

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων και Συστημάτων
Αποφάσεων

Ανάπτυξη Ολοκληρωμένου Διαδικτυακού Συστήματος Υποστήριξης Προβλέψεων

Διπλωματική Εργασία της ΗΛΕΚΤΡΑΣ ΣΚΕΠΕΤΑΡΗ

Επιβλέπων: Βασίλειος Ασημακόπουλος Καθηγητής Ε.Μ.Π.

Μονάδα Προβλέψεων και Στρατηγικής Αθήνα, Ιούλιος 2018

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων
Μονάδα Προβλέψεων και Στρατηγικής

Ανάπτυξη Ολοκληρωμένου Διαδικτυακού Συστήματος Υποστήριξης Προβλέψεων

Διπλωματική Εργασία της ΗΛΕΚΤΡΑΣ ΣΚΕΠΕΤΑΡΗ
Επιβλέπων: Βασίλειος Ασημακόπουλος Καθηγητής Ε.Μ.Π.
Μονάδα Προβλέψεων και Στρατηγικής Αθήνα, Ιούλιος 2018

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την Ιουλίου 2018.

(Υπογραφή)

(Υπογραφή)

(Υπογραφή)

.....

.....

.....

Βασίλειος Ασημακόπουλος

Ιωάννης Ψαρράς

Δημήτριος Ασκούνης

Καθηγητής Ε.Μ.Π.

Καθηγητής Ε.Μ.Π.

Καθηγητής Ε.Μ.Π.

Αθήνα, Ιούλιος 2018
(Υπογραφή)

.....

Ηλέκτρα Σκεπετάρη

Διπλωματούχος Ηλεκτρολόγος Μηχανικός και Μηχανικός Υπολογιστών Ε.Μ.Π.

Εθνικό Μετσόβιο Πολυτεχνείο
Σχολή Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών
Τομέας Ηλεκτρικών Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων
Μονάδα Προβλέψεων και Στρατηγικής

Copyright ©–All rights reserved Ηλέκτρα Σκεπετάρη. Με επιφύλαξη παντός δικαιώματος. 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 τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτή την εργασία εκφράζουν τον συγγραφέα και δεν πρέπει να ερμηνευθεί ότι αντιπροσωπεύουν επίσημες θέσεις του Εθνικού Μετσόβιου Πολυτεχνείου συμπεριλαμβανόμενων Σχολών, Τομέων και Μονάδων αυτού.

Πρόλογος

Η διπλωματική αυτή εργασία εκπονήθηκε στα πλαίσια των ερευνητικών δραστηριοτήτων της Μονάδας Προβλέψεων και Στρατηγικής κατά το ακαδημαϊκό έτος 2017-2018. Η μονάδα υπάγεται στον Τομέα Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων της Σχολής Ηλεκτρολόγων Μηχανικών & Μηχανικών Η/Υ, του Εθνικού Μετσόβιου Πολυτεχνείου. Θα ήθελα να ευχαριστήσω τον επιβλέποντα καθηγητή κ. Βασίλειο Ασημακόπουλο για την ευκαιρία που μου έδωσε να συμμετέχω στην ερευνητική δραστηριότητα του εργαστηρίου καθώς και να εκπονήσω τη παρούσα διπλωματική. Η στήριξη και η καθοδήγηση του όλα αυτά τα χρόνια που εργάζομαι στην Μονάδα ήταν πολύτιμη. Θα ήθελα ακόμα να ευχαριστήσω τον Καθηγητή κ. Ιωάννη Ψαρρά και τον Καθηγητή κ. Δημήτριο Ασκούνη για τη συμμετοχή τους στην επιτροπή εξέτασης της εργασίας. Ακόμα θα ήθελα να ευχαριστήσω θερμά τον Δρ. Βαγγέλη Σπηλιώτη, τόσο για την βοήθεια του στην εκπόνηση αυτής της εργασίας όσο και για την βοήθεια, καθοδήγηση και στήριξη που μου έχει προσφέρει όλα αυτά τα χρόνια στα πλαίσια της συνεργασίας μας στη Μονάδα Προβλέψεων και Στρατηγικής. Για την καθοδήγηση και τις συμβουλές τις τόσο στα πλαίσια του μαθήματος όσο και στα ερευνητικά έργα του εργαστηρίου θα ήθελα, επίσης, να ευχαριστήσω την φίλη και συνεργάτιδα Ζαμπέτα Λεγάκη. Τέλος, θα ήθελα να ευχαριστήσω και τα υπόλοιπα μέλη της Μονάδας Προβλέψεων και Στρατηγικής, Αχιλλέα, Κατερίνα, Χρήστο, Ανδρέα, Νίκη, Φώτη και Φωτεινή για την υποστήριξη τους σε κάθε επίπεδο, καθώς και την οικογένεια και τους φίλους μου.

Ηλέκτρα Σκεπετάρη

Περίληψη

Η παρούσα διπλωματική εργασία στοχεύει στο σχεδιασμό και την υλοποίηση μιας διαδικτυακής πλατφόρμας ανάλυσης δεδομένων σε μορφή χρονοσειρών και παραγωγής προβλέψεων. Τα τελευταία χρόνια η παραγωγή προβλέψεων αποτελεί προσόν υψίστης σημασίας όσον αφορά την οργάνωση κάθε επιτυχημένης εταιρείας. Προκειμένου να διευκολυνθούν οι εταιρείες στην λήψη αποφάσεων, έχουν αναπτυχθεί πολλά συστήματα υποστήριξης προβλέψεων. Παρά το γεγονός ότι αυτά τα συστήματα μπορούν σε γενικές γραμμές να αποδειχθούν αρκετά χρήσιμα, πολλές φορές παρουσιάζουν αρκετά ελαττώματα τόσο από τεχνολογική όσο και από μεθοδολογική άποψη.

Στη παρούσα εργασία αναλύονται αρχικά τα σχεδιαστικά χαρακτηριστικά των προϋπαρχόντων συστημάτων και δίνεται έμφαση στην ελλιπή διαδικτυακή παρουσία τους, στη μη προσαρμογή τους στα ιδιαίτερα χαρακτηριστικά της κάθε εταιρείας και χρήστη, καθώς και στις μη φιλικές τους διεπαφές. Τα χαρακτηριστικά αυτά είναι ιδιαίτερα σημαντικά δεδομένου ότι οι διαδικτυακές εφαρμογές που προσαρμόζονται στις ανάγκες του χρήστη και διαθέτουν φιλικό γραφικό περιβάλλον μπορούν να γίνουν εύκολα αποδεκτές, ακόμα και από χρήστες με ανεπαρκείς τεχνικές γνώσεις. Η μεθοδολογική ανεπάρκεια των προϋπαρχόντων συστημάτων, που συνίσταται στις παρωχημένες μεθόδους πρόβλεψης και στην έλλειψη παροχής δυνατοτήτων για παραγωγή κριτικών προβλέψεων, αποτελεί άλλο ένα σημαντικό αντικείμενο μελέτης.

Εν συνεχεία γίνονται κάποιες προτάσεις σχετικά με τον τρόπο με τον οποίο μπορεί να σχεδιαστεί ένα βελτιωμένο σύστημα υποστήριξης προβλέψεων, αξιοποιώντας αποκλειστικά εργαλεία ανοικτού κώδικα για την αντιμετώπιση των προβλημάτων που αναφέρθηκαν νωρίτερα. Στο πλαίσιο αυτό παρουσιάζεται το OMEN, ένα πλήρως προσαρμόσιμο και φιλικό προς τον χρήστη σύστημα πρόβλεψης, με διαδικτυακή αρχιτεκτονική, η οποία χρησιμοποιεί αποκλειστικά λύσεις ανοικτού κώδικα για την υποστήριξη εταιρειών και επαγγελματιών. Η εργασία παρουσιάζει αναλυτικά τα σενάρια χρήσης του συστήματος για κάθε λειτουργικότητά του, συμπεριλαμβανομένης της εξόρυξης δεδομένων, της επεξεργασίας δεδομένων και της παραγωγής και αναπροσαρμογής προβλέψεων. Η εφαρμογή διατίθεται ελεύθερα προς τον χρήστη στην ηλεκτρονική διεύθυνση omen-project.eu.

Η παρούσα εργασία ολοκληρώνεται με την απαρίθμηση ιδεών που στοχεύουν στην μελλοντική αναβάθμιση των υπηρεσιών του συστήματος και που δεν υλοποιήθηκαν στην παρούσα φάση.

Λέξεις κλειδιά: Διαδικτυακή Εφαρμογή, Λήψη Αποφάσεων, Σύστημα παραγωγής προβλέψεων

Abstract

The aim of this thesis is the design and development of a web-based platform that performs analysis of time-series data and generates forecasts. Nowadays, forecasting is of utmost importance when it comes to the planning of every successful company. In order to help companies in their forecasting and decision-making procedures, many Forecasting Support Systems have been developed through the years. While these systems can be proven quite helpful in general, it seems that they tend to be deficient both in technological and methodological aspect.

In the present thesis we indicate the main design vulnerabilities of typical solutions, that are their limited customizability, non web-based architecture and poor user interface. Those aspects are especially important as customizable, user-friendly applications can be easily accepted even by users with inadequate technical backgrounds. Furthermore, the methodological deficiency of these solutions will be analyzed, which consists of outdated methods and lack of the judgmental component.

Moreover, some suggestions are proposed regarding the way a system can be involved and set open-source solutions as a key parameter for dealing with all the problems mentioned above. In this respect OMEN is presented, a fully customizable web-based forecasting tool which uses exclusively open-source solutions and a modern interface to support companies and practitioners. The features of OMEN are demonstrated through use case scenarios for all of its modules, such as data mining, data pre-processing and forecasting. The application is available online for free at the web address omen-project.eu.

Finally, services which could not be implemented during this work are suggested for future updates of the platform.

Keywords: Web-based application, Decision Support Systems, Forecasting support systems

Περιεχόμενα

Πρόλογος	ii
Περίληψη	iv
Abstract	vi
Περιεχόμενα	viii
Κεφάλαιο 1	1
Ευρεία Περίληψη	1
Ανάγκη Παραγωγής Προβλέψεων	1
Προβλήματα Συστημάτων Παραγωγής Προβλέψεων	1
Σύγχρονες Διαδικτυακές Εφαρμογές	2
Λογισμικό Ανοιχτού Κώδικα	3
Έρευνα Εργαλείων Παραγωγής Προβλέψεων	3
Σχεδιασμός Αρχιτεκτονικής της Πλατφόρμας	3
Υλοποίηση της Πλατφόρμας	4
Δυνατότητες Επέκτασης της Πλατφόρμας	5
Ανάγκη Παραγωγής Προβλέψεων	6
Βήματα Παραγωγής Πρόβλεψης	6
Παράγοντες που οδηγούν στην ανάγκη για προβλέψεις	7
Κατηγορίες Προβλέψεων Βάσει των Γενικών Χαρακτηριστικών τους	8
Κατηγορίες Προβλέψεων Βάσει των Σκοπών τους	9
Κεφάλαιο 2	12
Συστήματα Παραγωγής Προβλέψεων	12
Προαπαιτούμενα για Χρήση FSS	12
Διαθέσιμες πληροφορίες και δεδομένα	13
Σχεδιασμός	13
Γνώση και Εκπαίδευση	14
Ιδανική έναντι πραγματικής χρήσης FSS	14
Σχεδιασμός FSS για Βελτίωση της Αποτελεσματικότητας	15
Σύγχρονες Διαδικτυακές Εφαρμογές	17
Πλεονεκτήματα Διαδικτυακών Εφαρμογών	17
Λογισμικό Ανοιχτού Κώδικα	19
Έρευνα Εργαλείων Παραγωγής Προβλέψεων	21
ForecastPRO	21
SAS® Forecast Studio	22
SAS® University Edition	23

Κεφάλαιο 3	26
Σχεδιασμός Αρχιτεκτονικής της Πλατφόρμας	26
Γενικός Σχεδιασμός	26
R Statistical Software	27
MySQL	28
CSS-HTML-JavaScript	29
Πακέτα R	30
RShiny	30
Shinydashboard	30
Σενάρια Χρήσης της Πλατφόρμας	31
Εγγραφή Νέου Χρήστη	32
Σύνδεση Χρήστη	32
Επιλογή Χρονοσειράς για Πρόβλεψη	33
Στατιστική Ανάλυση Δεδομένων	34
Συμπλήρωση Κενών Τιμών	35
Απαλοιφή Ακραίων Τιμών	35
Αποσύνθεση Χρονοσειράς	36
Εξομάλυνση Χρονοσειράς	36
Χρονική Συνάθροιση Χρονοσειράς	37
Παραγωγή Στατιστικής Πρόβλεψης	37
Παραγωγή Κριτικής Πρόβλεψης	38
Δείκτες Σφάλματος και Αξιολόγηση Πρόβλεψης	38
Εξαγωγή Αναφοράς	39
Κεφάλαιο 4	41
Ενσωματωμένες Μέθοδοι και Συναρτήσεις	41
Ανάλυση χρονοσειρών	41
Συμπλήρωση Κενών Τιμών	41
Last non-zero value	41
Linear regression	41
Multivariate imputations (trend & period)	42
Bootstrap EM Algorithm	42
Απαλοιφή Ακραίων Τιμών	43
Μέθοδος Chen & Liu	43
Μέθοδος Προσαρμογής Μοντέλου Πρόβλεψης & Εντοπισμού Ασυνήθιστων Τιμών	43
Μέθοδος Εξέτασης της Κατανομής των Σφαλμάτων Πρόβλεψης	44
Μέθοδος Εντοπισμού Ακραίων Τιμών	44

Αποσύνθεση	45
Κλασσική Μέθοδος Αποσύνθεσης	45
Εξομάλυνση	48
Κινητός Μέσος Όρος	48
Μη Γραμμική Εξομάλυνση	48
Τεχνικές προβλέψεων	49
Μέθοδοι Συνεχούς Ζήτησης	49
Naive	49
SES	49
ETS	50
Holt	52
Damped	52
Theta Classic	53
ARIMA	53
MAPA	55
Μέθοδοι Διακοπτόμενης Ζήτησης	58
Croston	58
SBA	58
TSB	59
iMAPA	59
Αξιολόγηση	60
Μέσο σφάλμα (Mean Error - ME)	60
Μέσο απόλυτο σφάλμα (Mean Absolute Error - MAE)	60
Μέσο τετραγωνικό σφάλμα (Mean Squared Error - MSE)	60
Ρίζα μέσου τετραγωνικού σφάλματος(Root Mean Squared Error - RMSE)	60
Μέσο απόλυτο ποσοστιαίο σφάλμα (Mean Absolute Percentage Error - MAPE)	61
Συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα (Symmetric Mean Absolute Percentage Error - sMAPE)	61
Κεφάλαιο 5	63
Υλοποίηση της πλατφόρμας	63
Αρχική Σελίδα	64
Επιλογή ή Μεταφόρτωση Χρονοσειρών	65
Στατιστική Ανάλυση	67
Συμπλήρωση Κενών Τιμών	69
Απαλοιφή Ακραίων Τιμών	70
Αποσύνθεση	72

Εξομάλυνση	74
Χρονική Συνάθροιση	75
Πρόβλεψη	76
Κριτικές Προσαρμογές	77
Αξιολόγηση Προβλέψεων	78
Εξαγωγή Αναφοράς	79
Κεφάλαιο 6	88
Αξιολόγηση Συστήματος	88
Σύνοψη	97
Μελλοντικές Προεκτάσεις	99
Βιβλιογραφία	101

Κεφάλαιο 1

Ευρεία Περίληψη

Ανάγκη Παραγωγής Προβλέψεων

Η παραγωγή προβλέψεων είναι απαραίτητη για την οργάνωση και το σχεδιασμό κάθε επιτυχημένης επιχείρησης ή οργανισμού. Η ανάγκη για πρόβλεψη πηγάζει από την ανάγκη για προγραμματισμό και σχεδιασμό στρατηγικών με γνώμονα τις μελλοντικές εξελίξεις.

Σε κάθε οργανισμό μπορούν να πραγματοποιηθούν διάφορων ειδών προβλέψεις όπως για παράδειγμα προβλέψεις πωλήσεων, παραγωγικών πόρων, προϋπολογισμού, και γενικά οικονομικών δεικτών ή αποθεμάτων.

Προκειμένου να παραχθούν ακριβείς προβλέψεις που θα είναι χρήσιμες για την εκάστοτε επιχείρηση θα πρέπει πρώτα να γίνει σωστή συλλογή και οργάνωση των ιστορικών στοιχείων. Στη συνέχεια, τα δεδομένα αυτά θα πρέπει να αξιοποιηθούν με τον καλύτερο δυνατό τρόπο, συνήθως μέσω κάποιου συστήματος παραγωγής προβλέψεων.

Για την διευκόλυνση των στελεχών των επιχειρήσεων, έχουν αναπτυχθεί διάφορα συστήματα με στόχο την καλύτερη διαχείριση του μεγάλου όγκου δεδομένων και την αυτοματοποίηση της διαδικασίας της πρόβλεψης με το μικρότερο δυνατό σφάλμα, στο λιγότερο δυνατό χρόνο.

Προβλήματα Συστημάτων Παραγωγής Προβλέψεων

Προκειμένου να παράγουν οι επιχειρήσεις ακριβείς προβλέψεις στον λιγότερο δυνατό χρόνο μέσω συστημάτων παραγωγής προβλέψεων είναι απαραίτητο, οι άνθρωποι να μπορούν να διαχειριστούν την πλατφόρμα με τον πιο αποτελεσματικό τρόπο.

Ωστόσο, παρά την βοήθεια που προσφέρουν τα συστήματα τέτοιας φύσης, τα περισσότερα είναι ελλιπή όσον αφορά την διαδικτυακή αρχιτεκτονική, την δυνατότητα προσαρμογής στις ανάγκες του χρήστη και την φιλική προς τον χρήστη διεπαφή.

Στην σημερινή εποχή, που οι επιχειρήσεις έχουν μεγάλο όγκο δεδομένων και πολλές φορές οι εργαζόμενοι που διαχειρίζονται τις εφαρμογές τέτοιου τύπου δεν έχουν επαρκές τεχνικό υπόβαθρο, μειονεκτήματα σαν και τα παραπάνω αποκτούν μεγάλη σημασία.

Σύγχρονες Διαδικτυακές Εφαρμογές

Με την πάροδο του χρόνου όλο και περισσότερες διαδικτυακές εφαρμογές κάνουν την εμφάνιση τους. Συγκρίνοντας τις παραδοσιακές με τις διαδικτυακές εφαρμογές οι χρήστες μπορούν εύκολα να εντοπίσουν τα χαρακτηριστικά που αιτιολογούν την άνοδο της δημοτικότητας τους.

Αρχικά, τέτοιου είδους εφαρμογές προσφέρουν δυνατότητα άμεσης πρόσβασης από οποιαδήποτε συσκευή δηλαδή δεν περιορίζονται στον υπολογιστή αλλά επεκτείνονται σε οποιαδήποτε συσκευή έχει σύνδεση στο διαδίκτυο χωρίς την εγκατάσταση κάποιου επιπρόσθετου λογισμικού. Η μόνη απαραίτητη εφαρμογή είναι ο περιηγητής διαδικτύου ο οποίος είναι προεγκατεστημένος σε όλα τα λειτουργικά συστήματα ακόμα και σε φορητές συσκευές όπως τα σύγχρονα κινητά τηλέφωνα. Η ιδιότητα αυτή των διαδικτυακών εφαρμογών είναι ιδιαίτερα σημαντική για μεγάλες επιχειρήσεις με πολλούς χρήστες που στην περίπτωση της τοπικής εφαρμογής θα έπρεπε να εγκατασταθεί η εφαρμογή σε κάθε ένα υπολογιστή ξεχωριστά. Όπως προκύπτει από το παραπάνω χαρακτηριστικό, δίνεται εύκολα η δυνατότητα ακόμα και εκτός του χώρου εργασίας του κάθε χρήστη. Η δυνατότητα αυτή δίνει ευελιξία στους χρήστες ώστε να χρησιμοποιούν τις εφαρμογές οπουδήποτε αυτοί επιθυμούν επιτρέποντας τους ακόμα και να εργάζονται από απομακρυσμένες περιοχές ή και από το σπίτι τους, χαρακτηριστικό ιδιαίτερα χρήσιμο σε περίπτωση κάποιου εκτάκτου προβλήματος της επιχείρησης.

Σημαντικό πλεονέκτημα συγκριτικά με τις τοπικές εφαρμογές εμφανίζεται στις περιπτώσεις που η εφαρμογή χρειάζεται κάποια αναβάθμιση. Σε μια κλασική τοπική εφαρμογή η αναβάθμιση του συστήματος θα πρέπει να γίνει σε κάθε ένα υπολογιστή ξεχωριστά από τον χρήστη. Αντίθετα σε μια διαδικτυακή εφαρμογή η αναβάθμιση πραγματοποιείται μόνο στον εξυπηρετητή που φιλοξενεί την εφαρμογή και το αναβαθμισμένο πρόγραμμα είναι διαθέσιμο σε όλους τους χρήστες άμεσα. Με τον τρόπο αυτό εξοικονομείται χρόνος ο οποίος είναι ιδιαίτερα πολύτιμος κυρίως για τις μεγάλες επιχειρήσεις ενώ ως συνέπεια του παραπάνω σημαντικά μειωμένο είναι και το κόστος της αναβάθμισης μιας και απαιτείται λιγότερο εργατικό δυναμικό για την διεκπεραίωση της αναβάθμισης.

Ένα ακόμα πλεονέκτημα των διαδικτυακών εφαρμογών είναι ότι πλέον με την εμφάνιση εργαλείων διαδικτυακής ανάπτυξης εφαρμογών όπως για παράδειγμα της [HTML5](#) είναι δυνατό ο δημιουργός της εφαρμογής να την εμπλουτίσει έτσι ώστε να είναι πιο φιλική, εύχρηστη και ευχάριστη προς τον χρήστη. Παλαιότερα οι εφαρμογές αυτές υστερούσαν στην εμφάνιση ωστόσο πλέον είναι ιδιαίτερα εύκολο να προσαρμοστούν στα σύγχρονα δεδομένα και ανάγκες του χρήστη.

Ένα ακόμα πλεονέκτημα των διαδικτυακών εφαρμογών είναι ότι είναι συμβατές με όλα τα λειτουργικά συστήματα, πράγμα που είναι λογικό καθότι η εφαρμογή δεν εγκαθίσταται ποτέ στην συσκευή του χρήστη. Η ιδιότητα αυτή οφείλεται επίσης και στην προτυποποίηση των γλωσσών προγραμματισμού τις οποίες χρησιμοποιεί η εφαρμογή. Ως συνέχεια του παραπάνω και εφόσον οι διαδικτυακές εφαρμογές δεν εκτελούνται στον υπολογιστή του χρήστη δεν καταναλώνουν και πόρους από το σύστημα. Για τον λόγο αυτό, αλλά και επειδή δεν καταλαμβάνουν καθόλου ή σχεδόν καθόλου χώρο στον δίσκο του χρήστη για την εγκατάσταση, παρά μόνο κατά την χρήση της εφαρμογής (και μόνο στην περίπτωση που ο χρήστης το επιθυμεί) οι εφαρμογές διαδικτύου είναι ιδιαίτερα ελαφριές για την υπολογιστική μονάδα.

Επιπρόσθετα, παρότι τέτοιου είδους εφαρμογές δεν εγκαθίστανται ποτέ στην συσκευή του χρήστη, οι σύγχρονοι περιηγητές προσφέρουν τη δυνατότητα της εκτός διαδικτύου χρήσης με την προϋπόθεση

ότι η εφαρμογή έχει κατασκευαστεί με ανάλογο τρόπο. Για παράδειγμα αν για κάποιο λόγο η σύνδεση στο διαδίκτυο διακοπεί αυτό δεν επηρεάζει τον χρήστη, ο οποίος συνεχίζει να χρησιμοποιεί την εφαρμογή κανονικά. Αυτό επιτυγχάνεται μέσω του περιηγητή ο οποίος κρατάει ένα αντίγραφο από τα αρχεία, τα οποία είναι απαραίτητα για την εκτός δικτύου χρήση της εφαρμογής, στον υπολογιστή του χρήστη, και τα χρησιμοποιεί όταν αυτό κριθεί απαραίτητο. Η συγκεκριμένη δυνατότητα δεν είναι διαθέσιμη σε όλες τις εφαρμογές που χρησιμοποιούν [HTML5](#) αλλά μόνο σε αυτές που έχει υπάρξει πρόβλεψη για χρήση της εφαρμογής και εκτός διαδικτύου ή ενδοδικτύου.

Λογισμικό Ανοιχτού Κώδικα

Στα πλαίσια της εργασίας θα πρέπει να γίνει αναφορά και στο λογισμικό ανοικτού κώδικα (Open Source). Ως Open Source χαρακτηρίζεται το μοντέλο διάθεσης λογισμικού, όπου ο πηγαίος κώδικας είναι διαθέσιμος σε όποιον ενδιαφέρεται για αυτόν. Οι όροι διανομής του λογισμικού αυτού πρέπει να πληρούν συγκεκριμένα κριτήρια, τα βασικότερα από τα οποία είναι: (α) η ελεύθερη χρήση, (β) η αντιγραφή/αναδιανομή και μεταβολή/βελτίωσή του. Από τα παραπάνω ενδέχεται να δημιουργηθεί η απορία για το ποια είναι η διαφορά μεταξύ Ελεύθερου Λογισμικού και Λογισμικού Ανοιχτού Κώδικα. Καταρχήν πρέπει να διευκρινιστεί ότι οι δύο όροι αναφέρονται σε παρόμοιες “ιδεολογίες”. Τόσο η μια “ιδεολογία” όσο και η άλλη, προσπαθούν να προωθήσουν τη δημιουργία καλύτερου λογισμικού και γενικά να προωθήσουν την “κοινή γνώση”. Η κύρια διαφορά τους είναι ότι ενώ το Ελεύθερο Λογισμικό δίνει έμφαση στην ελευθερία (για μάθηση, προσαρμογή, βελτίωση), το Λογισμικό Ανοιχτού Κώδικα δίνει έμφαση στη δημιουργία καλύτερου λογισμικού μέσω της συνεργασίας των προγραμματιστών και των μηχανικών όλου του κόσμου. Συνεπώς, το Λογισμικό Ανοιχτού Κώδικα είναι λιγότερο αυστηρό και πιο “φιλικό” προς τις εταιρίες που επιθυμούν να αξιοποιήσουν αλγόριθμους που υπάρχουν σε έργα τύπου Λογισμικού Ανοιχτού Κώδικα.

Έρευνα Εργαλείων Παραγωγής Προβλέψεων

Στα πλαίσια της παρούσας διπλωματικής, πραγματοποιήθηκε αναζήτηση παρόμοιων εργαλείων με θέμα την παραγωγή προβλέψεων. Οι εφαρμογές που βρέθηκαν προσφέρονται από μεγάλες, εδραιωμένες εταιρίες έναντι ετήσιων, συνήθως αρκετά ακριβών, συνδρομών. Δεν βρέθηκε κάποια διαθέσιμη εφαρμογή στο διαδίκτυο, η οποία να απευθύνεται σε επιχειρήσεις και να είναι ανεπτυγμένη στα πλαίσια ερευνητικής μελέτης.

Ένα ακόμη χαρακτηριστικό των διαθέσιμων εργαλείων, πέρα από την υψηλή τιμολόγηση αφορά την δομή και οργάνωση των εφαρμογών οι οποίες διαθέτουν μεγάλο αριθμό προσφερόμενων επιλογών και εισόδων, αυξάνοντας επομένως την πολυπλοκότητα χρήσης, ιδιαίτερα από απλούς χρήστες, χωρίς αυξημένες προϋπάρχουσες γνώσεις τόσο στον τομέα των προβλέψεων όσο και συγκεκριμένα στην χρήση πληροφοριακών συστημάτων.

Σχεδιασμός Αρχιτεκτονικής της Πλατφόρμας

Ο σχεδιασμός της αρχιτεκτονικής είναι ένα πολύ σημαντικό στάδιο για την υλοποίηση της εφαρμογής, αφού εάν είναι επιτυχημένος και ακριβής μπορεί να αποτρέψει πολλά προβλήματα, όπως σφάλματα λειτουργίας αλλά και γενικά κόστη. Είναι, επομένως, σημαντικό, κατά τον σχεδιασμό να υλοποιηθεί μια οργανωμένη, κατανοητή δομή που θα είναι εύκολα συντηρήσιμη.

Για να επιτευχθεί το παραπάνω, υλοποιήσαμε αρχικά ένα πολύ βασικό σύστημα με τις κύριες μόνο λειτουργίες και βασισμένοι σε αυτό επαναλάβαμε το σχεδιασμό αρκετές φορές για να καλύψουμε κενά και να βελτιώσουμε το πρωτότυπο. Η παραπάνω μεθοδολογία υπακούει στην μέθοδο σχεδιασμού LUCID (Logic User-Centered Interactive Design). Σύμφωνα με τα παραπάνω υλοποιήθηκε αρχικά η εφαρμογή ForWarD, που περιείχε μόνο τις δυνατότητες παραγωγής προβλέψεων και αξιολόγησης τους, και βασισμένοι στις ελλείψεις της, υλοποιήσαμε την εφαρμογή Omen.

Η διαδικασία που ακολουθεί ο χρήστης για να παράξει προβλέψεις έχει ως εξής:

- Σύνδεση με τον λογαριασμό του, δημιουργία νέου λογαριασμού ή συνέχεια ως επισκέπτης
- Επιλογή δεδομένων ή φόρτωση χρονοσειράς από .csv αρχείο
- Επεξεργασία δεδομένων
- Στατιστική ανάλυση δεδομένων
- Συμπλήρωση κενών τιμών
- Απαλοιφή ακραίων τιμών
- Απο-εποχικοποίηση χρονοσειράς
- Εξομάλυνση χρονοσειράς
- Πολλαπλή χρονική συνάθροιση χρονοσειράς
- Πρόβλεψη συνεχούς χρονοσειράς ή χρονοσειράς με κενές τιμές
- Παραγωγή κριτικής πρόβλεψης
- Αξιολόγηση και σφάλματα πρόβλεψης
- Παραγωγή αναφοράς διαδικασίας πρόβλεψης

Με χρήση της γλώσσας UML έγινε η σχεδίαση των λειτουργιών σε διαγράμματα περιπτώσεων χρήσης. Επιπλέον, παρέχονται mockups της εφαρμογής για την οπτικοποίηση της διεπαφής.

Υλοποίηση της Πλατφόρμας

Στη παρούσα εργασία θα προσπαθήσουμε να αντιμετωπίσουμε τα προβλήματα των μέχρι τώρα συστημάτων παραγωγής προβλέψεων και να σχεδιάσουμε ένα νέο, βελτιωμένο σύστημα που θα είναι απλό στη χρήση και θα παράγει ακριβείς προβλέψεις.

Αρχικά, το σύστημα θα είναι διαδικτυακό που σημαίνει ότι θα είναι διαθέσιμο προς χρήση κάθε στιγμή και από κάθε είδους συσκευή. Ακόμη, το σύστημα θα είναι προσαρμοσσιμο στις εκάστοτε ανάγκες κάθε χρήστη και επιχείρησης, έτσι ώστε να είναι πιο εύκολο στην χρήση και θα έχει φιλικό προς τον χρήστη γραφικό περιβάλλον, με απλές επιλογές και λογική δομή.

Στα πλαίσια αυτά θα χρησιμοποιηθεί η γλώσσα προγραμματισμού R, που αποτελεί ένα ισχυρό περιβάλλον για την ανάλυση αριθμητικών δεδομένων. Ενώ υπάρχουν πολλές άλλες εφαρμογές και εργαλεία που επιτρέπουν την ανάπτυξη εφαρμογών, η R επιτρέπει επίσης τον εύκολο υπολογισμό περίπλοκων στατιστικών συναρτήσεων, παρέχοντας παράλληλα ένα ισχυρό περιβάλλον για την εκτέλεση αριθμητικών προσομοιώσεων και δυνατότητες γραφικών παραστάσεων για την καλύτερη κατανόηση των αποτελεσμάτων. Η R αποτελεί ελεύθερη εφαρμογή λογισμικού της γλώσσας S (<http://www.r-project.org>) και αναπτύχθηκε για πρώτη φορά από τους R. Gentleman και R. Ihaka (U του Auckland, NZ) κατά τη διάρκεια της δεκαετίας του 1990. Από τότε, μέχρι σήμερα εξελίσσεται

συνεχώς και αποτελεί ένα από τα πιο προηγμένα συστήματα στατιστικής πληροφορικής, ελεύθερα διαθέσιμο για τις περισσότερες πλατφόρμες υπολογιστών. Ενημερώνεται συνεχώς από μέλη της ακαδημαϊκής κοινότητας έτσι ώστε να περιέχει τις state-of-the-art μεθόδους ανά πάσα στιγμή.

Η R λειτουργεί με βιβλιοθήκες, που ονομάζονται πακέτα. Στα πλαίσια της εφαρμογής μας, χρησιμοποιήθηκε σαν βασικό πακέτο το R Shiny, που δίνει την δυνατότητα ανάπτυξης εφαρμογών μέσω της R.

Όσον αφορά την αποθήκευση δεδομένων, χρησιμοποιήσαμε MySQL, ένα σύστημα διαχείρισης σχεσιακών βάσεων ανοικτού κώδικα (relational database management system - RDBMS), που χρησιμοποιεί την Structured Query Language (SQL), την πιο γνωστή γλώσσα για την προσθήκη, την πρόσβαση και την επεξεργασία δεδομένων σε μία Βάση Δεδομένων.

Ο λόγος που προτιμήθηκε είναι γιατί είναι ανοικτού κώδικα (open source), οποιοσδήποτε μπορεί να κατεβάσει τη MySQL και να την διαμορφώσει με βάση τις ανάγκες του, σύμφωνα πάντα με την γενική άδεια χρήσης και μέσω του πακέτου `mysql` παρέχεται η δυνατότητα διαχείρισης της μέσω R. Η MySQL είναι γνωστή κυρίως για την ταχύτητα, την αξιοπιστία, και την ευελιξία που παρέχει. Οι περισσότεροι συμφωνούν ωστόσο ότι δουλεύει καλύτερα όταν διαχειρίζεται περιεχόμενο και όχι όταν εκτελεί συναλλαγές. Η MySQL αυτή τη στιγμή μπορεί να λειτουργήσει σε περιβάλλον Linux, Unix, και Windows.

Δυνατότητες Επέκτασης της Πλατφόρμας

Παρά το γεγονός ότι η εφαρμογή περιλαμβάνει όλα τα παραπάνω στοιχεία που επιτρέπουν την παραγωγή ακριβών στατιστικών προβλέψεων, ένα μέρος που θα μπορούσε να ενσωματωθεί και να την κάνει πιο ολοκληρωμένη είναι η προσθήκη δυνατότητας για παραγωγή κριτικών προβλέψεων. Στη παρούσα φάση της πλατφόρμας, δίνεται μόνο η επιλογή προσθήκης κριτικής πρόβλεψης μέσω απλής κρίσης, ενώ θα ήταν χρήσιμο να ενσωματωθούν και άλλες μέθοδοι όπως η Delphi ή οι αναλογίες.

Από τεχνικής απόψεως, θα ήταν χρήσιμη η μεταφορά της εφαρμογής σε cloud-based υποδομή, αφού με τον τρόπο αυτό θα μειωθεί σημαντικά ο χρόνος εκτέλεσης των πράξεων και επομένως θα αυξηθεί η χρηστικότητα της εφαρμογής και η ικανοποίηση των χρηστών.

Ανάγκη Παραγωγής Προβλέψεων

Η σημασία της πρόβλεψης ως επιχειρηματικού εργαλείου είναι μεγάλη. Αρχικά, η πρόβλεψη είναι το πρώτο βήμα στη διαδικασία του σχεδιασμού. Μπορεί να επιτρέψει στην επιχείρηση ή τον οργανισμό να πάρει αποφάσεις για μελλοντικές δραστηριότητες που σχετίζονται με νέα προϊόντα στην αγορά παρά την αβεβαιότητα στις παγκόσμιες αγορές. Ο κίνδυνος κατά τη διαμόρφωση μιας πρόβλεψης λόγω των πιθανών επιπτώσεων περιβαλλοντικών, διαφημιστικών, ανταγωνιστικών, ή άλλων παραγόντων συνεχίζει ωστόσο να υπάρχει και να είναι εκτός του ελέγχου της επιχείρησης. Έτσι, η πρόβλεψη είναι μια εκτίμηση που χρησιμοποιεί στις περισσότερες περιπτώσεις τη μαθηματική μοντελοποίηση και την πορεία του παρελθόντος, προκειμένου να δώσει συμπεράσματα για το μέλλον.

Βήματα Παραγωγής Πρόβλεψης

Σε αυτό το σημείο θα αναφερθούν τα βασικότερα βήματα σε κάθε διαδικασία παραγωγής προβλέψεων, τα οποία διατηρήθηκαν και χρησιμοποιήθηκαν στο σχεδιασμό της εφαρμογής που αποτελεί το αντικείμενο της εν λόγω διπλωματικής για τον σχεδιασμό της πλατφόρμας:

Βήμα 1. Ορισμός του προβλήματος: Αρχικά θα πρέπει να γίνει κατανόηση της κατάστασης των δεδομένων και του λόγου που θα πρέπει να γίνει η πρόβλεψη.

Βήμα 2. Συλλογή Πληροφοριών. Η συλλογή των ιστορικών δεδομένων αλλά και των ιδιαίτερων χαρακτηριστικών τους είναι πολύ σημαντική διότι αποτελούν την βάση στην οποία θα βασιστεί και θα προσαρμοστεί η αντίστοιχη μέθοδος πρόβλεψης.

Βήμα 3. Προκαταρκτική - Διερευνητική Ανάλυση. Ο σκοπός αυτού του βήματος είναι η εύρεση του προτύπου που ακολουθούν τα δεδομένα και γενικά η εξαγωγή σημαντικών πληροφοριών από τα διαθέσιμα ιστορικά δεδομένα. Συνήθης τρόπος επίτευξης των παραπάνω είναι εν γένει η οπτικοποίηση των δεδομένων μέσω της γραφικής τους απεικόνισης, ο υπολογισμός στατιστικών δεικτών και αναλύσεων. Αυτό το βήμα υποδεικνύει συχνά ποιες μέθοδοι πρόβλεψης ενδείκνυνται για χρήση.

Βήμα 4. Επιλογή και Προσαρμογή του Μοντέλου. Αυτό το βήμα περιλαμβάνει την επιλογή και την προσαρμογή του αντίστοιχου ποσοτικού μοντέλου στα εκάστοτε δεδομένα. Υπάρχουν πολλοί παράγοντες που πρέπει να ληφθούν υπόψη. Συχνά είναι απαραίτητες διάφορες προσομοιώσεις.

Βήμα 5. Χρήση και Αξιολόγηση του Μοντέλου Πρόβλεψης. Μετά τη συνετή επιλογή του κατάλληλου ανά περίπτωση μοντέλου πρόβλεψης και τη σωστή εκτίμηση των παραμέτρων του, το μοντέλο καλείται να παράγει τις προβλέψεις και οι οποίες θα αξιολογηθούν. Η διαδικασία της πρόβλεψης δεν είναι ολοκληρωμένη όταν το μοντέλο απλά προσαρμοστεί στα δεδομένα αλλά μόνο όταν οι παραχθείσες προβλέψεις αξιολογηθούν, δηλαδή γίνουν γνωστές οι αντίστοιχες πραγματικές τιμές. Πρέπει να τονιστεί όμως, ότι η ακρίβεια στις μελλοντικές προβλέψεις δεν είναι το μοναδικό κριτήριο που πρέπει να λαμβάνεται ως κατευθυντήριος για την αξιολόγηση των μεθόδων πρόβλεψης. Όπως είναι λογικό, μια πρόβλεψη πωλήσεων είναι ένα ανεκτίμητο εργαλείο για κάθε επιχείρηση, επειδή προβλέπει μελλοντικά γεγονότα. Για παράδειγμα μια πρόβλεψη πωλήσεων μπορεί να δείξει το

επίπεδο πωλήσεων που θα πρέπει να αναμένει μια επιχείρηση να πετύχει από μήνα σε μήνα ή ακόμα και από έτος σε έτος, αλλά μπορεί επίσης να προσφέρει πληθώρα πρόσθετων στοιχείων - όπως να δείξει δυνατότητες ανάπτυξης σε συγκεκριμένα τμήματα ή να αποκαλύψει νέες ευκαιρίες στην αγορά και να διασαφηνίσει τις πιθανότητες για ανάπτυξη εάν βελτιωθούν ορισμένα χαρακτηριστικά και λειτουργίες των προϊόντων ή υπηρεσιών που προσφέρει.

Παράγοντες που οδηγούν στην ανάγκη για προβλέψεις

Μια πρόβλεψη μπορεί να διαδραματίσει σημαντικό ρόλο στην επιτυχία ή την αποτυχία της επιχείρησης. Σε γενικό επίπεδο, μπορούμε να ισχυριστούμε ότι μια ακριβής πρόβλεψη βελτιστοποιεί την επιχειρηματική λειτουργία σε ότι αφορά την παραγωγή, τους ανθρώπινους πόρους και την διαχείριση των οικονομικών στοιχείων μίας επιχείρησης. Συμβάλλει στη μείωση της αβεβαιότητας και στην πρόβλεψη τυχόν αλλαγών στην αγορά καθώς και στη βελτίωση της επικοινωνίας στα πλαίσια μιας επιχείρησης ή και μεταξύ μιας επιχείρησης και των πελατών της. Βοηθά, επίσης, στην αύξηση της γνώσης της αγοράς για τις επιχειρήσεις. Επιπλέον, μια ελπιδοφόρα πρόβλεψη μπορεί να αποτελέσει κίνητρο για τους επενδυτές που μπορεί να ενδιαφέρονται επενδύσουν σε μια επιχείρηση.

