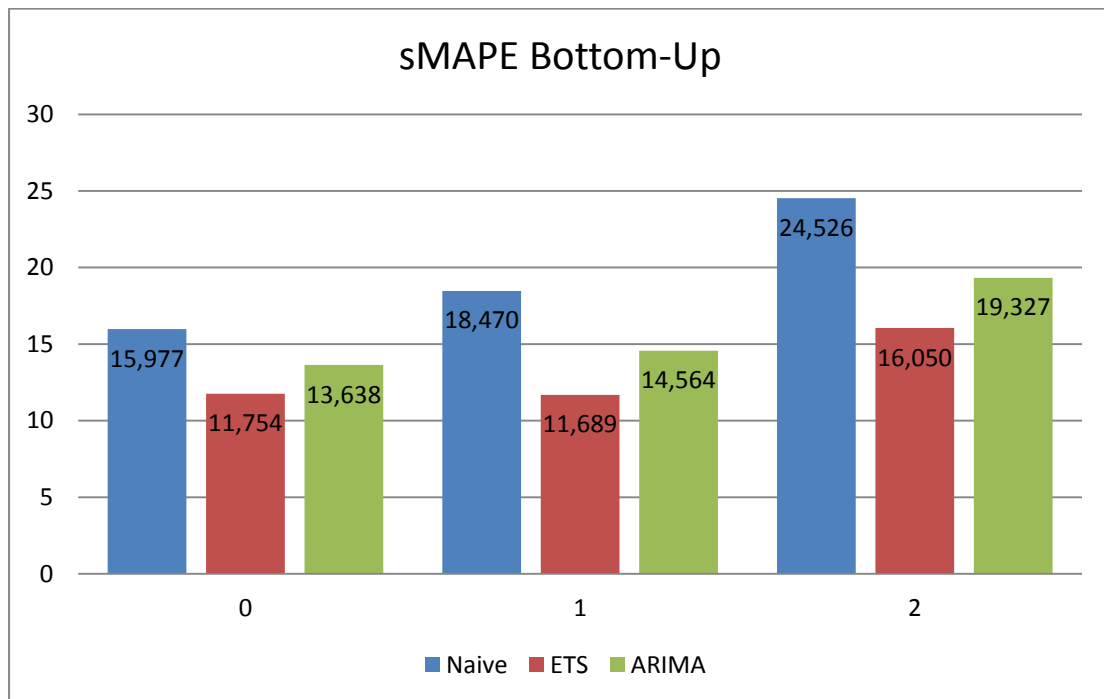
Πίνακας 1 : Σφάλματα μεθόδων αναφοράς

Στον παραπάνω πίνακα φαίνονται τα σφάλματα για κάθε ιεραρχική μέθοδο με την παραγωγή των βασικών προβλέψεων από τις μεθόδους Naive,ETS,ARIMA. Αναγράφονται τα σφάλματα σε κάθε επίπεδο, καθώς και ο μέσος όρος των τριών επιπέδων, ενώ σημειώνονται πιο έντονα οι χαμηλότερες τιμές ανά επίπεδο σε κάθε ιεραρχική μέθοδο.

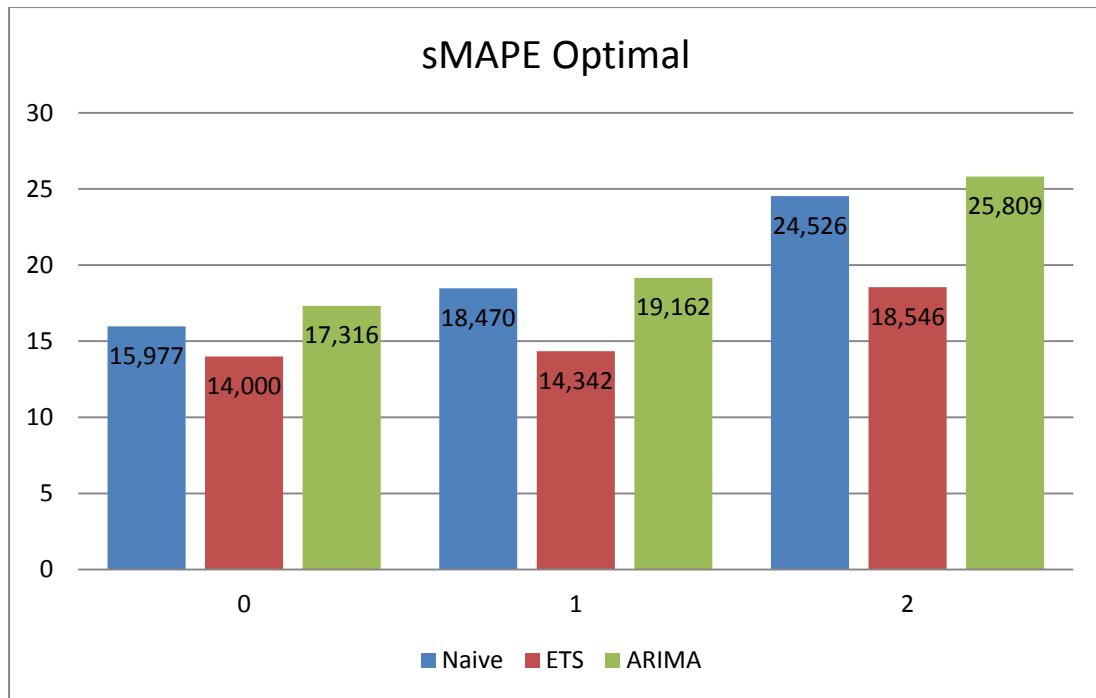
Τα παρακάτω τρία διαγράμματα παρουσιάζουν συγκριτικά τα σφάλματα των τριών μεθόδων βασικής πρόβλεψης, ανά επίπεδο, σε κάθε ιεραρχική μέθοδο.



Εικόνα 15 : Σφάλματα ανά επίπεδο, μεθόδου top-down με προβλέψεις naive,ets και arima

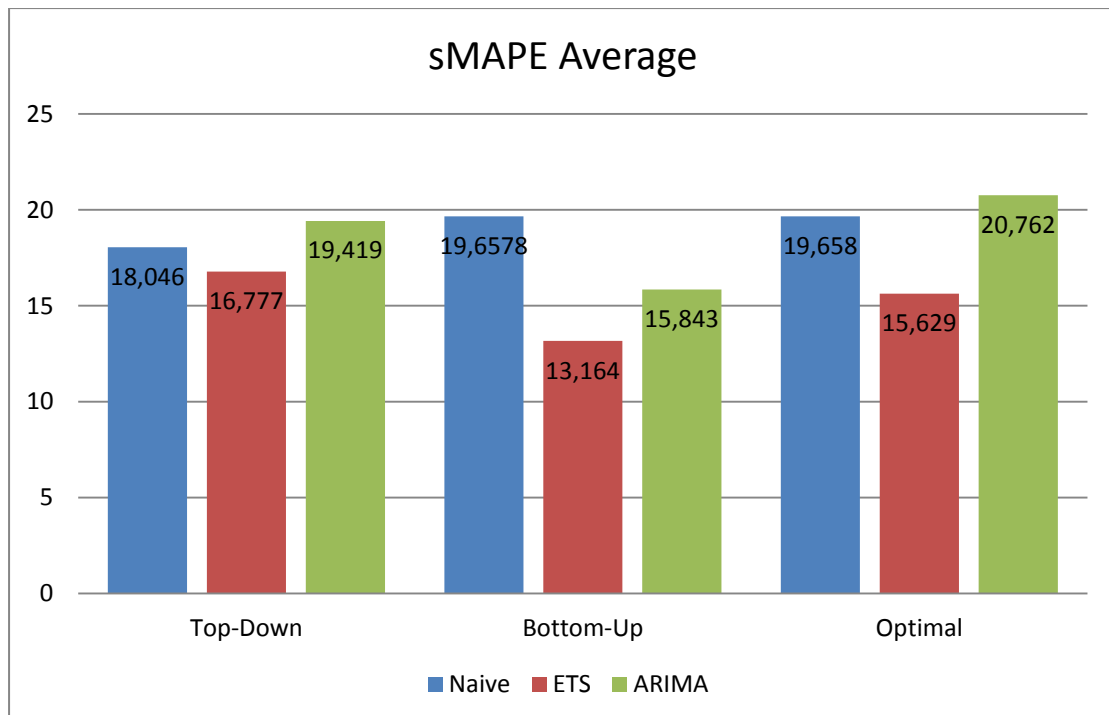


Εικόνα 16 : Σφάλματα ανά επίπεδο, μεθόδου bottom-up με προβλέψεις naive,ets και arima



