



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»



ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**ΠΡΟΒΛΕΨΗ ΕΚΤΑΚΤΩΝ ΚΑΙΡΙΚΩΝ ΦΑΙΝΟΜΕΝΩΝ ΜΕ ΧΡΗΣΗ
ΤΑΞΙΝΟΜΗΤΩΝ ΓΙΑ ΤΗΝ ΕΓΚΑΙΡΗ ΠΡΟΕΙΔΟΠΟΙΗΣΗ ΤΩΝ ΠΟΛΙΤΩΝ ΣΕ
ΜΕΓΑΛΕΣ ΠΟΛΕΙΣ ΤΗΣ ΕΛΛΑΔΑΣ**

ΔΡΗΣ ΑΓΓΕΛΟΣ
ΣΑΣΣΑΝΗ ΣΟΦΙΑ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ
ΑΣΗΜΑΚΟΠΟΥΛΟΣ ΒΑΣΙΛΕΙΟΣ

ΦΕΒΡΟΥΑΡΙΟΣ 2021



ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ ΚΑΙ Μ/Υ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΠΕΙΡΑΙΩΣ
ΣΧΟΛΗ ΝΑΥΤΙΛΙΑΣ ΚΑΙ ΒΙΟΜΗΧΑΝΙΑΣ
ΤΜΗΜΑΤΟΣ ΒΙΟΜΗΧΑΝΙΚΗΣ ΔΙΟΙΚΗΣΗΣ & ΤΕΧΝΟΛΟΓΙΑΣ
ΔΙΑΠΑΝΕΠΙΣΤΗΜΙΑΚΟ ΠΡΟΓΡΑΜΜΑ ΜΕΤΑΠΤΥΧΙΑΚΩΝ ΣΠΟΥΔΩΝ
«ΤΕΧΝΟ-ΟΙΚΟΝΟΜΙΚΑ ΣΥΣΤΗΜΑΤΑ»



ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

**ΠΡΟΒΛΕΨΗ ΕΚΤΑΚΤΩΝ ΚΑΙΡΙΚΩΝ ΦΑΙΝΟΜΕΝΩΝ ΜΕ ΧΡΗΣΗ
ΤΑΞΙΝΟΜΗΤΩΝ ΓΙΑ ΤΗΝ ΕΓΚΑΙΡΗ ΠΡΟΕΙΔΟΠΟΙΗΣΗ ΤΩΝ ΠΟΛΙΤΩΝ ΣΕ
ΜΕΓΑΛΕΣ ΠΟΛΕΙΣ ΤΗΣ ΕΛΛΑΔΑΣ**

ΔΡΗΣ ΑΓΓΕΛΟΣ
ΣΑΣΣΑΝΗ ΣΟΦΙΑ

ΕΠΙΒΛΕΠΩΝ ΚΑΘΗΓΗΤΗΣ
ΑΣΗΜΑΚΟΠΟΥΛΟΣ ΒΑΣΙΛΕΙΟΣ

Εγκρίθηκε από την τριμελή εξεταστική επιτροπή την

2021

.....
(Υπογραφή)

Ασημακόπουλος Βασίλειος
Καθηγητής Ε.Μ.Π.

.....
(Υπογραφή)

Ψαρράς Ιωάννης
Καθηγητής Ε.Μ.Π.

.....
(Υπογραφή)

Δούκας Χρυσόστομος
Επίκουρος Καθηγητής Ε.Μ.Π.