Επομένως, για να λειτουργεί μια επιχείρηση αποτελεσματικά, χρειάζεται να έχει μία ιδέα για το πώς θα μοιάζει το μέλλον. Κάθε κλάδος και ομάδα μιας επιχείρησης επωφελείται από τις προβλέψεις. Οι πωλητές και οι έμποροι επωφελούνται από τις προβλέψεις, επειδή οι αριθμοί πρόβλεψης πωλήσεων επηρεάζουν τον τρόπο διαχείρισης των πωλήσεων και τον τρόπο με τον οποίο εμπλέκονται οι πελάτες, καθώς και τις δαπάνες μάρκετινγκ. Δεδομένου ότι οι προβλέψεις εκτιμούν έναν αναμενόμενο όγκο πωλήσεων για μια συγκεκριμένη χρονική περίοδο, οι πωλητές μπορούν να τα χρησιμοποιήσουν για να ορίσουν τους στόχους δραστηριότητας τους και μπορούν να γίνουν μεταγενέστερες προσαρμογές για την επίτευξη των στόχων πωλήσεων. Οι έμποροι μπορούν να χρησιμοποιήσουν τις προβλέψεις για να μετρήσουν την αποτελεσματικότητα των εκστρατειών τους, να αποφασίσουν ποιες αγορές θα εισέλθουν και να εξέλθουν και να καθορίσουν τον κύκλο ζωής των προϊόντων τους.

Τα ανώτερα διευθυντικά στελέχη και οι οικονομικοί φορείς χρησιμοποιούν τις προβλέψεις για την προετοιμασία και την αξιολόγηση των οικονομικών σχεδίων, την αξιοποίηση της παραγωγής και την αξιολόγηση των αναγκών και της εφοδιαστικής. Μια πρόβλεψη μπορεί να βοηθήσει στην λήψη κρίσιμων αποφάσεων σχετικά με τον τρόπο κατανομής πόρων όπως το προσωπικό, το ενοίκιο, τα βοηθητικά προγράμματα και άλλα γενικά έξοδα σε μια επιχείρηση.

Δεδομένου ότι οι προβλέψεις έχουν σκοπό να εξετάσουν το μέλλον, πρέπει να γίνουν ορισμένες παραδοχές που αποτελούν τη βάση της πρόβλεψης. Όσο πιο ακριβείς είναι οι υποθέσεις, τόσο πιο ακριβείς είναι οι προβλέψεις, αλλά οι μεταβαλλόμενες συνθήκες μπορούν να επηρεάσουν δραματικά την ακρίβεια της πρόβλεψης. Οι επιχειρήσεις πρέπει να λάβουν υπόψη τα εξής:

- Αλλαγές στην αγορά . Πόσο θα μεγαλώσει η αγορά γενικά; Ποιες άλλες δυνάμεις ανταγωνισμού ενδέχεται να επηρεάσουν το μερίδιο αγοράς;
- Συμπεριφορά πελατών . Οι επιχειρήσεις χρειάζονται μια ιδέα για τον αριθμό των πελατών που κερδίζουν και χάνουν κάθε χρόνο, αν υπάρχουν ορισμένες περιόδους μεγαλύτερης διακύμανσης των πωλήσεων (π.χ. εποχικότητα) και επίγνωση του μέσου αριθμού και διακύμανσης των πωλήσεων.

- Πόροι . Οι επιχειρήσεις πρέπει επίσης να γνωρίζουν πόσο μπορεί να χρειαστεί να αναπτυχθούν και τους περιορισμούς τους, καθώς και τι ποσοστό των πόρων τους μπορούν να ξοδέψουν στη διαφήμιση καθώς και τι μπορεί να αυξήσει ή να μειώσει τον αριθμό των πωλήσεων.

Κατηγορίες Προβλέψεων Βάσει των Γενικών Χαρακτηριστικών τους

Στη συνέχεια, οι επιχειρήσεις πρέπει να καθορίσουν ποιο είδος πρόβλεψης είναι το καταλληλότερο. Υπάρχουν τρία γενικά είδη:

- Προβλέψεις ποιοτικού χαρακτήρα
- Προβλέψεις ποσοτικού χαρακτήρα
- Αιτιοκρατικά μοντέλα πρόβλεψης

Οι *ποιοτικές* προσεγγίσεις χρησιμοποιούνται γενικά όταν τα δεδομένα δεν είναι άμεσα διαθέσιμα - σε περιπτώσεις όπου μια επιχείρηση, ένα προϊόν ή μια υπηρεσία είναι καινούργια ή όταν υπάρχει εξωτερική πληροφορία που δεν μπορεί να συμπεριληφθεί σε ένα ποσοτικό μοντέλο πρόβλεψης . Συνήθως αυτή η τεχνική χρησιμοποιεί εκτιμήσεις εμπειρογνομόνων ή άλλα μέσα που πρέπει να είναι λογικά, συστηματικά και αμερόληπτα στις εκτιμήσεις τους, οι οποίες στη συνέχεια ποσοτικοποιούνται. Όπως υποδηλώνει το όνομα, δεν είναι τόσο αυστηρές γενικά όσο οι ποσοτικές προβλέψεις.

Οι ποσοτικές μέθοδοι βασίζονται σε ιστορικά δεδομένα ή δεδομένα "χρονοσειράς", επομένως χρησιμοποιούνται συχνότερα όταν το προϊόν ή η υπηρεσία είναι σταθερό και διαθέσιμο για κάποιο χρονικό διάστημα. Νέες επιχειρήσεις ή επιχειρήσεις με νέα προϊόντα ενδέχεται να μην μπορούν να χρησιμοποιήσουν αυτήν τη μέθοδο. Η πρόβλεψη γίνεται με βάση την αναγνώριση προτύπων, τάσεων και αλλαγών στα δεδομένα.

Η αιτιακή μοντελοποίηση , το τελευταίο από τα τρία εργαλεία πρόβλεψης, λαμβάνει υπόψη όλα όσα επηρεάζουν τις πωλήσεις, ακόμη και χρησιμοποιώντας κάποια ανάλυση χρονοσειρών και περιορίζει τον αριθμό των υποθέσεων στην πρόβλεψη. Εάν γίνουν παραδοχές, παρακολουθούνται καθ 'όλη τη διάρκεια της μοντελοποίησης για να διασφαλιστεί ότι είναι έγκυρες και το μοντέλο βελτιώνεται συνεχώς όταν διατίθενται περισσότερες πληροφορίες για την επιχείρηση.

Κατηγορίες Προβλέψεων Βάσει των Σκοπών τους

Πέρα από την παραπάνω κατηγοριοποίηση των προβλέψεων, μπορεί να γίνει κατηγοριοποίηση σύμφωνα με το στόχο τους. Η διαδικασία της πρόβλεψης είναι κάτι περισσότερο από ένα περίγραμμα των στρατηγικών. Οι προβλέψεις αποτελούν την καρδιά του σχεδίου μάρκετινγκ και μπορεί να έχουν τους παρακάτω στόχους:

- Προώθηση της ανάπτυξης της βιομηχανίας. Το σχέδιο μάρκετινγκ περιλαμβάνει τη παραγωγή προβλέψεων για τον ρυθμό ανάπτυξης της βιομηχανίας στην οποία δραστηριοποιείται η εταιρεία. Αυτή η πληροφορία είναι κρίσιμη, διότι οι στρατηγικές που ένας ιδιοκτήτης επιχείρησης θα εφαρμόσει σε περιόδους γρήγορης ανάπτυξης της βιομηχανίας είναι διαφορετικές από αυτές που θα χρησιμοποιούσε σε μια πιο αργή αγορά. Εάν ο κλάδος αναμένεται να αναπτυχθεί γρήγορα, το σχέδιο μάρκετινγκ θα επικεντρωθεί στην απόκτηση όσο το δυνατόν περισσότερων νέων πελατών. Εάν η ανάπτυξη πρόκειται να παραμείνει σταθερή, η εστίαση θα πρέπει να είναι περισσότερο στη διατήρηση των πελατών που έχει ήδη η επιχείρηση - επειδή οι ανταγωνιστές της θα προσπαθήσουν να προσελκύσουν τους πελάτες της μακριά.
- Εκτίμηση του μεγέθους της αγοράς. Ο ιδιοκτήτης μιας μικρής επιχείρησης πρέπει να έχει μια λογική εκτίμηση του μεγέθους της αγοράς της εταιρείας του. Το μέγεθος της αγοράς που πρέπει να εκτιμηθεί δεν αφορά το συνολικό αριθμό των πελατών που αγοράζουν αγαθά ή υπηρεσίες προσφερόμενα από την εταιρία, αλλά το μέγεθος της απευθυνόμενης αγοράς του - εκείνων των ομάδων πελατών που μπορεί να φτάσει με μια επιτυχημένη καμπάνια μάρκετινγκ και οι οποίες είναι πιθανότερο να αγοράσουν από την εταιρεία. Ένας ιδιοκτήτης μικρής επιχείρησης μερικές φορές είναι εύκολο να υπερεκτιμά το πραγματικό μέγεθος της αγοράς του, γεγονός που μπορεί να οδηγήσει σε απογοητευτικά αποτελέσματα των πωλήσεων, επειδή οι πελάτες δεν έφταναν το επίπεδο που είχε εκτιμηθεί ότι φτάνουν.
- Προσδιορισμός αναγκών σε πόρους. Τα έσοδα προκύπτουν έπειτα από ενέργειες της ομάδας διαχείρισης και του προσωπικού της επιχείρησης, όπως είναι οι κλήσεις προώθησης πωλήσεων, οι διαφημιστικές εκστρατείες και η κοινωνική δικτύωση. Αυτές οι ενέργειες απαιτούν τόσο χρήματα όσο και προσωπικό για την εφαρμογή τους. Ο ιδιοκτήτης της επιχείρησης, ιδιαίτερα στην περίπτωση των μικρομεσαίων επιχειρήσεων, πρέπει να προβλέπει με ακρίβεια τις ανάγκες αυτές, καθώς και το κόστος τους για να εξασφαλίσει ότι τόσο τα κεφάλαια όσο και το προσωπικό είναι διαθέσιμα για την εφαρμογή των στρατηγικών στο σχέδιο μάρκετινγκ. Οι μικρές επιχειρήσεις πρέπει να χρησιμοποιούν κάθε πόρο για μάρκετινγκ με σύνεση και να διαθέτουν χρόνο και πόρους σε αυτές τις στρατηγικές έτσι ώστε να αυξάνουν τις πιθανότητες να δημιουργήσουν πωλήσεις.
- Ανάπτυξη μετρήσιμων μεγεθών για την αξιολόγηση της πορείας της επιχείρησης. Ένας ιδιοκτήτης μιας επιχείρησης πρέπει να προβλέπει περισσότερα από τα έσοδα και το κόστος. Πρέπει, για παράδειγμα, να παρακολουθεί την παραγωγικότητα του προσωπικού και την επιτυχία των στρατηγικών μάρκετινγκ, δημιουργώντας μεγέθη μέσω των οποίων θα μπορεί να ελέγχει τους παραπάνω παράγοντες κατά τη διάρκεια του έτους. Μια σημαντική περίπτωση μεγέθους που πρέπει να παρακολουθείται είναι ο ρυθμός με τον οποίο η εταιρία αποκτά πελάτες, ο οποίος μπορεί να αφορά πόσες από τις κλήσεις προωθητικού χαρακτήρα έγιναν και πόσες από αυτές είχαν ως αποτέλεσμα πραγματικές πωλήσεις. Μια χαμηλότερη

τιμή αυτού του μεγέθους, από αυτήν που η επιχείρηση είχε προβλέψει μπορεί να σημαίνει ότι πρέπει να παρέχει πρόσθετη κατάρτιση στο προσωπικό πωλήσεων του ή ακόμα και ότι το προϊόν που παρουσιάζει στους υποψήφιους πελάτες θα πρέπει να βελτιωθεί.

- Προβλέψεις cash flows. Η πρόβλεψη ροών εσόδων και η πρόβλεψη ροών δαπανών μάρκετινγκ, παίζουν μεγάλο ρόλο στη συνολική πρόβλεψη ταμειακών ροών της εταιρείας, οι οποίες περιλαμβάνουν και άλλα λειτουργικά έξοδα, όπως το διοικητικό κόστος προσωπικού, τα έξοδα γραφείου και τυχόν πληρωμές τόκων που υποχρεούται να καταβάλει η εταιρεία. Η μη ρεαλιστική πρόβλεψη εσόδων - ή η σοβαρή υποτίμηση του κόστους μάρκετινγκ που απαιτείται για την επίτευξη των πωλήσεων - μπορεί να οδηγήσει σε έλλειμμα μετρητών για την εταιρεία. Για μια μικρή επιχείρηση, το έλλειμμα μετρητών μπορεί να σημαίνει ότι κάποιες δαπάνες θα πρέπει να περικοπούν, πράγμα που απαιτεί δύσκολες επιλογές, όπως περικοπές προσωπικού.

Από τα παραπάνω, είναι φανερό το πόσο αναγκαίες είναι οι προβλέψεις για μια επιχείρηση, ανεξαρτήτως του μεγέθους της, ή άλλων χαρακτηριστικών της στοιχείων.

Κεφάλαιο 2

Συστήματα Παραγωγής Προβλέψεων

Οι προβλέψεις έχουν βασικό ρόλο στη διαχείριση της αλυσίδας εφοδιασμού στο πλαίσιο μιας επιχείρησης ή οργανισμού. Στις περισσότερες περιπτώσεις οι προβλέψεις αυτές προκύπτουν μέσω κάποιου πληροφοριακού συστήματος το οποίο βοηθά στη λήψη αποφάσεων για λειτουργίες όπως ο προγραμματισμός πόρων, η παραγωγή και οι πωλήσεις. Τα συστήματα αυτού του είδους έχουν, λοιπόν, ορισμένα κοινά χαρακτηριστικά, τα οποία θα αναλυθούν περαιτέρω στο παρόν κεφάλαιο, ορισμένα εκ των οποίων θα μπορούσαν να βελτιωθούν, έτσι ώστε να βελτιστοποιηθεί η διαδικασία λήψης αποφάσεων.

Τα συστήματα παραγωγής προβλέψεων αποτελούν ένα κομμάτι μιας ευρύτερης κατηγορίας πληροφοριακών συστημάτων, γνωστών ως Συστήματα Υποστήριξης Αποφάσεων (Decision Support Systems). Παρά το γεγονός ότι αποτελούν μια υποκατηγορία συστημάτων, το 2001 η αξία τέτοιου είδους πληροφοριακών συστημάτων ήταν περίπου 250 εκατομμύρια δολάρια στις ΗΠΑ και το Ηνωμένο Βασίλειο. Ωστόσο, όσο απαραίτητα και αν είναι τα συστήματα, δεν θα πρέπει να απουσιάζει ο ανθρώπινος παράγοντας τόσο στην επίβλεψη της διαδικασίας όσο και στην τελική κρίση και λήψη της απόφασης. Τα στατιστικά μοντέλα είναι έμπειρα στην εύρεση μοτίβων, ιδιαίτερα μέσα σε μεγάλους όγκους δεδομένων, που μπορεί να προκαλούν σύγχυση στον άνθρωπο, ενώ οι manager μπορούν να λάβουν μέρος υπόψη τα αποτελέσματα ειδικών γεγονότων όπως οι διαφημιστικές καμπάνιες. Τέτοιοι παράγοντες μπορεί να αντιμετωπιστούν λανθασμένα ως θόρυβος από στατιστικά μοντέλα, επειδή οι επιπτώσεών δεν προκύπτουν από ιστορικά δεδομένα.

Τα βασικά χαρακτηριστικά που έχει κάθε Σύστημα Παραγωγής Προβλέψεων είναι τα εξής:

- Η βάση δεδομένων που περιλαμβάνει ιστορικά στοιχεία, μαζί με συναφή δεδομένα όπως οι τιμές και το ιστορικό συμβάντων όπως η προώθηση πωλήσεων
- Οι τεχνικές προβλέψεων, που μπορεί να αποτελούν στατιστικές μεθόδους όπως η εκθετική εξομάλυνση, ή μεθόδους τεχνητής νοημοσύνης όπως νευρωνικά δίκτυα
- Η πλατφόρμα μέσω της οποίας ο χρήστης μπορεί να εφαρμόσει τις τεχνικές προβλέψεων στα δεδομένα της επιχείρησης στην οποία ανήκει.

Το τελικό χαρακτηριστικό που έχουν αυτά τα συστήματα είναι ότι διαθέτουν στον χρήστη την δυνατότητα της επαναλαμβανόμενης ανά τακτά χρονικά διαστήματα, παραγωγής προβλέψεων επιτρέποντας την σύγκριση μεγεθών και την πιο αποτελεσματική εκτίμηση σφαλμάτων.

Προαπαιτούμενα για Χρήση FSS

Αυτή η ενότητα περιγράφει μερικά βασικά στοιχεία, που είναι απαραίτητα προκειμένου να έχουμε ένα καλά σχεδιασμένο FSS. Τα αναφερόμενα χαρακτηριστικά πηγάζουν από τον σχεδιασμό γενικότερων πληροφοριακών συστημάτων αλλά και ειδικών Συστημάτων Υποστήριξης Αποφάσεων.

Διαθέσιμες πληροφορίες και δεδομένα

Τα δεδομένα που θα είναι σχετικά με την πρόβλεψη προέρχονται από διάφορα είδη πηγών επομένως ένα FSS χρειάζεται να παρέχει εύκολη πρόσβαση σε αυτές τις πηγές και τα δεδομένα που περιέχουν. Επίσης, τα ίδια τα δεδομένα μπορούν να έχουν διάφορες μορφές τις οποίες το σύστημα θα πρέπει να ξεχωρίζει, έτσι ώστε ο χρήστης να μπορεί να τα χειριστεί με τον πιο αποτελεσματικό τρόπο.

Τα είδη δεδομένων που μπορεί να έχει μία επιχείρηση ή οργανισμός μπορεί να είναι:

- δεδομένα χρονοσειρών σε διάφορα επίπεδα συνάθροισης (π.χ. παλαιότερες πωλήσεις κατά ομάδα προϊόντων, μέγεθος συσκευασίας, χώρα, πελάτη, τρίμηνο, μήνα ή εβδομάδα). Συχνά χρειάζονται οι χρονοσειρές να καθαριστούν για να αφαιρέσουν τα αποτελέσματα του παρελθόντος, όπως π.χ. ειδικά γεγονότα που διαφορετικά θα μπορούσαν να αλλοιώσουν στατιστικές προεκτάσεις. Αυτός ο καθαρισμός δεδομένων μπορεί να ξεχωρίσει χαρακτηριστικά που κυμαίνονται από λευκό θόρυβο έως εποχιακά μοτίβα, ενώ είναι επίσης πιθανό να αποκαλυφθεί κάποια αλλαγή επιπέδου στα δεδομένα.
- πληροφορίες σχετικά με τις δραστηριότητες που έχουν να κάνουν με τους πελάτες όπως προωθήσεις τιμών και αποσύρσεις προϊόντων.
- πληροφορίες σχετικά με άλλες σχετικές μεταβλητές, π.χ. τις καιρικές προβλέψεις, το χρονοδιάγραμμα των πωλήσεων ή προωθητικών κινήσεων των ανταγωνιστών.
- προβλέψεις που έγιναν σε προηγούμενες περιόδους. Οι εταιρείες χρησιμοποιούν συνήθως μια κυλιόμενη πρόβλεψη όπου οι προηγούμενες προβλέψεις ενημερώνονται ως προσεγγίσεις περιόδου πρόβλεψης.
- προβλέψεις όπως, στατιστικές προβλέψεις, κριτικές προβλέψεις των manager βάσει των επαφών τους με τους πελάτες, προβλέψεις των επιπτώσεων των μειώσεων των τιμών που προέρχονται από οικονομετρικά μοντέλα κ.λ.π.
- πληροφορίες σχετικά με τα σφάλματα που σχετίζονται με τις παρελθοντικές προβλέψεις που μπορούν να παρέχουν ανατροφοδότηση στο μέλλον.

Σχεδιασμός

Η διαδικασία παραγωγής προβλέψεων, που αφορά το στάδιο μετά την συγκέντρωση και προεπεξεργασία των δεδομένων μπορεί να χωριστεί σε δύο μέρη. Το πρώτο αφορά την παραγωγή των στατιστικών (βάσει μοντέλων) προβλέψεων. Εξαρτάται άμεσα από το FSS, και γίνεται συνήθως αυτόματα (με αλγορίθμους που χρησιμοποιούνται για την εκτίμηση της βέλτιστης μεθόδου και αντίστοιχων παραμέτρων) ή ανάλογα με την επιλογή που εισάγει ο χρήστης σχετικά με τη μέθοδο πρόβλεψης και, ενδεχομένως, τις τιμές των παραμέτρων. Το δεύτερο στάδιο περιλαμβάνει προσαρμογή του στατιστικού αποτελέσματος από τον άνθρωπο για να ληφθούν υπόψη ειδικοί παράγοντες και άλλες διαθέσιμες πληροφορίες.

Σε πολλές εταιρείες μεγάλος αριθμός των προβλέψεων πρέπει να πραγματοποιούνται σε τακτά χρονικά διαστήματα, (π.χ. σε περίπτωση που οι προβλέψεις για πάνω από 10.000 προϊόντα είναι κοινές). Οι προβλέψεις συχνά χρειάζονται εβδομαδιαία, και μερικές φορές ακόμα και καθημερινά άρα το σύστημα υποστήριξης τους θα πρέπει να εξασφαλίζει ότι ένα μεγάλο μέρος της διαδικασίας έχει αυτοματοποιηθεί και είναι διαθέσιμο να τρέξει ανά πάσα στιγμή, με μικρή ή ακόμη και καθόλου ανθρώπινη παρέμβαση.

Γνώση και Εκπαίδευση

Πολλοί από τους ανθρώπους που είναι υπεύθυνοι για τον χειρισμό των FSS σήμερα, έχουν λάβει ελάχιστη εκπαίδευση σχετική με μεθόδους στατιστικής πρόβλεψης, αν και μπορεί να έχουν πολυετή εμπειρία παραγωγής προβλέψεων. Σε μερικούς οργανισμούς τα FSS χρησιμοποιούνται ακόμα και σε συναντήσεις διευθυντών για την αναθεώρηση των τακτικών τους. Αυτό σημαίνει ότι το FSS πρέπει να είναι σε θέση να παρουσιάζει την παρούσα παραγωγή και να προβάλλει τα αποτελέσματα εύκολα σε μορφή κατανοητή για όλους τους χρήστες, όπως για παράδειγμα μορφή πινάκων και διαγραμμάτων.

Ιδανική έναντι πραγματικής χρήσης FSS

Μπορούμε να χαρακτηρίσουμε μια χρονοσειρά ως μία σειρά δεδομένων που αποτελείται από τα εξής βασικά συστατικά:

1. Κανονικά πρότυπα ή σχέσεις (π.χ. τάσεις, εποχικότητα, σταθερές σχέσεις μεταξύ δαπανών διαφημίσεων και πωλήσεις) ·
2. Θόρυβος - ο οποίος είναι απρόβλεπτος.

Όπου υπάρχουν επαρκή στοιχεία για την εφαρμογή της μεθόδου που εντοπίζει τα ποιοτικά χαρακτηριστικά, ιδανικά, το σύστημα θα πρέπει να εξηγεί πλήρως τον εντοπισμό τους έτσι ώστε ο χρήστης να μη προσπαθήσει να αλλάξει την στατιστική πρόβλεψη, προσαρμόζοντας την κριτικά με αρνητικό αποτέλεσμα. Η κριτική πρόβλεψη δεν είναι τόσο ακριβής όσο η στατιστική πρόβλεψη τις περισσότερες φορές. Αυτό το γεγονός προέρχεται από τις διάφορες προκαταλήψεις ή συγχύσεις ανθρώπινων πληροφοριών που έχουν εντοπιστεί σε ευρεία κλίμακα στα πλαίσια επιχειρήσεων. Προκύπτουν από το γεγονός ότι κάθε χρήστης έχει το δικό του τρόπο σκέψης και προσπαθεί να εξηγήσει διάφορα φαινόμενα, σύμφωνα με τη προσωπική του εμπειρία, πολλές φορές λανθασμένα. Ομοίως, ένα στατιστικό μοντέλο δεν είναι σε θέση να προβλέψει με ακρίβεια το αντικανονικό στοιχείο, αφού εξ ορισμού περιλαμβάνει μόνο εκείνες τις περιπτώσεις όπου υπάρχουν ανεπαρκή δεδομένα μέσω των οποίων υπολογίζει την μελλοντική πορεία διαφόρων μεγεθών. Υπάρχουν πολλές περιπτώσεις που ακόμη και ένας συνδυασμός στατιστικής και κριτικής μεθόδου μπορούν να οδηγήσουν σε ακόμα πιο ακριβείς προβλέψεις. Για παράδειγμα, σύμφωνα με τους Blattberg & Hoch και Lawrence διαπιστώθηκε πως ο μέσος όρος των κριτικών και στατιστικών προβλέψεων ήταν ακριβέστερος από την κάθε μέθοδο ξεχωριστά. Ομοίως, όταν υπάρχουν κριτικές προσαρμογές πάνω στις στατιστικές προβλέψεις, υπάρχουν ενδείξεις ότι η ακρίβεια θα βελτιωθεί σε περίπτωση που ο forecaster έχει κάποια παραπάνω ένδειξη ή γνώση που δεν μπόρεσε να συμπεριληφθεί στο στατιστικό μοντέλο.

Ωστόσο, υπάρχουν επίσης πολλές ενδείξεις τα σημερινά συστήματα παραγωγής και υποστήριξης προβλέψεων στην πράξη απέχουν πολύ από το ιδανικό. Όταν πρόκειται για τη προσθήκη του κριτικού παράγοντα στις στατιστικές προβλέψεις που παρέχονται αυτόματα από το σύστημα, αντιμετωπίζουν διάφορα προβλήματα. Οι άνθρωποι προσπαθούν να προσθέσουν κριτική πρόβλεψη,

ακόμα και όταν αυτό είναι περιττό. Αυτό σημαίνει ότι κάνουν περιττές και επιζήμιες προσαρμογές στις στατιστικές προβλέψεις και στη συνέχεια είναι υπερβολικά σίγουροι για την ακρίβεια των διορθωμένων προβλέψεών τους. Συχνά αυτές οι προσαρμογές δεν αποθηκεύονται ως εξωτερικές προσθήκες στο μοντέλο, πράγμα το οποίο μπορεί να προκαλέσει προβλήματα. Για παράδειγμα, περίπτωση που δύο άτομα είναι υπεύθυνα για την προσαρμογή, μπορεί ο δεύτερος να πάρει το ήδη διορθωμένο μοντέλο που περιλαμβάνει το στατιστικό αποτέλεσμα και τους επιμέρους υπολογισμούς του συστήματος (π.χ. εκτιμήσεις επιπέδου, τάσης ή εποχιακής παράγοντες) μαζί με τη διόρθωση του πρώτου υπεύθυνου. Έτσι, μπορεί τελικά ενώ ο πρώτος έχει προσαρμόσει το αποτέλεσμα, αυξάνοντας το λίγο επειδή π.χ. τον προηγούμενο μήνα είχε υπάρξει μια επιτυχημένη διαφημιστική καμπάνια της εταιρείας, ο δεύτερος να επιχειρήσει να το ξανά αυξήσει, δίνοντας ένα υπερβολικά αισιόδοξο αποτέλεσμα.

Επίσης, στις κριτικές προθήκες που προβλέπουν ειδικά γεγονότα, όπως οι πωλήσεις έπειτα από προωθητικές ενέργειες, οι άνθρωποι αγνοούν τις ποσοτικές προβλέψεις ακόμα και όταν προσφέρουν μια ακριβή πρόβλεψη για το κανονικό πρότυπο χρονοσειρών.

Ένα ακόμα σενάριο αποτυχημένης χρήσης του συστήματος αφορά την περίπτωση των χρηστών που όταν τους δίνεται η δυνατότητα να επιλέξουν στατιστικό μοντέλο για την παραγωγή των προβλέψεων που χρειάζονται, συνήθως επιλέγουν λάθος. Συγκεκριμένα, προτιμούν τις προεπιλεγμένες παραμέτρους ή μεθόδους και στη συνέχεια προσπαθούν να αντισταθμίσουν αυτό το φαινόμενο με μεγάλη κριτική προσαρμογή στη στατιστική πρόβλεψη, πράγμα που φανερά αποτελεί λανθασμένη χρήση του συστήματος.

Αν η ιδανική ισορροπία μεταξύ των στατιστικών προβλέψεων του συστήματος και των ανθρωπίνων εισροών μπορούσαν να επιτευχθούν, είναι αναμενόμενο ότι θα είχαμε καλύτερα και ακριβέστερα αποτελέσματα. Αυτό προκύπτει από το γεγονός ότι οι στατιστικές μέθοδοι είναι καλύτερες στην απομόνωση της τυχαιότητας πράγμα που σημαίνει ότι είναι λιγότερο πιθανό να εντοπίσουν ψευδή μοτίβα. Ελλείπει ειδικών και υποθέτοντας ότι υπάρχουν επαρκή παρελθοντικά δεδομένα, είναι ικανές να προβλέψουν αξιόπιστα τα επίπεδα και τη διακύμανση του μεγέθους. Ο ρόλος του αναλυτή θα είναι να ενσωματώσει στο αποτέλεσμα αυτό, παράγοντες που δεν έχουν ληφθεί υπ' όψιν, με την προϋπόθεση ότι προσέχει και δεν υπερεκτιμά τις εκτιμήσεις του.

Σχεδιασμός FSS για Βελτίωση της Αποτελεσματικότητας

Οι λόγοι που προαναφέρθηκαν οδηγούν σε δύο βασικές συνιστώσες που πρέπει να ληφθούν υπ' όψιν για τον σχεδιασμό ενός πιο αποτελεσματικού Συστήματος Υποστήριξης Προβλέψεων:

- Βελτίωση των προβλέψεων ή ακόμα και προσθήκη της ικανότητας να αντιλαμβάνεται πότε είναι απαραίτητη η ανθρώπινη παρέμβαση και
- Δυνατότητα στον χρήστη, ακόμη και εάν δεν έχει μεγάλη σχετική εκπαίδευση, να αντιλαμβάνεται όλα τα βήματα που οδήγησαν στη πρόβλεψη

Μέχρι σήμερα τα περισσότερα συστήματα τέτοιου είδους ήταν Windows εφαρμογές, συνήθως εγκατεστημένες τοπικά. Ένα βήμα, λοιπόν, προς τη βελτίωση της αποτελεσματικότητας είναι η αντικατάσταση αυτών των παρωχημένων εφαρμογών με εφαρμογές που βασίζονται στο διαδίκτυο. Αυτή η αντικατάσταση θα προσφέρει πολλά πλεονεκτήματα, όπως η προσβασιμότητα και η διαθεσιμότητα 24/7, η συμβατότητα μεταξύ των πλατφορμών και των συσκευών και θα δώσει τέλος στην ανάγκη για εγκατάσταση πρόσθετου λογισμικού ή των ενημερώσεων. Η προσαρμοστικότητα

είναι επίσης ένα χαρακτηριστικό γνώρισμα των διαδικτυακών λύσεων, που αποτελεί ζήτημα μεγάλης σημασίας δεδομένου ότι κάθε εταιρεία έχει τις δικές της ανάγκες πρόβλεψη. Τα FSS πρέπει να είναι σε θέση να προσαρμόζονται ανάλογα με τις απαιτήσεις, τα προϊόντα, τα συστήματα και τη φύση της εταιρείας που εξετάζουν με διαφορετικές αρχιτεκτονικές, διεπαφές, μοντέλα πρόβλεψης κλπ. Αυτό είναι πολύ πιο εύκολο να γίνει μέσω εφαρμογών που βασίζονται στο διαδίκτυο, δεδομένης της ευελιξίας τους και της δυνατότητας προσαρμογής οποιασδήποτε αλλαγής από τη πλευρά του διακομιστή.

Επιπλέον, τα υπάρχοντα συστήματα FSS δεν παρέχουν στον χρήστη τις πιο σύγχρονες μεθόδους τόσο όσο αφορά τη συλλογή δεδομένων, όσο και την προεπεξεργασία, τη μοντελοποίηση της πρόβλεψης, την επιλογή μοντέλου και την αξιολόγηση απόδοσης τη στιγμή που γίνονται διαθέσιμες στο κοινό. Αυτό οφείλεται κυρίως στη δυσκολία των προγραμματιστών να εφαρμόσουν ή να ενημερώσουν τις μεθόδους υψηλής πολυπλοκότητας με τα συστήματά τους που, ακόμη και στις λίγες περιπτώσεις που γίνεται, είναι μια διαδικασία μεγάλης χρονικής διάρκειας. Ως αποτέλεσμα, οι δυνατότητες πρόβλεψης των FSS μένουν πίσω σε σύγκριση με τα ακαδημαϊκά δεδομένα, γεγονός παράλογο αφού η έρευνα έχει ως σκοπό τη βελτίωση και την υποστήριξη διαδικασιών σε πραγματικές εφαρμογές. Για να επιτευχθεί η προσαρμογή και να μειωθεί ο χρόνος που απαιτείται είναι απαραίτητη η χρήση του ανοικτού πηγαίου κώδικα. Για παράδειγμα, το στατιστικό λογισμικό R είναι δημοσίως διαθέσιμο και έχει πολλά πλεονεκτήματα. Αρχικά, η ενθουσιώδης και εκθετικά αναπτυσσόμενη κοινότητα χρηστών προσφέρει δωρεάν όχι μόνο έτοιμες μεθόδους και συναρτήσεις τελευταίας τεχνολογίας αλλά και ανεκτίμητες συμβουλές. Επί του παρόντος, υπάρχουν περίπου 180 πακέτα που σχετίζονται με χρονοσειρές, συμπεριλαμβανομένων μεταξύ άλλων πακέτα για πρόβλεψη, μοντελοποίηση, αποσύνθεση και μοντέλα δυναμικής παλινδρόμησης. Οι συναρτήσεις είναι καλά τεκμηριωμένες και έχουν πολύ καλές οδηγίες χρήσης, γεγονός που αυξάνει την αποδοχή των χρηστών. Η Python είναι επίσης μια εξαιρετική λύση καθώς η χρήση της έχει αυξηθεί με την πάροδο του χρόνου και περιλαμβάνει μεθόδους σχεδόν για οποιαδήποτε στατιστική λειτουργία, ειδικά μετά την εισαγωγή των `pandas` για την υποστήριξη δομών δεδομένων.

Σύγχρονες Διαδικτυακές Εφαρμογές

Στο πλαίσιο της εξέλιξης στον τομέα της τεχνολογίας και του διαδικτύου, καθώς και του ηλεκτρονικού εμπορίου στο διαδίκτυο, οι διαδικτυακές εφαρμογές κέρδισαν δημοτικότητα συγκριτικά με τις τοπικές εφαρμογές. Η μετάβαση αυτή έγινε σταδιακά, και γίνεται αντιληπτή με πολλούς τρόπους.

Αρχικά, οι τοπικές εφαρμογές αποτελούν λογισμικό που εγκαθίσταται σε έναν μόνο υπολογιστή ή συσκευή. Αυτό το εγκατεστημένο λογισμικό χρησιμοποιείται για την εκτέλεση ορισμένων λειτουργιών και εργασιών, όπως επεξεργασία κειμένου και αναπαραγωγή πολυμέσων κλπ. Σε δικτυακό περιβάλλον, οι εφαρμογές που υπάρχουν μπορούν να χρησιμοποιούνται ταυτόχρονα από πολλαπλούς χρήστες. Όπως είναι λογικό, με την εμφάνιση της κοινωνικής δικτύωσης και του ηλεκτρονικού εμπορίου και μάρκετινγκ, οι εφαρμογές που βασίζονται στο διαδίκτυο έχουν κερδίσει δημοτικότητα σε σχέση με την εφαρμογή γραφείου. Ορισμένα βασικά πλεονεκτήματα και αποτελεσματικά χαρακτηριστικά των εφαρμογών που βασίζονται στο διαδίκτυο που τους καθιστούν ανώτερες από εφαρμογές που βασίζονται σε desktop έχουν ως εξής.

Πλεονεκτήματα Διαδικτυακών Εφαρμογών

- Συντήρηση. Η συντήρηση είναι ένα από τα πιο αποτελεσματικά, ελκυστικά και συναρπαστικά χαρακτηριστικά μιας διαδικτυακής εφαρμογής. Απλά αρκεί η web-based εφαρμογή να εγκατασταθεί μία φορά και στη συνέχεια ο χρήστης μπορεί να απολαμβάνει τα οφέλη της για μεγάλο χρονικό διάστημα, σε αντίθεση με τις παλαιότερες, windows-based εφαρμογές που πρέπει να εγκατασταθούν χωριστά σε κάθε υπολογιστή.
- Ευκολία των ενημερώσεων. Η ενημέρωση των web εφαρμογών είναι πολύ εύκολη. Χρειάζεται απλά να γίνει εγκατάσταση των ενημερώσεων στον κεντρικό υπολογιστή-client και κάθε χρήστης λαμβάνει την ενημερωμένη έκδοση της εφαρμογής, σε αντίθεση με τις windows-based εφαρμογές που πρέπει να ενημερώνονται ξεχωριστά σε κάθε υπολογιστή.
- Ευκολία στη χρήση. Οι εφαρμογές που βασίζονται τοπική εγκατάσταση έχουν περιορισμούς ευχρηστίας, ενώ οι εφαρμογές που βασίζονται στον ιστό είναι πιο βολικές για τον χρήστη, αφού του δίνει τη δυνατότητα να έχει πρόσβαση σε οποιαδήποτε εφαρμογή από οποιαδήποτε τοποθεσία.
- Δεν απαιτούνται δικαιώματα διαχειριστή. Δεν απαιτούνται δικαιώματα διαχειριστή για την εγκατάσταση και την ενημέρωση εφαρμογών μέσω διαδικτύου.
- Διατίθενται ανά πάσα στιγμή και οπουδήποτε. Κάθε χρήστης μπορεί εύκολα να έχει πρόσβαση στα δεδομένα και τις εξελίξεις σε πραγματικό χρόνο με τη δυνατότητα του διαδικτύου και με τη βοήθεια εφαρμογών μέσω διαδικτύου.
- Λιγότεροι περιορισμοί εκτός του τείχους προστασίας. Συνήθως, οι εφαρμογές που βασίζονται στο διαδίκτυο χρησιμοποιούν δύο θύρες HTTP και HTTPS και δεν απαιτούν πρόσθετη θύρα για να ανοίξουν ένα τείχος προστασίας, ενώ οι εφαρμογές γραφείου δεν χρησιμοποιούν αυτές τις θύρες και απαιτούν πρόσθετες θύρες για πρόσβαση στο εξωτερικό ενός τείχους προστασίας.
- Πολλαπλές πλατφόρμες. Ένα από τα πιο αποτελεσματικά, ελκυστικά και σημαντικά χαρακτηριστικά των web-based εφαρμογών είναι ότι είναι δυνατόν να χρησιμοποιηθούν από πολλαπλές πλατφόρμες εύκολα. Γενικά, μια εφαρμογή που βασίζεται στον ιστό λειτουργεί σε

διαφορετικά προγράμματα περιήγησης ιστού όπως Opera, Mozilla Firefox και Internet Explorer κλπ.

- Πολλαπλά λειτουργικά. Οι εφαρμογές που βασίζονται στο Web είναι συμβατές με τα περισσότερα λειτουργικά συστήματα υπολογιστών. Η ποιότητα, οι επιδόσεις και οι λειτουργίες των εφαρμογών παραμένουν οι ίδιες όλη την ώρα. Αντιθέτως, οι παλαιότερες εφαρμογές είναι διαθέσιμες για ορισμένα λειτουργικά και πολλές φορές μια επιχείρηση χρειάζεται να έχει περισσότερους υπολογιστές για να καλύψει τις ανάγκες της.
- Απαιτείται λιγότερος χώρος στο δίσκο. Μία εφαρμογή που βασίζεται στο Web συνήθως απαιτεί μικρό χώρο στο δίσκο αφού τα δεδομένα αποθηκεύονται εξ αποστάσεως. Ένα από τα πιο συναρπαστικά χαρακτηριστικά είναι ότι οι εφαρμογές που βασίζονται στον ιστό είναι ελεύθερες από ιούς, επειδή δεν είναι εγκατεστημένες σε έναν υπολογιστή.
- Κόστος. Οι εφαρμογές ιστού είναι συνήθως πολύ φθηνότερες από την επιφάνεια εργασίας. Και αυτό στην περίπτωση που πρέπει να αγοραστούν, πράγμα που δεν αποτελεί σύνηθες φαινόμενο. Ταυτόχρονα, η ανάπτυξη μιας εφαρμογής για επιτραπέζιους υπολογιστές είναι μερικές φορές πιο δαπανηρή σε σύγκριση με μία web-based και αυτό επειδή μπορείτε να αναπτύξετε μια εφαρμογή για τον ιστό χρησιμοποιώντας HTML, CSS και JS. Η δημιουργία εφαρμογών με αυτά τα απλά πρότυπα καθιστά τη ζωή των προγραμματιστών ευκολότερη.

