



ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ

ΣΧΟΛΗ ΤΕΧΝΟΛΟΓΙΩΝ ΠΛΗΡΟΦΟΡΙΚΗΣ ΚΑΙ ΤΗΛΕΠΙΚΟΙΝΩΝΙΩΝ

ΤΜΗΜΑ ΨΗΦΙΑΚΩΝ ΣΥΣΤΗΜΑΤΩΝ

ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ

ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ ΚΑΙ ΥΠΗΡΕΣΙΕΣ

ΚΑΤΕΥΘΥΝΣΗ: ΠΡΟΗΓΜΕΝΑ ΠΛΗΡΟΦΟΡΙΑΚΑ ΣΥΣΤΗΜΑΤΑ

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**Μηχανισμός έκδοσης αναφορών και παρακολούθησης συστήματος ηλεκτρονικών
προμηθειών ESPD.**

Σωτήριος Ν. Καραγεωργόπουλος

Επιβλέπουσα:

Ανδριάνα Πρέντζα, Καθηγήτρια

ΠΕΙΡΑΙΑΣ

Φεβρουάριος 2023

ΜΕΤΑΠΤΥΧΙΑΚΗ ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

Μηχανισμός έκδοσης αναφορών και παρακολούθησης συστήματος ηλεκτρονικών προμηθειών ESPD.

Σωτήριος Ν. Καραγεωργόπουλος

A.M.: ME2137

ΠΕΡΙΛΗΨΗ

Η παρούσα διπλωματική εργασία πραγματεύεται τη σχεδίαση και την υλοποίηση ενός μηχανισμού έκδοσης αναφορών και παρακολούθησης της λειτουργίας του συστήματος ESPD (European Single Procurement Document – Ευρωπαϊκό Ενιαίο Έγγραφο Σύμβασης). Πιο συγκεκριμένα, θα παρουσιαστεί η καταγραφή των εξαγόμενων εγγράφων ESPD σε αρχεία καταγραφής (log). Επιπλέον, θα αναφερθούν οι τεχνικές ανάλυσης δεδομένων για τη δημιουργία προβλέψεων που εμφανίζονται στην αναφορά του ESPD συστήματος. Τέλος, θα γίνει περιγραφή της αρχιτεκτονικής του μηχανισμού έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD και των μερών που το απαρτίζουν. Ο σκοπός εκπόνησης της παρούσας διπλωματικής εργασίας είναι η σχεδίαση και η υλοποίηση ενός συστήματος επιχειρησιακής ευφυΐας (business intelligence) που θα προσφέρει χρήσιμες πληροφορίες για διάφορες πτυχές της λειτουργίας του συστήματος ESPD και θα αποτελεί ένα χρήσιμο εργαλείο για τη λήψη αποφάσεων.

ΘΕΜΑΤΙΚΗ ΠΕΡΙΟΧΗ: Επεξεργασία και ανάλυση δεδομένων (data analytics) στις Ηλεκτρονικές Προμήθειες.

ΛΕΞΕΙΣ ΚΛΕΙΔΙΑ: Business Intelligence, Time Series Analysis, Prophet Model, ESPD, eProcurement, Reporting, Monitoring.

ΕΥΧΑΡΙΣΤΙΕΣ

Θα ήθελα να ευχαριστήσω την αγαπητή Καθηγήτρια Ανδριάννα Πρέντζα για την άψογη συνεργασία, καθώς και για την εμπιστοσύνη που μου δείχνει, κάτι το οποίο αποτελεί ιδιαίτερη τιμή για μένα. Τέλος, θα ήθελα να ευχαριστήσω τους γονείς μου για την αμέριστη στήριξη που μου παρέχουν για την εκπλήρωση των ακαδημαϊκών στόχων μου.

ΠΕΡΙΕΧΟΜΕΝΑ

ΠΕΡΙΛΗΨΗ	2
ΕΥΧΑΡΙΣΤΙΕΣ.....	3
ΠΕΡΙΕΧΟΜΕΝΑ	4
ΛΙΣΤΑ ΠΙΝΑΚΩΝ.....	9
ΠΙΝΑΚΑΣ ΣΥΝΤΟΜΕΥΣΕΩΝ	10
1. ΕΙΣΑΓΩΓΗ	12
1.1 Ορισμός του προβλήματος	12
1.2 Δομή της Διπλωματικής Εργασίας	12
1.3 Συνεισφορά της Διπλωματικής Εργασίας.....	12
2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ	14
2.1 Εισαγωγή.....	14
2.2 Επιχειρησιακή Ευφυΐα.....	14
2.2.1 Ορισμός Επιχειρησιακής Ευφυΐας.....	14
2.2.2 Οφέλη της Επιχειρησιακής Ευφυΐας για κάθε οργανισμό.....	14
2.2.3 Κατηγορίες Συστημάτων Επιχειρησιακής Ευφυΐας.....	15
2.2.4 Ηλεκτρονικές Προμήθειες και Επιχειρησιακή Ευφυΐα.....	16
2.2.5 Περιπτώσεις χρήσης της Επιχειρησιακής Ευφυΐας	17
2.3 Ανάλυση Χρονοσειρών	18
2.3.1 Ορισμός Χρονοσειράς και περιπτώσεις χρήσης.....	18
2.3.2 Τάση και Εποχικότητα Χρονοσειράς.....	19

2.3.3 Τεχνικές Εξομάλυνσης Χρονοσειρών.....	22
2.4 Μοντέλα Μηχανικής Μάθησης.....	24
2.4.1 Τι είναι «Μοντέλο Μηχανικής Μάθησης».....	24
2.4.2 Μεθοδολογία Δημιουργίας Μοντέλων Μηχανικής Μάθησης.....	25
2.4.3 Το αυτοπαλινδρομικό μοντέλο (Autoregressive Model)	26
2.4.4 Το μοντέλο κινητών μέσων (Moving Average Model)	28
2.4.5 Το μοντέλο ARIMA	28
2.4.6 Το μοντέλο Prophet.....	29

3. ΜΕΘΟΔΟΛΟΓΙΑ ΑΝΑΛΥΣΗΣ, ΣΧΕΔΙΑΣΗΣ ΚΑΙ ΑΝΑΠΤΥΞΗΣ ΣΥΣΤΗΜΑΤΟΣ 32

3.1 Εισαγωγή.....	32
3.2 Το μοντέλο μεταδεδομένων του «Ευρωπαϊκού Ενιαίου Έγγραφου Σύμβασης» (ESPD) για τη δημιουργία αρχείων καταγραφής.....	32
3.3 Μηχανισμός Αποθήκευσης Αρχείων Καταγραφής ESPD.....	33
3.4 Παρουσίαση της υπηρεσίας ανάλυσης δεδομένων των αρχείων καταγραφής του συστήματος ESPD (Analytics Service).....	37
3.5 Η αρχιτεκτονική του μηχανισμού έκδοσης αναφορών και παρακολούθησης σε πραγματικό χρόνο του συστήματος ESPD.....	42

4. ΥΛΟΠΟΙΗΣΗ ΣΥΣΤΗΜΑΤΟΣ 44

4.1 Εισαγωγή.....	44
4.2 Τεχνολογίες	44
4.2.1 Γλώσσες Προγραμματισμού	44
4.2.1.1 Java.....	44
4.2.1.2 Python.....	44
4.2.1.3 TypeScript	45
4.2.2 Βιβλιοθήκες	45
4.2.2.1 Angular	45

4.2.2.2 Apache ECharts	45
4.2.2.3 Flask.....	45
4.2.2.4 Pandas	46
4.2.2.5 Prophet	46
4.2.2.6 Apache FreeMarker	47
4.2.3 Βάσεις Δεδομένων	48
4.2.3.1 ksqIDB.....	48
4.2.3.2 MongoDB.....	49
4.2.3.3 Redis	49
5. ΠΑΡΟΥΣΙΑΣΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΩΝ	50
5.1 Εισαγωγή.....	50
5.2 Παρουσίαση της υπηρεσίας ανάλυσης δεδομένων (Analytics Service) και των διαθέσιμων REST διεπαφών	50
5.3 Παρουσίαση του μηχανισμού έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD	65
6. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ	74
6.1 Τελικά Συμπεράσματα	74
6.2 Μελλοντικές βελτιώσεις.....	74
ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ	75

ΛΙΣΤΑ ΕΙΚΟΝΩΝ

ΕΙΚΟΝΑ 1 ΧΡΟΝΟΣΕΙΡΑ ΤΗΣ ΤΙΜΗΣ ΠΡΟΣΦΟΡΑΣ ΑΝΑ ΜΗΝΑ ΓΙΑ ΤΗΝ ΕΛΛΑΔΑ.	19
ΕΙΚΟΝΑ 2 ΑΠΟΣΥΝΘΕΣΗ ΤΗΣ ΧΡΟΝΟΣΕΙΡΑΣ ΓΙΑ ΤΗΝ ΤΙΜΗ ΠΡΟΣΦΟΡΑΣ ΣΕ ΕΥΡΩ ΣΤΗΝ ΕΛΛΑΔΑ ΤΟ ΧΡΟΝΙΚΟ ΔΙΑΣΤΗΜΑ 2012-2022.	20
ΕΙΚΟΝΑ 3 Η ΕΠΟΧΙΚΟΤΗΤΑ ΤΗΣ ΤΙΜΗΣ ΠΡΟΣΦΟΡΑΣ ΣΕ ΕΥΡΩ ΤΗ ΧΡΟΝΙΚΗ ΠΕΡΙΟΔΟ 2012 -2022.	21
ΕΙΚΟΝΑ 4 ΧΡΟΝΟΣΕΙΡΑ ΜΕ ΠΟΛΛΑΠΛΑΣΙΑΣΤΙΚΗ ΕΠΟΧΙΚΟΤΗΤΑ. ΤΑ ΒΕΛΑΚΙΑ ΔΕΙΧΝΟΥΝ ΤΟ ΑΥΞΑΝΟΜΕΝΟ ΠΛΑΤΟΣ ΤΗΣ ΧΡΟΝΟΣΕΙΡΑΣ ΚΑΘΕ ΕΤΟΣ. ΠΗΓΗ: [18].	21
ΕΙΚΟΝΑ 5 Η ΧΡΟΝΟΣΕΙΡΑ ΤΗΣ ΤΙΜΗΣ ΠΡΟΣΦΟΡΑΣ ΣΤΗΝ ΕΛΛΑΔΑ ΤΟ 2012-2022 (ΜΠΛΕ ΧΡΩΜΑ) ΚΑΙ Η ΕΞΟΜΑΛΥΜΕΝΗ ΧΡΟΝΟΣΕΙΡΑ ΜΕ ΤΗΝ ΤΕΧΝΙΚΗ ΑΠΛΟΥ ΚΙΝΗΤΟΥ ΜΕΣΟΥ (ΠΟΡΤΟΚΑΛΙ ΧΡΩΜΑ).	23
ΕΙΚΟΝΑ 6 Η ΧΡΟΝΟΣΕΙΡΑ ΤΗΣ ΤΙΜΗΣ ΠΡΟΣΦΟΡΑΣ ΣΤΗΝ ΕΛΛΑΔΑ ΤΟ 2012-2022 (ΜΠΛΕ ΧΡΩΜΑ) ΚΑΙ Η ΕΞΟΜΑΛΥΜΕΝΗ ΧΡΟΝΟΣΕΙΡΑ ΤΟΥ ΚΕΝΤΡΑΡΙΣΜΕΝΟΥ ΚΙΝΗΤΟΥ ΜΕΣΟΥ (ΠΟΡΤΟΚΑΛΙ ΧΡΩΜΑ).	23
ΕΙΚΟΝΑ 7 Η ΧΡΟΝΟΣΕΙΡΑ ΤΗΣ ΤΙΜΗΣ ΠΡΟΣΦΟΡΑΣ ΣΤΗΝ ΕΛΛΑΔΑ ΤΟ 2012-2022 (ΜΠΛΕ ΧΡΩΜΑ) ΚΑΙ Η ΕΞΟΜΑΛΥΜΕΝΗ ΧΡΟΝΟΣΕΙΡΑ ΤΟΥ ΚΙΝΗΤΟΥ ΔΙΑΜΕΣΟΥ (ΠΟΡΤΟΚΑΛΙ ΧΡΩΜΑ).	24
ΕΙΚΟΝΑ 8 ΤΟ ΜΟΝΤΕΛΟ ΜΕΤΑΔΕΔΟΜΕΝΩΝ ΤΟΥ ESPD ΓΙΑ ΤΗ ΔΗΜΙΟΥΡΓΙΑ ΑΡΧΕΙΩΝ ΚΑΤΑΓΡΑΦΗΣ (LOG).	33
ΕΙΚΟΝΑ 9 Η ΡΟΗ ΤΩΝ ΔΕΔΟΜΕΝΩΝ ΑΠΟ ΤΟΥΣ PRODUCERS ΣΤΟΥΣ CONSUMERS.	34
ΕΙΚΟΝΑ 10 ΠΑΡΑΔΕΙΓΜΑ ΑΡΧΙΤΕΚΤΟΝΙΚΗΣ ΕΝΟΣ ΚΑΦΚΑ CLUSTER. ΠΗΓΗ: [27].	35
ΕΙΚΟΝΑ 11 ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΜΗΧΑΝΙΣΜΟΥ ΑΠΟΘΗΚΕΥΣΗΣ ΑΡΧΕΙΩΝ ΚΑΤΑΓΡΑΦΗΣ ΤΟΥ ΕΠΙΧΕΙΡΗΣΙΑΚΟΥ ΣΥΣΤΗΜΑΤΟΣ ESPD.	35
ΕΙΚΟΝΑ 12 UML ΔΙΑΓΡΑΜΜΑ ΔΡΑΣΤΗΡΙΟΤΗΤΑΣ ΤΟΥ ΜΗΧΑΝΙΣΜΟΥ ΑΠΟΘΗΚΕΥΣΗΣ ΑΡΧΕΙΩΝ ΚΑΤΑΓΡΑΦΗΣ ESPD.	36
ΕΙΚΟΝΑ 13 Η ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΤΗΣ ΥΠΗΡΕΣΙΑΣ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ ΤΟΥ ΜΗΧΑΝΙΣΜΟΥ ΕΚΔΟΣΗΣ ΑΝΑΦΟΡΩΝ ΚΑΙ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ ΣΕ ΠΡΑΓΜΑΤΙΚΟ ΧΡΟΝΟ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ ESPD.	38
ΕΙΚΟΝΑ 14 UML ΔΙΑΓΡΑΜΜΑ ΔΡΑΣΤΗΡΙΟΤΗΤΑΣ ΓΙΑ ΤΗΝ ΥΠΗΡΕΣΙΑ ΑΝΑΛΥΣΗΣ ΔΕΔΟΜΕΝΩΝ.	39
ΕΙΚΟΝΑ 15 Η ΑΡΧΙΤΕΚΤΟΝΙΚΗ ΤΟΥ ΜΗΧΑΝΙΣΜΟΥ ΕΚΔΟΣΗΣ ΑΝΑΦΟΡΩΝ ΚΑΙ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ ΣΕ ΠΡΑΓΜΑΤΙΚΟ ΧΡΟΝΟ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ ESPD.	42
ΕΙΚΟΝΑ 16 UML ΔΙΑΓΡΑΜΜΑ ΠΕΡΙΠΤΩΣΕΩΝ ΧΡΗΣΗΣ ΓΙΑ ΤΟΝ ΜΗΧΑΝΙΣΜΟ ΕΚΔΟΣΗΣ ΑΝΑΦΟΡΩΝ ΚΑΙ ΠΑΡΑΚΟΛΟΥΘΗΣΗΣ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ ESPD.	43
ΕΙΚΟΝΑ 17 Η ΔΙΑΔΙΚΑΣΙΑ ANALYST-IN-THE-LOOP. ΠΗΓΗ: [33].	47
ΕΙΚΟΝΑ 18 ΛΕΙΤΟΥΡΓΙΑ ΤΟΥ ΑΡΑΧΕ FREE MARKER. ΠΗΓΗ: [40].	48
ΕΙΚΟΝΑ 19 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΗΣ ΤΟΥ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ΕΞΑΧΘΕΝΤΩΝ ESPD ΕΓΓΡΑΦΩΝ.	50
ΕΙΚΟΝΑ 20 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΗΣ ΤΟΥ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ΕΞΑΧΘΕΝΤΩΝ ESPD REQUEST ΕΓΓΡΑΦΩΝ.	51
ΕΙΚΟΝΑ 21 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΗΣ ΤΟΥ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ΕΞΑΧΘΕΝΤΩΝ ESPD RESPONSE ΕΓΓΡΑΦΩΝ.	51
ΕΙΚΟΝΑ 22 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΗΣ ΤΟΥ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ΕΞΑΧΘΕΝΤΩΝ ESPD REQUEST XML ΕΓΓΡΑΦΩΝ.	52
ΕΙΚΟΝΑ 23 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΗΣ ΤΟΥ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ΕΞΑΧΘΕΝΤΩΝ ESPD REQUEST HTML ΕΓΓΡΑΦΩΝ.	53
ΕΙΚΟΝΑ 24 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΗΣ ΤΟΥ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ΕΞΑΧΘΕΝΤΩΝ ESPD REQUEST PDF ΕΓΓΡΑΦΩΝ.	54
ΕΙΚΟΝΑ 25 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΗΣ ΤΟΥ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ΕΞΑΧΘΕΝΤΩΝ ESPD RESPONSE XML ΕΓΓΡΑΦΩΝ.	55
ΕΙΚΟΝΑ 26 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΗΣ ΤΟΥ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ΕΞΑΧΘΕΝΤΩΝ ESPD RESPONSE HTML ΕΓΓΡΑΦΩΝ.	56
ΕΙΚΟΝΑ 27 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΗΣ ΤΟΥ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ΕΞΑΧΘΕΝΤΩΝ ESPD RESPONSE PDF ΕΓΓΡΑΦΩΝ.	57
ΕΙΚΟΝΑ 28 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΗΣ ΤΟΥ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ΕΞΑΧΘΕΝΤΩΝ ESPD ΕΚΔΟΣΗΣ 1.x.x.	58
ΕΙΚΟΝΑ 29 ΑΠΟΤΕΛΕΣΜΑΤΑ ΠΡΟΒΛΕΨΗΣ ΤΟΥ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ΕΞΑΧΘΕΝΤΩΝ ESPD ΕΚΔΟΣΗΣ 2.x.x.	59
ΕΙΚΟΝΑ 30 ΤΟ ΜΟΝΤΕΛΟ ΠΡΟΒΛΕΨΗΣ ΓΙΑ ΤΑ ESPD REQUEST HTML ΕΓΓΡΑΦΑ.	60
ΕΙΚΟΝΑ 31 ΤΟ ΜΟΝΤΕΛΟ ΠΡΟΒΛΕΨΗΣ ΓΙΑ ΤΑ ESPD RESPONSE HTML ΕΓΓΡΑΦΑ.	60
ΕΙΚΟΝΑ 32 ΤΟ ΜΟΝΤΕΛΟ ΠΡΟΒΛΕΨΗΣ ΓΙΑ ΤΑ ESPD REQUEST PDF ΕΓΓΡΑΦΑ.	61
ΕΙΚΟΝΑ 33 ΤΟ ΜΟΝΤΕΛΟ ΠΡΟΒΛΕΨΗΣ ΓΙΑ ΤΑ ESPD RESPONSE PDF ΕΓΓΡΑΦΑ.	61
ΕΙΚΟΝΑ 34 ΤΟ ΜΟΝΤΕΛΟ ΠΡΟΒΛΕΨΗΣ ΓΙΑ ΤΑ ESPD REQUEST XML ΕΓΓΡΑΦΑ.	62
ΕΙΚΟΝΑ 35 ΤΟ ΜΟΝΤΕΛΟ ΠΡΟΒΛΕΨΗΣ ΓΙΑ ΤΑ ESPD RESPONSE XML ΕΓΓΡΑΦΑ.	62
ΕΙΚΟΝΑ 36 ΤΟ ΜΟΝΤΕΛΟ ΠΡΟΒΛΕΨΗΣ ΓΙΑ ΤΑ ESPD ΕΓΓΡΑΦΑ ΕΚΔΟΣΗΣ 1.x.x.	63
ΕΙΚΟΝΑ 37 ΤΟ ΜΟΝΤΕΛΟ ΠΡΟΒΛΕΨΗΣ ΓΙΑ ΤΑ ESPD ΕΓΓΡΑΦΑ ΕΚΔΟΣΗΣ 2.x.x.	63
ΕΙΚΟΝΑ 38 ΤΟ ΜΟΝΤΕΛΟ ΠΡΟΒΛΕΨΗΣ ΓΙΑ ΤΟΝ ΣΥΝΟΛΙΚΟ ΑΡΙΘΜΟ ΤΩΝ ESPD ΕΓΓΡΑΦΩΝ.	64
ΕΙΚΟΝΑ 39 ΤΟ ΜΟΝΤΕΛΟ ΠΡΟΒΛΕΨΗΣ ΓΙΑ ΤΟΝ ΣΥΝΟΛΙΚΟ ΑΡΙΘΜΟ ΤΩΝ ESPD REQUEST ΕΓΓΡΑΦΩΝ.	64
ΕΙΚΟΝΑ 40 ΤΟ ΜΟΝΤΕΛΟ ΠΡΟΒΛΕΨΗΣ ΓΙΑ ΤΟ ΣΥΝΟΛΙΚΟ ΑΡΙΘΜΟ ΤΩΝ ESPD RESPONSE ΕΓΓΡΑΦΩΝ.	65
ΕΙΚΟΝΑ 41 ΑΡΧΙΚΗ ΣΕΛΙΔΑ ΕΙΣΟΔΟΥ.	65
ΕΙΚΟΝΑ 42 ΕΙΣΑΓΩΓΗ USERNAME ΚΑΙ PASSWORD ΓΙΑ ΤΗΝ ΕΙΣΟΔΟ ΣΤΟ ΣΥΣΤΗΜΑ.	66
ΕΙΚΟΝΑ 43 ΔΙΑΓΡΑΜΜΑ ΧΡΟΝΟΣΕΙΡΑΣ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ΕΞΑΧΘΕΝΤΩΝ ESPD ΑΡΧΕΙΩΝ.	66
ΕΙΚΟΝΑ 44 ΕΠΙΛΟΓΗ ΣΥΓΚΕΚΡΙΜΕΝΟΥ ΧΡΟΝΙΚΟΥ ΔΙΑΣΤΗΜΑΤΟΣ ΑΠΟ ΤΗ ΧΡΟΝΟΣΕΙΡΑ.	67
ΕΙΚΟΝΑ 45 ΑΠΟΘΗΚΕΥΣΗΣ ΤΟΥ ΔΙΑΓΡΑΜΜΑΤΟΣ ΧΡΟΝΟΣΕΙΡΑΣ ΩΣ ΕΙΚΟΝΑ.	67

ΕΙΚΟΝΑ 46 ΔΙΑΓΡΑΜΜΑ ΠΙΤΑΣ ΚΑΙ ΓΡΑΜΜΗΣ ΑΝΑ ΕΤΟΣ ΓΙΑ ΚΑΘΕ ΕΚΔΟΣΗ.....	68
ΕΙΚΟΝΑ 47 ΔΙΑΓΡΑΜΜΑ ΠΙΤΑΣ ΓΙΑ ΤΟ 2021, ΕΠΙΛΕΓΟΝΤΑΣ ΤΟ ΕΤΟΣ 2021 ΜΕ ΤΟΝ ΚΕΡΣΟΡΑ ΠΑΝΩ ΣΤΟ ΔΙΑΓΡΑΜΜΑ ΓΡΑΜΜΗΣ ΓΙΑ ΤΟ 2021.	68
ΕΙΚΟΝΑ 48 ΡΑΒΔΟΓΡΑΜΜΑ ΜΕ ΤΙΣ ΕΞΑΓΩΓΕΣ ESPD ΑΡΧΕΙΩΝ ΑΝΑ ΕΚΔΟΣΗ.	69
ΕΙΚΟΝΑ 49 ΡΑΒΔΟΓΡΑΜΜΑ ΓΙΑ ΤΑ ESPD REQUEST ΚΑΙ RESPONSE ΤΗΣ ΕΚΔΟΣΗΣ 2.Χ.Χ.	69
ΕΙΚΟΝΑ 50 ΡΑΒΔΟΓΡΑΜΜΑ ΜΕ ΤΟ ΠΛΗΘΟΣ ESPD ΑΡΧΕΙΩΝ ΠΟΥ ΕΧΟΥΝ ΕΞΑΧΘΕΙ ΑΝΑ ΗΜΕΡΑ.	70
ΕΙΚΟΝΑ 51 ΔΙΑΓΡΑΜΜΑ ΔΟΝΟΥΤ ΜΕ ΤΟ ΠΛΗΘΟΣ ΤΩΝ ESPD ΑΡΧΕΙΩΝ ΓΙΑ ΚΑΘΕ ΓΛΩΣΣΑ.	70
ΕΙΚΟΝΑ 52 ΔΙΑΓΡΑΜΜΑ ΔΟΝΟΥΤ ΜΕ ΤΟΝ ΣΥΝΟΛΙΚΟ ΑΡΙΘΜΟ ΕΞΑΓΩΓΩΝ ΓΙΑ ΚΑΘΕ ΤΥΠΟ ΑΡΧΕΙΟΥ ESPD.....	71
ΕΙΚΟΝΑ 53 ΕΠΙΛΟΓΗ ΤΟΥ REPORTING ΑΠΟ ΤΟ ΝΑΥΒΑΡ.	71
ΕΙΚΟΝΑ 54 ΚΛΙΚ ΣΤΟ ΚΟΥΜΠΙ PDF ΓΙΑ ΤΗΝ ΕΚΔΟΣΗ ΤΗΣ ΑΝΑΦΟΡΑΣ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ ESPD.....	71
ΕΙΚΟΝΑ 55 ΕΝΔΕΙΚΤΙΚΗ ΕΙΚΟΝΑ ΤΟΥ ΜΕΡΟΥΣ I ΜΕΤΑ ΤΗΝ ΕΚΔΟΣΗ ΤΗΣ ΑΝΑΦΟΡΑΣ ΓΙΑ ΤΟ ΣΥΣΤΗΜΑ ESPD.....	72
ΕΙΚΟΝΑ 56 ΕΝΔΕΙΚΤΙΚΗ ΕΙΚΟΝΑ ΤΟΥ ΜΕΡΟΥΣ II ΜΕΤΑ ΤΗΝ ΕΚΔΟΣΗ ΤΗΣ ΑΝΑΦΟΡΑΣ ΓΙΑ ΤΟ ΣΥΣΤΗΜΑ ESPD.....	73

ΛΙΣΤΑ ΠΙΝΑΚΩΝ

ΠΙΝΑΚΑΣ 1 ΠΡΟΒΛΕΨΗ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ESPD ΑΡΧΕΙΩΝ, ΜΕΧΡΙ ΤΗΝ ΕΠΟΜΕΝΗ ΕΒΔΟΜΑΔΑ.	39
ΠΙΝΑΚΑΣ 2 ΠΡΟΒΛΕΨΗ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ESPD REQUEST ΑΡΧΕΙΩΝ, ΜΕΧΡΙ ΤΗΝ ΕΠΟΜΕΝΗ ΕΒΔΟΜΑΔΑ.....	40
ΠΙΝΑΚΑΣ 3 ΠΡΟΒΛΕΨΗ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ESPD RESPONSE ΑΡΧΕΙΩΝ, ΜΕΧΡΙ ΤΗΝ ΕΠΟΜΕΝΗ ΕΒΔΟΜΑΔΑ.	40
ΠΙΝΑΚΑΣ 4 ΠΡΟΒΛΕΨΗ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ESPD REQUEST ΑΡΧΕΙΩΝ ΜΕ ΒΑΣΗ ΤΟ ΤΥΠΟ ΑΡΧΕΙΟΥ, ΜΕΧΡΙ ΤΗΝ ΕΠΟΜΕΝΗ ΕΒΔΟΜΑΔΑ.....	40
ΠΙΝΑΚΑΣ 5 ΠΡΟΒΛΕΨΗ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ESPD RESPONSE ΑΡΧΕΙΩΝ ΜΕ ΒΑΣΗ ΤΟ ΤΥΠΟ ΑΡΧΕΙΟΥ, ΜΕΧΡΙ ΤΗΝ ΕΠΟΜΕΝΗ ΕΒΔΟΜΑΔΑ.	41
ΠΙΝΑΚΑΣ 6 ΠΡΟΒΛΕΨΗ ΣΥΝΟΛΙΚΟΥ ΑΡΙΘΜΟΥ ESPD ΑΡΧΕΙΩΝ ΜΕ ΒΑΣΗ ΤΗΝ ΕΚΔΟΣΗ, ΜΕΧΡΙ ΤΗΝ ΕΠΟΜΕΝΗ ΕΒΔΟΜΑΔΑ.	41

ΠΙΝΑΚΑΣ ΣΥΝΤΟΜΕΥΣΕΩΝ

ESPD	European Single Procurement Document
ESPD-EDM	European Single Procurement Document – Exchange Data Model
PDF	Portable Document Format
RDF	Resource Description Framework
XML	Extensible Markup Language
HTML	HyperText Markup Language
ARIMA	Autoregressive Integrated Moving Average
SARIMA	Seasonal Autoregressive Integrated Moving Average
KPSS	Kwiatkowski-Phillips-Schmidt-Shin
ADF	Augmented-Dickey-Fuller
AR(p)	Autoregressive Model of order p
MA(q)	Moving Average Model of order q
PACF	Partial Autocorrelation Function
ROI	Return Of Investment
CSV	Comma Separated Values
JSON	JavaScript Object Notation
NDJSON	Newline Delimited JSON
BI	Business Intelligence
GIS	Geographic Information System
CRM	Customer Relationship Management
DSS	Decision Support Systems
RFM	Recency Frequency Monetary
ETL	Extract Transform Load
RMSE	Root Mean Square Error
UI	User Interface
REST	Representational State Transfer

API	Application Programming Interface
WSGI	Web Server Gateway Interface
MVC	Model View Controller
SQL	Structured Query Language
KSQL	Kafka Structured Query Language
PNG	Portable Network Graphics
UML	Unified Modeling Language
DBSCAN	Density-based Spatial Clustering With Noise
HTTP	Hypertext Transfer Protocol
LOF	Local Outlier Factor

1. ΕΙΣΑΓΩΓΗ

1.1 Ορισμός του προβλήματος

Στη σύγχρονη εποχή, ο όγκος των δεδομένων που παράγεται καθημερινά από τους χρήστες είναι τεράστιος, κάτι το οποίο αποτελεί συγχρόνως πρόβλημα αλλά και μεγάλο πλεονέκτημα για κάθε οργανισμό. Το πρόβλημα εντοπίζεται στη διαχείριση και στην ανάλυση μεγάλου όγκου δομημένων, ημιδομημένων και αδόμητων δεδομένων. Τα δομημένα δεδομένα έχουν σαφώς προκαθορισμένη δομή που συμμορφώνεται σε συγκεκριμένο σχήμα, σε αυτή την κατηγορία εντάσσονται κυρίως οι σχεσιακές βάσεις δεδομένων [1, p. 75]. Τα ημιδομημένα δεδομένα δεν έχουν σαφώς προκαθορισμένη δομή, αλλά η δομή τους μπορεί να είναι δυναμική, χωρίς να συμμορφώνεται σε συγκεκριμένο σχήμα, σε αυτή την κατηγορία εντάσσονται τύποι αρχείων όπως XML, RDF κ.α [1, p. 76]. Τα αδόμητα δεδομένα είναι δεδομένα που δεν έχουν αναγνωρίσιμη δομή σε αυτή την κατηγορία εντάσσονται τύποι αρχείων όπως εικόνες, PDF κ.α [1, p. 74]. Επομένως, η πληροφορία που εξάγεται από τα δεδομένα ενός συστήματος μπορεί να βοηθήσει στη λήψη ορθών αποφάσεων για τον οργανισμό, διατηρώντας την ανταγωνιστικότητα του στο επιχειρηματικό περιβάλλον. Συνεπώς, ο σκοπός αυτής της διπλωματικής εργασίας είναι η δημιουργία ενός συστήματος επιχειρησιακής ευφυΐας (business intelligence), που θα χρησιμοποιείται ως συμβουλευτικό εργαλείο για τη λήψη κρίσιμων αποφάσεων σχετικά με το σύστημα ESPD.

1.2 Δομή της Διπλωματικής Εργασίας

Η παρούσα διπλωματική εργασία εστιάζει στη θεματική περιοχή της επεξεργασίας και ανάλυσης δεδομένων (data analytics) στις ηλεκτρονικές προμήθειες. Στην αρχή, παρουσιάζεται η ανάλυση χρονοσειρών για τη δημιουργία προβλέψεων, η έννοια του όρου επιχειρησιακή ευφυΐα (business intelligence), η συμβολή της επιχειρησιακής ευφυΐας στη λήψη κρίσιμων αποφάσεων για ένα οργανισμό, καθώς και οι διάφορες αρχιτεκτονικές που ακολουθούνται για το σχεδιασμό συστημάτων επιχειρησιακής ευφυΐας. Στη συνέχεια, θα γίνει ιδιαίτερη αναφορά στο μοντέλο μεταδεδομένων του ESPD, που χρησιμοποιείται για τη δημιουργία αρχείων καταγραφής, καθώς και περιγραφή του μηχανισμού έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD. Επιπρόσθετα, θα παρατεθούν οι τεχνολογίες που χρησιμοποιήθηκαν για την υλοποίηση του μηχανισμού και θα γίνει παρουσίαση του συστήματος. Τέλος, συνάγονται τελικά συμπεράσματα και μελλοντικές βελτιώσεις για το μηχανισμό έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD.

1.3 Συνεισφορά της Διπλωματικής Εργασίας

Η συμβολή της παρούσας διπλωματικής εργασίας είναι η υλοποίηση ενός μηχανισμού παρακολούθησης του συστήματος ESPD, ο οποίος δεν υπήρχε και είναι ιδιαίτερα αναγκαίος για τη λήψη αποφάσεων δεδομένου ότι στη σημερινή εποχή ο τεράστιος όγκος δεδομένων μπορεί να προσφέρει αξιόπιστη και χρήσιμη πληροφορία. Πιο συγκεκριμένα, υλοποιήθηκε ένας μηχανισμός αποθήκευσης αρχείων καταγραφής (log), σχετικά με τις εξαγωγές ESPD

εγγράφων από το σύστημα ESPD. Στη συνέχεια, με τη χρήση του μοντέλου Prophet, δίνεται η δυνατότητα να προβλεφθεί σε ένα εύλογο χρονικό διάστημα το πλήθος εγγράφων που θα εξαχθούν από το ESPD σύστημα, με βάση διάφορα κριτήρια όπως η έκδοση του εγγράφου, η γλώσσα του εγγράφου κ.α . Τέλος, παρέχεται η παρακολούθηση των εξαγόμενων εγγράφων του συστήματος ESPD σε πραγματικό χρόνο (real-time), με βάση διάφορα κριτήρια όπως η ημέρα, η γλώσσα κ.α.

2. ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΕΠΙΣΚΟΠΗΣΗ

2.1 Εισαγωγή

Στο Κεφάλαιο 2 διατυπώνεται ο όρος επιχειρησιακή ευφυΐα και παρατίθενται τα οφέλη ενός οργανισμού από την εφαρμογή της επιχειρησιακής ευφυΐας στη λήψη αποφάσεων, οι κατηγορίες και οι περιπτώσεις χρήσεις συστημάτων επιχειρησιακής ευφυΐας. Επίσης, παρουσιάζονται η έννοια των χρονοσειρών και η χρησιμότητα τους στην επίλυση επιχειρησιακών προβλημάτων ενός οργανισμού, τα βασικά στοιχεία που αποτελούν μία χρονοσειρά και οι τεχνικές εξομάλυνσης χρονοσειράς. Στη συνέχεια, γίνεται αναφορά στη μεθοδολογία δημιουργίας ενός προβλεπτικού μοντέλου. Τέλος, περιγράφεται το αυτοπαλινδρομικό μοντέλο (autoregressive model), το μοντέλο κινητών μέσων (moving average model), το μοντέλο ARIMA και το μοντέλο Prophet.