Περίληψη

Τα τελευταία χρόνια έχει παρατηρηθεί, ανά τακτά χρονικά διαστήματα και ως απόρροια της κλιματικής αλλαγής, η όλο και πιο συχνή εμφάνιση έντονων καιρικών φαινομένων, τόσο στον ευρωπαϊκό όσο και στον ελλαδικό χώρο, γεγονός που επισύρει δυσμενείς κοινωνικο - οικονομικές συνέπειες για τις περιοχές που υφίστανται τα φαινόμενα. Ειδικότερα, στην Ελλάδα τόσο στην πρωτεύουσα όσο και στην περιφέρεια, η αδυναμία της πολιτείας να παρουσιάσει τον κατάλληλο βαθμό ετοιμότητας μπροστά σε τέτοια συμβάντα, και τα συνοδά γεγονότα, όπως οι απώλειες ανθρώπινων ζωών και περιουσιών, οι καταστροφές οδικών δικτύων κ.λπ., αποτέλεσαν το κίνητρο για τη συγκεκριμένη μελέτη. Στόχος της παρούσας εργασίας υπήρξε η πρόβλεψη μελλοντικής εμφάνισης έντονων καιρικών φαινομένων σε μεγάλες πόλεις της Ελλάδας, με τη χρήση κατάλληλων αλγορίθμων μηχανικής μάθησης. Χρησιμοποιήθηκαν τρεις δυαδικοί ταξινομητές, Logistic Regression, Random Forests και $k - NN$, οι οποίοι εκπαιδεύτηκαν να αναγνωρίζουν την ύπαρξη καιρικού φαινομένου ή ηλιοφάνειας, ανάλογα με τα μετεωρολογικά δεδομένα που τους δόθηκαν. Τα μετεωρολογικά δεδομένα, που χρησιμοποιήθηκαν για την εκπαίδευση των ταξινομητών και τα οποία συνδυάστηκαν με καταγραφές έντονων καιρικών φαινομένων, όπως οι θυελλώδεις άνεμοι, οι χαλαζοπτώσεις, οι καταιγίδες κ.λπ., αφορούσαν σε τιμές ύψους υετού, θερμοκρασίας και ανέμου οι οποίες ελήφθησαν για διάφορες πόλεις της Ελλάδας, σε βάθος εικοσαετίας (2000 - 2019), από την Εθνική Μετεωρολογική Υπηρεσία. Βάσει των δεικτών απόδοσης για τις προβλέψεις των τριών αλγορίθμων, ο αλγόριθμος Logistic Regression ήταν εκείνος που απέδωσε τις καλύτερες προβλέψεις εμφάνισης φαινομένου για όλες τις πόλεις και κυρίως για τα δυτικά προάστια της Αττικής και την Κοζάνη. Ο Logistic Regression, επίσης, έδειξε να επηρεάζεται λιγότερο από ανακριβείς μετεωρολογικές προβλέψεις, στο πλαίσιο της ανάλυσης ευαισθησίας που εκπονήθηκε και κατά την οποία μεταβλήθηκαν οι τιμές των μετεωρολογικών παραμέτρων διαδοχικά και όλες μαζί. Τα αποτελέσματα που ελήφθησαν από τη συγκεκριμένη μελέτη δείχνουν ότι οι δύο από τους τρεις αλγόριθμους που εξετάστηκαν είναι ικανοί, εάν τους παρασχεθούν μελλοντικές μετεωρολογικές προβλέψεις, να δώσουν σωστή πρόβλεψη για την ύπαρξη κάποιου επικείμενου επικίνδυνου φαινομένου καθιστώντας τους σημαντικό κομμάτι του πυρήνα ενός Συστήματος Έγκαιρης Προειδοποίησης, για την έγκαιρη προειδοποίηση των πολιτών με στόχο την προστασία τους από επικίνδυνα φαινόμενα. Η περαιτέρω διερεύνηση των περιορισμών που αφορούν στα σύνολα δεδομένων εκπαίδευσης των ταξινομητών θα μπορούσε να ισχυροποιήσει ακόμη περισσότερο την απόδοση και την αξιοπιστία τους.

Λέξεις Κλειδιά: μετεωρολογικά και υδρολογικά φαινόμενα, συστήματα έγκαιρης προειδοποίησης, μηχανική μάθηση, αλγόριθμοι ταξινόμησης, ανάλυση ευαισθησίας



Abstract

Nowadays, the frequent appearance of severe weather phenomena, has been observed, as a result of climate change. Their increasing occurrence has caused socio - economic consequences globally, including our country. The lack of adequate preparation of citizens and infrastructures both in the capital and the provinces of Greece, and the accompanied events, such as deaths of people, losses of personal properties, disasters of the road networks etc. was the motivation of this study. Hence, the aim of this thesis was the forecasting of future severe phenomena in major cities, in Greece, using appropriate machine learning algorithms. Three binary classifiers were used, Logistic Regression, Random Forests, and k – NN. They were trained to recognize the occurrence of either weather phenomenon or sunshine, depending on the given meteorological data. The aforementioned meteorological data (i.e. precipitation height, temperature, and wind data) were obtained for different greek cities, from the Hellenic National Meteorological Service, over a period of twenty years (2000 - 2019) and were combined with the recordings of severe weather phenomena, such as stormy winds, hailstorms, storms, etc. According to the evaluation of the performance of the binary classifiers applied, it was revealed that the Logistic Regression algorithm was the most reliable among the three for all cities and especially for the western suburbs of Attica and Kozani. A sensitivity analysis was also conducted by changing the values of some meteorological parameters, firstly consecutively and then simultaneously. According to this analysis, Logistic Regression was also the less affected algorithm by the inaccurate meteorological forecasts. The results obtained from this study demonstrate that two of the three algorithms examined, are able to give reliable predictions for unexpected phenomena; in the case of given future meteorological forecasts. This ability of the classifiers could be useful in the Early Warning Systems, for the timely and accurate warning and for the proper preparation of communities; against severe phenomena. Further investigation of the limitations of training datasets could further enhance the performance and reliability of the classifiers.

Keywords: meteorological and hydrological phenomena, early warning systems, machine learning, classification algorithms, sensitivity analysis



Πρόλογος

Στο σημείο αυτό, θεωρούμε απαραίτητο να ευχαριστήσουμε τους ανθρώπους που συνέδραμαν στην εκπόνηση αυτής της εργασίας.