Λογισμικό Ανοιχτού Κώδικα

Στον χώρο της πληροφορικής και των ηλεκτρονικών υπολογιστών, με τον όρο λογισμικό ανοικτού κώδικα εννοείται λογισμικό του οποίου ο πηγαίος κώδικας διατίθεται στο κοινό. Κατά καιρούς έχουν εμφανιστεί αρκετές διαφορετικές άδειες χρήσης σχεδιασμένες να συνοδεύουν λογισμικό ανοικτού κώδικα. Μερικές από αυτές επιτρέπουν στους χρήστες και να τροποποιήσουν τον κώδικα ή και να τον αξιοποιήσουν σε άλλες εφαρμογές. “Ανοιχτός κώδικας” δε σημαίνει ότι ο οποιοσδήποτε μπορεί να έχει πρόσβαση στο λογισμικό. Σημαίνει, ωστόσο, ότι ο πηγαίος κώδικας του λογισμικού είναι διαθέσιμος στον οποιοδήποτε θέλει να τον χρησιμοποιήσει. Πίσω από κάθε ΕΛ/ΛΑΚ βρίσκεται μια ολόκληρη κοινότητα προγραμματιστών που το αναπτύσσουν, το δοκιμάζουν, το ελέγχουν. Όσο περισσότεροι άνθρωποι βλέπουν και ελέγχουν ένα κώδικα, τόσο πιο πιθανό είναι τα σφάλματα να γίνουν αντιληπτά και να διορθωθούν γρήγορα.

Καθώς υπάρχουν αμέτρητοι προγραμματιστές και χρήστες που εργάζονται για τη βελτίωση της ασφάλειας των λογισμικών ανοικτού κώδικα, αντίστοιχα πολλές είναι και οι καινοτομίες και οι αναβαθμίσεις των προϊόντων αυτών. Σχεδόν κάθε λογισμικό επιδέχεται παραμετροποίησης μέχρι κάποιο βαθμό. Η διαφορά του ανοικτού λογισμικού είναι ότι η παραμετροποίηση αυτή δεν έχει όριο. Με τις απαραίτητες τεχνικές γνώσεις και ικανότητες γίνεται να δημιουργηθεί ένα εντελώς νέο προϊόν, που να ανταποκρίνεται στις ανάγκες του κάθε προβλήματος. Η κοινότητα χρηστών και προγραμματιστών που χτίζεται γύρω από κάθε λογισμικό είναι αυτή που προτείνει και δοκιμάζει νέες ιδέες με βάση τις ανάγκες των χρηστών. Αυτή η πηγή πληροφόρησης είναι ανεκτίμητη και έχει πολύ μεγαλύτερη δυναμική από κάθε ανατροφοδότηση που θα μπορούσε να έχει ένα λογισμικό κλειστού κώδικα. Σε εκατοντάδες διαδικτυακά φόρουμ, λίστες αλληλογραφίας, ή ακόμα και “ζωντανά” chat rooms υπάρχουν χρήστες που έχουν συναντήσει κάποιο παρόμοιο πρόβλημα καθώς και οι λύσεις στις οποίες κατέφυγαν.

Μια επιπλέον κρίσιμη παράμετρος είναι το “κλειδώμα” σε συγκεκριμένο λογισμικό. Αν το ψηφιακό περιεχόμενο (content) που παράγεται εξαρτάται άμεσα από το λογισμικό που χρησιμοποιεί ο καθένας και δεν δίνεται η δυνατότητα να το επεξεργασίας μέσω κάποιου άλλου αντίστοιχου λογισμικού τότε δημιουργείται μία εξάρτηση από το πρώτο. Και αν για κάποιο λόγο σταματήσει η υποστήριξή του από την εταιρία που το αναπτύσσει τότε όλο το υλικό απαξιώνεται. Τα περισσότερα ανοιχτά λογισμικά είναι εγγενώς διαλειτουργικά. Χρησιμοποιώντας ανοιχτά πρότυπα παρέχουν τη δυνατότητα να μεταφερθούν τα δεδομένα σε κάποιο άλλο αντίστοιχο λογισμικό. Ακόμα και αν το αρχικό λογισμικό ανταποκρίνεται στην λύση κάποιου προβλήματος, με το ελεύθερο λογισμικό δίνεται η δυνατότητα μέσω μίας συνεργασίας με προγραμματιστές, να προσαρμοστεί στο εκάστοτε πρόβλημα και τις απαιτήσεις του.

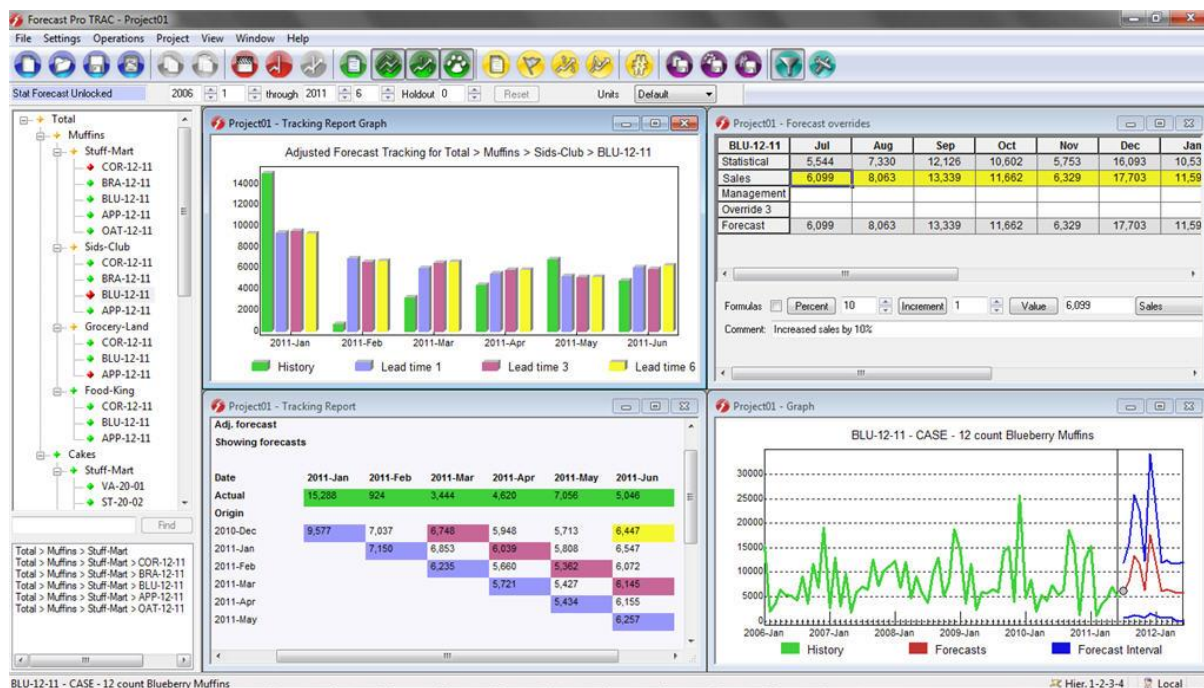
Χρησιμοποιώντας ανοιχτό λογισμικό η διάρκεια ζωής των παγίων στα οποία έχει επενδύσει κάθε επιχείρηση, αυξάνεται. Όταν μια επιχείρηση χρησιμοποιεί λογισμικά κλειστού κώδικα όπως Microsoft Windows και Office, εισέρχεται σε ένα κύκλο εξάρτησης που απαιτεί την ταυτόχρονη αναβάθμιση του hardware με το λογισμικό επ’ αόριστον. Με το λογισμικό ανοικτού κώδικα από την άλλη υπάρχουν επιλογές μικρότερων απαιτήσεων σε πόρους, οι οποίες επιμηκύνουν τη ζωή του hardware παλαιότερης γενιάς. Εξαρτάται από την ίδια την επιχείρηση και όχι από κάποια εταιρεία – πάροχο λογισμικού το πότε θα αναβαθμίσει τους υπολογιστές της. Το παραπάνω έχει ως αποτέλεσμα να αυξάνονται οι δυνατότητες ελέγχου της επιχείρησης. Πιο συγκεκριμένα, η διαφορά στη δυνατότητα ελέγχου που παρέχει το ανοιχτό λογισμικό απέναντι στο “κλειστό” μπορεί να αποδοθεί σχηματικά με τη σύγκριση μιας μηχανής στις οποίας τα εξαρτήματα δίνεται πρόσβαση με ένα

“μαύρο κουτί”. Στην πρώτη περίπτωση γίνεται να επισημανθεί τι δε λειτουργεί σωστά, ενώ στη δεύτερη θα πρέπει η κατασκευάστρια εταιρία να κληθεί για υποστήριξη, έναντι κάποιου χρηματικού ποσού. Ως αποτέλεσμα από το παραπάνω το κυριότερο κίνητρο για χρήση ανοικτού λογισμικού είναι το κόστος του: είναι δωρεάν. Και αυτό δεν ισχύει μόνο για το κόστος απόκτησης του λογισμικού, αλλά και για εργασίες που σχετίζονται με αυτό, όπως η υποστήριξη και η αναβάθμιση.

Έρευνα Εργαλείων Παραγωγής Προβλέψεων

Στα πλαίσια της αξιολόγησης του συστήματος που δημιουργήθηκε, είναι απαραίτητο να υπάρχει μέτρο σύγκρισης. Για τον λόγο αυτό εξετάστηκαν τα χαρακτηριστικά των ήδη υαρχόντων συστημάτων τα οποία παρουσιάζονται πιο αναλυτικά παρακάτω.

ForecastPRO

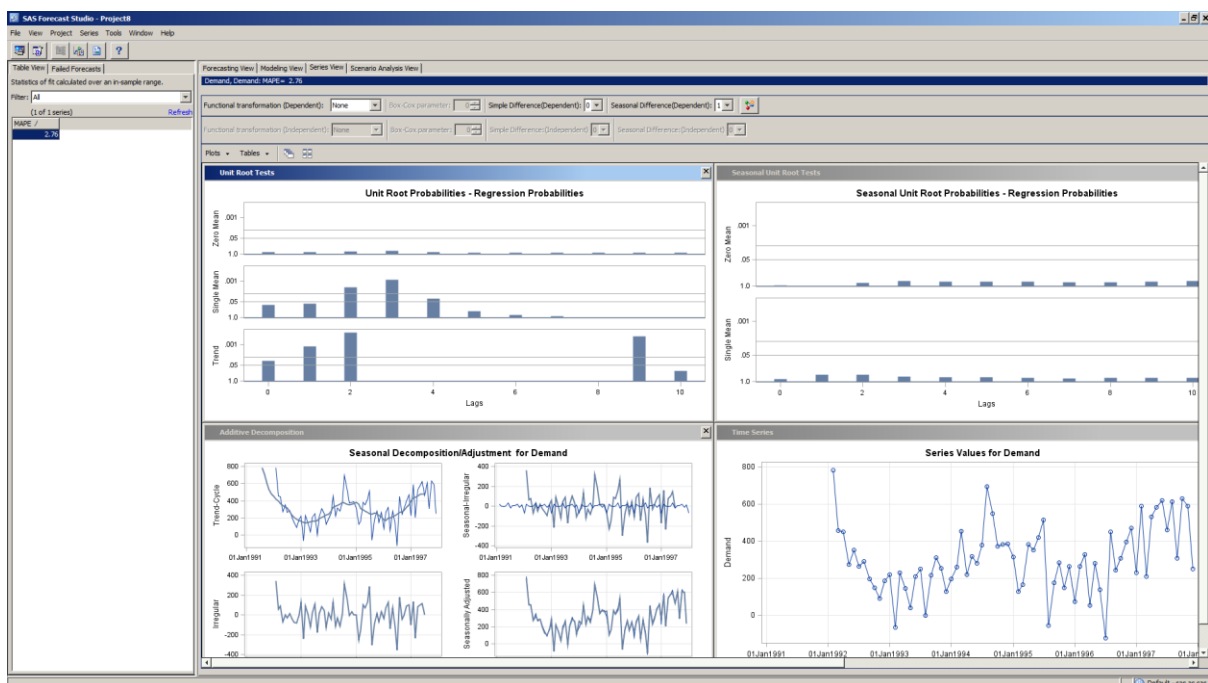


Το ForecastPro είναι ένα ολοκληρωμένο σύστημα διαχείρισης προβλέψεων που παρέχει σε επαγγελματίες τη δυνατότητα παραγωγής αξιόπιστων προβλέψεων, και τα εργαλεία για την αποτελεσματική διαχείριση, παρακολούθηση και βελτίωση της διαδικασίας των επιχειρηματικών προβλέψεων. Το σύστημα αυτό είναι σχεδιασμένο ειδικά για επαγγελματίες. Η διεπαφή χρήστη του ForecastPro εμφανίζει ιεραρχικά τα δεδομένα σε δομή που μοιάζει με δέντρο, επιτρέποντας στους χρήστες να δουλεύουν σε οποιοδήποτε επίπεδο εύκολα. Επίσης, το σύστημα διαθέτει δυνατότητα παραγωγής γραφικών παραστάσεων και πινάκων που ενημερώνονται άμεσα και επιτρέπει σε στελέχη μεγάλων εταιριών να παρουσιάζουν τα αποτελέσματα των προβλέψεων τους μέσω της επιλογής εξαγωγής αναφορών σε αρχεία Excel. Το ForecastPro περιλαμβάνει εργαλεία για συνεργασία κατά τη διάρκεια της διαδικασίας παραγωγής προβλέψεων, δυνατότητες για ιεραρχικές προβλέψεις, ευέλικτες μορφές αναφορών, παρακολούθησης και αξιολόγησης, δυνατότητα για ενσωμάτωση με βάσεις δεδομένων και εταιρικά συστήματα, εγγενή υποστήριξη 32-bit και 64-bit μηχανημάτων. Επίσης διατίθεται σε πολλές γλώσσες όπως τα αγγλικά, τα κινέζικα, τα γαλλικά, τα γερμανικά, τα ιαπωνικά, τα πορτογαλικά και τα ισπανικά. Τέλος, το ForecastPro έχει ενσωματωθεί επισήμως σε διάφορα εμπορικά συστήματα, συμπεριλαμβανομένων των ERP, DP, SCM και άλλων συστημάτων. Είναι, γενικότερα, η πιο δημοφιλής επιλογή σε ότι αφορά συστήματα υποστήριξης προβλέψεων. Παρά τη δημοτικότητα του συστήματος, υπάρχουν τα εξής μειονεκτήματα στη λύση αυτή:

- Μη φιλική διεπαφή χρήστη

- Υπερβολικά μεγάλο πλήθος επιλογών που μπορεί να οδηγήσει τον χρήστη σε σύγχυση
- Εφαρμογή με τοπική εγκατάσταση
- Έλλειψη δυνατότητας για εκτέλεση στο σύννεφο
- Υψηλό κόστος, ειδικά για μικρές επιχειρήσεις
- Έλλειψη δυνατότητας για ενσωμάτωση των τελευταίων μεθόδων

SAS® Forecast Studio



Το SAS Forecast Studio έχει σχεδιαστεί για να επιτρέπει ακόμα και σε αρχάριους χρήστες να προχωρούν εύκολα στη διαδικασία πρόβλεψης με πολύ αυτοματοποιημένο τρόπο. Ο χρήστης δεν χρειάζεται να γνωρίζει κανένα προγραμματισμό SAS ή να έχει γνώσεις σχετικές μοντέλα χρονοσειρών αφού το σύστημα είναι εντελώς αυτοματοποιημένο. Φυσικά, καθώς οι γνώσεις του χρήστη για τις μεθόδους πρόβλεψης μεγαλώνουν το σύστημα μπορεί να λειτουργεί ακόμα πιο αποτελεσματικά. Παρόλο που η διεπαφή χρήστη είναι εξορθολογισμένη ώστε να επιτρέπει στους λιγότερο έμπειρους προγνώστες να μετακινούνται γρήγορα μέσω του της διαδικασίας πρόβλεψης, η εφαρμογή έχει όλο το βάθος που αναμένουν οι πιο προηγμένοι αναλυτές.

Το SAS Forecast Studio παρέχει μια σειρά από εργαλεία δημιουργίας μοντέλων, επιτρέποντας σε πιο έμπειρους αναλυτές να επωφεληθούν από την παραγωγικότητα του περιβάλλοντος χρήστη. Όλες οι εργασίες που γίνονται στο SAS Forecast Studio αποθηκεύονται σε έργα. Ένα έργο SAS Forecast Studio περιέχει πληροφορίες σχετικά με τα δεδομένα που χρησιμοποιούνται για την πρόβλεψη, τις λεπτομέρειες σχετικά με τη διαδικασία πρόβλεψης, τις προτιμήσεις που έχουν καθοριστεί από το χρήστη, τυχόν σημειώσεις που εισάγονται από τους χρήστες, τον προκύπτοντα κώδικα SAS Forecast Studio και το σύνολο των δεδομένων εξόδου. Για να βοηθήσει τους χρήστες να ξεκινήσουν, το SAS Forecast Studio περιλαμβάνει οδηγίες για να καθοδηγήσει τους χρήστες μέσω των βημάτων που απαιτούνται για δημιουργία και χειρισμό ενός έργου.

Τα βήματα, συνοπτικά, περιλαμβάνουν:

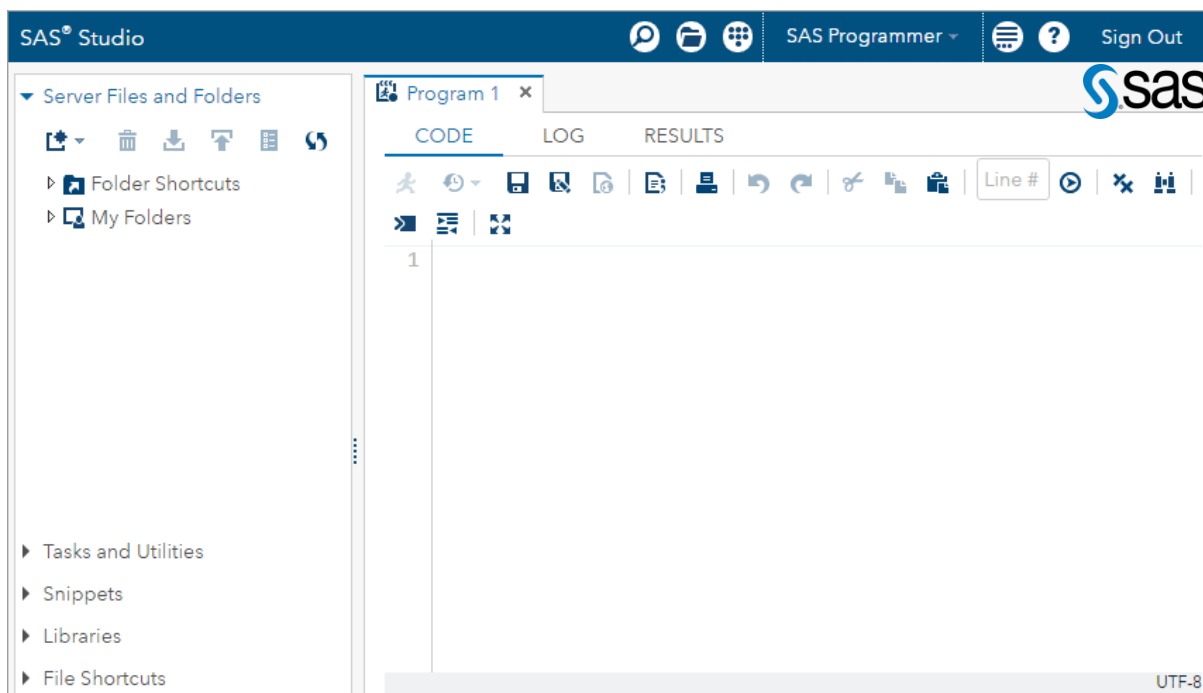
- Επιλογή του συνόλου δεδομένων εισόδου που πρέπει να προβλεφθεί
- Ανάθεση μεταβλητών ρόλων
- Καθορισμός ιεραρχίας πρόβλεψης (προαιρετικό)
- Καθορισμός επιλογών πρόβλεψης σε επίπεδο έργου
- Επιλογή θέσης για την αποθήκευση του έργου
- Δημιουργία μεταβλητών συμβάντων για χρήση στη διαδικασία πρόβλεψης (προαιρετικό)

Τα παραπάνω είναι τα δύο δημοφιλέστερα συστήματα για χρήση από επιχειρηματίες. Είναι και τα δύο συστήματα βασισμένα σε τοπική εγκατάσταση και δίνουν πολλές και εξειδικευμένες επιλογές που σε πολλές περιπτώσεις μπορεί να δημιουργήσουν σύγχυση σε έναν χρήστη χωρίς ειδικές γνώσεις στον τομέα των προβλέψεων. Πέρα από τα συστήματα για εταιρίες, είναι σημαντικό να αναφερθεί και ένα σύστημα προορισμένο για χρήση από ακαδημαϊκούς.

Τα μειονεκτήματα του συστήματος αυτού είναι τα παρακάτω:

- Μη φιλική διεπαφή χρήστη
- Υπερβολικά μεγάλο πλήθος επιλογών που μπορεί να οδηγήσει τον χρήστη σε σύγχυση
- Εφαρμογή με τοπική εγκατάσταση
- Υψηλό κόστος, ειδικά για μικρές επιχειρήσεις
- Για την ενσωμάτωση μεθόδων απαιτείται η γνώση της γλώσσας προγραμματισμού του συστήματος

SAS® University Edition



Το SAS University Edition περιλαμβάνει το SAS Studio, το SAS Base, το SAS / STAT, το SAS / IML, το SAS / ACCESS και διάφορες διαδικασίες χρονοσειρών από το SAS / ETS. Είναι το ίδιο

λογισμικό αναλυτικών στοιχείων παγκόσμιας κλάσης που χρησιμοποιείται από περισσότερες από 83.000 επιχειρηματικές, κυβερνητικές και πανεπιστημιακές τοποθεσίες σε όλο τον κόσμο. Μερικά από τα χαρακτηριστικά του είναι η διεπαφή που επιτρέπει στον χρήστη να αλληλεπιδρά με το λογισμικό από τον σταθμό εργασίας PC, Mac ή Linux, τα αξιόπιστα εργαλεία του και οι σύγχρονες στατιστικές μέθοδοι που περιλαμβάνουν διάφορες διαδικασίες προβλέψεων χρονοσειρών - TIMEDATA, TIMESERIES, ARIMA, ESM, UCM και TIMEID - περιλαμβάνονται από το SAS / ETS. Ακόμη το σύστημα αυτό δίνει τη δυνατότητα για πιο εμπειριστατωμένη, εξειδικευμένη ανάλυση και εξερεύνηση μέσω μιας δικής του ευέλικτης γλώσσας προγραμματισμού. Το SAS University Edition διατίθεται μέσω του Amazon Web Services, έτσι ώστε να γίνεται να τρέξει απευθείας στο σύννεφο, αντί να χρειάζεται να εγκατασταθεί τοπικά.

Συνοψίζοντας, υπάρχουν αδιαμφισβήτητα πολλά ισχυρά εργαλεία για την παραγωγή προβλέψεων, έχουν, όμως, όλα διάφορα αρνητικά χαρακτηριστικά. Στο παρακάτω κεφάλαιο, θα παρουσιαστεί πως προσπαθήσαμε να αντιμετωπίσουμε τις αδυναμίες των ήδη υπαρχόντων συστημάτων και να δημιουργήσουμε μια νέα ενιαία λύση, για χρήση από επαγγελματίες και ερευνητές.

Κεφάλαιο 3

Σχεδιασμός Αρχιτεκτονικής της Πλατφόρμας

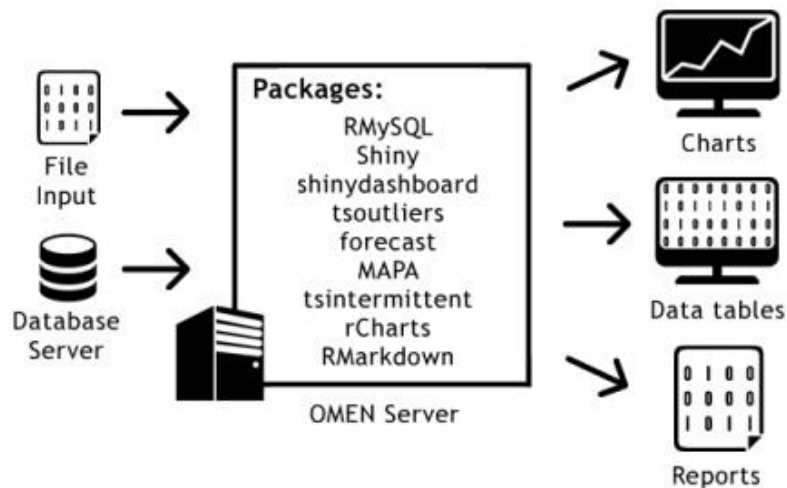
Γενικός Σχεδιασμός

Ο σχεδιασμός του συστήματος βασίστηκε σε μία απλή αρχιτεκτονική χωρισμένη σε επίπεδα ανάλογα με το πόσο κοντά ή μακριά γίνονται οι διεργασίες του κάθε επιπέδου με τον χρήστη. Έγινε προσπάθεια να ενσωματωθούν σύγχρονες τεχνικές για τη γρήγορη και ορθή ανάπτυξη της πλατφόρμας με όσο πιο δομημένο και λογικό τρόπο.

Το πρώτο επίπεδο αποτελείται από τους χρήστες του συστήματος. Όλο το σύστημα έχει σχεδιαστεί με βάση τις ανάγκες και τις απαιτήσεις των χρηστών με στόχο τη βέλτιστη αλληλεπίδραση με τα εργαλεία που προσφέρονται και την αποφυγή της ανάγκης για επαφή μεταξύ κώδικα και χρήστη. Οι χρήστες έρχονται σε αλληλεπίδραση με το σύστημα μέσω του browser από οποιαδήποτε διαθέσιμη συσκευή.

Συνεχίζοντας δεύτερο επίπεδο του σχεδιασμού, έχουμε τις ενέργειες και τα στοιχεία που εκτελούνται στον browser και αποτελούν ό,τι βλέπει ο χρήστης. Ο σχεδιασμός σε αυτό το επίπεδο βασίστηκε στις πιο διαδεδομένες γλώσσες για front-end περιεχόμενο και σε νέες τάσεις σχεδίασης. Κατά βάση χρησιμοποιήθηκε το Bootstrap Framework και η CSS σαν style sheet γλώσσα και η JavaScript για τις script ανάγκες του συστήματος. Ο συνδυασμός αυτός προσφέρεται μέσω των πακέτων R RShiny και shinydashboard, που δημιουργούν εύκολα μία διεπαφή χρήστη απλή στη χρήση και προσαρμόσιμη σε διάφορα μεγέθη οθονών. Σε νοοτροπία σχεδιασμού ακολουθήσαμε την νέα τάση που είναι τα dashboard με μορφή πίνακα ελέγχου με πλάγιο μενού και άλλα στοιχεία όπως ημερολόγια, επιλογή διαστημάτων και άλλα.

Τα επόμενα επίπεδα αποτελούν το back end κομμάτι του όλου συστήματος. Στον web server υλοποιείται όλη η λογική. Είναι ο σύνδεσμος μεταξύ πελάτη και βάσης δεδομένων και από εκείνον περνάει όλη η επεξεργασία των αιτήσεων και των δεδομένων. Αποφασίσαμε να χρησιμοποιήσουμε έναν Linux Web Server, μέσω του RStudio Server, που επιτρέπει την δημιουργία browser-based διεπαφών, σε εφαρμογές γραμμένες σε R. Η κύρια γλώσσα σε αυτό το επίπεδο είναι η R. Επιλέχθηκε λόγω του πλήθους δυνατοτήτων που προσφέρει για ανάπτυξη στατιστικών εφαρμογών. Στο τελευταίο επίπεδο της σχεδίασης βρίσκεται η βάση δεδομένων. Η βάση είναι μία sql βάση και συγκεκριμένα χρησιμοποιήθηκε μια MySQL βάση, η οποία αποτελεί εργαλείο ανοιχτού κώδικα. Στην παρακάτω εικόνα φαίνεται εποπτικά η γενική αρχιτεκτονική του συστήματος την οποία θα αναλύσαμε με λεπτομέρεια.



R Statistical Software

Η R είναι μια γλώσσα προγραμματισμού ειδικά σχεδιασμένη για ανάπτυξη εφαρμογών στατιστικού περιεχομένου. Πρόκειται για ένα GNU project, παρόμοιο με τη γλώσσα S που αναπτύχθηκε από τους Bell Laboratories (πρώην AT & T, τώρα Lucent Technologies) από τον John Chambers και τους συναδέλφους του. Η R μπορεί να θεωρηθεί ως μια διαφορετική εφαρμογή της S. Υπάρχουν μερικές σημαντικές διαφορές, αλλά μεγάλο μέρος του κώδικα που γράφτηκε για την S παραμένει αμετάβλητο και στην R. Η R παρέχει μια μεγάλη στατιστικών μεθόδων (γραμμική και μη γραμμική μοντελοποίηση, ανάλυση χρονοσειρών, ταξινόμηση, συστοιχία, ...) και τεχνικών γραφικής παρουσίασης και είναι εξαιρετικά επεκτάσιμη. Η γλώσσα S συχνά το πρώτο βήμα για ανάπτυξη στατιστικών εφαρμογών και ουσιαστικά μέσω της R παρέχεται μια επιλογή ανοιχτού κώδικα για τη συμμετοχή σε αυτή την ανάπτυξη εφαρμογών.

Ένα από τα πλέον δυνατά σημεία της R είναι η ευκολία με την οποία μπορούν να παραχθούν καλά σχεδιασμένα γραφήματα, με επίπεδα ποιότητας δημοσίευσης, καθώς και μαθηματικές πράξεις και τύποι.

Η R διατίθεται ως Ελεύθερο Λογισμικό σύμφωνα με τους όρους της Γενικής Δημόσιας Άδειας GNU του [GNU Free Software Foundation](#) σε μορφή πηγαίου κώδικα. Συγκεντρώνει και τρέχει σε μεγάλη ποικιλία πλατφορμών UNIX και παρόμοιων συστημάτων (συμπεριλαμβανομένων των FreeBSD και Linux), Windows και MacOS.

Αποτελεί μία ολοκληρωμένη σύνθεση λογισμικού για την επεξεργασία δεδομένων, τον υπολογισμό πολύπλοκων παραστάσεων και τη γραφική απεικόνιση. Περιλαμβάνει μεθοδολογίες και εργαλεία για:

- αποτελεσματική διαχείριση και αποθήκευση δεδομένων,
- χρήση μαθηματικών τελεστών για υπολογισμούς σε συστοιχίες δεδομένων (συγκεκριμένα πίνακες, λίστες κ.ά.)
- ανάλυση δεδομένων

- δημιουργία και επεξεργασία γραφικών παραστάσεων για την ανάλυση και την απεικόνιση δεδομένων, είτε επί της οθόνης είτε σε έντυπη μορφή, και
- ανάπτυξη απλών προγραμματιστικών διεργασιών προγραμματισμού που όπως μεταβλητές, βρόχους, αναδρομικές λειτουργίες που ορίζονται από το χρήστη και εγκαταστάσεις εισόδου και εξόδου.

Ο όρος "περιβάλλον" ο οποίος χρησιμοποιείται συχνά για τον χαρακτηρισμό της R έχει ως στόχο να την αναδείξει ως πλήρως προγραμματισμένο και συνεκτικό σύστημα, και όχι ως μια συλλογή πολύ συγκεκριμένων άκαμπτων εργαλείων, όπως συμβαίνει συχνά με άλλα λογισμικά ανάλυσης δεδομένων.

Μεγάλο μέρος του περιβάλλοντος είναι γραμμένο στη ίδια διάλεκτο R του S, πράγμα που διευκολύνει τους χρήστες να ακολουθήσουν τις αλγοριθμικές επιλογές που έχουν γίνει. Για πιο εξειδικευμένες υπολογιστικές εργασίες, δίνεται η δυνατότητα σύνδεσης και κλήσης σε χρόνο εκτέλεσης, κομματιών κώδικα σε C, C++ και Fortran, ενώ οι πιο προχωρημένοι χρήστες μπορούν να γράψουν κώδικα C για να χειριστούν άμεσα αντικείμενα R.

Η R είναι βασισμένη στην ανάπτυξη μέσω της κοινότητας χρηστών της και μπορεί να επεκταθεί (εύκολα) μέσω πακέτων. Υπάρχουν περίπου οκτώ πακέτα που παρέχονται με την αρχική διανομή R και πολλά άλλα πακέτα που καλύπτουν ένα ευρύ φάσμα σύγχρονων στατιστικών εργασιών είναι διαθέσιμα μέσω του CRAN.

Τέλος, η R έχει τη δική της μορφή τεκμηρίωσης που μοιάζει με LaTeX, η οποία χρησιμοποιείται για την παροχή περιεκτικών οδηγιών, τόσο on-line σε διάφορες μορφές όσο και σε έντυπη μορφή.

MySQL

Η MySQL είναι μία γλώσσα προγραμματισμού για Συστήματα Διαχείρισης Βάσεων Δεδομένων (DBMS). Προκειμένου τα DBMS να μπορούν να έχουν κάποιους κοινούς άξονες αναφοράς έχει αναπτυχθεί η γλώσσα SQL (Structured Query Language) ή οποία επιτρέπει μέσα από ένα σύστημα εντολών, την προσπέλαση των δομημένων δεδομένων που περιέχει ένα DBMS. Κάθε DBMS ουσιαστικά δέχεται και εκτελεί ένα σετ εντολών SQL για να διαχειριστεί τα δεδομένα του. Η MySQL είναι μία διανομή της γλώσσας SQL και ένα MySQL DBMS είναι ένα σύστημα που δέχεται εντολές της MySQL.

Η MySQL αποτελεί ελεύθερο λογισμικό που μπορεί να χρησιμοποιήσει κανείς ερασιτεχνικά ή επαγγελματικά. Ο κωδικός του εγχειρήματος είναι διαθέσιμος μέσω της GNU General Public License, καθώς και μέσω ορισμένων ιδιόκτητων συμφωνιών. Ανήκει και χρηματοδοτείται από μία και μοναδική κερδοσκοπική εταιρία, τη σουηδική MySQL AB, η οποία σήμερα ανήκει στην Oracle. Καθώς η τεχνολογία των DBMS εξελίχθηκε στο χρόνο προέκυψαν διάφοροι διαχωρισμοί αυτών, ανάλογα με τον τύπο των δεδομένων που αποθηκεύονται σε αυτά. Οι οργανωμένες δομές δεδομένων αποθηκεύουν σύνολα δεδομένων(πλειάδες) και τον τρόπο με τον οποίο αυτά συσχετίζονται. Επειδή η λογική πίσω από τη διαχείριση αυτών των δεδομένων είναι η συσχέτιση (σχέση) των δεδομένων, αυτές οι βάσεις αποκαλούνται Σχεσιακές Βάσεις Δεδομένων και τα συστήματά τους σχεσιακά (RDBMS - Relational Database Managements Systems).

CSS-HTML-JavaScript

Η CSS (Cascading Style Sheets-Διαδοχικά Φύλλα Στυλ) είναι μια γλώσσα υπολογιστή που ανήκει στην κατηγορία των γλωσσών φύλλων στυλ και χρησιμοποιείται για τον έλεγχο της εμφάνισης ενός εγγράφου που έχει γραφτεί με μια γλώσσα σήμανσης. Χρησιμοποιείται, δηλαδή, για τον έλεγχο της εμφάνισης ενός εγγράφου που γράφτηκε στις γλώσσες HTML και XHTML. Η CSS είναι μια γλώσσα υπολογιστή προορισμένη να αναπτύσσει στυλιστικά μια ιστοσελίδα δηλαδή να διαμορφώνει περισσότερα χαρακτηριστικά, χρώματα, στοίχιση και δίνει περισσότερες δυνατότητες σε σχέση με την html. Για μια όμορφη και καλοσχεδιασμένη ιστοσελίδα η χρήση της CSS κρίνεται ως απαραίτητη.

Η HTML (αρχικοποίηση του αγγλικού HyperText Markup Language-Γλώσσα Σήμανσης Υπερκειμένου) είναι η κύρια γλώσσα σήμανσης για τις ιστοσελίδες, και τα στοιχεία της είναι τα βασικά δομικά στοιχεία των ιστοσελίδων. Η HTML γράφεται υπό μορφή στοιχείων HTML τα οποία αποτελούνται από ετικέτες (tags), οι οποίες περικλείονται μέσα σε σύμβολα «μεγαλύτερο από» και «μικρότερο από» (για παράδειγμα `<html>`), μέσα στο περιεχόμενο της ιστοσελίδας. Οι ετικέτες HTML συνήθως λειτουργούν ανά ζεύγη (για παράδειγμα `<h1>` και `</h1>`), με την πρώτη να ονομάζεται ετικέτα έναρξης και τη δεύτερη ετικέτα λήξης (ή σε άλλες περιπτώσεις ετικέτα ανοίγματος και ετικέτα κλεισίματος αντίστοιχα). Ανάμεσα στις ετικέτες, οι σχεδιαστές ιστοσελίδων μπορούν να τοποθετήσουν κείμενο, πίνακες, εικόνες κλπ.

Ο σκοπός ενός web browser είναι να διαβάσει τα έγγραφα HTML και τα συνθέτει σε σελίδες που μπορεί κανείς να διαβάσει ή να ακούσει. Ο browser δεν εμφανίζει τις ετικέτες HTML, αλλά τις χρησιμοποιεί για να ερμηνεύσει το περιεχόμενο της σελίδας.

Τα στοιχεία της HTML χρησιμοποιούνται για να κτίσουν όλους του ιστότοπους. Η HTML επιτρέπει την ενσωμάτωση εικόνων και άλλων αντικειμένων μέσα στη σελίδα, και μπορεί να χρησιμοποιηθεί για να εμφανίσει διαδραστικές φόρμες. Παρέχει τις μεθόδους δημιουργίας δομημένων εγγράφων (δηλαδή εγγράφων που αποτελούνται από το περιεχόμενο που μεταφέρουν και από τον κώδικα μορφοποίησης του περιεχομένου) καθορίζοντας δομικά σημαντικά στοιχεία για το κείμενο, όπως κεφαλίδες, παραγράφους, λίστες, συνδέσμους, παραθέσεις και άλλα. Μπορούν επίσης να ενσωματώνονται σενάρια εντολών σε γλώσσες όπως η JavaScript, τα οποία επηρεάζουν τη συμπεριφορά των ιστοσελίδων HTML.

Η JavaScript (JS) είναι διερμηνευμένη γλώσσα προγραμματισμού για ηλεκτρονικούς υπολογιστές. Αρχικά αποτελεί μέρος της υλοποίησης των φυλλομετρητών Ιστού, ώστε τα σενάρια από την πλευρά του πελάτη (client-side scripts) να μπορούν να επικοινωνούν με τον χρήστη, να ανταλλάσσουν δεδομένα ασύγχρονα και να αλλάζουν δυναμικά το περιεχόμενο του εγγράφου που εμφανίζεται. Η JavaScript είναι μια γλώσσα σεναρίων που βασίζεται στα πρωτότυπα (prototype-based), είναι δυναμική, με ασθενείς τύπους και έχει συναρτήσεις ως αντικείμενα πρώτης τάξης. Η σύνταξή της είναι επηρεασμένη από τη C. Η JavaScript αντιγράφει πολλά ονόματα και συμβάσεις ονοματοδοσίας από τη Java, αλλά γενικά οι δύο αυτές γλώσσες δε σχετίζονται και έχουν πολύ διαφορετική σημασιολογία. Οι βασικές αρχές σχεδιασμού της JavaScript προέρχονται από τις γλώσσες προγραμματισμού Self και Scheme. Είναι γλώσσα βασισμένη σε διαφορετικά προγραμματιστικά παραδείγματα (multi-paradigm), υποστηρίζοντας αντικειμενοστραφές, προστακτικό και συναρτησιακό στυλ προγραμματισμού.