2.2 Επιχειρησιακή Ευφυΐα

2.2.1 Ορισμός Επιχειρησιακής Ευφυΐας

Ο όρος επιχειρησιακή ευφυΐα (business intelligence) παρουσιάστηκε πρώτη φορά με την έννοια που είναι ευρέως γνωστή σήμερα από τον Howard Dresner το έτος 1989, περιγράφοντας την επιχειρησιακή ευφυΐα ως μια συλλογή τεχνολογιών που βοηθούν τις επιχειρήσεις να λάβουν τις κατάλληλες αποφάσεις με βάση γεγονότα [2]. Στη βιβλιογραφία εμφανίζονται πολλοί ορισμοί του όρου επιχειρησιακή ευφυΐα όπως των Zeng et al [3], που αναφέρει ότι η επιχειρησιακή ευφυΐα είναι μία διαδικασία συλλογής, διαχείρισης και διάδοσης της πληροφορίας με σκοπό τη μείωση της αβεβαιότητας κατά τη διαδικασία λήψης αποφάσεων. Επίσης, ένας άλλος ορισμός παρατίθεται από τους Stackowiak et al. [4], που περιγράφει την επιχειρησιακή ευφυΐα ως μία διαδικασία συλλογής μεγάλου όγκου δεδομένων με σκοπό την ανάλυση τους συνοψίζοντας την πληροφορία σε αναφορές (reports). Τέλος, ένας ακόμη ορισμός για τον όρο επιχειρησιακή ευφυΐα που αξίζει να αναφερθεί είναι του Negash et al [5], ότι η επιχειρησιακή ευφυΐα είναι μια διαδικασία μετατροπής των επιχειρησιακών δεδομένων σε χρήσιμη πληροφορία για τον οργανισμό η οποία μέσω ανάλυσης από ειδικούς θα εξάγει την απαραίτητη γνώση για τον οργανισμό. Συνεπώς, από τους προαναφερθέντες ορισμούς μπορεί να συναχθεί ότι η επιχειρησιακή ευφυΐα αποτελεί μία διαδικασία ανάκτησης δεδομένων από ετερογενείς πηγές, με σκοπό την εφαρμογή τεχνικών ανάλυσης δεδομένων για την παραγωγή χρήσιμης πληροφορίας, η οποία με την κατάλληλη ανάλυση από ειδικούς θα οδηγήσει στη δημιουργία απαραίτητης γνώσης για τη λήψη των κατάλληλων επιχειρησιακών αποφάσεων από τη διοίκηση του οργανισμού.

2.2.2 Οφέλη της Επιχειρησιακής Ευφυΐας για κάθε οργανισμό

Η συνεισφορά της επιχειρησιακής ευφυΐας στη σημερινή εποχή είναι ιδιαίτερα σημαντική για κάθε οργανισμό, καθώς του δίνει τη δυνατότητα εξοικονόμησης χρημάτων και

μεγιστοποίησης του κέρδους. Πιο συγκεκριμένα, σύμφωνα με τους Jourdan et al. [6], οργανισμοί που εφάρμοσαν τεχνολογίες επιχειρησιακής ευφυΐας είχαν έως και 1000% ROI. Το προϊόν που προσφέρει η επιχειρησιακή ευφυΐα στον οργανισμό είναι η πληροφορία που πηγάζει μέσα από τα επιχειρησιακά δεδομένα. Οι Wieder και Ossimitz [7] απέδειξαν ότι η ποιότητα των επιχειρησιακών δεδομένων επηρεάζει την ποιότητα της πληροφορίας που παράγεται από ένα σύστημα επιχειρησιακής ευφυΐας. Επομένως, η πληροφορία για να είναι αξιόπιστη, χρειάζεται αφενός η ορθή διαχείριση των δεδομένων από ετερογενείς πηγές, όπως βάσεις δεδομένων, αρχεία JSON κλπ. και αφετέρου η εφαρμογή των κατάλληλων τεχνικών ανάλυσης δεδομένων. Επιπρόσθετα, η επιχειρησιακή ευφυΐα μπορεί να προσφέρει στη βελτίωση των επιχειρησιακών διαδικασιών ενός οργανισμού σε πραγματικό χρόνο (real-time) ή σε σχετικά πραγματικό χρόνο (near real-time). Οι Vuksic et al. [8] αναφέρουν ότι ένα σύστημα επιχειρησιακής ευφυΐας μπορεί να προσφέρει σε πραγματικό χρόνο (real-time) την κατάσταση κάθε επιχειρησιακής διαδικασίας (business process), μετρώντας τον χρόνο εκτέλεσης της καθώς και το κόστος σε χρηματικούς όρους που απαιτείται για την εκτέλεση της, με συνέπεια η διοίκηση του οργανισμού να έχει τη δυνατότητα βελτίωσης των επιχειρησιακών διαδικασιών του οργανισμού. Συνεπώς, διαπιστώνεται ότι η χρήση της επιχειρησιακής ευφυΐας συνεισφέρει όχι μόνο στους επιχειρηματικούς σκοπούς ενός οργανισμού, όπως η διατήρηση ή προσέλκυση νέων πελατών, αλλά και στη διάρθρωση των επιχειρησιακών διαδικασιών του.

2.2.3 Κατηγορίες Συστημάτων Επιχειρησιακής Ευφυΐας

Στη σημερινή εποχή, τα συστήματα επιχειρησιακής ευφυΐας χωρίζονται σε δύο βασικές κατηγορίες: τα παραδοσιακά συστήματα επιχειρησιακής ευφυΐας (traditional BI Systems) και τα νέα γενιάς (next generation BI Systems). Σύμφωνα με τους Quoc Duy Vo et al. [9, p. 2], ένα παραδοσιακό σύστημα επιχειρησιακής ευφυΐας αποτελείται από τρία επίπεδα:

- Επίπεδο Παρουσίασης (Presentation Layer). Αποτελεί τη διεπαφή του χρήστη με το σύστημα, προσφέροντας αναφορές για τη λειτουργία επιχειρησιακού συστήματος ή οπτικοποίηση των δεδομένων με χρήσιμα διαγράμματα, τα οποία προσφέρουν σημαντική πληροφορία για τη λειτουργία του επιχειρησιακού συστήματος.
- Επίπεδο Εφαρμογής (Application Layer). Σε αυτό το επίπεδο πραγματοποιούνται ερωτήματα (queries) για την ανάκτηση δεδομένων από αποθήκες δεδομένων (data warehouses), την εκπαίδευση και αξιολόγηση μοντέλων μηχανικής μάθησης, τα οποία είναι χρήσιμα για την παροχή αξιόπιστων προβλέψεων.
- Επίπεδο Δεδομένων (Data Layer). Αυτό το επίπεδο αποτελείται από αποθήκες δεδομένων (data warehouses), που λαμβάνουν δεδομένα από βάσεις δεδομένων επιχειρησιακών συστημάτων του οργανισμού, με σκοπό την εφαρμογή τεχνικών ανάλυσης δεδομένων και την εξαγωγή χρήσιμης πληροφορίας, χωρίς οι βάσεις δεδομένων των επιχειρησιακών συστημάτων του οργανισμού να απασχολούνται με τη διαδικασία της ανάλυσης των δεδομένων.

Μειονέκτημα των παραδοσιακών συστημάτων επιχειρησιακής ευφυΐας αποτελεί η αργή εξυπηρέτηση των χρηστών του συστήματος. Επίσης, η συντήρησή τους είναι ιδιαίτερα

δύσκολη καθώς απαιτεί άτομα με εξειδικευμένες γνώσεις σε κάθε επίπεδο του συστήματος. Αντίθετα, τα συστήματα επιχειρησιακής ευφυΐας νέας γενιάς εξαλείφουν τα μειονεκτήματα των παραδοσιακών συστημάτων επιχειρησιακής ευφυΐας, προσφέροντας πληροφορία σε πραγματικό χρόνο (real-time) ή σε σχετικά πραγματικό χρόνο (near real-time). Οι κατηγορίες των συστημάτων επιχειρησιακής ευφυΐας νέας γενιάς είναι [9, p. 2]:

- **Επιχειρησιακά Συστήματα (Operational BI).** Αποτελούν συστήματα επιχειρησιακής ευφυΐας πραγματικού χρόνου (real-time) ή σχετικά πραγματικού χρόνου (near real-time), που αποσκοπούν στη μείωση του χρόνου της ανάκτησης και ανάλυσης δεδομένων. Τα επιχειρησιακά συστήματα βασίζονται (Operational BI) σε συμβάντα, τα οποία όταν πραγματοποιηθούν ενεργοποιούν διάφορες διαδικασίες, όπως ανάκτηση δεδομένων και ενημέρωση της διεπαφής χρήστη (user interface) σε πραγματικό χρόνο κ.α. Τα δεδομένα που χρησιμοποιούνται είναι streaming από συστήματα του οργανισμού.
- **Συστήματα Κατάστασης (Situational BI).** Τα συστήματα κατάστασης (S BI) παρουσιάζουν τις πρόσφατες ενημερώσεις, αντλώντας ημιδομημένα δεδομένα από το εταιρικό δίκτυο, από το διαδίκτυο κ.α και δομημένα δεδομένα από αποθήκες δεδομένων (data warehouses) με σκοπό την παραγωγή πληροφορίας σε πραγματικό χρόνο, για την υποστήριξη της διαδικασίας λήψης αποφάσεων.
- **Συστήματα Αυτοεξυπηρέτησης (Self-Service BI).** Τα συστήματα αυτοεξυπηρέτησης (SSBI) προσφέρουν όμορφες και φιλικές προς τον χρήστη διεπαφές (user-friendly interfaces) χωρίς να χρειάζονται γνώσεις προγραμματισμού, με σκοπό τη δημιουργία αναφορών (reports), την οπτικοποίηση δεδομένων (data visualization) με τη χρήση καλαίσθητων διαγραμμάτων που αποδίδουν την πληροφορία με κατανοητό τρόπο στον χρήστη. Τέλος, δίνουν τη δυνατότητα στον χρήστη να επεξεργαστεί τα δεδομένα που βρίσκονται σε αποθήκες δεδομένων ή αρχεία, μέσω διεπαφής χρήστη (user interface).

2.2.4 Ηλεκτρονικές Προμήθειες και Επιχειρησιακή Ευφυΐα

Τα τελευταία χρόνια έχει αυξηθεί ραγδαία η εφαρμογή της επιχειρησιακής ευφυΐας σε συστήματα ηλεκτρονικών προμηθειών. Πιο αναλυτικά, σύμφωνα με την εταιρεία συμβούλων Gartner προβλέπεται ότι πάνω από το 75% των επιχειρήσεων που ασχολούνται με διαχείριση ηλεκτρονικών προμηθειών θα χρησιμοποιούν επιχειρησιακή ευφυΐα στα συστήματά τους μέχρι το 2026 [10]. Μερικά από τα προτερήματα εφαρμογής της επιχειρησιακής ευφυΐας σε συστήματα ηλεκτρονικών προμηθειών είναι τα ακόλουθα [11]:

- Δίνει τη δυνατότητα σε υψηλής ποιότητας πληροφορία που πηγάζει από τα δεδομένα του οργανισμού, διευκολύνοντας τη λήψη αποφάσεων με την οπτικοποίηση δεδομένων, τη δημιουργία προβλέψεων και την έκδοση αναφορών.
- Διασφαλίζει τη διαφάνεια στις οικονομικές δόσοληψίες που αφορούν προμήθειες ενός οργανισμού.

- Συνεισφέρει στον εντοπισμό των κατάλληλων προμηθευτών με βάση την ποιότητα, την τιμή και τον χρόνο παράδοσης των προϊόντων.
- Βοηθάει στην αυτοματοποίηση της διαδικασίας ηλεκτρονικών προμηθειών, όταν οι αποθήκες έχουν φτάσει σε ένα επίπεδο διαθεσιμότητας όπου κρίνεται αναγκαία η παραγγελία νέων προϊόντων.

Στην παρούσα διπλωματική εργασία δημιουργήθηκε ένας μηχανισμός επιχειρησιακής ευφυΐας για την παρακολούθηση του συστήματος ESPD σε πραγματικό χρόνο καθώς και την έκδοση αναφοράς της κατάστασης του συστήματος. Επίσης, περιλαμβάνει προβλέψεις για το πλήθος των εγγράφων ESPD που θα εκδοθούν από το σύστημα. Το σύστημα ESPD αποτελεί ένα σύστημα για τη δημιουργία ηλεκτρονικά του Ευρωπαϊκού Ενιαίου Εγγράφου Σύμβασης (ESPD), το οποίο περιλαμβάνει την οικονομική κατάσταση, τις ικανότητες και την καταλληλότητα των επιχειρήσεων για διαδικασίες σύναψης δημόσιων συμβάσεων στην Ευρωπαϊκή Ένωση [12]. Το πλεονέκτημα αυτού του μηχανισμού είναι ότι δίνει τη δυνατότητα άμεσης πληροφόρησης του διαχειριστή του συστήματος σχετικά με την έκδοση των ESPD εγγράφων, μέσα από ένα φιλικό περιβάλλον χρήστη και οπτικοποίησης των δεδομένων με καλαίσθητα διαγράμματα.

2.2.5 Περιπτώσεις χρήσης της Επιχειρησιακής Ευφυΐας

Η χρήση συστημάτων επιχειρησιακής ευφυΐας από οργανισμούς πληθαίνει τα τελευταία χρόνια, καθώς βασικό γνώρισμα της είναι η εξαγωγή γνώσης για το μέλλον, η οποία εξαλείφει σε μεγάλο ποσοστό την αβεβαιότητα στην επιλογή των κατάλληλων στρατηγικών που θα παρθούν από τη διοίκηση του οργανισμού. Αξίζει να σημειωθεί ότι η χρήση συστημάτων επιχειρησιακής ευφυΐας δεν αφορά μόνο ιδιωτικούς οργανισμούς, αλλά και δημόσιους οργανισμούς σε τομείς, όπως ο τουρισμός, η υγεία, η οικονομία κλπ. Σύμφωνα με την Ranjan [13, p. 66], υπάρχουν πολλές κατηγορίες συστημάτων στις οποίες είναι χρήσιμη η επιχειρησιακή ευφυΐα, όπως Γεωγραφικά Πληροφοριακά Συστήματα (GIS), Συστήματα Διαχείρισης Πελατών (CRM), Συστήματα Λήψης Αποφάσεων (DSS) κλπ. Μερικές χαρακτηριστικές περιπτώσεις χρήσης είναι:

- Εντοπισμός προβλημάτων στην εξυπηρέτηση πελατών. Θα ήταν δυνατό στα σχόλια των πελατών να πραγματοποιείται συναισθηματική ανάλυση (sentiment analysis) και σε συνδυασμό με την αξιολόγηση που δίνουν οι ίδιοι χρήστες για το προϊόν, να ενημερώνεται σε πραγματικό χρόνο ο οργανισμός για πιθανές αλλαγές που θα πρέπει να γίνουν στο προϊόν προκειμένου να αυξηθεί η ικανοποίηση των πελατών.
- Κατηγοριοποίηση πελατών (customer segmentation), με βάση τη συμπεριφορά τους ως προς την πιο πρόσφατη αγορά (recency) που έχουν πραγματοποιήσει, τη συχνότητα (frequency) που αγοράζουν προϊόντα και τα χρηματικά ποσά που έχουν ξοδέψει συνολικά (monetary) (RFM analysis) [14].

- Εμφάνιση επικίνδυνων περιοχών για πυρκαγιές ή φυσικές καταστροφές σε πραγματικό χρόνο, από γεωλογικά και καιρικά δεδομένα.
- Ανίχνευση χρηματικών δοσοληψιών που πιθανώς είναι ύποπτες για ξέπλυμα μαύρου χρήματος.
- Κατηγοριοποίηση πελατών μίας τράπεζας σε ικανούς να εξοφλήσουν ένα δάνειο και μη, αξιοποιώντας δεδομένα τους σε πραγματικό χρόνο.

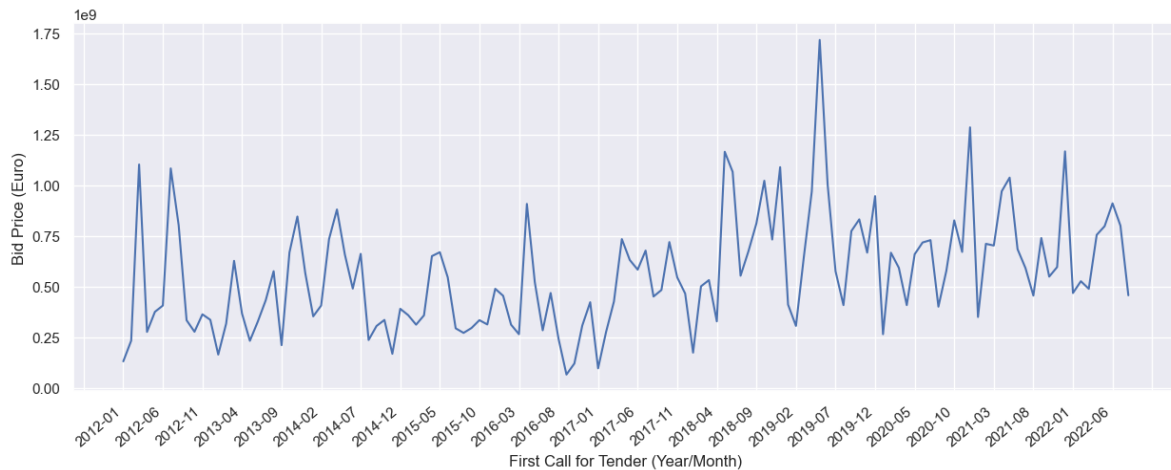
Συνεπώς, τα συστήματα επιχειρησιακής ευφυΐας δίνουν ιδιαίτερα σημαντικές δυνατότητες, που έχουν ενισχυθεί από τη ραγδαία ανάπτυξη των υπολογιστικών νεφών (cloud computing) και του μεγάλου όγκου δεδομένου (big data), με συνέπεια οι οργανισμοί να αποφασίζουν και να σχεδιάζουν με μεγαλύτερη ασφάλεια το μέλλον τους, ενισχύοντας την ανταγωνιστικότητά τους στο επιχειρηματικό περιβάλλον.

2.3 Ανάλυση Χρονοσειρών

Για τη δημιουργία των γραφημάτων που παρουσιάζονται στο κεφάλαιο 2.3, συλλέχθηκαν δεδομένα για τη διαδικασία προσφορών στην Ελλάδα από το portal orentender, που υλοποιήθηκε με τη χρηματοδότηση του προγράμματος Horizon 2020 της Ευρωπαϊκής Ένωσης [15]. Το orentender περιλαμβάνει χρήσιμα γραφήματα και στατιστικά στοιχεία για τη διαδικασία προσφορών συνολικά στην Ευρωπαϊκή Ένωση, καθώς και για κάθε χώρα ξεχωριστά. Τα σύνολα δεδομένων που προσφέρει είναι σε μορφή CSV, JSON και NDJSON. Σκοπός του έργου είναι η εξάλειψη της διαφθοράς σε διαδικασίες προμηθειών, διότι κοστίζει ετησίως πολλά χρήματα στους κρατικούς προϋπολογισμούς των κρατών μελών της Ευρωπαϊκής Ένωσης.

2.3.1 Ορισμός Χρονοσειράς και περιπτώσεις χρήσης

Οι χρονοσειρές εμφανίστηκαν πρώτη φορά στο βιβλίο του διάσημου μηχανικού και οικονομολόγου William Playfair στο βιβλίο του Statistical Breviary το 1801, παρουσιάζοντας τις εξαγωγές και εισαγωγές προϊόντων για διάφορα βασίλεια της εποχής [16]. Πιο συγκεκριμένα, η χρονοσειρά είναι μία ακολουθία παρατηρήσεων, που συλλέγονται σε ίσα χρονικά διαστήματα για μία μεταβλητή ενδιαφέροντος (variable of interest) [17]. Στην εικόνα 1 εμφανίζεται η χρονοσειρά, που περιλαμβάνει τις τιμές των προσφορών για κάθε μήνα από το 2012 έως το 2022 για την Ελλάδα.



Εικόνα 1 Χρονοσειρά της τιμής προσφοράς ανά μήνα για την Ελλάδα.

Στην Εικόνα 1, η μεταβλητή ενδιαφέροντος είναι η τιμή προσφοράς (Bid Price) για κάθε μήνα. Αξίζει να σημειωθεί ότι η δημιουργία μίας χρονοσειράς αφενός μπορεί να είναι χρήσιμη για την περιεκτική και κατατοπιστική παρουσίαση των παρατηρήσεων μίας μεταβλητής ενδιαφέροντος στον χρόνο, αφετέρου μπορεί να προσφέρει στη δημιουργία προβλεπτικών μοντέλων που θα βοηθούν τη διοίκηση του οργανισμού στη σωστή λήψη σωστών επιχειρηματικών αποφάσεων. Μερικές χαρακτηριστικές περιπτώσεις είναι:

- Η πρόβλεψη πωλήσεων ενός προϊόντος. Είναι σημαντικό για ένα οργανισμό να γνωρίζει την ποσότητα των προϊόντων που θα πωλήσει στους πελάτες του. Αναλυτικότερα, είναι αναγκαία η κατάλληλη διαχείριση των προϊόντων, που βρίσκονται στις αποθήκες προκειμένου να μην υπάρξει έλλειψη προϊόντων αλλά ούτε και να έχει μεγάλο πλεόνασμα προϊόντων. Επίσης, είναι σημαντικό για τον οργανισμό να γνωρίζει τα έσοδα που θα έχει από την πώληση των προϊόντων για να καθορίσει τις επιχειρηματικές κινήσεις του με περισσότερη ασφάλεια για το μέλλον.
- Η προώθηση ενός προϊόντος στους κατάλληλους πελάτες. Η διαφημιστική προώθηση ενός οργανισμού κοστίζει πολλά χρήματα από το προϋπολογισμό του, γι' αυτό είναι αναγκαίο να γνωρίζει πόσοι πελάτες θα ανταποκριθούν θετικά σε μία πιθανή διαφήμιση του προϊόντος. Με αυτό το τρόπο, ο οργανισμός μπορεί να σχεδιάσει τη διαφημιστική προώθηση, χωρίς να δαπανήσει περισσότερα χρήματα από όσα απαιτούνται.
- Δημογραφικοί δείκτες. Είναι σημαντικό για κάθε κυβέρνηση να έχει τη δυνατότητα να προβλέψει τους δημογραφικούς δείκτες στο μέλλον, όπως ανεργία, γεννήσεις, ρυθμός ανάπτυξης της οικονομίας κ.α, προκειμένου να είναι σε θέση να καθορίσει τα κατάλληλα μέτρα και πολιτικές που θα είναι προς όφελος του εθνικού συμφέροντος.

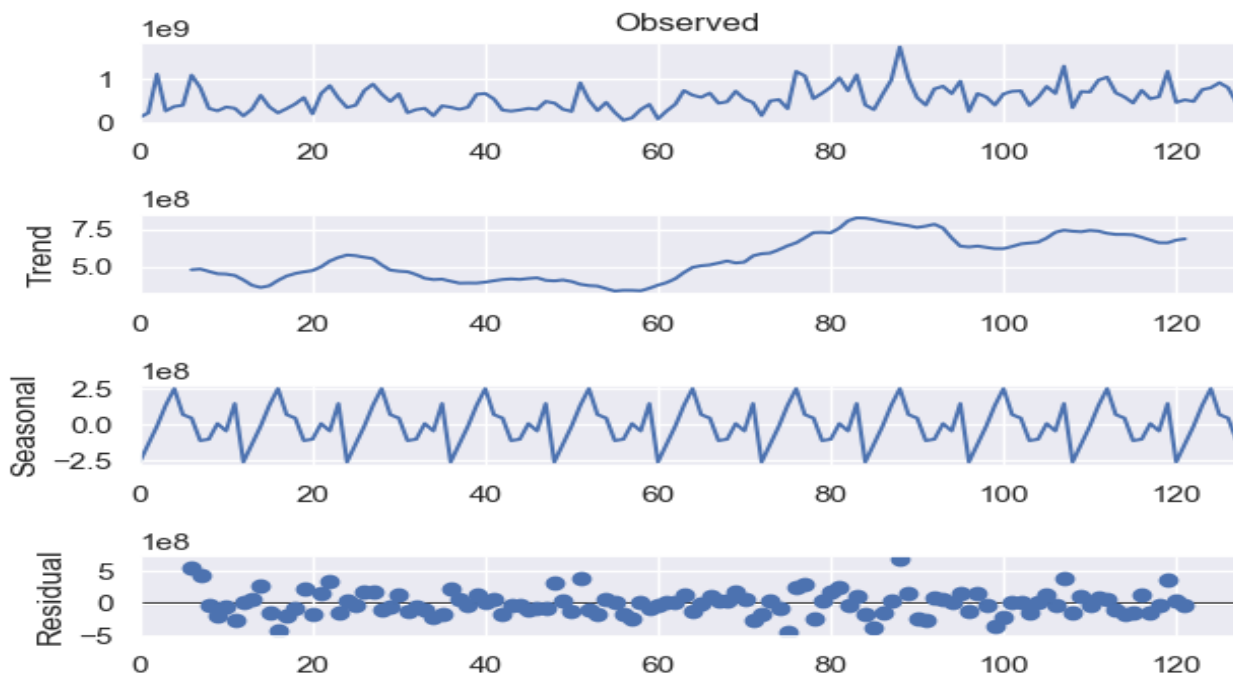
2.3.2 Τάση και Εποχικότητα Χρονοσειράς

Μία χρονοσειρά αποτελείται από τρεις βασικές συνιστώσες: την τάση (trend), την εποχικότητα (seasonality) και τα υπόλοιπα (residuals). Η τάση (trend) ορίζεται ως ένα μοτίβο που παρουσιάζει την αυξομείωση της συχνότητας μίας μεταβλητής ενδιαφέροντος στην

πάροδο του χρόνου. Ο όρος εποχικότητα (seasonality) διατυπώνεται ως η επαναλαμβανόμενη αυξομείωση της συχνότητας μίας μεταβλητής ενδιαφέροντος σε κάθε συγκεκριμένο χρονικό διάστημα. Το προσθετικό μαθηματικό μοντέλο που περιγράφει μία χρονοσειρά:

$$y_t = S_t + T_t + R_t \quad (I)$$

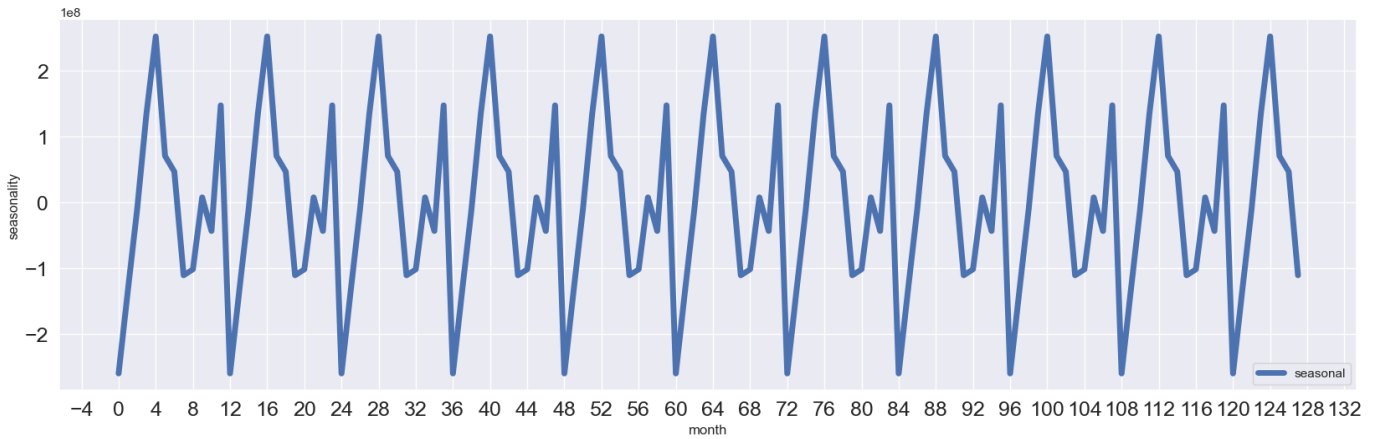
, όπου t είναι η χρονική στιγμή t , S_t : η εποχικότητα τη χρονική στιγμή t , T_t : η τάση τη χρονική στιγμή t και R_t : είναι η εναπομένουσα χρονοσειρά τη στιγμή t αν αφαιρέσουμε την τάση και την εποχικότητα [17]. Στην Εικόνα 2 παρουσιάζεται η αποσύνθεση της χρονοσειράς της Εικόνας 1 σε τάση (trend), εποχικότητα (seasonality) και υπόλοιπα (residuals).



Εικόνα 2 Αποσύνθεση της χρονοσειράς για την τιμή προσφοράς σε Ευρώ στην Ελλάδα το χρονικό διάστημα 2012-2022.

Από την τάση (trend) στην εικόνα 2 γίνεται αντιληπτό ότι οι τιμές προσφοράς σε Ευρώ στην Ελλάδα αυξήθηκαν από το 2021, ενώ τα τελευταία χρόνια παραμένουν σταθερά υψηλές. Η τιμή προσφοράς αυξάνεται από τον 60ο μήνα, δηλαδή από το έτος 2017, ενώ στη συνέχεια σταθεροποιείται. Επίσης, από την εποχικότητα (seasonality) παρατηρείται ετησίως ότι το μήνα Ιανουάριο οι τιμές προσφοράς είναι πολύ χαμηλές, ενώ οι μεγαλύτερες τιμές προσφοράς παρουσιάζονται τον μήνα Μάιο και Δεκέμβριο. Αυτό γίνεται εμφανές στην εικόνα 3.

Μηχανισμός έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD.

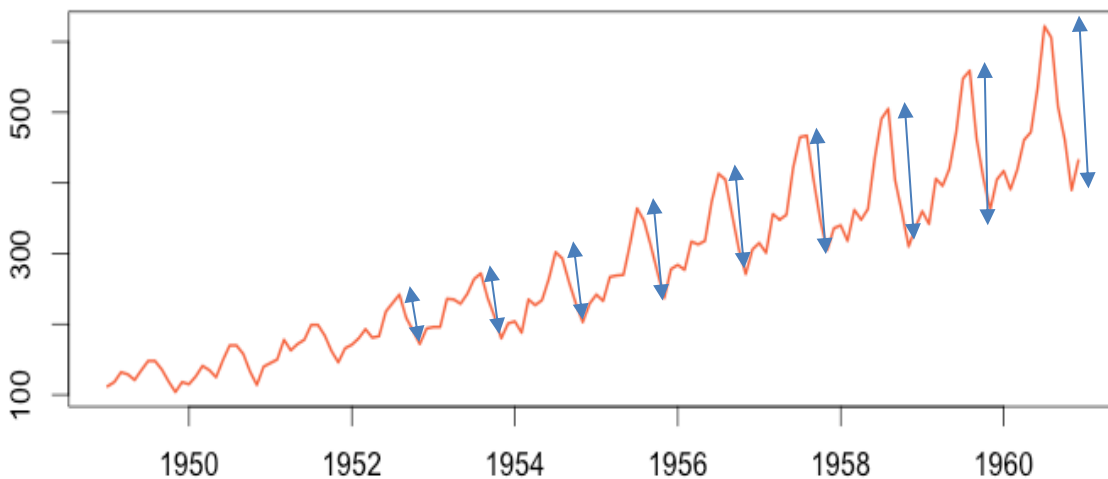


Εικόνα 3 Η εποχικότητα της τιμής προσφοράς σε Ευρώ τη χρονική περίοδο 2012 -2022.

Τέλος, το πολλαπλασιαστικό μαθηματικό μοντέλο μίας χρονοσειράς είναι:

$$y_t = S_t T_t R_t \quad (II)$$

Η διαφορά μίας χρονοσειράς που εκφράζεται με το πολλαπλασιαστικό μοντέλο και μίας χρονοσειράς που εκφράζεται με το προσθετικό μοντέλο, είναι ότι το πλάτος μίας χρονοσειράς με το πολλαπλασιαστικό μοντέλο αυξάνεται αναλογικά κάθε χρονική περίοδο. Στην εικόνα 4 παρουσιάζεται ένα παράδειγμα χρονοσειράς με πολλαπλασιαστικό μοντέλο [19].



Εικόνα 4 Χρονοσειρά με πολλαπλασιαστική εποχικότητα. Τα βελόνια δείχνουν το αυξανόμενο πλάτος της χρονοσειράς κάθε έτος. Πηγή: [18].

2.3.3 Τεχνικές Εξομάλυνσης Χρονοσειρών

Οι τεχνικές εξομάλυνσης χρονοσειρών δημιουργούν μία νέα χρονοσειρά απαλλαγμένη από θόρυβο, προκειμένου να γίνουν εμφανή διάφορα μοτίβα, όπως η τάση μίας χρονοσειράς. Μια βασική τεχνική εξομάλυνσης είναι ο απλός κινητός μέσος όρος (Simple Moving Average). Το μαθηματικό μοντέλο που το περιγράφει είναι:

$$m_T = \frac{1}{N} \sum_{t=T-N+1}^T y_t$$

όπου το T είναι μία χρονική στιγμή της χρονοσειράς, N είναι το πλήθος παρατηρήσεων που θα εφαρμοστεί ο κινητός μέσος όρος και ονομάζεται πλάτος παραθύρου (window width), και y_t η παρατήρηση τη χρονική στιγμή t [19]. Για παράδειγμα, έστω 5 παρατηρήσεις $y_t = \{2, 8, 2, 5, 9\}$, $t = 1, \dots, 5$ και πλάτος παραθύρου $N=3$ για κάθε χρονική στιγμή θα είναι:

Τη χρονική στιγμή $T=1$, δεν μπορεί να υπολογιστεί ο κινητός μέσος όρος, καθώς το $N=3$.

Τη χρονική στιγμή $T=2$, δεν μπορεί να υπολογιστεί ο κινητός μέσος όρος, καθώς το $N=3$.

Τη χρονική στιγμή $T=3$, ο απλός κινητός μέσος όρος θα είναι:

$$m_3 = \frac{1}{N} \sum_{t=1}^3 y_t = \frac{y_1 + y_2 + y_3}{3} = \frac{2 + 8 + 2}{3} = \frac{12}{3} = 4$$

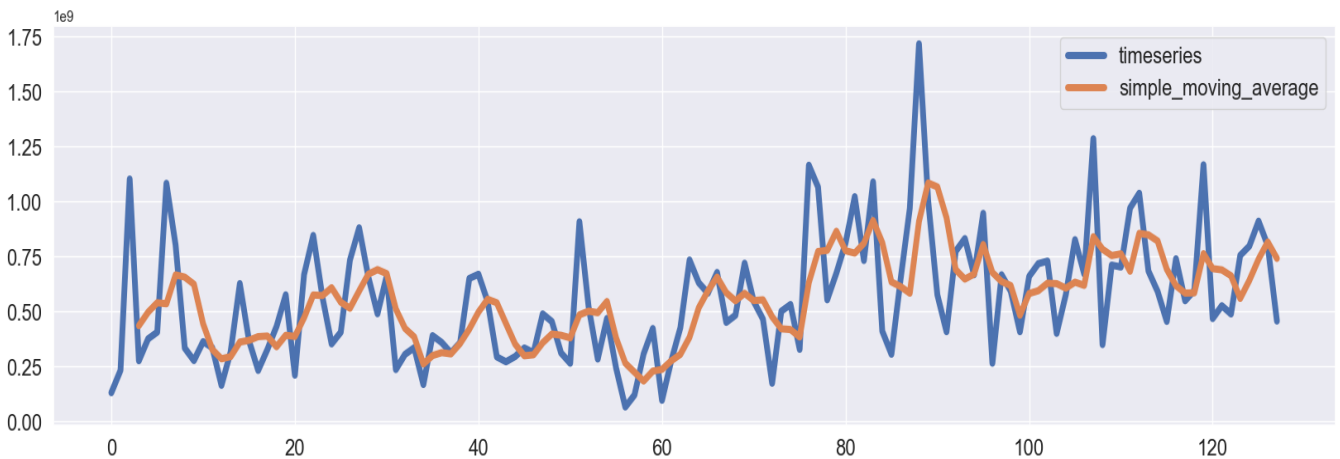
Τη χρονική στιγμή $T=4$, ο απλός κινητός μέσος όρος θα είναι:

$$m_3 = \frac{1}{N} \sum_{t=2}^4 y_t = \frac{y_2 + y_3 + y_4}{3} = \frac{8 + 2 + 5}{3} = \frac{15}{3} = 5$$

Τη χρονική στιγμή $T=5$, ο απλός κινητός μέσος όρος θα είναι:

$$m_4 = \frac{1}{N} \sum_{t=3}^5 y_t = \frac{y_3 + y_4 + y_5}{3} = \frac{2 + 5 + 9}{3} = \frac{16}{3} = 5.33$$

Επομένως, η εξομαλυμένη χρονοσειρά που θα προκύψει είναι οι παρατηρήσεις $\{4, 5, 5.33\}$. Παρατηρούμε ότι από 5 παρατηρήσεις προέκυψαν 3 παρατηρήσεις με βάση την τεχνική εξομάλυνσης του απλού κινητού μέσου (simple moving average), όμως αυτό δεν είναι πρόβλημα σε μία χρονοσειρά με πολύ μεγάλο αριθμό παρατηρήσεων, καθώς η απώλεια πληροφορίας είναι πολύ μικρή. Στην Εικόνα 5 παρουσιάζεται η εξομαλυμένη χρονοσειρά για την τιμή προσφοράς στην Ελλάδα την περίοδο 2012-2022 με την τεχνική του απλού κινητού μέσου (simple moving average) με πλάτος παραθύρου (window width) 4 μήνες.