Εικόνα 17 : Σφάλματα ανά επίπεδο, μεθόδου optimal με προβλέψεις naive,ets και arima

Όπως φαίνεται και από τα διαγράμματα η ETS δίνει εμφανώς καλύτερα αποτελέσματα σε όλα τα επίπεδα της ιεραρχίας και σε όλες τις ιεραρχικές μεθόδους. Στις top-down και optimal reconciliation μεθόδους τη δεύτερη καλύτερη απόδοση έχει η naive και ακολουθούν τα μοντέλα arima ενώ γίνεται ακριβώς το αντίθετο στην bottom-up, για όλα τα επίπεδα της ιεραρχίας.



Εικόνα 18 : Σύγκριση naive, ets και arima για κάθε ιεραρχική μέθοδο

Το παραπάνω διάγραμμα παρουσιάζει συγκριτικά τα μέσα σφάλματα όλων των επιπέδων των μεθόδων πρόβλεψης και για τι τρεις μεθόδους ιεραρχικής συνάθροισης. Είναι φανερό ότι για την συμφιλίωση των προβλέψεων με την ιεραρχική μέθοδο bottom-up έχουμε καλύτερα αποτελέσματα ενώ οι optimal και top-down έχουν μικρές διαφορές ανάλογα με το ποια μέθοδος χρησιμοποιείται για την παραγωγή των βασικών προβλέψεων.

### **5.3 Μέθοδοι Μηχανικής Μάθησης**

Μετά την υλοποίηση των μεθόδων αναφοράς εξετάστηκαν διάφορες προσεγγίσεις βασισμένες στα μοντέλα μηχανικής μάθησης και πιο συγκεκριμένα στα πολυεπίπεδα αντίληπτρα(MLP). Η πρώτη προσέγγιση ήταν να δημιουργηθεί ένα διαφορετικό νευρωνικό δίκτυο για κάθε χρονοσειρά ώστε να παράγουμε τις βασικές προβλέψεις και μετά να τις συμφιλιώσουμε με τις παραπάνω ιεραρχικές μεθόδους. Αντίθετα η δεύτερη προσέγγιση εξέταζε τη δημιουργία ενός ενιαίου πολυεπίπεδου αντίληπτρου για την παραγωγή προβλέψεων σε όλες τις χρονοσειρές και την μετέπειτα εφαρμογή των ιεραρχικών μεθόδων σε αυτές. Όσον αφορά την εκπαίδευση των μοντέλων πραγματοποιήθηκαν διάφορες δοκιμές ώστε να προκύψει ποια είναι η καλύτερη προσέγγιση για τα δεδομένα μας. Οι δοκιμές αφορούσαν κυρίως τον αριθμό των εισόδων και τη συνάρτηση ενεργοποίησης. Πιο συγκεκριμένα δοκιμάστηκε η εκπαίδευση των νευρωνικών δικτύων με 3, 6 και 12 εισόδους και συνάρτηση ενεργοποίησης υπερβολική εφαπτομένη και σιγμοειδή.

Οι δύο επόμενοι πίνακες αφορούν την προσέγγιση στην οποία δημιουργήσαμε ένα πολυεπίπεδο αντίληπτρο για την κάθε χρονοσειρά ξεχωριστά, παράγαμε τις βασικές προβλέψεις και τις συμφιλιώσαμε με τις τρεις γνωστές ιεραρχικές μεθόδους. Ο πρώτος πίνακας αφορά τα νευρωνικά δίκτυα με συνάρτηση ενεργοποίησης την υπερβολική εφαπτομένη ενώ ο δεύτερος με συνάρτηση ενεργοποίησης την σιγμοειδή. Παρουσιάζονται τα σφάλματα ανά επίπεδο και κατά μέσο όρο για όλους τους συνδυασμούς που ελέγχθηκαν.

Σφάλμα sMAPE-MLP with TanH			
Top-Down			
	3 inputs	6 inputs	12 inputs
<b>Level 0</b>	10,65398	<b>8,77322</b>	10,29355
<b>Level 1</b>	13,02951	<b>11,19566</b>	12,25199
<b>Level 2</b>	17,66469	<b>16,33708</b>	17,05121
Average	13,78273	<b>12,10199</b>	13,19892
Bottom-Up			
	3 inputs	6 inputs	12 inputs
<b>Level 0</b>	11,17531	11,75452	<b>11,00123</b>
<b>Level 1</b>	13,22584	13,14609	<b>12,48466</b>
<b>Level 2</b>	17,25478	16,21522	<b>16,20212</b>
Average	13,88531	13,70528	<b>13,22934</b>
Optimal			
	3 inputs	6 inputs	12 inputs
<b>Level 0</b>	10,60975	<b>9,08128</b>	10,26991
<b>Level 1</b>	13,29232	<b>11,67173</b>	12,30627
<b>Level 2</b>	17,79584	18,69915	<b>16,46377</b>
Average	13,89930	13,15072	<b>13,01332</b>

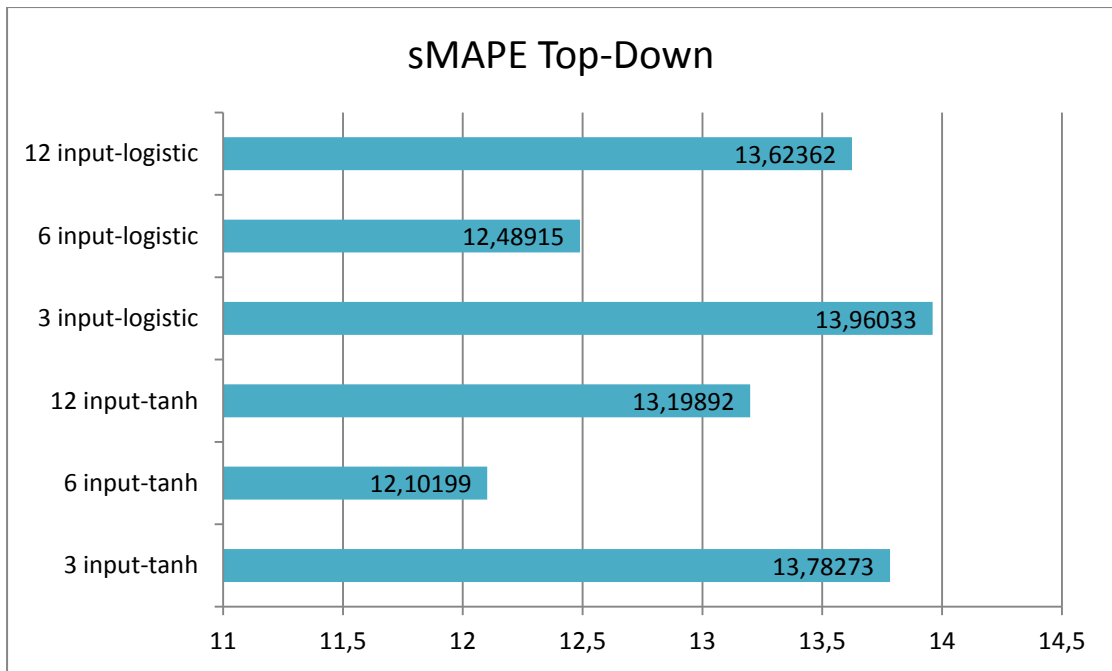
Πίνακας 2 : Σφάλματα για χρήση 74 mlp με tanh

Παρατηρούμε ότι η top-down λειτουργεί σε όλα τα επίπεδα καλύτερα, όταν τα νευρωνικά δίκτυα έχουν εκπαιδευθεί με εξάδες εισόδων ενώ η bottom-up λειτουργεί καλύτερα με νευρωνικά 12 εισόδων. Η optimal δίνει καλύτερα αποτελέσματα στα επίπεδα για  $K=0,1$  για νευρωνικά 6 εισόδων, ενώ στο επίπεδο για  $K=2$  λειτουργεί καλύτερα με νευρωνικά 12 εισόδων. Τα ίδια ισχύουν και για τις δύο συναρτήσεις ενεργοποίησης με μικρές διαφορές στις τιμές των σφαλμάτων.