Πακέτα R

RShiny

Το shiny είναι ένα πακέτο R με σκοπό την ανάπτυξη διαδραστικών εφαρμογών ιστού (apps) κατευθείαν μέσω R. Οι εφαρμογές shiny συγκεντρώνονται σε ένα αρχείο που ονομάζεται app.R. Το script app.R τοποθετείται σε έναν κατάλογο και η εφαρμογή μπορεί να τρέξει εφόσον ικανοποιηθούν τα τρία παρακάτω στοιχεία:

- ένα αρχείο διεπαφής χρήστη
- μια λειτουργία διακομιστή
- μια κλήση στη συνάρτηση shinyApp

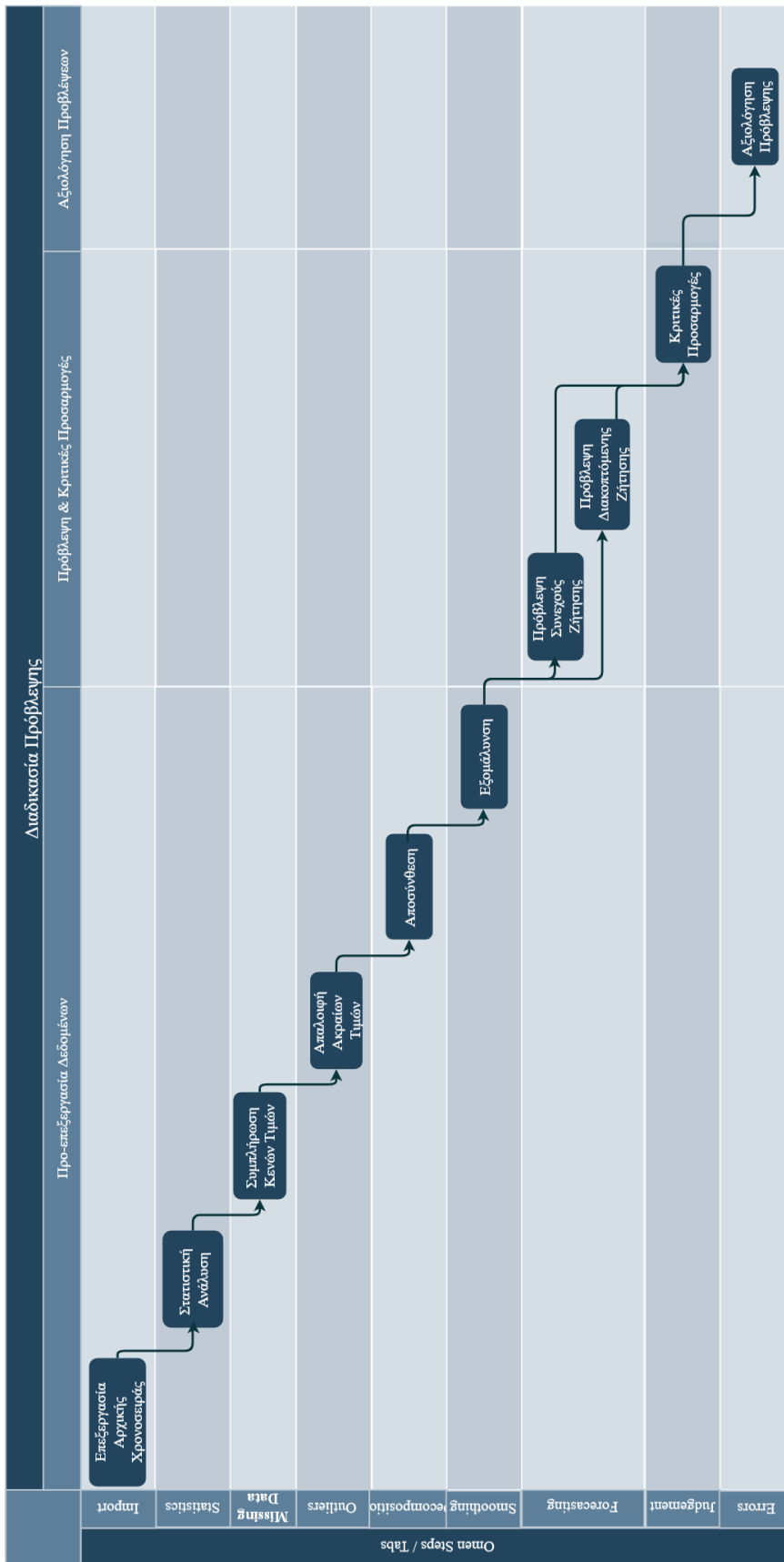
Το αντικείμενο διεπαφής χρήστη (ui) ελέγχει τη διάταξη και την εμφάνιση της εφαρμογής σας και αποτελεί ουσιαστικά το front-end της εφαρμογής. Η λειτουργία διακομιστή περιέχει τις οδηγίες που χρειάζεται ο υπολογιστής για να δημιουργήσει και να τρέξει την εφαρμογή. Τέλος, η λειτουργία shinyApp δημιουργεί αντικείμενα Shiny app χρησιμοποιώντας ένα ζεύγος UI / διακομιστή.

Shinydashboard

Το shinydashboard αποτελεί ένα πακέτο R που διευκολύνει τη δημιουργία shiny εφαρμογών με διεπαφή χρήστη με μορφή πίνακα ελέγχου. Το Shinydashboard υποστηρίζει όλες τις κανονικές διαδραστικές λειτουργίες που κάνει το Shiny. Εκτός από όλες τις shiny δυνατότητες εισόδου και εξόδου που μπορεί να υπάρχουν σε μια εφαρμογή shinydashboard, υπάρχουν μερικά πράγματα που προσφέρει αυτό το πακέτο στη δομή τέτοιων εφαρμογών. Συγκεκριμένα, οι περισσότερες εφαρμογές shinydashboard έχουν πλάγιο μενού.

Τα υπόλοιπα πακέτα που χρησιμοποιήθηκαν θα αναφερθούν παρακάτω με τις αντίστοιχες λειτουργίες στις οποίες συμβάλλουν.

Σενάρια Χρήσης της Πλατφόρμας



Για την οργάνωση και δομή του συστήματος λήφθηκε υπ' όψιν το διάγραμμα ροής κάθε διαδικασίας πρόβλεψης που φαίνεται στην προηγούμενη σελίδα.

Εγγραφή Νέου Χρήστη

Περιγραφή

Ένας νέος χρήστης θέλει να δημιουργήσει νέο λογαριασμό για το OMEN.

Προϋποθέσεις

Ο χρήστης βρίσκεται ήδη στο μενού Home.

Σενάριο

1. Ο χρήστης επιλέγει το radio button με την ετικέτα Sign Up
2. Ο χρήστης εισάγει όλα τα υποχρεωτικά πεδία (όνομα χρήστη, κωδικό, επαλήθευση κωδικού, email)
3. Ο χρήστης υποβάλλει την αίτηση
4. Ο web server ελέγχει αν υπάρχει χρήστης με το ίδιο όνομα χρήστη
5. Ο web server ενημερώνει τη βάση με τις πληροφορίες του χρήστη
6. Επιβεβαίωση επιτυχούς ή μη εγγραφής με μήνυμα στην οθόνη

Σύνδεση Χρήστη

Περιγραφή

Ένας χρήστης θέλει να συνδεθεί στο σύστημα ώστε να έχει πρόσβαση στο Dashboard

Προϋποθέσεις

Ο χρήστης βρίσκεται στο μενού Home

Σενάριο

- 1η επιλογή - Σύνδεση ως επισκέπτης (Continue without an account)
 1. Ο χρήστης επιλέγει το radio button με την ετικέτα Continue without an account
 2. Ο χρήστης βλέπει στο δεξί μέρος της οθόνης τα στοιχεία του guest account και έχει πρόσβαση στο dashboard
- 2η επιλογή - Σύνδεση με ήδη υπάρχων λογαριασμό χρήστη
 1. Ο χρήστης επιλέγει το radio button με την ετικέτα Sign in
 2. Ο χρήστης εισάγει όλα τα υποχρεωτικά πεδία (username, κωδικό)
 3. Ο χρήστης υποβάλλει την αίτηση
 4. Ο web server ελέγχει αν είναι έγκυρα τα στοιχεία
 5. Επιβεβαίωση επιτυχούς ή μη εγγραφής
 6. Ανακατεύθυνση στο Dashboard

Επιλογή Χρονοσειράς για Πρόβλεψη

Περιγραφή

Ένας χρήστης θέλει να επιλέξει μια χρονοσειρά για να αρχίσει τη διαδικασία της πρόβλεψης

Προϋποθέσεις

Ο χρήστης χρήστης έχει συνδεθεί στο σύστημα

Σενάριο

- 1η επιλογή - Επιλογή ήδη μεταφορτωμένης χρονοσειράς
 1. Ο χρήστης βρίσκεται στην αρχική σελίδα (Home)
 2. Ο χρήστης επιλέγει από το δεξί παράθυρο το radio button με ετικέτα την επιθυμητή χρονοσειρά
 3. Η χρονοσειρά έχει επιλεγεί και ο χρήστης μπορεί να προχωρήσει στην σελίδα Import Time Series
 4. Ο χρήστης βλέπει την χρονοσειρά στο παράθυρο με ετικέτα “View & Edit Imported Data” και την γραφική της στο παράθυρο με ετικέτα “View Imported & Edited Time Series”
 5. Ο χρήστης μπορεί να ορίσει τις ημερομηνίες που καθορίζει το εύρος του in-sample μέσω του ημερολογιακού μενού με ετικέτα “Date Selection”
 6. Η χρονοσειρά που έχει επιλεγεί ανανεώνεται και εμφανίζεται στο παράθυρο με ετικέτα “View & Edit Imported Data” και την γραφική της στο παράθυρο με ετικέτα “View Imported & Edited Time Series”
 7. Ο χρήστης μπορεί να επεξεργαστεί τις τιμές της χρονοσειράς στο παράθυρο με ετικέτα “View & Edit Imported Data”
 8. Η χρονοσειρά που έχει επεξεργαστεί ανανεώνεται και εμφανίζεται στο παράθυρο με ετικέτα “View & Edit Imported Data” και την γραφική της στο παράθυρο με ετικέτα “View imported & Edited Time Series”

- 2η επιλογή - Μεταφόρτωση χρονοσειράς
 1. Ο χρήστης βρίσκεται στην σελίδα Import Time Series
 2. Ο χρήστης επιλέγει το κουμπί με την ετικέτα Upload
 3. Ένα αναδυόμενο παράθυρο εμφανίζεται στο οποίο ο χρήστης μπορεί να επιλέξει μία χρονοσειρά από τον υπολογιστή του σε μορφή .csv αρχείου
 4. Ο χρήστης επιλέγει μία χρονοσειρά
 5. Ο χρήστης επιλέγει το radio button με ετικέτα το σύμβολο που έχει χρησιμοποιηθεί για να διαχωρίζει τις τιμές στο μεταφορτωμένο αρχείο
 6. Ο χρήστης βλέπει την χρονοσειρά στο παράθυρο με ετικέτα “View & Edit Imported Data” και την γραφική της στο παράθυρο με ετικέτα “View Imported & Edited Time Series”
 7. Ο χρήστης μπορεί να ορίσει τις ημερομηνίες που καθορίζει το εύρος του in-sample μέσω του ημερολογιακού μενού με ετικέτα “Date Selection”
 8. Η χρονοσειρά που έχει επιλεγεί ανανεώνεται και εμφανίζεται στο παράθυρο με ετικέτα “View & Edit Imported Data” και την γραφική της στο παράθυρο με ετικέτα “View Imported & Edited Time Series”
 9. Ο χρήστης μπορεί να επεξεργαστεί τις τιμές της χρονοσειράς στο παράθυρο με ετικέτα “View & Edit Imported Data”

10. Η χρονοσειρά που έχει επεξεργαστεί ανανεώνεται και εμφανίζεται στο παράθυρο με ετικέτα “View & Edit Imported Data” και την γραφική της στο παράθυρο με ετικέτα “View imported & Edited Time Series”

Στατιστική Ανάλυση Δεδομένων

Περιγραφή

Ένας χρήστης θέλει να δει την στατιστική ανάλυση της χρονοσειράς του.

Προϋποθέσεις

Ο χρήστης έχει συνδεθεί, έχει επιλέξει μία χρονοσειρά και βρίσκεται στο μενού Statistics.

Σενάριο

1. Ο χρήστης βλέπει τους στατιστικούς δείκτες και τις τιμές που έχουν για την χρονοσειρά που έχει επιλέξει. (Min, Mean, Max, 1st quantile, Median, 3rd quantile, Standard deviation, Skewness, Kurtosis, Coefficient of variation, Linear correlation coefficient, Observations, Hurst, Normality, Stationarity, Trend intensity (%), Seasonal intensity (%), Randomness intensity (%), ACF & PACF)
2. Ο χρήστης βλέπει τα στατιστικά διαγράμματα και τις τιμές που έχουν για τη χρονοσειρά που έχει επιλέξει (ACF diagram, PACF diagram, Boxplot, Kernel density estimates)

Συμπλήρωση Κενών Τιμών

Περιγραφή

Ένας χρήστης θέλει να συμπληρώσει τις κενές τιμές της χρονοσειράς του.

Προϋποθέσεις

Ο χρήστης έχει συνδεθεί, έχει επιλέξει μία χρονοσειρά και βρίσκεται στο μενού Missing Data.

Σενάριο

1. Ο χρήστης επιλέγει το radio button με ετικέτα την μέθοδο συμπλήρωσης κενών τιμών που επιθυμεί (Use last non-zero value, Use linear regression, Multivariate imputations (trend & period), Bootstrap EM Algorithm)
2. Ο χρήστης βλέπει τις τιμές της αρχικής και την ανανεωμένης χρονοσειράς σε μορφή πίνακα στο παράθυρο με ετικέτα “Original & Filled Data”
3. Ο χρήστης βλέπει τη γραφική παράσταση της αρχικής και την ανανεωμένης χρονοσειράς στο παράθυρο με ετικέτα “Original & Filled Time Series”

Απαλοιφή Ακραίων Τιμών

Περιγραφή

Ένας χρήστης θέλει να αφαιρέσει τις ακραίες τιμές της χρονοσειράς του.

Προϋποθέσεις

Ο χρήστης έχει συνδεθεί, έχει επιλέξει μία χρονοσειρά, και βρίσκεται στο μενού Handling Outliers.

Σενάριο

1. Ο χρήστης επιλέγει την μέθοδο απαλοιφής ακραίων τιμών που επιθυμεί από το dropdown menu (by the Chen & Liu procedure, By fitting a forecasting model & observing unusual values, By examining the distribution of the forecasting errors, By identifying extreme values)
2. Ανάλογα με τη μέθοδο που έχει επιλέξει πρέπει να επιλέξει τις τιμές παραμέτρων της μεθόδου, που εμφανίζονται αυτόματα μετά την επιλογή
3. Ο χρήστης βλέπει τις τιμές της αρχικής και την ανανεωμένης χρονοσειράς σε μορφή πίνακα στο παράθυρο με ετικέτα “Original & Normalised Data”
4. Ο χρήστης βλέπει τη γραφική παράσταση της αρχικής και την ανανεωμένης χρονοσειράς στο παράθυρο με ετικέτα “Original & Normalised Time Series”

Αποσύνθεση Χρονοσειράς

Περιγραφή

Ένας χρήστης θέλει να πραγματοποιήσει αποσύνθεση της χρονοσειράς του.

Προϋποθέσεις

Ο χρήστης έχει συνδεθεί, έχει επιλέξει μία χρονοσειρά, έχει περάσει τα βήματα Missing Data και Handling Outliers και βρίσκεται στο μενού Decomposition.

Σενάριο

1. Ο χρήστης βλέπει τη συχνότητα που είτε έχει επιλεγεί αυτόματα είτε από τον χρήστη στο πρώτο βήμα καθώς και το επίπεδο εμπιστοσύνης των υπολογιζόμενων τιμών
2. Ο χρήστης βλέπει σε διαγράμματα τα τέσσερα μέρη της χρονοσειράς (Randomness, Cycle, Trend, Seasonality) καθώς και το Seasonplot της
3. Ο χρήστης βλέπει σε διάγραμμα με ετικέτα “Seasonally Adjusted Time Series” τις μέχρι τώρα επιλογές που έχει κάνει ανανεωμένες με την αποεποχικοποιημένη χρονοσειρά
4. Ο χρήστης βλέπει τις τιμές της αρχικής και την ανανεωμένης χρονοσειράς σε μορφή πίνακα στο παράθυρο με ετικέτα “Original & Deseasonalized Data”
5. Ο χρήστης επιλέγει αν θέλει να λάβει υπόψη την αποσύνθεση της χρονοσειράς στα επόμενα βήματα του κάνοντας κλικ στο checkbox με ετικέτα “Consider seasonal adjustments”

Εξομάλυνση Χρονοσειράς

Περιγραφή

Ένας χρήστης θέλει να πραγματοποιήσει εξομάλυνση της χρονοσειράς του.

Προϋποθέσεις

Ο χρήστης έχει συνδεθεί, έχει επιλέξει μία χρονοσειρά, έχει περάσει τα βήματα Missing Data, Handling Outliers και Decomposition και βρίσκεται στο μενού Smoothing.

Σενάριο

1. Ο χρήστης επιλέγει την μέθοδο εξομάλυνσης που επιθυμεί από το dropdown menu (Non Linear Smooth, Simple Moving Average)
2. Ανάλογα με τη μέθοδο που έχει επιλέξει μπορεί να επιλέξει τις τιμές παραμέτρων της μεθόδου, που εμφανίζονται αυτόματα μετά την επιλογή
3. Ο χρήστης βλέπει τις τιμές της αρχικής και την ανανεωμένης χρονοσειράς σε μορφή πίνακα στο παράθυρο με ετικέτα “Original & Smoothened Data”
4. Ο χρήστης βλέπει τη γραφική παράσταση της αρχικής και την ανανεωμένης χρονοσειράς στο παράθυρο με ετικέτα “Original & Smoothened Time Series”
5. Ο χρήστης επιλέγει αν θέλει να λάβει υπ’ όψιν την αποσύνθεση της χρονοσειράς στα επόμενα βήματα του κάνοντας κλικ στο checkbox με ετικέτα “Consider Smoothing Adjustments”

Χρονική Συνάθροιση Χρονοσειράς

Περιγραφή

Ένας χρήστης θέλει να πραγματοποιήσει χρονική συνάθροιση της χρονοσειράς του.

Προϋποθέσεις

Ο χρήστης έχει συνδεθεί, έχει επιλέξει μία χρονοσειρά, έχει περάσει τα βήματα Missing Data, Handling Outliers, Decomposition και Smoothing και βρίσκεται στο μενού Temporal Aggregation.

Σενάριο

1. Ο χρήστης επιλέγει το επίπεδο χρονικής συνάθροισης που επιθυμεί από το dropdown menu (Original Time Series, Annual Model, 2-Month Model, Trimester Model, Quarterly Model, Half-Year Model)
2. Ανάλογα με την επιλογή του εμφανίζεται το καλύτερο και το χειρότερο κομμάτι
3. Ο χρήστης βλέπει τις τιμές της αρχικής και των συναθροισμένων σε διάφορα επίπεδα χρονοσειρών σε μορφή πίνακα στο παράθυρο με ετικέτα “Original & Temporally Aggregated Data”
4. Ο χρήστης βλέπει τη γραφική παράσταση της αρχικής και των συναθροισμένων σε διάφορα επίπεδα χρονοσειρών στο παράθυρο με ετικέτα “Time series across different temporal aggregation levels”

Παραγωγή Στατιστικής Πρόβλεψης

Περιγραφή

Ένας χρήστης θέλει να παράξει προβλέψεις για τη χρονοσειρά του.

Προϋποθέσεις

Ο χρήστης έχει συνδεθεί, έχει επιλέξει μία χρονοσειρά, έχει περάσει τα βήματα Missing Data, Handling Outliers, Decomposition και Smoothing και βρίσκεται στο μενού Forecasting.

Σενάριο

1. Εμφανίζεται ο διάλογος Διακοπτόμενης Ζήτησης και ο χρήστης απαντά αν θέλει να χρησιμοποιήσει μεθόδους διακοπτόμενης ζήτησης
2. Ο χρήστης επιλέγει την μέθοδο πρόβλεψης που επιθυμεί από το dropdown menu ή την επιλογή Optimal αν θέλει να πραγματοποιήσει cross validation διαγωνισμό για τις προβλέψεις του
3. Ο χρήστης επιλέγει αν θέλει να υπολογιστούν διαστήματα εμπιστοσύνης καθώς και το επίπεδο εμπιστοσύνης από το dropdown με ετικέτα “Confidence Levels”
4. Ο χρήστης επιλέγει τον ορίζοντα πρόβλεψης που επιθυμεί από το slider με ετικέτα Horizon
5. Ο χρήστης βλέπει τις τιμές της αρχικής χρονοσειράς και της χρονοσειράς με τις προβλέψεις σε μορφή πίνακα στο παράθυρο με ετικέτα “Historical data & Forecasts”
6. Ο χρήστης βλέπει τη γραφική παράσταση της αρχικής χρονοσειράς και των προβλέψεων στο παράθυρο με ετικέτα “Historical Time Series & Forecasts”

Παραγωγή Κριτικής Πρόβλεψης

Περιγραφή

Ένας χρήστης θέλει να παράξει κριτικές προβλέψεις για τη χρονοσειρά του.

Προϋποθέσεις

Ο χρήστης έχει συνδεθεί, έχει επιλέξει μία χρονοσειρά, έχει περάσει τα βήματα Missing Data, Handling Outliers, Decomposition, Smoothing, Forecasting και βρίσκεται στο μενού Judgemental Adjustments.

Σενάριο

1. Ο χρήστης βλέπει τις στατιστικές προβλέψεις τόσο σε διάγραμμα όσο και σε κουτιά τιμών
2. Ο χρήστης προσαρμόζει τις στατιστικές προβλέψεις κάνοντας drag & drop στο διάγραμμα ή επεξεργάζοντας τα κουτιά τιμών

Δείκτες Σφάλματος και Αξιολόγηση Πρόβλεψης

Περιγραφή

Ένας χρήστης θέλει να δει τους δείκτες σφάλματος για την αξιολόγηση της χρονοσειράς του.

Προϋποθέσεις

Ο χρήστης έχει συνδεθεί, έχει επιλέξει μία χρονοσειρά, έχει περάσει τα βήματα Missing Data, Handling Outliers, Decomposition, Smoothing, Forecasting και Judgemental Adjustments και βρίσκεται στο μενού Error Metrics & Assessment.

Σενάριο

1. Ο χρήστης βλέπει τους δείκτες σφάλματος και τις τιμές που έχουν για την χρονοσειρά που έχει επιλέξει. (ME, MPE, MAPE, MSE, sMAPE)
2. Ο χρήστης βλέπει τα στατιστικά διαγράμματα και τις τιμές που έχουν για τη χρονοσειρά που έχει επιλέξει (In Sample-Forecast Model και Out of Sample-Forecasts Plots, το τελευταίο αν υπάρχει out of sample)

Εξαγωγή Αναφοράς

Περιγραφή

Ένας χρήστης θέλει να εξάγει μία αναφορά όλης της διαδικασίας που πραγματοποίησε.

Προϋποθέσεις

Ο χρήστης έχει συνδεθεί, έχει επιλέξει μία χρονοσειρά, έχει περάσει τα βήματα Missing Data, Handling Outliers, Decomposition, Smoothing, Forecasting, Judgemental Adjustments και Error Metrics & Assessment και βρίσκεται στο μενού Report.

Σενάριο

1. Ο χρήστης εισάγει τον τίτλο που θέλει να έχει η αναφορά του στο πεδίο Report Title
2. Ο χρήστης κάνει κλικ στο κουμπί και κατεβάζει αρχείο Word με την αναφορά του

Κεφάλαιο 4

Ενσωματωμένες Μέθοδοι και Συναρτήσεις

Ανάλυση χρονοσειρών

Συμπλήρωση Κενών Τιμών

Πολύ συχνά, οι χρονοσειρές σε διάφορες εταιρίες έχουν κενές τιμές, που μπορεί να οφείλονται είτε στην αδυναμία συλλογής τους, είτε στην έλλειψη τους. Σε κάθε περίπτωση, οι κενές αυτές τιμές μπορεί να οδηγήσουν το μοντέλο στην παραγωγή λανθασμένων προβλέψεων καθώς τα γεγονότα αυτά δεν αντικατοπτρίζουν την μελλοντική πραγματικότητα. Για το λόγο αυτό, θα πρέπει με κάποιο τρόπο να συμπληρωθούν έτσι ώστε να αφαιρεθεί κάθε τυχούσα επίδραση τους που μπορεί να οδηγήσει σε σφάλματα. Το OMEN διαθέτει τέσσερις μεθόδους για συμπλήρωση τυχόν κενών τιμών της χρονοσειράς. Οι μέθοδοι αναφέρονται αναλυτικά παρακάτω:

Last non-zero value

Αρχικά, η πρώτη επιλογή για συμπλήρωση τιμών με βάση τα υπάρχοντα δεδομένα ορίζεται στο σύστημα ως “By using the last non-missing value”. Μια μέθοδος για να γίνει αυτό είναι η μέθοδος Last-Observation-Carried-Forward (LOCF). Η τελευταία μεταγενέστερη παρατήρηση (LOCF) είναι μια μέθοδος που χρησιμοποιείται συχνά στην πράξη επειδή είναι πολύ εύκολη στη χρήση. Στη μέθοδο αυτή, η τελευταία έγκυρη τιμή χρησιμοποιείται ως τιμή όλων των μεταγενέστερων περιόδων που πρέπει να υπολογιστούν. Αυτή η μέθοδος εφαρμόζεται κυρίως στις κατηγορικές μεταβλητές για τις οποίες είναι γνωστό ότι αλλάζουν ελάχιστα ή καθόλου με την πάροδο του χρόνου. Ένα παράδειγμα μιας τέτοιας μεταβλητής είναι το γένος. Για άλλες κατηγορικές ή ποσοτικές μεταβλητές, αυτή η μέθοδος συχνά δημιουργεί κατά λάθος μια υπερβολικά σταθερή εικόνα της πραγματικής κατάστασης. Για παράδειγμα, για τα στοιχεία κάποιου δείκτη χρηματιστηρίου, η μέθοδος αυτή μπορεί να οδηγήσει σε λανθασμένα συμπεράσματα σε ότι αφορά τη σταθερότητα του. Στην LOCF, η τελευταία παρατηρούμενη τιμή y_{t-1} χρησιμοποιείται για να υπολογιστεί η τιμή y_t της χρονικής τιμής t που λείπει. Μια άλλη παραλλαγή της μεθόδου είναι η τελευταία έγκυρη παρατήρηση που μεταφέρεται προς τα πίσω (LOCB), στην οποία χρησιμοποιούμε την πρώτη έγκυρη τιμή y_{t+1} μετά την κενή τιμή y_t που πρέπει να υπολογιστεί. Η μέθοδος LOCF, μπορεί να χρησιμοποιηθεί και για την συμπλήρωση πολλών διαδοχικές κενών τιμών. Το πρόβλημα με το LOCF είναι ότι συχνά δεν δίνει ρεαλιστικό αποτέλεσμα και καταλήγει να συμπληρώνει την χρονοσειρά με μία σταθερή τιμή, ακόμα και όταν κάτι τέτοιο απέχει από την πραγματικότητα. Η συνάρτηση που χρησιμοποιείται για να πραγματοποιηθεί η λειτουργία αυτή είναι η `na.locf` του πακέτου R, `zoo`.

Linear regression

Ένα βασικό χαρακτηριστικό των δεδομένων χρονοσειρών είναι η τάση. Χρησιμοποιώντας την παλινδρόμηση μπορούμε να μοντελοποιήσουμε και να προβλέψουμε την τάση στα δεδομένα χρονοσειρών θεωρώντας $t = 1, \dots, T$ ως μεταβλητή πρόβλεψης του $y_t = \beta_0 + \beta_1 t + \varepsilon_t$. Προφανώς, οι παρατηρήσεις που λείπουν δεν βρίσκονται στην ευθεία αλλά είναι διάσπαρτες γύρω από αυτήν. Μπορούμε να σκεφτούμε κάθε παρατήρηση y_t που αποτελείται από το συστηματικό ή

εξηγημένο μέρος του μοντέλου, $\beta_0 + \beta_1 x_i \beta_0 + \beta_1 x_i$, και το τυχαίο "σφάλμα", ϵ_i . Ο όρος "σφάλμα" δεν υποδηλώνει λάθος, αλλά απόκλιση από το υποκείμενο μοντέλο ευθείας γραμμής.

Multivariate imputations (trend & period)

Η μέθοδος αυτή εφαρμόζει μία παλινδρόμηση στις παρατηρήσεις, τους δείκτες εποχιακότητας και τη τάση. Με τον τρόπο αυτό συμπληρώνει τυχόν κενές τιμές λαμβάνοντας υπόψη το εποχιακό στοιχείο και την τάση των δεδομένων.

Bootstrap EM Algorithm

Η μέθοδος αυτή παρέχεται από το πακέτο της R, AmeliaII και υπολογίζει τις κενές τιμές χρησιμοποιώντας μια νέα προσέγγιση bootstrapping τεχνικής, τον αλγόριθμο EMB (μεγιστοποίηση προσδοκίας με bootstrapping). Ο αλγόριθμος χρησιμοποιεί το γνωστό EM (expectation-maximization) αλγόριθμο σε πολλαπλά bootstrapped δείγματα των αρχικών ελλιπών δεδομένων για την παραγωγή των τιμών των παραμέτρων για τα δεδομένα. Ο αλγόριθμος στη συνέχεια αντλεί τις imputed τιμές από κάθε σειρά bootstrapped παραμέτρων, και αντικαθιστά τις τιμές που λείπουν. Κάποια από τα πλεονεκτήματα αυτής της μεθόδου είναι ότι είναι πολύ γρήγορη και σύμφωνα με την επίσημη τεκμηρίωση της δεν διακόπτεται από σφάλματα.

Απαλοιφή Ακραίων Τιμών

Οι ανωμαλίες στις χρονοσειρές μπορούν να χαλάσουν τα μοτίβα τους και να εισαγάγουν σημαντικά σφάλματα στην απόδοση πρόβλεψης. Αυτό οφείλεται κυρίως στη μεταφερόμενη επίδραση της απόκλισης στην πρόβλεψη και στη μεροληψία στον υπολογισμό των παραμέτρων του μοντέλου. Με τον όρο ακραίες τιμές μπορεί να εννοούμε αλλαγές επιπέδου στις χρονοσειρές, ή ασυνήθιστες τιμές που μπορεί να επηρεάσουν το μοντέλο πρόβλεψης ενώ δεν πρέπει. Θα πρέπει να αφαιρεθούν, λοιπόν, πριν προχωρήσουμε στη διαδικασία πρόβλεψης. Παρακάτω αναφέρονται οι ενσωματωμένες μέθοδοι του OMEN που αφαιρούν τις ακραίες τιμές από τα δεδομένα.

Μέθοδος Chen & Liu

Μπορούν να ληφθούν υπόψη τρεις τύποι ακραίων τιμών. Συγκεκριμένα οι επιλογές είναι οι Αθροιστικές ακραίες τιμές "AO", μετατοπίσεις επιπέδου "LS" και οι προσωρινές αλλαγές "TC". Η μεθοδολογία που περιγράφεται από τους Chen & Liu (1993) ακολουθείται για να εντοπιστούν οι τιμές. Το αρχικό πλαίσιο βασίζεται σε μοντέλα χρονοσειρών ARIMA.

Ας προσδιορίσουμε ένα μοντέλο ARIMA για τη σειρά y^*_t με m outliers είναι $L_j(B)$ με βάρη w_j :

$$y^*_t = \sum_{j=1}^m \omega_j L_j(B) I_t(t_j) + \frac{\theta(B)}{\phi(B)\alpha(B)} a_t$$

όπου $I_t(t_j)$ είναι μεταβλητή δείκτη που περιέχει την τιμή 1 στην παρατήρηση t_j όπου το j -στο outlier βρίσκεται, το $\phi(B)$ είναι ένα αυτορρυθμιζόμενο πολυώνυμο με όλες τις ρίζες εκτός του κύκλου της μονάδας, το $\theta(B)$ ένα πολυώνυμο κινητού μέσου όρου με όλες τις ρίζες έξω από τον κύκλο της μονάδας και το $\alpha(B)$ είναι ένα αυτορρυθμιζόμενο πολυώνυμο με όλες τις ρίζες στον κύκλο της μονάδας.

Η παρουσία των ακραίων τιμών ελέγχεται μέσω t -στατιστικών που εφαρμόζονται στην ακόλουθη εξίσωση παλινδρόμησης:

$$\pi(B)y^*_t \equiv \hat{e}_t = \sum_{j=1}^m \omega_j \pi(B) L_j(B) I_t(t_j) + a_t$$

Οι regressors της παραπάνω εξίσωσης δημιουργούνται από τη μέθοδο outliers.regressors.arima

Ορισμένοι κανόνες στίλβωσης εφαρμόζονται από το locate.outliers:

- Εάν οι μετατοπίσεις στάθμης βρίσκονται σε διαδοχικά χρονικά σημεία, τότε μόνο το σημείο με υψηλότερη τιμή t κατά απόλυτη τιμή διατηρείται.
- Εάν περισσότεροι από ένας τύποι απόκλισης υπερβαίνουν το κατώφλι `eval` σε ένα δεδομένο χρονικό σημείο, ο τύπος outlier με υψηλότερη τιμή t κατά απόλυτη τιμή διατηρείται και τα υπόλοιπα αφαιρούνται.

Μέθοδος Προσαρμογής Μοντέλου Πρόβλεψης & Εντοπισμού Ασυνήθιστων Τιμών

Κατά τη μέθοδο αυτή αρχικά γίνεται διαγωνισμός εποχιακότητας για το μοντέλο για να βρεθεί κατά πόσο επηρεάζει το εποχιακό στοιχείο τα δεδομένα. Στη συνέχεια εφαρμόζεται το μοντέλο ets με παραμέτρους "ZZN" ή "ZZZ", που σημαίνει ότι οι τύποι σφάλματος και τάσης επιλέγονται αυτόματα,

ενώ η εποχιακότητα επιλέγεται αυτόματα μόνο στην περίπτωση που το τεστ υποδεικνύει ότι υπάρχει. Το μοντέλο θα αναλυθεί παρακάτω αναλυτικότερα. Τέλος, η μέθοδος εξετάζει για κάθε τιμή αν απέχει πολύ από την τιμή στην αντίστοιχη θέση του μοντέλου πρόβλεψης που έχει επιλεγθεί και προσαρμόσται στα δεδομένα μέσω της συνθήκης:

$$(R1[k] \geq (1.1 - ta/100)) \text{ ή } (R2[k] \leq 0.75 + ta/100) \text{ ή } (R1[k] \leq (0.9 + ta/100)) \& (R2[k] \geq (1.25 - tb/100))$$

όπου R1 : Η αποεποχικοποιημένη χρονοσειρά προς την τάση και
R2: Η αποεποχικοποιημένη χρονοσειρά προς το μοντέλο πρόβλεψης

Αν κάτι τέτοιο ισχύει θεωρεί ότι εντοπίστηκε κάποια ακραία τιμή την οποία διορθώνει υπολογίζοντας τη γραμμική παρεμβολή μεταξύ του προηγούμενου και του επόμενου σημείου. Αυτό γίνεται με την συνάρτηση `na.approx` του πακέτου `R, zoo`.

Μέθοδος Εξέτασης της Κατανομής των Σφαλμάτων Πρόβλεψης

Κατά τη μέθοδο αυτή αρχικά γίνεται διαγωνισμός εποχιακότητας για το μοντέλο, και όπως και στην προηγούμενη μέθοδο εφαρμόζεται το μοντέλο `ets` με παραμέτρους “ZZZ” ή “ZZN” ανάλογα με την ύπαρξη ή μη εποχιακότητας. Στη συνέχεια υπολογίζεται η μέση τιμή τόσο των αποεποχικοποιημένων τιμών του `insample` όσο και των τιμών του μοντέλου πρόβλεψης. Για να εντοπιστεί αν μία τιμή αποτελεί outlier ή όχι, η μέθοδος αυτή συγκρίνει για κάθε παρατήρηση την παρατήρηση με τη μέση τιμή της αποεποχικοποιημένης χρονοσειράς συν την τυπική απόκλιση της χρονοσειράς του μοντέλου πρόβλεψης πολλαπλασιασμένη επί ένα βάρος.

$$D[k] \geq (\underline{D} + (3 - ta) * std(f)) \text{ ή } D[k] \leq (\underline{D} - (3 - ta) * std(f))$$

Αν η παραπάνω συνθήκη ισχύει θεωρείται ότι εντοπίστηκε κάποια ακραία τιμή την οποία η μέθοδος θα συμπληρώσει με γραμμική παρεμβολή μεταξύ του προηγούμενου και του επόμενου σημείου μέσω της συνάρτησης `na.approx` του πακέτου `R, zoo`.

Μέθοδος Εντοπισμού Ακραίων Τιμών

Κατά τη μέθοδο αυτή αρχικά γίνεται διαγωνισμός εποχιακότητας για το μοντέλο, και όπως και στην προηγούμενη μέθοδο εφαρμόζεται το μοντέλο `ets` με παραμέτρους “ZZZ” ή “ZZN” ανάλογα με την ύπαρξη ή μη εποχιακότητας. Αρχικά υπολογίζεται το Ratio ως ο κινητός μέσος όρος 7 των [μοντέλο πρόβλεψης, αποεποχικοποιημένο μοντέλο, αντίστροφο μοντέλο πρόβλεψης] προς τον κινητό μέσο όρο 5 του ίδιου σετ δεδομένων. Στη συνέχεια εντοπίζεται η θέση των ακραίων τιμών μέσω της συνθήκης:

$$(Ratio[k] \geq (1.05 - ta/100)) \text{ | } (Ratio[k] \leq (0.95 + ta/100))$$

Τέλος η μέθοδος συμπληρώνει και πάλι τις τιμές που αφαιρέθηκαν ως outliers με γραμμική παρεμβολή μεταξύ του προηγούμενου και του επόμενου σημείου μέσω της συνάρτησης `na.approx` του πακέτου `R, zoo`.

Στις παραπάνω συναρτήσεις οι τιμές των παραμέτρων `ta`, `tb` ρυθμίζονται κάθε φορά από τον χρήστη, ενώ η αρχική τιμή τους είναι `ta=tb=0`.

Αποσύνθεση

Το OMEN μπορεί να χρησιμοποιηθεί για να αποσυνθέσει τη χρονοσειρά και να παράσχει μια ακλόνητη εικόνα των συστατικών του τα οποία, όπως ορίζονται από την κλασσική αποσύνθεση, διακρίνονται στην τάση και την εποχική, την κυκλική και την τυχαία συνιστώσα. Η αποσύνθεση μπορεί να ενισχύσει τις επιδόσεις των προβλέψεων, καθώς σε περιπτώσεις σημαντικής εποχικότητας και υιοθέτησης μη εποχιακών μοντέλων πρόβλεψης, η αφαίρεση και η πρόσθεση του εποχιακού στοιχείου των χρονοσειρών πριν και μετά την πρόβλεψη μπορούν να αποδειχθούν πολύ ωφέλιμες. Από την άποψη αυτή, η κλασσική πολλαπλασιαστική αποσύνθεση με κινητούς μέσους όρους εφαρμόζεται και οι εποχιακοί δείκτες υπολογίζονται για την επιλεγμένη συχνότητα.

Κλασσική Μέθοδος Αποσύνθεσης

Η κλασσική μέθοδος αποσύνθεσης αποτελεί την πιο διαδεδομένη μέθοδο την πιο απλή διαδικασία για την απομόνωση των 4 συνιστωσών της χρονοσειράς. Πιο συγκεκριμένα, το μοντέλο της πολλαπλασιαστικής αποσύνθεσης το οποίο υπολογίζει το σύστημα έχει την εξής μορφή:

$$Y_t = S_t * T_t * C_t * R_t$$

Για τον υπολογισμό των επιμέρους συνιστωσών ακολουθούνται τα επόμενα βήματα:

- Υπολογισμός ενός κινητού μέσου όρου ο οποίος βασίζεται στο μήκος της εποχικότητας Αυτό σημαίνει ότι αν πρόκειται για μηνιαία δεδομένα τότε πρέπει να υπολογιστεί κεντρικός κινητός μέσος όρος μήκους 12, ενώ αν πρόκειται για τριμηνιαία δεδομένα, τότε πρέπει να υπολογιστεί κινητός μέσος όρος μήκους 4. Με αυτόν τον τρόπο ο υπολογιζόμενος μέσος όρος δεν περιέχει το εποχιακό στοιχείο ενώ περιέχει πολύ μικρή ή και καθόλου τυχαιότητα, δεδομένου ότι η τυχαιότητα αντιπροσωπεύεται από τυχαίες διακυμάνσεις που κυμαίνονται γύρω από το μέσο όρο των παρατηρήσεων. Οι τιμές των μέσων όρων που υπολογίζονται με αυτόν τον τρόπο είναι εξομαλυμένες και δίνουν μία καλή εκτίμηση της συμπεριφοράς της χρονοσειράς όσον αφορά την τάση και την κυκλικότητα. Οπότε μπορούμε να θεωρήσουμε ότι:

$$KMO(n) = T * C$$

Όπου $KMO(n)$ είναι ένας κινητός μέσος όρος μήκους n και T, C είναι οι συνιστώσες τάσης κύκλου αντίστοιχα. Στην περίπτωση που τα δεδομένα έχουν άρτιου μήκους εποχικότητα τότε χρησιμοποιείται ο κεντρικός κινητός μέσος όρος.