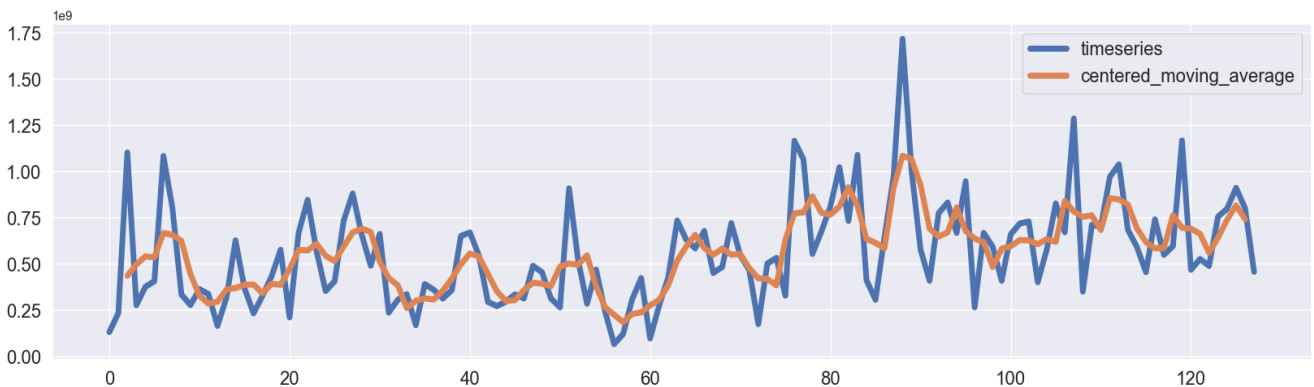


Εικόνα 5 Η χρονοσειρά της τιμής προσφοράς στην Ελλάδα το 2012-2022 (μπλέ χρώμα) και η εξομαλυμένη χρονοσειρά με την τεχνική απλού κινητού μέσου (πορτοκαλί χρώμα).

Άλλη μία σημαντική τεχνική εξομάλυνσης χρονοσειρών είναι ο κεντραρισμένος κινητός μέσος όρος (centered moving average), ο οποίος υπολογίζεται με τον τύπο:

$$m_T = \frac{1}{K+1} \sum_{i=-K}^K y_{T-i}$$

Στην Εικόνα 6 παρουσιάζεται η εξομαλυμένη χρονοσειρά για την τιμή προσφοράς στην Ελλάδα την περίοδο 2012-2022 με την τεχνική του του κεντραρισμένου κινητού μέσου (centered moving average) με πλάτος παραθύρου (window width) 4 μήνες.



Εικόνα 6 Η χρονοσειρά της τιμής προσφοράς στην Ελλάδα το 2012-2022 (μπλέ χρώμα) και η εξομαλυμένη χρονοσειρά του κεντραρισμένου κινητού μέσου (πορτοκαλί χρώμα).

Όπου το πλάτος παραθύρου (window width) $N = 2K + 1$. Το μειονέκτημα του απλού κινητού μέσου όρου (simple moving average) και του κεντραρισμένου μέσου όρου (centered moving average) είναι η ευαισθησία τους σε ακραίες τιμές (outliers). Για την εξάλειψη των outliers μπορεί να χρησιμοποιηθεί η τεχνική του κινητού διάμεσου (moving median), ο οποίος περιγράφεται από τον τύπο:

$$m_t^{[N]} = \text{med}(y_{t-k}, \dots, y_t, \dots, y_{t+k})$$

όπου $N = 2k$ είναι το πλάτος παραθύρου (window size) για το κινητό διάμεσο. Για παράδειγμα, αν έχουμε 5 παρατηρήσεις $y_t = \{2, 8, 2, 5, 9\}$, $t = 1, \dots, 5$ και πλάτος παραθύρου $N=3$ για κάθε χρονική στιγμή θα είναι:

Για $t=1$, δεν μπορεί να υπολογιστεί η κινητή διάμεσος, καθώς το $N=3$.

Για $t=2$, η διάμεσος θα είναι:

$$m_2^{[3]} = \text{med}(y_1, y_2, y_3) = \text{med}(2, 8, 2) = 2$$

Για $t=3$, η διάμεσος θα είναι:

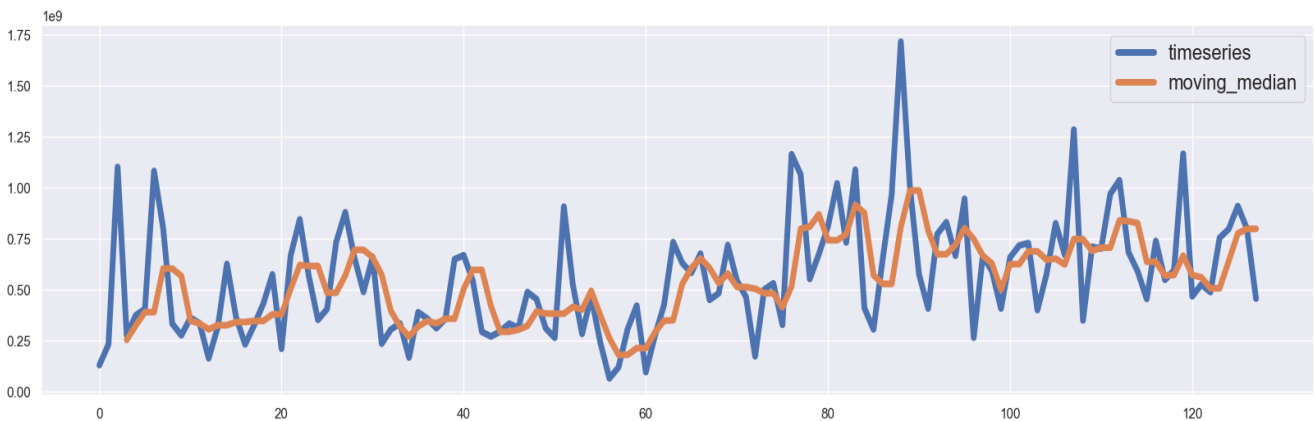
$$m_3^{[3]} = \text{med}(y_2, y_3, y_4) = \text{med}(8, 2, 5) = 5$$

Για $t=4$, η διάμεσος θα είναι:

$$m_4^{[3]} = \text{med}(y_3, y_4, y_5) = \text{med}(2, 5, 9) = 5$$

Για $t=5$, δεν μπορεί να υπολογιστεί η κινητή διάμεσος, καθώς το $N=3$.

Επομένως, οι παρατηρήσεις που προέκυψαν από την εφαρμογή της κινητής διάμεσου (moving median) είναι $\{8, 5, 5\}$. Στην Εικόνα 7 παρουσιάζεται η εξομαλυμένη χρονοσειρά για την τιμή προσφοράς στην Ελλάδα την περίοδο 2012-2022 με την τεχνική του του κινητού διάμεσου (moving median) με πλάτος παραθύρου (window width) 4 μήνες.



Εικόνα 7 Η χρονοσειρά της τιμής προσφοράς στην Ελλάδα το 2012-2022 (μπλέ χρώμα) και η εξομαλυμένη χρονοσειρά του κινητού διάμεσου (πορτοκαλί χρώμα).

2.4 Μοντέλα Μηχανικής Μάθησης

2.4.1 Τι είναι «Μοντέλο Μηχανικής Μάθησης».

Ένα μοντέλο μηχανικής μάθησης είναι το αποτέλεσμα εκτέλεσης ενός αλγόριθμου μηχανικής μάθησης. Πιο συγκεκριμένα, ένα προβλεπτικό μοντέλο δημιουργείται με την εκτέλεση ενός αλγόριθμου μηχανικής μάθησης στα δεδομένα εκπαίδευσης, με σκοπό την πρόβλεψη μελλοντικών τιμών ή την εύρεση των μοτίβων στα δεδομένα [19]. Τα προβλεπτικά μοντέλα διαχωρίζονται συνήθως σε τέσσερις (4) κατηγορίες [20]:

- Μοντέλα Συσταδοποίησης (Clustering Models). Ο σκοπός δημιουργίας των μοντέλων συσταδοποίησης είναι η ομαδοποίηση δεδομένων με κοινά χαρακτηριστικά. Οι ομάδες που θα προκύψουν μπορούν να αναλυθούν, προκειμένου να ερμηνευτεί η συμπεριφορά τους και να καθοριστούν μελλοντικά συμπεράσματα. Αλγόριθμοι που χρησιμοποιούνται για τη δημιουργία μοντέλων συσταδοποίησης είναι ο K-Means, DBSCAN κλπ.
- Μοντέλα Κατηγοριοποίησης (Classification Models). Τα μοντέλα κατηγοριοποίησης εκπαιδεύονται σε δεδομένα με προκαθορισμένες ετικέτες για κάθε μία παρατήρηση, προκειμένου να προβλέψουν την ετικέτα μίας νέας παρατήρησης. Αλγόριθμοι που χρησιμοποιούνται για τη δημιουργία μοντέλων κατηγοριοποίησης είναι η Λογιστική Παλινδρόμηση (Logistic Regression), το Δένδρο Απόφασης (Decision Tree) κλπ.
- Μοντέλα Ακραίων Τιμών (Outlier Models). Τα μοντέλα ακραίων τιμών χρησιμοποιούνται για τον εντοπισμό ακραίων τιμών σε ένα σύνολο δεδομένων, δηλαδή παρατηρήσεις που δεν ακολουθούν το μοτίβο των δεδομένων και είναι ιδιαίτερα χρήσιμα για την ανίχνευση ανωμαλιών σε ένα σύστημα ή περιπτώσεων απάτης. Αλγόριθμοι που χρησιμοποιούνται για τη δημιουργία μοντέλων ακραίων τιμών είναι ο Isolation Forest, ο LOF κλπ.
- Μοντέλα Χρονοσειρών (Time Series Model). Τα μοντέλα χρονοσειρών εκπαιδεύονται με ιστορικά δεδομένα για την ανίχνευση μοτίβων, που μπορούν να χρησιμοποιηθούν για τη δημιουργία προβλέψεων στο μέλλον. Συνεπώς, τα μοντέλα χρονοσειρών χρησιμοποιούνται για την εύρεση μελλοντικών μοτίβων με βάση τα διαθέσιμα ιστορικά δεδομένα. Αλγόριθμοι που χρησιμοποιούνται για τη δημιουργία μοντέλων χρονοσειρών είναι ο ARIMA, MA κλπ.

2.4.2 Μεθοδολογία Δημιουργίας Μοντέλων Μηχανικής Μάθησης

Η μεθοδολογία δημιουργίας μοντέλων μηχανικής μάθησης περιλαμβάνει τα ακόλουθα βήματα [21]:

1. **Συλλογή δεδομένων (data collection)** από διάφορες πηγές, όπως csv αρχεία, βάσεις δεδομένων κλπ.
2. **Καθαρισμός δεδομένων (data cleaning)**, αφαιρώντας ακραίες τιμές (filtering), συμπλήρωση ελλειπουσών τιμών (imputation) και πιθανώς αλλαγή της κλίμακας των δεδομένων (transformation), προκειμένου η δομή των δεδομένων να βοηθήσει στη δημιουργία προβλεπτικού μοντέλου με μικρότερα σφάλματα στις προβλέψεις του.

3. **Επιλογή των χαρακτηριστικών (feature selection)** από το σύνολο δεδομένων, χρησιμοποιώντας διάφορες τεχνικές, όπως χ^2 test, Fischer Score κλπ, διατηρώντας την αρχή του parsimony. Η αρχή του parsimony δίνει έμφαση στη δημιουργία ερμηνεύσιμων και αποδοτικών προβλεπτικών μοντέλων.
4. **Εκπαίδευση του προβλεπτικού μοντέλου (model training)**. Διαχωρισμός των ήδη υπάρχοντων δεδομένων σε εκπαίδευσης (training data), δεδομένα αξιολόγησης (validation data) και δεδομένα δοκιμής (testing data). Τα δεδομένα εκπαίδευσης (training data) χρησιμοποιούνται για την εύρεση υπολανθανόντων μοτίβων στα δεδομένα που θα βοηθήσουν στη δημιουργία ενός μοντέλου. Τα δεδομένα αξιολόγησης (validation data) είναι χρήσιμα για τη βελτιστοποίηση του μοντέλου, ρυθμίζοντας τις παραμέτρους του (hyperparameter tuning), χωρίς να υπάρχει υπερπροσαρμογή (overfitting) του μοντέλου στα δεδομένα.
5. **Αξιολόγηση του προβλεπτικού μοντέλου (model evaluating)**. Σε αυτή τη φάση χρησιμοποιούνται τα δεδομένα δοκιμής (testing data) τα οποία εφαρμόζονται στο προβλεπτικό μοντέλο και με διάφορες μετρικές αξιολογείται η απόδοση του, όπως για παράδειγμα η ρίζα του μέσου τετραγωνικού σφάλματος (RMSE), ακρίβεια (accuracy) των προβλέψεων του μοντέλου κλπ.
6. **Εφαρμογή του προβλεπτικού μοντέλου σε παραγωγική διαδικασία (deploy model)**. Το καλύτερο μοντέλο που επιλέχθηκε από τη φάση της αξιολόγησης, χρησιμοποιείται για προβλέψεις σε παραγωγικό περιβάλλον.

2.4.3 Το αυτοπαλινδρομικό μοντέλο (Autoregressive Model)

Το αυτοπαλινδρομικό μοντέλο (autoregressive model) είναι ένας γραμμικός συνδυασμός των τιμών προηγούμενων χρονικών στιγμών μίας στάσιμης χρονοσειράς. Ένα αυτοπαλινδρομικό μοντέλο (autoregressive model) τάξης p το οποίο συμβολίζεται ως AR(p) [22], περιγράφεται με το ακόλουθο μαθηματικό μοντέλο:

$$y_t = c + \varphi_1 y_{t-1} + \varphi_2 y_{t-2} + \dots + \varphi_p y_{t-p} + \varepsilon_t$$

, όπου $\varphi_1, \varphi_2, \dots, \varphi_p$ είναι παράμετροι οι οποίοι υπολογίζονται ανάλογα το μοντέλο, και $y_{t-1}, y_{t-2}, \dots, y_{t-p}$ είναι παρελθούσες τιμές της χρονοσειράς και ε_t ο λευκός θόρυβος. Με τη χρήση του τελεστή οπισθοχώρησης (backshift operator), ο τύπος του αυτοπαλινδρομικού μοντέλου (autoregressive model) τάξης p , γράφεται ως εξής:

$$y_t = c + \sum_{i=1}^p \varphi_i B^i y_t + \varepsilon_t$$

Για παράδειγμα, ένα αυτοπαλινδρομικό μοντέλο τάξης 2, δηλαδή AR(2) αναπτύσσεται ως εξής:

$$y_t = c + \sum_{i=1}^2 \varphi_i B^i y_t = c + \varphi_1 B y_t + \varphi_2 B^2 y_t + \varepsilon_t = c + \varphi_1 y_{t-1} + \varphi_1 y_{t-2} + \varepsilon_t$$

Επομένως, αν η παρούσα χρονική στιγμή είναι η T , τότε ο υπολογισμός μίας μελλοντικής τιμής τη χρονική $T+1$ σε AR(2) μοντέλο θα είναι:

$$y_{T+1} = c + \sum_{i=1}^2 \varphi_i B^i y_{T+1} = c + \varphi_1 B y_t + \varphi_2 B^2 y_t + \varepsilon_{T+1} = c + \varphi_1 y_T + \varphi_1 y_{T-1} + \varepsilon_{T+1}$$

Επειδή δεν είναι γνωστή η τιμή του λευκού θορύβου τη χρονική στιγμή $T+1$, τότε θεωρείται ότι αυτή είναι μηδέν. Άρα προκύπτει ότι:

$$y_{T+1} = c + \varphi_1 y_T + \varphi_1 y_{T-1}$$

Με όμοιο τρόπο υπολογίζονται και για μελλοντικές χρονικές στιγμές $T+h$, όπου $h = 1, \dots, n$. Η εφαρμογή ενός αυτοπαλινδρομικού μοντέλου γίνεται σε μία στασιμή χρονοσειρά, δηλαδή μία χρονοσειρά απαλλαγμένη από τάση (trend) και εποχικότητα (seasonality). Η διαπίστωση αν μία χρονοσειρά είναι στάσιμη ή όχι μπορεί να γίνει με τη χρήση στατιστικών τεστ, όπως το KPSS και το ADF τεστ. Αν η χρονοσειρά δεν είναι στάσιμη, τότε μπορεί να χρησιμοποιηθεί η τεχνική της εποχικής διαφοροποίησης (seasonal differencing). Η εποχική διαφοροποίηση (seasonal differencing) πρώτης τάξης [17], περιγράφεται από τον τύπο:

$$y'_t = y_t - B y_t = y_t - y_{t-1}$$

Το αποτέλεσμα είναι μία νέα πιθανώς στάσιμη χρονοσειρά, από τη διαφορά της τιμής της τρέχουσας χρονικής στιγμής t με την τιμή της προηγούμενης χρονικής στιγμής $t-1$ της αρχικής χρονοσειράς. Αν η νέα χρονοσειρά που θα προκύψει μετά την εποχική διαφοροποίηση (seasonal differencing) πρώτης τάξης δεν είναι στάσιμη, τότε εφαρμόζεται εποχική διαφοροποίηση (seasonal differencing) δεύτερης τάξης [17], η οποία περιγράφεται από τον τύπο:

$$\begin{aligned} y''_t &= y'_t - B y'_t = y'_t - y'_{t-1} \\ &= y_t - y_{t-1} - (y_{t-1} - y_{t-2}) \\ &= y_t - 2y_{t-1} - y_{t-2} \\ &= y_t - 2B y_t - B^2 y_t \end{aligned}$$

Ομοίως συνεχίζεται η διαδικασία έως ότου να προκύψει μία στάσιμη χρονοσειρά. Η χρονοσειρά είναι συνήθως στάσιμη με εποχική διαφοροποίηση (seasonal differencing) πρώτης τάξης. Τέλος, το αυτοπαλινδρομικό μοντέλο είναι χρήσιμο αν είναι γνωστό ότι υπάρχει μεγάλη συσχέτιση μεταξύ της τρέχουσας τιμής και των προηγούμενων τιμών, το οποίο μπορεί να γίνει οπτικά αντιληπτό με το διάγραμμα μερικής αυτοσυσχέτισης (PACF).

2.4.4 Το μοντέλο κινητών μέσων (Moving Average Model)

Το μοντέλο κινητών μέσων (Moving Average Model) αποτελεί ένα γραμμικό συνδυασμό των σφαλμάτων της τρέχουσας χρονικής, καθώς και των προηγούμενων χρονικών στιγμών μίας στάσιμης χρονοσειράς. Ένα μοντέλο κινητών μέσων τάξης q , συμβολίζεται ως MA(q) [17] και περιγράφεται από το ακόλουθο μαθηματικό μοντέλο:

$$y_t = c + \varepsilon_t + \sum_{i=1}^{q-1} m_i B^i \varepsilon_{t-i} = c + \varepsilon_t + m_1 \varepsilon_{t-1} + \dots + m_q \varepsilon_{t-q-1}$$

Για παράδειγμα, ένα μοντέλο κινητών μέσων τάξης MA(3) γράφεται ως εξής:

$$y_t = c + \varepsilon_t + \sum_{i=1}^2 m_i B^i \varepsilon_{t-i} = c + \varepsilon_t + m_1 \varepsilon_{t-1} + m_2 \varepsilon_{t-2}$$

Ένα μοντέλο κινητών μέσων είναι χρήσιμο όταν θεωρείται ότι υπάρχει μεγάλη επίδραση των σφαλμάτων στη χρονοσειρά, κάτι το οποίο θα επηρεάζει και τις μελλοντικές τιμές της.

2.4.5 Το μοντέλο ARIMA

Το μοντέλο ARIMA αποτελείται από την ενοποίηση του αυτοπαλινδρομικού μοντέλου (autoregressive model) και του μοντέλου κινητών μέσων (moving average model), χρησιμοποιώντας και την τεχνική της εποχικής διαφοροποίησης (seasonal differencing). Το μοντέλο ARIMA συμβολίζεται ως ARIMA(p, d, q), όπου p είναι η τάξη του αυτοπαλινδρομικού μοντέλου (autoregressive model), d είναι η τάξη της εποχικής διαφοροποίησης (seasonal differencing) και q είναι η τάξη του μοντέλου κινητών μέσων (moving average model). Ένα ARIMA(p, d, q) [17] περιγράφεται με το ακόλουθο μοντέλο:

$$\left(1 - \sum_{i=1}^p \varphi_i B^i\right) (1 - B)^d y_t = c + \left(1 + \sum_{i=1}^q \varphi_i B^i\right) \varepsilon_t$$

Για παράδειγμα ένα μοντέλο ARIMA(3,1,1) μπορεί να γραφεί ως:

$$\left(1 - \sum_{i=1}^3 \varphi_i B^i\right) (1 - B)^1 y_t = c + \left(1 + \sum_{i=1}^1 m_i B^i\right) \varepsilon_t \Leftrightarrow$$

$$\Leftrightarrow (1 - \varphi_1 B - \varphi_2 B^2 - \varphi_3 B^3)(1 - B)y_t = c + (1 + m_1 B)\varepsilon_t \Leftrightarrow$$

$$\Leftrightarrow (1 - \varphi_1 B - \varphi_2 B^2 - \varphi_3 B^3)y_t - (1 - \varphi_1 B - \varphi_2 B^2 - \varphi_3 B^3)B y_t = c + \varepsilon_t + m_1 B \varepsilon_t \Leftrightarrow$$

$$\Leftrightarrow y_t - \varphi_1 B y_t - \varphi_2 B^2 y_t - \varphi_3 B^3 y_t - B y_t + \varphi_1 B^2 y_t + \varphi_2 B^3 y_t + \varphi_3 B^4 y_t$$

$$= c + \varepsilon_t + m_1 B \varepsilon_t \Leftrightarrow$$

$$\Leftrightarrow y_t - \varphi_1 y_{t-1} - \varphi_2 y_{t-2} - \varphi_3 y_{t-3} - y_{t-1} + \varphi_1 y_{t-2} + \varphi_2 y_{t-3} + \varphi_3 y_{t-4} = c + \varepsilon_t + m_1 \varepsilon_{t-1} \Leftrightarrow$$

$$\Leftrightarrow y_t - (\varphi_1 + 1)y_{t-1} + (\varphi_1 - \varphi_2)y_{t-2} + (\varphi_2 - \varphi_3)y_{t-3} + \varphi_3 y_{t-4} = c + \varepsilon_t + m_1 \varepsilon_{t-1} \Leftrightarrow$$

$$\Leftrightarrow y_t = (\varphi_1 + 1)y_{t-1} - (\varphi_1 - \varphi_2)y_{t-2} - (\varphi_2 - \varphi_3)y_{t-3} - \varphi_3 y_{t-4} + c + \varepsilon_t + m_1 \varepsilon_{t-1}$$

Η πρόβλεψη τη χρονική στιγμή T+1 για το μοντέλο ARIMA(3,1,1) θα είναι:

$$y_{T+1} = (\varphi_1 + 1)y_T - (\varphi_1 - \varphi_2)y_{T-1} - (\varphi_2 - \varphi_3)y_{T-2} - \varphi_3 y_{T-3} + c + \varepsilon_{T+1} + m_1 \varepsilon_T$$

Επειδή δεν είναι γνωστό τη μελλοντική χρονική T+1 η τιμή του σφάλματος αυτό αντικαθίσταται με το μηδέν, άρα προκύπτει ότι:

$$y_{T+1} = (\varphi_1 + 1)y_T - (\varphi_1 - \varphi_2)y_{T-1} - (\varphi_2 - \varphi_3)y_{T-2} - \varphi_3 y_{T-3} + c + m_1 \varepsilon_T$$

Ομοίως, υπολογίζονται οι τιμές για μελλοντικές τιμές $T + h$, όπου $h = 2, \dots, n$.

Το πλεονέκτημα του μοντέλου ARIMA σε σχέση με το αυτοπαλινδρομικό μοντέλο (autoregressive model) και το μοντέλο κινητών μέσων (moving average model) είναι ότι περιλαμβάνει εποχική διαφοροποίηση (seasonal differencing), καθώς οι χρονοσειρές σε πραγματικά προβλήματα επηρεάζονται από τάση και εποχικότητα. Τέλος, λαμβάνει υπόψη τις προηγούμενες τιμές της χρονοσειράς και τον λευκό θόρυβο, τα οποία επιδρούν στη διαμόρφωση μίας μελλοντικής τιμής της χρονοσειράς, με συνέπεια η πρόβλεψη συνήθως να είναι κοντά στην πραγματική τιμή της χρονοσειράς και το διάστημα πρόβλεψης (prediction interval) του μοντέλου να εμπεριέχει τη μελλοντική τιμή.

2.4.6 Το μοντέλο Prophet

Το μοντέλο Prophet αποτελεί ένα μη γραμμικό μοντέλο που συνδυάζει την τάση ,που ονομάζεται ανάπτυξη (growth term), την εποχικότητα (seasonality), τις διακοπές (holidays) και λευκό θόρυβο [23, p. 7]. Το προσθετικό μοντέλο (additive model) εκφράζεται ως :

$$y_t = g(t) + s(t) + h(t) + \varepsilon_t$$

Είναι ευρέως γνωστό ότι η τάση δεν είναι γραμμική στις περισσότερες χρονοσειρές, και συνήθως υπάρχουν απότομες αλλαγές λόγω εξωγενών παραγόντων, για παράδειγμα η εμφάνιση της πανδημίας του covid-19 ελαχιστοποίησε παγκοσμίως τον αριθμό πτήσεων. Επομένως, ένα μοντέλο σαν το ARIMA δεν θα μπορούσε να προβλέψει μία απότομη αλλαγή στην τάση, με συνέπεια το σφάλμα των προβλέψεων να είναι εξαιρετικά μεγάλο. Η λύση σε αυτό το πρόβλημα δίνεται από το μοντέλο Prophet, το οποίο λαμβάνει υπόψη τις απότομες αλλαγές στην τάση μοντελοποιώντας την είτε με το τμηματικό γραμμικό μοντέλο (piecewise linear model) είτε με το τμηματικό λογιστικό μοντέλο (piecewise logistic growth model). Το τμηματικό λογιστικό μοντέλο ανάπτυξης (piecewise logistic growth model) στη γενική μορφή του ορίζεται ως [23, p. 8]:

$$g(t) = \frac{C(t)}{1 + e^{\alpha(t)}}$$

,όπου $C(t)$ είναι η φέρουσα ικανότητα (carrying capacity) η οποία ορίζει τη μέγιστη τιμή της καμπύλης η οποία αλλάζει κάθε χρονική στιγμή. Η συνάρτηση $\alpha(t)$ μοντελοποιεί τις αλλαγές στις τάσεις για το σημείο αλλαγής c (changepoint) και ορίζεται ως:

$$\alpha(t) = \begin{cases} km - kt, & t < c \\ -(k + \delta)(t - (m + \gamma)), & t \geq c \end{cases}$$

,όπου m είναι μία παράμετρος που εκφράζει το πλήθος των προηγούμενων χρονικών στιγμών, το k είναι η κλίση της καμπύλης (logistic growth rate) και γ μια παράμετρος που ρυθμίζει τα σωστά σημεία αλλαγής.

Το τμηματικό γραμμικό μοντέλο (piecewise linear model) για ένα σημείο αλλαγής c είναι:

$$g(t) = \begin{cases} kt + m, & t < c \\ (k + \delta)t + (m + \gamma), & t \geq c \end{cases}$$

Στη συνέχεια, η εποχικότητα (seasonality) ορίζεται ως μία σειρά Fourier, ομαλοποιώντας την εποχικότητα. Το μαθηματικό μοντέλο της εποχικότητας [18, p. 11] είναι:

$$s(t) = \sum_{n=1}^N \left(a_n \cos\left(\frac{2\pi nt}{P}\right) + b_n \sin\left(\frac{2\pi nt}{P}\right) \right)$$

Όπου P είναι η περίοδος της χρονοσειράς, για $P=7$ η περίοδος είναι εβδομαδιαία, ενώ για $P=365.25$ είναι ετήσια. Το δεκαδικό μέρος του 365.25 είναι λόγω της ύπαρξης δίσεκτων ετών. Αν οι παράμετροι είναι το διάνυσμα $\gamma = [a_1, b_1, \dots, a_N, b_N]$, τότε το διάνυσμα L για κάθε τιμή του χρόνου t για N παρατηρήσεις, είναι της μορφής:

$$L(t) = \left[\cos\left(\frac{2\pi t}{P}\right), \dots, \sin\left(\frac{2\pi Nt}{P}\right) \right]$$

Άρα η εποχικότητα εκφράζεται ως:

$$s(t) = L(t)\gamma$$

με $\gamma \sim N(0, \sigma^2)$, όπου σ^2 είναι η τυπική απόκλιση των παραμέτρων $a_1, b_1, \dots, a_N, b_N$. Η τιμή του σ^2 είναι παράμετρος η οποία αφορά την προγενέστερη κλίμακα εποχικότητας (seasonality prior scale), η οποία όσο μεγαλύτερη είναι η τιμή τόσο ευρύτερη είναι η κατανομή για τα $2N$ δείγματα του Fourier. Για την αναπαράσταση του μοντέλου των διακοπών (holidays), χρησιμοποιείται ένα σύνολο T_j που περιλαμβάνει τις χρονικές στιγμές για κάθε διακοπή (holiday) j . Επομένως, οι διακοπές (holidays) εκφράζονται ως [23, p. 12]:

$$h(t) = \sum_{t \in T_j} k_j$$

για $j = 1, 2, \dots, L$, όπου L είναι το πλήθος των διακοπών σε ένα έτος (holidays) και k_j είναι η μεταβολή της πρόβλεψης για τη διακοπή (holiday) j . Το προσθετικό Prophet μοντέλο με τμηματικό γραμμικό μοντέλο (piecewise linear model) της τάσης εφαρμόστηκε για τη δημιουργία προβλεπτικών μοντέλων στην υπηρεσία ανάλυσης δεδομένων που αποτελεί μέρος του μηχανισμού έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD. Ο βασικός λόγος επιλογής του μοντέλου Prophet είναι η δυνατότητα πιο ακριβών προβλέψεων σε σχέση με τα υπόλοιπα μοντέλα προβλέψεων, λαμβάνοντας υπόψη απότομες αλλαγές στην τάση μίας χρονοσειράς, κάτι το οποίο αποτελεί συχνό φαινόμενο σε εφαρμογές που βρίσκονται σε παραγωγικό περιβάλλον.

3. ΜΕΘΟΔΟΛΟΓΙΑ ΑΝΑΛΥΣΗΣ, ΣΧΕΔΙΑΣΗΣ ΚΑΙ ΑΝΑΠΤΥΞΗΣ ΣΥΣΤΗΜΑΤΟΣ

3.1 Εισαγωγή

Στο κεφάλαιο 3 περιγράφεται το μοντέλο μεταδεδομένων «Ευρωπαϊκού Ενιαίου Έγγραφου Σύμβασης» (ESPD), που υλοποιήθηκε για τη δημιουργία αρχείων καταγραφής (log). Στη συνέχεια, γίνεται αναφορά στον μηχανισμό αποθήκευσης των αρχείων καταγραφής του επιχειρησιακού συστήματος ESPD. Επίσης, παρουσιάζεται η δομή και η λειτουργία του μηχανισμού έκδοσης αναφορών για το σύστημα ESPD, καθώς και της υπηρεσίας ανάλυσης δεδομένων των αρχείων καταγραφής (log) του ESPD. Τέλος, γίνεται περιγραφή του μηχανισμού έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD, ο οποίος αποτελεί το σύνολο των προαναφερθέντων μηχανισμών, υλοποιώντας ένα σύστημα επιχειρησιακής ευφυΐας, το οποίο προσφέρει έκδοση αναφορών (reporting) και παρακολούθηση (monitoring) των εξαγωγών ESPD αρχείων σε πραγματικό χρόνο (real-time) για το σύστημα ESPD.

3.2 Το μοντέλο μεταδεδομένων του «Ευρωπαϊκού Ενιαίου Έγγραφου Σύμβασης» (ESPD) για τη δημιουργία αρχείων καταγραφής

Το Ευρωπαϊκό Ενιαίο Έγγραφο Έγγραφο Σύμβασης (ESPD) αποτελεί μία δήλωση των οικονομικών φορέων που αποδεικνύει ότι πληρούν ή όχι τα κριτήρια επιλογής και αποκλεισμού για τη συμμετοχή τους σε ένα διαγωνισμό σύμβασης προμηθειών που πραγματοποιείται σε μία χώρα της Ευρωπαϊκής Ένωσης. Η δομή ενός ESPD εγγράφου, περιγράφεται από το ESPD-EDM. Πιο συγκεκριμένα, το ESPD-EDM αποτελεί ένα μοντέλο δεδομένων που στοχεύει στη σημασιολογική διαλειτουργικότητα, μεταξύ των ευρωπαϊκών υπηρεσιών ESPD. Αυτό το μοντέλο δεδομένων διαχωρίζεται σε δύο κατηγορίες τα ESPD Request και ESPD Response. Το ESPD Request μοντέλο δεδομένων περιλαμβάνει τα απαραίτητα δεδομένα, με τα οποία η αναθέτουσα αρχή γνωστοποιεί τα κριτήρια αποκλεισμού και επιλογής στους οικονομικούς φορείς, ενώ το ESPD Response μοντέλο δεδομένων περιλαμβάνει τα απαραίτητα δεδομένα, με τα οποία οι οικονομικοί φορείς απαντούν στις απαιτήσεις των αναθετουσών αρχών με πειστήρια (evidence) [24]. Κάθε χώρα της Ευρωπαϊκής Ένωσης έχει υλοποιήσει ένα σύστημα ESPD, το οποίο είναι διαλειτουργικό με τα συστήματα ESPD των υπόλοιπων κρατών μελών. Στην Ελλάδα, έχει υλοποιηθεί το σύστημα Promitheus ESPDint για την έκδοση εγγράφων ESPD. Στην παρούσα εργασία κρίθηκε αναγκαία η δημιουργία αρχείων καταγραφής (log) των εξαγωγών ESPD αρχείων. Στην εικόνα 8 φαίνεται το μοντέλο μεταδεδομένων για τη δημιουργία των αρχείων καταγραφής

ESPDLog
-logId: UUID
-year: Integer
-date: LocalDate
-time: LocalTime
-day: String
-lang: EULanguageCodeEnum
-fileFormat: DocumentType
-artifactType: ArtefactType
-version: EDMVersion
-filesCount: Long

Εικόνα 8 Το μοντέλο μεταδεδομένων του ESPD για τη δημιουργία αρχείων καταγραφής (log).