Ευχαριστούμε τον καθηγητή του Τομέα Ηλεκτρικών Βιομηχανικών Διατάξεων και Συστημάτων Αποφάσεων και επιβλέποντα καθηγητή της διπλωματικής εργασίας μας, κ. Ασημακόπουλο Βασίλειο καθώς και τον ερευνητικό συνεργάτη της Μονάδας Προβλέψεων και Στρατηγικής, και επίσης επιβλέποντα της εργασίας μας, κ. Σπηλιώτη Ευάγγελο, για την καθοδήγησή του στην υλοποίηση της σκέψης μας.

Ξεχωριστές ευχαριστίες θα θέλαμε να απευθύνουμε στον ερευνητή κ. Σασσάνη Βασίλειο για τις συμβουλές και τη βοήθειά του σε θέματα προγραμματισμού, που συμπεριελήφθησαν στην εργασία μας.

Τέλος, απευθύνουμε ευχαριστίες στη Διεύθυνση Κλιματολογίας - Εφαρμογών της ΕΜΥ για τα στοιχεία που μας παρείχε.



Πίνακας Περιεχομένων

Περιεχόμενα

Περίληψη.....	i
Abstract.....	ii
Πρόλογος.....	iii
Πίνακας Περιεχομένων.....	iv
1 Εισαγωγή.....	1
1.1 Φυσικές Καταστροφές.....	1
1.1.1 Κίνδυνοι & Επιπτώσεις.....	1
1.1.2 Αποτίμηση των επιπτώσεων των φυσικών καταστροφών.....	2
1.1.3 Το κλίμα ως συνάρτηση των φυσικών καταστροφών.....	4
1.1.4 Οι φυσικές καταστροφές τα τελευταία χρόνια στον Ελλαδικό Χώρο.....	4
1.2 Συστήματα Έγκαιρης Προειδοποίησης.....	6
1.2.1 Αναδρομή στη δημιουργία των EWS.....	6
1.2.2 Ορισμός & Χαρακτηριστικά των EWS.....	9
1.3 Τα EWS για μετεωρολογικές προβλέψεις και προειδοποιήσεις στην Ελλάδα.....	12
1.3.1 Το σύστημα Meteoalarm.....	12
1.3.2 Το σύστημα ΖΕΥΣ.....	13
1.3.3 Το σύστημα ΠΟΣΕΙΔΩΝ.....	14
1.3.4 Το σύστημα EFFIS.....	14
1.3.5 Το σύστημα SEVIRI.....	15
1.3.6 Το σύστημα Disaster Management Framework.....	15
1.3.7 Η περίπτωση της Μάνδρας Αττικής.....	15
1.4 Κόστη και πλεονεκτήματα των EWS παγκοσμίως.....	15
2 Θεωρητικά Στοιχεία.....	24
2.1 Ορισμοί.....	24
2.1.1 Ταξινόμηση.....	24
2.1.2 Η έννοια του αλγορίθμου.....	24
2.1.3 Μηχανική Μάθηση.....	24
2.1.4 Επιτηρούμενη Μάθηση.....	25
2.2 Προγνωστική Μοντελοποίηση (Predictive Modelling).....	25
2.2.1 Προγνωστικά Μοντέλα.....	26
2.3 Προγνωστική Μοντελοποίηση Ταξινόμησης (Classification Predictive Modeling).....	27
2.3.1 Αξιολόγηση δυαδικών ταξινομητών (binary classifiers).....	27
2.4 Ταξινομητές.....	30
2.4.1 Λογιστική Παλινδρόμηση.....	30
2.4.2 Δένδρα Απόφασης.....	32
2.4.3 k - Nearest Neighbors.....	35
2.4.4 Τεχνική σύνθεσης υπερ - δειγματοληψίας μειονοτήτων SMOTE.....	39
2.5 Προ - επεξεργασία Δεδομένων.....	41
3 Μεθοδολογία.....	43
3.1 Πειραματικά Δεδομένα.....	43
3.2 Προ - επεξεργασία των δεδομένων.....	46
3.2.1 Μη ισορροπημένα σύνολα δεδομένων.....	46
3.2.2 Τεχνική Σύνθεσης Πειραματικών Συνόλων Δεδομένων με λίγες καταγραφές.....	46
3.3 Σύνολα Εκπαίδευσης & Σύνολα Δοκιμής.....	48



4	Αποτελέσματα	49
4.1	Απόδοση Ταξινομητών ανά Πόλη	49
4.1.1	Δυτική Αττική	49
4.1.2	Κοζάνη	50
4.1.3	Αλεξανδρούπολη	51
4.1.4	Λάρισα	52
4.1.5	Ιωάννινα	53
4.1.6	Ηράκλειο Κρήτης	54
4.1.7	Τρίπολη	55
4.1.8	Καλαμάτα	56
4.1.9	Θεσσαλονίκη	57
4.2	Συγκριτική Απεικόνιση των Ταξινομητών	58
4.3	Ανάλυση Ευαισθησίας	62
4.3.1	Ανάλυση Ευαισθησίας στη μεταβολή του Ύψους Υετού	62
4.3.2	Ανάλυση Ευαισθησίας στη μεταβολή της Μέγιστης και της Ελάχιστης Θερμοκρασίας	62
4.3.3	Ανάλυση Ευαισθησίας στη μεταβολή της ταχύτητας του ανέμου	62
4.3.4	Ανάλυση Ευαισθησίας στη ταυτόχρονη μεταβολή όλων των παραμέτρων	63
5	Συμπεράσματα- Συζήτηση	67
5.1	Σύγκριση των ταξινομητών	67
5.2	Η περίπτωση της Θεσσαλονίκης	68
5.3	Ανάλυση Ευαισθησίας	68
5.4	Συμπεράσματα	69
5.5	Περιορισμοί των μοντέλων πρόβλεψης	69
5.5.1	Η ανάγκη για μαζικά εκπαιδευτικά σύνολα δεδομένων	69
5.5.2	Σωστή ταξινόμηση δεδομένων	69
5.5.3	Εφαρμογή εκπαιδεύσεων σε διαφορετικές περιπτώσεις	70
5.6	Προγνωστικά μοντέλα στο μέλλον	70
6	Βιβλιογραφία	71