- Διαίρεση των πραγματικών δεδομένων με τις αντίστοιχες τιμές των κινητών μέσων όρων. Με τον τρόπο αυτό προκύπτουν οι λόγοι εποχικότητας, οι οποίοι όμως περιέχουν τυχαιότητα.

$$\frac{Y}{KMO(n)} = \frac{S * T * C * R}{T * C} = S * R$$

- Απαλοιφή τυχαιότητας από τους λόγους εποχικότητας του προηγούμενου βήματος. Η διαδικασία αυτή επιτυγχάνεται με εύρεση της μέσης τιμής των αντίστοιχων λόγων εποχικότητας, δηλαδή των λόγων που αναφέρονται σε αντίστοιχες περιόδους. Οι υπολογιζόμενοι μέσοι όροι αποτελούν τους δείκτες εποχικότητας της χρονοσειράς. Σε ορισμένες περιπτώσεις, οι δείκτες εποχικότητας χρίζουν κανονικοποίησης έτσι ώστε το άθροισμά τους να ισούται με το μήκος της εποχικότητας. Αν η χρονοσειρά περιέχει αρκετή τυχαιότητα, τότε προτείνεται ο υπολογισμός ενδιάμεσων μέσων όρων, προκειμένου να προκύψουν δείκτες εποχικότητας. Η διαφοροποίηση έγκειται στη μη χρήση του ελάχιστου και του μέγιστου της εποχικότητας στον υπολογισμό κάθε δείκτη εποχικότητας.
- Διαίρεση των πραγματικών τιμών της χρονοσειράς με τους αντίστοιχες δείκτες της εποχικότητας αυτής για την εύρεση της αποεποχικοποιημένης χρονοσειράς. Η χρονοσειρά που προκύπτει περιέχει μόνο τάση, κύκλο και τυχαιότητα.

$$\frac{Y}{S} = \frac{S * T * C * R}{S} = T * C * R$$

- Απαλοιφή τυχαιότητας από την εποχικοποιημένη χρονοσειρά Αυτό επιτυγχάνεται, υπολογίζοντας το κινητό μέσο όρο μήκους 3 ή 5 παρατηρήσεων την αποεποχικοποιημένης χρονοσειράς. Η σειρά των μέσων όρων που προκύπτει, αποτελεί μία αρκετά ομαλή και ακριβή σειρά της τάσης κύκλου. Για βέλτιστη εξομάλυνση και απαλοιφή της τυχαιότητας συνίσταται η χρήση κινητού μέσου όρου 3x3. Αν το ζητούμενο είναι ο υπολογισμός της τυχαιότητας, όπως στην περίπτωση της απεικόνισης των ποιοτικών χαρακτηριστικών της χρονοσειράς, που γίνεται στην εφαρμογή, τότε αυτή μπορεί να προκύψει από της διαίρεση της αποεποχικοποιημένης χρονοσειράς και της τάσης κύκλου:

$$\frac{T * C * R}{KMO(3 \times 3)} = \frac{T * C * R}{T * C} = R$$

- Υπολογισμός της τάσης από της χρονοσειρά τάσης – κύκλου του προηγούμενου βήματος. Στο στάδιο αυτό, η σειρά που προκύπτει έχει μόνο τάση και κυκλικότητα. Σε περιπτώσεις που είναι επιθυμητός ο διαχωρισμός των δύο συνιστωσών είτε για την απεικόνισή τους, όπως στην περίπτωση της εφαρμογής, είτε για πρόβλεψη, θα πρέπει να επιλεγθεί το μοντέλο τάσης που περιγράφει καλύτερα τη χρονοσειρά. Αν θεωρηθεί γραμμική τάση, τότε ο υπολογισμός της συνιστώσας τάσης επιτυγχάνεται από εφαρμογή της απλής γραμμικής παλινδρόμησης, της οποίας οι εξισώσεις περιγράφονται και αναλύονται και στη συνέχεια.

$$T = a + bt, \text{ Όπου } a = \underline{TC} - b\underline{t},$$

$$b = \frac{\frac{\sum_{i=1}^n (t_i * TC_i)}{n} - \underline{t} * \underline{TC}}{\frac{\sum_{i=1}^n (t_i)^2}{n} - \underline{t}^2}, \quad \underline{TC} = \frac{\sum_{i=1}^n (t_i)}{n}, \quad \underline{t} = \frac{\sum_{i=1}^n (t_i)}{n}$$

Το T δηλώνει την τάση της χρονοσειράς, το TC δηλώνει τις τιμές της σειράς τάσης κύκλου, το t δηλώνει τον αύξοντα αριθμό της χρονικής περιόδου. Οι συντελεστές α και β, δηλώνουν το αρχικό επίπεδο και την τάση αντίστοιχα. Αφού υπολογιστεί η χρονοσειρά της τάσης τότε μπορεί να ορισθεί και η συνιστώσα του κύκλου ως εξής:

$$C = \frac{T * C}{C}$$

- Έχοντας αναλύσει την αρχική χρονοσειρά στις συνιστώσες της, είναι εφικτό να συντεθούν πολλαπλασιαστικά έτσι ώστε να εκτιμηθούν μελλοντικές τιμές της χρονοσειράς. Η διαδικασία λέγεται πρόβλεψη μέσω αποσύνθεσης.

$$F_{n+1} = T_i * S_i * C_i$$

Ωστόσο, αυτή η μέθοδος δεν έχει υλοποιηθεί στα πλαίσια της παρούσας πτυχιακής εργασίας, δεδομένης της αυξημένης χρήσης των μεθόδων εκθετικής εξομάλυνσης και την ενσωμάτωση αυτών. Οι εν λόγω μέθοδοι θα περιγραφούν στην συνέχεια.

Εξομάλυνση

Μέσω της εξομάλυνσης, το συστατικό του θορύβου θα συρρικνωθεί ή θα εξαλειφθεί τελείως, υπογραμμίζοντας έτσι τα χρήσιμα χαρακτηριστικά των χρονοσειρών (επίπεδο και τάση) και οδηγώντας σε δυνητικά καλύτερες προβλέψεις. Η πρώτη από τις παρεχόμενες μεθόδους υπολογίζει έναν απλό κινητό μέσο όρο. Η δεύτερη επιλογή του χρήστη είναι ένας μη γραμμικός μηχανισμός που βασίζεται στον μετασχηματισμό θεμάτων για τον προσδιορισμό των μακροπρόθεσμων τάσεων που εισήγαγε ο Ασημακόπουλος, σύμφωνα με την οποία εντοπίζονται ισχυρές τοπικές διακυμάνσεις και σταδιακά περιορίζονται με βάση τις τιμές των γειτονικών παρατηρήσεων

Κινητός Μέσος Όρος

Ένας κινητός μέσος όρος χρησιμοποιείται συνήθως για την εξομάλυνση των βραχυπρόθεσμων διακυμάνσεων και την επισήμανση μακροπρόθεσμων τάσεων ή κύκλων. Στη συγκεκριμένη περίπτωση διατίθεται ως επιλογή ένας απλός κινητός μέσος όρος (SMA) που είναι ο μη σταθμισμένος μέσος όρος των προηγούμενων n στοιχείων. Ένα παράδειγμα ενός απλού ισοδύναμου σταθμισμένου μέσου όρου για ένα δείγμα n παρατηρήσεων είναι η μέση τιμή των τιμών των προηγούμενων n παρατηρήσεων. Εάν αυτές οι τιμές είναι $p_M, p_{M-1}, \dots, p_{M-(n-1)}$ τότε ο τύπος είναι:

$$\begin{aligned}\bar{p}_{SM} &= \frac{p_M + p_{M-1} + \dots + p_{M-(n-1)}}{n} \\ &= \frac{1}{n} \sum_{i=0}^{n-1} p_{M-i}\end{aligned}$$

Μη Γραμμική Εξομάλυνση

Η μέθοδος αυτή έχει ως σκοπό την εξομάλυνση των δεδομένων με τη βοήθεια της παραμέτρου meanmetric. Πιο συγκεκριμένα, το meanmetric υπολογίζεται ως η μέση τιμή της χρονοσειράς που προκύπτει από τη συνθήκη:

$$Difs(i) = 300 * \left| \frac{B(i-1) - 2 * B(i) + B(i+1)}{B(i-1) + B(i) + B(i+1)} \right|$$

Όπου B : η χρονοσειρά που προκύπτει από backcasting της αρχικής. Στη συνέχεια συγκρίνεται ο δείκτης αυτός με το όριο που έχουμε θέσει για να εξετάσουμε αν έχουμε φτάσει το επιθυμητό σημείο ομαλότητας. Εάν κάτι τέτοιο ισχύει τότε σημαίνει ότι η αρχική χρονοσειρά ήταν αρκετά ομαλή και δεν χρειάζεται να επεξεργαστεί περαιτέρω. Στην αντίθετη περίπτωση, υπολογίζονται οι δείκτες alpha και beta ως εξής:

$$\begin{aligned}beta &= B(index): index = index \text{ of } \max(Difs) \\ alpha &= \frac{\frac{B(index+1) + B(index-1)}{2} - beta}{100}\end{aligned}$$

Τελικά, επεξεργασόμαστε τη χρονοσειρά B αντικαθιστώντας το $B(index)$ κάθε φορά με το $beta + 50 * alpha$ και υπολογίζουμε εκ νέου τη χρονοσειρά $Difs$, όσο η συνθήκη meanmetric δείχνει ότι δεν έχουμε φτάσει την επιθυμητή ομαλότητα.

Τεχνικές προβλέψεων

Μέθοδοι Συνεχούς Ζήτησης

Ανάλογα με τις επιλογές που έγιναν σε όλη τη διαδικασία πρόβλεψης, η χρονοσειρά μπορεί να είναι είτε η αρχική, είτε μία προσαρμοσμένη μετά την πραγματοποίηση των διαδικασιών που προηγήθηκαν όπως της αποσύνθεσης και της εξομάλυνσης. Το σύστημα περιλαμβάνει τόσο παραδοσιακές όσο και σύγχρονες μεθόδους ως εξής:

Naive

Η συγκεκριμένη μέθοδος δίνει ως πρόβλεψη για την επόμενη χρονική περίοδο την ίδια τιμή με την παρατήρηση που είχε σημειωθεί την προηγούμενη ακριβώς χρονική περίοδο. Έχει καλή απόδοση για προβλέψεις ενός βήματος σε αποεποχικοποιημένες χρονοσειρές καθώς η αναμενόμενη τιμή της πρόβλεψης δεν διαφέρει σημαντικά από την τελευταία παρατήρηση που είναι διαθέσιμη. Η μαθηματική σχέση που περιγράφει αυτή τη μέθοδο πρόβλεψης είναι:

$$X_t(m) = X_t$$

SES

Η SES ή απλή εκθετική εξομάλυνση, αποτελεί μια μέθοδο πρόβλεψης που ενδείκνυται για προβλέψεις ενός βήματος ή σε θορυβώδεις και με μεγάλη τυχαιότητα χρονοσειρές. Περιγράφεται από τις ακόλουθες εξισώσεις:

$$\begin{aligned} e_t &= Y_t - F_t \\ S_t &= S_{t-1} + a e_t \\ F_{t+1} &= S_t \end{aligned}$$

Στις παραπάνω εξισώσεις e_t είναι το σφάλμα της περιόδου t , δηλαδή η διαφορά της πραγματικής τιμής από την πρόβλεψη, S_t το επίπεδο για την περίοδο t και F_t η πρόβλεψη για την περίοδο t . Η παράμετρος a ονομάζεται συντελεστής εξομάλυνσης και λαμβάνει τιμές εντός του διαστήματος $[0,1]$. Για την έναρξη της διαδικασίας υπολογισμού του μοντέλου απαιτείται ο ορισμός του αρχικού επιπέδου (S_0). Το αρχικό επίπεδο πρέπει να είναι αντιπροσωπευτικό των δεδομένων και επιλέγεται με βάση τα ποιοτικά χαρακτηριστικά της εκάστοτε χρονοσειράς. Επηρεάζει την επιλογή του συντελεστή εξομάλυνσης, καθώς διαφορετικές τιμές του αρχικού επιπέδου μπορεί να οδηγήσουν σε εντελώς διαφορετικούς συντελεστές εξομάλυνσης. Ως αρχικό επίπεδο μπορούν να χρησιμοποιηθούν: Ο μέσος όρος όλων των παρατηρήσεων, ο μέσος όρος ορισμένων αρχικών παρατηρήσεων της χρονοσειράς, η πρώτη παρατήρηση, ή το σταθερό επίπεδο από το μοντέλο της απλής γραμμικής παλινδρόμησης. Η επιλογή του βέλτιστου συντελεστή εξομάλυνσης είναι μια διαδικασία πολύ σημαντική για την παραγωγή ακριβέστερων προβλέψεων και καθορίζεται από δύο παράγοντες. Αρχικά, εξαρτάται από το ποσοστό θορύβου στη χρονοσειρά. Όσο περισσότερος είναι ο θόρυβος τόσο μικρότερος πρέπει να είναι ο συντελεστής εξομάλυνσης ώστε να αποφεύγεται η υπερβολική αντίδραση στο θόρυβο. Ο δεύτερος παράγοντας, έχει να κάνει με τη σταθερότητα του μέσου όρου της χρονοσειράς. Αν ο μέσος όρος μιας χρονοσειράς μεταβάλλεται, ο συντελεστής εξομάλυνσης θα πρέπει να είναι μεγάλος ώστε οι προβλέψεις να παρακολουθούν τις μεταβολές που παρουσιάζουν τα δεδομένα. Για τιμές του

συντελεστή εξομάλυνσης κοντά στη μονάδα, το σφάλμα της τελευταίας περιόδου συμβάλλει περισσότερο στον υπολογισμό της επόμενης πρόβλεψης. Για την ακραία περίπτωση $\alpha=1$ το μοντέλο ταυτίζεται με τη μέθοδο Naive, ενώ για $\alpha=0$ κάθε πρόβλεψη ισούται με το αρχικό επίπεδο. Γενικά μικρές τιμές της παραμέτρου α έχουν ως αποτέλεσμα μεγαλύτερη εξομάλυνση, εφόσον το μοντέλο της πρόβλεψης βρίσκεται κοντά στο αρχικό επίπεδο και αργεί να ακολουθήσει μεγάλες μεταβολές των ιστορικών δεδομένων. Αντίθετα για μεγάλες τιμές του α το μοντέλο πρόβλεψης ακολουθεί γρηγορότερα τη χρονοσειρά. Η επιλογή της παραμέτρου α , συνεπώς, μπορεί να γίνει είτε προσεγγιστικά, είτε με μεγαλύτερη ακρίβεια με τη χρήση υπολογιστικών εργαλείων. Στην περίπτωση μας, η παράμετρος α υπολογίζεται αυτόματα μέσω της ελαχιστοποίησης του σφάλματος MSE.

ETS

Η ETS (Hyndman, 2008) είναι ένα State Space Model (SSM), δηλαδή αναφέρεται σε μια κατηγορία πιθανολογικών γραφικών μοντέλων, και περιγράφει την πιθανολογική εξάρτηση μεταξύ της μεταβλητής λανθάνουσας κατάστασης και την παρατηρούμενη μέτρηση. Ουσιαστικά η μέθοδος ETS διαθέτει τρία βασικά στοιχεία για την εξαγωγή προβλέψεων, το είδος των σφαλμάτων, το είδος της τάσης και το είδος της εποχιακότητας (*Error, Trend, Seasonal*). Το στοιχείο των σφαλμάτων και το στοιχείο της εποχιακότητας μπορεί να είναι είτε προσθετικά (Additive-A), είτε πολλαπλασιαστικά (Multiplicative-M). Επίσης το στοιχείο της εποχιακότητας μπορεί να μην λαμβάνεται υπ' όψιν (None-N). Αν ληφθεί υπόψη και το στοιχείο της τάσης, τότε μπορεί να είναι ένα από τα παρακάτω είδη: Δεν λαμβάνεται υπόψη το στοιχείο της τάσης-None (N) Προσθετική - Additive (A) Προσθετική φθίνουσα τάση - Additive damped (Ad) Πολλαπλασιαστική τάση - Multiplicative (M) Πολλαπλασιαστική φθίνουσα τάση - Multiplicative Damped (Md) Επομένως οι περιπτώσεις ανάλογα με αυτές τις συνιστώσες χωρίζονται αρχικά με βάση το είδος των σφαλμάτων, δηλαδή σε προσθετικά ή πολλαπλασιαστικά. Για την απλή γραμμική παλινδρόμηση ισχύουν οι τύποι:

$$et = Yt - Ft \quad St = St-1 + a \cdot et \quad Ft+1 = St$$

Όπου επιλύοντας τις παραπάνω εξισώσεις η εξίσωση της πραγματικής τιμής γράφεται ως εξής:

$$Yt = St-1 + et$$

Για να γίνει αυτό σε ένα state model, το μόνο που χρειάζεται είναι να καθοριστεί η κατανομή πιθανοτήτων για et . Στο μοντέλο με προσθετικά σφάλματα, υποθέτουμε ότι τα σφάλματα πρόβλεψης ενός βήματος (one-step forecast errors) et διανέμονται κανονικά με μέση τιμή 0 και διακύμανση σ^2 . Μία συντομογραφία για αυτό είναι το $et = et \sim NID(0, \sigma^2)$, όπου NID σημαίνει πως είναι κατανομημένα με κανονικό και ανεξάρτητο τρόπο (Normally and Independently Distributed). Επομένως τώρα οι εξισώσεις γράφονται:

$$Yt = St-1 + et \quad (1)$$

$$St = St-1 + a \cdot et \quad (2)$$

Η πρώτη εξίσωση αναφέρεται ως εξίσωση μέτρησης και η δεύτερη ως εξίσωση κατάστασης ή μετάβασης. Αυτές οι δύο εξισώσεις, μαζί με την στατιστική κατανομή των σφαλμάτων, σχηματίζουν ένα πλήρως καθορισμένο στατιστικό μοντέλο. Η εξίσωση μέτρησης δείχνει τη σχέση μεταξύ των παρατηρήσεων και των μη παρατηρούμενων καταστάσεων. Σε αυτή την περίπτωση η παρατήρηση Yt είναι μια γραμμική συνάρτηση του επιπέδου St και το τυχαίο σφάλμα et , το απρόβλεπτο μέρος του

Y_t . Για τα άλλα μοντέλα, αυτή η σχέση μπορεί να είναι μη γραμμική. Η εξίσωση της μετάβασης δείχνει την εξέλιξη της κατάστασης μέσα στο χρόνο. Όσο υψηλότερη είναι η τιμή του α , τόσο πιο γρήγορες θα είναι οι μεταβολές του επιπέδου.

Με τον ίδιο τρόπο με πριν μπορούμε να εξάγουμε το μοντέλο με χρήση πολλαπλασιαστικών σφαλμάτων. Αυτό γίνεται γράφοντας τα τυχαία σφάλματα σαν σχετικά σφάλματα:

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

και αντικαθιστώντας $F_t = S_{t-1}$ μας δίνει ότι:

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

Τότε μπορούμε να γράψουμε την πολλαπλασιαστική μορφή του μοντέλου ως εξής:

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

Τότε μπορούμε να γράψουμε την πολλαπλασιαστική μορφή του μοντέλου ως εξής:

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

Προσδιορισμός παραμέτρων

Μια εναλλακτική λύση για τον υπολογισμό των παραμέτρων, εκτός από την ελαχιστοποίηση του αθροίσματος των τετραγώνων των σφαλμάτων, είναι να μεγιστοποιήσουμε την «πιθανότητα». Με τον όρο αυτό εννοούμε συγκεκριμένα την πιθανότητα των στοιχείων που προκύπτουν από το καθορισμένο μοντέλο. Έτσι, μία μεγάλη πιθανότητα συνδέεται με ένα καλό μοντέλο. Οι παράμετροι εξομάλυνσης α , β , γ και ϕ , και οι αρχικές καταστάσεις S_t και T_t υπολογίζονται με βάση τη μεγιστοποίηση της πιθανότητας. Οι παράμετροι εξομάλυνσης παίρνουν τιμές στο διάστημα $[0,1]$. Η επιλογή του κατάλληλου μοντέλου γίνεται επιλέγοντας μέσα από τρία κριτήρια.

- Το πρώτο κριτήριο είναι το κριτήριο πληροφορίας του Akaike (Akaike's Information Criteria) και δίνεται από τον τύπο: $AIC = -2 \log(L) + 2 \cdot k$ Όπου το L είναι η πιθανότητα του μοντέλου και k ο συνολικός αριθμός των παραμέτρων και των αρχικών καταστάσεων που θα πρέπει να εκτιμηθούν.
- Το δεύτερο είναι η διορθωμένη μορφή του προηγούμενου κριτηρίου και υπολογίζεται ως εξής:

$$AIC_c = AIC + 2(k+1)(k+2) / T - k$$

Το τρίτο είναι το κριτήριο πληροφορίας του Bayes (Bayes Information Criterion):

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

Holt

Το μοντέλο γραμμικής τάσης Holt, είναι η επέκταση της απλής εκθετικής εξομάλυνσης και περιγράφεται από τις εξισώσεις:

$$\begin{aligned}e_t &= Y_t - F_t \\S_t &= S_{t-1} + T_{t-1} + ae_t \\T_t &= t_{t-1} + be_t \\F_{t+m} &= S_{t+m}T_t\end{aligned}$$

Στις παραπάνω εξισώσεις e_t είναι το σφάλμα της περιόδου t , S_t το επίπεδο για την περίοδο t , T_t η τάση για την περίοδο t και F_t η πρόβλεψη για την περίοδο t . Η παράμετρος a είναι ο συντελεστής εξομάλυνσης του επιπέδου, ενώ η παράμετρος b ο συντελεστής εξομάλυνσης της τάσης και λαμβάνουν τιμές εντός του διαστήματος $[0,1]$. Με m συμβολίζεται ο χρονικός ορίζοντας της πρόβλεψης. Όπως και στην περίπτωση της μεθόδου SES, έτσι και σε αυτή την περίπτωση η επιλογή του βέλτιστου συνδυασμού τιμών για τις παραμέτρους a και b βασίζεται στην ελαχιστοποίηση του μέσου τετραγωνικού σφάλματος (MSE). Συνήθως η βέλτιστη τιμή του συντελεστή β για την τάση είναι μικρότερη από την τιμή του συντελεστή α για το επίπεδο, κάτι που συμβαίνει διότι η τιμή της τάσης είναι μικρότερη από εκείνη του επιπέδου για κάθε περίοδο. Για την έναρξη της μεθόδου απαραίτητα είναι το αρχικό επίπεδο (S_0) και η αρχική τάση (T_0), η επιλογή των οποίων είναι ιδιαίτερα σημαντική για την ακρίβεια των παραγόμενων προβλέψεων. Το αρχικό επίπεδο υπολογίζεται όπως και στην απλή εκθετική εξομάλυνση. Ως αρχική τάση μπορούν να χρησιμοποιηθούν η διαφορά της δεύτερης από την πρώτη παρατήρηση ($Y_2 - Y_1$), η διαφορά της v -οστής παρατήρησης από την πρώτη διαιρεμένη με $v-1$ ή η σταθερά της κλίσης από το μοντέλο απλής γραμμικής εξομάλυνσης.

Damped

Το μοντέλο μη γραμμικής τάσης (Gardner και McKenzie 1985) αποτελεί μια προσαρμογή του μοντέλου γραμμικής τάσης που χρησιμοποιείται σε περιπτώσεις μη γραμμικών τάσεων. Αυτό επιτυγχάνεται με την προσθήκη της παραμέτρου διόρθωσης της τάσης (trendmodification parameter) ϕ , η οποία ελέγχει το ρυθμό αύξησης των τιμών της τάσης σε μια χρονοσειρά. Οι εξισώσεις που περιγράφουν το μοντέλο μη γραμμικής τάσης είναι οι εξής:

$$\begin{aligned}e_t &= Y_t - F_t \\S_t &= S_{t-1} + T_{t-1} + ae_t \\T_t &= t_{t-1} + be_t \\F_{t+m} &= S_{t+m}T_t + \sum_{i=1}^m \phi^i T_t\end{aligned}$$

Όπως γίνεται φανερό οι εξισώσεις είναι ίδιες με εκείνες της μεθόδου Holt, πλην της τελευταίας, όπου αντί για τον υπολογισμό μιας γραμμικής αύξησης της τάσης μέσω του συντελεστή m , γίνεται ένας μη γραμμικός υπολογισμός της, με τη χρήση της παραμέτρου εξομάλυνσης ϕ . Η παράμετρος ϕ (σε αντίθεση με τις παραμέτρους α και β) μπορεί να λάβει και τιμές μεγαλύτερες της μονάδας. Στη περίπτωση που $0 < \phi < 1$ προκύπτει το μοντέλο της φθίνουσας τάσης (damped exponential smoothing).

Το μοντέλο αυτό χρησιμοποιείται κυρίως στην παραγωγή μεσοπρόθεσμων προβλέψεων, καθώς χαρακτηρίζεται από έλλειψη τάσης για υπεραισιοδοξία.

Theta Classic

Η μέθοδος Theta (Assimakopoulos και Nikolopoulos, 2000) είναι μία μέθοδος πρόβλεψης, η οποία βασίζεται στην μεταβολή των τοπικών καμπυλοτήτων μιας χρονοσειράς μέσα από την παράμετρο θ που εφαρμόζεται πολλαπλασιαστικά στις διαφορές δεύτερης τάξης των δεδομένων. Η καινούργια χρονοσειρά που δημιουργείται διατηρεί την μέση τιμή και κλίση της αρχικής χρονοσειράς αλλά όχι και τις τοπικές καμπυλότητες και τη διακύμανση. Οι χρονοσειρές που παράγονται με αυτή τη διαδικασία ονομάζονται *theta lines*. Βασικό χαρακτηριστικό αυτών των γραμμών είναι η καλύτερη προσέγγιση της μακροπρόθεσμης συμπεριφοράς των δεδομένων ή ανάδειξη και τονισμός των βραχυπρόθεσμων χαρακτηριστικών, ανάλογα με την τιμή της παραμέτρου θ .

Η μέθοδος διαχωρίζει την αρχική χρονοσειρά σε δύο ή περισσότερες γραμμές *theta*. Η κάθε γραμμή *theta* προεκτείνεται στο μέλλον ξεχωριστά, με την ίδια ή και με διαφορετικές μεθόδους πρόβλεψης και οι παραγόμενες προβλέψεις συνδυάζονται για να προκύψει η τελική πρόβλεψη. Ο απλός συνδυασμός δύο γραμμών *theta*, για $\theta=0$ (ευθεία γραμμή) και για $\theta=2$ (διπλασιασμός των τοπικών καμπυλοτήτων) αποτελούν την κλασική *theta* που είναι αυτή που είναι ενσωματωμένη στο σύστημα. Τα βήματα που ουσιαστικά περιγράφουν τη μεθοδολογία της κλασικής μεθόδου *theta* είναι τα παρακάτω:

- Έλεγχος εποχιακότητας. Ελέγχεται η κάθε χρονοσειρά για στατιστικά σημαντική εποχιακή συμπεριφορά.
- Αποεποχικοποίηση. Μέσω της κλασικής μεθόδου πολλαπλασιαστικής αποσύνθεσης
- Αποσύνθεση. Η κάθε χρονοσειρά αποσυντίθεται σε γραμμές *theta*, την ευθεία γραμμικής παλινδρόμησης ($\theta=0$) και τη γραμμή *theta* με παράμετρο $\theta=2$.
- Πρόβλεψη. Η γραμμή *theta* με παράμετρο $\theta = 0$, που αναπαριστά την ευθεία γραμμικής παλινδρόμησης, προεκτείνεται με τον συνηθισμένο τρόπο, ενώ η δεύτερη γραμμή προεκτείνεται μέσω της απλής γραμμικής εξομάλυνσης.
- Συνδυασμός. Οι παραγόμενες προβλέψεις των δύο γραμμών *theta* συνδυάζονται με ίσα βάρη.
- Εποχικοποίηση. Οι τελικές προβλέψεις εποχικοποιούνται, χρησιμοποιώντας τους δείκτες εποχιακότητας που υπολογίστηκαν στο βήμα 1. Ακολούθως περιγράφεται ο υπολογισμός των δύο γραμμών *theta* που προτείνονται στην αναφερόμενη μεθοδολογία. Δεδομένου πως η *Theta Line(0)* ισοδυναμεί με την ευθεία ελαχίστων τετραγώνων (LRL) που περιγράφηκε στην προηγούμενη παράγραφο, απομένει ο υπολογισμός της *Theta Line(2)*. Έτσι προκύπτουν οι εξής σχέσεις: Η τελευταία εξίσωση οδηγεί σε έναν εναλλακτικό τρόπο παραγωγής της *Theta Line(2)* αφού η LRL μπορεί εύκολα να παραχθεί σύμφωνα με τη θεωρία της παλινδρόμησης.

ARIMA

Τα ολοκληρωμένα αυτοπαλινδρομικά μοντέλα κινητών μέσων όρων (ARIMA) είναι μια γενίκευση από τα αυτοπαλινδρομικά μοντέλα κινητών μέσων όρων (ARMA). Τα μοντέλα αυτά εφαρμόζονται σε ορισμένες περιπτώσεις όπου τα δεδομένα παρουσιάζουν ενδείξεις μη-στασιμότητας, όπου ένα πρώτο differencing step (που αντιστοιχεί στο "ολοκληρωμένο" μέρος του μοντέλου) μπορεί να εφαρμοστεί για την απομάκρυνση της μη-στασιμότητας. Το μοντέλο συνήθως αναφέρεται ως

ARIMA (p, d, q) όπου οι παράμετροι p, d, q είναι μη-αρνητικοί ακέραιοι, που αναφέρονται με τη σειρά τους στο αυτοπαλινδρομικό μέρος, ολοκληρωμένο μέρος, και μέρος κινητού μέσου όρου, αντίστοιχα. Τα μοντέλα ARIMA αποτελούν ένα σημαντικό μέρος της προσέγγισης Box-Jenkins. Όταν δύο από τους τρεις όρους είναι μηδενικοί, είναι σύνηθες να λείπουν τα "AR", "I" ή "MA" από το ακρωνύμιο που περιγράφουν το μοντέλο. Για παράδειγμα, όταν ARIMA (0,1,0) τότε ισχύει I (1), και όταν είναι ARIMA (0,0,1) ισχύει MA (1). Σε αντίθεση με τα ντετερμινιστικά μοντέλα, η χρήση των οποίων απαιτεί γνώση των παραγόντων από τις οποίες εξαρτάται το μέγεθος, η εφαρμογή των μοντέλων ARIMA βασίζεται στον υπολογισμό της πιθανότητας για την οποία η τιμή του μεγέθους βρίσκεται σε κάποιο διάστημα.

Η διαδικασία εφαρμογής των μεθόδων ARIMA, προϋποθέτει μια προεπεξεργασία δεδομένων, έτσι ώστε η χρονοσειρά που μελετάται να είναι στάσιμη, δηλαδή η μέση τιμή, η διακύμανση και η συνάρτηση αυτοσυσχετίσής της να είναι σταθερές στην πάροδο του χρόνου. Αυτό μπορεί να εξασφαλισθεί με τους παρακάτω τρόπους:

- **Μετασχηματισμός των δεδομένων:** Η εν λόγω διαδικασία περιορίζει την τυχαιότητα και τις ακραίες τιμές (ασυνέχειες) που τυχόν υπάρχουν, οδηγώντας έτσι σε μία νέα χρονοσειρά μικρότερης και σταθερότερης διακύμανσης. Έτσι, το επίπεδο της χρονοσειράς δεν μεταβάλλεται το ίδιο απότομα με πριν και μπορεί να υποτεθεί στασιμότητα. Η λογαρίθμηση της χρονοσειράς αποτελεί την απλούστερη και πλέον διαδεδομένη μορφή μετασχηματισμού καθώς οδηγεί σε ικανοποιητικά αποτελέσματα με αρκετά απλό τρόπο και μικρό υπολογιστικό κόστος. Άλλη επιλογή είναι η χρήση μετασχηματισμών δυνάμεων, όπως π.χ. οι μετασχηματισμοί Box-Cox.
- **Διαφόριση (differencing) της χρονοσειράς:** Η διαφόριση, ανάλογα με την μορφή που λαμβάνει, περιορίζει τις διακυμάνσεις επιπέδου αφαιρώντας τάση και εποχιακότητα. Έτσι, παράγεται μία νέα χρονοσειρά σταθερότερου επιπέδου και διακύμανσης. Στην ουσία, κατά τη διαφόριση μίας χρονοσειράς n παρατηρήσεων, δημιουργείται μία νέα με στοιχεία της τις διαφορές των παρατηρήσεων της πρώτης.
1η τάξη: $Y'_t = Y_t - Y_{t-1}$, 2η τάξη: $Y''_t = Y'_t - Y'_{t-1} = Y_t - 2Y_{t-1} - Y_{t-2}$ κ.ο.κ.
- **Εποχιακή διαφόριση σε περιπτώσεις χρονοσειρών έντονης εποχιακότητας:** Εδώ η χρονοσειρά που παράγεται είναι αποτέλεσμα της διαφοράς μεταξύ των παρατηρήσεων της αρχικής χρονοσειράς και εκείνων των προηγούμενων αντίστοιχων εποχιακών περιόδων.
1η τάξη: $Y'_t = Y_t - Y_{t-m}$, 2η τάξη: $Y''_t = Y'_t - Y'_{t-m} = Y_t - 2Y_{t-m} - Y_{t-2m}$ κ.ο.κ. όπου m η περίοδος εποχιακότητας.

Στην συνέχεια εκτιμούνται τα μοντέλα ARIMA ως εξής:

- **Μοντέλα αυτοπαλινδρόμησης - AR (p) :** Τα μοντέλα αυτοπαλινδρόμησης θεωρούν γραμμικές σχέσεις ανάμεσα στις παρατηρήσεις της ίδιας της χρονοσειράς και τις αξιοποιούν για την περιγραφή της και την παραγωγή προβλέψεων. Ένα τέτοιο μοντέλο αναφέρεται ως μοντέλο αυτοπαλινδρόμησης AR (p), όπου p η τάξη του, δηλαδή το πλήθος των παρελθοντικών παρατηρήσεων που αξιοποιούνται για την παραγωγή της πρόβλεψης. Αλγεβρικά αυτό αναπαρίσταται ως εξής:
$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p}$$
, όπου φ_i οι συντελεστές αυτοσυσχετίσης του μοντέλου για υστέρηση i και c μία σταθερά $\square = \square(1 - \varphi_1 - \varphi_2 - \dots - \varphi_p)$
- **Μοντέλα κινητού μέσου όρου - MA (q):** Τα μοντέλα κινητού μέσου όρου θεωρούν γραμμικές σχέσεις ανάμεσα στην παρατήρηση της χρονοσειράς που εξετάζεται και στα

σφάλματα που εμφάνισε το μοντέλο MA σε προηγούμενες περιόδους. Ένα τέτοιο μοντέλο γράφεται αλγεβρικά ως εξής:

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

όπου θ_i οι συντελεστές μερικής αυτοσυσχέτισης του μοντέλου MA για υστέρηση i . Στην ουσία η τιμή της παρατήρησης y_t εξαρτάται κατά παράγοντα θ_1 από το σφάλμα που παρήγαγε το μοντέλο την προηγούμενη περίοδο, κατά παράγοντα θ_2 από το σφάλμα που παρήγαγε το μοντέλο την προ-προηγούμενη περίοδο ... και κατά παράγοντα θ_q από το σφάλμα του μοντέλου q περιόδους πίσω. Υπολογίζεται ως γραμμικός συνδυασμός αυτών προσ αυξανοντάς την –προαιρετικά– κατά μία σταθερά c .

Τα μοντέλα AR και MA μπορούν να συνδυαστούν αποδοτικά για την ανάλυση και πρόβλεψη στάσιμων χρονοσειρών. Έτσι, εισάγοντας στην εξίσωση και τα μοντέλα διαφόρισης για τη διασφάλιση της στασιμότητας, προκύπτουν τα μοντέλα ARIMA(p,d,q), όπου p , d , q η τάξη του αντίστοιχου μοντέλου. Το συνολικό μοντέλο αναπαρίσταται με τη χρήση του τελεστή ολίσθησης B ως εξής:

$$(1 - \varphi_1 B - \varphi_2 B^2 - \dots - \varphi_p B^p)(1 - B)^d (1 - \theta_1 B - \theta_2 B^2 - \dots - \theta_q B^q) y_t = c + e_t$$

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

MAPA

Σύμφωνα με την μέθοδο αυτή, παράγονται πολλαπλές χρονοσειρές από τα αρχικά δεδομένα της χρονοσειράς, χρησιμοποιώντας χρονική συνάθροιση (temporal aggregation). Οι νέες χρονοσειρές που προκύπτουν εκφράζουν και τονίζουν διαφορετικά χαρακτηριστικά της αρχικής χρονοσειράς καθώς η χρονική συνάθροιση συμβάλλει στην επισήμανση ή στην μείωση της σημασίας σημάτων που προέρχονται από συνιστώσες διαφορετικών χρονοσειρών. Σε κάθε χρονοσειρά εφαρμόζεται η κατάλληλη εκθετική μέθοδος εξομάλυνσης και παράγονται προβλέψεις για τις συνιστώσες κάθε χρονοσειράς. Αυτή η διαδικασία μας επιτρέπει τον τελικό συνδυασμό των συνιστωσών των χρονοσειρών διαφορετικής χρονικής συνάθροισης και την παραγωγή των τελικών προβλέψεων. Με αυτήν την προσέγγιση, επιτυγχάνεται καλύτερη πρόβλεψη των συνιστωσών των διαφορετικών χρονοσειρών μέσω της χρονικής συνάθροισης, ενώ μειώνεται η σημασία επιλογής μοντέλου αφού έχουμε συνδυασμό των προβλέψεων. Στα χαμηλότερα επίπεδα συνάθροισης, όπου ασχολούμαστε με χρονοσειρές υψηλής συχνότητας, δίνεται έμφαση στις περιοδικές συνιστώσες της χρονοσειράς, όπως η εποχιακότητα. Καθώς αυξάνεται το επίπεδο συνάθροισης, τα σήματα υψηλής συχνότητας φιλτράρονται με αποτέλεσμα να μετριάζεται η σημασία τους ενώ δίνεται έμφαση στις συνιστώσες χαμηλής συχνότητας των δεδομένων όπως το επίπεδο και η τάση της χρονοσειράς. Συνεπώς, κάνοντας χρήση της παραπάνω διαδικασίας συνάθροισης, διευκολύνεται ο προσδιορισμός, η επιλογή και η παραμετροποίηση των προγνωστικών μοντέλων.

Το πλαίσιο παραγωγής προβλέψεων της μεθόδου μπορεί να διαχωριστεί στα εξής τρία στάδια: (α) το στάδιο της συνάθροισης (β) το στάδιο της πρόβλεψης (γ) το στάδιο του συνδυασμού.

- Στάδιο Συνάθροισης: Έστω Y_t μια χρονοσειρά με παρατηρήσεις y_t (όπου $t=1, \dots, n$) που λαμβάνονται με μια δεδομένη συχνότητα. Μπορούμε να συναθροίσουμε αυτή την χρονοσειρά θεωρώντας διαδοχικά σύνολα τιμών της αρχικής χρονοσειράς ομαδοποιημένα σε

σύνολα μήκους k . Συμβολίζουμε τις χρονικά συναθροισμένες χρονοσειρές με $Y[k]$ με παρατηρήσεις $y_i[k]$ και $i=1,2,\dots,n/k$ έτσι ώστε :

$$y_i[k] = \sum_{t=1+(i-1)k}^{ik} y_t$$

Προφανώς για $i=1$ προκύπτει $y_i[k]=y_t$. Θεωρούμε ως τελεστή συνάθροισης τον αριθμητικό μέσο όρο. Η προκύπτουσα χρονοσειρά $Y[k]$ έχει (n/k) παρατηρήσεις. Αν το υπόλοιπο της διαίρεσης n/k είναι μη μηδενικό για ένα συγκεκριμένο επίπεδο συνάθροισης k , τότε κάποιες παρατηρήσεις δεν θα συμπεριληφθούν κατά τον σχηματισμό συνόλων χρονικών συναθροίσεων (aggregation buckets). Στην περίπτωση αυτή αφαιρούμε από την αρχική χρονοσειρά $(n-n/k)$ παρατηρήσεις ξεκινώντας από την αρχή ούτως ώστε να σχηματίσουμε πλήρη σύνολα συνάθροισης. Θεωρητικά, μπορούμε να συναθροίσουμε μια χρονοσειρά για όλα τα $k=2,\dots,n$. Οστόσο, στην πράξη είναι προτιμότερο να κάνουμε συνάθροιση για $k \ll n$, ούτως ώστε να αφήσουμε ένα επαρκές δείγμα για παραμετροποίηση της μεθόδου πρόβλεψης που εφαρμόζεται στην χρονοσειρά $Y[k]$.

- Στάδιο Πρόβλεψης: Οι επιμέρους προβλέψεις υπολογίζονται με χρήση της μεθόδου ets που θα αναλυθεί περαιτέρω παρακάτω. Στη συνέχεια, οι προβλεπόμενες τιμές της $Y[k]$ προκύπτουν με απλή άθροιση των προβλεπόμενων συνιστωσών, δηλαδή :

$$\hat{y}_{i+h[k]}[k] = \hat{I}_{i+h[k]}[k] + \hat{b}_{i+h[k]}[k] + \hat{s}_{i-s+h[k]}[k]$$

Η παραγωγή προβλέψεων διαφορετικών επιπέδων συνάθροισης για μια δοσμένη χρονική περίοδο έχει ως έξοδο διαφορετικές μήκος τιμών. Για παράδειγμα, αν η αρχική χρονοσειρά $Y[1]$ έχει μηνιαία συχνότητα, τότε η πρόβλεψη για έναν χρόνο μπροστά απαιτεί 12 τιμές προβλέψεων ενώ για τη συναθροισμένη χρονοσειρά $Y[12]$ απαιτείται μόνο μία τιμή για την ίδια πρόβλεψη. Προκειμένου να διευκολυνθεί ο συνδυασμός των προβλέψεων, είναι επιθυμητό να έχουμε όλες τις συναθροισμένες προβλέψεις εκφρασμένες στην χρονική κλίμακα της αρχικής χρονοσειράς $Y[1]$. Για κάθε τιμή $z_i[k]$ αυτό επιτυγχάνεται κάνοντας χρήση της παρακάτω σχέσης :

$$z_i[k] = \sum_{j=1}^k \omega_j z_i[k], \text{ όπου } t = 1, 2, \dots, n \text{ και } i = \lceil t/k \rceil$$

Ουσιαστικά, η παραπάνω εξίσωση διαμοιράζει την i -τιμή σε k παρατηρήσεις στην αρχική χρονική βάση λειτουργώντας ως τμηματικά συνεχής παρεμβολή.