Τα πεδία από τα οποία αποτελείται το μοντέλο μεταδεδομένων ESPDLog είναι το αναγνωριστικό του αρχείου καταγραφής (logId), το έτος που δημιουργήθηκε το αρχείο καταγραφής (year), την ημερομηνία που δημιουργήθηκε το αρχείο καταγραφής (date), την ώρα που δημιουργήθηκε το αρχείο καταγραφής (time), την ημέρα που δημιουργήθηκε το αρχείο καταγραφής (day), τη γλώσσα (lang), το τύπο αρχείου (fileFormat), δηλαδή PDF, HTML ή XML, το τύπο του ESPD μοντέλου (artifactType), δηλαδή ESPD Request ή ESPD Response, και την έκδοση (version) του ESPD αρχείου, που εξήχθη από το σύστημα ESPD. Το μοντέλο μεταδεδομένων ESPDLog χρησιμοποιεί το Builder design pattern, παρέχοντας μία δυναμική δομή [25, p. 97], η οποία δεν είναι σαφώς καθορισμένη από τη δημιουργία του αντικειμένου. Επομένως, το πεδίο που περιέχει το πλήθος των ESPD αρχείων (filesCount) χρησιμοποιείται σε περιπτώσεις, όπου πραγματοποιείται ανάκτηση δεδομένων με συγκεντρωτικές επερωτήσεις (aggregated queries) στη βάση δεδομένων ksqIDB, που αναφέρεται στο κεφάλαιο 4.2.3.1 καθώς και για την αποθήκευση των αρχείων καταγραφής. Συνεπώς, παρέχει τα απαραίτητα μεταδεδομένα που χρειάζονται από το μοντέλο δεδομένων του ESPD-EDM, προκειμένου να υλοποιηθεί η έκδοση αναφορών (reporting) και η παρακολούθηση (monitoring) των εξαγωγών ESPD αρχείων σε πραγματικό χρόνο.

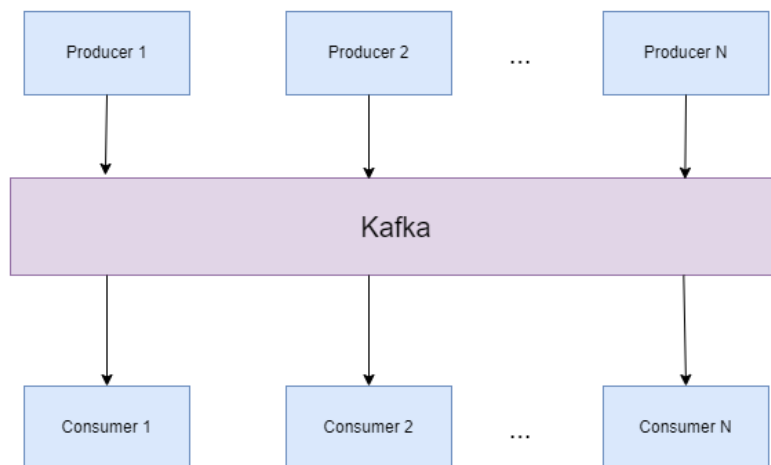
3.3 Μηχανισμός Αποθήκευσης Αρχείων Καταγραφής ESPD

Ο μηχανισμός αποθήκευσης αρχείων καταγραφής για το επιχειρησιακό σύστημα ESPD υλοποιήθηκε με τη χρήση του Kafka, το οποίο αποτελεί μια κατανεμημένη event stream processing platform, δηλαδή λαμβάνει δεδομένα από διάφορες πηγές, όπως βάσεις δεδομένων, σενσορες κλπ σε πραγματικό χρόνο (real-time), και στέλνει αυτά τα δεδομένα σε άλλες εφαρμογές. Οι 3 βασικές δυνατότητες που παρέχει το Kafka είναι [26]:

- Αποστολή και λήψη εγγραφών (Publish/Subscribe)

- Αποθήκευση των εγγραφών με ανοχή σε σφάλματα (fault tolerance)
- Επεξεργασία ροών δεδομένων (stream processing), όπου απαιτείται.

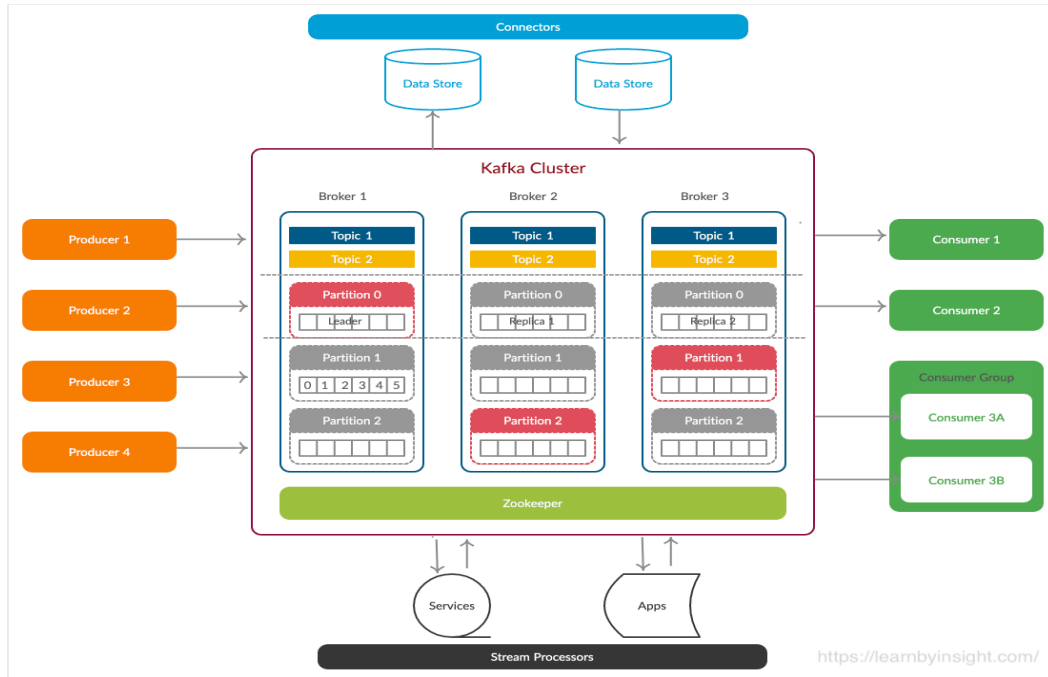
Οι producers αποτελούν την πηγή δεδομένων για ένα Kafka cluster και επικοινωνούν με ένα ή περισσότερους broker, ενώ οι consumers αποτελούν το προορισμό των δεδομένων που στέλνουν οι producers σε ένα Kafka cluster. Οι producers μπορεί να είναι εφαρμογές, σενσορες, IoT συσκευές κλπ., ενώ οι consumers μπορεί να είναι βάσεις δεδομένων, εφαρμογές κλπ., που λαμβάνουν σε πραγματικό χρόνο (real-time) τα νέα δεδομένα. Στην Εικόνα 9 παρουσιάζεται η ροή των δεδομένων από τους Producers στους Consumers.



Εικόνα 9 Η ροή των δεδομένων από τους Producers στους Consumers.

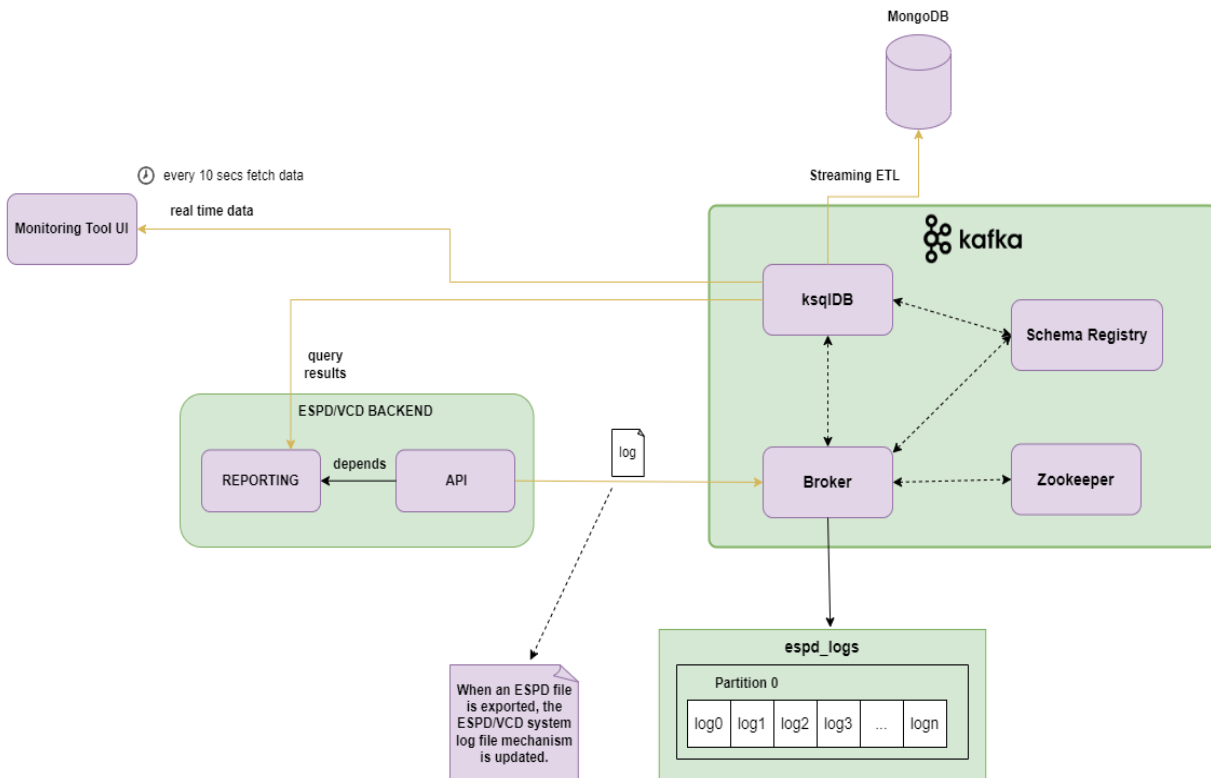
Η επικοινωνία μεταξύ των Producers και των Consumers γίνεται μέσω των Brokers. Ένα Kafka cluster μπορεί να περιλαμβάνει τουλάχιστον ένα Broker, επιβάλλονται συνήθως περισσότεροι από ένας Broker σε παραγωγικό περιβάλλον, προκειμένου σε περίπτωση που ένας από αυτούς παρουσιάσει πρόβλημα τότε ένα άλλος Broker να αναλάβει στη θέση του, χωρίς να υπάρξει απώλεια δεδομένων. Αυτό ορίζεται από το replication factor, που αποτελεί τον αριθμό των Broker που περιέχουν ένα συγκεκριμένο topic. Οι brokers επειδή δεν διατηρούν την κατάσταση του Kafka cluster, συντονίζονται από ένα server που ονομάζεται Zookeeper. Ο Zookeeper συντονίζει κάθε φορά ποιος Broker θα εξυπηρετήσει ένα Producer ή ένα Consumer. Επίσης, διατηρεί πληροφορίες σχετικά με τα Topics και τα Partitions ενός Kafka Cluster. Ένα topic είναι μια συλλογή από εγγραφές (records), που διαχωρίζονται σε ένα ή περισσότερα partitions. Κάθε partition περιέχει ένα υποσύνολο εγγραφών (records) ενός topic. Τα partitions είναι ο τρόπος με τον οποίο το Kafka επιτυγχάνει κλιμάκωση (scalability) και ταχύτητα τόσο στη λήψη νέων εγγραφών (records) από τους Producers όσο και στην αποστολή εγγραφών (records) στους Consumers, καθώς διανέμονται σε όλους τους Brokers του Kafka cluster. Τέλος, κάθε partition περιλαμβάνει μία ουρά (queue), που ονομάζεται logs. Τα logs είναι append-only, δηλαδή κάθε νέο log αποθηκεύεται στο τέλος της ουράς και δεν επιτρέπεται η διαγραφή ή η τροποποίηση του. Αξίζει να σημειωθεί ότι οι όροι logs ή εγγραφές (records) αφορούν τα δεδομένα που είναι αποθηκευμένα σε ένα Kafka Partition και είναι ισοδύναμοι. Στην εικόνα 10 παρουσιάζεται η αρχιτεκτονική ενός Kafka cluster.

Μηχανισμός έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD.



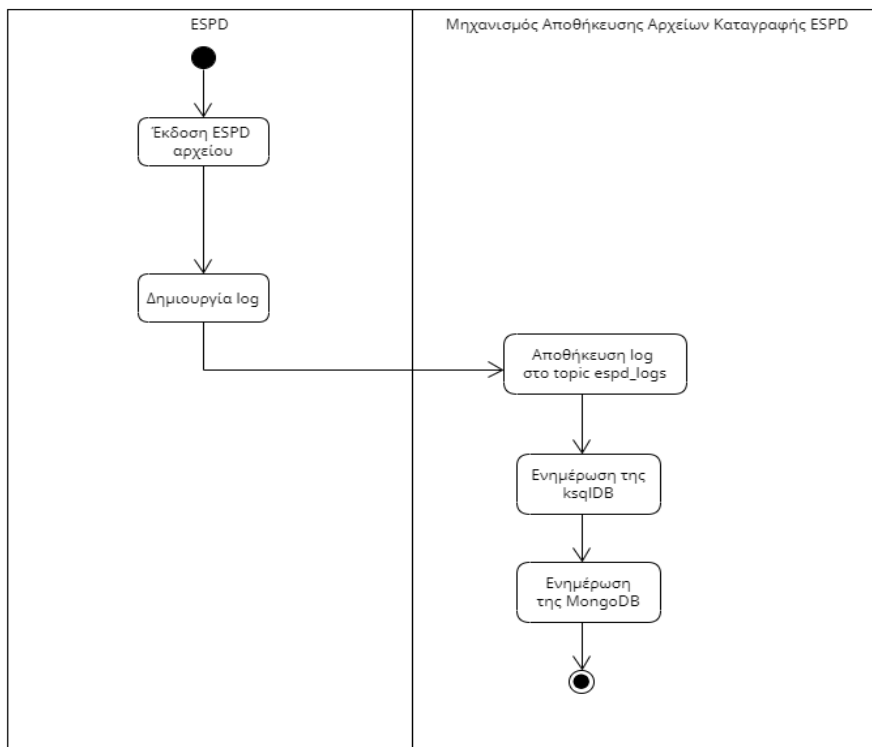
Εικόνα 10 Παράδειγμα αρχιτεκτονικής ενός Kafka Cluster. Πηγή: [27].

Η αρχιτεκτονική του μηχανισμού αρχείων καταγραφής παρουσιάζεται στην Εικόνα 11.



Εικόνα 11 Αρχιτεκτονική μηχανισμού αποθήκευσης αρχείων καταγραφής του επιχειρησιακού συστήματος ESPD.

Στην Εικόνα 11, παρατηρείται ότι χρησιμοποιείται ένας broker, ο οποίος αποθηκεύει τα δεδομένα σε ένα topic που ονομάζεται espd_logs. Το espd_logs περιλαμβάνει ένα partition, το οποίο περιλαμβάνει τα logs του ESPD συστήματος. Σε παραγωγικό περιβάλλον προφανώς επιβάλλεται να υπάρχουν περισσότεροι brokers και περισσότερα partition, τα οποία να διανέμονται στους brokers του μηχανισμού αποθήκευσης αρχείων καταγραφής, για την επίτευξη κλιμάκωσης (scalability) του μηχανισμού. Επίσης, το συμβάν που ενεργοποιεί τον μηχανισμό αποθήκευσης αρχείων καταγραφής είναι η εξαγωγή ενός ESPD αρχείου από το επιχειρησιακό σύστημα ESPD, με συνέπεια την αποστολή ενός μηνύματος στον Broker, ο οποίος με τη σειρά του θα το αποθηκεύσει στο partition του espd_logs topic. Στη συνέχεια, η βάση δεδομένων ksqiDB, η οποία βασίζεται στα δεδομένα, που βρίσκονται στο topic espd_logs θα πραγματοποιήσει ενημέρωση σε πραγματικό χρόνο της βάσης δεδομένων MongoDB, που χρησιμοποιείται για την ανάλυση δεδομένων των αρχείων καταγραφής. Η διαδικασία αυτή ονομάζεται streaming ETL [28, p. 20]. Επίσης, η ksqiDB χρησιμοποιείται για τη δημιουργία επερωτήσεων (queries) πάνω στα Kafka topics, κάτι το οποίο είναι χρήσιμο για τη δημιουργία της αναφοράς (report) του ESPD συστήματος σε πραγματικό χρόνο (real-time), επεκτείνοντας έτσι τις δυνατότητες του Kafka. Στην εικόνα 12, παρουσιάζεται το διάγραμμα δραστηριότητας του μηχανισμού αποθήκευσης αρχείων καταγραφής (log) του συστήματος ESPD. Πιο συγκεκριμένα, η έκδοση ενός ESPD εγγράφου προκαλεί τη δημιουργία ενός αρχείου καταγραφής (log), το οποίο αποθηκεύεται σε ένα topic του Kafka cluster που ονομάζεται espd_logs. Τέλος, η ksqiDB ενημερώνεται για την εισαγωγή του νέου αρχείου καταγραφής (log) και πραγματοποιεί streaming ETL ενημερώνοντας τη MongoDB που τα δεδομένα της χρησιμοποιούνται για τη δημιουργία προβλεπτικών μοντέλων από την υπηρεσία ανάλυσης δεδομένων.



Εικόνα 12 UML διάγραμμα δραστηριότητας του μηχανισμού αποθήκευσης αρχείων καταγραφής ESPD.

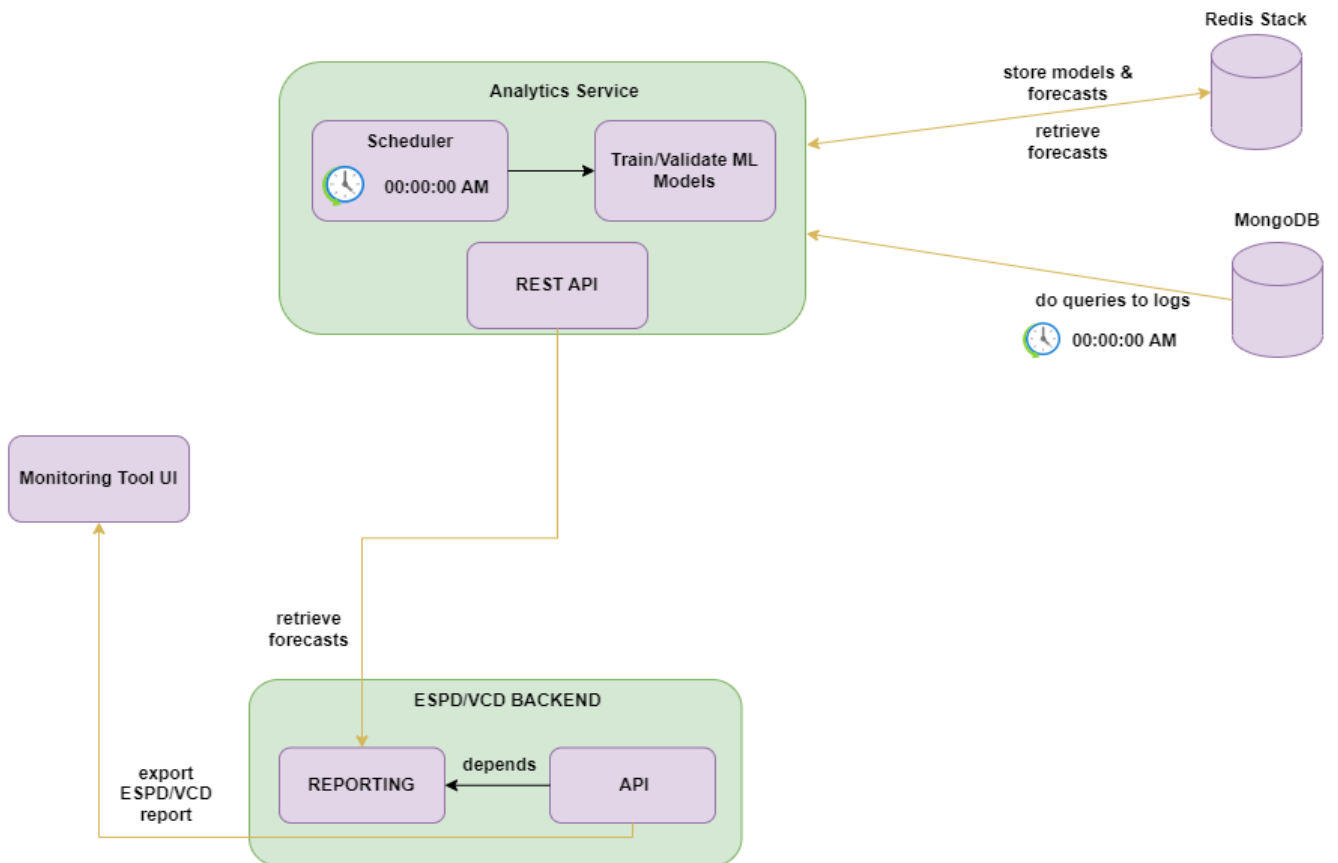
3.4 Παρουσίαση της υπηρεσίας ανάλυσης δεδομένων των αρχείων καταγραφής του συστήματος ESPD (Analytics Service)

Η υπηρεσία ανάλυσης δεδομένων, είναι υπεύθυνη για τη δημιουργία προβλέψεων σε διάστημα μίας εβδομάδας. Η μεθοδολογία που ακολουθείται για τη δημιουργία νέων προβλέψεων είναι ακόλουθη:

- Ανάκτηση των αρχείων καταγραφής από τη βάση δεδομένων MongoDB κάθε μέρα στις 12:00 π.μ.
- Εκπαίδευση των μοντέλων μηχανικής μάθησης με βάση τις καλύτερες παραμέτρους σημείου αλλαγής προηγούμενης κλίμακας (changepoint prior scale) και προγενέστερη κλίμακα εποχικότητας (seasonality prior scale) με την τεχνική του cross validation.
- Αξιολόγηση των μοντέλων μηχανικής μάθησης με βάση τη ρίζα του μέσου τετραγωνικού σφάλματος (RMSE).
- Επιλογή του μοντέλου μηχανικής μάθησης με τη μικρότερη ρίζα του μέσου τετραγωνικού σφάλματος (RMSE).
- Αποθήκευση του καλύτερου μοντέλου μηχανικής μάθησης με βάση το RMSE, σε μία βάση δεδομένων Redis καθώς και τις προβλέψεις του.

Η τεχνική του cross validation για χρονοσειρές σύμφωνα με τη βιβλιοθήκη Prophet βασίζεται σε τρεις παραμέτρους την αρχική περίοδο (initial period), περίοδος (period) και ορίζοντας πρόβλεψης (horizon). Η αρχική περίοδο (initial period) περιλαμβάνει τα αρχικά χρονικά σημεία που περιλαμβάνονται στο σύνολο δοκιμής (testing dataset), η περίοδος (period) είναι ο αριθμός των χρονικών σημείων μετά την αρχική περίοδο (initial period) και ορίζοντας πρόβλεψης (horizon) ορίζει το πλήθος των χρονικών σημείων για τα οποία θα προβλεφθούν οι τιμές τους [29]. Για παράδειγμα αν η αρχική περίοδο (initial period) είναι 200 ημέρες, η περίοδος είναι 30 ημέρες και ο ορίζοντας πρόβλεψης (horizon) 60 ημέρες, τότε το σύνολο δοκιμών (testing dataset) θα έχει συνολικά 230 ημέρες και με βάση αυτές θα προβλέψει τις τιμές της μεταβλητής ενδιαφέροντος (variable of interest) για τις επόμενες 60 ημέρες, δηλαδή μέχρι και την 290^η ημέρα. Η διαδικασία αυτή συνεχίζεται προσθέτοντας τα επόμενα χρονικά σημεία με βάση την περίοδο (period) στο (testing dataset). Στην επόμενη επανάληψη θα είναι 260 ημέρες και ο ορίζοντας πρόβλεψης (horizon) θα είναι 60 ημέρες, άρα η πρόβλεψη θα είναι μέχρι την 320^η ημέρα. Όταν δεν υπάρχουν πλέον αρκετά χρονικά σημεία για να συνεχίσει τις προβλέψεις, τότε σταματάει. Για τη δημιουργία μοντέλων πρόβλεψης στην υπηρεσία ανάλυσης δεδομένων χρησιμοποιείται αρχική περίοδο (initial period) 270 ημέρες, περίοδος 30 ημερών (period) και ορίζοντας πρόβλεψης (horizon) 90 ημέρες για την εφαρμογή της τεχνικής cross validation. Αξίζει να αναφερθεί ότι η παράμετρος αλλαγής προηγούμενης κλίμακας (changepoint prior scale) ορίζει την ευελιξία στην αυτόματη επιλογή σημείων αλλαγής της τάσης. Μεγάλη τιμή στο changepoint prior scale επιτρέπει την επιλογή πολλών σημείων αλλαγής (changepoints), ενώ μία μικρή τιμή επιτρέπει την επιλογή μικρού αριθμού σημείων αλλαγής (changepoints) [30].

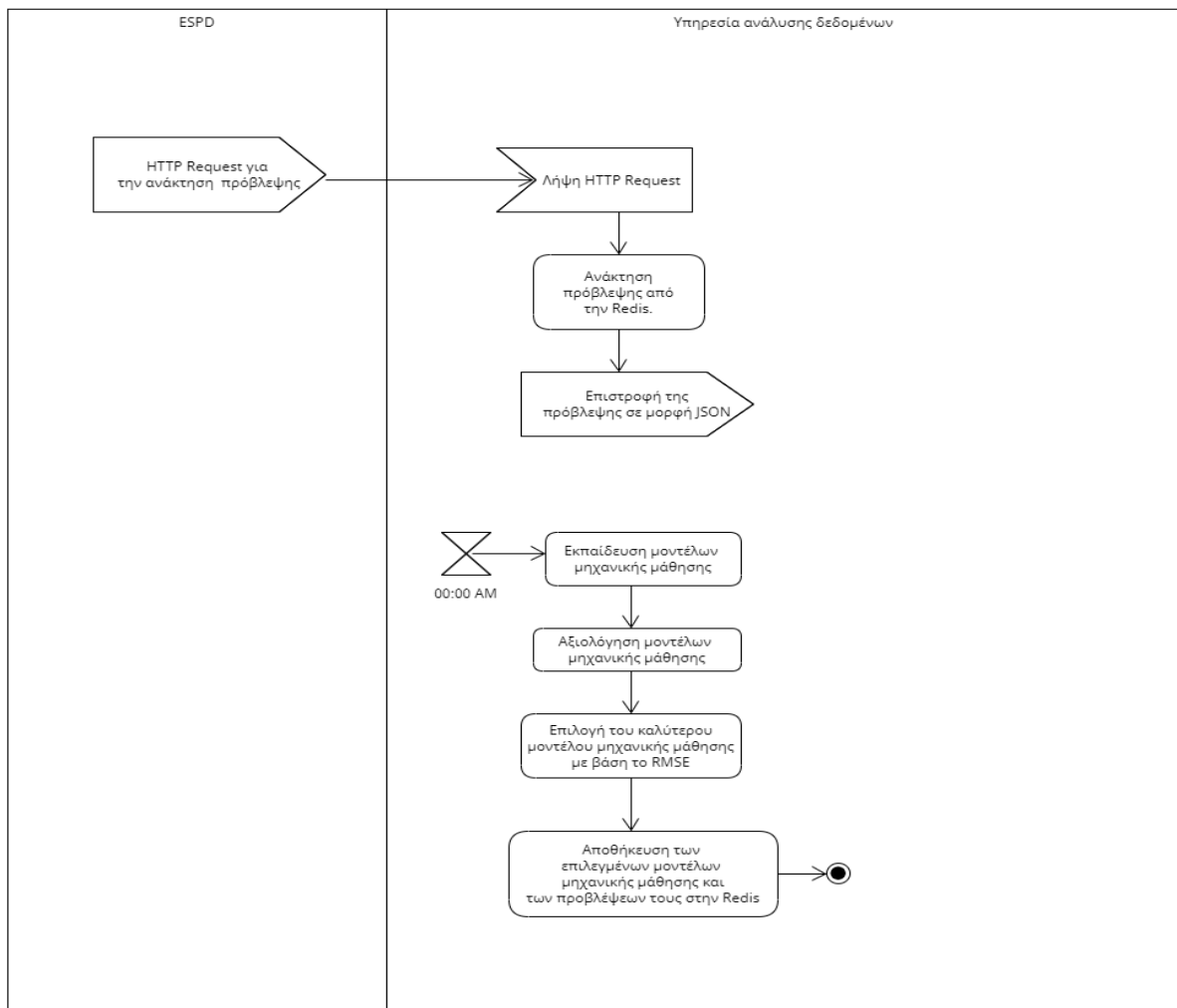
Μηχανισμός έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD.



Εικόνα 13 Η αρχιτεκτονική της υπηρεσίας ανάλυσης δεδομένων του μηχανισμού έκδοσης αναφορών και παρακολούθησης σε πραγματικό χρόνο του συστήματος ESPD.

Στην εικόνα 13, παρατηρείται ότι όταν ένας χρήστης ζητήσει την εξαγωγή αναφοράς για το σύστημα ESPD, τότε η βιβλιοθήκη reporting του ESPD ανακτά τα δεδομένα από την υπηρεσία ανάλυσης δεδομένων, προκειμένου η αναφορά να περιλαμβάνει τις προβλέψεις μέχρι την επόμενη εβδομάδα. Οι προβλέψεις περιλαμβάνουν την ακριβή τιμή πρόβλεψης (prediction), το διάστημα πρόβλεψης (prediction interval) και το επίπεδο σημαντικότητας (significance level), που είναι εξ' ορισμού 95%. Στην εικόνα 14, παρουσιάζεται το διάγραμμα δραστηριότητας για την υπηρεσία ανάλυσης δεδομένων, περιγράφοντας τη διαδικασία ανάκτησης προβλέψεων για τις εξαγωγές εγγράφων από το σύστημα ESPD, καθώς και τη διαδικασία εκπαίδευσης και αξιολόγησης για την επιλογή των καλύτερων προβλεπτικών μοντέλων.

Μηχανισμός έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD.



Εικόνα 14 UML διάγραμμα δραστηριότητας για την υπηρεσία ανάλυσης δεδομένων.

Τέλος, οι διαθέσιμες διεπαφές REST για την υπηρεσία ανάλυσης δεδομένων (AnalyticsService) παρουσιάζονται στους πίνακες 1 έως 6.

Πίνακας 1 Πρόβλεψη συνολικού αριθμού ESPD αρχείων, μέχρι την επόμενη εβδομάδα.

AnalyticsService	
Περιγραφή	Οι προβλέψεις για τον συνολικό αριθμό ESPD εγγράφων που θα εξαχθούν από το σύστημα ESPD μέχρι την επόμενη εβδομάδα.
Endpoint	/api/v1/forecast/weekly/total/espd
HTTP μέθοδος	GET
Παράμετροι	N/A

Επιστρέφει	Μία λίστα με τις προβλέψεις.
------------	------------------------------

Πίνακας 2 Πρόβλεψη συνολικού αριθμού ESPD Request αρχείων, μέχρι την επόμενη εβδομάδα.

AnalyticsService	
Περιγραφή	Οι προβλέψεις για τον συνολικό αριθμό ESPD Request εγγράφων που θα εξαχθούν από το σύστημα ESPD μέχρι την επόμενη εβδομάδα.
Endpoint	/api/v1/forecast/weekly/total/espd/requests
HTTP μέθοδος	GET
Παράμετροι	N/A
Επιστρέφει	Μία λίστα με τις προβλέψεις.

Πίνακας 3 Πρόβλεψη συνολικού αριθμού ESPD Response αρχείων, μέχρι την επόμενη εβδομάδα.

AnalyticsService	
Περιγραφή	Οι προβλέψεις για τον συνολικό αριθμό ESPD Response εγγράφων που θα εξαχθούν από το σύστημα ESPD μέχρι την επόμενη εβδομάδα.
Endpoint	/api/v1/forecast/weekly/total/espd/responses
HTTP μέθοδος	GET
Παράμετροι	N/A
Επιστρέφει	Μία λίστα με τις προβλέψεις.

Πίνακας 4 Πρόβλεψη συνολικού αριθμού ESPD Request αρχείων με βάση το τύπο αρχείου, μέχρι την επόμενη εβδομάδα.

AnalyticsService	
Περιγραφή	Οι προβλέψεις για τον συνολικό αριθμό ESPD Request εγγράφων με βάση το τύπο αρχείου, δηλαδή PDF, XML και HTML, που θα εξαχθούν από το σύστημα ESPD μέχρι την επόμενη εβδομάδα.

Endpoint	/api/v1/forecast/weekly/espdc/request/<file>
HTTP μέθοδος	GET
Παράμετροι	Ο τύπος αρχείου.
Επιστρέφει	Μία λίστα με τις προβλέψεις.

Πίνακας 5 Πρόβλεψη συνολικού αριθμού ESPD Response αρχείων με βάση το τύπο αρχείου, μέχρι την επόμενη εβδομάδα.

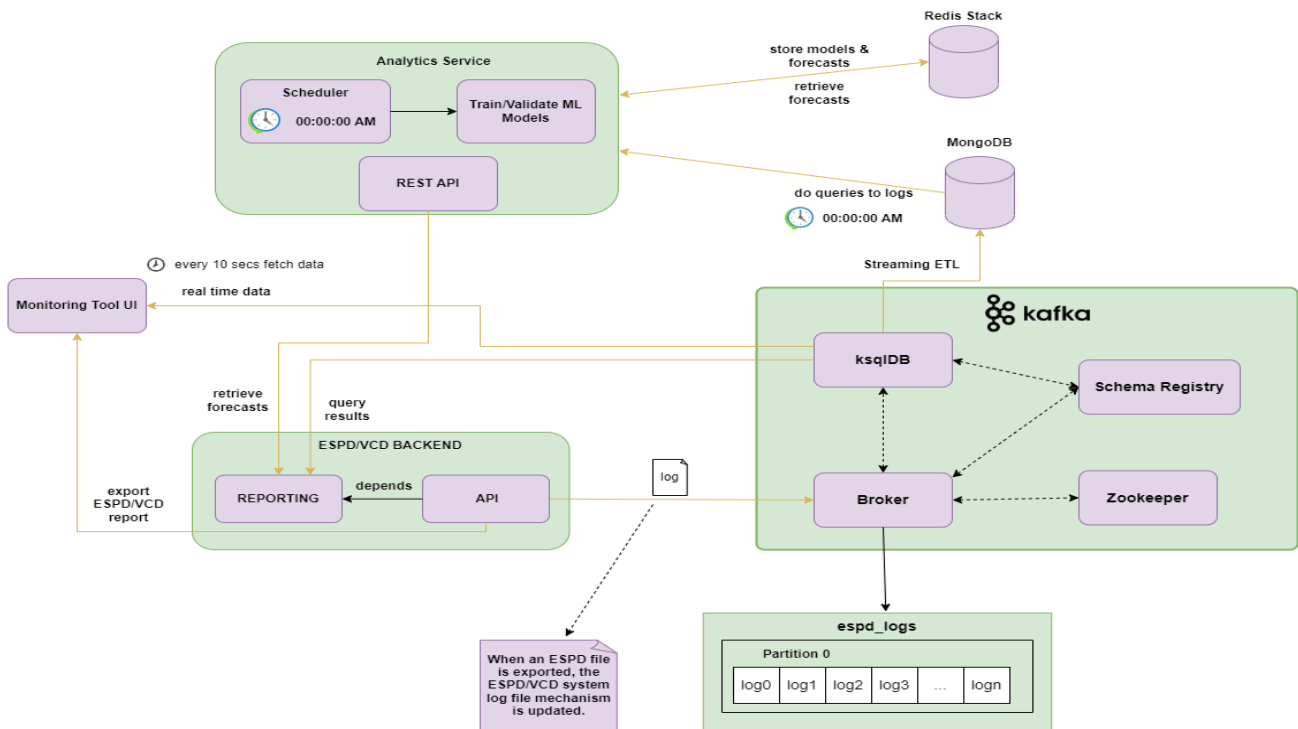
AnalyticsService	
Περιγραφή	Οι προβλέψεις για τον συνολικό αριθμό ESPD Response εγγράφων με βάση το τύπο αρχείου, δηλαδή PDF, XML και HTML, που θα εξαχθούν από το σύστημα ESPD μέχρι την επόμενη εβδομάδα.
Endpoint URL	/api/v1/forecast/weekly/espdc/response/<file>
HTTP μέθοδος	GET
Παράμετροι	Ο τύπος αρχείου.
Επιστρέφει	Μία λίστα με τις προβλέψεις.