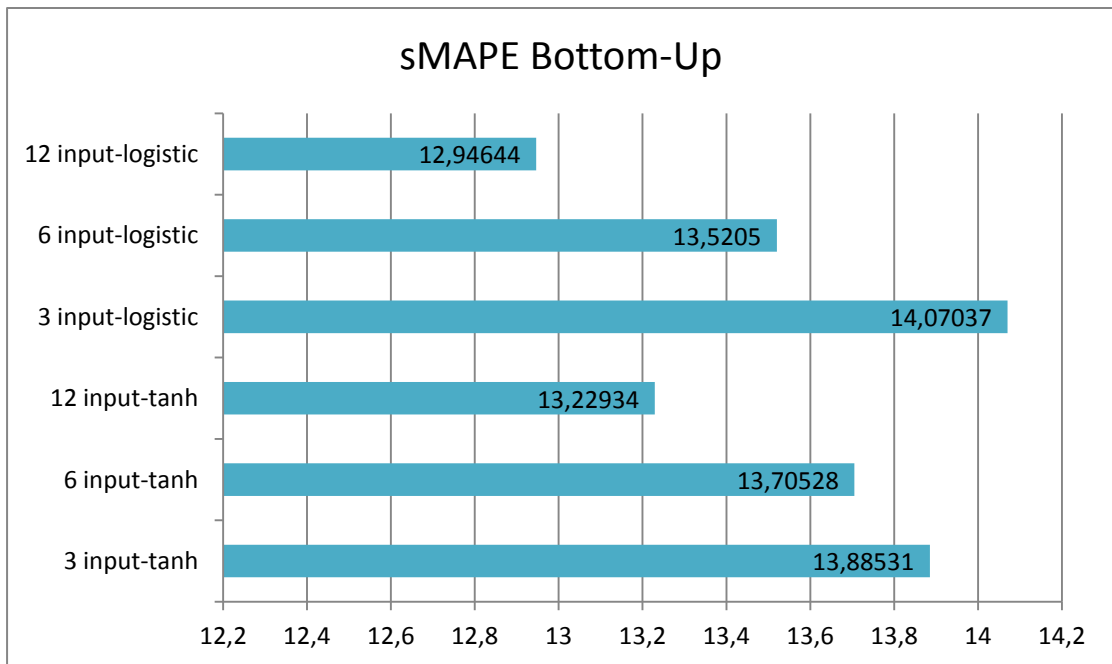
Σφάλμα sMAPE-MLP with Logistic			
Top-Down			
	3 inputs	6 inputs	12 inputs
<b>Level 0</b>	10,81971	<b>9,32700</b>	10,98919
<b>Level 1</b>	13,23192	<b>11,55486</b>	12,56321
<b>Level 2</b>	17,82935	<b>16,58560</b>	17,31846
Average	13,96033	<b>12,48915</b>	13,62362
Bottom-Up			
	3 inputs	6 inputs	12 inputs
<b>Level 0</b>	11,48388	11,48336	<b>10,62593</b>
<b>Level 1</b>	13,47412	12,93554	<b>12,25087</b>
<b>Level 2</b>	17,25312	16,14261	<b>15,96251</b>
Average	14,07037	13,52050	<b>12,94644</b>
Optimal			
	3 inputs	6 inputs	12 inputs
<b>Level 0</b>	10,71721	<b>9,48815</b>	10,82154
<b>Level 1</b>	13,90565	<b>11,64139</b>	12,61442
<b>Level 2</b>	18,16016	18,90183	<b>16,42602</b>
Average	14,26101	13,34379	<b>13,28733</b>

Πίνακας 3 : Σφάλματα για χρήση 74 mlp με logistic

Τα τρία επόμενα διαγράμματα παρουσιάζουν τα μέσα σφάλματα για κάθε ιεραρχική μέθοδο με τους 6 τρόπους εκπαίδευσης των πολυεπίπεδων αντίληπτρων.

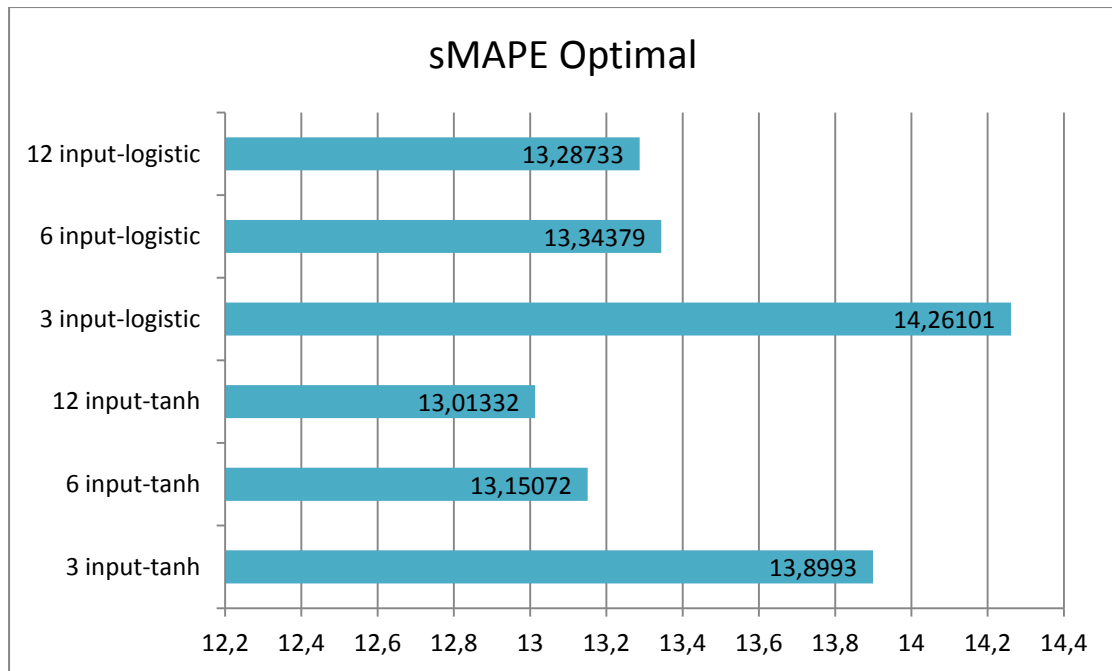


Εικόνα 19 : Σύγκριση σφαλμάτων της top-down με 74 mlp για διάφορα input sets



Εικόνα 20 : Σύγκριση σφαλμάτων της bottom-up με 74 mlp για διάφορα input sets





Εικόνα 21: Σύγκριση σφαλμάτων της optimal με 74 mlp για διάφορα input sets

Προκύπτει από τα παραπάνω ότι καλύτερα αποτελέσματα δίνει η top-down όταν οι βασικές προβλέψεις παράγονται με νευρωνικά δίκτυα 6 εισόδων. Ακολουθούν οι optimal και bottom-up για νευρωνικά με εκπαίδευση 12 εισόδων, ενώ η εκπαίδευση με 3 εισόδους δίνει τα χειρότερα αποτελέσματα ανεξάρτητα από την ιεραρχική μέθοδο συμφιλίωσης των προβλέψεων.