1 Εισαγωγή

Σε ένα γενικό πλαίσιο κλιματικών μεταβολών, η εμφάνιση των φυσικών φαινομένων, και οι καταστροφές που σε πολλές περιπτώσεις επιφέρουν, είναι ζημιογόνες για τις κοινωνίες που τις υφίστανται, ενώ σε κάθε περίπτωση συνοδεύονται από σημαντικές κοινωνικό-οικονομικές επιπτώσεις. Εάν αναλογιστεί κανείς τις απώλειες ανθρώπινων ζωών – που σε τέτοιες περιπτώσεις είναι μαζικές-, τις σημαντικές καταστροφές σε οικήματα αλλά και σε δημόσιες υποδομές, μπορεί να αντιληφθεί ότι υπάρχει επιτακτική ανάγκη να υπάρχει σωστή πρόληψη. Στην πρόληψη, που αποτελεί ένας από τους σημαντικότερους στόχους στον τομέα διαχείρισης φυσικών καταστροφών παγκοσμίως, εντάσσεται η ανάγκη άμεσης έκδοσης σχετικής προειδοποίησης, η οποία θα δώσει τη δυνατότητα και το χρόνο στους ανθρώπους που θα πληγούν, να προετοιμαστούν κατάλληλα ώστε να έχουν τις λιγότερες δυνατές απώλειες. Της προειδοποίησης, ωστόσο, προηγείται η σωστή και ακριβής πρόβλεψη εμφάνισης έκτακτης συνθήκης, που θα δώσει την πληροφορία και θα σημάνει τον συναγερμό ώστε να ενεργοποιηθούν οι απαραίτητοι μηχανισμοί, οι οποίοι αρχικά θα προειδοποιήσουν και στη συνέχεια θα βρίσκονται σε ετοιμότητα για την εμφάνιση του φαινομένου που έχει προβλεφθεί.

Στόχος, λοιπόν της συγκεκριμένης εργασίας είναι η αναζήτηση ενός αποτελεσματικού τρόπου εντοπισμού έντονων καιρικών φαινομένων, στο άμεσο μέλλον, για κάποιες μεγάλες πόλεις της Ελλάδας, ώστε εάν υπάρχει διαθέσιμη μετεωρολογική πρόβλεψη να μπορεί να γίνει και ο εντοπισμός του επικείμενου φαινομένου. Με τη χρήση αλγόριθμων μηχανικής μάθησης, και τη γνώση από παλιότερες καταγεγραμμένες εμφανίσεις έντονων καιρικών φαινομένων, επιδιώκουμε να εκπαιδύσουμε τρεις δυαδικούς ταξινομητές προκειμένου να είναι σε θέση να δίνουν προβλέψεις εντοπισμού επικείμενων έντονων φαινομένων. Επιπλέον, για να εξετάσουμε πόσο επηρεάζονται οι ταξινομητές που επιλέξαμε από ανακριβείς προβλέψεις, τροποποιήσαμε τέσσερις μετεωρολογικές παραμέτρους, το ύψος υετού, τη μέγιστη και την ελάχιστη θερμοκρασία καθώς και τους κόμβους ταχύτητας του ανέμου.

Απώτερος στόχος αυτής της διαδικασίας είναι να εξετασθεί το κατά πόσο οι ταξινομητές που χρησιμοποιήθηκαν μπορούν να δώσουν ακριβείς προβλέψεις εμφάνισης φαινομένων προκειμένου μελλοντικά να αποτελέσουν τμήμα ενός Συστήματος Έγκαιρης Προειδοποίησης, βοηθώντας έτσι στην ανάπτυξη και τη βελτίωση της ποιότητας του καθώς και στη διεύρυνση της χρήσης του.