Πίνακας 6 Πρόβλεψη συνολικού αριθμού ESPD αρχείων με βάση την έκδοση, μέχρι την επόμενη εβδομάδα.

AnalyticsService	
Περιγραφή	Οι προβλέψεις για τον συνολικό αριθμό ESPD εγγράφων με βάση την έκδοση, που θα εξαχθούν από το σύστημα ESPD μέχρι την επόμενη εβδομάδα.
Endpoint URL	/api/v1/forecast/weekly/espdc/<version>
HTTP μέθοδος	GET
Παράμετροι	Η έκδοση του αρχείου.
Επιστρέφει	Μία λίστα με τις προβλέψεις.

3.5 Η αρχιτεκτονική του μηχανισμού έκδοσης αναφορών και παρακολούθησης σε πραγματικό χρόνο του συστήματος ESPD

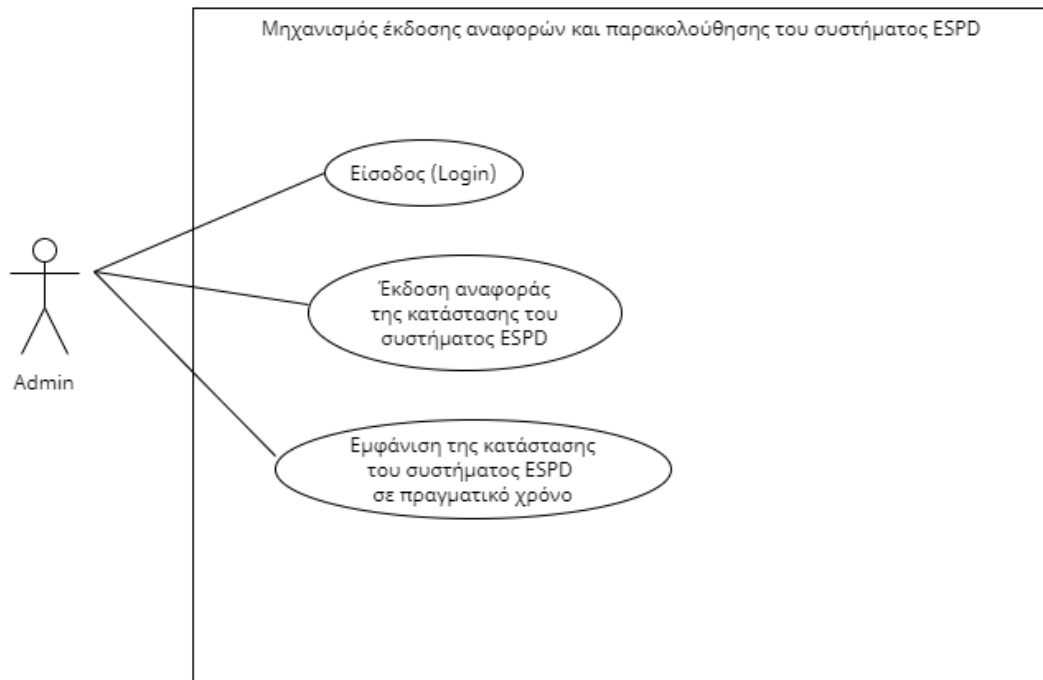
Ο μηχανισμός έκδοσης αναφορών και παρακολούθησης σε πραγματικό χρόνο του συστήματος ESPD αποτελεί ένα νέας γενιάς σύστημα επιχειρησιακής ευφυΐας, το οποίο ανήκει στην κατηγορία των επιχειρησιακών συστημάτων (Operational BI). Το συμβάν με το οποίο ενεργοποιείται ο μηχανισμός είναι η εξαγωγή ενός εγγράφου ESPD Request ή ESPD Response.



Εικόνα 15 Η αρχιτεκτονική του μηχανισμού έκδοσης αναφορών και παρακολούθησης σε πραγματικό χρόνο του συστήματος ESPD.

Στη συνέχεια, αποθηκεύονται τα μεταδεδομένα του ESPD εγγράφου ως αρχείο καταγραφής (log) στο Kafka topic espd_logs και ταυτόχρονα ενημερώνεται η ksqldb για το νέο αρχείο καταγραφής (log), πραγματοποιώντας streaming ETL στη βάση δεδομένων MongoDB. Τέλος, ενημερώνονται σε πραγματικό χρόνο τα γραφήματα στη διεπαφή χρήστη (UI), τα οποία παρουσιάζουν με περιεκτικό και όμορφο τρόπο την πληροφορία που πηγάζει από τα μεταδεδομένα των ESPD εγγράφων που εξήχθησαν από το σύστημα ESPD. Αν ένας χρήστης ζητήσει την έκδοση της αναφοράς για το σύστημα ESPD, τότε το HTTP αίτημα στέλνεται στο REST API του ESPD backend, το οποίο χρησιμοποιεί τη βιβλιοθήκη reporting, για να δημιουργήσει την αναφορά του ESPD συστήματος σε μορφή PDF. Η βιβλιοθήκη reporting ανακτά τα δεδομένα από τη βάση δεδομένων ksqldb και από το REST API της υπηρεσίας ανάλυσης δεδομένων (analytics service), προκειμένου να δημιουργήσει την αναφορά. Τέλος, η υπηρεσία ανάλυσης δεδομένων (analytics service) κάθε μέρα στις 12:00 π.μ. πραγματοποιεί την εκπαίδευση, την αξιολόγηση και την επιλογή των καλύτερων μοντέλων μηχανικής μάθησης και τα αποθηκεύει στη Redis, μαζί με τις προβλέψεις τους, οι

οποίες έχουν ορίζοντα 7 ημερών από την ημέρα που δημιουργήθηκε το μοντέλο. Στην Εικόνα 16, παρουσιάζεται το διάγραμμα περιπτώσεων χρήσης του μηχανισμού έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD. Πιο συγκεκριμένα, παρατηρείται ότι ο διαχειριστής του συστήματος μπορεί να πραγματοποιήσει είσοδο στο σύστημα με τη χρήση συνθηματικού και όνομα χρήστη, να εκδώσει αναφορά της κατάστασης του συστήματος ESPD και να παρακολουθήσει την κατάσταση του συστήματος σε πραγματικό χρόνο.



Εικόνα 16 UML διάγραμμα περιπτώσεων χρήσης για τον μηχανισμό έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD.

4. ΥΛΟΠΟΙΗΣΗ ΣΥΣΤΗΜΑΤΟΣ

4.1 Εισαγωγή

Στο κεφάλαιο 4 θα αναφερθούν οι κυριότερες τεχνολογίες που χρησιμοποιήθηκαν για την υλοποίηση του μηχανισμού έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD. Πιο συγκεκριμένα, θα παρουσιαστούν οι γλώσσες προγραμματισμού που χρησιμοποιήθηκαν για την εκπόνηση της παρούσας διπλωματικής εργασίας, δηλαδή Java, Python και TypeScript, καθώς και οι πιο σημαντικές βιβλιοθήκες τους. Τέλος, παρατίθενται οι βάσεις δεδομένων που χρησιμοποιήθηκαν για την αποθήκευση των δεδομένων του μηχανισμού, όπως η `ksqlDB`, `MongoDB` και `Redis Stack`.

4.2 Τεχνολογίες

4.2.1 Γλώσσες Προγραμματισμού

4.2.1.1 Java

Η Java [31] αποτελεί μία γλώσσα αντικειμενοστραφούς προγραμματισμού, γενικού σκοπού, η οποία χρησιμοποιείται παγκοσμίως από πολλές επιχειρήσεις για την ανάπτυξη συστημάτων. Η ανάπτυξη μίας εφαρμογής σε Java απαιτεί ένα JDK (Java Development Kit) για τη μεταγλώττιση του κώδικα και ένα JRE (Java Runtime Environment) για την εκτέλεση του κώδικα. Δεν εξαρτάται από το είδος του λειτουργικού συστήματος, και άρα είναι συμβατή σε όλα τα υπολογιστικά συστήματα. Η Java δημοσιεύτηκε στις 23 Μαΐου του 1995 από τον James Gosling της εταιρίας Sun Microsystems, η οποία εξαγοράστηκε στις 20 Απριλίου του 2009 από την Oracle [32]. Η τελευταία έκδοση της είναι η Java SE 19 που δημοσιεύθηκε στις 20 Σεπτεμβρίου του 2022. Η χρήση της Java είναι απαραίτητη για την υλοποίηση της λειτουργικότητας της βιβλιοθήκης έκδοσης αναφορών (reporting).

4.2.1.2 Python

Η γλώσσα προγραμματισμού Python [33] είναι μία γλώσσα γενικού σκοπού, η οποία υλοποιήθηκε για να λύσει το πρόβλημα της μη αναγνωσιμότητας του κώδικα, χρησιμοποιώντας εσοχές (indentation) για τον ορισμό ενός μπλοκ κώδικα. Η Python δημοσιεύτηκε στις 20 Φεβρουαρίου του 1991 από τον Guido Van Rossum και υποστηρίζει τον αντικειμενοστραφή προγραμματισμό (object-oriented), τον διαδικαστικό προγραμματισμό (procedural programming) και τον λειτουργικό προγραμματισμό (functional programming). Η μεταγλώττιση (compilation) περιλαμβάνει τη μετατροπή του προγράμματος που είναι αποθηκευμένο σε `.py` σε bytes code, δημιουργώντας ένα αρχείο με κατάληξη `.pyc`. Τέλος, ο κώδικας που βρίσκεται στο αρχείο `.pyc` εκτελείται στη Python Virtual Machine. Η χρήση της γλώσσας Python στην παρούσα εργασία ήταν ιδιαίτερα σημαντική για τη δημιουργία μοντέλων μηχανικής μάθησης καθώς και για την υλοποίηση του REST API υπηρεσίας ανάλυσης δεδομένων.

4.2.1.3 TypeScript

Η TypeScript [34] είναι μία γλώσσα προγραμματισμού, η οποία χαρακτηρίζεται για το αυστηρό συντακτικό της στυλ πάνω στη γλώσσα προγραμματισμού JavaScript. Η TypeScript δημοσιεύτηκε στις 1 Οκτωβρίου 2012 από τη Microsoft, με σκοπό τη δημιουργία διαδικτυακών εφαρμογών. Η TypeScript επεκτείνει τη γλώσσα προγραμματισμού JavaScript, προσφέροντας την απαραίτητη αξιοπιστία για την υλοποίηση ενός έργου. Τέλος, χρησιμοποιείται από το web framework Angular για την υλοποίηση της διεπαφής χρήστη (UI) του μηχανισμού έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD.

4.2.2 Βιβλιοθήκες

4.2.2.1 Angular

Το Angular [35] αποτελεί ένα web framework για τη σχεδίαση και ανάπτυξη single-page εφαρμογών (SPA) με αποδοτικό και εκλεπτυσμένο τρόπο. Δημοσιεύτηκε πρώτη φορά από τη Google στις 14 Σεπτεμβρίου 2016 και πλέον η τελευταία έκδοση είναι η 15.0.2, που δημοσιεύτηκε στις 30 Νοεμβρίου του 2022. Η επιλογή αυτής της τεχνολογίας έγινε με βάση κριτήρια όπως καλύτερη εμπειρία χρήστη, εύκολη αποσφαλμάτωση και καλύτερη απόδοση, που είναι μερικά πλεονεκτήματα από τη χρήση single-page apps.

4.2.2.2 Apache ECharts

Η βιβλιοθήκη Apache ECharts [36] προσφέρει μία πλούσια συλλογή από μοντέρνα και δυναμικά διαγράμματα για τις γλώσσες προγραμματισμού TypeScript και JavaScript. Η Apache Echarts αναπτύχθηκε αρχικά από τη Baidu και στη συνέχεια έγινε μέρος του Apache Software Foundation. Πιο συγκεκριμένα, η βιβλιοθήκη Apache ECharts προσφέρει πολλούς διαφορετικούς τύπους διαγραμμάτων, responsive διαγράμματα, δηλαδή διαγράμματα που ανταποκρίνονται σωστά σε όλα τα είδη συσκευών, όπως κινητά, tablets, laptop κλπ., διαγράμματα φιλικά προς τον χρήστη (user-friendly), προσφέροντας μεγέθυνση/σμίκρυνση σε ένα συγκεκριμένο σημείο του γραφήματος, ή αποθήκευση του διαγράμματος σε εικόνα κλπ., καθώς και τη δημιουργία 3D γραφημάτων. Στην παρούσα διπλωματική εργασία χρησιμοποιήθηκε για την υλοποίηση των διαγραμμάτων της διεπαφής χρήστη (UI) του μηχανισμού έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD.

4.2.2.3 Flask

Το Flask [37] αποτελεί ένα web framework της Python, το οποίο χρησιμοποιείται κυρίως για την ανάπτυξη REST API. Πρώτη φορά δημοσιεύτηκε στις 1 Απριλίου του 2010 και η τελευταία έκδοση του είναι 2.2.2 στις 9 Αυγούστου του 2022. Τα βασικά δομικά στοιχεία (components) του Flask είναι το Werkzeug, το οποίο είναι μία βιβλιοθήκη για την υλοποίηση WSGI εφαρμογών, δίνοντας τη δυνατότητα επικοινωνίας μίας εφαρμογής γραμμένης σε Python με ένα web server, το Jinja το οποίο είναι μία template engine για τη δημιουργία δυναμικού περιεχομένου σε διεπαφές χρήστη (UI), το MarkupSafe που είναι μία βιβλιοθήκη

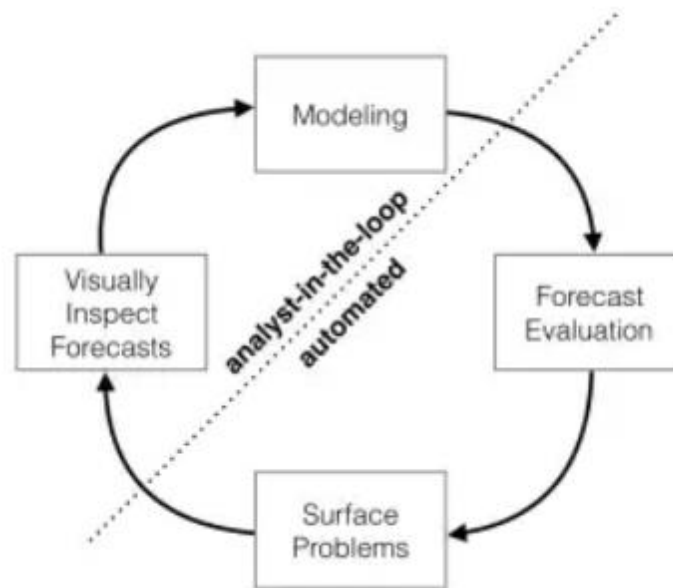
για τον χειρισμό αλφαριθμητικών (string) που περιέχουν ειδικούς χαρακτήρες, όπως ένα HTML ή XML, και για την προστασία από επιθέσεις με έγχυση (injection attacks) και το ItsDangerous που είναι μία βιβλιοθήκη για τη δημιουργία sessions και cookies, χωρίς να επιτρέπει στους χρήστες να τροποποιήσουν τα δεδομένα. Το Flask χρησιμοποιήθηκε στην παρούσα διπλωματική εργασία για τη δημιουργία του REST API της υπηρεσίας ανάλυσης δεδομένων.

4.2.2.4 Pandas

Η Pandas [38] είναι μια βιβλιοθήκη ανάλυσης δεδομένων για τη γλώσσα προγραμματισμού Python. Η πρώτη δημοσίευση έγινε στις 11 Ιανουαρίου του 2008 και η τελευταία έκδοση είναι η 1.5.2 στις 22 Νοεμβρίου του 2022. Πιο συγκεκριμένα, προσφέρει μια πλούσια συλλογή από μεθόδους οι οποίες διευκολύνουν τη διαδικασία της ανάλυσης και της επεξεργασίας των δεδομένων. Επίσης, προσθέτει διάφορες χρήσιμες και ευέλικτες δομές δεδομένων, όπως τα DataFrame, Series κλπ. που είναι απαραίτητες για τη διαδικασία της ανάλυσης δεδομένων. Στην παρούσα διπλωματική εργασία, η βιβλιοθήκη Pandas ήταν ιδιαίτερα χρήσιμη για την επεξεργασία και την αποθήκευση δεδομένων της υπηρεσίας ανάλυσης δεδομένων.

4.2.2.5 Prophet

Η Prophet [39] είναι μία βιβλιοθήκη για την υλοποίηση της διαδικασίας δημιουργίας προβλέψεων σε χρονοσειρές. Η πρώτη δημοσίευση έγινε στις 17 Φεβρουαρίου 2017 από τη Facebook και η τελευταία έκδοση είναι η 1.1.1 στις 9 Σεπτεμβρίου 2022. Πιο συγκεκριμένα, αυτοματοποιεί τη μεθοδολογία analyst-in-the-loop, η οποία αναφέρει ότι ένας αναλυτής δεδομένων (data analyst) όταν λάβει ένα σύνολο δεδομένων πραγματοποιεί τις κατάλληλες διαδικασίες επεξεργασίας δεδομένων, ξεκινάει τη διαδικασία υλοποίησης του μοντέλου πρόβλεψης δοκιμάζοντας διάφορες παραμέτρους, αξιολογεί τις προβλέψεις του μοντέλου και αντιμετωπίζει τα προβλήματα που μπορούν να προκύψουν. Η διαδικασία αυτή συνεχίζεται μέχρι να μην υπάρχει πρόβλημα. Στην εικόνα 17, παρουσιάζεται η διαδικασία analyst-in-the-loop που ακολουθεί η βιβλιοθήκη Prophet.

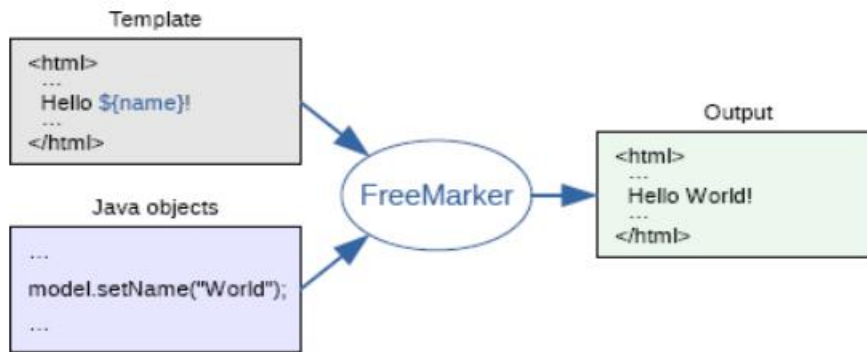


Εικόνα 17 Η διαδικασία analyst-in-the-loop. Πήγη: [33].

Στην παρούσα διπλωματική εργασία η βιβλιοθήκη Prophet χρησιμοποιήθηκε για την εκπαίδευση και αξιολόγηση των μοντέλων μηχανικής μάθησης που υλοποιήθηκαν για την υπηρεσία ανάλυσης δεδομένων.

4.2.2.6 Apache FreeMarker

Το Apache FreeMarker [40] αποτελεί μία template engine, για την εμφάνιση δυναμικών δεδομένων σε μορφή HTML, με βάση ένα μοντέλο δεδομένων Java. Πιο συγκεκριμένα, το Apache FreeMarker είναι μία βιβλιοθήκη Java, η οποία δημιουργήθηκε το 2000 και εστιάζει στην αρχιτεκτονική MVC. Η τελευταία σταθερή έκδοση του είναι η 2.3.31 στις 16 Φεβρουαρίου του 2021. Η λειτουργία του Apache FreeMarker είναι απλή, καθώς χρησιμοποιείται ένα μοντέλο δεδομένων Java και το template που χαρακτηρίζει τη δομή του HTML αρχείου. Τέλος το FreeMarker, μέσω κατάλληλων διαδικασιών, παράγει το HTML. Στην παρούσα διπλωματική εργασία ήταν ιδιαίτερα χρήσιμη για τη δημιουργία των αναφορών του συστήματος ESPD σε μορφή PDF. Τέλος, η λειτουργία του Apache FreeMarker παρουσιάζεται στην Εικόνα 18.



Εικόνα 18 Λειτουργία του Apache FreeMarker. Πηγή: [40].

4.2.3 Βάσεις Δεδομένων

4.2.3.1 ksqlDB

Η ksqlDB [41] είναι μια σχεσιακή βάση δεδομένων σε πραγματικό χρόνο (real-time), η οποία βασίζεται πάνω στο Kafka. Πρώτη φορά δημοσιεύτηκε το 2017 και η τελευταία έκδοση είναι η 0.28.2 στις 2 Αυγούστου 2022. Το συντακτικό της ksqlDB είναι βασισμένο στην SQL, και δίνει τη δυνατότητα δημιουργίας πινάκων (table), όψεων (views) και ροών (streams). Κάθε πίνακας (table) ή όψη (view) ή ροή (stream) αποθηκεύεται ως topic στο Kafka. Κάθε νέα προσθήκη στο topic, με το οποίο έχει δημιουργηθεί ένας πίνακας (table) ή μία όψη (view) ή μία ροή (stream) ενημερώνει και το topic τους στο Kafka. Οι επερωτήσεις στην ksqlDB διακρίνονται σε τρεις (3) κατηγορίες τα persistent queries, τα push queries και τα pull queries. Τα persistent queries αφορούν τη δημιουργία πινάκων (tables) ή ροών (streams) και τρέχουν επ' αόριστων, προκειμένου να λαμβάνουν κάθε αλλαγή που πραγματοποιείται σε ένα Kafka topic. Τα push queries γίνονται για τον εντοπισμό κάθε νέας αλλαγής που γίνεται στο topic, πάνω σε μία long-lived σύνδεση κάτι το οποίο είναι ιδανικό για εφαρμογές πραγματικού χρόνου (real-time). Τα pull queries γίνονται σε όψεις (views) ανοίγοντας και κλείνοντας μία σύνδεση με τη βάση δεδομένων, όπως ακριβώς σε μία παραδοσιακή σχεσιακή βάση δεδομένων. Στην ksqlDB υποστηρίζονται connectors δύο κατηγοριών είτε source connectors, οι οποίοι χρησιμοποιούνται για τη φόρτωση των δεδομένων από μία βάση δεδομένων στην ksqlDB και τους sink connectors που χρησιμοποιούνται για τη φόρτωση δεδομένων από την ksqlDB σε μία βάση δεδομένων. Η διαδικασία αυτή ονομάζεται streaming ETL καθώς συμβαίνει σε πραγματικό χρόνο (real-time), με την πραγματοποίηση κάποιου συμβάντος (event). Τέλος, δίνει τη δυνατότητα συνδέσεων (joins) δύο ή περισσότερων topics, όπως ακριβώς η σύνδεση δύο ή περισσότερων πινάκων στην SQL. Στην παρούσα διπλωματική εργασία, η βάση δεδομένων ksqlDB ήταν ιδιαίτερα χρήσιμη για την υλοποίηση συγκεντρωτικών επερωτήσεων (aggregated queries) σε πραγματικό χρόνο (real-time) πάνω στο Kafka topic espd_logs.

4.2.3.2 MongoDB

Η MongoDB [42] είναι μία document-oriented NoSQL βάση δεδομένων, η οποία δεν έχει αυστηρή δομή, καθώς διαχειρίζεται έγγραφα (documents) με μορφή που ομοιάζει με JSON και είναι καταναμημένα σε συλλογές (collections). Η MongoDB δημοσιεύτηκε πρώτη φορά στις 11 Φεβρουαρίου του 2009 και η τελευταία έκδοση της είναι η 6.0.1 στις 19 Αυγούστου του 2022. Οι δυνατότητες που παρέχει είναι η ευελιξία στη δομή που έχει κάθε έγγραφο, κάτι το οποίο είναι ιδανικό για την εξοικονόμηση χώρου στο δίσκο, διότι πολλές σύγχρονες εφαρμογές δεν απαιτούν την αποθήκευση όλων των δεδομένων, π.χ σε μία εφαρμογή αξιολόγησης ενός προϊόντος ο χρήστης μπορεί να αξιολογήσει το προϊόν αλλά να μην συμπληρώσει τη φόρμα σχολίων, άρα δεν είναι απαραίτητο να αποθηκευτεί ένα κενό σχόλιο, σε αντίθεση με μία σχεσιακή βάση δεδομένων που έχει αυστηρή δομή και θα αποθηκευόταν μία κενή τιμή. Επιπλέον, δίνει τη δυνατότητα κλιμάκωσης (scalability), είναι ιδιαίτερα ταχύτατη στην ανάγνωση (read) με τη χρήση ευρετηρίων (indexes) και προσφέρει τη δυνατότητα εφαρμογής πολύπλοκων queries με τη χρήση του aggregation pipeline επιστρέφοντας τα αποτελέσματα τους σε πολύ μικρό χρόνο. Στην παρούσα διπλωματική εργασία αποθηκεύονται τα δεδομένα που βρίσκονται στο Kafka topic espd_logs στην MongoDB με streaming ETL μέσω της ksqlDB, για την πραγματοποίηση της διαδικασίας ανάλυσης δεδομένων, χωρίς να απασχολείται η βάση δεδομένων ksqlDB.

4.2.3.3 Redis

Η Redis [43] είναι μία in-memory δομή δεδομένων, η οποία μπορεί να χρησιμοποιηθεί είτε ως βάση δεδομένων κλειδιού-τιμής (key-value), είτε ως cache είτε ως message broker για την αποστολή μηνυμάτων σε πραγματικό χρόνο (real-time). Η Redis δημοσιεύτηκε πρώτη φορά στις 10 Μαΐου του 2009 και πλέον βρίσκεται στην έκδοση 7.0.7 που δημοσιεύτηκε στις 16 Δεκεμβρίου του 2022. Μεγάλο πλεονέκτημα της Redis είναι η ταχύτητα που προσδίδει στην ανάκτηση των δεδομένων, διότι τα δεδομένα αποθηκεύονται στη μνήμη και όχι στο σκληρό δίσκο. Επίσης, είναι πολύ εύκολο να υλοποιηθεί μία βάση δεδομένων Redis καθώς απαιτείται η δημιουργία ενός ζευγαριού κλειδιού-τιμής, όπου η τιμή μπορεί να είναι μία δομή δεδομένων όπως αλφαριθμητικό (string), λίστα (list), σύνολο (set) κλπ. Η βάση δεδομένων χρησιμοποιήθηκε για την αποθήκευση των μοντέλων μηχανικής μάθησης που δημιουργήθηκαν από την υπηρεσία ανάλυσης δεδομένων καθώς και των προβλέψεων τους. Με αυτό το τρόπο επιτυγχάνεται άμεση απόκριση του REST API της υπηρεσίας ανάλυσης δεδομένων, σε μόλις κάποια χιλιοστά του δευτερολέπτου (milliseconds).

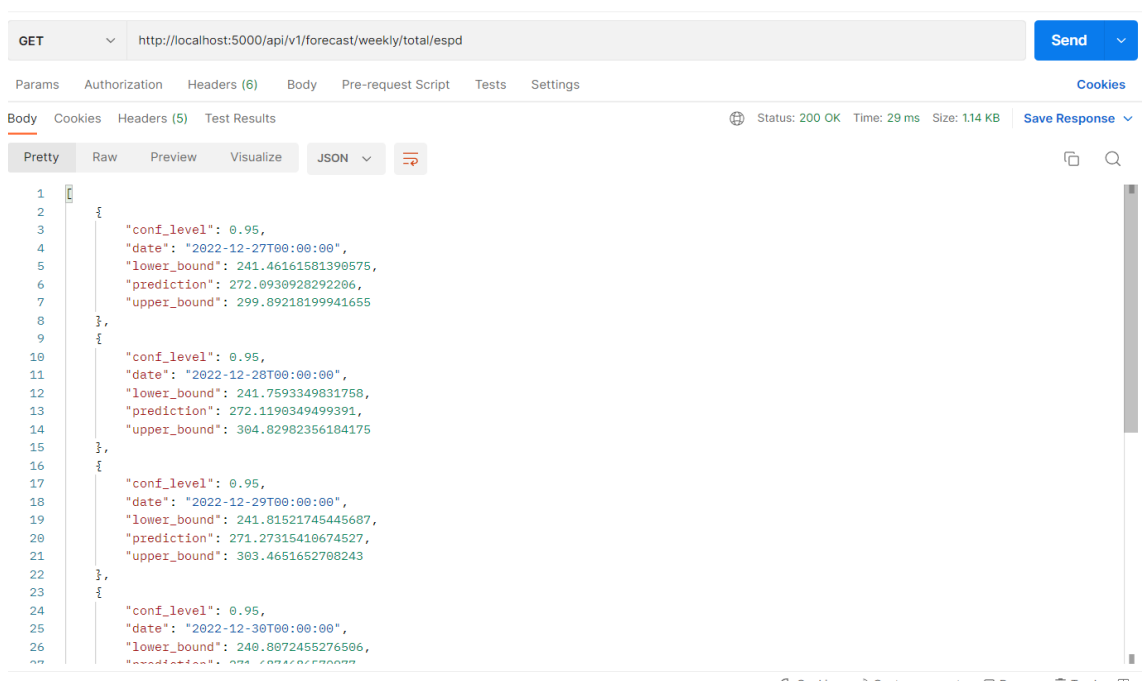
5. ΠΑΡΟΥΣΙΑΣΗ ΤΟΥ ΣΥΣΤΗΜΑΤΟΣ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΩΝ

5.1 Εισαγωγή

Στο κεφάλαιο 5 παρουσιάζεται η λειτουργικότητα της υπηρεσίας ανάλυσης δεδομένων και συνολικά του μηχανισμού έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD σε πραγματικό χρόνο. Πιο συγκεκριμένα, θα παρατεθούν οι διαθέσιμες διεπαφές της υπηρεσίας ανάλυσης δεδομένων, μέσω των οποίων θα ανακτηθούν οι προβλέψεις, καθώς και η διεπαφή χρήστη για την εμφάνιση των διαγραμμάτων χρονοσειρών, ώστε με οπτικό τρόπο να γίνεται αντιληπτό κατά πόσο τα μοντέλα μηχανικής μάθησης προβλέπουν ορθά τις μελλοντικές εξαγωγές ESPD εγγράφων, χωρίς να επηρεάζονται από υπερπροσαρμογή (overfitting) ή υποπροσαρμογή (underfitting). Τέλος, θα γίνει παρουσίαση της συνολικής λειτουργικότητας του μηχανισμού έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD, με την εμφάνιση της κατάστασης των εξαγωγών ESPD εγγράφων από το σύστημα ESPD σε πραγματικό χρόνο μέσω των διαγραμμάτων από τη διεπαφή χρήστη και την εξαγωγή αναφοράς από το σύστημα.

5.2 Παρουσίαση της υπηρεσίας ανάλυσης δεδομένων (Analytics Service) και των διαθέσιμων REST διεπαφών

Η υπηρεσία ανάλυσης δεδομένων (AnalyticsService) παρέχει διεπαφές REST για την ανάκτηση προβλέψεων. Στην εικόνα 19 παρουσιάζονται τα αποτελέσματα από την ανάκτηση των προβλέψεων του συνολικού αριθμού ESPD αρχείων, μέχρι την επόμενη εβδομάδα.



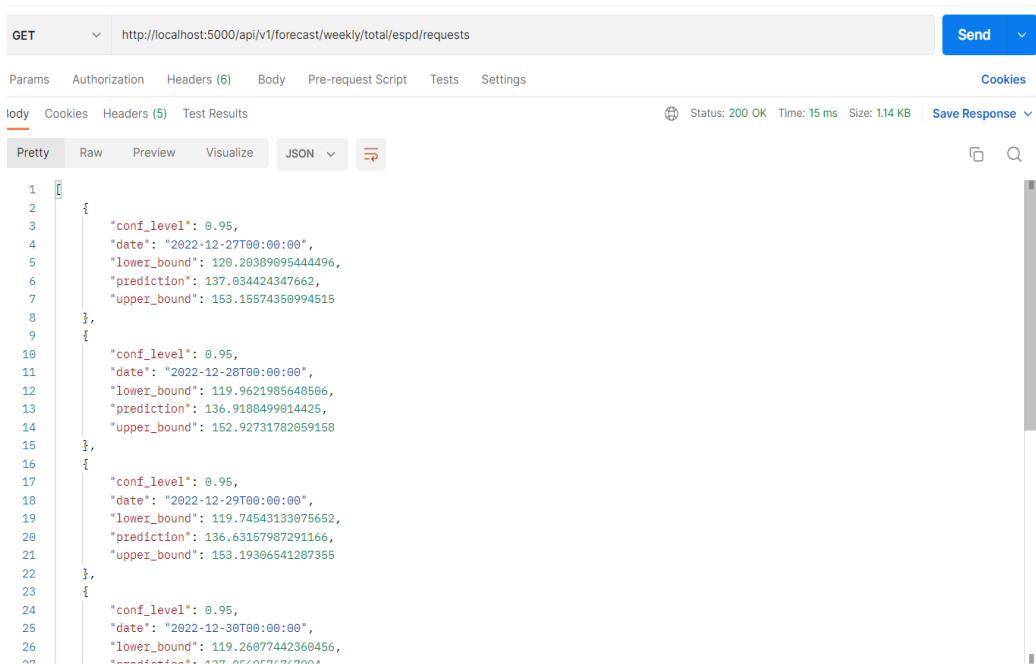
```
GET http://localhost:5000/api/v1/forecast/weekly/total/espdc
Status: 200 OK Time: 29 ms Size: 1.14 KB

{
  "conf_level": 0.95,
  "date": "2022-12-27T00:00:00",
  "lower_bound": 241.46161581390575,
  "prediction": 272.0930928292206,
  "upper_bound": 299.09218199941655
},
{
  "conf_level": 0.95,
  "date": "2022-12-28T00:00:00",
  "lower_bound": 241.7593349831758,
  "prediction": 272.1190349499391,
  "upper_bound": 304.02982356184175
},
{
  "conf_level": 0.95,
  "date": "2022-12-29T00:00:00",
  "lower_bound": 241.81521745445687,
  "prediction": 271.27315410674527,
  "upper_bound": 303.4651652708243
},
{
  "conf_level": 0.95,
  "date": "2022-12-30T00:00:00",
  "lower_bound": 240.8072455276506,
  "prediction": 271.2001065900000,
  "upper_bound": 303.4651652708243
}
```

Εικόνα 19 Αποτελέσματα πρόβλεψης του συνολικού αριθμού εξαχθέντων ESPD εγγράφων.

Μηχανισμός έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD.

Στην εικόνα 20 παρουσιάζονται τα αποτελέσματα από την ανάκτηση των προβλέψεων του συνολικού αριθμού ESPD Request αρχείων, μέχρι την επόμενη εβδομάδα.