Όσον αφορά τη δεύτερη προσέγγιση, δηλαδή τη δημιουργία ενός ενιαίου αντίληπτρου για την παραγωγή των βασικών προβλέψεων έγιναν δύο κύριες δοκιμές. Στη πρώτη εκπαιδεύθηκε ένα νευρωνικό από όλα τα δεδομένα όλων των χρονοσειρών, ενώ στη δεύτερη η εκπαίδευση έγινε από λιγότερα τυχαία επιλεγμένα δεδομένα από όλες τις χρονοσειρές. Όπως είναι φανερό στην πρώτη περίπτωση έχουμε περισσότερα δεδομένα και κατ' επέκταση η διαδικασία της εκπαίδευσης του δικτύου είναι αρκετά χρονοβόρα, κάτι που δεν ισχύει στην δεύτερη περίπτωση αφού δεν χρησιμοποιούμε όλες τις τιμές των χρονοσειρών. Ακολουθούν τα σφάλματα των μεθόδων που εξετάστηκαν.

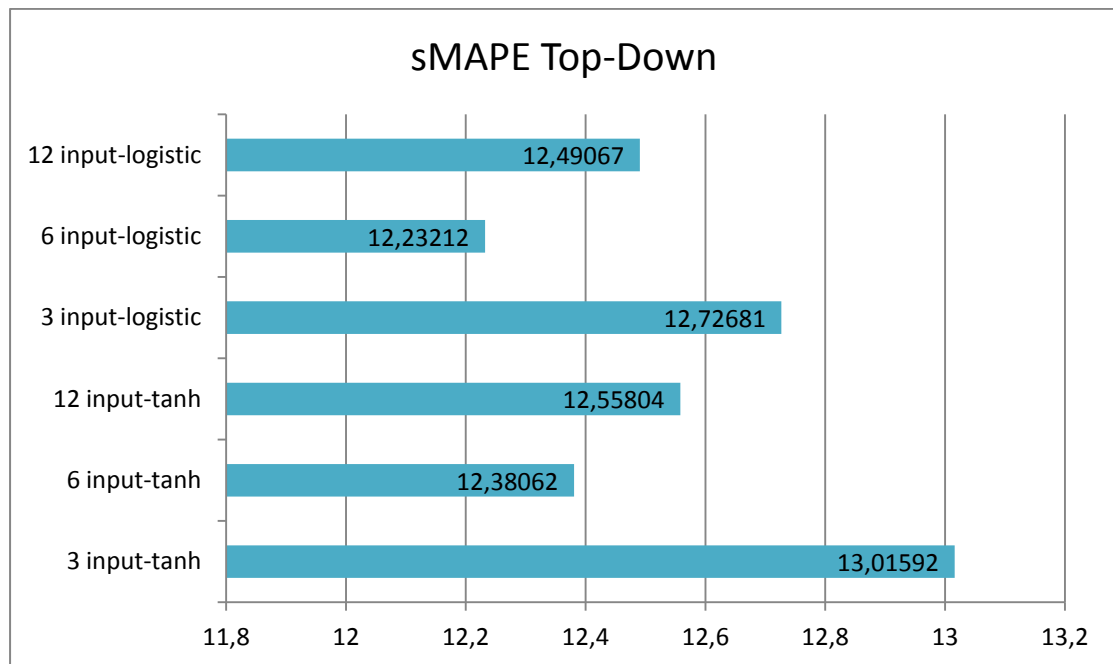
Σφάλμα sMAPE-one MLP with all inputs(tanh)			
Top-Down			
	3 inputs	6 inputs	12 inputs
Level 0	9,90746	<b>9,18995</b>	9,60405
Level 1	12,15563	<b>11,48164</b>	11,49534
Level 2	16,98468	<b>16,47027</b>	16,57472
Average	13,01592	<b>12,38062</b>	12,55804
Bottom-Up			
	3 inputs	6 inputs	12 inputs
Level 0	10,35785	9,88501	<b>9,66255</b>
Level 1	12,80456	11,65239	<b>11,28006</b>
Level 2	16,76403	<b>15,49505</b>	15,62335
Average	13,30881	12,34415	<b>12,18865</b>
Optimal			
	3 inputs	6 inputs	12 inputs
Level 0	9,98891	<b>9,23063</b>	9,58096
Level 1	12,20831	<b>11,11345</b>	11,15876
Level 2	16,50763	<b>15,20199</b>	15,90418
Average	12,90162	<b>11,84869</b>	12,21463

Πίνακας 4 : Σφάλματα για χρήση ενός mlp εκπαιδευμένου από όλα τα δεδομένα με tanh

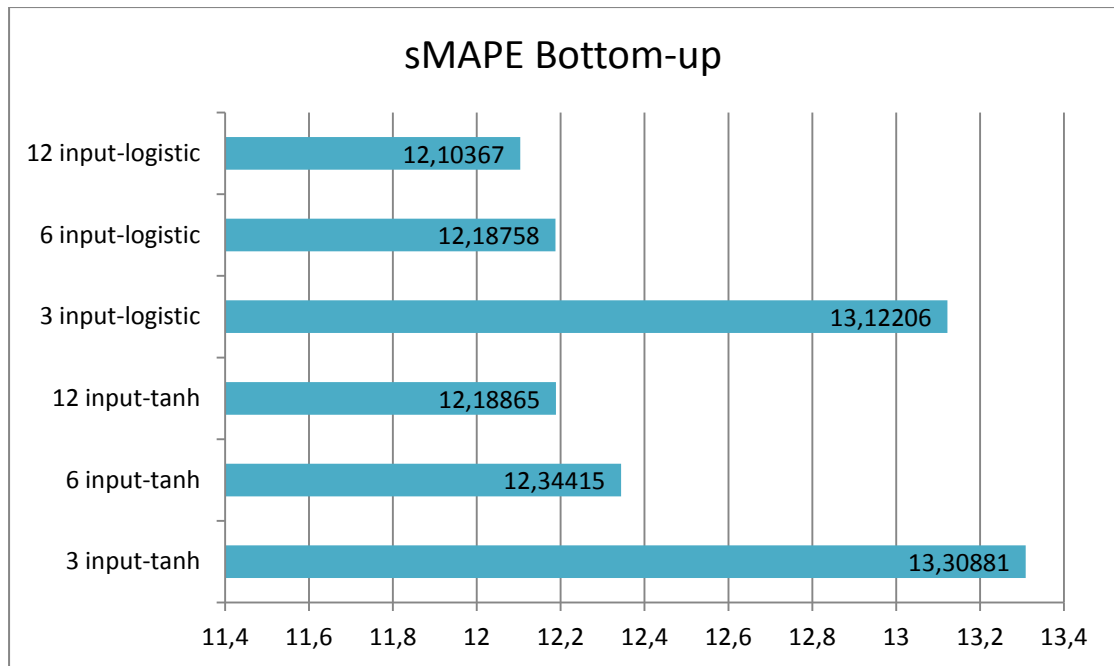
Σφάλμα sMAPE-one MLP with all inputs(logistic)			
Top-Down			
	3 inputs	6 inputs	12 inputs
Level 0	9,53425	<b>8,99807</b>	9,47698
Level 1	11,85724	<b>11,32346</b>	11,44950
Level 2	16,78893	<b>16,37484</b>	16,54552
Average	12,72681	<b>12,23212</b>	12,49067
Bottom-Up			
	3 inputs	6 inputs	12 inputs
Level 0	10,13872	9,68984	<b>9,67213</b>
Level 1	12,63969	11,53233	<b>11,22455</b>
Level 2	16,58778	<b>15,34058</b>	15,41432
Average	13,12206	12,18758	<b>12,10367</b>
Optimal			
	3 inputs	6 inputs	12 inputs
Level 0	9,62438	<b>9,04889</b>	9,46600
Level 1	12,00541	<b>10,94860</b>	11,06796
Level 2	16,33356	<b>14,97964</b>	15,48710
Average	12,65445	<b>11,65904</b>	12,00702

Πίνακας 5 : Σφάλματα για χρήση ενός mlp εκπαιδευμένου από όλα τα δεδομένα με logistic