1.1 Φυσικές Καταστροφές

1.1.1 Κίνδυνοι & Επιπτώσεις

Έχει παρατηρηθεί παγκοσμίως ότι ο αριθμός και η σοβαρότητα καταστροφών, ως επακόλουθο φυσικών φαινομένων, αυξάνει συνεχώς. Μία καταστροφή, που προκαλείται από ένα φυσικό κίνδυνο, μπορεί να οριστεί ως η σοβαρή διαταραχή της λειτουργίας μίας κοινότητας ή μίας κοινωνίας που μπορεί να προκαλέσει εκτεταμένες ανθρώπινες, υλικές, οικονομικές ή/και περιβαλλοντικές απώλειες, οι οποίες υπερβαίνουν την ικανότητα των



πληγέντων να τις αντιμετωπίσουν, κάνοντας χρήση των ίδιων πόρων [1]. Η ταυτόχρονη συνύπαρξη τριών στοιχείων, δηλ. του συμβάντος κινδύνου, των συνθηκών τρωτότητας απέναντι στον κίνδυνο και της ανεπάρκειας της ικανότητας ή των μέτρων αντιμετώπισης του κινδύνου, οδηγούν στην εμφάνιση μίας καταστροφής λόγω φυσικών παραγόντων [2]. Οι φυσικοί κίνδυνοι έχουν συνήθως τα εξής χαρακτηριστικά: εκδηλώνονται σε σύντομο χρονικό διάστημα, δίνουν άμεσα και ακούσια αποτελέσματα, αντιπροσωπεύουν πιθανές καταστροφές και δημιουργούν καταστάσεις έκτακτης ανάγκης [3].

Οι επιπτώσεις των καταστροφών αυξάνονται συνεχώς ως αποτέλεσμα του συνδυασμού των αλλαγών που υφίσταται το κλίμα, του πολλαπλασιασμού του παγκόσμιου πληθυσμού [4], των μεγαλύτερων συγκεντρώσεων πληθυσμών σε ευάλωτες περιοχές καθώς και της τροποποίησης και υποβάθμισης του φυσικού περιβάλλοντος, μέσω δόμησης σε κοίτες ποταμών και λεκανών απορροής, παράκτιας εκμετάλλευσης, καταστροφής υγροτόπων, τροποποίηση της ροής ποταμών, αποψίλωσης των δασών, διάβρωσης του εδάφους και μείωσης της γονιμότητας του εδάφους. Η ευπάθεια στους κινδύνους επιδεινώνεται από το χαμηλό οικονομικό επίπεδο, τις ασθένειες, τις κατά τόπους διαμάχες και τις μεταφορές πληθυσμών.

Είναι γεγονός ότι οι φυσικές καταστροφές επιβαρύνουν, κυρίως, το τμήμα του παγκόσμιου πληθυσμού που βρίσκεται σε πιο δυσμενή κοινωνικοοικονομική κατάσταση, καθώς το 69% όλων των ανθρώπων που σκοτώθηκαν από καταστροφές μεταξύ του 1996 και του 2015 ταξινομήθηκαν ως χαμηλού ή κατώτερου μεσαίου εισοδήματος [5]. Ειδικότερα, οι φτωχές κοινωνίες που αναπτύσσονται σε ακατάλληλες τοποθεσίες, όπως πλημμυρικές περιοχές, όχθες ποταμών, απότομες πλαγιές, οικισμούς με πρόχειρες κατασκευές/ακατάλληλα υλικά για υποδομές και στέγαση σε συνδυασμό με εφαρμογή αμφίβολων γεωργικών πρακτικών που συχνά δύνανται να δημιουργήσουν ή/και να επαυξήσουν νέες συνθήκες κινδύνου. Ακολούθως, η μειωμένη δυνατότητα που έχει αυτή η μερίδα του πληθυσμού να ανταποκριθεί στην αποκατάσταση των επιπτώσεων που οι καταστροφές προκαλούν, την καθιστά ακόμη πιο ευάλωτη.

Έχει επίσης παρατηρηθεί, ότι οι γυναίκες και τα παιδιά αποτελούν πιο ευάλωτες ομάδες συγκριτικά με τους άνδρες [6], καθώς κατά τη διάρκεια του τσουνάμι του Ινδικού Ωκεανού το 2004, πάνω από το 70% των θυμάτων ήταν γυναίκες, ενώ το 2008, εξαιτίας του κυκλώνα Nargis στη Μιανμάρ, σκοτώθηκαν διπλάσιες γυναίκες από ότι άνδρες. Το 2005, ο τυφώνας Katrina επηρέασε κυρίως γυναίκες και δη αφροαμερικανές, καταδεικνύοντας την αδυναμία τόσο του φύλου όσο και των ασθενέστερων ομάδων που δεν έχουν πρόσβαση σε πληροφόρηση και υγειονομική περίθαλψη.