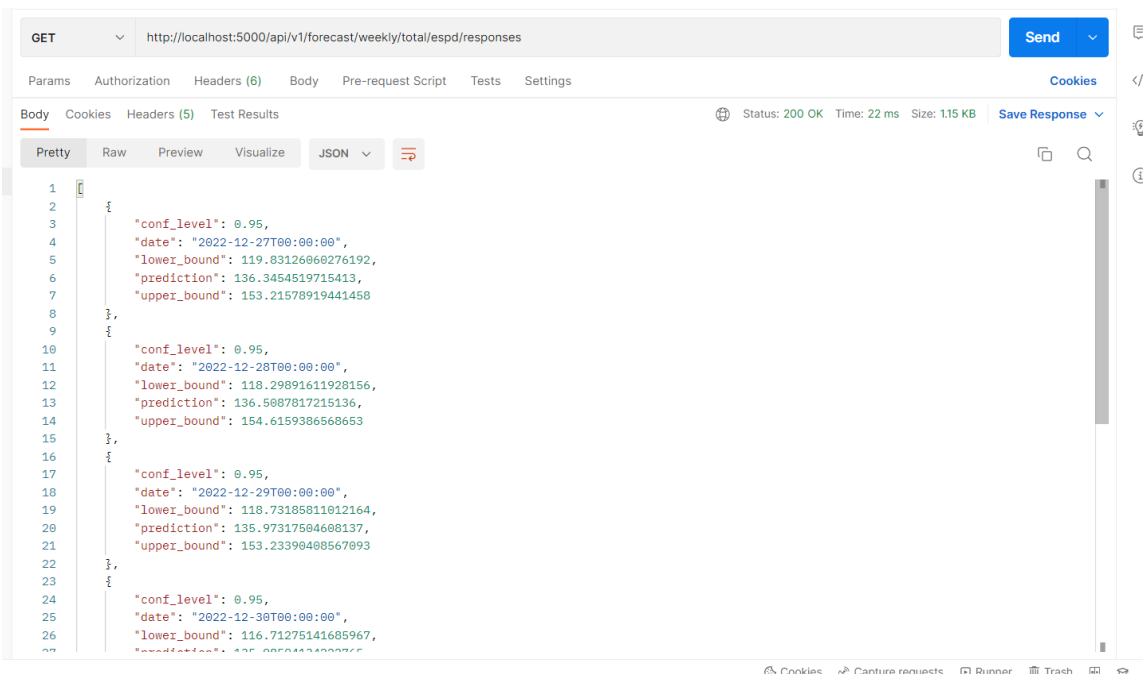


```
GET http://localhost:5000/api/v1/forecast/weekly/total/espd/requests
Status: 200 OK Time: 15 ms Size: 114 KB

[{"date": "2022-12-27T00:00:00", "prediction": 137.034424347662, "lower_bound": 120.20389095444496, "upper_bound": 153.15574350994515}, {"date": "2022-12-28T00:00:00", "prediction": 136.9188499014425, "lower_bound": 119.9621985648506, "upper_bound": 152.92731782059158}, {"date": "2022-12-29T00:00:00", "prediction": 136.63157987291166, "lower_bound": 119.74543133075652, "upper_bound": 153.19386541287355}, {"date": "2022-12-30T00:00:00", "prediction": 119.26877442360456, "lower_bound": 119.26877442360456, "upper_bound": 119.26877442360456}]]
```

Εικόνα 20 Αποτελέσματα πρόβλεψης του συνολικού αριθμού εξαχθέντων ESPD Request εγγράφων.

Στην εικόνα 21 παρουσιάζονται τα αποτελέσματα από την ανάκτηση των προβλέψεων του συνολικού αριθμού ESPD Response αρχείων, μέχρι την επόμενη εβδομάδα.

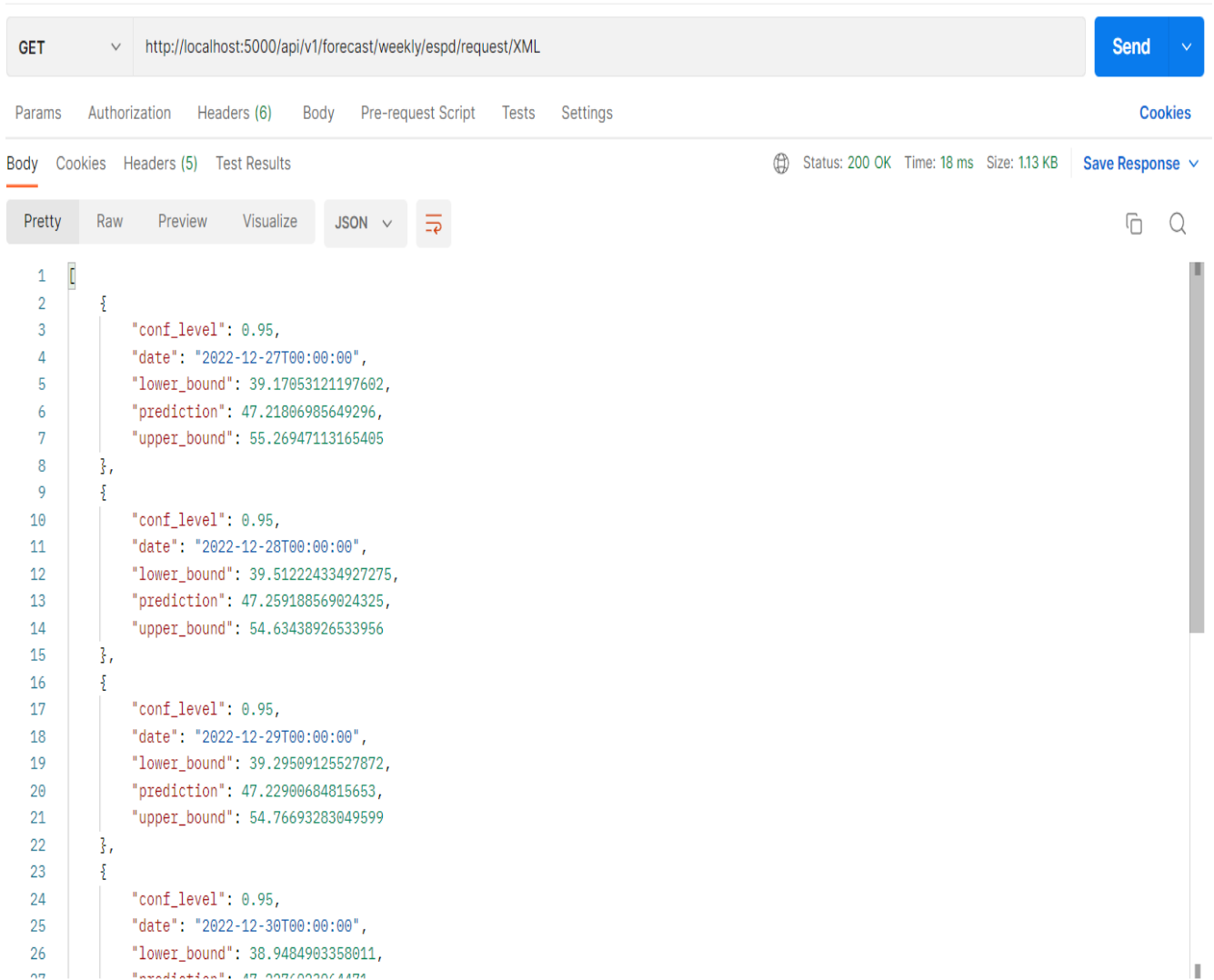


```
GET http://localhost:5000/api/v1/forecast/weekly/total/espd/responses
Status: 200 OK Time: 22 ms Size: 115 KB

[{"date": "2022-12-27T00:00:00", "prediction": 136.3454519715413, "lower_bound": 119.83126060276192, "upper_bound": 153.21578919441458}, {"date": "2022-12-28T00:00:00", "prediction": 136.5087817215136, "lower_bound": 118.29891611928156, "upper_bound": 154.6159386568653}, {"date": "2022-12-29T00:00:00", "prediction": 135.97317504608137, "lower_bound": 118.73185811012164, "upper_bound": 153.23399408567093}, {"date": "2022-12-30T00:00:00", "prediction": 116.71275141685967, "lower_bound": 116.71275141685967, "upper_bound": 116.71275141685967}]]
```

Εικόνα 21 Αποτελέσματα πρόβλεψης του συνολικού αριθμού εξαχθέντων ESPD Response εγγράφων.

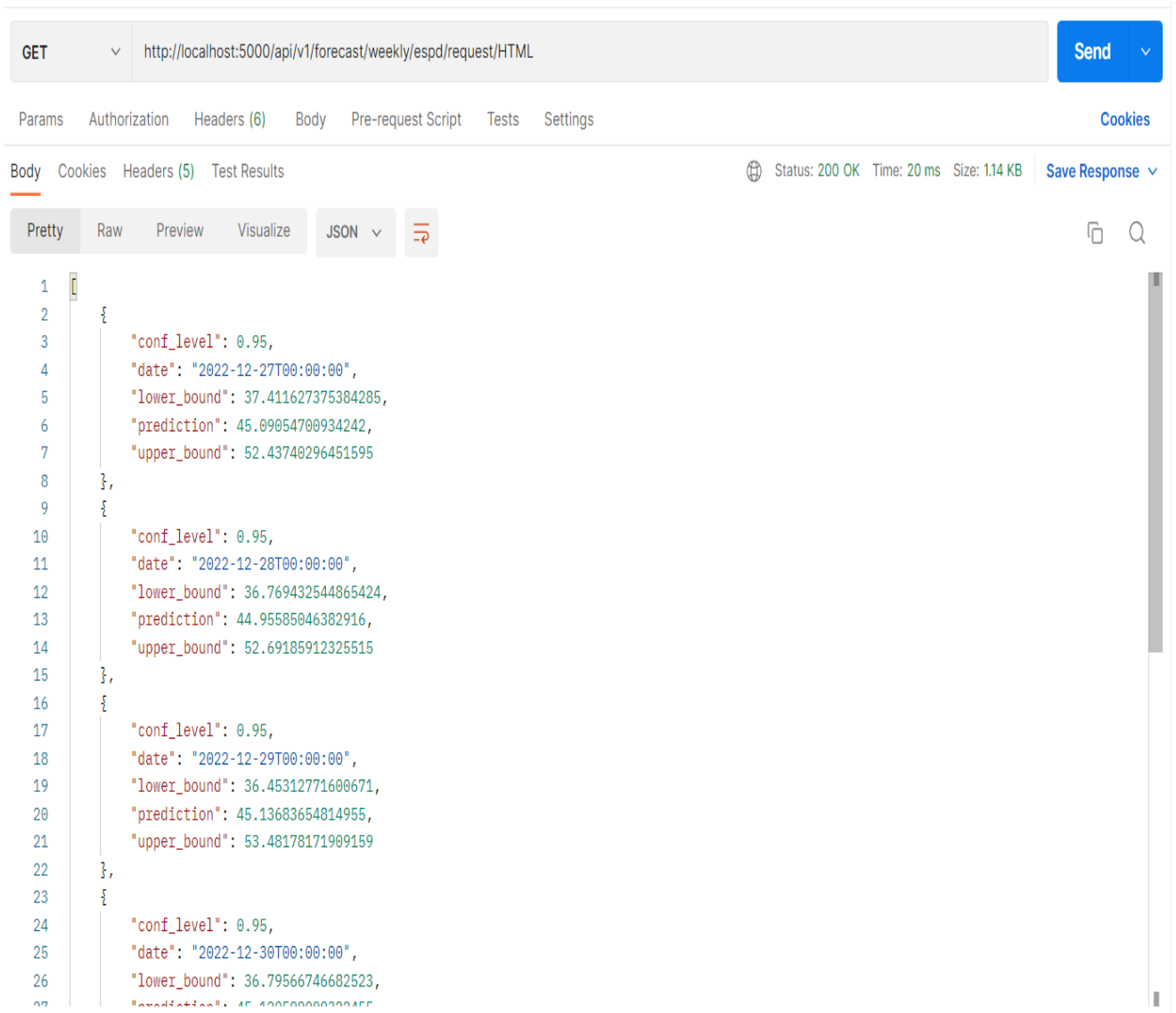
Στην εικόνα 22 παρουσιάζονται τα αποτελέσματα από την ανάκτηση των προβλέψεων του συνολικού αριθμού ESPD Request αρχείων με βάση τον τύπο αρχείου XML, μέχρι την επόμενη εβδομάδα.



```
1  {
2  }
3  "conf_level": 0.95,
4  "date": "2022-12-27T00:00:00",
5  "lower_bound": 39.17053121197602,
6  "prediction": 47.21806985649296,
7  "upper_bound": 55.26947113165405
8  },
9  {
10 "conf_level": 0.95,
11 "date": "2022-12-28T00:00:00",
12 "lower_bound": 39.512224334927275,
13 "prediction": 47.259188569024325,
14 "upper_bound": 54.63438926533956
15 },
16 {
17 "conf_level": 0.95,
18 "date": "2022-12-29T00:00:00",
19 "lower_bound": 39.29509125527872,
20 "prediction": 47.22900684815653,
21 "upper_bound": 54.76693283049599
22 },
23 {
24 "conf_level": 0.95,
25 "date": "2022-12-30T00:00:00",
26 "lower_bound": 38.9484903358011,
27 "prediction": 47.007600064471
```

Εικόνα 22 Αποτελέσματα πρόβλεψης του συνολικού αριθμού εξαχθέντων ESPD Request XML εγγράφων.

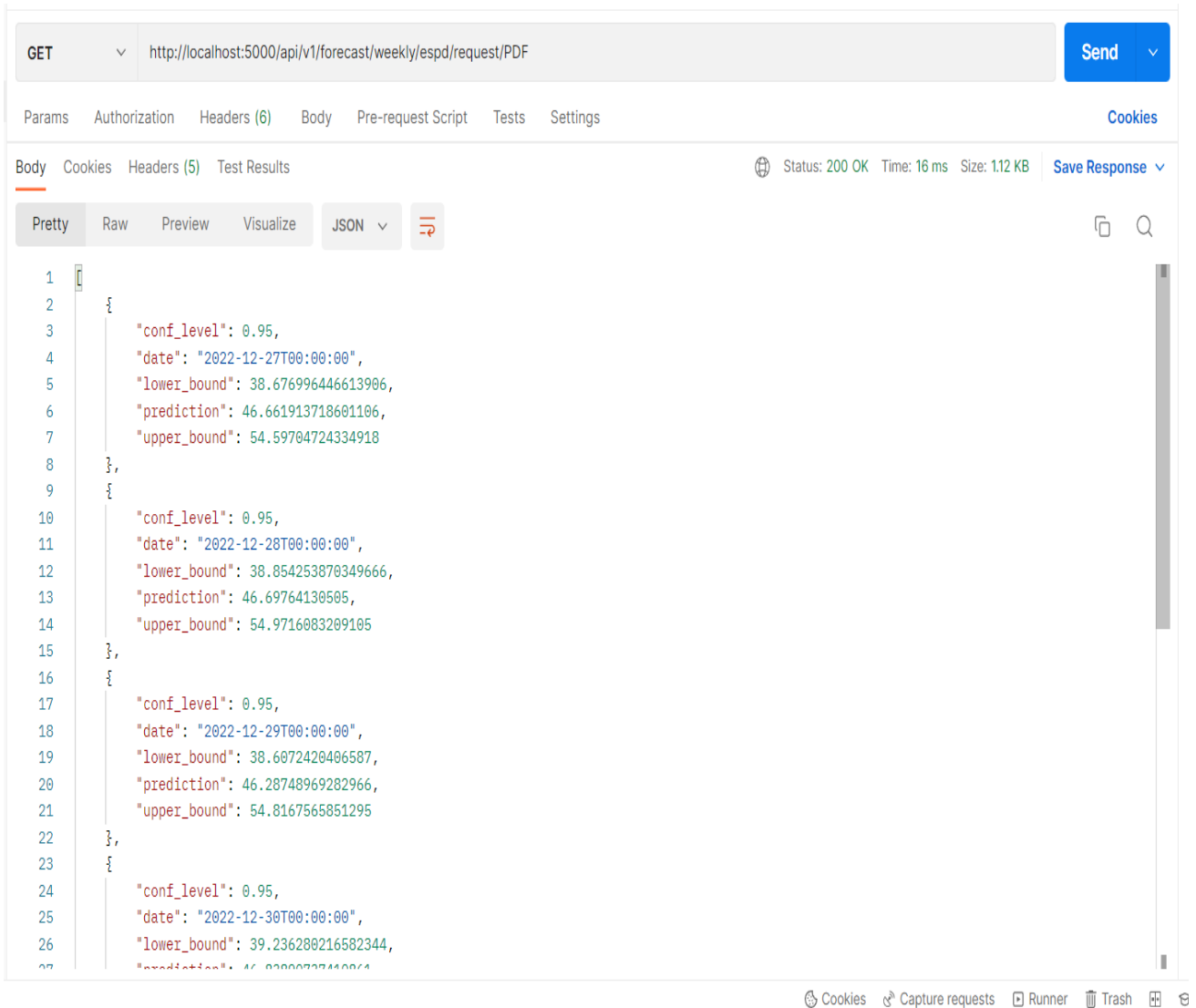
Στην εικόνα 23 παρουσιάζονται τα αποτελέσματα από την ανάκτηση των προβλέψεων του συνολικού αριθμού ESPD Request αρχείων με βάση τον τύπο αρχείου HTML, μέχρι την επόμενη εβδομάδα.



```
1 [
2   {
3     "conf_level": 0.95,
4     "date": "2022-12-27T00:00:00",
5     "lower_bound": 37.411627375384285,
6     "prediction": 45.09054700934242,
7     "upper_bound": 52.43740296451595
8   },
9   {
10    "conf_level": 0.95,
11    "date": "2022-12-28T00:00:00",
12    "lower_bound": 36.769432544865424,
13    "prediction": 44.95585046382916,
14    "upper_bound": 52.69185912325515
15  },
16  {
17    "conf_level": 0.95,
18    "date": "2022-12-29T00:00:00",
19    "lower_bound": 36.45312771600671,
20    "prediction": 45.13683654814955,
21    "upper_bound": 53.48178171909159
22  },
23  {
24    "conf_level": 0.95,
25    "date": "2022-12-30T00:00:00",
26    "lower_bound": 36.79566746682523,
27    "prediction": 45.42050000000000
```

Εικόνα 23 Αποτελέσματα πρόβλεψης του συνολικού αριθμού εξαχθέντων ESPD Request HTML εγγράφων.

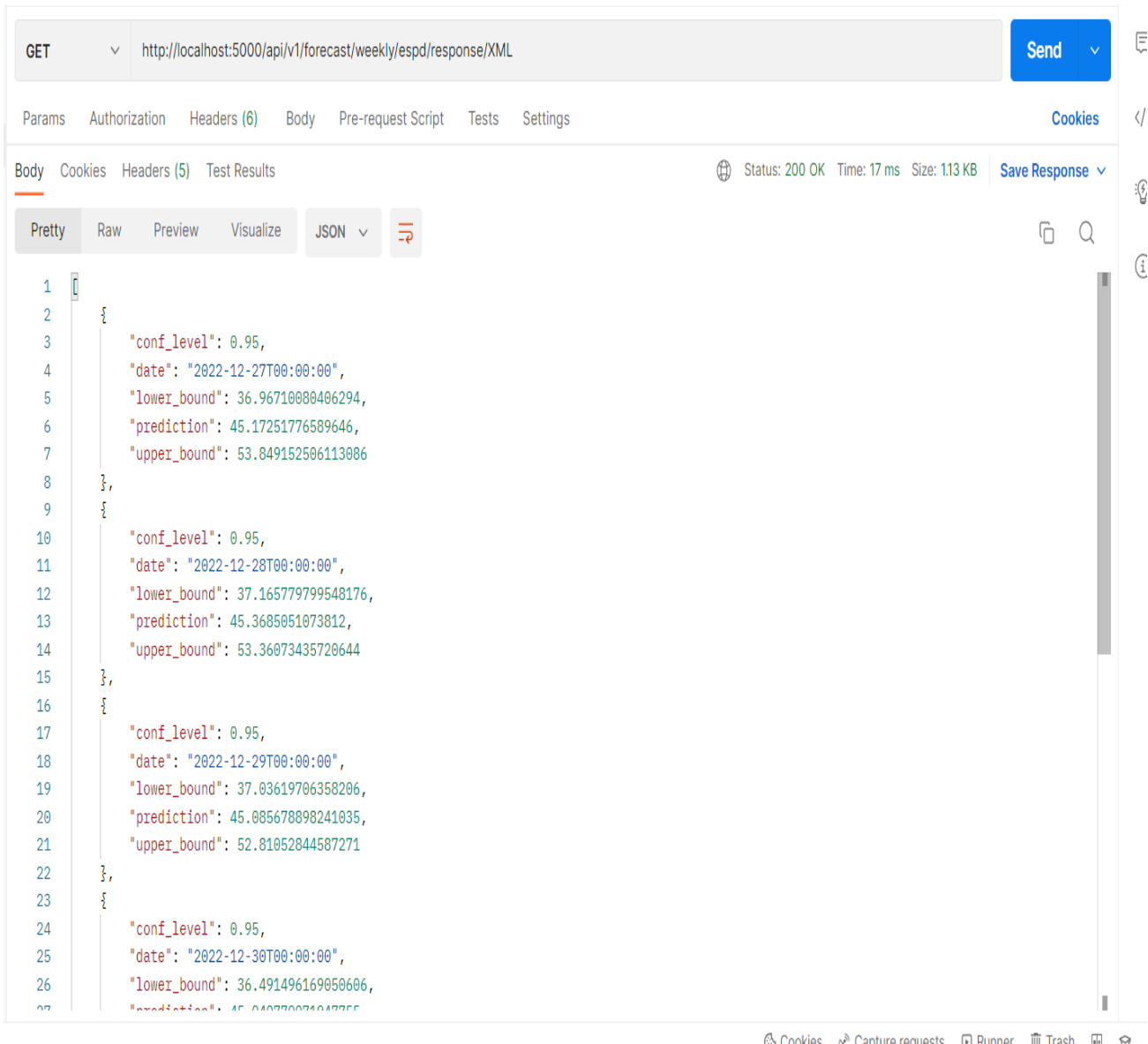
Στην εικόνα 24 παρουσιάζονται τα αποτελέσματα από την ανάκτηση των προβλέψεων του συνολικού αριθμού ESPD Request αρχείων με βάση τον τύπο αρχείου PDF, μέχρι την επόμενη εβδομάδα.



```
GET http://localhost:5000/api/v1/forecast/weekly/espdlrequest/PDF
Status: 200 OK Time: 16 ms Size: 1.12 KB
Body: JSON
[
  {
    "conf_level": 0.95,
    "date": "2022-12-27T00:00:00",
    "lower_bound": 38.676996446613906,
    "prediction": 46.661913718601106,
    "upper_bound": 54.59704724334918
  },
  {
    "conf_level": 0.95,
    "date": "2022-12-28T00:00:00",
    "lower_bound": 38.854253870349666,
    "prediction": 46.69764130505,
    "upper_bound": 54.9716083209105
  },
  {
    "conf_level": 0.95,
    "date": "2022-12-29T00:00:00",
    "lower_bound": 38.6072420406587,
    "prediction": 46.28748969282966,
    "upper_bound": 54.8167565851295
  },
  {
    "conf_level": 0.95,
    "date": "2022-12-30T00:00:00",
    "lower_bound": 39.236280216582344,
    "prediction": 46.8200777440074
  }
]
```

Εικόνα 24 Αποτελέσματα πρόβλεψης του συνολικού αριθμού εξαθθέντων ESPD Request PDF εγγράφων.

Στην εικόνα 25 παρουσιάζονται τα αποτελέσματα από την ανάκτηση των προβλέψεων του συνολικού αριθμού ESPD Response αρχείων με βάση τον τύπο αρχείου XML, μέχρι την επόμενη εβδομάδα.

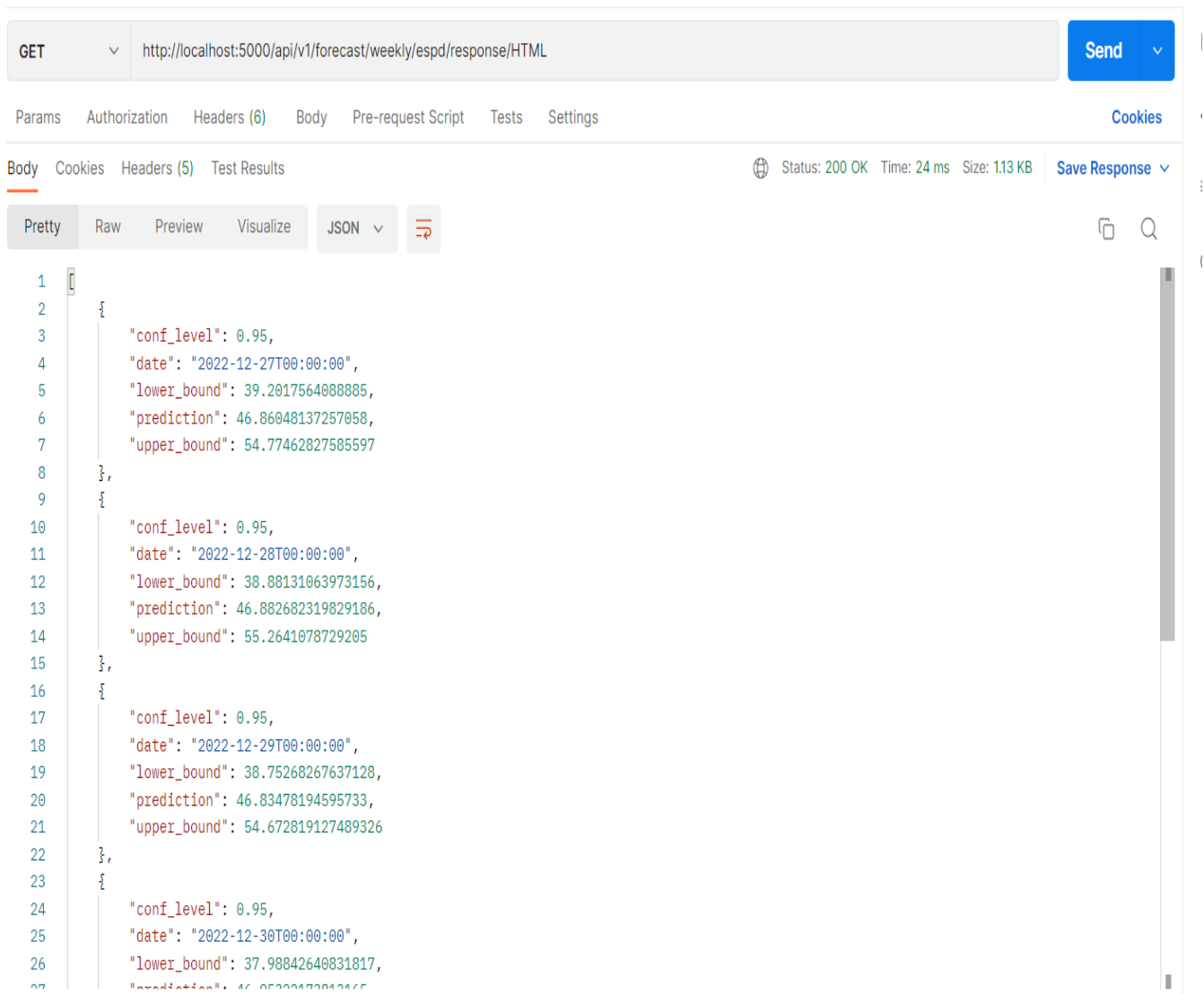


```
1 {
2   {
3     "conf_level": 0.95,
4     "date": "2022-12-27T00:00:00",
5     "lower_bound": 36.96710080406294,
6     "prediction": 45.17251776589646,
7     "upper_bound": 53.849152506113086
8   },
9   {
10    "conf_level": 0.95,
11    "date": "2022-12-28T00:00:00",
12    "lower_bound": 37.165779799548176,
13    "prediction": 45.3685051073812,
14    "upper_bound": 53.36073435720644
15  },
16  {
17    "conf_level": 0.95,
18    "date": "2022-12-29T00:00:00",
19    "lower_bound": 37.03619706358206,
20    "prediction": 45.085678898241035,
21    "upper_bound": 52.81052844587271
22  },
23  {
24    "conf_level": 0.95,
25    "date": "2022-12-30T00:00:00",
26    "lower_bound": 36.491496169050606,
27    "prediction": 45.040770074047755
```

Εικόνα 25 Αποτελέσματα πρόβλεψης του συνολικού αριθμού εξαχθέντων ESPD Response XML εγγράφων.

Μηχανισμός έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD.

Στην εικόνα 26 παρουσιάζονται τα αποτελέσματα από την ανάκτηση των προβλέψεων του συνολικού αριθμού ESPD Response αρχείων με βάση τον τύπο αρχείου HTML, μέχρι την επόμενη εβδομάδα.

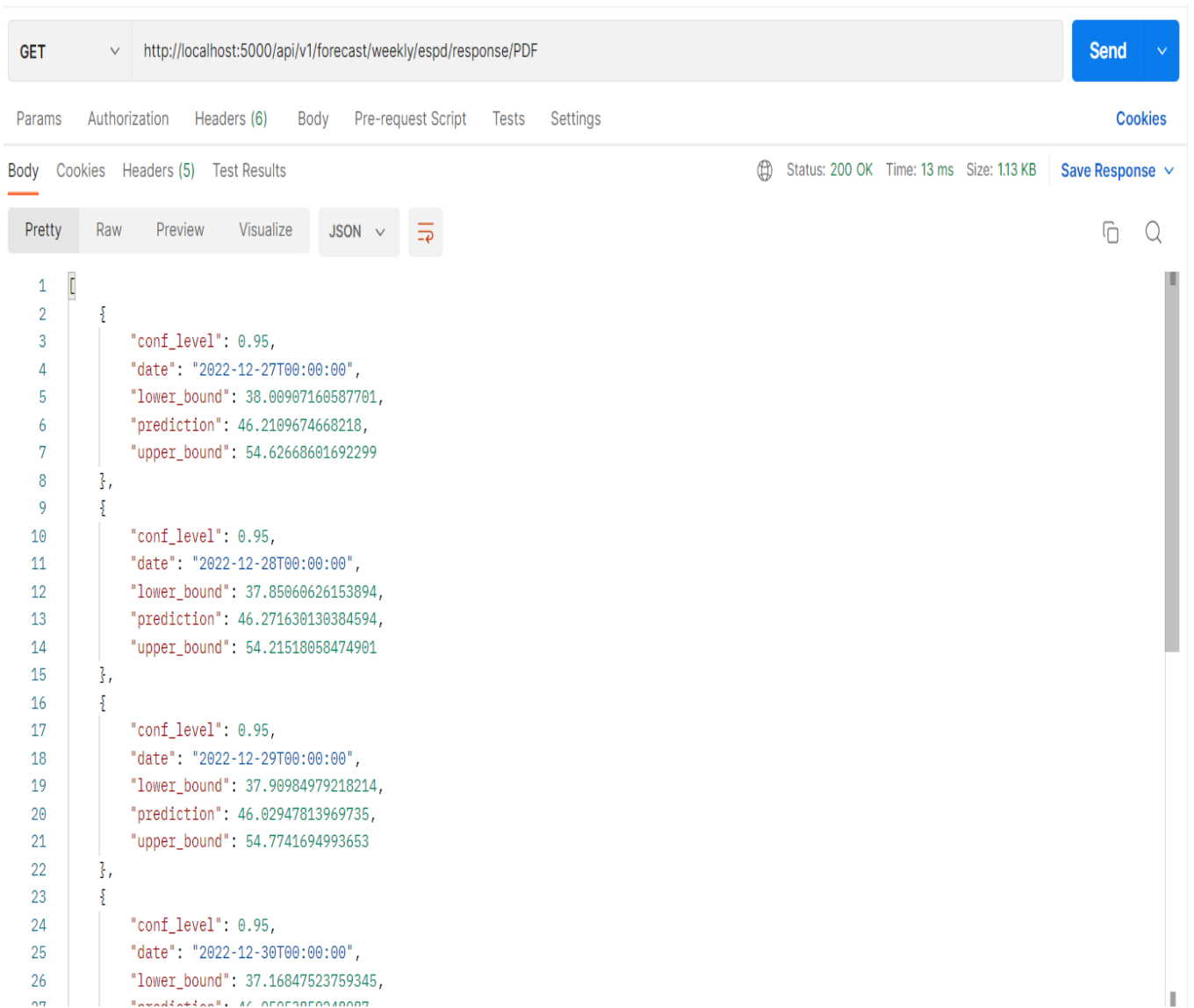


The screenshot shows a web browser's developer tools interface. At the top, a GET request is shown to the URL `http://localhost:5000/api/v1/forecast/weekly/esp/resp/html`. The response status is 200 OK, with a time of 24 ms and a size of 1.13 KB. The response body is displayed in JSON format, showing a list of forecast data points. The JSON structure is as follows:

```
1 {
2   {
3     "conf_level": 0.95,
4     "date": "2022-12-27T00:00:00",
5     "lower_bound": 39.2017564088885,
6     "prediction": 46.86048137257058,
7     "upper_bound": 54.77462827585597
8   },
9   {
10    "conf_level": 0.95,
11    "date": "2022-12-28T00:00:00",
12    "lower_bound": 38.88131063973156,
13    "prediction": 46.882682319829186,
14    "upper_bound": 55.2641078729205
15  },
16  {
17    "conf_level": 0.95,
18    "date": "2022-12-29T00:00:00",
19    "lower_bound": 38.75268267637128,
20    "prediction": 46.83478194595733,
21    "upper_bound": 54.672819127489326
22  },
23  {
24    "conf_level": 0.95,
25    "date": "2022-12-30T00:00:00",
26    "lower_bound": 37.98842640831817,
27    "prediction": 46.05200472041415
```

Εικόνα 26 Αποτελέσματα πρόβλεψης του συνολικού αριθμού εξαχθέντων ESPD Response HTML εγγράφων.

Στην εικόνα 27 παρουσιάζονται τα αποτελέσματα από την ανάκτηση των προβλέψεων του συνολικού αριθμού ESPD Response αρχείων με βάση τον τύπο αρχείου PDF, μέχρι την επόμενη εβδομάδα.



```
GET http://localhost:5000/api/v1/forecast/weekly/espdc/response/PDF

Params Authorization Headers (6) Body Pre-request Script Tests Settings Cookies

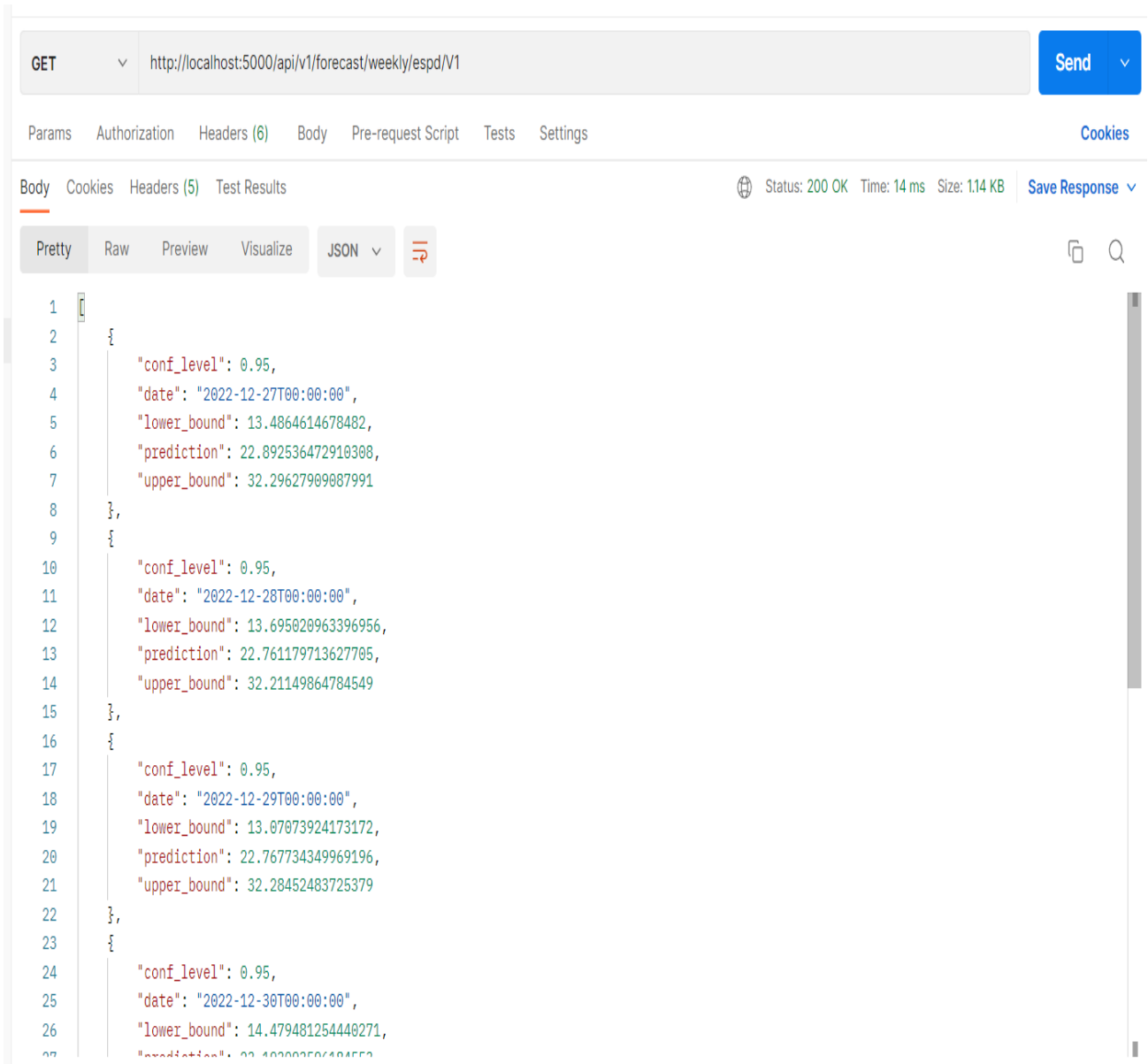
Body Cookies Headers (5) Test Results Status: 200 OK Time: 13 ms Size: 113 KB Save Response

Pretty Raw Preview Visualize JSON

1 {
2   {
3     "conf_level": 0.95,
4     "date": "2022-12-27T00:00:00",
5     "lower_bound": 38.00907160587701,
6     "prediction": 46.2109674668218,
7     "upper_bound": 54.62668601692299
8   },
9   {
10    "conf_level": 0.95,
11    "date": "2022-12-28T00:00:00",
12    "lower_bound": 37.85060626153894,
13    "prediction": 46.271630130384594,
14    "upper_bound": 54.21518058474901
15  },
16  {
17    "conf_level": 0.95,
18    "date": "2022-12-29T00:00:00",
19    "lower_bound": 37.90984979218214,
20    "prediction": 46.02947813969735,
21    "upper_bound": 54.7741694993653
22  },
23  {
24    "conf_level": 0.95,
25    "date": "2022-12-30T00:00:00",
26    "lower_bound": 37.16847523759345,
```

Εικόνα 27 Αποτελέσματα πρόβλεψης του συνολικού αριθμού εξαχθέντων ESPD Response PDF εγγράφων.