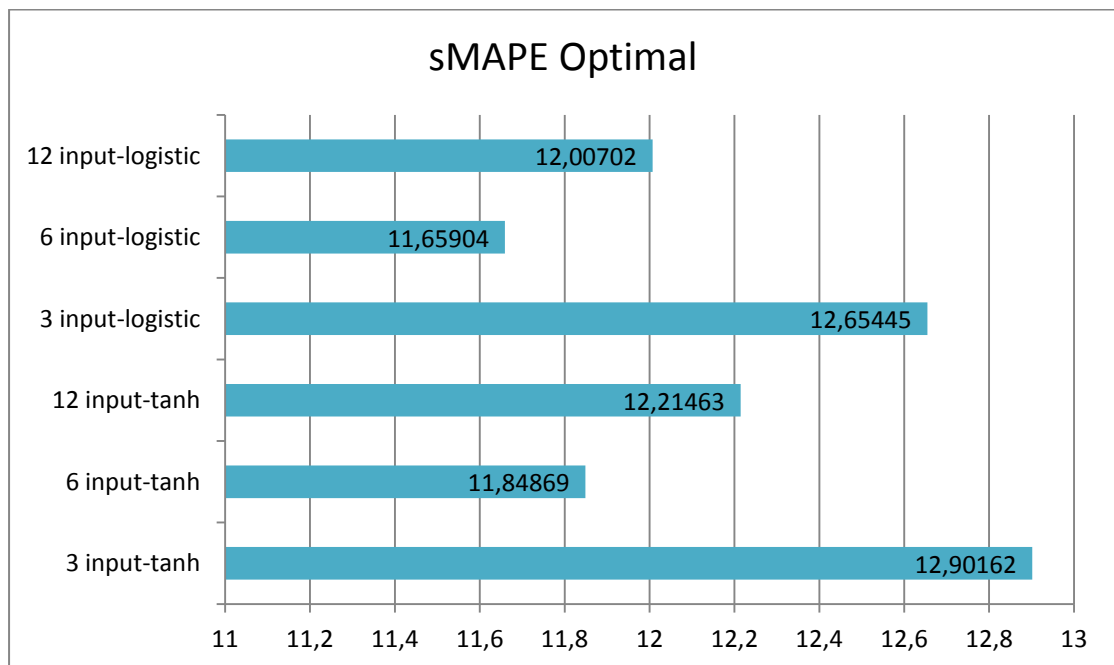
Επισημαίνονται πιο έντονα τα μικρότερα σφάλματα ανά επίπεδο κάθε ιεραρχικής μεθόδου. Από τα προηγούμενα φαίνεται πως πάλι η εκπαίδευση των νευρωνικών δικτύων με 3 εισόδους δεν αποδίδει τόσο καλά όσο οι άλλες δύο, ενώ πάλι παρατηρείται πως η bottom-up δουλεύει καλύτερα με νευρωνικό 12 εισόδων ενώ οι άλλες με 6 εισόδων. Ακόμα παρατηρείται μια μείωση στα σφάλματα όλων των μεθόδων ενώ αυτή τη φορά τα καλύτερα αποτελέσματα δίνει η optimal reconciliation σε αντίθεση με πριν. Όλα αυτά φαίνονται πιο καθαρά στα επόμενα διαγράμματα.



Εικόνα 22 : Σύγκριση σφαλμάτων της top-down με ένα mlp, εκπαιδευμένο από όλα τα δεδομένα, για διάφορα input sets



Εικόνα 23: Σύγκριση σφαλμάτων της bottom-up με ένα mlr,εκπαιδευμένο από όλα τα δεδομένα, για διάφορα input sets



Εικόνα 24: Σύγκριση σφαλμάτων της optimal με ένα mlr,εκπαιδευμένο από όλα τα δεδομένα, για διάφορα input sets

Από τα παραπάνω φαίνεται πως στην συγκεκριμένη μέθοδο παραγωγής προβλέψεων, όπου εκπαιδεύουμε μόνο ένα ενιαίο νευρωνικό δίκτυο, καλύτερα αποτελέσματα δίνει η logistic και όχι η tanh ως συνάρτηση ενεργοποίησης του κρυφού επιπέδου.

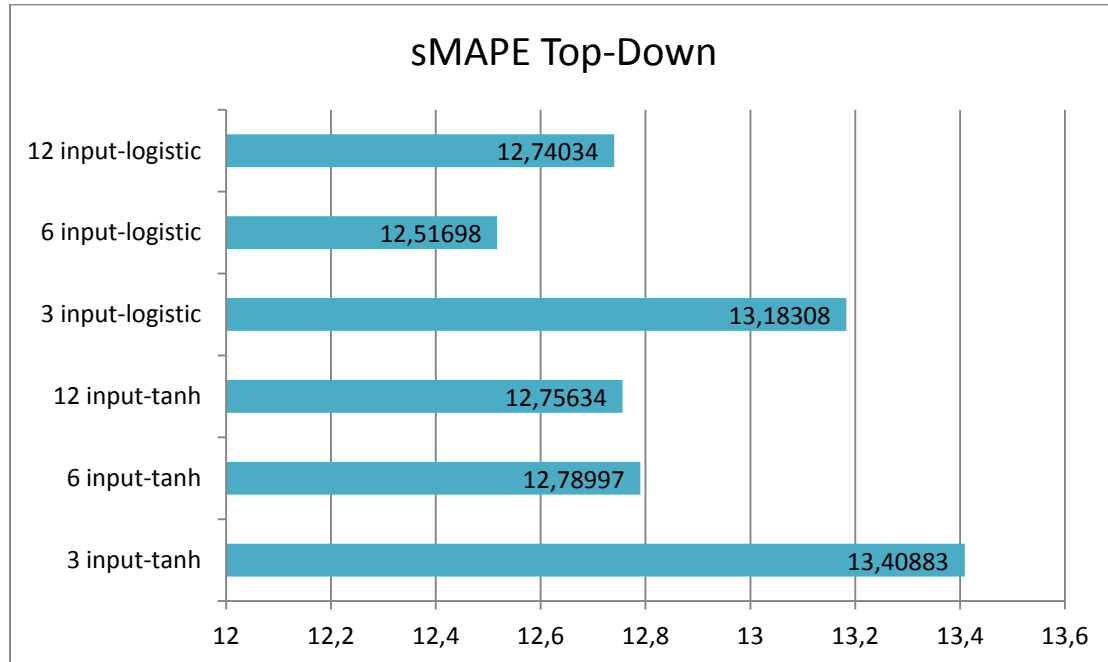
Τέλος ακολουθούν τα σφάλματα της προσέγγισης όπου το νευρωνικό δίκτυο εκπαιδεύθηκε με τυχαίες εισόδους από όλες τις χρονοσειρές και μικρότερο αριθμό από αυτές, αντί να χρησιμοποιηθούν όλες οι τιμές.

Σφάλμα sMAPE-one MLP with random inputs(tanh)			
Top-Down			
	3 inputs	6 inputs	12 inputs
<b>Level 0</b>	10,40363	<b>9,76051</b>	9,81403
<b>Level 1</b>	12,57494	11,86521	<b>11,73754</b>
<b>Level 2</b>	17,24793	16,74419	<b>16,71744</b>
Average	13,40883	12,78997	<b>12,75634</b>
Bottom-Up			
	3 inputs	6 inputs	12 inputs
<b>Level 0</b>	10,67448	<b>10,15092</b>	10,20119
<b>Level 1</b>	12,82577	<b>11,68978</b>	11,72751
<b>Level 2</b>	17,00082	<b>15,81944</b>	15,87645
Average	13,50036	<b>12,55338</b>	12,60172
Optimal			
	3 inputs	6 inputs	12 inputs
<b>Level 0</b>	10,43727	<b>9,79445</b>	9,81389
<b>Level 1</b>	12,42939	11,61866	<b>11,4871</b>
<b>Level 2</b>	16,85147	<b>15,93872</b>	16,02411
Average	13,23938	12,45061	<b>12,4417</b>