- Στάδιο Συνδυασμού: Στο τρίτο και τελευταίο στάδιο του αλγόριθμου MAPA συνδυάζονται οι συνιστώσες που υπολογίστηκαν προηγουμένως με στόχο την παραγωγή εκτιμήσεων για την τάση, το επίπεδο και την εποχιακότητα, οι οποίες μεταγενέστερα θα χρησιμοποιηθούν για την παραγωγή των τελικών προβλέψεων. Δεδομένου ότι η συνάθροιση των χρονοσειρών έγινε κάνοντας χρήση του αριθμητικού μέσου όρου, δεν χρειάζεται περαιτέρω ρύθμιση του επιπέδου των συνιστωσών που προέκυψαν για τα διάφορα επίπεδα συνάθροισης. Στη συνέχεια υπάρχει η περίπτωση που οι μέθοδοι συνδυάζονται με μη σταθμισμένο μέσο όρο, ως εξής:

$$\hat{I}_{t+h[1]}[k] = K^{-1} \sum_{k=1}^K I_{t+h[1]}[k]$$

$$\hat{b}_{t+h[1]}[k] = K^{-1} \sum_{k=1}^K b_{t+h[1]}[k]$$

$$\hat{s}_{t+h[1]}[k] = K'^{-1} \sum_{k=1}^{K'} s_{t+h[1]}[k], S/k \in \mathbb{Z} \text{ και } k < S$$

Στις παραπάνω σχέσεις, K είναι το μέγιστο επίπεδο συνάθροισης και K' είναι το πλήθος των επιπέδων συνάθροισης, στα οποία διακρίνεται η συνιστώσα της εποχιακότητας.

Μέθοδοι Διακοπτόμενης Ζήτησης

Οι διακοπτόμενες χρονοσειρές, οι οποίες κυριαρχούν στις υπηρεσίες και τα αποθέματα σε πολλές βιομηχανίες, περιλαμβάνουν λίγες τιμές που οδηγούν σε σποραδική ζήτηση. Τέτοιες χροσειρές δεν μπορούν να αντιμετωπιστούν αποτελεσματικά από τις μεθόδους που περιλαμβάνονται στον προηγούμενο κατάλογο για συνεχή δεδομένα. Όπως υποδηλώνει η βιβλιογραφία, η Croston είναι η πιο συνήθης μέθοδος για την πρόβλεψη διακοπτόμενης ζήτησης. Εντούτοις, η SBA μπορεί να χρησιμοποιηθεί για να περιορίσει περίπου τη θετική της μεροληψία και η TSB για να δημιουργήσει αμερόληπτες προβλέψεις για όλα τα χρονικά σημεία, επικαιροποιώντας την πιθανότητα ζήτησης αντί για το διάστημα ζήτησης. Τέλος, η iMAPA είναι επίσης μια εναλλακτική λύση για την εισαγωγή των πλεονεκτημάτων της χρονικής συνάθροισης και της μεθόδου MAPA σε διακοπτόμενη ζήτηση. Το OMEN περιλαμβάνει όλες τις προαναφερθείσες μεθόδους και τις εφαρμόζει μέσα από το πακέτο tsintermittent ενώ ταυτόχρονα χρησιμοποιούνται αυτοματοποιημένες μέθοδοι για τη βελτιστοποίηση των παραμέτρων τους και την ενίσχυση της απόδοσής τους.

Croston

Η μέθοδος Croston αφορά την εξαγωγή προβλέψεων εφαρμόζοντας ανεξάρτητα απλή εκθετική εξομάλυνση τόσο στις μη μηδενικές τιμές των χρονοσειρών όσο και στα χρονικά διαστήματα μεταξύ των μη μηδενικών τιμών των χρονοσειρών. Πιο αναλυτικά, η μέθοδος που εισήγαγε ο Croston, διαχωρίζει αρχικά την χρονοσειρά των πραγματικών παρατηρήσεων σε δύο επιμέρους, όπου η μία αποτελείται από τα χρονικά διαστήματα μεταξύ των μη μηδενικών ζητήσεων (intervals) και η άλλη από το πλήθος των ανεξάρτητων μη μηδενικών ζητήσεων (demands). Οι δύο χρονοσειρές προεκτείνονται ανεξάρτητα με χρήση της μεθόδου εκθετικής εξομάλυνσης σταθερού επιπέδου. Ως παράμετρος εξομάλυνσης στην εφαρμογή της μεθόδου είναι συνήθης στη βιβλιογραφία η χρήση της τιμής $\alpha=0.05$. Η προέκταση των χρονοσειρών ακολουθείται από τον υπολογισμό της πρόβλεψης Croston μέσω εύρεσης του πηλίκου των δύο ανεξάρτητων προβλέψεων ως εξής:

$$F_{Croston} = \frac{F_{demands}}{F_{intervals}}$$

SBA

Οι Συντετός και Boylan (2001), διαπίστωσαν ότι η μέθοδος Croston είναι μια θετικά προκατειλημμένη μέθοδος (positive biased), δηλαδή παρουσιάζει μια αισιόδοξη τάση στα αποτελέσματα των προβλέψεων της. Ένα άλλο σημαντικό επίσης επίτευγμα της έρευνας τους ήταν η σύνδεση του επιπέδου της αισιόδοξης τάσης της μεθόδου με την τιμή της παραμέτρου εξομάλυνσης α που χρησιμοποιείται για την προέκταση των δύο αποσυντιθέμενων χρονοσειρών. Η μέγιστη προκατάληψη μάλιστα παρατηρήθηκε όταν η παράμετρος εξομάλυνσης α λάβει τη μέγιστη τιμή της δηλαδή $\alpha=1$. Γενικά παρατηρήθηκε μία αναλογία της θετικής προκατάληψης της μεθόδου με τις μεγάλες τιμές της παραμέτρου α , και η μέθοδος Croston να ενδείκνυται μόνο όταν το α είναι μικρότερο του 0.15. Η μέθοδος λοιπόν που αναπτύχθηκε, πήρε το όνομά της από τους

προαναφερόμενους, ονομάζεται SBA (Syntetos and Boylan Approximation) και αποτελεί μία τροποποίηση της Croston, στην οποία η πρόβλεψη υπολογίζεται από τον ακόλουθο μαθηματικό τύπο:

$$F_{SBA} = \left(1 - \frac{a}{2}\right) \frac{F_{demands}}{F_{intervals}}$$

Μεταγενέστερες εμπειρικές μελέτες και μελέτες προσομοίωσης (Syntetos & Boylan, 2005; Teunter & Sani, 2009) έδειξαν πως η αρχική μέθοδος Croston παρουσιάζει μικρότερη θετική προκατάληψη εάν υπάρχουν λίγες μη μηδενικές τιμές στη ζήτηση ενώ η τροποποιημένη μέθοδος των Σύντετου και Boylan έχει μικρότερη προκατάληψη εάν παρατηρούνται περισσότερες μηδενικές τιμές στη ζήτηση.

TSB

Η μέθοδος Teunter, Syntetos και Babai (TSB) προκύπτει από τις εξισώσεις:

$$\begin{aligned} \text{Αν } Dt = 1 \text{ τότε } P_{t+1} &= \beta(1) + (1 - \beta)P_t, \\ Z_{t+1} &= aX_t + (1 - a)Z_t, \\ Y_{t+1} &= P_{t+1}Z_{t+1} \end{aligned}$$

$$\begin{aligned} \text{Αν } Dt = 0 \text{ τότε } P_{t+1} &= (1 - \beta)P_t, \\ Z_{t+1} &= Z_t, \\ Y_{t+1} &= P_{t+1}Z_{t+1} \end{aligned}$$

Το βασικό πλεονέκτημα αυτής της μεθόδου είναι ότι η ενημέρωση των προβλέψεων σε κάθε χρονική περίοδο (είτε υπάρχει ζήτηση ή όχι) επιτρέπει την εκτίμηση P_t να προσεγγίσει το μηδέν εάν υπάρχει μια μακρά περίοδος περιόδων χωρίς ζήτηση. Αντιθέτως, η εκτίμηση διαστήματος που υπολογίζεται από την Croston θα παραμείνει αμετάβλητη. Σε ένα πρακτικό πλαίσιο, η TSB επιτρέπει να λαμβάνονται αποφάσεις για το κατά πόσον θα συνεχίσει να αποθηκεύει τα στοιχεία ή όχι. Μία ιδέα ίσως είναι να ορίσετε κάποιο κατώτατο όριο για το p , έτσι ώστε εάν ξεπεραστεί (μόνο μία φορά ή για αρκετές διαδοχικές περιόδους), θα αποφασιστεί ότι το προϊόν είναι πλέον παρωχημένο.

iMAPA

Η μέθοδος iMAPA αποτελεί την έκδοση της μεθόδου MAPA που αναφέρεται σε προβλέψεις δεδομένων διακοπτόμενης ζήτησης. Η μέθοδος αυτή δέχεται διαφορετικά επιχειρήματα για το ελάχιστο και το μέγιστο επίπεδο χρονικής συνάθροισης, τη μέθοδο συνδυασμού των επιμέρους χρονοσειρών και τις παραμέτρους του μοντέλου.

Αξιολόγηση

Αυτό το στοιχείο μενού παρέχει στον χρήστη μια ολοκληρωμένη αναφορά σχετικά με την απόδοση της συνολικής διαδικασίας πρόβλεψης, αξιολογώντας την ακρίβεια και τη μεροληψία των προβλέψεων μέσω πολλαπλών μετρήσεων σφάλματος που παρατίθενται αναλυτικότερα παρακάτω.

Μέσο σφάλμα (Mean Error - ME)

Υπολογίζεται από τον προσημασμένο μέσο όρο των σφαλμάτων και εκφράζει ένα μέτρο συστηματικότητας του σφάλματος. Τιμές του δείκτη αυτού κοντά στο μηδέν υποδηλώνουν ότι τα σφάλματα είναι τυχαία και όχι συστηματικά.

Θετικές τιμές του ME δηλώνουν απαισιοδοξία στις προβλέψεις, ενώ αρνητικές τιμές αισιοδοξία.

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

Μέσο απόλυτο σφάλμα (Mean Absolute Error - MAE)

Είναι ένα μέτρο αστοχίας της πρόβλεψης σε σχέση με την πραγματική τιμή, χωρίς να λαμβάνεται υπόψη η κατεύθυνση της πρόβλεψης. Διατηρεί τις μονάδες μέτρησης της αρχικής χρονοσειράς. Μεγαλύτερες τιμές του δείκτη MAE υποδηλώνουν μικρότερη ακρίβεια στις προβλέψεις.

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

Μέσο τετραγωνικό σφάλμα (Mean Squared Error - MSE)

Όπως και ο προηγούμενος δείκτης, είναι ένα μέτρο της ακρίβειας των προβλέψεων. Δίνει αρκετά μεγαλύτερο βάρος σε μεγάλα σφάλματα (λόγω τετραγωνισμού των σφαλμάτων) και μικρότερο βάρος στα μικρά σφάλματα. Ο δείκτης MSE χρησιμοποιείται για τον υπολογισμό των βέλτιστων συντελεστών εξομάλυνσης.

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

Ρίζα μέσου τετραγωνικού σφάλματος (Root Mean Squared Error - RMSE)

Για να εκφράζεται σε μονάδες της αρχικής χρονοσειράς χρησιμοποιείται η ρίζα του MSE: RMSE = (Root Mean Squared Error).

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

Μέσο απόλυτο ποσοστιαίο σφάλμα (Mean Absolute Percentage Error - MAPE)

Ενδείκνυται για τη σύγκριση της ακρίβειας χρονοσειρών με διαφορετικά επίπεδα μέσης τιμής. Μικρότερες τιμές του εν λόγω δείκτη υποδηλώνουν μεγαλύτερη ακρίβεια της μεθόδου πρόβλεψης. Ο δείκτης MAPE δεν μπορεί να εφαρμοσθεί σε χρονοσειρές διακοπτόμενης ζήτησης διότι καταλήγει σε απροσδιοριστία.

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

Συμμετρικό μέσο απόλυτο ποσοστιαίο σφάλμα (Symmetric Mean Absolute Percentage Error - sMAPE)

Η διαφορά με τον προηγούμενο δείκτη είναι ότι το σφάλμα δε διαιρείται με την πραγματική τιμή, αλλά με το ημίθροισμα της πραγματικής τιμής και της πρόβλεψης. Έτσι ο δείκτης sMAPE αποκτά ανώτατο όριο και παίρνει τιμές στο διάστημα [0%, 200%].

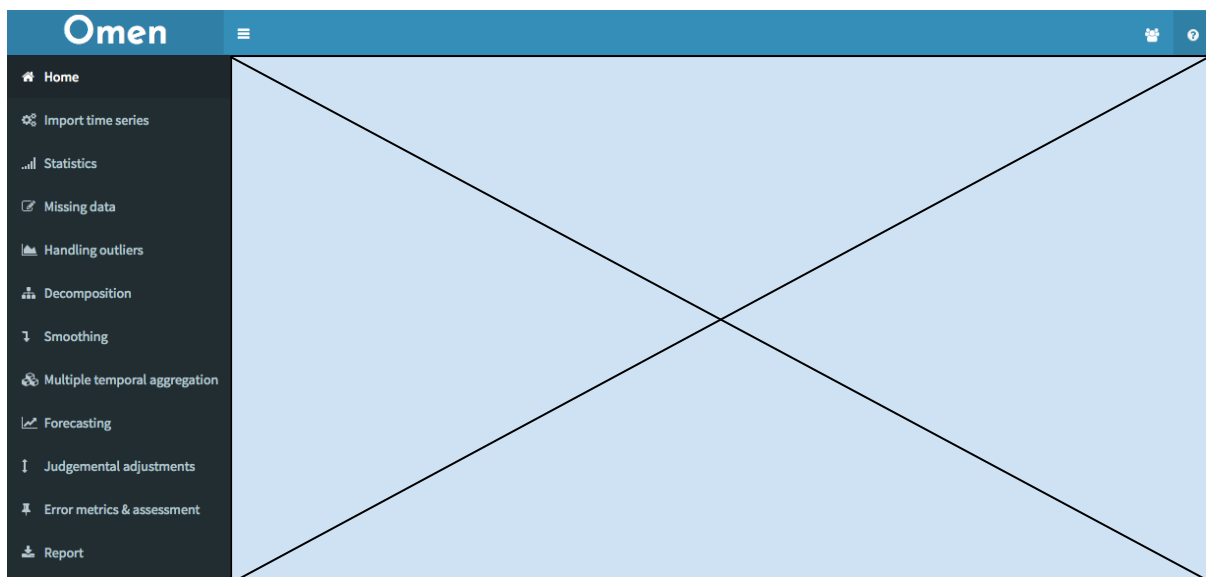
$$sMAPE = \frac{1}{n} \sum_{i=1}^n \frac{|Y_i - F_i|}{(Y_i + F_i)/2}$$

Κεφάλαιο 5

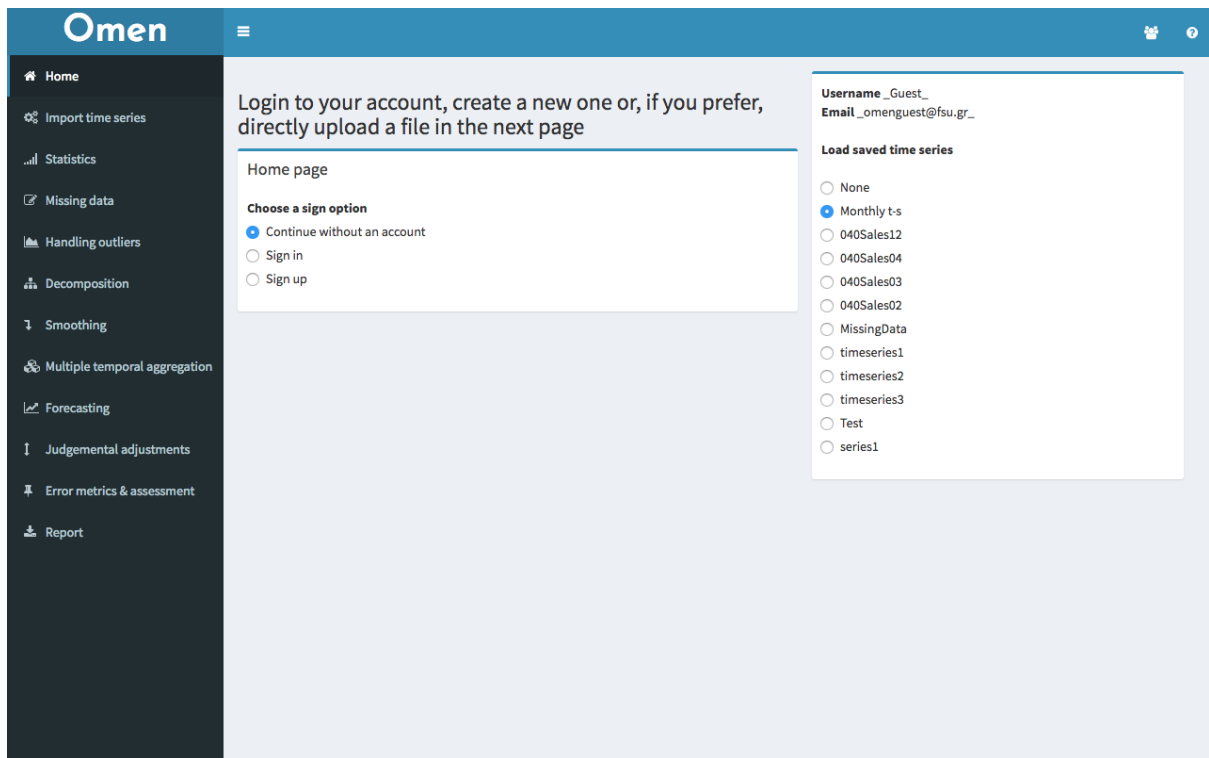
Υλοποίηση της πλατφόρμας

Το σύστημα έχει την μορφή πίνακα ελέγχου με πλάγιο μενού. Είναι σχεδιασμένο με τρόπο τέτοιο ώστε να καθοδηγεί τον χρήστη καθ' όλη τη διαδικασία της πρόβλεψης. Συγκεκριμένα τα μενού είναι δυναμικά και οι διαθέσιμες κινήσεις αλλάζουν ανάλογα με τις προηγούμενες επιλογές του χρήστη έτσι ώστε να μην υπάρχουν αχρείαστα μενού ή στοιχεία στην επιφάνεια εργασίας. Το γενικό μοτίβο του σχεδιασμού έχει εκτός από το πλάγιο διασπώμενο μενού που βρίσκεται στην αριστερή πλευρά της οθόνης, μία μπάρα στο πάνω μέρος της οθόνης με το όνομα του συστήματος και δύο επιλογές. Η μία δείχνει την ομάδα πίσω από το σύστημα σε ένα αναδυόμενο παράθυρο ενώ η άλλη παρέχει στον χρήστη ότι πληροφορίες και βοήθημα μπορεί να χρειαστεί σε κάθε σελίδα που βρίσκεται.

Παρακάτω παρουσιάζεται το βασικό πλαίσιο της εφαρμογής



Αρχική Σελίδα



Η αρχική σελίδα έχει ως σκοπό την σύνδεση του χρήστη έτσι ώστε να μπορέσει να χρησιμοποιήσει το σύστημα. Είναι δυναμική και ανάλογα με την επιλογή του χρήστη στο αριστερό παράθυρο εμφανίζονται ή εξαφανίζονται τα υπόλοιπα παράθυρα.

Αν ο χρήστης επιλέξει να συνεχίσει χωρίς λογαριασμό παραμένει συνδεδεμένος ως Guest και στο δεξί μέρος βλέπει τις χρονοσειρές του guest χρήστη. Αντιθέτως, αν επιλέξει την επιλογή Sign In ή Sign Up εμφανίζονται τα παράθυρα για τις αντίστοιχες λειτουργίες.

Login to your account, create a new one or, if you prefer, directly upload a file in the next page

Home page

Choose a sign option

Continue without an account

Sign in

Sign up

Username:

Password:

Confirm Password:

Email Address:

Create Account

Login to your account, create a new one or, if you prefer, directly upload a file in the next page

Home page

Choose a sign option

Continue without an account

Sign in

Sign up

Login

Username:

Password:

Login

Επιλογή ή Μεταφόρτωση Χρονοσειρών

The screenshot displays the 'Import time series page' in the Omen application. On the left is a dark sidebar with navigation links. The main area is divided into three sections: a file upload section, a chart, and a data table.

File Upload Section:

- Text: "Please choose a file to import"
- File selected: "N1408.csv"
- Status: "Upload complete"
- Options:
 - Header
 - Separator:
 - Comma
 - Semicolon
 - Tab
 - Data frequency: "12"
 - Date range: "1990-01-01" to "1995-08-01"

View imported & edited time series: A line chart showing data points over time, with values ranging from approximately 1k to 8k.

View & edit imported data: A table with 18 rows and 3 columns: Rownames, Original Data, and Edited Data.

Rownames	Original Data	Edited Data
1	1350	1350
2	2400	2400
3	1950	2250
4	2250	2250
5	3000	3000
6	2400	2700
7	3000	3000
8	2850	2850
9	2550	2950
10	3600	3300
11	2850	2850
12	2250	2250
13	2100	2100
14	2100	2100
15	3750	3750
16	2250	2250
17	2400	2400
18	2400	2400

Υπάρχουν δύο τρόποι να επιλέξει ένας χρήστης χρονοσειρά για να αρχίσει τη διαδικασία της πρόβλεψης. Ο πρώτος είναι να επιλέξει μια χρονοσειρά από αυτές που έχει αποθηκεύσει στην βάση σε κάποια προηγούμενη συνεδρία του συστήματος. Αυτό γίνεται πολύ εύκολα από το δεξί παράθυρο στην αρχική σελίδα. Εναλλακτικά, το σύστημα διαθέτει την δυνατότητα μεταφόρτωσης μίας χρονοσειράς από τον υπολογιστή του χρήστη σε μορφή .csv αρχείου. Η λειτουργία αυτή είναι διαθέσιμη στην καρτέλα με ετικέτα "Import Time Series".

Αυτό το στοιχείο μενού επιτρέπει στο χρήστη να εισαγάγει μια χρονοσειρά στο OMEN για να ξεκινήσει η διαδικασία πρόβλεψης. Η χρονοσειρά μπορεί να εισαχθεί είτε μέσω ενός αρχείου csv το οποίο πρέπει να περιλαμβάνει δύο στήλες: μία για τη χρονική σήμανση των παρατηρήσεων και μία για τα δεδομένα. Ο χρήστης πρέπει επίσης να καθορίσει αν το αρχείο περιέχει κεφαλίδες (headers), καθώς και τον χαρακτήρα που χρησιμοποιείται για να διαχωρίσει τα πεδία (Comma, Semicolon ή Tab).

Με οποιονδήποτε από τους παραπάνω τρόπους και αν επιλέξει τα δεδομένα του ο χρήστης, το σύστημα του παρέχει τη δυνατότητα να καθορίσει ορισμένα από τα χαρακτηριστικά της χρονοσειράς του. Αρχικά, η συχνότητα της επιλεγμένης χρονοσειράς μπορεί να οριστεί από τον χρήστη ή να υπολογιστεί αυτόματα από το σύστημα με βάση την αυτοσυσχέτιση των δεδομένων. Στην τελευταία περίπτωση, διεξάγεται ένας διαγωνισμός εποχικότητας με βάση τους συντελεστές αυτοσυσχέτισης για κάθε πιθανή συχνότητα και επιλέγεται η συχνότητα με την υψηλότερη τιμή αυτοσυσχέτισης. Ωστόσο, προκειμένου να θεωρηθούν εποχιακές οι χρονοσειρές, πρέπει να υπάρχει εμπιστοσύνη τουλάχιστον 90%.

Το δείγμα των δεδομένων που θα χρησιμοποιηθούν κατά τη διάρκεια της διαδικασίας πρόβλεψης (insample) μπορεί να επιλεγεί χρονικά μέσω ενός ημερολογίου. Αυτό το δείγμα είναι που θα

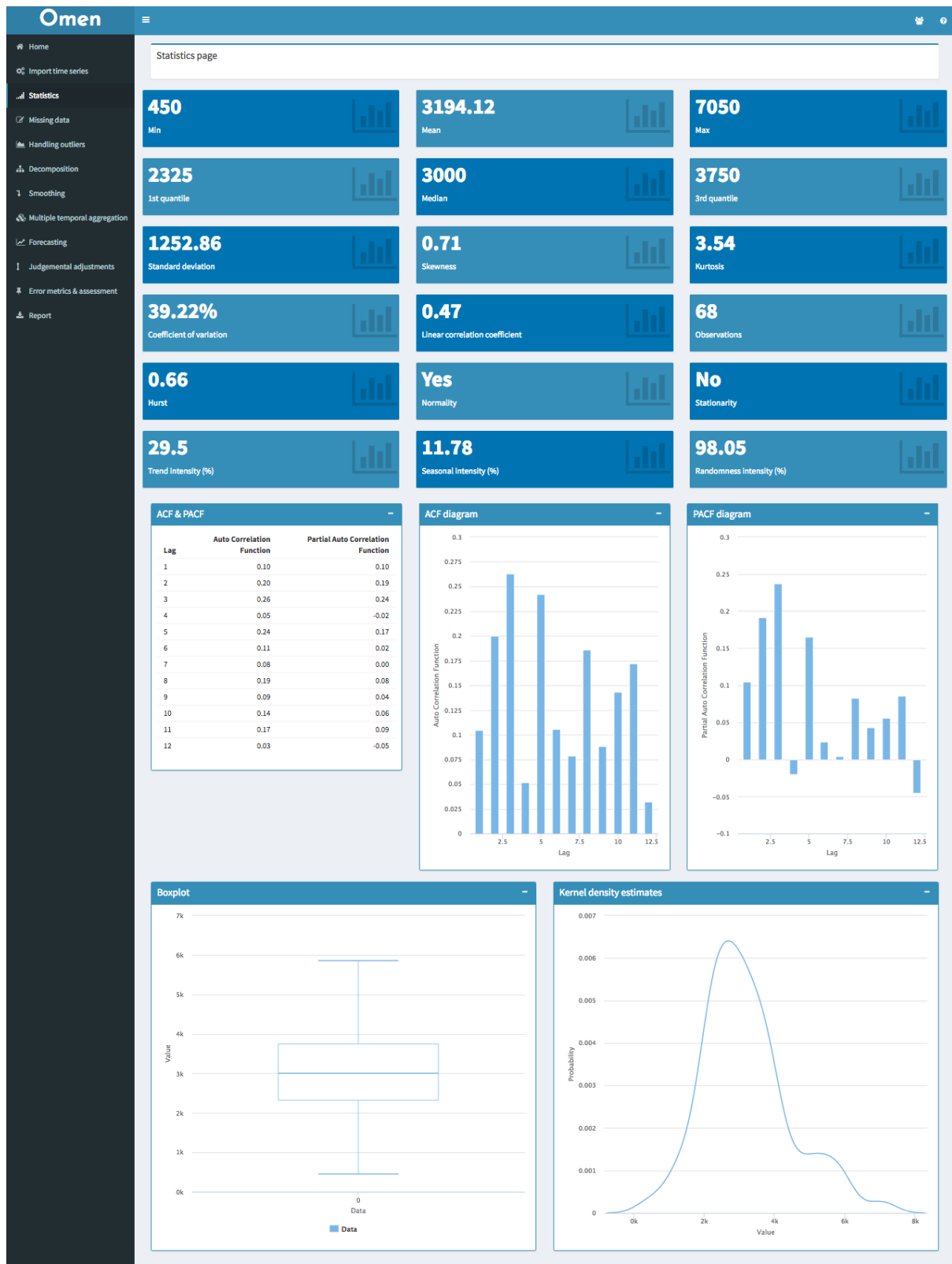
καθορίσει τις τιμές των προβλέψεων. Σε περίπτωση που είναι διαθέσιμα περισσότερα δεδομένα μετά την επιλεγμένη χρονική σφραγίδα (out-sample), το OMEN θα τα χρησιμοποιήσει για να αξιολογήσει καλύτερα την απόδοση πρόβλεψης.

Στο δεξί μέρος της σελίδας “Import Time Series” ο χρήστης μπορεί να επεξεργαστεί τις τιμές της χρονοσειράς του στο παράθυρο View & Edit Imported Data σε έναν πίνακα που θυμίζει αρχείο Excel. Στη συνέχεια μπορεί να δει την αρχική και την επεξεργασμένη χρονοσειρά γραφικά στο διάγραμμα που παρέχεται στο παράθυρο View & Edit Imported Time Series.

Σε αυτό το σημείο είναι σημαντικό να αναφερθεί ότι η παραπάνω λογική έχει ακολουθηθεί για λόγους συνέχειας σε ολόκληρο το σύστημα. Πιο συγκεκριμένα στο δεξί μέρος κάθε σελίδας υπάρχει ένας πίνακας με τα αρχικά σε σύγκριση με τα τελευταία δεδομένα και στο αριστερό το διάγραμμα με τις χρονοσειρές όλων των βημάτων μέχρι εκείνο το σημείο.

Τέλος, στην καρτέλα αυτή ο χρήστης έχει τη δυνατότητα να αποθηκεύσει την χρονοσειρά που μεταφόρτωσε στην βάση δεδομένων του OMEN, έτσι ώστε να μπορεί να την ανακτήσει άμεσα σε κάποια επόμενη συνεδρία.

Στατιστική Ανάλυση



Σε αυτό το σημείο πραγματοποιείται στατιστική ανάλυση υψηλού επιπέδου στη χρονοσειρά ώστε ο χρήστης να επιθεωρήσει τα κύρια χαρακτηριστικά των δεδομένων. Οι στατιστικοί δείκτες που δίνονται έχουν ως σκοπό να βοηθήσουν το χρήστη να γνωρίσει παραπάνω τα δεδομένα του έτσι ώστε να μπορέσει να πάρει σωστές αποφάσεις. Εκτός από τους δείκτες δίνονται επίσης και διαγράμματα, αφού οι γραφικές παραστάσεις μπορούν να βοηθήσουν στην κατανόηση των δεδομένων ακόμα και από χρήστες χωρίς πολλές γνώσεις στον τομέα των προβλέψεων.

Συμπλήρωση Κενών Τιμών

The screenshot displays the Omen software interface. On the left is a navigation menu with options like Home, Import time series, Statistics, Missing data, Handling outliers, Decomposition, Smoothing, Multiple temporal aggregation, Forecasting, Judgemental adjustments, Error metrics & assessment, and Report. The main area is titled 'Missing data page' and contains a section 'Choose a method to fill missing data' with four radio button options: 'Use last non-zero value', 'Use linear regression', 'Multivariate imputations (trend & period)', and 'Bootstrap EM Algorithm'. Below this is a chart titled 'Original & filled time series' showing 'value' on the y-axis (0k to 6k) and 'Observations' on the x-axis (0 to 120). The chart compares 'Adjusted Timeseries' (blue line) and 'Original Timeseries' (purple line). To the right is a table titled 'Original & filled data' with two columns: 'Original Timeseries' and 'Adjusted Timeseries'.

Original Timeseries	Adjusted Timeseries
4590.00	4590.00
4930.00	4930.00
5300.00	5300.00
0.00	4967.10
0.00	5017.52
5000.00	5000.00
5085.00	5085.00
0.00	5131.17
5065.00	5065.00
5165.00	5165.00
0.00	5186.74
4970.00	4970.00
5250.00	5250.00
0.00	5009.19
5675.00	5675.00
0.00	5017.05
0.00	5054.08
5680.00	5680.00
5525.00	5525.00
5180.00	5180.00
5415.00	5415.00
0.00	4755.21
0.00	5060.76
4570.00	4570.00
0.00	4741.11
0.00	4806.98
4845.00	4845.00
0.00	4891.14
0.00	4563.35
4315.00	4315.00
0.00	4990.99

Η σελίδα αυτή έχει ως σκοπό να βοηθήσει τον χρήστη να συμπληρώσει τυχούσες κενές τιμές που μπορεί να έχει η χρονοσειρά του. Η καρτέλα είναι διαμορφωμένη με την κλασική δομή του συστήματος. Είναι χωρισμένη σε δύο στήλες την δεξιά που αποτελείται από έναν πίνακα που διαθέτει αριθμητικά δεδομένα για τη χρονοσειρά και την αριστερή στήλη που αποτελείται από τις επιλογές για συμπλήρωση κενών τιμών καθώς και την γραφική παράσταση των δεδομένων.

Οι επιλογές που διατίθενται δίνονται σε μορφή radio button και αποτελούνται από τέσσερις μεθόδους συμπλήρωσης κενών τιμών. Με κάθε επιλογή του χρήστη τόσο ο πίνακας όσο και η γραφική παράσταση ανανεώνονται με τις ανάλογες τιμές έτσι ώστε να μπορεί ο χρήστης να έχει πλήρη εποπτεία των δεδομένων του.

Απαλοιφή Ακραίων Τιμών

Handling outliers page

Choose a method to handle your outliers
By identifying extreme values

Ta Value: 2

Consider Outlier Adjustments

Original & normalised time series

6000
5500
5000
4500
4000
3500
3000

0 10 20 30 40 50 60 70 80 90 100 110 120 130

Index

→ A. Original Timeseries → B. Adjusted Timeseries

Original & normalised data

Original Timeseries	Adjusted Timeseries
4590.00	4590.00
4930.00	4930.00
5300.00	5300.00
5465.00	5465.00
5395.00	5395.00
5000.00	5000.00
5085.00	5085.00
4935.00	4935.00
5065.00	5065.00
5165.00	5165.00
5300.00	5300.00
4970.00	4970.00
5250.00	5250.00
5515.00	5515.00
5675.00	5675.00
5805.00	5805.00
5820.00	5820.00
5680.00	5680.00
5525.00	5525.00
5180.00	5180.00
5415.00	5415.00
5075.00	5075.00
4775.00	4775.00
4570.00	4570.00
4620.00	4620.00
4790.00	4790.00

Download

Handling outliers page

Choose a method to handle your outliers
By fitting a forecasting model & observing unusual values

Ta Value: 9

Tb Value: 25

Consider Outlier Adjustments

Original & normalised time series

6000
5500
5000
4500
4000
3500
3000

0 10 20 30 40 50 60 70 80 90 100 110 120 130

Index

→ A. Original Timeseries → B. Adjusted Timeseries

Original & normalised data

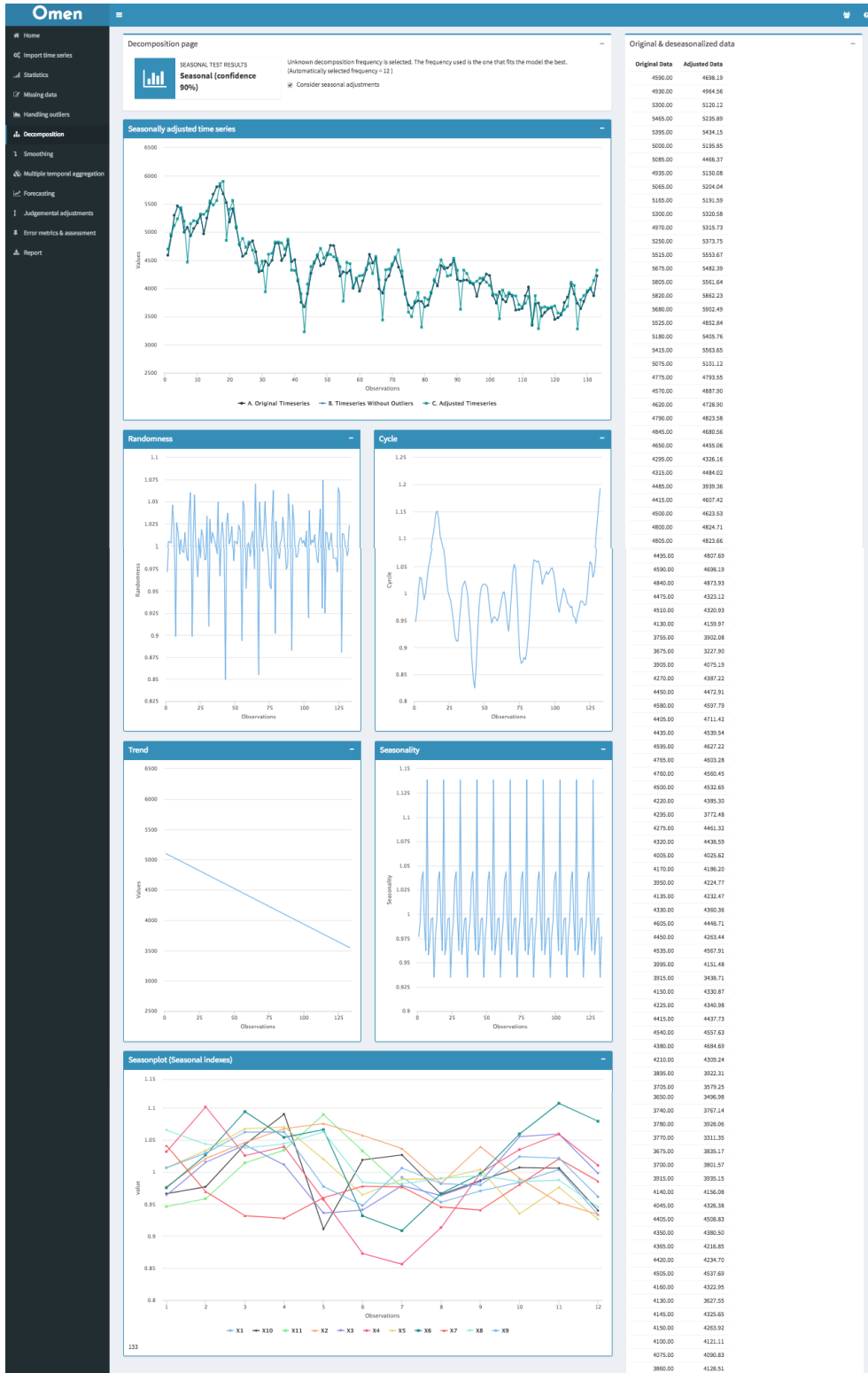
Original Timeseries	Adjusted Timeseries
4590.00	4590.00
4930.00	4930.00
5300.00	5300.00
5465.00	5465.00
5395.00	5395.00
5000.00	5000.00
5085.00	5085.00
4935.00	4935.00
5065.00	5065.00
5165.00	5165.00
5300.00	5300.00
4970.00	4970.00
5250.00	5250.00
5515.00	5515.00
5675.00	5675.00
5805.00	5805.00
5820.00	5820.00
5680.00	5519.94
5525.00	5525.00
5180.00	5180.00
5415.00	5075.50
5075.00	5075.00
4775.00	4775.00
4570.00	4570.00
4620.00	4620.00
4790.00	4790.00

Download

Η καρτέλα αυτή έχει ως σκοπό την απαλοιφή ακραίων τιμών από την αρχική χρονοσειρά. Η σελίδα είναι διαμορφωμένη με την κλασική δομή του OMEN δηλαδή στην δεξιά στήλη παρατίθεται ο πίνακας με τα δεδομένα και στην αριστερή η γραφική παράσταση και οι επιλογές. Οι μέθοδοι που διατίθενται, έχουν την μορφή dropdown list και ανάλογα με την επιλογή του χρήστη σε ότι αφορά την μέθοδο, ανανεώνονται οι διαθέσιμες επιλογές για παραμέτρους εφόσον διαφορετικές μέθοδοι έχουν διαφορετικό αριθμό παραμέτρων. Με κάθε επιλογή του χρήστη τόσο ο πίνακας όσο και η γραφική

παράσταση αναεώνονται με τις ανάλογες τιμές έτσι ώστε να μπορεί ο χρήστης να έχει πλήρη εποπτεία των δεδομένων του.

Αποσύνθεση



Στην καρτέλα της αποσύνθεσης η κλασική δομή διατηρείται. Στο κουτί των επιλογών παρατίθεται η συχνότητα που έχει επιλεγεί με το τεστ εποχιακότητας που πραγματοποιείται αυτόματα από το σύστημα καθώς και το ποσοστό εμπιστοσύνης της επιλεγμένης συχνότητας.

Τα διαγράμματα που παρατίθενται αποτελούνται από τα τέσσερα βασικά συστατικά της χρονοσειράς που είναι η τυχαιότητα ο κύκλος η τάση και η εποχιακότητα. Τέλος, δίνεται το seasonplot που δείχνει την χρονοσειρά χωρισμένη σε κομμάτια και έχει ως σκοπό να κάνει πιο εύκολη την κατανόηση του πως έγινε ο χωρισμός της χρονοσειράς.

Εξομάλυνση

The screenshots show the 'Smoothing page' in the Omen application. The interface includes a sidebar with navigation options, a main control area for selecting a smoothing method and adjusting parameters, a time series plot, and a table of original and smoothed data.