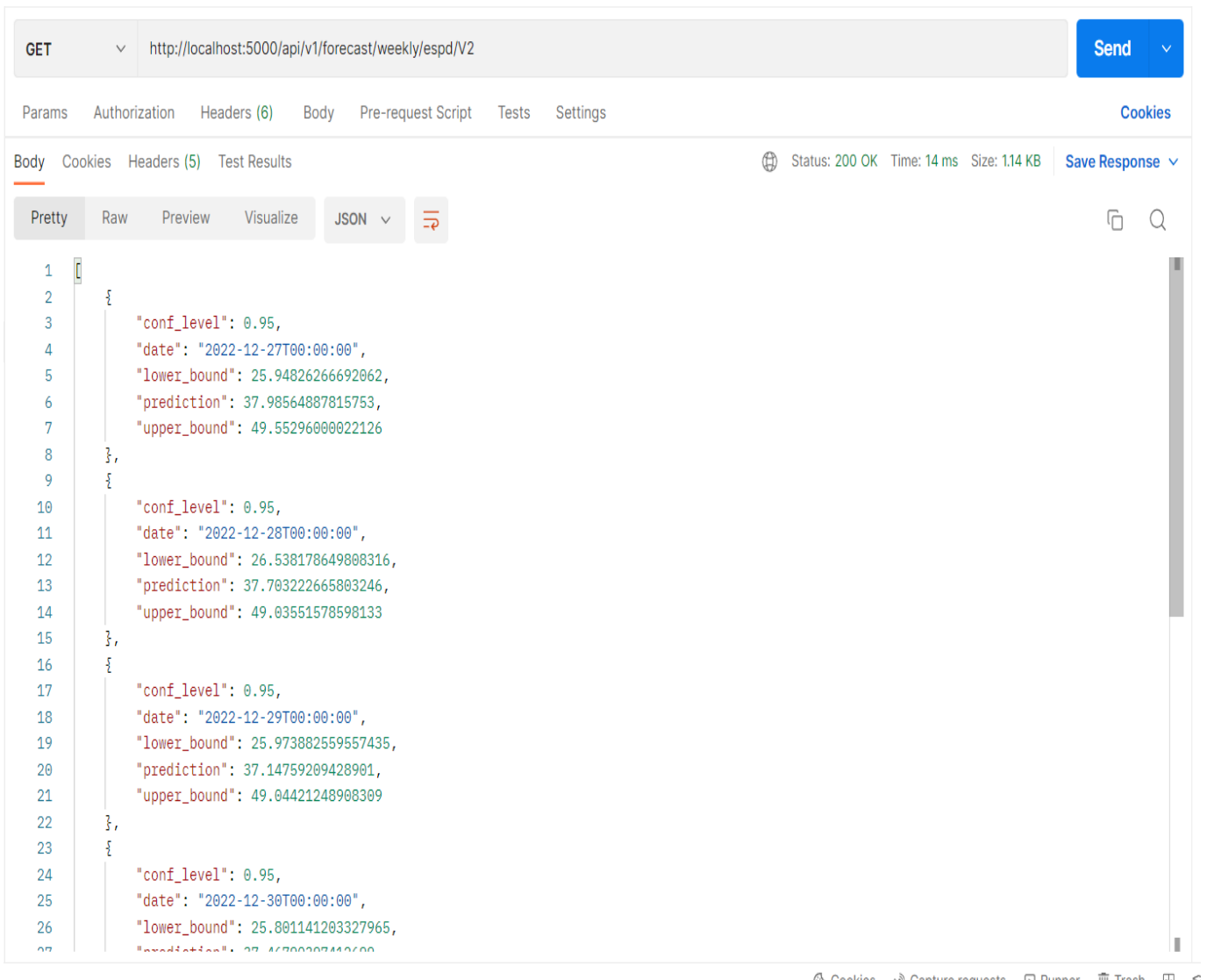
Στην εικόνα 28 παρουσιάζονται τα αποτελέσματα από την ανάκτηση των προβλέψεων του συνολικού αριθμού ESPD αρχείων της έκδοσης 1.χ.χ, μέχρι την επόμενη εβδομάδα.



```
1  {
2    "conf_level": 0.95,
3    "date": "2022-12-27T00:00:00",
4    "lower_bound": 13.4864614678482,
5    "prediction": 22.892536472910308,
6    "upper_bound": 32.29627909087991
7  },
8  {
9    "conf_level": 0.95,
10   "date": "2022-12-28T00:00:00",
11   "lower_bound": 13.695020963396956,
12   "prediction": 22.761179713627705,
13   "upper_bound": 32.21149864784549
14  },
15  {
16   "conf_level": 0.95,
17   "date": "2022-12-29T00:00:00",
18   "lower_bound": 13.07073924173172,
19   "prediction": 22.767734349969196,
20   "upper_bound": 32.28452483725379
21  },
22  {
23   "conf_level": 0.95,
24   "date": "2022-12-30T00:00:00",
25   "lower_bound": 14.479481254440271,
26   "prediction": 22.402002504404559
27  }
```

Εικόνα 28 Αποτελέσματα πρόβλεψης του συνολικού αριθμού εξαχθέντων ESPD έκδοσης 1.χ.χ.

Στην εικόνα 29 παρουσιάζονται τα αποτελέσματα από την ανάκτηση των προβλέψεων του συνολικού αριθμού ESPD αρχείων της έκδοσης 2.χ.χ, μέχρι την επόμενη εβδομάδα.



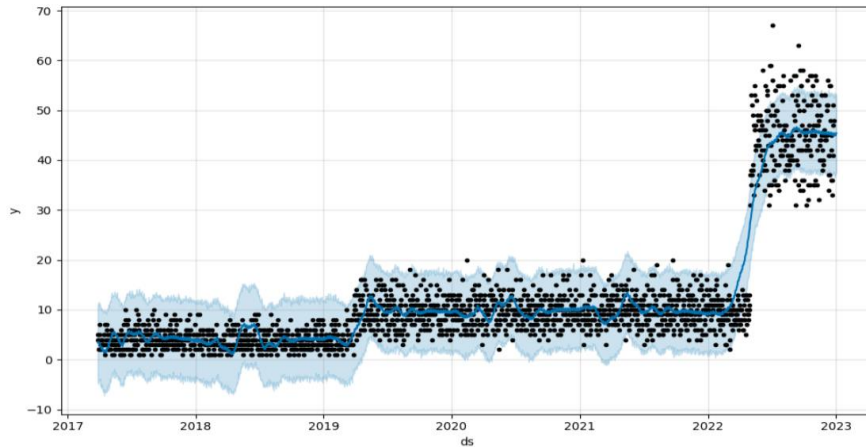
```
1 {
2   {
3     "conf_level": 0.95,
4     "date": "2022-12-27T00:00:00",
5     "lower_bound": 25.94826266692062,
6     "prediction": 37.98564887815753,
7     "upper_bound": 49.5529600022126
8   },
9   {
10    "conf_level": 0.95,
11    "date": "2022-12-28T00:00:00",
12    "lower_bound": 26.538178649808316,
13    "prediction": 37.703222665803246,
14    "upper_bound": 49.03551578598133
15  },
16  {
17    "conf_level": 0.95,
18    "date": "2022-12-29T00:00:00",
19    "lower_bound": 25.973882559557435,
20    "prediction": 37.14759209428901,
21    "upper_bound": 49.04421248908309
22  },
23  {
24    "conf_level": 0.95,
25    "date": "2022-12-30T00:00:00",
26    "lower_bound": 25.801141203327965,
27    "prediction": 37.46700207442600
```

Εικόνα 29 Αποτελέσματα πρόβλεψης του συνολικού αριθμού εξαθθέντων ESPD έκδοσης 2.χ.χ.

Τέλος, τα διαγράμματα χρονοσειρών που παρουσιάζουν τις προβλέψεις του κάθε μοντέλου μηχανικής μάθησης βρίσκονται στο path /models/info. Η μπλε γραμμή είναι οι προβλέψεις του μοντέλου κάθε χρονική στιγμή, ενώ η γαλάζια περιοχή το διάστημα πρόβλεψης (prediction interval).

Time Series Models.

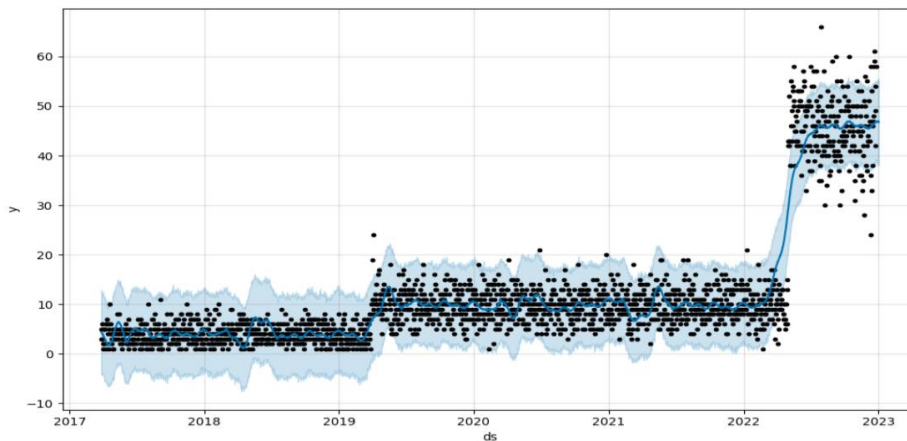
ESPD Requests in HTML format.



1. The time series model for exported ESPD Requests in HTML format. The horizontal axis is dates and the vertical axis is the exported files per date. The blue line is the forecasts of model.

Εικόνα 30 Το μοντέλο πρόβλεψης για τα ESPD Request HTML έγγραφα.

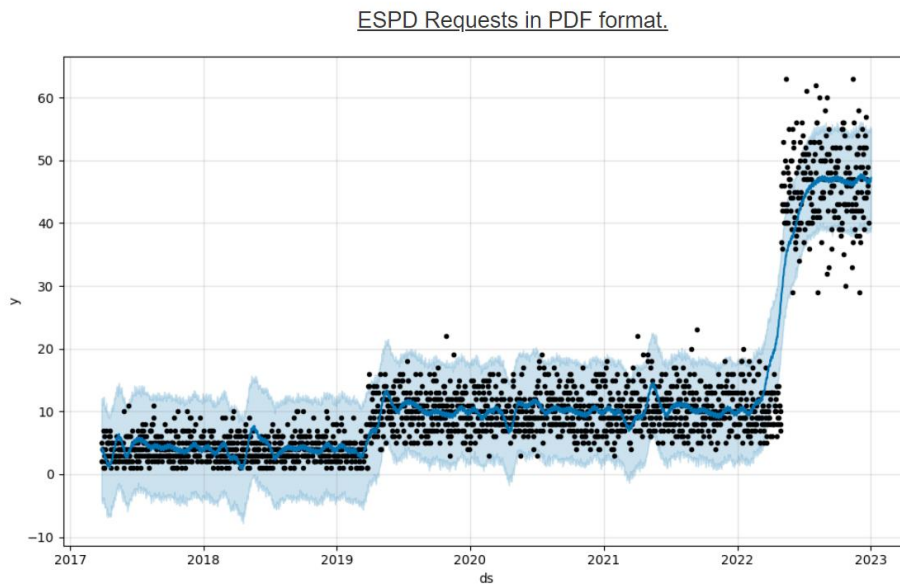
ESPD Responses in HTML format.



2. The time series model for exported ESPD Responses in HTML format. The horizontal axis is dates and the vertical axis is the exported files per date. The blue line is the forecasts of model.

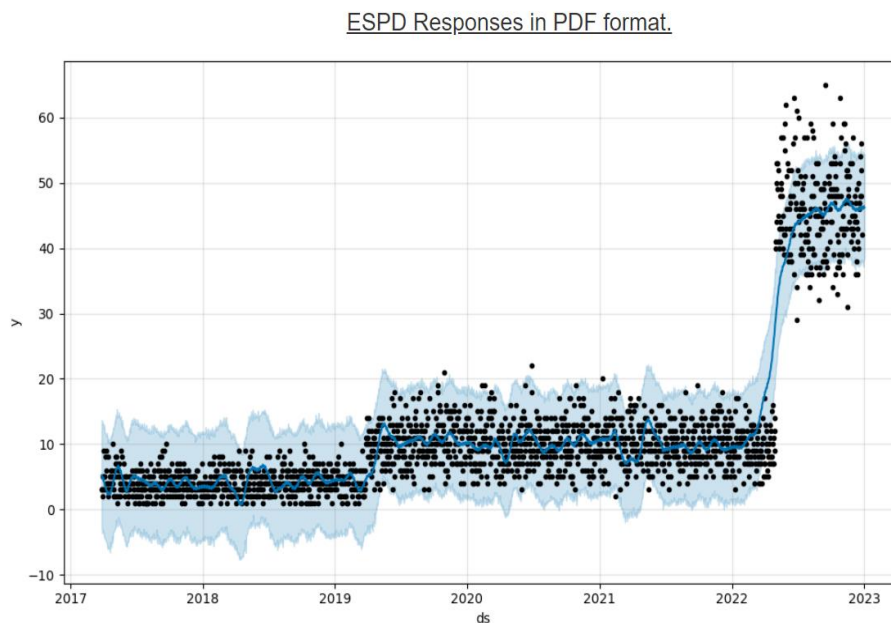
Εικόνα 31 Το μοντέλο πρόβλεψης για τα ESPD Response HTML έγγραφα.

Μηχανισμός έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD.



3. The time series model for exported ESPD Requests in PDF format. The horizontal axis is dates and the vertical axis is the exported files per date. The blue line is the forecasts of model.

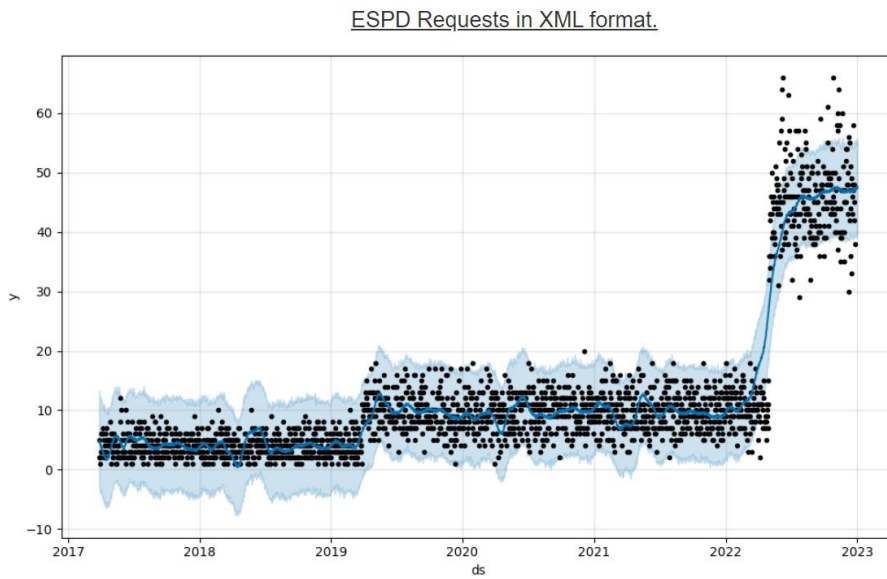
Εικόνα 32 Το μοντέλο πρόβλεψης για τα ESPD Request PDF έγγραφα.



4. The time series model for exported ESPD Responses in PDF format. The horizontal axis is dates and the vertical axis is the exported files per date. The blue line is the forecasts of model.

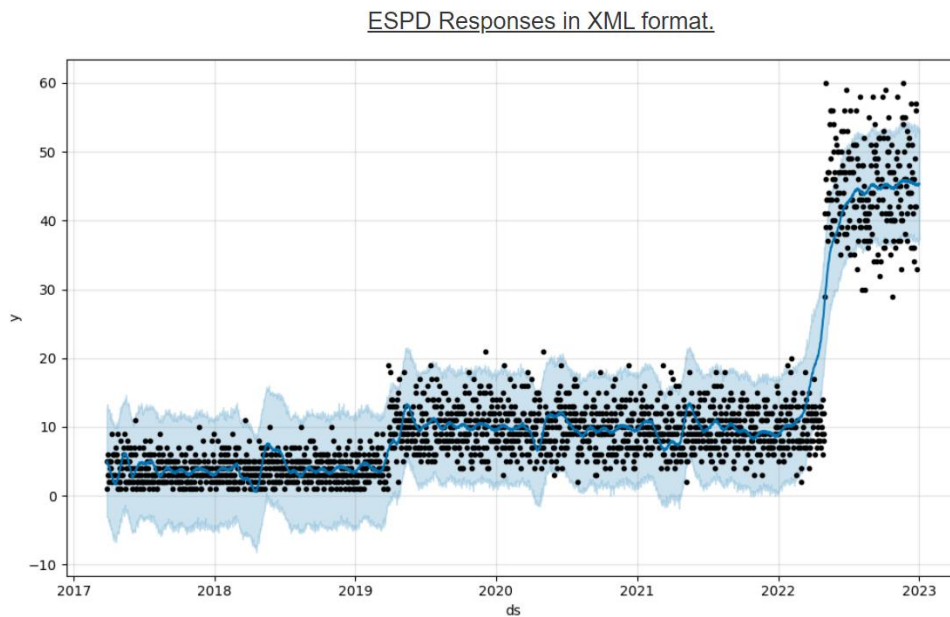
Εικόνα 33 Το μοντέλο πρόβλεψης για τα ESPD Response PDF έγγραφα.

Μηχανισμός έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD.



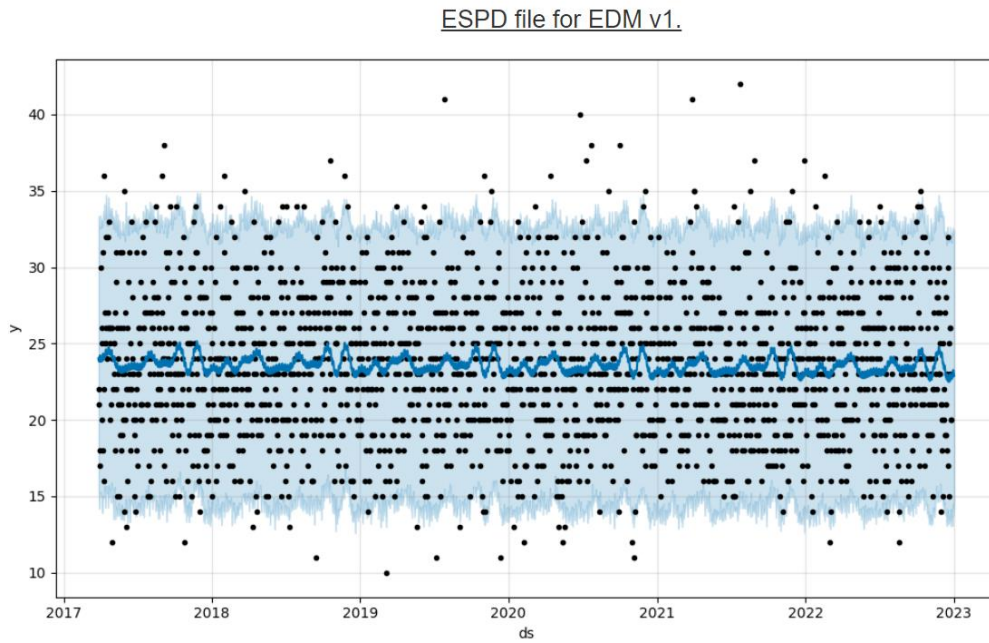
5. The time series model for exported ESPD Requests in XML format. The horizontal axis is dates and the vertical axis is the exported files per date. The blue line is the forecasts of model.

Εικόνα 34 Το μοντέλο πρόβλεψης για τα ESPD Request XML έγγραφα.



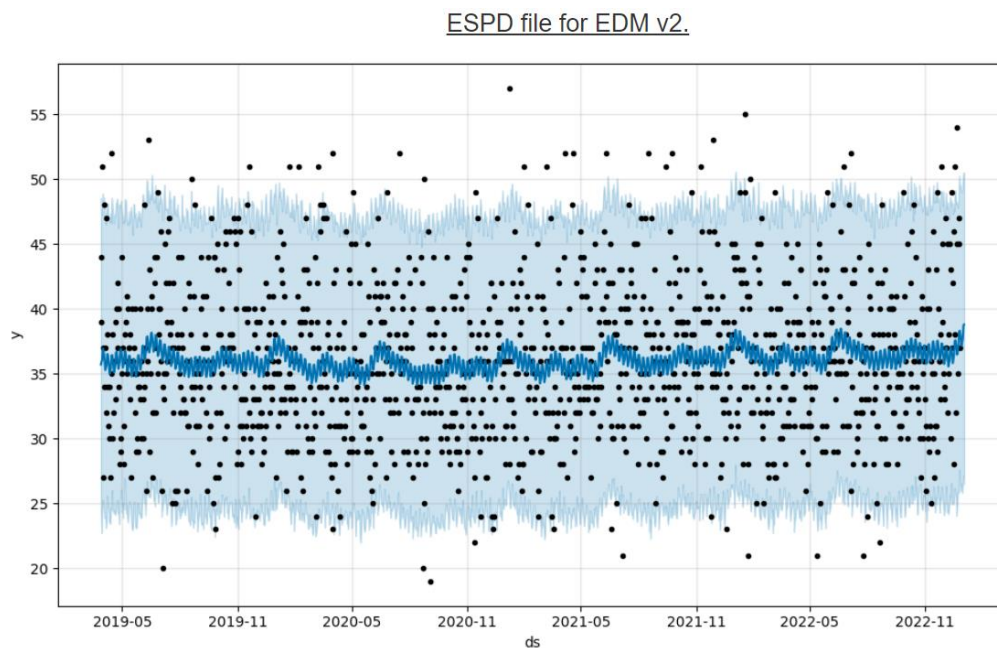
6. The time series model for exported ESPD Responses in XML format. The horizontal axis is dates and the vertical axis is the exported files per date. The blue line is the forecasts of model.

Εικόνα 35 Το μοντέλο πρόβλεψης για τα ESPD Response XML έγγραφα.



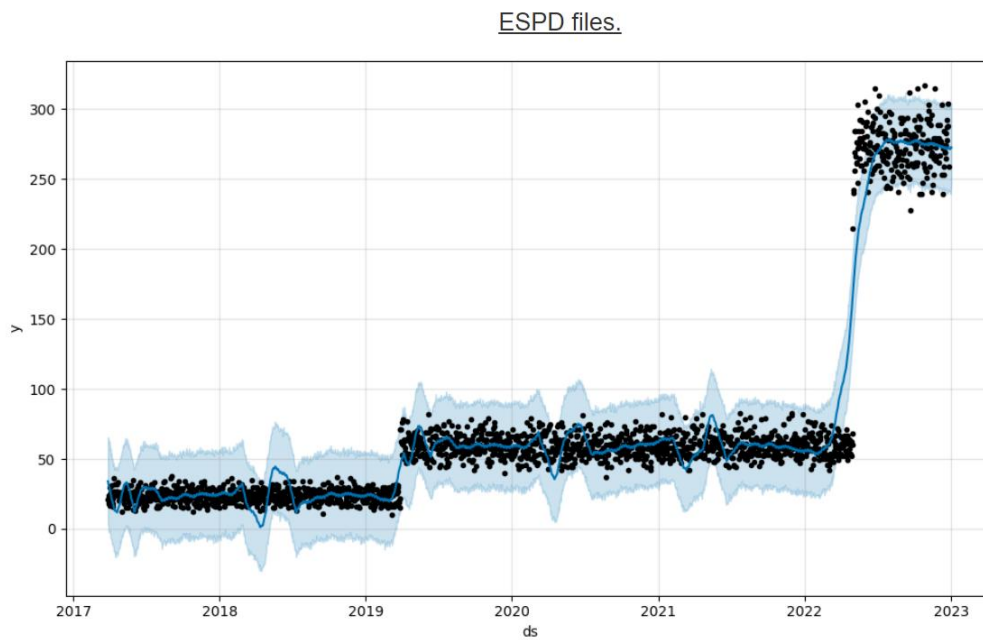
7. The time series model for exported ESPD files for EDM version 1. The horizontal axis is dates and the vertical axis is the exported files per date. The blue line is the forecasts of model.

Εικόνα 36 Το μοντέλο πρόβλεψης για τα ESPD έγγραφα έκδοσης 1.κ.κ.



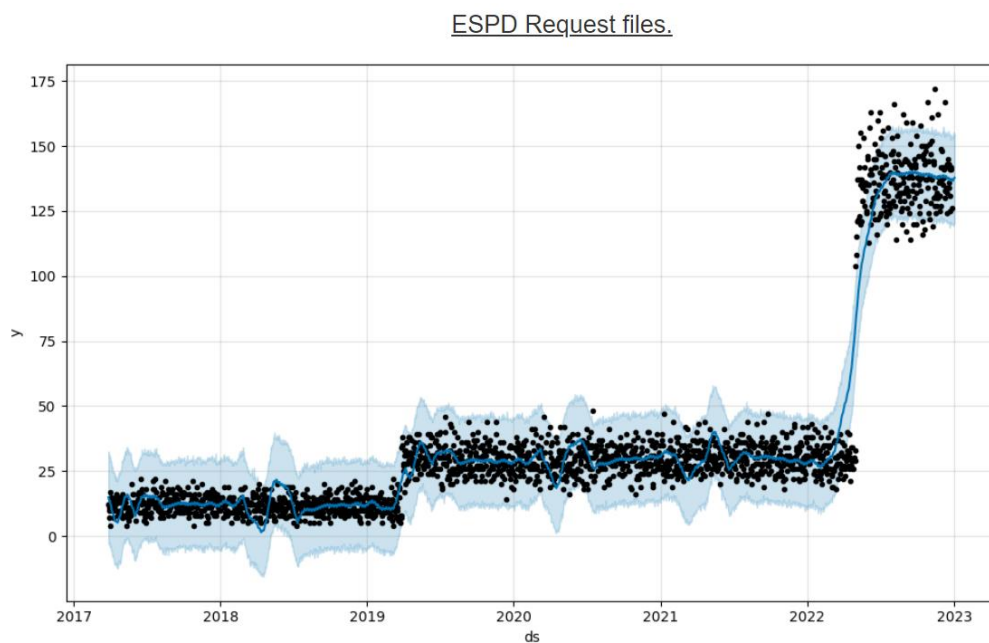
8. The time series model for exported ESPD files for EDM version 2. The horizontal axis is dates and the vertical axis is the exported files per date. The blue line is the forecasts of model.

Εικόνα 37 Το μοντέλο πρόβλεψης για τα ESPD έγγραφα έκδοσης 2.κ.κ.



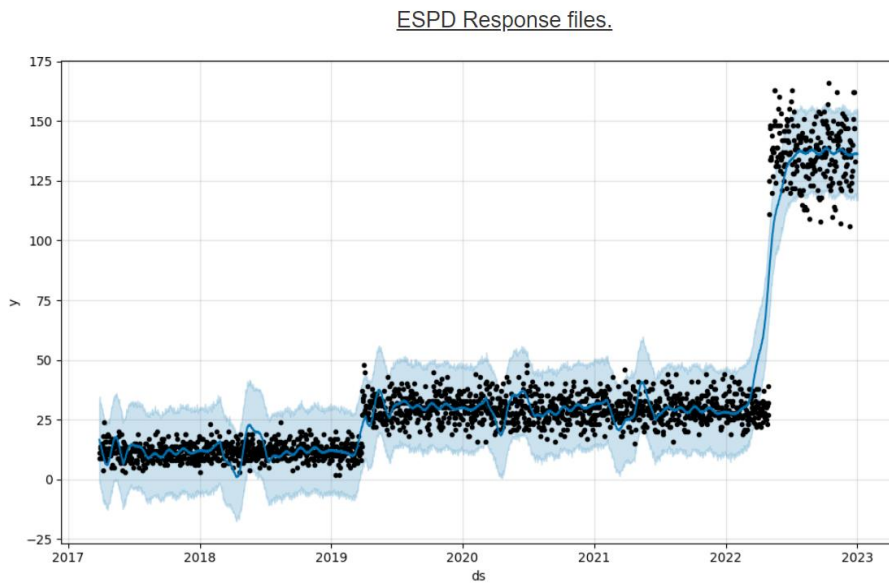
9. The time series model for exported ESPD files. The horizontal axis is dates and the vertical axis is the exported files per date. The blue line is the forecasts of model.

Εικόνα 38 Το μοντέλο πρόβλεψης για τον συνολικό αριθμό των ESPD εγγράφων.



10. The time series model for exported ESPD Requests. The horizontal axis is dates and the vertical axis is the exported files per date. The blue line is the forecasts of model.

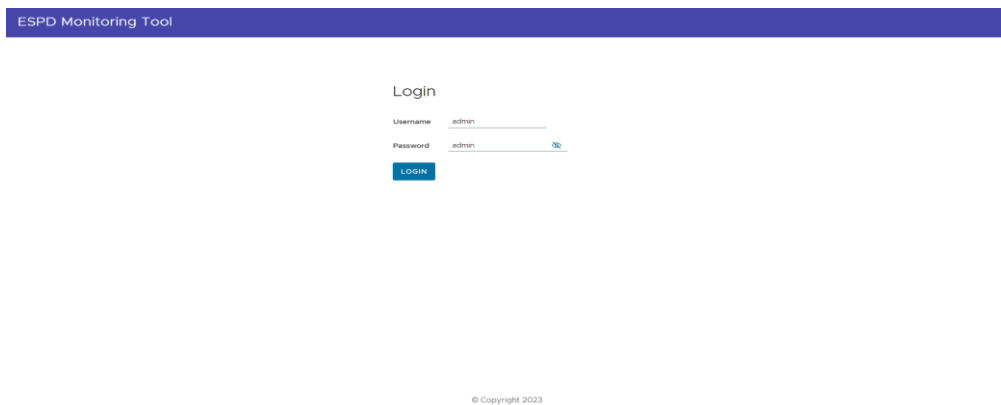
Εικόνα 39 Το μοντέλο πρόβλεψης για τον συνολικό αριθμό των ESPD Request εγγράφων.



Εικόνα 40 Το μοντέλο πρόβλεψης για το συνολικό αριθμό των ESPD Response εγγράφων.

5.3 Παρουσίαση του μηχανισμού έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD

Αρχικά, ο διαχειριστής του συστήματος συνδέεται στο σύστημα με όνομα χρήστη (username) admin και συνθηματικό (password) admin. Στην εικόνα 41, παρουσιάζεται η αρχική σελίδα εισόδου των χρηστών.




Εικόνα 41 Αρχική σελίδα εισόδου.

Στην Εικόνα 42, ο χρήστης πραγματοποιεί είσοδο στην εφαρμογή, χρησιμοποιώντας για όνομα χρήστη και συνθηματικό τη λέξη admin.

Login

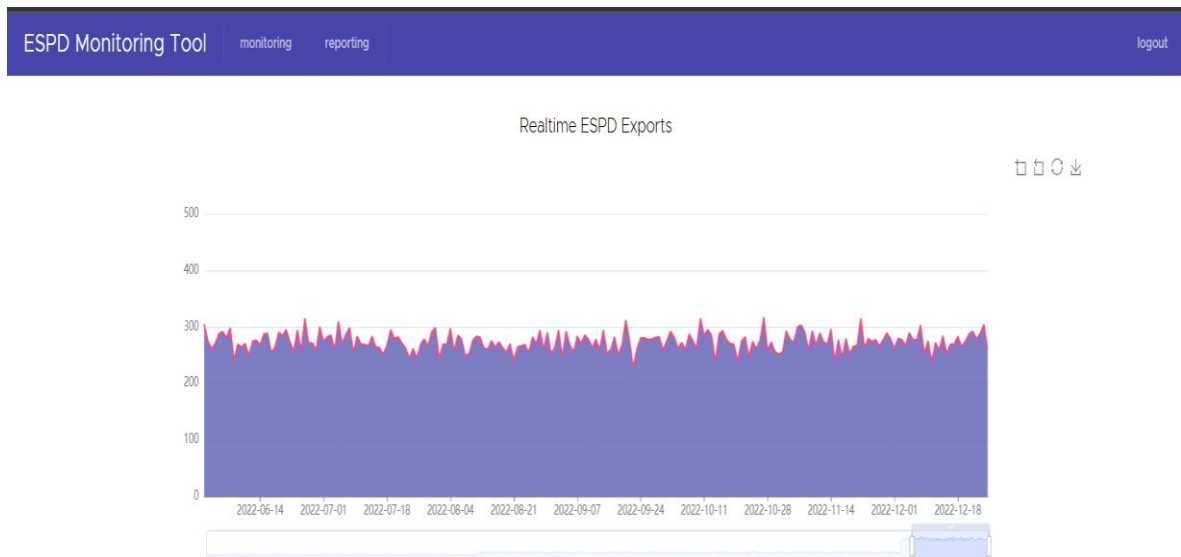
Username

Password 

LOGIN

Εικόνα 42 Εισαγωγή username και password για την είσοδο στο σύστημα.

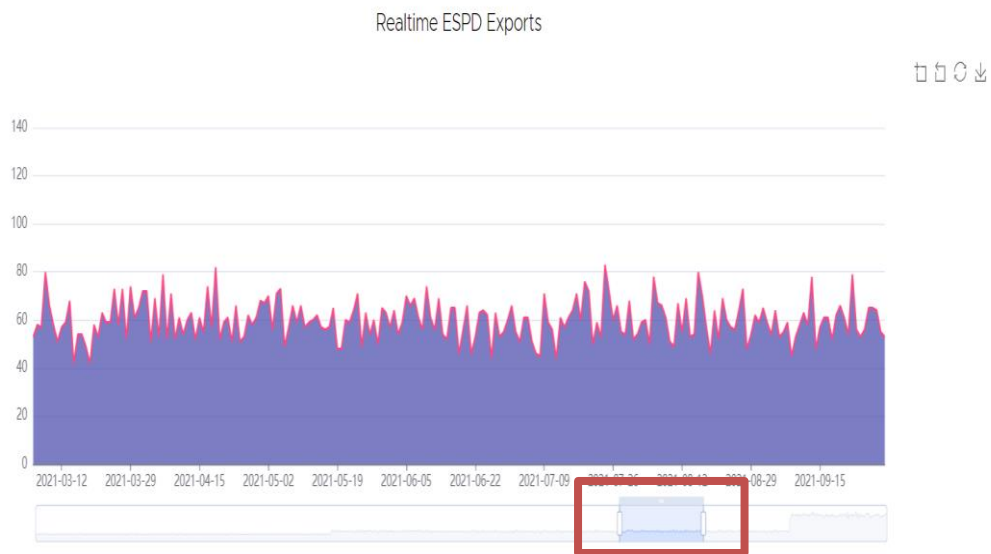
Η διεπαφή χρήστη (UI) του μηχανισμού έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD περιλαμβάνει ένα διάγραμμα χρονοσειράς όπως εμφανίζεται στην εικόνα 43.