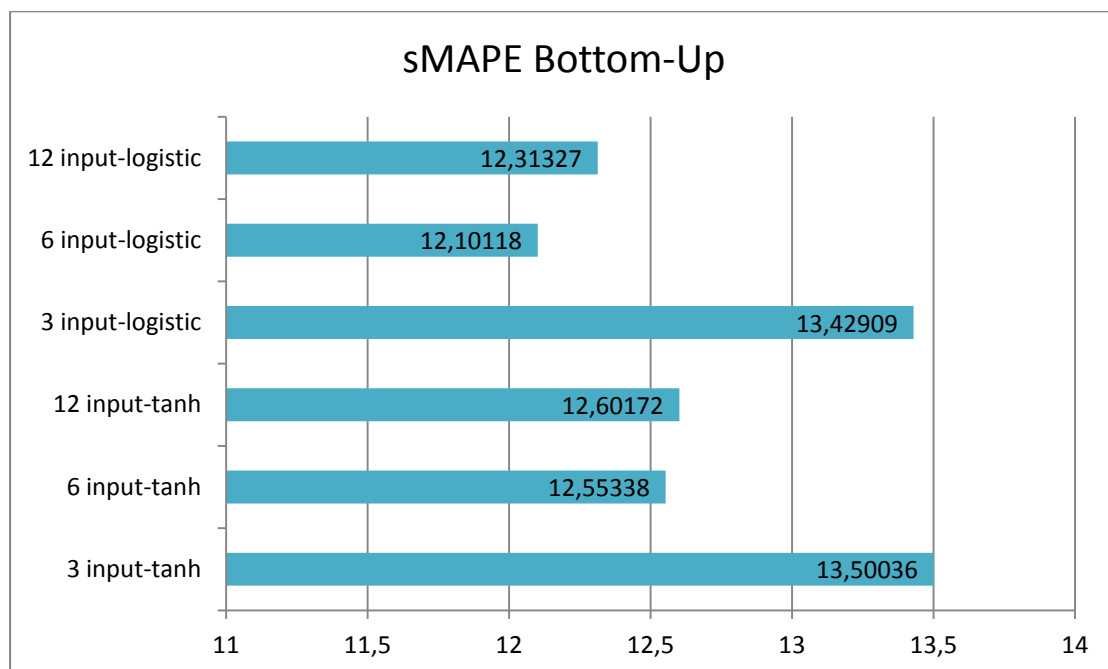
Πίνακας 6 : Σφάλματα για χρήση ενός mlp εκπαιδευμένου από τυχαία δεδομένα με tanh

Σφάλμα sMAPE-one MLP with random inputs(logistic)			
Top-Down			
	3 inputs	6 inputs	12 inputs
Level 0	10,13719	<b>9,37659</b>	9,82764
Level 1	12,31804	<b>11,59825</b>	11,67402
Level 2	17,09401	<b>16,57609</b>	16,71937
Average	13,18308	<b>12,51698</b>	12,74034
Bottom-Up			
	3 inputs	6 inputs	12 inputs
Level 0	10,53764	<b>9,49942</b>	9,89508
Level 1	12,78932	11,38693	<b>11,38496</b>
Level 2	16,96032	<b>15,41719</b>	15,65976
Average	13,42909	<b>12,10118</b>	12,31327
Optimal			
	3 inputs	6 inputs	12 inputs
Level 0	10,19281	<b>9,35864</b>	9,80135
Level 1	12,29358	11,38554	<b>11,36143</b>
Level 2	16,76728	<b>15,56355</b>	15,8778
Average	13,08456	<b>12,10258</b>	12,34686

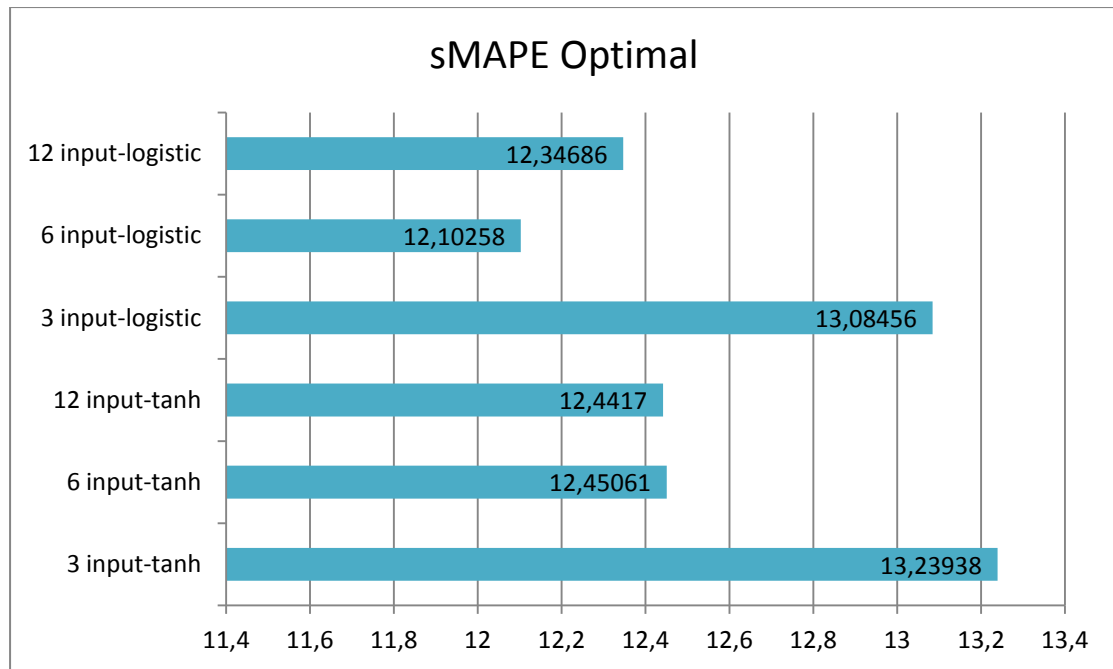
Πίνακας 7 : Σφάλματα για χρήση ενός mlp εκπαιδευμένου από τυχαία δεδομένα με logistic



Εικόνα 25 : Σύγκριση σφαλμάτων της top-down με ένα mlp, εκπαιδευμένο από τυχαία δεδομένα, για διάφορα input sets



Εικόνα 26 : Σύγκριση σφαλμάτων της bottom-up με ένα mlp, εκπαιδευμένο από τυχαία δεδομένα, για διάφορα input sets



Εικόνα 27: Σύγκριση σφαλμάτων της optimal γιμ ένα mlp,εκπαιδευμένο από τυχαία δεδομένα, για διάφορα input sets

Όπως ήταν λογικό, οι top-down μέθοδοι δεν έδωσαν καλύτερα αποτελέσματα, εφόσον στην εκπαίδευση του νευρωνικού συμμετείχαν και τιμές από τις χρονοσειρές των ψαμηλότερων επιπέδων. Αυτό ήταν αναμενόμενο και οι μέθοδοι εξετάστηκαν κυρίως για λόγους επαλήθευσης. Όσον αφορά τις bottom-up και optimal μεθόδους δίνουν αποτελέσματα λίγο χειρότερα σε σχέση με την εκπαίδευση του νευρωνικού δικτύου από όλες τις τιμές. Το πλεονέκτημα της εκπαίδευσης του νευρωνικού με τυχαίες τιμές από όλες τις χρονοσειρές αντί για όλες τις τιμές τους είναι πως μειώνεται σε μεγάλο βαθμό ο χρόνος εκπαίδευσης ενώ ταυτόχρονα δίνουν αρκετά καλά αποτελέσματα. Το μειονέκτημα όμως της προσέγγισης αυτής είναι πως το μοντέλο δεν έχει τόσο σταθερή απόδοση εφόσον εξαρτάται από τα τυχαία δεδομένα που εκπαιδεύθηκε.



## 5.4 Σύγκριση Μεθόδων

Παραπάνω παρουσιάστηκαν τα αποτελέσματα των μεθόδων που ελέγχθηκαν ως σημείο αναφοράς, καθώς και των μεθόδων που αναπτύχθηκαν με multilayer perceptrons. Στη συνέχεια παρουσιάζονται συγκριτικά οι μέθοδοι που λειτούργησαν καλύτερα κατά μέσο όρο σε κάθε προσέγγιση και ακολουθούν μερικές παρατηρήσεις όσον αφορά το πείραμα.