1.1.2 Αποτίμηση των επιπτώσεων των φυσικών καταστροφών

Οι άμεσες και εμφανείς επιπτώσεις των φυσικών καταστροφών συνοψίζονται σε τραυματισμούς και απώλειες ανθρώπινων ζώων, καταστροφές ιδιωτικών και δημόσιων δομών, καλλιεργειών, φυσικών πόρων κλπ., ενώ έμμεσες συνέπειες αποτελούν η προσωρινή ανεργία, οι διαταραχές/διακοπή στη λειτουργία επιχειρήσεων, οι μειώσεις



εισοδημάτων και μισθών, οι περιβαλλοντικές επιπτώσεις, οι ψυχολογικές συνέπειες κλπ. [3]. Ωστόσο, η συνολική ποσοτικοποίηση τους αποτελεί μία δύσκολη εξίσωση και για το λόγο αυτό παρουσιάζονται αποκλίσεις στις μελέτες που αποτιμούν τις φυσικές καταστροφές, οι οποίες προκύπτουν από τον τρόπο αποτίμησης και από το είδος των απωλειών που προσμετρώνται. Παρόλα αυτά, στη γενική αποτίμηση περιλαμβάνονται σχεδόν πάντα οι ανθρώπινες απώλειες (νεκροί και πληγέντες) καθώς και οι οικονομικές απώλειες σε οικονομικούς όρους.

Ως προς τις ανθρώπινες απώλειες, βάσει στοιχείων του 2005 [3], στις φτωχές και αναπτυσσόμενες χώρες σκοτώθηκαν 30 φορές περισσότεροι άνθρωποι, μέγεθος που αναλογεί σε περίπου 13 περισσότερα άτομα ανά καταστροφή, σε σχέση με τις αναπτυγμένες χώρες. Βασική επίπτωση αυτού είναι η καθυστέρηση στη συνολική πρόοδο των συγκεκριμένων κοινωνιών, καθώς δεν υπάρχουν οι διαθέσιμοι πόροι που θα οδηγήσουν στη γρήγορη ανάκαμψη. Στον αντίποδα, οι οικονομικές απώλειες βαραίνουν περισσότερο τις αναπτυγμένες χώρες, λόγω των επενδύσεων που έχουν γίνει σε βιομηχανικές εγκαταστάσεις, εξοπλισμούς κ.λπ., αν και ο αντίκτυπος είναι μικρός σε επίπεδο συνολικής οικονομίας.

Σύμφωνα με στατιστικά δεδομένα [7], το μεγαλύτερο ποσοστό των καταγεγραμμένων μεγάλων καταστροφών που προκλήθηκαν από φυσικούς κινδύνους, για το χρονικό διάστημα από το 1995 έως το 2015, περιελάμβανε πλημμύρες, καταιγίδες, καύσωνες και ξηρασίες. Ειδικά για τη δεκαετία 1995 – 2004 [8, 9], καταγράφηκαν σχεδόν 6.000 καταστροφές, που άφησαν πίσω περίπου 900.000 νεκρούς, 2,5 δισεκατομμύρια ανθρώπων που επηρεάστηκαν και υλικές απώλειες ύψους 738 δισεκατομμύρια δολαρίων. Τα τελευταία χρόνια έχει καταγραφεί εκθετική αύξηση στο κόστος των άμεσων ζημιών [10], με χαρακτηριστικό παράδειγμα εκείνο των Ηνωμένων Πολιτειών όπου από περίπου 75,5 δισεκατομμύρια δολάρια το 1960 έχει ανέλθει σε κάποια τρισεκατομμύρια δολάρια τη δεκαετία 2007 έως 2019.

Η εκτίμηση των μακροπρόθεσμων τάσεων στις καταστροφές εξαρτάται από το σύνολο των δεδομένων αλλά και από τη χρονική περίοδο που εξετάζεται. Φερ ειπείν, συγκρίνοντας τη δεκαετία 1995–2004 με την προηγούμενη δεκαετία 1985-1994, φαίνεται ότι ο αριθμός των ατόμων που επηρεάστηκε αυξήθηκε κατά 1,5 φορές, η οικονομική ζημιά αυξήθηκε κατά 1,8 φορές και οι συνολικοί θάνατοι διπλασιάστηκαν [7]. Βέβαια, στο διπλασιασμό των θανάτων θα πρέπει να ληφθεί υπόψη ότι στη δεκαετία 1995-2004 συμπεριλαμβάνεται το τσουνάμι της 26ης Δεκεμβρίου 2004, που άφησε πίσω του πολλούς νεκρούς. Πριν από αυτή την ημερομηνία, η τάση στα ποσοστά θανάτου έβαινε συνεχώς μειούμενη, ξεκινώντας από τη δεκαετία του 1950, γεγονός που αποδίδεται στην ανάπτυξη και βελτίωση συστημάτων προειδοποίησης για φυσικές καταστροφές, στην αυξημένη ετοιμότητα και απόκριση, συμπεριλαμβανομένων και συστημάτων επισιτιστικής βοήθειας. Μία πιο γενική εικόνα για την πιο πρόσφατη 30-ετία (1980-2008), δίνεται από τα στοιχεία της Διεθνούς Στρατηγικής του ΟΗΕ για τη μείωση του κινδύνου των καταστροφών σύμφωνα με τα οποία, ο συνολικός απολογισμός είναι 2,1 εκατομμύρια νεκροί, 5,3 δισεκατομμύρια πληγέντες και 1,6 τρισεκατομμύρια \$ Η.Π.Α. οικονομική ζημιά παγκοσμίως. Τέλος, επίσημα στοιχεία της βάσης



EM-DAT, αναφέρουν ότι το 2017, ο συνολικός απολογισμός 335 φυσικών καταστροφών, ήταν 9.697 νεκροί, περισσότεροι από 95,6 εκατομμύρια πληγέντες και συνολικό κόστος 335 δισεκατομμύρια \$ Η.Π.Α [3].