Top Screenshot (Non Linear Smooth):

- Method: Non Linear Smooth
- Option: Consider Smoothing Adjustments

Original data	Smoothed data
4590.00	4885.77
4930.00	4965.00
5300.00	5119.24
5465.00	5208.90
5395.00	5223.40
5000.00	5196.65
5085.00	5150.27
4935.00	5147.99
5065.00	5204.25
5165.00	5225.84
5300.00	5294.19
4970.00	5314.01
5250.00	5376.75
5515.00	5460.36
5675.00	5481.45
5805.00	5561.95
5820.00	5588.07
5680.00	5542.15
5525.00	5453.43
5180.00	5403.57
5415.00	5290.74
5075.00	5099.88
4775.00	4967.92
4570.00	4886.32
4620.00	4793.35
4790.00	4758.41

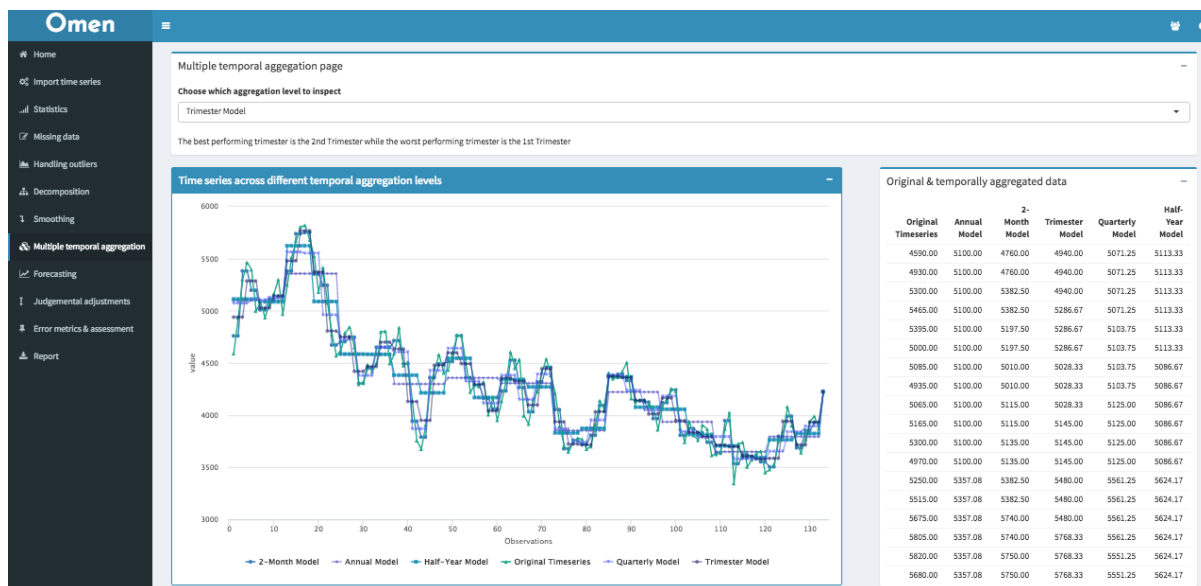
Bottom Screenshot (Simple Moving Average):

- Method: Simple Moving Average
- Order of moving average smoother: 4
- Option: Consider Smoothing Adjustments

Original data	Smoothed data
4590.00	5155.04
4930.00	5138.67
5300.00	5097.36
5465.00	5218.36
5395.00	5165.76
5000.00	5073.18
5085.00	5033.03
4935.00	5003.12
5065.00	5108.79
5165.00	5236.01
5300.00	5278.32
4970.00	5345.36
5250.00	5411.22
5515.00	5462.58
5675.00	5554.67
5805.00	5659.43
5820.00	5624.56
5680.00	5526.24
5525.00	5468.73
5180.00	5330.58
5415.00	5222.44
5075.00	5150.08
4775.00	4981.38
4570.00	4842.86
4620.00	4794.39
4790.00	4726.53

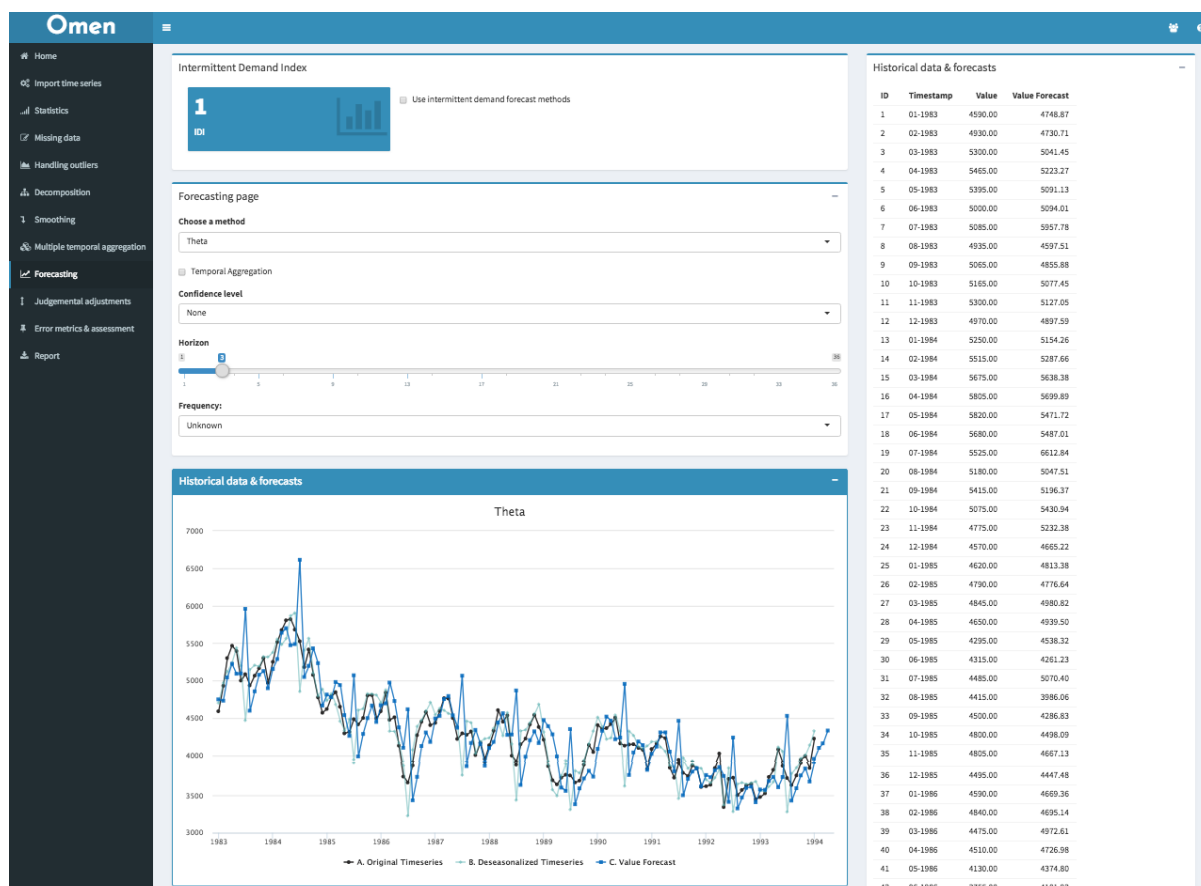
Η κλασική δομή του συστήματος διατηρείται και στην καρτέλα εξομάλυνσης. Πιο συγκεκριμένα, και εδώ δίνονται οι επιλογές μεθόδου εξομάλυνσης σε μορφή dropdown selectbox και ανανεώνονται οι επιλογές παραμέτρων ανάλογα με τη μέθοδο. Για παράδειγμα το Non Linear Smooth δεν έχει παραμέτρους, όπως φαίνεται στην πρώτη εικόνα ενώ ο ΚΜΟ έχει και επιλέγεται από τον χρήστη, τι ΚΜΟ θέλει.

Χρονική Συνάθροιση



Σε αυτή τη καρτέλα ανάλογα με την επιλογή που κάνει ο χρήστης για το επίπεδο χρονικής συνάθροισης βλέπει στο κουτί επιλογών ένα μήνυμα που του δίνει πληροφορίες για το επιλεγμένο επίπεδο χρονικής συνάθροισης. Ακόμη δίνονται τα δεδομένα τόσο σε γραφική παράσταση όσο και σε πίνακα.

Πρόβλεψη

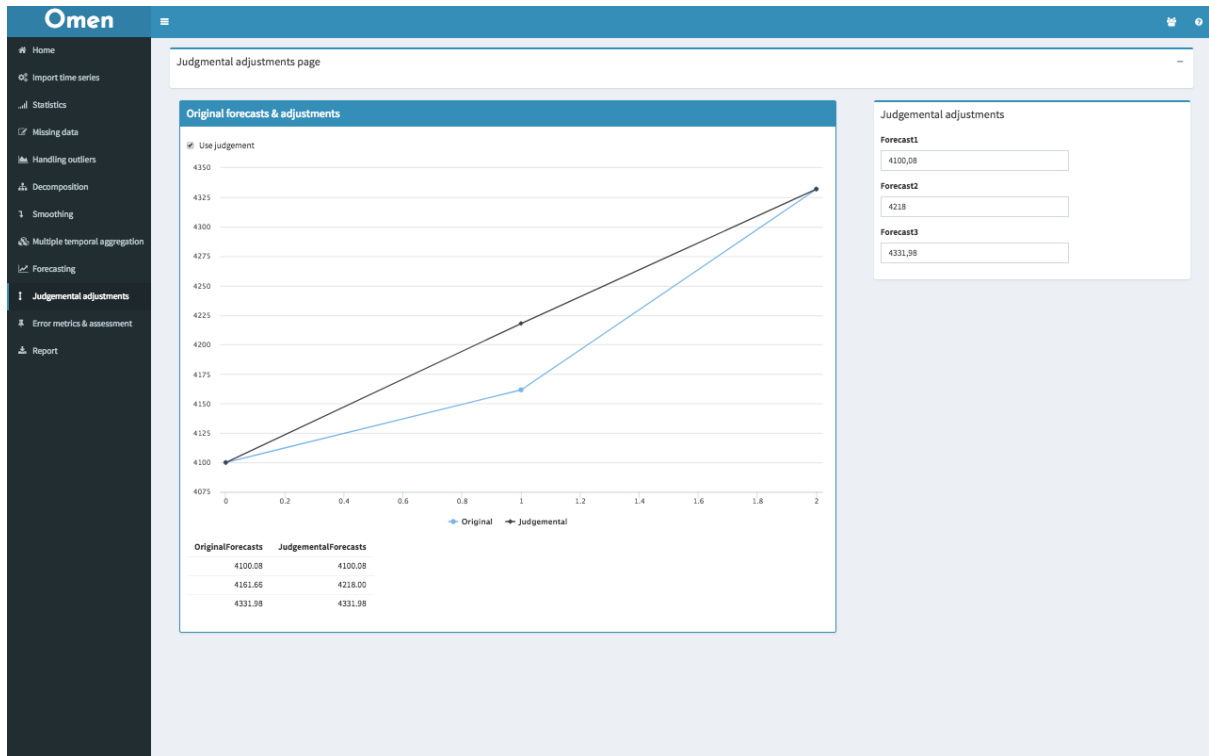


Η καρτέλα των προβλέψεων έχει την κλασική δομή του συστήματος. Στο κουτί των επιλογών μπορεί να επιλεγεί η μέθοδος πρόβλεψης, ο υπολογισμός επίπεδων εμπιστοσύνης, ο ορίζοντας πρόβλεψης καθώς και η συχνότητα της χρονοσειράς.

Η γραφική παράσταση απεικονίζει όχι μόνο τη πρόβλεψη αλλά και όλα τα ενδιάμεσα βήματα από τα οποία έχει περάσει ο χρήστης μέχρι να οδηγηθεί στην πρόβλεψη. Επίσης, ο τίτλος του διαγράμματος είναι ο τίτλος της μεθόδου που εφαρμόστηκε, για παράδειγμα, στο παραπάνω στιγμιότυπο theta.

Τέλος, πέρα από την κλασική δομή, στην καρτέλα προβλέψεων έχουμε ένα παραπάνω κουτί το οποίο αφορά την χρήση μεθόδων διακοπτόμενης ζήτησης. Πιο συγκεκριμένα δίνεται ο δείκτης IDI που δείχνει κατά πόσο η χρονοσειρά είναι διακοπτόμενη και εάν η τιμή του είναι μεγαλύτερη από 1.2, επιλέγεται αυτόματα να θεωρηθεί χρονοσειρά διακοπτόμενης ζήτησης και οι μέθοδοι που φαίνονται στο dropdown παρακάτω είναι μόνο μέθοδοι διακοπτόμενης ζήτησης. Δίνεται ακόμα ένα checkbox σε περίπτωση που ο χρήστης θέλει να επιβάλλει την εφαρμογή μεθόδων διακοπτόμενης ζήτησης.

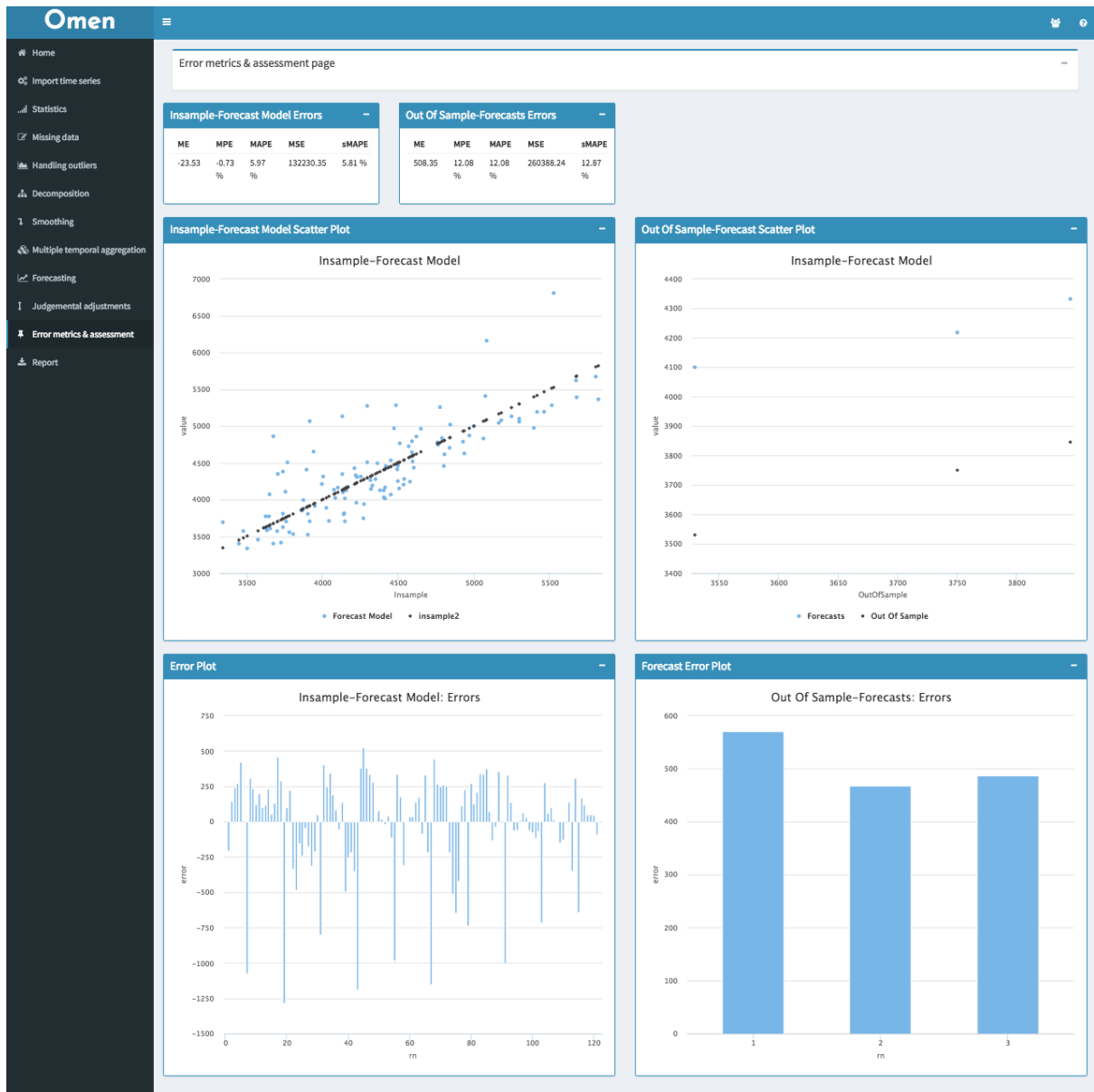
Κριτικές Προσαρμογές



Στην καρτέλα κριτικών προσαρμογών δίνονται δύο τρόποι με τους οποίους ο χρήστης μπορεί να εφαρμόσει τις δικές του προσαρμογές στις στατιστικές προβλέψεις. Αρχικά δίνεται η επιλογή να γίνουν απευθείας πάνω στην γραφική παράσταση κάνοντας απλά drag - drop την παρατήρηση που θέλει να αλλάξει. Εναλλακτικά μπορεί να επιλέξει να συμπληρώσει ανάλογα το κουτί της παρατήρησης στο δεξί μέρος της σελίδας.

Σε κάθε περίπτωση ανανεώνεται το διάγραμμα αλλά και ο πίνακας με τις τελευταίες τιμές που έχουν επιλεγεί.

Αξιολόγηση Προβλέψεων



Στην καρτέλα αυτή δίνονται εκτός από τους υπολογισμένους δείκτες σφαλμάτων και διαγράμματα που βοηθούν στην αξιολόγηση των προβλέψεων.

Εξαγωγή Αναφοράς

Η αναφορά αποτελεί μία από τις σημαντικότερες λειτουργίες του συστήματος. Πιο συγκεκριμένα, αποτελεί το εργαλείο το οποίο χρησιμοποιείται από τους αναλυτές για την εξαγωγή συμπερασμάτων. Παρακάτω φαίνεται μία αναφορά που έχει παραχθεί από το σύστημα μετά από μία διαδικασία πρόβλεψης.

Omen:

Imported file:

January	February	Marc h	April	May	June	July	August	September
4590	4970	4775	4800	4270	4275	3915	3780	4505
4930	5250	4570	4805	4450	4320	4150	3770	4160
5300	5515	4620	4295	4280	4305	4225	3675	4130
5465	5675	4790	4590	4405	4170	4115	3700	4145
5395	5805	4845	4840	4435	3950	4540	3915	4150
5000	5820	4650	4475	4595	4135	4380	4140	4100
5085	5680	4295	4510	4865	4330	4210	4045	4075
4935	5525	4315	4130	4760	4605	3895	4405	3860
5165	5180	4285	3755	4500	4450	3705	4350	4085
5165	5215	4415	3675	4220	4535	3650	4365	4155
5200	5075	4500	3905	4295	3995	3740	4420	4255

Table continues below

October	November	December
4230	3870	3530
3875	4025	3750
3740	3345	3845
3940	3725	4080
3805	3740	3900
3760	3505	3735
3905	3575	3640
3865	3635	3770
3615	3655	3940
3625	3450	3990
3645	3480	3870

Statistics:

Mean: 4308.3

Median: 4227.5

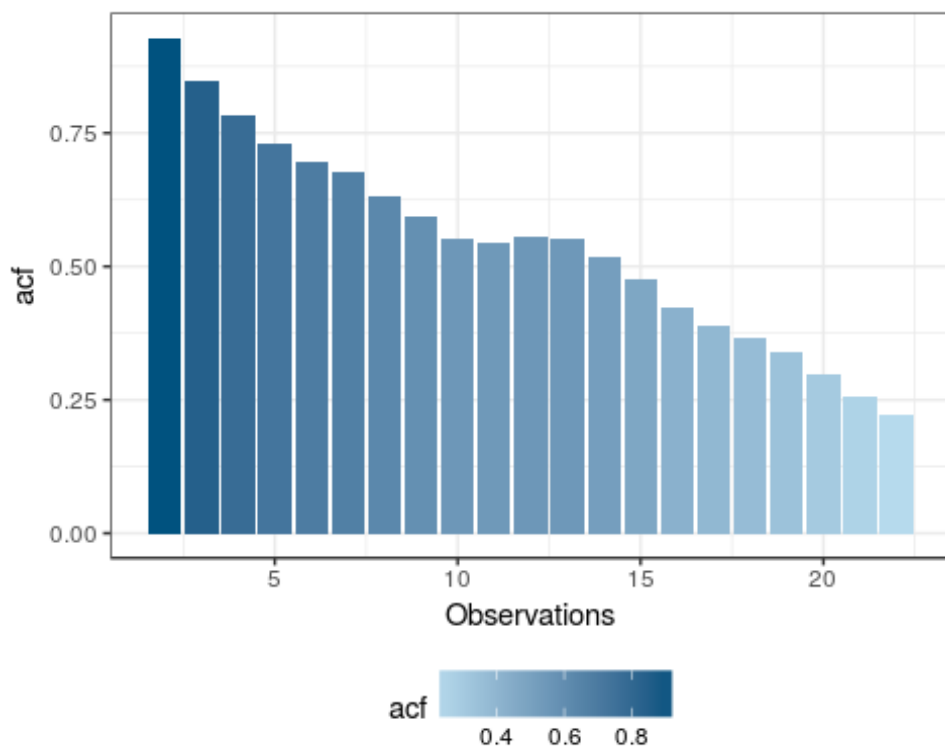
Standard Deviation: 561.46

Coefficient of Variation: 13.03%

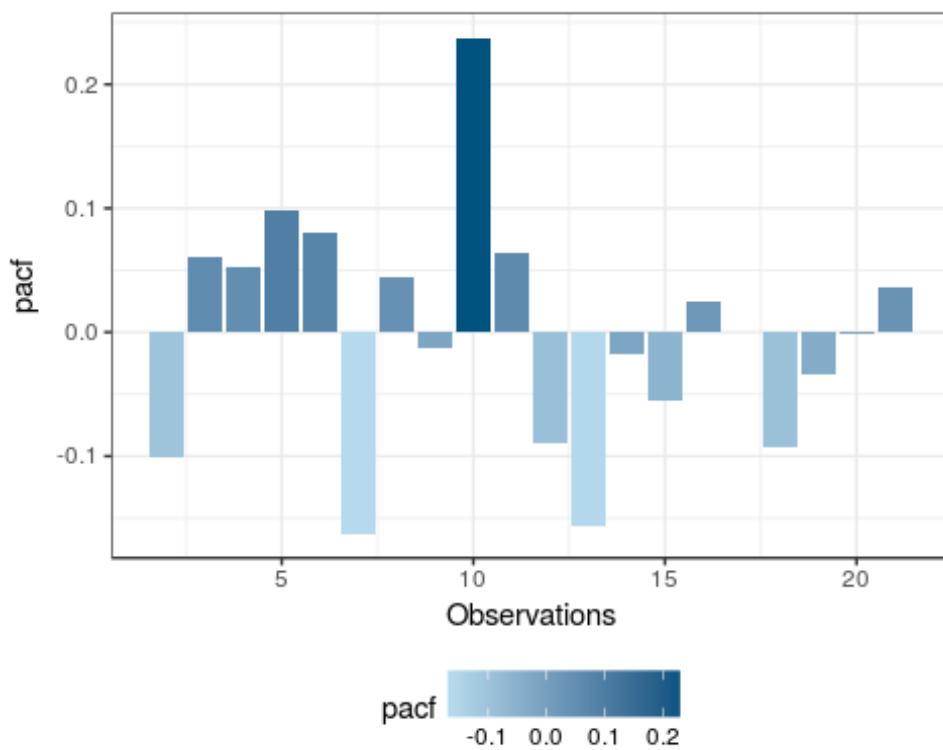
Variance: $3.152403910^{\{5\}}$

Linear Correlation Coefficient: -0.82

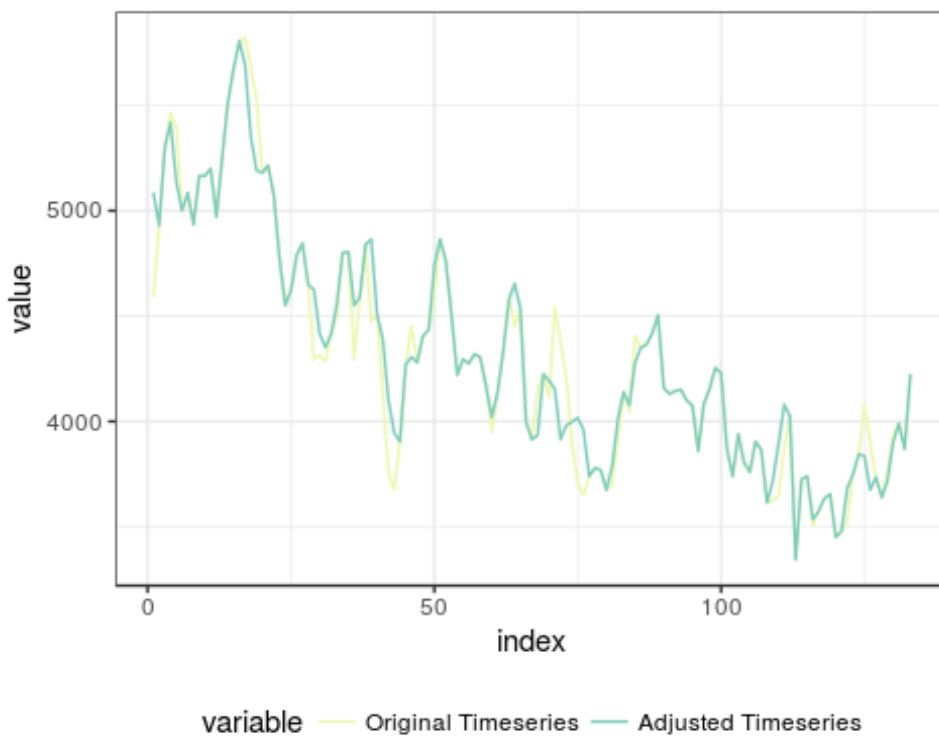
ACF Plot



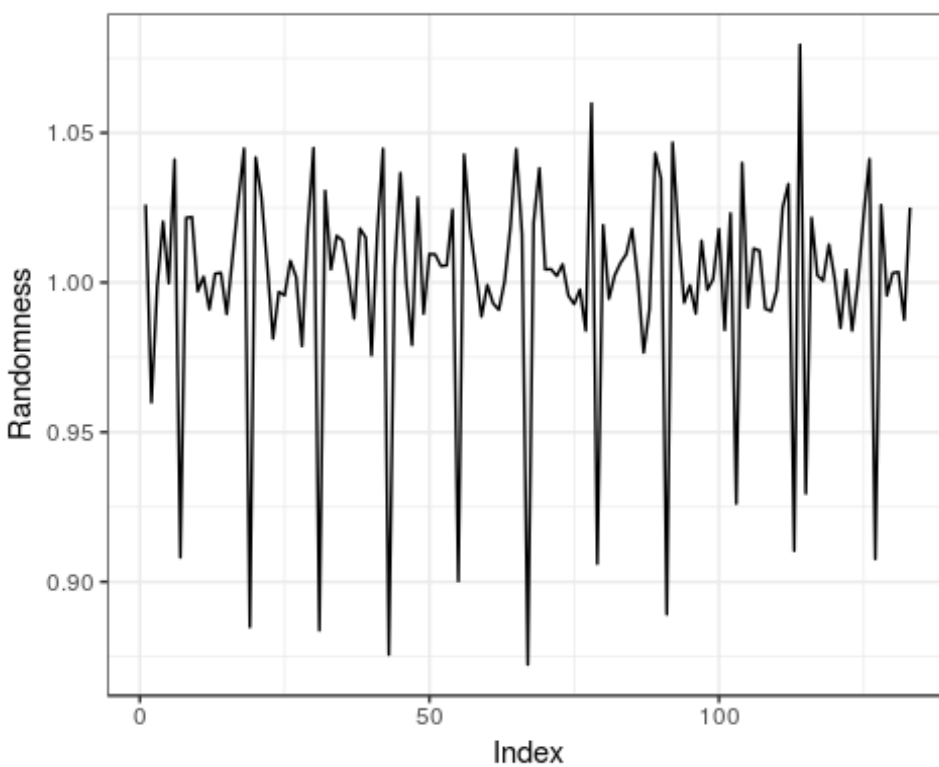
PACF Plot



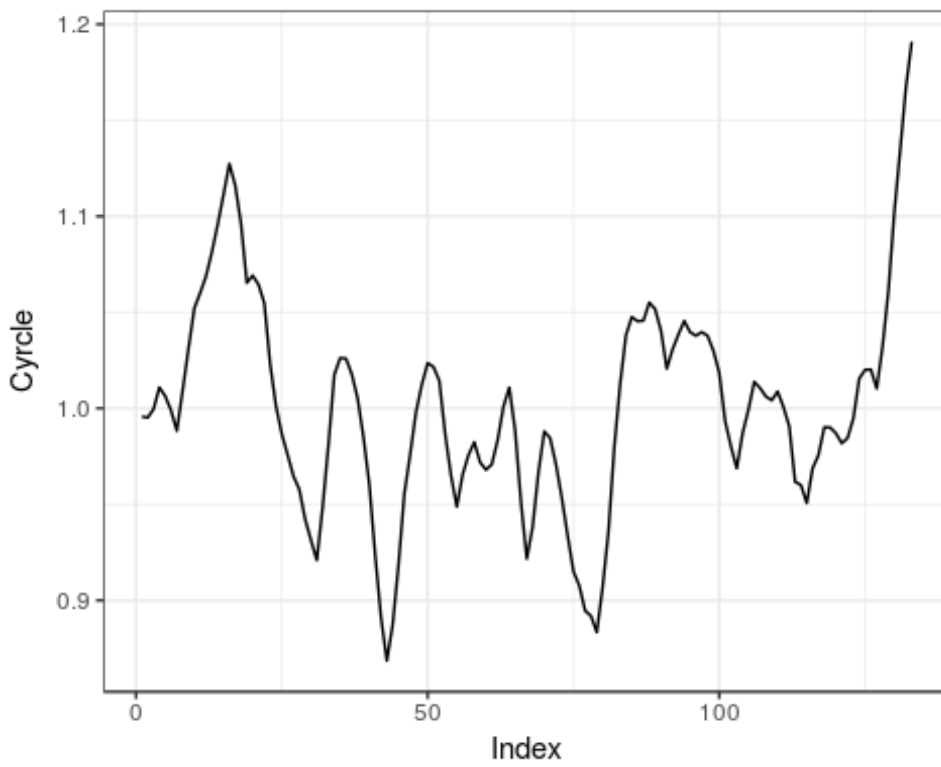
Outliers



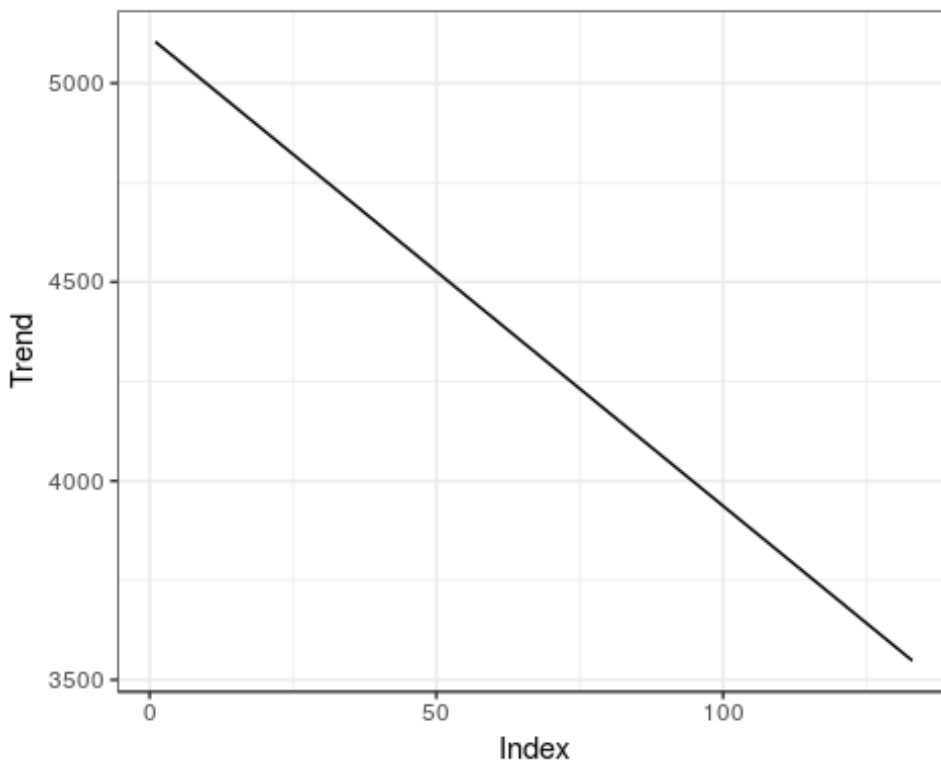
Decompose Randomness



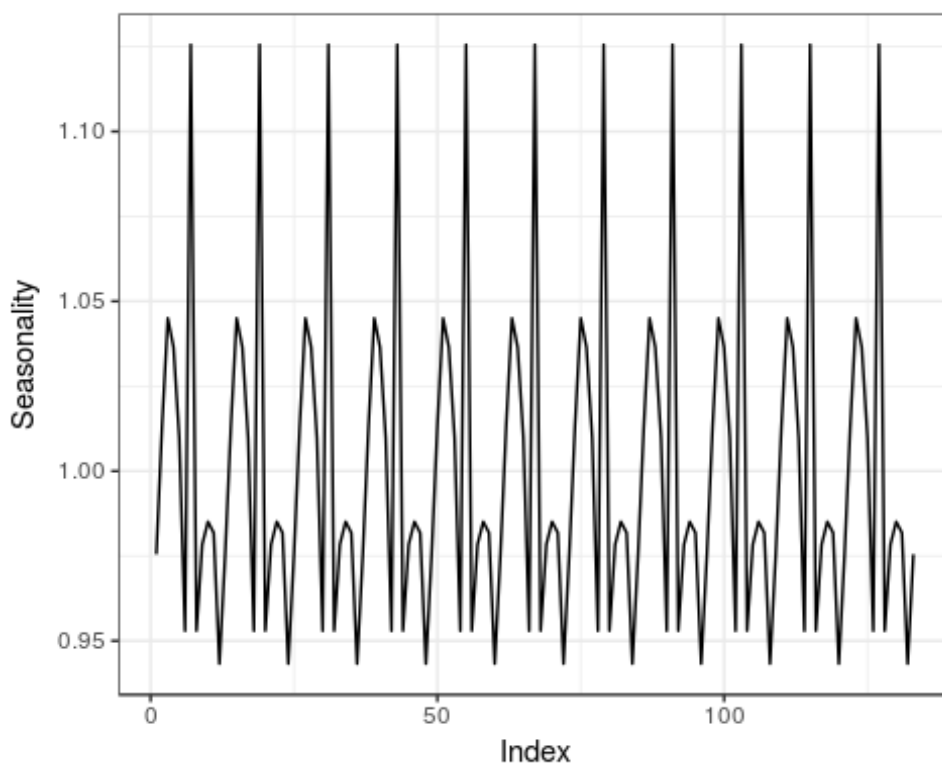
Cycle



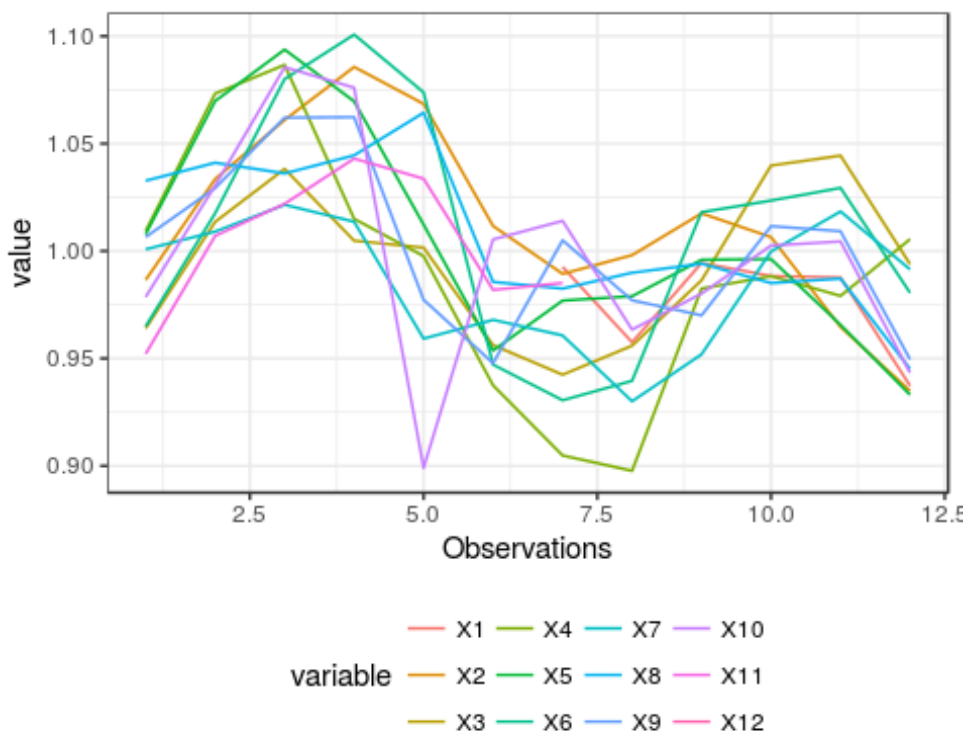
Trend



Seasonality

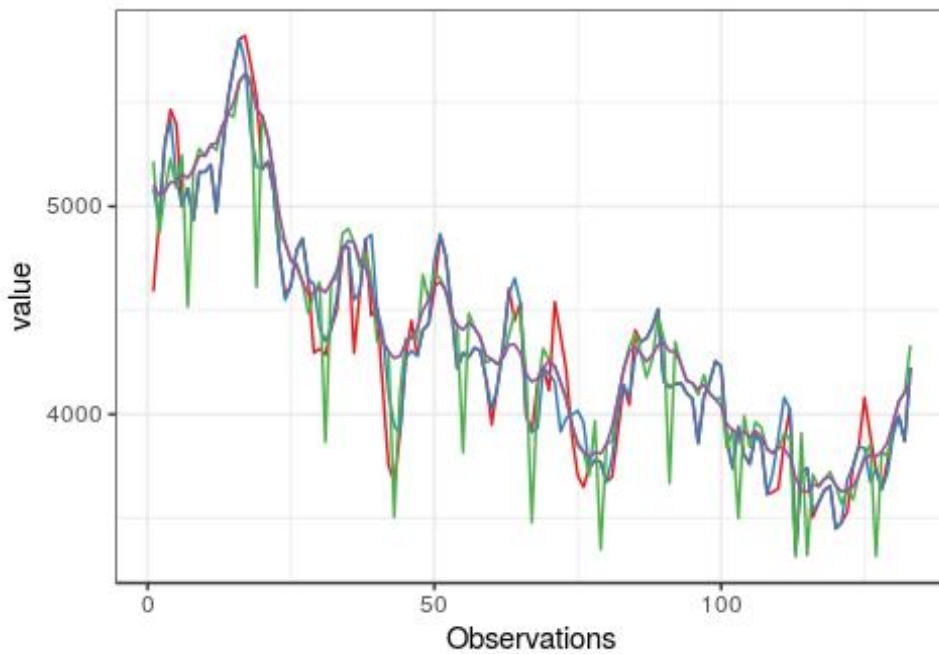


Seasonplot

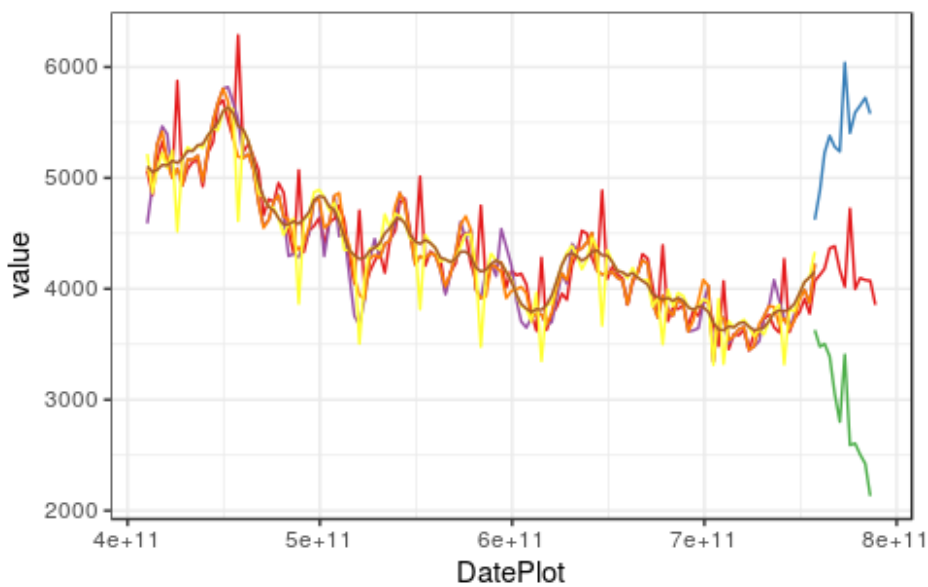


Smoothing

Chosen Smoothing Method: Non Linear Smooth

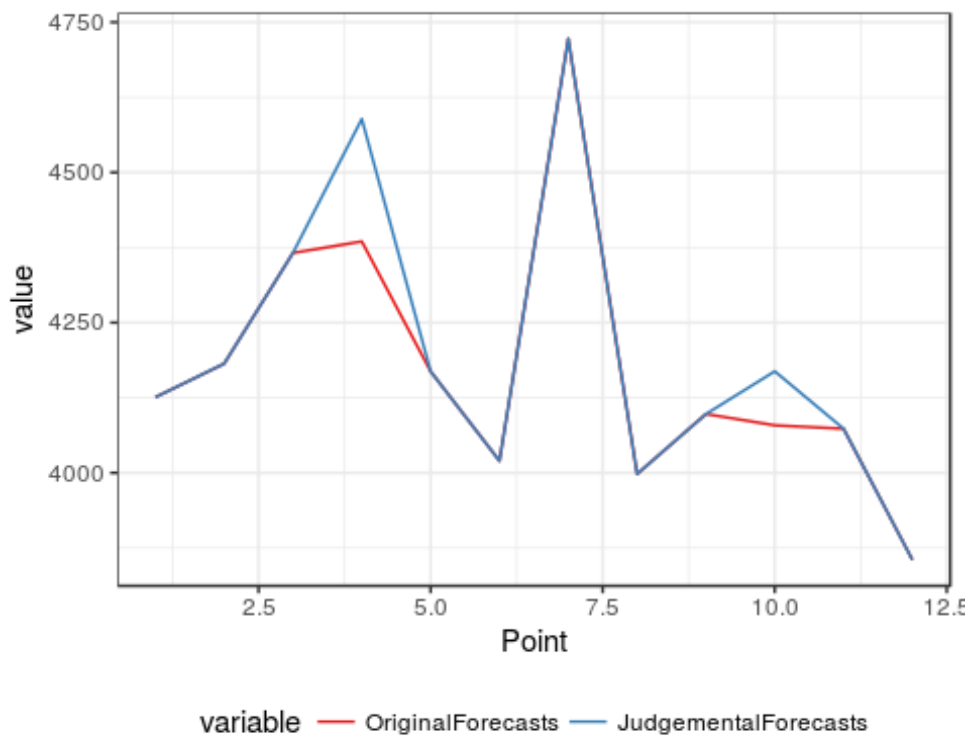


Forecasts



E. Value Forecast G. Confidence Level 2 B. Timeseries Without Outliers —
 F. Confidence Level 1 A. Original Timeseries C. Deseasonalized Timeseries

Judgemental Adjustments

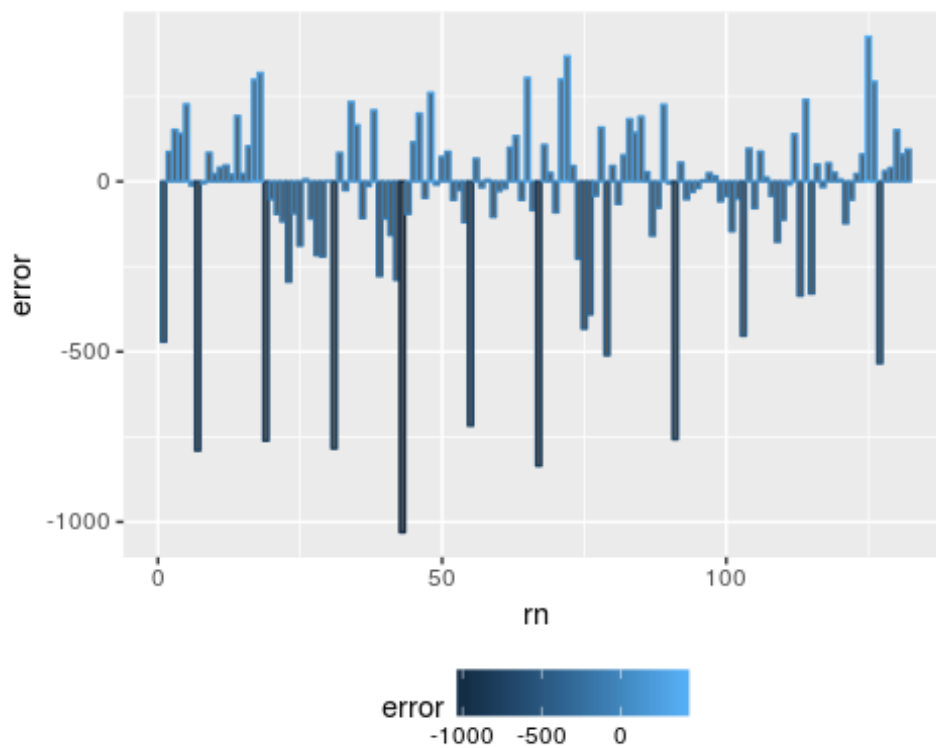
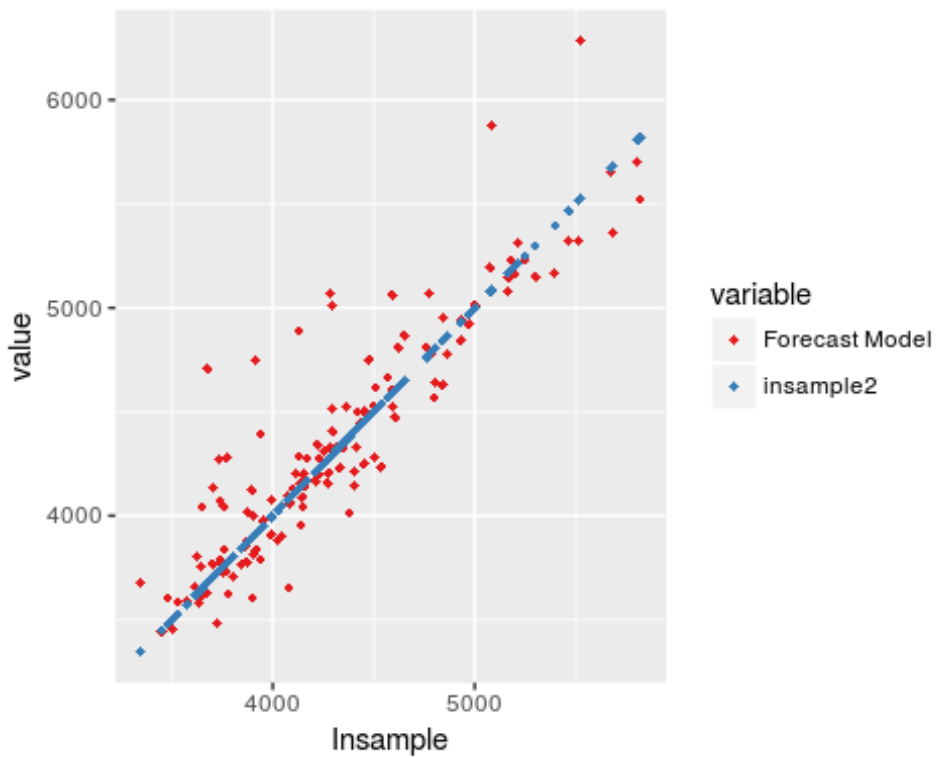


Errors

Insample-Forecast Model Errors

ME	MPE	MAPE	MSE	sMAPE
-45.07	-1.19 %	3.85 %	64114.81	3.74 %

Errors - Visualization



The imported time-series has no out-of-sample data to compare with the forecasts.