Σφάλμα sMAPE				
	Level 0	Level 1	Level 2	Average
<b>Benchmark Methods</b>	11,75367	11,68944	16,04994	13,16435
<b>One NN per time series</b>	8,77322	11,19566	16,33708	12,10199
<b>One NN(train with all inputs)</b>	9,04889	10,9486	14,97964	11,65904
<b>One NN(train with random inputs)</b>	9,49942	11,38693	15,41719	12,10118

Πίνακας 8 : Σύγκριση καλύτερων μεθόδων κάθε προσέγγισης, ανά επίπεδο και μέσο όρο

Η πρώτη γραμμή αντιστοιχεί στις μεθόδους αναφοράς, η δεύτερη στην προσέγγιση που εκπαιδεύθηκε ένα νευρωνικό για κάθε χρονοσειρά, η τρίτη στην προσέγγιση που εκπαιδεύθηκε ένα νευρωνικό με όλες τις τιμές, ενώ η τέταρτη στο νευρωνικό με την εκπαίδευση από τυχαίες τιμές. Από τον παραπάνω πίνακα καθώς και από όλα τα προηγούμενα προκύπτουν οι εξής παρατηρήσεις.

Η εκπαίδευση ενός νευρωνικού δικτύου για κάθε χρονοσειρά και η συμφιλίωση των προβλέψεων με τις μεθόδους bottom-up και optimal ήταν ιδιαίτερα χρονοβόρα και δεν έδωσε αισθητά καλύτερα αποτελέσματα σε σχέση με την καλύτερη benchmark μέθοδο. Αντίθετα, η εκπαίδευση ενός δικτύου για την Total χρονοσειρά και η εφαρμογή της μεθόδου top-down παρουσίασε καλύτερη απόδοση και ήταν αρκετά γρηγορότερη στην διαδικασία παραγωγής προβλέψεων. Αξίζει να σημειωθεί ότι αυτή η μέθοδος έδωσε τα καλύτερα αποτελέσματα για το επίπεδο  $k=0$  από όλες τις προσεγγίσεις.

Η χρήση ενός ενιαίου νευρωνικού δικτύου εκπαιδευμένου από όλες τις τιμές των χρονοσειρών έδωσε κατά μέσο όρο τα καλύτερα αποτελέσματα από όλες τις προσεγγίσεις. Παρουσίασαν εμφανή βελτίωση οι μέθοδοι bottom-up και optimal κατά μέσο όρο, κυρίως λόγω της μεγάλης βελτίωσης των προβλέψεων στα επίπεδα  $k=1$  και  $k=2$ .

Η εκπαίδευση του νευρωνικού με λιγότερες, αλλά τυχαίες τιμές από κάθε χρονοσειρά είχε επίσης αρκετά καλή απόδοση. Δεν έδωσε τόσο καλά αποτελέσματα

όσο η προηγούμενη προσέγγιση, όμως λειτούργησε πολύ καλύτερα από τις μεθόδους αναφοράς και από την χρήση ενός δικτύου για κάθε χρονοσειρά. Πλεονέκτημα της είναι ότι ήταν πολύ γρήγορη η εκπαίδευση του νευρωνικού δικτύου.

Τέλος είναι φανερό ότι μπορεί να βελτιωθεί η απόδοση και των τριών ιεραρχικών μεθόδων με τη χρήση νευρωνικών δικτύων για την παραγωγή των βασικών προβλέψεων, σε σχέση με τις προβλέψεις από τις κλασικές μεθόδους. Την πιο μεγάλη βελτίωση για τα συγκεκριμένα δεδομένα παρουσίασε η optimal reconciliation, η οποία ενώ στις μεθόδους αναφοράς έδινε τα χειρότερα αποτελέσματα, με τη χρήση νευρωνικών δικτύων έδωσε τις πιο κοντινές προβλέψεις.

## Κεφάλαιο 6 : Αποτελέσματα και Προεκτάσεις

### 6.1 Συνοπτικά Αποτελέσματα της Εργασίας

Στην παρούσα εργασία έγινε πρόβλεψη ζήτησης σε καταστήματα λιανικής με χρήση ιεραρχικών μεθόδων και μεθόδων μηχανικής μάθησης. Σκοπός ήταν να εξετάσουμε κατά πόσο ο συνδυασμός των ιεραρχικών μεθόδων πρόβλεψης με τα πολυεπίπεδα αντίληπτρα μπορεί να βελτιώσει τις προβλέψεις των χρονοσειρών, σε σχέση με τη χρήση των ιεραρχικών μεθόδων με κλασικές μεθόδους πρόβλεψης. Για το λόγο αυτό εξετάστηκαν οι τρεις βασικότερες ιεραρχικές μέθοδοι top-down, bottom-up και optimal για την συμφιλίωση των βασικών προβλέψεων. Ως σημείο αναφοράς χρησιμοποιήθηκαν οι προβλέψεις από τα μοντέλα naive, ets και arima σε συνδυασμό με τις παραπάνω ιεραρχικές μεθόδους.

Τα βασικά συμπεράσματα που προκύπτουν από την παραπάνω διαδικασία είναι :

- Από τις μεθόδους αναφοράς τα καλύτερα αποτελέσματα δίνει η bottom-up όταν οι βασικές προβλέψεις παράγονται από το μοντέλο ets. Αυτό ήταν αναμενόμενο λόγω της φύσης των δεδομένων και του γεγονότος ότι χειριζόμασταν κυρίως τις χρονοσειρές του χαμηλότερου επιπέδου.
- Το δείγμα train παίζει σημαντικό ρόλο και βοηθάει όσο μεγαλώνει. Αυτό γίνεται εμφανές από τις προσεγγίσεις που εκπαιδεύεται ένα νευρωνικό με όλες τις τιμές και ένα με λιγότερες τυχαίες, με την πρώτη περίπτωση να δίνει πιο αξιόπιστες προβλέψεις.
- Ο αριθμός των εισόδων κατά την εκπαίδευση των νευρωνικών επηρεάζει τα αποτελέσματα. Οι 3 εισοδοί δεν λειτουργούν τόσο καλά, ενώ οι 6 και 12 εισοδοί φαίνονται να δίνουν καλύτερα αποτελέσματα. Όταν τα νευρωνικά εκπαιδεύονται με 6 εισόδους λειτουργούν καλύτερα σε συνδυασμό με την top-down, ενώ με 12 εισόδους λειτουργούν καλύτερα με την bottom-up. Η optimal φαίνεται να λειτουργεί καλά και στις δύο περιπτώσεις, όμως επηρεάζεται και από άλλες παραμέτρους.
- Η συνάρτηση ενεργοποίησης υπερβολική εφαιπτομένη λειτουργεί καλύτερα στην προσέγγιση με τα 74 MLP, ενώ η λογιστική σιγμοειδής στις προσεγγίσεις που εκπαιδεύεται μόνο ένα ενιαίο MLP.
- Στο level 0 τα καλύτερα αποτελέσματα δίνει η top-down όταν οι βασικές προβλέψεις παράγονται από ένα νευρωνικό εκπαιδευμένο μόνο από την total χρονοσειρά. Αυτό συμβαίνει γιατί με τη μέθοδο αυτή παράγονται οι καλύτερες προβλέψεις για την total χρονοσειρά και κατ' επέκταση μειώνονται τα σφάλματα στο υψηλότερο επίπεδο της ιεραρχίας.

- Στα επίπεδα 1 και 2 δουλεύει καλύτερα η optimal, όταν οι βασικές προβλέψεις προκύπτουν από ένα ενιαίο νευρωνικό δίκτυο εκπαιδευμένο από όλες τις τιμές όλων των χρονοσειρών.
- Η χρήση των νευρωνικών δικτύων βελτιώνει τις προβλέψεις και για τις τρεις ιεραρχικές μεθόδους. Τη μικρότερη βελτίωση εμφανίζει η bottom-up, ενώ οι άλλες δύο μέθοδοι δίνουν αισθητά καλύτερα αποτελέσματα σε σχέση με τις μεθόδους αναφοράς.
- Τα νευρωνικά δίκτυα φαίνεται να επηρεάζουν περισσότερο τα επίπεδα 0 και 2 και λιγότερο το επίπεδο 1 σε σχέση με την bottom-up(ets) που ήταν η καλύτερη μέθοδος αναφοράς.