Αξίζει να σημειωθεί ότι, ο μεγαλύτερος αριθμός καταστροφών της 30-ετίας καταγράφεται στην Ασία (3.341 συμβάντα) και, ακολούθως, στην Αμερική (2.101 συμβάντα), την Αφρική (1.699 συμβάντα), την Ευρώπη (1.190 συμβάντα) και την Ωκεανία (380 συμβάντα). Οι υδρολογικές καταστροφές - πλημμύρες (2.887 συμβάντα) και οι μετεωρολογικές καταστροφές - καταιγίδες (2.381 συμβάντα), εντοπίστηκαν σε μεγάλο βαθμό στην Ασία και την Αμερική. Ακολουθούν οι βιολογικές καταστροφές (επιδημίες) (1.039 συμβάντα), με κυριότερη περιοχή εμφάνισης την Αφρική, οι γεωφυσικές καταστροφές (σεισμοί, ηφαίστεια) με κύρια περιοχή εμφάνισης την Ασία και οι κλιματολογικές καταστροφές (ακραίες θερμοκρασίες, ξηρασίες, πυρκαγιές) [3].

1.1.3 Το κλίμα ως συνάρτηση των φυσικών καταστροφών

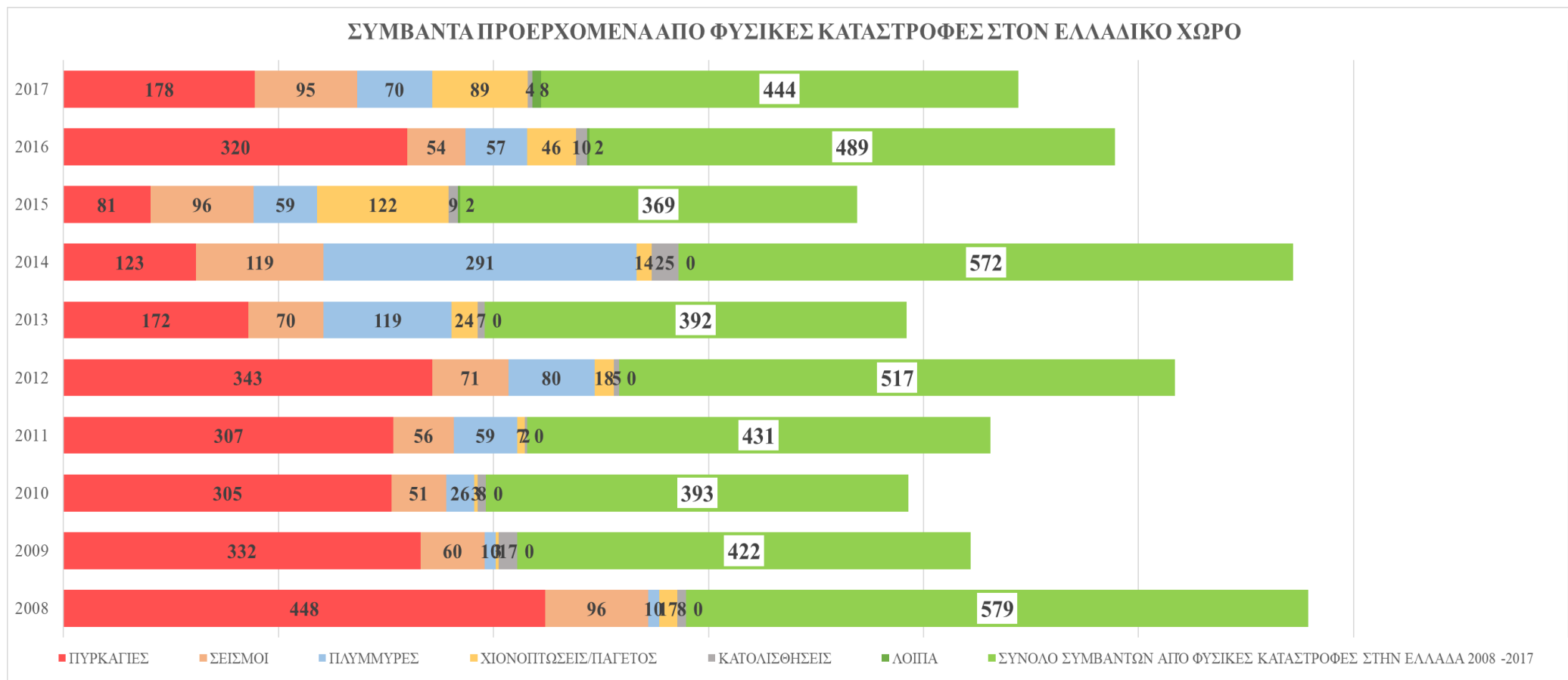
Παρά το γεγονός ότι τα σεισμικά φαινόμενα συνδέονται, παγκοσμίως με μεγαλύτερα ποσοστά θνησιμότητας, λόγω της άναρχης ή ελλιπούς δόμησης που συναντά κανείς σε πολλές περιοχές του κόσμου, έχει αποδειχθεί ότι οι φυσικές καταστροφές έχουν ως επί το πλείστον υδρο - μετεωρολογική προέλευση, λόγω ακραίων τιμών θερμοκρασίας ανέμου, και βροχόπτωσης. Κατά την εικοσαετία 1995 - 2015, οι μετεωρολογικές, κλιματολογικές και υδρολογικές καταστροφές έχουν καταγραφεί, σε ποσοστό 90% ,ως η αιτία πρόκλησης 6.457 καταστροφών, υπερδιπλάσιων από εκείνες τις εικοσαετίας 1976-1995, με 606.000 νεκρούς και 4,1 δισεκατομμύρια πληγέντες [3].