Εικόνα 43 Διάγραμμα χρονοσειράς συνολικού αριθμού εξαχθέντων ESPD αρχείων.

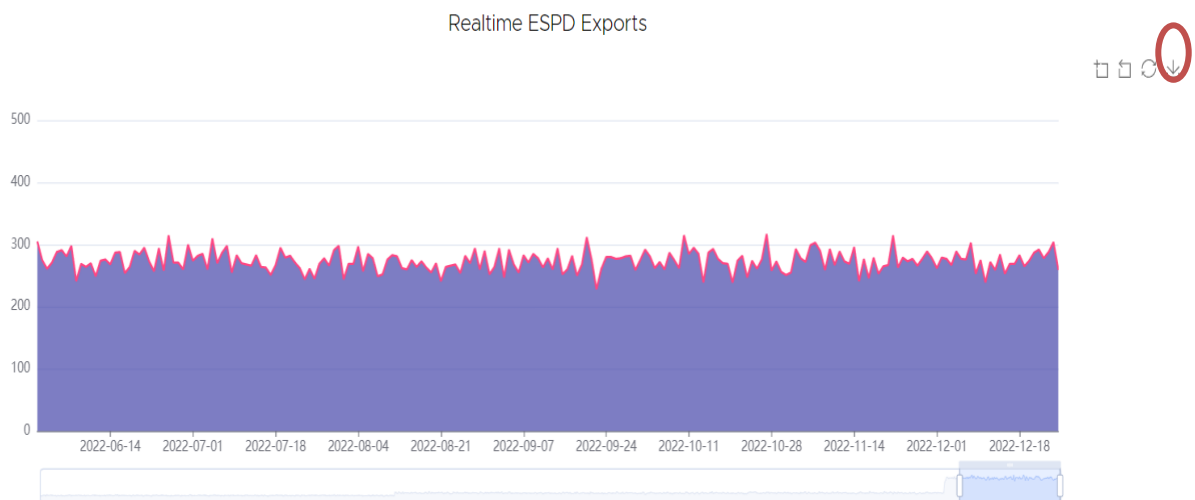
Στο συγκεκριμένο διάγραμμα γίνεται να επιλέξουμε το χρονικό διάστημα χρησιμοποιώντας τη μπάρα κάτω από το διάγραμμα.

Μηχανισμός έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD.



Εικόνα 44 Επιλογή συγκεκριμένου χρονικού διαστήματος από τη χρονοσειρά.

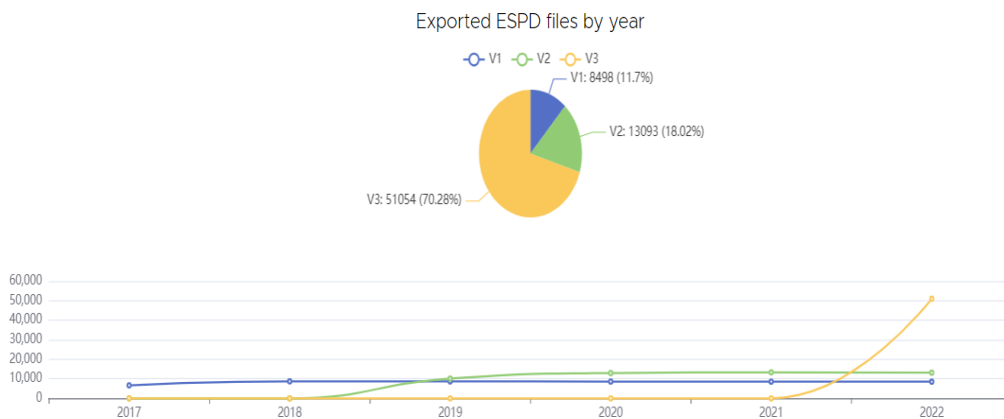
Επίσης, δίνεται η δυνατότητα αποθήκευσης σε εικόνα τύπου PNG πατώντας το εικονίδιο που έχει κυκλωθεί στην Εικόνα 45.



Εικόνα 45 Αποθήκευση του διαγράμματος χρονοσειράς ως εικόνα.

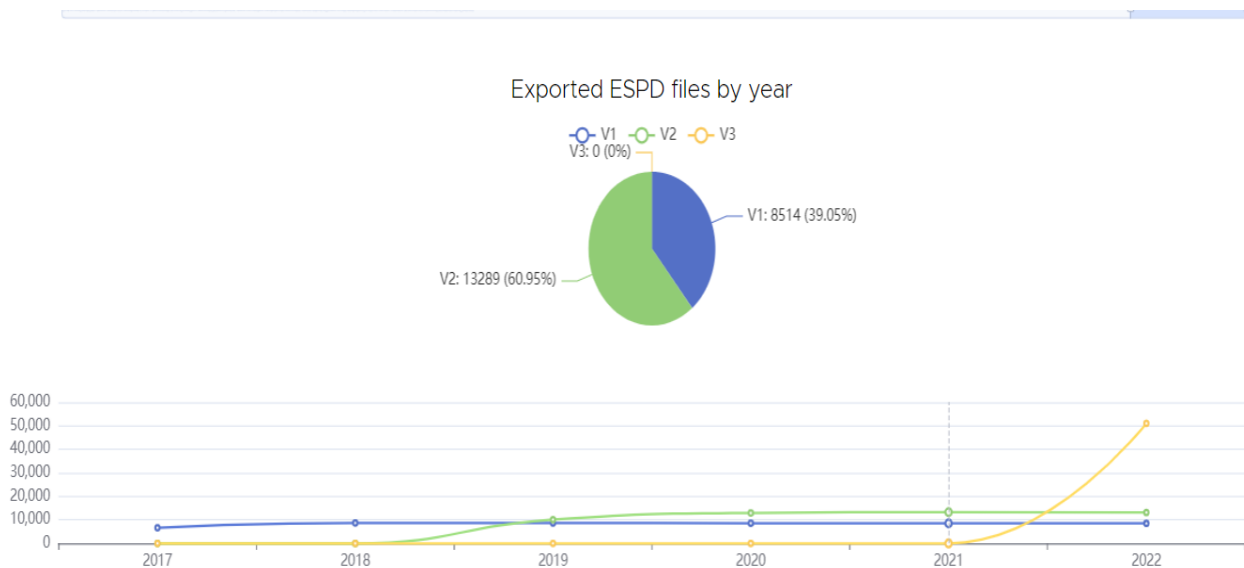
Στη συνέχεια, παρουσιάζεται ένα διάγραμμα γραμμής και μία πίτα, τα οποία παρουσιάζουν τις συνολικές εξαγωγές ESPD αρχείων ανά έτος για κάθε έκδοση.

Μηχανισμός έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD.



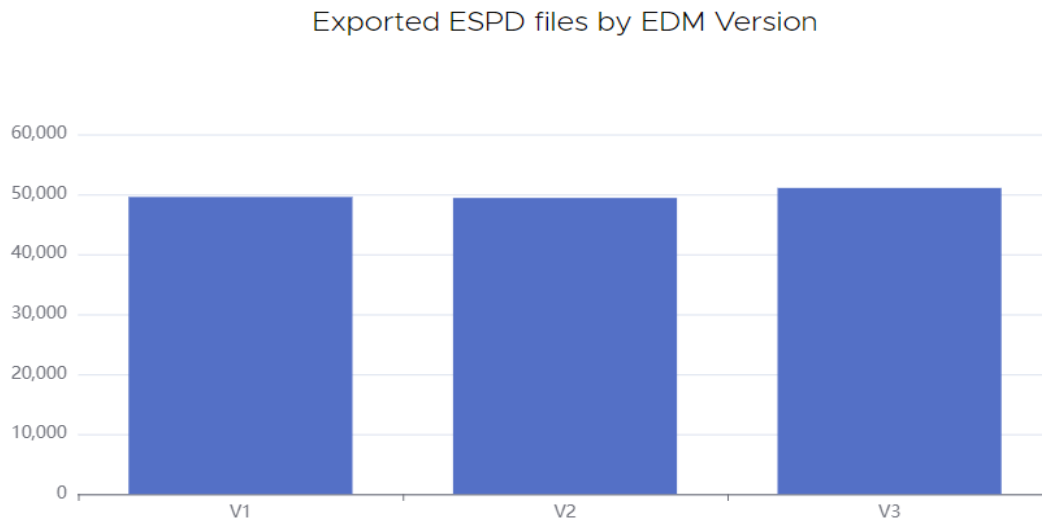
Εικόνα 46 Διάγραμμα πίτας και γραμμής ανά έτος για κάθε έκδοση.

Αν ο χρήστης επιλέξει ένα διαφορετικό έτος, τότε και τα περιεχόμενα της πίτας θα αλλάξουν, δείχνοντας τα ποσοστά και του πλήθους των εγγράφων ESPD που έχουν εξαχθεί για αυτό έτος. Στην εικόνα 47 επιλέγεται το έτος 2021.



Εικόνα 47 Διάγραμμα πίτας για το 2021, επιλέγοντας το έτος 2021 με τον κέρσορα πάνω στο διάγραμμα γραμμής για το 2021.

Επίσης, εμφανίζεται ένα ραβδόγραμμα με το πλήθος των ESPD εγγράφων που εξήχθησαν για κάθε έκδοση συνολικά από το σύστημα ESPD.



Εικόνα 48 Ραβδόγραμμα με τις εξαγωγές ESPD αρχείων ανά έκδοση.

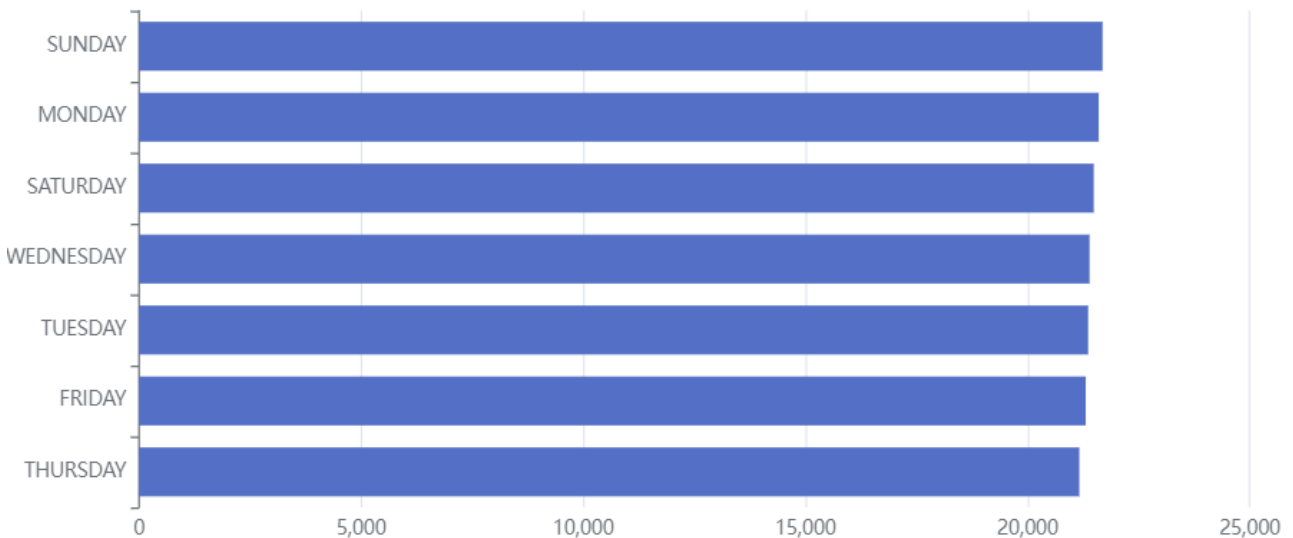
Ο χρήστης αν κάνει κλικ με τον κέρσορα πάνω σε κάποια από τη μπάρα, τότε θα εμφανιστεί ο συνολικός αριθμός ESPD Request και Response για τη συγκεκριμένη έκδοση.



Εικόνα 49 Ραβδόγραμμα για τα ESPD Request και Response της έκδοσης 2.x.x.

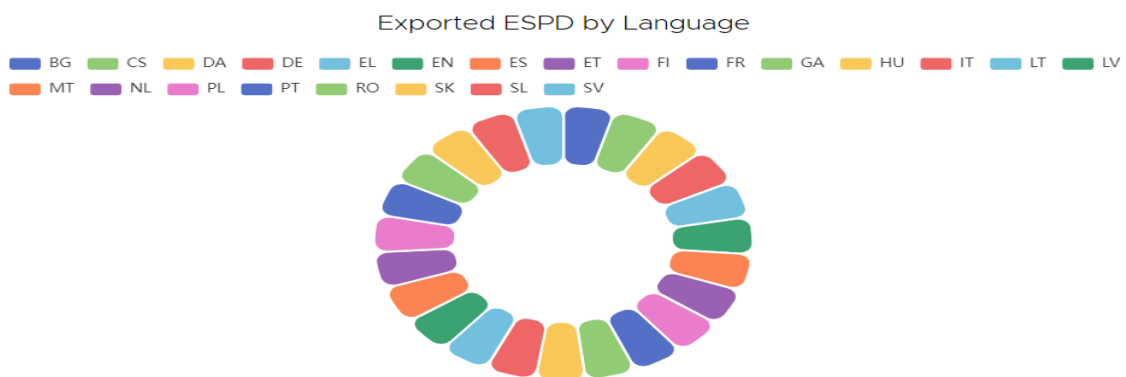
Στην εικόνα 50 παρουσιάζεται ένα ραβδόγραμμα, με την εξαγωγή ESPD αρχείων ανά ημέρα. Οι ράβδοι είναι ταξινομημένοι σε φθίνουσα σειρά με βάση την εξαγωγή ESPD αρχείων, για να γίνεται εύκολα αντιληπτό ποια ημέρα εξαγονται τα περισσότερα ESPD αρχεία.

ESPD Exports By Day



Εικόνα 50 Ραβδόγραμμα με το πλήθος ESPD αρχείων που έχουν εξαχθεί ανά ημέρα.

Στην εικόνα 51 εμφανίζεται ένα donut με τις διαφορετικές γλώσσες στις οποίες έχουν γραφεί τα ESPD έγγραφα που έχουν εξαχθεί από το ESPD.



Εικόνα 51 Διάγραμμα donut με το πλήθος των ESPD αρχείων για κάθε γλώσσα.

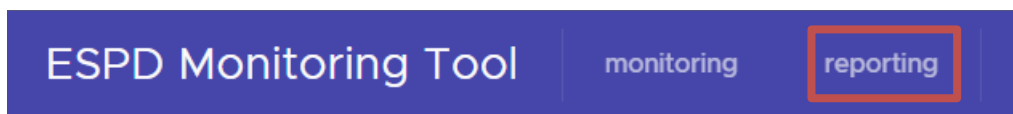
Στην εικόνα 52 παρουσιάζεται ένα διάγραμμα donut με το συνολικό αριθμό εξαγωγών, για κάθε τύπο αρχείου ESPD που έχει εξαχθεί από το σύστημα ESPD.

Exported ESPD by file format



Εικόνα 52 Διάγραμμα donut με τον συνολικό αριθμό εξαγωγών για κάθε τύπο αρχείου ESPD.

Ένας χρήστης για να εκδώσει μία αναφορά για το σύστημα ESPD θα πρέπει να πατήσει την επιλογή reporting από το πανβάρ και στη συνέχεια να πατήσει το κουμπί PDF.



Εικόνα 53 Επιλογή του reporting από το πανβάρ.

Reporting

Click the below button to export the report of ESPD/VCD System in PDF format.



Εικόνα 54 Κλίκ στο κουμπί PDF για την έκδοση της αναφοράς του Συστήματος ESPD.

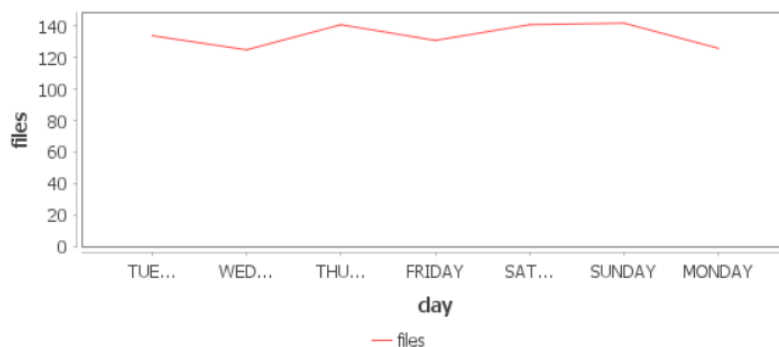
Η αναφορά του ESPD περιλαμβάνει δύο μέρη το Μέρος I: Δεδομένα σε πραγματικό χρόνο και το Μέρος II: Προβλέψεις.

Στο Μέρος I παρέχονται οι εξής πληροφορίες:

- Οι συνολικές εξαγωγές ESPD Request αρχείων για αυτή την εβδομάδα.
- Οι συνολικές εξαγωγές ESPD Response αρχείων για αυτή την εβδομάδα.
- Οι εξαγωγές για κάθε έκδοση ESPD Request μέχρι σήμερα.
- Οι εξαγωγές για κάθε έκδοση ESPD Response μέχρι σήμερα.
- Οι εξαγωγές ESPD Request ανά γλώσσα μέχρι σήμερα.
- Οι εξαγωγές ESPD Response ανά γλώσσα μέχρι σήμερα.
- Οι εξαγωγές ESPD Request ανά μορφή αρχείου μέχρι σήμερα.
- Οι εξαγωγές ESPD Response ανά μορφή αρχείου μέχρι σήμερα.
- Οι συνολικές εξαγωγές ESPD Request αρχείων ανά έτος για την έκδοση V1.
- Οι συνολικές εξαγωγές ESPD Response αρχείων ανά έτος για την έκδοση V1.
- Οι συνολικές εξαγωγές ESPD Request αρχείων ανά έτος για την έκδοση V2.
- Οι συνολικές εξαγωγές ESPD Response αρχείων ανά έτος για την έκδοση V2.
- Ο συνολικός αριθμός εξαγόμενων ESPD Request για την έκδοση V3 ανά έτος.
- Ο συνολικός αριθμός εξαγόμενων ESPD Response για την έκδοση V3 ανά έτος.

Μέρος I: Δεδομένα σε πραγματικό χρόνο

Οι συνολικές εξαγωγές ESPD Request αρχείων για αυτή την εβδομάδα παρουσιάζονται στην Εικόνα 1.



Εικόνα 1: Οι εξαγωγές ESPD Request αρχείων αυτή την εβδομάδα.

Επίσης, παρουσιάζονται αναλυτικά δεδομένα για κάθε ημέρα της εβδομάδας στον Πίνακα 1.

Πίνακα 1: Εξαγωγές ESPD Request ανά ημέρα της εβδομάδας.

Ημέρα	Πλήθος αρχείων
Τρίτη (2022-12-20)	134
Τετάρτη (2022-12-21)	125
Πέμπτη (2022-12-22)	141
Παρασκευή (2022-12-23)	131
Σάββατο (2022-12-24)	141
Κυριακή (2022-12-25)	142
Δευτέρα (2022-12-26)	126
Συνολικά	940

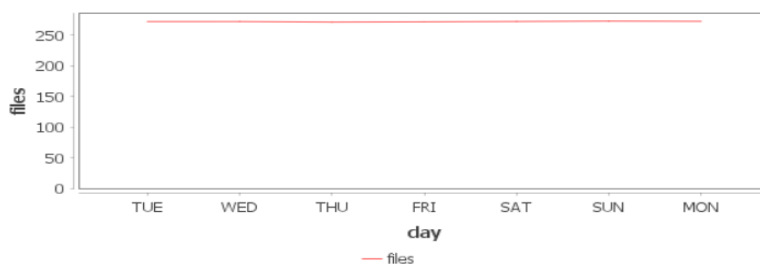
Εικόνα 55 Ενδεικτική εικόνα του Μέρους I μετά την έκδοση της αναφοράς για το σύστημα ESPD.

Στο Μέρος II παρέχονται οι εξής πληροφορίες:

- Οι συνολικές εξαγωγές ESPD αρχείων που προβλέπονται για κάθε ημέρα της επόμενης εβδομάδας.
- Οι συνολικές εξαγωγές ESPD Request αρχείων που προβλέπονται για κάθε ημέρα της επόμενης εβδομάδας.
- Οι συνολικές εξαγωγές ESPD Response αρχείων που προβλέπονται για κάθε ημέρα της επόμενης εβδομάδας.
- Τα εξαγόμενα ESPD Request σε μορφή XML που προβλέπονται για την επόμενη εβδομάδα.
- Τα εξαγόμενα ESPD Response σε μορφή XML που προβλέπονται για την επόμενη εβδομάδα.
- Τα εξαγόμενα ESPD Request σε μορφή HTML που προβλέπονται για την επόμενη εβδομάδα.
- Τα εξαγόμενα ESPD Response σε μορφή HTML που προβλέπονται για την επόμενη εβδομάδα.
- Τα εξαγόμενα ESPD Request σε μορφή PDF που προβλέπονται για την επόμενη εβδομάδα.
- Τα εξαγόμενα ESPD Response σε μορφή PDF που προβλέπονται για την επόμενη εβδομάδα.
- Τα εξαγόμενα ESPD αρχεία της έκδοσης 1 που προβλέπονται για την επόμενη εβδομάδα.
- Τα εξαγόμενα ESPD αρχεία της έκδοσης 2 που προβλέπονται για την επόμενη εβδομάδα.

Μέρος II: Προβλέψεις

Οι συνολικές εξαγωγές ESPD αρχείων που προβλέπεται για κάθε ημέρα της επόμενης εβδομάδας παρουσιάζονται στην Εικόνα 15.



Εικόνα 15: Οι εξαγωγές ESPD αρχείων την επόμενη εβδομάδα.

Οι συνολικές εξαγωγές ESPD αρχείων που προβλέπεται για κάθε ημέρα της επόμενης εβδομάδας παρουσιάζονται στον Πίνακα 15. Είναι βέβαιο κατά 95% ότι οι εξαγωγές αρχείων κυμαίνονται σύμφωνα με τη στήλη «Διάστημα Πρόβλεψης».

Πίνακα 15: Εξαγωγές ESPD αρχείων ανά ημέρα της επόμενης εβδομάδας.

Ημέρα	Πλήθος αρχείων	Διάστημα Πρόβλεψης
Τρίτη (2022-12-27)	272	241 - 300
Τετάρτη (2022-12-28)	272	242 - 305
Πέμπτη (2022-12-29)	271	242 - 303
Παρασκευή (2022-12-30)	272	241 - 305
Σάββατο (2022-12-31)	272	239 - 301
Κυριακή (2023-01-01)	273	243 - 305
Δευτέρα (2023-01-02)	273	247 - 303
Συνολικά	1.905	—

Εικόνα 56 Ενδεικτική εικόνα του Μέρους II μετά την έκδοση της αναφοράς για το σύστημα ESPD.

6. ΣΥΜΠΕΡΑΣΜΑΤΑ ΚΑΙ ΜΕΛΛΟΝΤΙΚΕΣ ΚΑΤΕΥΘΥΝΣΕΙΣ

6.1 Τελικά Συμπεράσματα

Η παρούσα διπλωματική εργασία ανέδειξε την αξία που προσδίδει η χρήση της επιχειρησιακής ευφυΐας (BI) στη λήψη αποφάσεων ενός οργανισμού. Πιο συγκεκριμένα, ο σκοπός της εργασίας ήταν η υλοποίηση μίας λύσης επιχειρησιακής ευφυΐας που ονομάζεται μηχανισμός έκδοσης αναφορών και παρακολούθησης του συστήματος ESPD, που παρουσιάζει την κατάσταση του συστήματος ESPD σε πραγματικό χρόνο (real-time) και παραθέτει προβλέψεις με τη συμβολή χρονοσειρών, συνεισφέροντας έτσι στη λήψη αποφάσεων με ορίζοντα όχι μόνο στο παρόν αλλά και στο μέλλον. Σε αυτό το πλαίσιο, έγινε παρουσίαση της ανάλυσης χρονοσειρών που αποτελούν ένα σημαντικό εργαλείο για τη δημιουργία προβλεπτικών μοντέλων που θα μπορούν να κατανοούν τις τάσεις και με βάση αυτές να προβλέπουν τις μελλοντικές τάσεις και συμπεριφορές μίας μεταβλητής ενδιαφέροντος (variable of interest). Συνεπώς, το συμπέρασμα που συνάγεται από την παρούσα διπλωματική εργασία είναι ότι η επιχειρησιακή ευφυΐα (BI) μπορεί να συνεισφέρει στην ανακάλυψη άγνωστων πτυχών για τη συμπεριφορά μίας μεταβλητής ενδιαφέροντος (variable of interest), όπως οι εξαγωγές ESPD εγγράφων, με τη χρήση κατάλληλων τεχνικών ανάλυσης δεδομένων.

6.2 Μελλοντικές βελτιώσεις

Οι εκδόσεις του συστήματος ESPD αλλάζουν συχνά με την πάροδο του χρόνου, γι' αυτό η ενημέρωση του μηχανισμού είναι απαραίτητη. Επίσης, θα ήταν ιδιαίτερα ωφέλιμο η αναφορά που εκδίδεται από το μηχανισμό να πιστοποιείται με ψηφιακή υπογραφή, διασφαλίζοντας έτσι την ταυτοποίηση του φορέα και την ακεραιότητα των δεδομένων που εμφανίζονται στο έγγραφο. Επίσης, αν τεθεί σε παραγωγικό περιβάλλον θα πρέπει να ληφθεί υπόψη η κλιμάκωση του μηχανισμού, προκειμένου να μην υπάρχουν προβλήματα διαθεσιμότητας. Τέλος, θα πρέπει να αποθηκεύονται κρυπτογραφημένα τα συνθηματικά των χρηστών του συστήματος σε μία βάση δεδομένων.

ΒΙΒΛΙΟΓΡΑΦΙΚΕΣ ΑΝΑΦΟΡΕΣ

- [1] Rolf Sint, Sebastian Schaffert, Stephanie Stroka and Roland Ferst. "Combining Unstructured, Fully Structured and Semi-Structured Information in Semantic Wikis," in *4th Semantic Wiki Workshop (SemWiki 2009) at the 6th European Semantic Web Conference (ESWC 2009)*, Hersonissos, Greece, 2009, pp.73-87. Accessed: Dec. 15, 2022. [Online]. Available: https://www.researchgate.net/profile/Max-Voelkel/publication/230824995_4th_Semantic_Wiki_Workshop_SemWiki_2009_at_the_6th_European_Semantic_Web_Conference_ESWC_2009_Hersonissos_Greece_June_1st_2009_Proceedings/links/59116752aca272ec9a4e7182/4th-Semantic-Wiki-Workshop-SemWiki-2009-at-the-6th-European-Semantic-Web-Conference-ESWC-2009-Hersonissos-Greece-June-1st-2009-Proceedings.pdf#page=83.
- [2] Andrea L. Nylund, "Tracing the BI Family Tree," Knowledge Management, July 1999. Accessed: Dec. 15, 2022. [Online]. Available: https://www.escholar.com/wp-content/uploads/2019/12/dw_family_tree_1999.pdf.
- [3] Li Zeng, Lida Xu, Zhongzhi Shi, Maoguang Wang, and Wenjuan Wu, "Techniques, Process, and Enterprise Solutions of Business Intelligence," *2006 IEEE Conference on Systems, Man, and Cybernetics October 8-11, 2006, Taipei, Taiwan*. Accessed: Dec. 15, 2022. [Online]. Available: https://www.researchgate.net/profile/Zhongzhi-Shi/publication/4262775_Techniques_Process_and_Enterprise_Solutions_of_Business_Intelligence/links/554c23f00cf29752ee7edaae/Techinques-Process-and-Enterprise-Solutions-of-Business-Intelligence.pdf.
- [4] Robert Stackowiak, Joseph Rayman and Rick Greenwald, *Oracle Data Warehousing and Business Intelligence Solutions*, Wiley.
- [5] Solomon Negash, Paul Gray, Frada Burstein and Clyde Holsapple, "Business Intelligence," in *Communications of the Association for Information Systems*, vol.13, pp.177-195, 2004. Accessed: Dec. 15, 2022. [Online]. Available: https://www.researchgate.net/publication/226752489_Business_Intelligence.
- [6] Zack Jourdan, R. Kelly Rainer and Thomas E. Marshall. "Business Intelligence: An analysis of the Literature," in *Information Systems Management*, vol.25, pp. 121-133, doi: 10.1080/10580530801941512.
- [7] Bernard Wieder, Maria-Luise Ossimitz. "The Impact of Business Intelligence on the quality of decision making – a mediation model," in *Procedia Computer Science*, vol. 64, pp. 1163-1171, 2015, doi: 10.1016/j.procs.2015.08.599.
- [8] Vesna Bosilj Vuksic, Mirjana Pejic Bacha and Ales Popovic. "Supporting performance management with business process management and business intelligence: A case analysis of integration and orchestration," in *International Journal of Information Management*, vol. 33, pp. 613-619, 2013, doi: 10.1016/j.ijinfomgt.2013.03.008.
- [9] Quoc Duy Vo et al. "Next Generation Business Intelligence and Analytics: A Survey," Dept. of Computer Sciences, Georgia Southern University, Georgia, USA, 2017, doi: 10.48550/ARXIV.1704.03402.
- [10] Mercado Eletronico. "Business Intelligence (BI): artificial intelligence in corporate procurement", blog.mercadoe.com. <https://blog.mercadoe.com/en/business-intelligence-bi-compras/> (accessed Jan. 20, 2023).
- [11] Lyle Del Vecchio. "Business Intelligence In Procurement". planergy.com. <https://planergy.com/blog/business-intelligence-procurement/> (accessed Jan. 20, 2023).
- [12] European Commision. "European Single Procurement Document (ESPD)". ec.europa.eu. https://ec.europa.eu/isa2/solutions/european-single-procurement-document-espdoc_en/ (accessed Jan. 20, 2023).
- [13] Jayanthi Ranjan, "Business Intelligence: Concepts, Components, Techniques and Benefits," in *Journal of Theoretical and Applied Information Technology*, vol. 9, pp. 60-70, 2009.

- [14] Hui-Chu Chang and Hsiao-Ping Tsai, "Group RFM analysis as a novel framework to discover better customer consumption behavior", in *Expert Systems with Applications*, vol. 38, pp. 14499-14513, 2011, doi: 10.1016/j.eswa.2011.05.034.
- [15] DIGIWHIST, "About Opentender". opentender.eu. <https://opentender.eu/gr/about/about-opentender> (accessed Dec 15, 2022).
- [16] Diana White, Joshua Eastes, Negar Janani and River Bond. "Playfair's Introduction of Time Series to Represent Data," University of Colorado, 2020. Accessed: Dec. 15, 2022. [Online]. Available: https://digitalcommons.ursinus.edu/cgi/viewcontent.cgi?article=1007&context=triumphs_statistics.
- [17] Douglas Montgomery, Cheryl Jennings and Murat Kulahci. *Introduction to Time Series Analysis and Forecasting 2nd edition*, Canada:Wiley, 2015.
- [18] Anomaly, "Extracting Seasonality and Trend from Data: Decomposition Using R". anomaly.io. <https://anomaly.io/seasonal-trend-decomposition-in-r/index.html> (accessed Dec. 3, 2022).
- [19] Chris Parsons, "What Is a Machine Learning Model?". blogs.nvidia.com. <https://blogs.nvidia.com/blog/2021/08/16/what-is-a-machine-learning-model/#:~:text=A%20machine%20learning%20model%20is,mathematical%20engines%20of%20artificial%20intelligence>. (accessed Jan. 20, 2023).
- [20] Jake Frankenfield, "Predictive Modeling Definition". investopedia.com. <https://www.investopedia.com/terms/p/predictive-modeling.asp> (accessed Jan. 20, 2023).
- [21] Ethem Alpaydin, *Introduction to Machine Learning 3rd edition*, Cambridge, MA, USA: MIT Press, 2014.
- [22] Rob J Hyndman and George Athanasopoulos. *Forecasting: principles and practice 3rd edition*, OTexts: Melbourne, Australia, 2021. Accessed: Dec. 15, 2022. [Online]. Available: <https://otexts.com/fpp3/>.
- [23] Sean J. Taylor and Benjamin Letham. "Forecasting at Scale," PeerJ Preprints, 2017, doi: 10.7287/peerj.preprints.3190v2.[24] Publications Office of the European Union, "Business Handbook v3.0.1". docs.ted.europa.eu. https://docs.ted.europa.eu/ESPD-EDM/latest/xml_business_handbook.html (accessed Dec. 17, 2022).
- [25] Erich Gamma, Richard Helm, Ralph Johnson and John Vlissides. *Design Patterns: Elements of Reusable Object Oriented Software 1st edition*, Addison-Wesley, 1995.
- [26] Dylan D. Scott, *Kafka in Action*, Manning Publications, 2020.
- [27] HEVO, "Kafka Clusters Architecture 101: A Comprehensive Guide". hevodata.com. <https://hevodata.com/learn/kafka-clusters/> (accessed Dec. 19, 2022).
- [28] Weiping Qu, "On-Demand ETL for Real-Time Analytics," Ph.D. dissertation, Dept. Computer Science, Technische Universität Kaiserslautern, Kaiserslautern, Germany, 2020. [Online]. Available: https://kluedo.ub.rptu.de/frontdoor/deliver/index/docId/6252/file/dissertation_weiping_qu.pdf (accessed Dec. 23, 2022).
- [29] Prophet. "Diagnostics". facebook.github.io. <https://facebook.github.io/prophet/docs/diagnostics.html> (accessed Dec. 21, 2022).
- [30] Prophet. "Trend Changepoints". facebook.github.io. https://facebook.github.io/prophet/docs/trend_changepoints.html (accessed Dec. 21, 2022).
- [31] Java. Oracle. Accessed: Dec. 23, 2022. [Online]. Available: <https://www.oracle.com/java/>.

- [32] Redwood Shores. “Oracle Buys Sun”. oracle.com.
<https://www.oracle.com/corporate/pressrelease/oracle-buys-sun-042009.html> (accessed Dec. 23, 2022).
- [33] Python. Python Software Foundation. Accessed: Dec. 23, 2022. [Online]. Available:
<https://www.python.org/>.
- [34] TypeScript. Microsoft. Accessed: Dec. 23, 2022. [Online]. Available: <https://www.typescriptlang.org/>.
- [35] Angular. Google. Accessed: Dec. 23, 2022. [Online]. Available: <https://angular.io/>.
- [36] Apache ECharts. Apache Software Foundation. Accessed: Dec. 23, 2022. [Online]. Available:
<https://echarts.apache.org/en/index.html>.
- [37] Flask. Armin Ronacher. Accessed: Dec. 25, 2022. [Online]. Available:
<https://flask.palletsprojects.com/en/2.2.x/>.
- [38] Pandas. Wes McKinney. Accessed: Dec. 25, 2022. [Online]. Available: <https://pandas.pydata.org/>.
- [39] Prophet. Facebook. Accessed: Dec. 25, 2022. [Online]. Available: <https://facebook.github.io/prophet/>.
- [40] Apache FreeMarker. Apache Software Foundation. Accessed: Dec. 25, 2022. [Online]. Available:
<https://freemarker.apache.org/>.
- [41] ksqlDB. Confluent Inc. Accessed: Dec. 25, 2022. [Online]. Available: <https://docs.ksqldb.io/en/latest/>.
- [42] MongoDB. MongoDB Inc. Accessed: Dec. 25, 2022. [Online]. Available:
<https://www.mongodb.com/home>.
- [43] Redis Stack. Redis. Accessed: Dec. 25, 2022. [Online]. Available: <https://redis.io/>.