Κύριο συμπέρασμα που προκύπτει είναι ότι η χρήση των νευρωνικών δικτύων σε συνδυασμό με τις ιεραρχικές μεθόδους πρόβλεψης παράγει πιο αξιόπιστα αποτελέσματα από την κλασική χρήση των ιεραρχικών μεθόδων, τόσο σε κάθε επίπεδο της ιεραρχίας όσο και κατά μέσο όρο. Τα ποσοστά βελτίωσης μεταβάλλονται ανάλογα με τον τρόπο εκπαίδευσης των νευρωνικών δικτύων καθώς και με τον τρόπο χρήσης τους. Προφανώς μεγάλο ρόλο παίζουν και τα δεδομένα τα οποία χρησιμοποιούνται κάθε φορά, ειδικά σε περιπτώσεις που περιέχουν special events.

## 6.2 Προεκτάσεις

Όπως αναφέραμε παραπάνω σκοπός της μελέτης ήταν ο συνδυασμός της ιεραρχικής συνάθροισης με τις μεθόδους μηχανικής μάθησης για την παραγωγή προβλέψεων. Τα βασικά στοιχεία της μελέτης είναι οι μέθοδοι που χρησιμοποιήθηκαν για τις προβλέψεις, ο τύπος των δεδομένων και οι ορίζοντες πρόβλεψης. Αλλάζοντας κάποιο από τα στοιχεία αυτά μπορεί να προκύψει μια διαφορετική μελέτη με διαφορετικά αποτελέσματα.

Επιλέξαμε συγκεκριμένες ιεραρχικές μεθόδους, μεθόδους πρόβλεψης και τύπο νευρωνικών δικτύων για την παρούσα μελέτη. Χρησιμοποιήθηκαν οι ιεραρχικές μέθοδοι με την περισσότερη βιβλιογραφία, δηλαδή οι top-down, bottom-up και optimal, ενώ ο τύπος νευρωνικού δικτύου που αναπτύχθηκε ήταν τα πολυεπίπεδα αντίληπτρα. Θα μπορούσε να προκύψει μια νέα μελέτη, με λιγότερο γνωστές ιεραρχικές μεθόδους ή με τη χρήση διαφορετικού τύπου νευρωνικού δικτύου ή μεθόδου μηχανικής μάθησης για την παραγωγή των προβλέψεων.

Ακόμα τα δεδομένα που χρησιμοποιήθηκαν στην παρούσα εργασία ήταν ημερήσια και ορίζοντας πρόβλεψης ήταν οι δύο εβδομάδες. Αλλάζοντας τον τύπο των δεδομένων θα μπορούσε να προκύψει μια νέα μελέτη για εβδομαδιαία, μηνιαία ή ετήσια δεδομένα. Αντίστοιχα για διαφορετικό ορίζοντα πρόβλεψης θα επηρεαζόταν διαφορετικά η αποδοτικότητα των μοντέλων.

Τέλος, βασική επέκταση θα αποτελούσε μια ηλεκτρονική εφαρμογή που θα λάμβανε τις ιεραρχικές χρονοσειρές, θα εκπαιδεύε τα αντίστοιχα νευρωνικά δίκτυα και θα παρήγαγε τις ιεραρχικές προβλέψεις για κάθε επίπεδο της ιεραρχίας. Η εφαρμογή θα ήταν ιδιαίτερα σε επιχειρήσεις με μεγάλο αριθμό καταστημάτων, σε διαφορετικές γεωγραφικές περιοχές, όπως ήταν τα δεδομένα που ελέγχθηκαν εδώ, ή σε οποιαδήποτε άλλη ιεραρχική δομή απαιτούσε την παραγωγή προβλέψεων.



## Βιβλιογραφία

- Aoki, M. (1987). *State space modelling of time series*. Berlin: Springer-Verla
- Armstrong, J. S. (Ed.). (2001). *Principles of forecasting: a handbook for researchers and practitioners* (Vol. 30). Springer Science & Business Media.
- Athanasopoulos, G., & Hyndman, R. J. (2008). Modelling and forecasting Australian domestic tourism. *Tourism Management*, 29, 19–31.
- Box, G. E., Jenkins, G. M., Reinsel, G. C., & Ljung, G. M. (2015). *Time series analysis: forecasting and control*. John Wiley & Sons.
- George Athanasopoulos , Roman A. Ahmed, Rob J. Hyndman, "Hierarchical forecasts for Australian domestic tourism" , *International Journal of Forecasting* 25 (2009) 146–166
- Giuseppe Ciaburro, Balaji Venkateswaran, "Neural Networks with R", (2017) Packt Publishing
- G. Peter Zhang, Idea Group Inc., "Neural Networks in Business Forecasting", 2004
- G. Zhang - B.E. Patuwo - M.Y. Hu, "Forecasting with artificial neural networks: The state of art", *International Journal of Forecasting*, 1997
- Hyndman, R.J., & Athanasopoulos, G. (2018) "Forecasting: principles and practice", OTexts
- Hyndman, R., Koehler, A. B., Ord, J. K., & Snyder, R. D. (2008). *Forecasting with exponential smoothing: the state space approach*. Springer Science & Business Media.
- Hyndman, R. J., Koehler, A. B., Snyder, R. D., & Grose, S. (2002)." A state space framework for automatic forecasting using exponential smoothing methods". *International Journal of Forecasting*, 18, 439–454.
- Jacek M. Zurada, West Publishing Company, "Introduction to Artificial Neural Networks", 1992
- Makridakis, S., & Hibon, M. (2000). The M3-competition: Results, conclusions and implications. *International Journal of Forecasting*, 16, 451–476
- Montgomery, D. C., Jennings, C. L., & Kulahci, M. (2015)." Introduction to time series analysis and forecasting". John Wiley & Sons.

Petropoulos, F., Kourentzes, N., 2014. "Improving forecasting via multiple temporal aggregation". The International Journal of Applied Forecasting In press.

Rob J. Hyndman , Roman A. Ahmed, George Athanasopoulos, Han Lin Shang, "Optimal combination forecasts for hierarchical time series" ,Computational Statistics and Data Analysis 55 (2011) 2579–2589

Shumway, R. H., & Stoffer, D. S. (2017). "Time series analysis and its applications: with R examples". Springer.

Slawek Smyl, " A hybrid method of Exponential Smoothing and Recurrent Neural Networks for time series forecasting", International Journal of Forecasting ,2019

## **Βιβλία - Διπλωματικές Εργασίες**

Γκολογιάννης Π.(2013)- Μελέτη ευρωπαϊκών τραπεζών για την πρόβλεψη της συμπεριφοράς τους στη σύγχρονη οικονομική πραγματικότητα με τη χρήση τεχνητών νευρωνικών δικτύων

"Νευρωνικά Δίκτυα και Μηχανική Μάθηση " - 3 η έκδοση, Simon Haykin, Εκδόσεις Παπασωτηρίου, 2009

Πετρόπουλος Φ., Ασημακόπουλος Β., (2011). Επιχειρησιακές Προβλέψεις, εκδόσεις συμμετρία, Αθήνα.

Φακίνος Νικ. Γεώργιος (2016) - Πρόβλεψη Ηλεκτρικής Κατανάλωσης σε Σύμπλεγμα Εμπορικών Κτιρίων με Χρήση Μεθόδων Χρονικής Συνάθροισης Δεδομένων

Φρισήρας Χρήστος Κ. (2018) - Αξιολόγηση Μεθόδων Πρόβλεψης Οριακής Τιμής Συστήματος Ηλεκτρικής Ενέργειας στην Ελληνική Αγορά