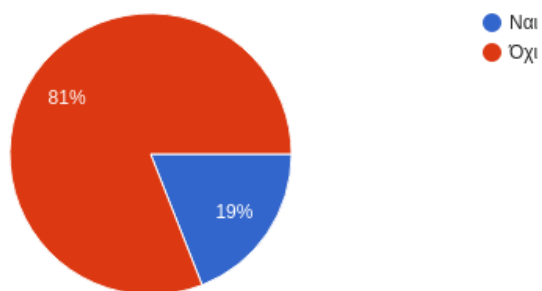
Κεφάλαιο 6

Αξιολόγηση Συστήματος

Για την αξιολόγηση του συστήματος δόθηκε ένα ερωτηματολόγιο σε 21 φοιτητές προπτυχιακού και μεταπτυχιακού επιπέδου που τους ζητήθηκε να χρησιμοποιήσουν τα συστήματα Open και SAS για την παραγωγή προβλέψεων σε τρεις χρονοσειρές. Οι χρονοσειρές αυτές ήταν αντίστοιχα, μία διακοπτόμενης ζήτησης, μία εποχιακή και μία με πολλές ακραίες τιμές, για να μπορέσουν να αξιολογηθούν αποτελεσματικά όλες οι δυνατότητες των συστημάτων. Οι απαντήσεις των φοιτητών συνοψίζονται παρακάτω.

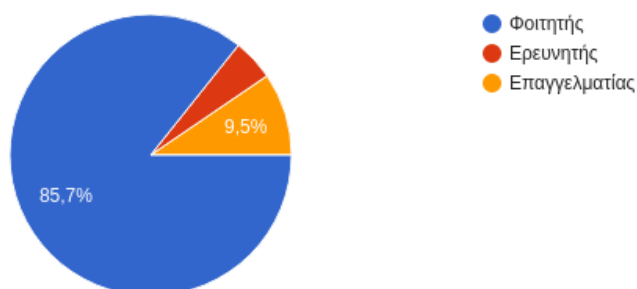
Έχετε κάποιο προηγούμενο statistical background;

21 απαντήσεις



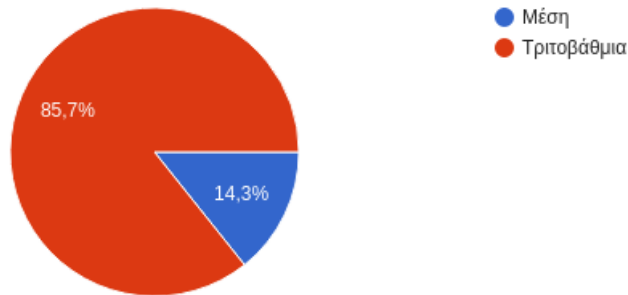
Ιδιότητα

21 απαντήσεις

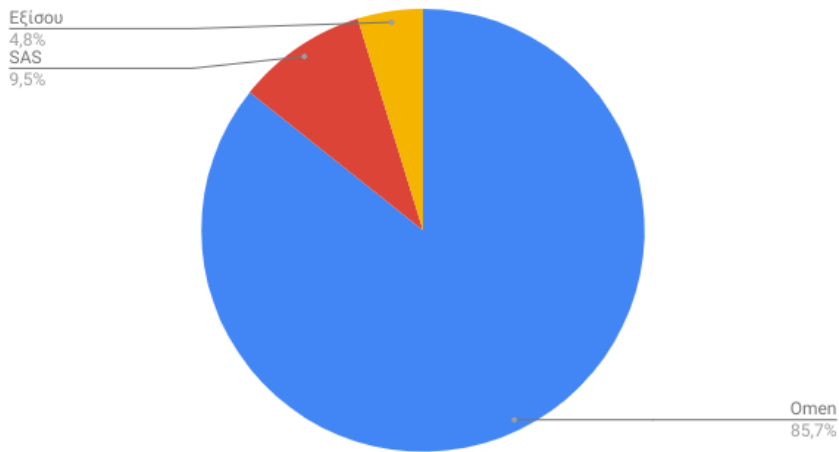


Εκπαίδευση

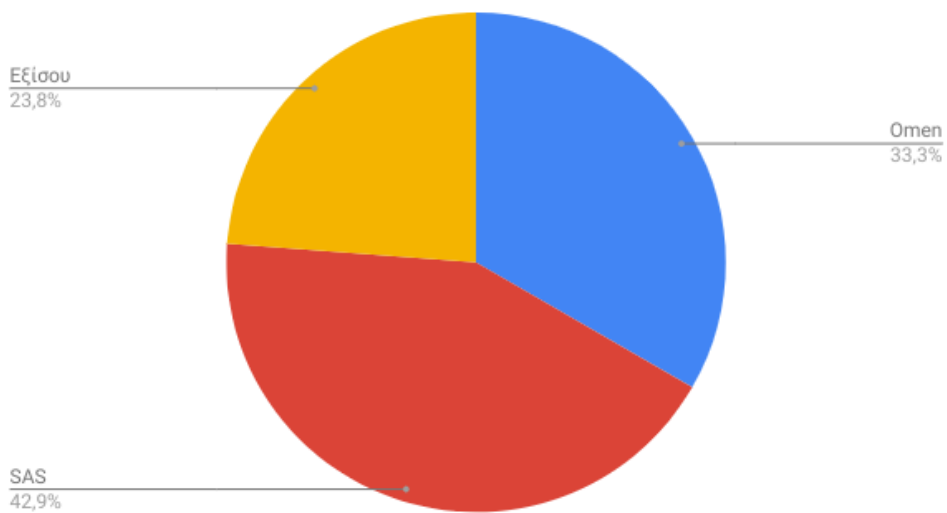
21 απαντήσεις



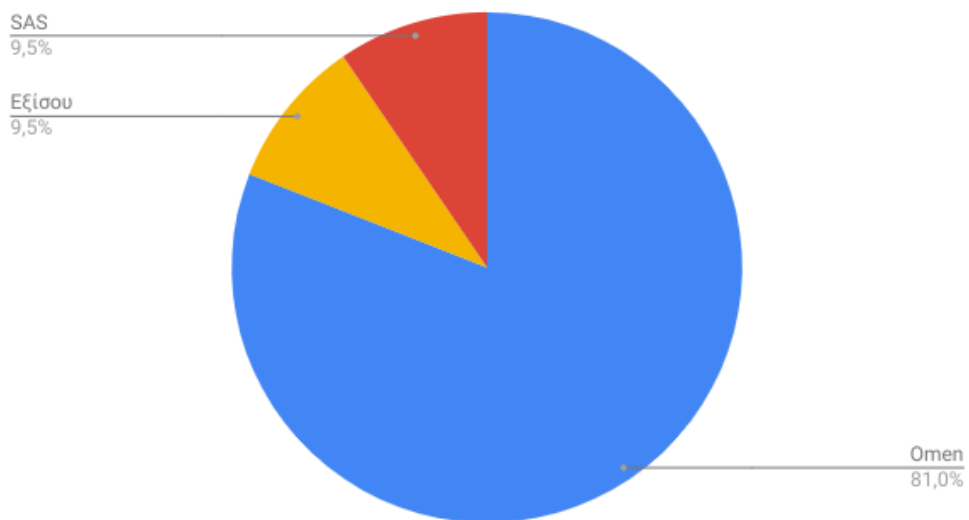
Με ποιο σύστημα εξοικειωθήκατε πιο γρήγορα και εύκολα;



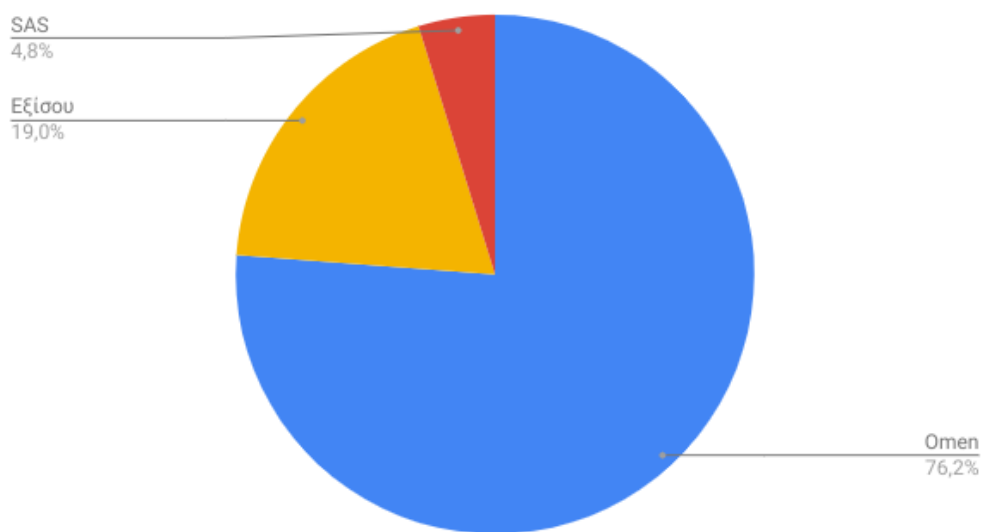
Ποιο σύστημα είχε καλύτερη υποστήριξη (documentation);



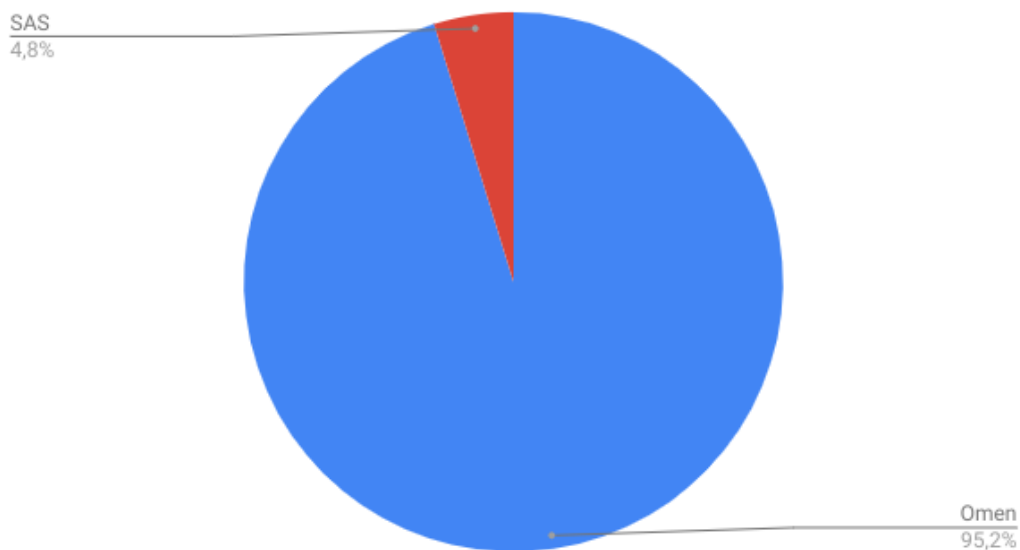
Ποιό σύστημα είχε πιο κατανοητή και λογική δομή;



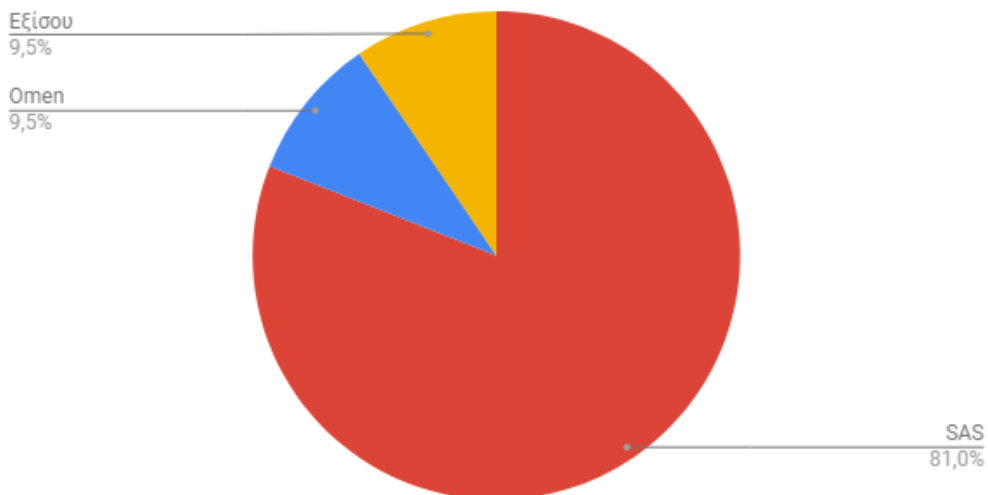
Ποιο σύστημα έδινε καλύτερες δυνατότητες πλοήγησης;



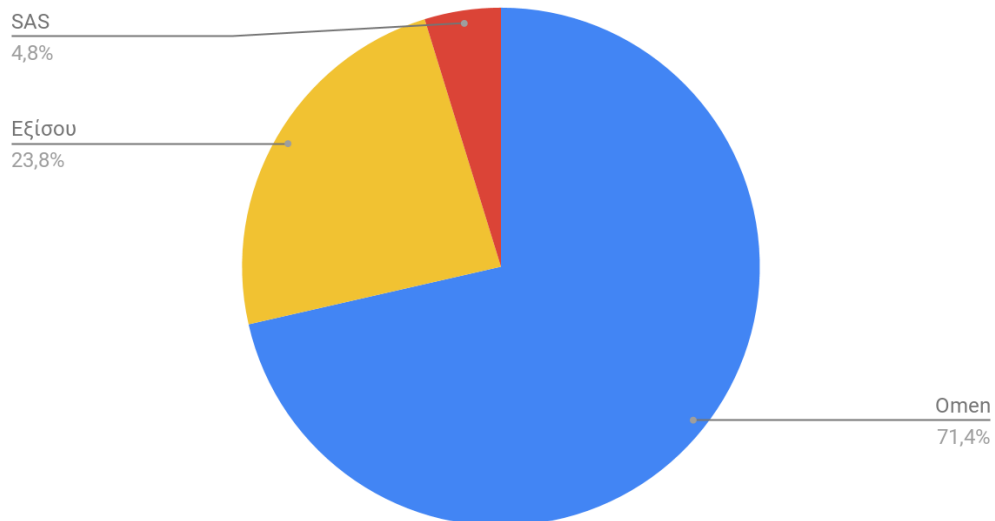
Ποιό σύστημα ήταν πιο ελκυστικό στη χρήση;



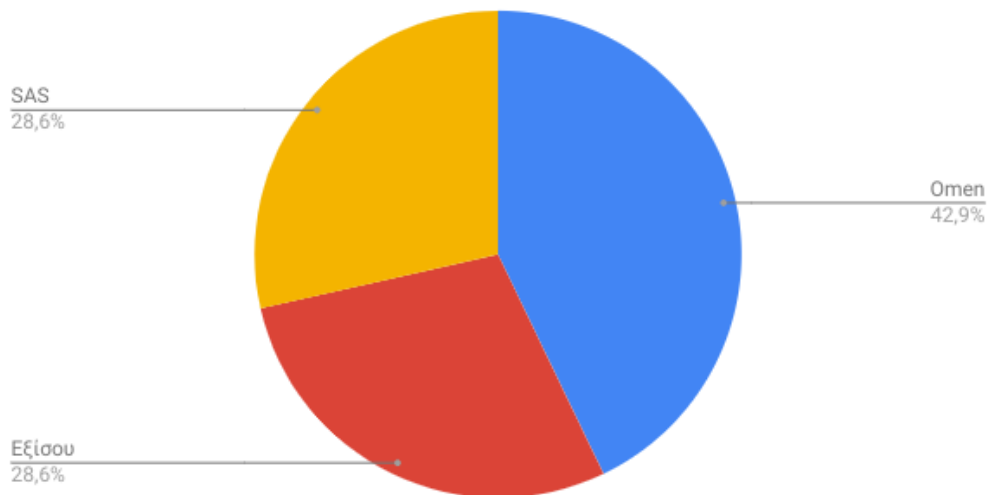
Σε ποιό σύστημα βρεθήκατε περισσότερες φορές να μην ξέρετε πως να προχωρήσετε;



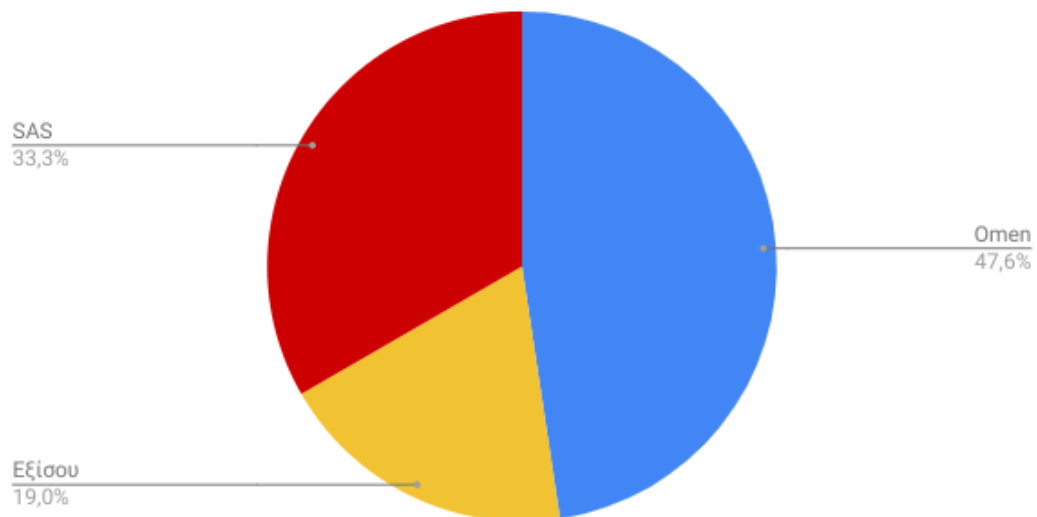
Ποιό σύστημα θα προτείνατε σε κάποιον συνάδελφο - φίλο;



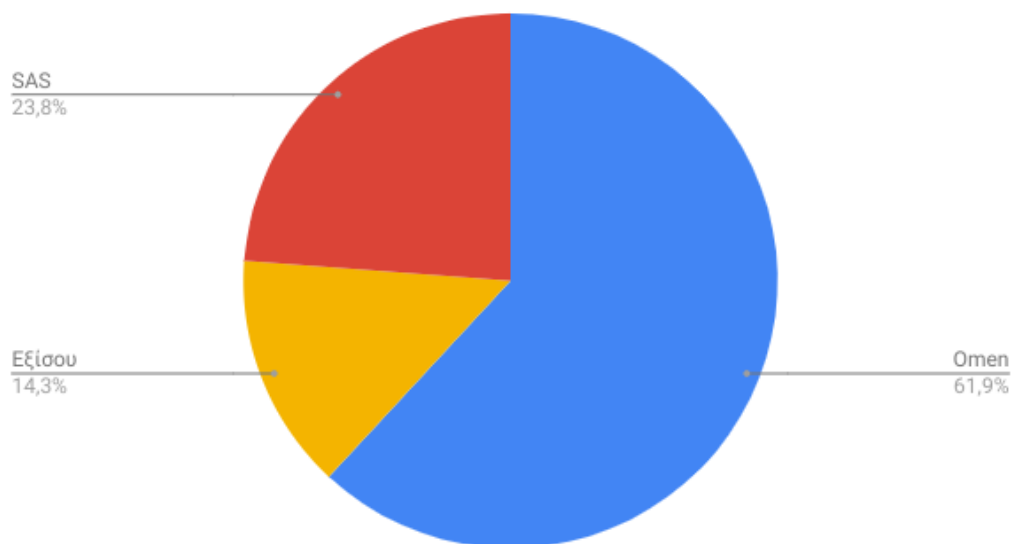
Με ποιό σύστημα νιώθατε μεγαλύτερη ασφάλεια κατά την παραγωγή των προβλέψεων;



Με ποιά σύστημα είχατε καλύτερη επίβλεψη της διαδικασίας πρόβλεψης;



Ποιό σύστημα είχε καλύτερες απεικονίσεις και γραφικά;



Συνολικά, η πλειοψηφία έκρινε ότι το Omen ήταν το πιο φιλικό προς τον χρήστη σύστημα με την πιο κατανοητή δομή. Ανάμεσα στα σχόλια των φοιτητών υπήρχαν πολλά που ανέφεραν προβλήματα σχετικά με την εγκατάσταση και εκτέλεση του SAS σε σύγκριση με το Omen που θεωρήθηκε πιο εύχρηστο και ‘ελαφρύ’ πρόγραμμα.

Πιο συγκεκριμένα, από το παρακάτω διάγραμμα φαίνεται ότι τα μειονεκτήματα του Omen εντοπίζονται ειδικά στην αξιολόγηση των προβλέψεων, όπου οι χρήστες έχουν δυσκολίες στην κατανόηση της out-of-sample αξιολόγησης που παρέχει το σύστημα, και στην αναφορά, που επεκτείνεται το προηγούμενο πρόβλημα. Επίσης, παρατηρήθηκε ότι η τεκμηρίωση των επιλογών του συστήματος δεν ήταν επαρκής σε ορισμένα σημεία και θα πρέπει να βελτιωθεί.

Η υπεροχή του συστήματος μας, εντοπίζεται στα σημεία της ολικής επίβλεψης της διαδικασίας, στις δυνατότητες επεξεργασίας, στην δομή, στις γραφικές απεικονίσεις κάθε βήματος. Το μεγαλύτερο προσόν του συστήματος είναι το πόσο ελκυστικό είναι για τον χρήστη, αφού εκεί το Omen είχε την προτίμηση όλων των φοιτητών.

Πέρα από την χρήση του συστήματος από τους φοιτητές, έχει γίνει προσπάθεια να προωθηθεί και σε επιχειρήσεις. Ενδεικτικά, μία έκδοση του έχει χρησιμοποιηθεί από την ΕΥΔΑΠ, όπου το σύστημα λειτούργησε πολύ αποτελεσματικά για παραγωγή προβλέψεων στη ζήτηση νερού. Επίσης, το Omen χρησιμοποιήθηκε για την παραγωγή προβλέψεων για τον πληθυσμό στον διαγωνισμό FabSpace HackOnEarth, όπου οι προβλέψεις αξιολογήθηκαν ως οι ακριβέστερες τόσο σε εθνικό επίπεδο, όσο και σε ευρωπαϊκό επίπεδο.



Σύνοψη

Τα συστήματα υποστήριξης προβλέψεων χρησιμοποιούνται στην καθημερινή ζωή για να βοηθήσουν τις διαδικασίες πρόβλεψης και να στηρίζουν σημαντικές αποφάσεις των εταιρειών. Παρά τη βελτίωση τους κατά τη διάρκεια των τελευταίων ετών, εξακολουθούν να παρατηρούνται μειονεκτήματα τους σε τρεις βασικές πτυχές: λύσεις που δεν βασίζονται στον ιστό, έλλειψη προσαρμοστικότητας και πολύπλοκες διεπαφές χρήστη.

Στην παρούσα εργασία παρουσιάζεται το OMEN, ένα πλήρως προσαρμόσιμο web-based FSS με μια σύγχρονη και αποδοτική διεπαφή που υποστηρίζει μεθόδους και μοντέλα τελευταίας τεχνολογίας. Μέσω του OMEN, επιδεικνύουμε τους περιορισμούς των τυπικών λύσεων και εξηγούμε πώς μπορούν να χρησιμοποιηθούν ορισμένες από τις δυνατότητές του για τη βελτίωση των σύγχρονων FSS. Όπως παρουσιάστηκαν, οι τυπικές λύσεις είναι οι εφαρμογές που βασίζονται σε τοπική εγκατάσταση, οι οποίες περιορίζουν τη διαθεσιμότητα, τη συμβατότητα του συστήματος, αυξάνοντας ταυτόχρονα την πολυπλοκότητα και τη χρήση των πόρων. Επιπλέον, δεδομένης της αρχιτεκτονικής και της δομής τους, τα FSS είναι δύσκολο να προσαρμοστούν και καταλήγουν να αποτελούν άκαμπτο και γενικό λογισμικό που δεν στοχεύει στις ειδικές ανάγκες κάθε εταιρείας και τελικά αποδεικνύεται ότι είναι αρκετά αναποτελεσματικό. Αυτό οδηγεί επίσης σε μια σύγχυση με πολύπλοκες επιλογές, μοντέλα και εργαλεία που καθιστούν δύσκολη τη χρήση του εργαλείου ακόμη και για έμπειρους χρήστες.

Τα προβλήματα εντοπίζονται επίσης στη μεθοδολογία, συμπεριλαμβανομένων των παρωχημένων μεθόδων και της έλλειψης της κριτικής συνιστώσας κατά τη διαδικασία πρόβλεψης. Με την εκμετάλλευση λύσεων ανοιχτού κώδικα υποδεικνύουμε πώς μπορεί να αναπτυχθεί ένα ισχυρό FSS, όπως το OMEN που προσφέρει πολλά πλεονεκτήματα τόσο τεχνολογικά όσο και μεθοδολογικά. Χρησιμοποιώντας το Shiny και την R, το OMEN αποτελεί ένα προσαρμόσιμο και εύκολο στην χρήση εργαλείο που βασίζεται στο Web, το οποίο επιτρέπει ταυτόχρονα την πρόσβαση των πολλαπλών χρηστών 24 ώρες το 24ωρο από τη συσκευή της επιλογής τους. Η απλοποιημένη διεπαφή που οργανώνεται σε περιεκτικά και καλά τεκμηριωμένα στοιχεία μενού διευκολύνει τη χρήση για τους επαγγελματίες, ενώ οι πλέον σύγχρονοι αλγόριθμοι, όπως η χρονική συνάθροιση, ο χειρισμός χρονοσειρών διακοπτόμενης ζήτησης και τα μοντέλα πρόβλεψης τόσο για συνεχή όσο και για διακοπτόμενα δεδομένα, διατίθενται εύκολα χρησιμοποιώντας τις state-of-the-art τεχνολογίες που παρέχονται από ακαδημαϊκούς και ερευνητές στην κοινότητα της R.

Μελλοντικές Προεκτάσεις

Αφού διορθωθούν οι ελλείψεις που παρατηρήθηκαν από την αξιολόγηση που πραγματοποιήθηκε από τους φοιτητές, το μελλοντικό μας έργο επικεντρώνεται στην περαιτέρω προώθηση του OMEN και στην προσέλκυση σύγχρονων FSS. Αυτό περιλαμβάνει την εκμετάλλευση λύσεων βασισμένων σε cloud για τη μείωση των χρόνων εκτέλεσης σε έντονο φόρτο εργασίας και την παροχή οικονομικών λύσεων για επιχειρήσεις χωρίς υποδομές. Σημειώνεται ότι το Cloud Computing είναι μια δομή με την οποία μάς δίνεται η δυνατότητα να έχουμε πρόσβαση και να χρησιμοποιούμε web εφαρμογές χωρίς να τις διαθέτουμε στον υπολογιστή μας ή σε κάποια άλλη συσκευή που είναι συνδεδεμένη με το Internet. Με άλλα λόγια το Cloud Computing σημαίνει μεγάλα κέντρα δεδομένων, τα οποία προσφέρουν οικονομίες κλίμακας, φθηνότερη υπολογιστική ισχύ και, κυρίως, την ευελιξία να πληρώνει κανείς μόνο για ό,τι χρησιμοποιεί. Επίσης η έρευνα διαπιστώνει ότι επιχειρήσεις που κάνουν ισορροπημένη χρήση του Cloud διαθέτουν την απαραίτητη ευελιξία που τους επιτρέπει να ανταποκρίνονται ταχύτερα στις ανάγκες των πελατών τους, ενώ ταυτόχρονα έχουν τη δυνατότητα να αυξήσουν τα έσοδά τους.

Θα πρέπει επίσης να εξεταστούν μέθοδοι ομαδοποίησης των δεδομένων για την αποτελεσματική αντιμετώπιση των ιεραρχικών χρονοσειρών. Πολλές εφαρμογές απαιτούν προβλέψεις για μεγάλες συλλογές σχετικών χρονικών σειρών που μπορεί να είναι οργανωμένες σε ιεραρχική δομή. Για παράδειγμα, οι πωλήσεις μιας πολυεθνικής εταιρείας μπορούν να χωριστούν σε μια γεωγραφική ιεραρχία σε χώρες, περιοχές, πόλεις και καταστήματα. Η εταιρεία συνήθως χρειάζεται προβλέψεις τόσο για τις συνολικές πωλήσεις, όσο και για τις εθνικές πωλήσεις, τις περιφερειακές πωλήσεις και τις πωλήσεις για καθένα μεμονωμένο κατάστημα, και αυτές οι προβλέψεις θα πρέπει να υπολογιστούν κατάλληλα σε όλη την ιεραρχία. Έχει γίνει ήδη μία προσπάθεια ενσωμάτωσης του ιεραρχικού στοιχείου στο σύστημα, αλλά θα πρέπει να επεκταθεί παραπάνω για να μπορεί να χρησιμοποιηθεί αποτελεσματικά.

Τέλος, η δυνατότητα προσθήκης κριτικών προβλέψεων μέσω του συστήματος είναι απαραίτητη. Κάτι τέτοιο μπορεί να γίνει με την ενσωμάτωση μεθοδολογιών όπως οι δομημένες αναλογίες ή η Delphi στο σύστημα.

Βιβλιογραφία

01. Allaire, J., Cheng, J., Xie, Y., McPherson, J., Chang, W., Allen, J., Wickham, H., Atkins, A., Hyndman, R., RStudio, jQuery Foundation, jQuery contributors, Otto, M., Thornton, J., Bootstrap contributors, Twitter, Farkas, A., Jehl, S., Sagalaev, I., MacFarlane, J., Google, Raggett, D., W3C, 2016.rmarkdown: Dynamic documents for r. R package, version 0.9.2.
02. Andrawis, R. R., Atiya, A. F., El-Shishiny, H., 2011. Combination of long term and short term forecasts,with application to tourism demand forecasting. *International Journal of Forecasting* 27 (3), 870 – 886.
03. Assimakopoulos, V., 1995. A successive filtering technique for identifying long- term trends. *Journal of Forecasting*, John Wiley 14 (1), 35 – 43.
04. Assimakopoulos, V., Nikolopoulos, K., 2000. The theta model: a decomposition approach to forecasting.*International Journal of Forecasting* 16 (4), 521 – 530.
05. Beaumont, A. N., 2014. Data transforms with exponential smoothing methods of forecasting. *International Journal of Forecasting* 30 (4), 918 – 927.
06. Chang, W., RStudio, Almasaeed Studio, Adobe Systems Incorporated, 2015. shinydashboard: Create dashboards with shiny. R package, version 0.5.1.
07. Chen, C., Liu, L.-M., 1993. Joint estimation of model parameters and outlier effects in time series.*Journal of the American Statistical Association* 88 (421), 284 – 297.
08. Croston, J., 1972. Forecasting and stock control for intermittent demand. *Operational Research Quarterly* 23 (1), 289 – 303.
09. Davey, A., Flores, B., 1993. Identification of seasonality in time series: A note. *Mathematical and Computer Modelling* 18 (6), 73 – 81.
10. Davydenko, A., Fildes, R., 2013. Measuring forecasting accuracy: The case of judgmental adjustments to sku-level demand forecasts. *International Journal of Forecasting* 29 (3), 510 – 522.
11. de Lacalle, J. L., 2015. Detection of outliers in time series. Version: 0.6, <http://jalobe.com>.
12. Fildes, R., Petropoulos, F., 2015. Simple versus complex selection rules for forecasting many time series.*Journal of Business Research* 68 (8), 1692 – 1701.
13. Franses, P. H., 2016. A note on the mean absolute scaled error. *International Journal of Forecasting* 32 (1), 20 – 22.
14. Gardner, E. S., 1985. Exponential smoothing: the state of the art. *Journal of Forecasting* 4 (1), 1 – 28.
15. Gardner, E. S., 2006. Exponential smoothing: The state of the artpart II. *International Journal of Forecasting* 22 (4), 637 – 666.
16. Guerrero, V. M., 1993. Time-series analysis supported by power transformations. *Journal of Forecasting* 12 (1), 37 – 48.
17. Hyndman, R., 2015. forecast: Forecasting functions for time series and linear models. R package version 5.8, <http://github.com/robjhyndman/forecast>.
18. Hyndman, R., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for r. *Journal of Statistical Software* 26 (3), 1 – 22.
19. Hyndman, R., Khandakar, Y., 2008. Automatic time series forecasting: the forecast package for R. *Journal of Statistical Software* 26 (3), 1 – 22.

20. Hyndman, R. J., Athanasopoulos, G., Bergmeir, C., Cinelli, C., Khan, Y., Mayer, Z., Razbash, S., Schmidt, D., Shaub, D., Tang, Y., Wang, E., Zhou, Z., 2015. forecast: Forecasting functions for timeseries and linear models. Version 6.2.
21. Hyndman, R. J., Koehler, A. B., 2006. Another look at measures of forecast accuracy. *International Journal of Forecasting* 22 (4), 679 – 688.
22. 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 (3), 439 – 454.
23. Kendall, M., Stuart, A., 1983. *The advanced theory of statistics* 3, 410–414.
24. Kourentzes, N., 2014. On intermittent demand model optimisation and selection. *International Journal of Production Economics* 156, 180 – 190.
25. Kourentzes, N., Petropoulos, F., 2014. MAPA: Multiple Aggregation Prediction Algorith. R package version 1.9. <http://CRAN.R-project.org/package=MAPA>.
26. Kourentzes, N., Petropoulos, F., 2014. Multiple Aggregation Prediction Algorithm, Version: 1.9, <http://kourentzes.com/forecasting/2014/04/19/multiple-aggregation-prediction-algorithm-mapa/>.
27. Kourentzes, N., Petropoulos, F., 2015. Forecasting with multivariate temporal aggregation: The case of promotional modelling. *International Journal of Production Economics*.
28. Kourentzes, N., Petropoulos, F., 2015. Intermittent time series forecasting. r package version 1.8.
29. Kourentzes, N., Petropoulos, F., Trapero, J. R., 2014. Improving forecasting by estimating time series structural components across multiple frequencies. *International Journal of Forecasting* 30 (2), 291 – 302.
30. Ledolter, J., 1989. The effect of additive outliers on the forecasts from arima models. *International Journal of Forecasting* 5 (2), 231 – 240.
31. Makridakis, S., Wheelwright, S. C., McGee, V. E., 1983. *Forecasting: Methods and applications* (Second Edition). New York: Wiley.
32. McKinney, W., 2010. Data structures for statistical computing in python. *Proceedings of the 9th Python in Science Conference*, 51 – 56.
33. Nadaraya, E. A., 1964. On estimating regression. *Theory of Probability and its Applications* 9 (1), 141 2.
34. Petropoulos, F., 2015. Forecasting support systems: Ways forward. *Foresight: The International Journal of Applied Forecasting*, issue 39, 5 – 11.
35. Petropoulos, F., Kourentzes, N., Jun 2015. Forecast combinations for intermittent demand. *J Oper Res Soc* 66 (6), 914–924.
36. Proietti, T., Ltkepohl, H., 2013. Does the boxcox transformation help in forecasting macroeconomic time series? *International Journal of Forecasting* 29 (1), 88 – 99.
37. R Development Core Team, 2008. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. ISBN 3-900051-07-0, <http://www.R-project.org>.
38. Sakamoto, Y., I. M., G., K., 1986. Akaike information criterion statistics. D. Reidel Publishing Company.
39. Syntetos, A. A., Boylan, J. E., 2005. The accuracy of intermittent demand estimates. *International Journal of Forecasting* 21 (2), 303 – 314.
40. Tashman, L. J., 2000. Out-of-sample tests of forecasting accuracy: an analysis and review. *International Journal of Forecasting* 16 (4), 437 – 450.

41. Teunter, R. H., Syntetos, A. A., Babai, M. Z., 2011. Intermittent demand: Linking forecasting to inventory obsolescence. *European Journal of Operational Research* 214 (3), 606 – 615.
42. Vaidyanathan, R., 2013. rCharts: Interactive Charts using Javascript Visualization Librarie. R package version 0.4.5.
43. Watson, G. S., 1964. Smooth regression analysis. *Sankhy: The Indian Journal of Statistics, Series A* 26 (4), 359 372