Την εν λόγω εικοσαετία, οι μετεωρολογικές καταστροφές οδήγησαν σε 242.000 ανθρώπινες απώλειες, ενώ οι υδρολογικές καταστροφές αποτέλεσαν το 47% του συνόλου των -σχετικών με το κλίμα- καταστροφών, αφήνοντας 2,3 δισεκατομμύρια πληγέντες. Οι καταστροφές λόγω ακραίων θερμοκρασιών (καύσωνες και παγωνιές) ήταν ιδιαίτερα θανατηφόρες ως προς τον αριθμό των θανάτων ανά περιστατικό (περίπου 405 θάνατοι ανά περιστατικό, κατά μέσο όρο) [3].

Εξετάζοντας τις επόμενες μελλοντικές δεκαετίες, οι προβλέψεις των επιστημόνων μιλούν για αύξηση των ακραίων καιρικών συνθηκών και άμεση συνεπαγωγή της ανοδικής τάσης των εξ αυτών προκαλούμενων καταστροφών.

1.1.4 Οι φυσικές καταστροφές τα τελευταία χρόνια στον Ελλαδικό Χώρο

Στην Ελλάδα, βάσει των στοιχείων που αντλήθηκαν από την ιστοσελίδα της Γενικής Γραμματείας Πολιτικής Προστασίας (Γ.Γ.Π.Π.) [11] και πιο συγκεκριμένα από τους δημοσιευμένους απολογισμούς και τα στατιστικά στοιχεία δράσης από το Κέντρο Επιχειρήσεων της Πολιτικής Προστασίας [12, 13] για τη δεκαετία 2008 - 2017, τα συμβάντα που προέκυψαν και αποδίδονται στα φυσικά φαινόμενα ανήλθαν σε 4608 (2609 πυρκαγιές, 768 σεισμοί, 781 πλημμύρες, 343 χιονοπτώσεις / παγετοί, 95 κατολισθήσεις και 12 λοιπά φαινόμενα, όπως τσουνάμι, ανεμοστρόβιλοι κ.λπ.). Για τα εν λόγω συμβάντα υπάρχουν καταγεγραμμένες ανθρώπινες απώλειες, πληγέντες καθώς και υλικές ζημιές. Στο γράφημα που ακολουθεί ([Εικ. 1](#)) παρουσιάζονται ανά έτος τα φαινόμενα που καταγράφηκαν.



Εικόνα 1.1: Συμβάντα από φυσικές καταστροφές στην Ελλάδα 2008 -2017, βάσει των στοιχείων της Γ.Γ.Π.Π



Είναι χαρακτηριστικό το πρόσφατο συμβάν, εξαιτίας θυελλωδών ανέμων, πολύ ισχυρών καταγίδων και βιβλικών πλημμυρών, που έπληξε τους Δήμους Ελευσίνας, Μάνδρας -Ειδυλλίας, Μεγαρέων, Ασπροπύργου Αττικής στις 14, 15, 16 και 17 Νοεμβρίου 2017. Ειδικά για την περιοχή ανάντη της Μάνδρας, έχει αναφερθεί ότι δέχθηκε περί τα 150mm βροχής, με περιορισμένη χωρική εξάπλωση αναφορικά με την έντασή της, σε διάστημα περίπου 7 ωρών, σχεδόν το 40% της βροχόπτωσης που θα δεχόταν η περιοχή κατά τη διάρκεια ενός έτους. Λόγω της ισχυρής έντασης με την οποία εκδηλώθηκε η βροχή στα ορεινά, σε σχέση με τις πιο χαμηλές – κατοικημένες – περιοχές, δημιουργήθηκε και η σημαντική απορροή των υδάτων προς χαμηλότερα υψόμετρα. Η συγκεκριμένη καταγίδα, εάν και σύντομης διάρκειας, έδωσε υψηλή ένταση κατακρημνισμάτων, χωρικά οριοθετημένη. Η επαναληψιμότητα του συγκεκριμένου φαινομένου υπολογίζεται σε 100 έτη. Ο απολογισμός από το συμβάν αυτό ανέρχεται σε 26 νεκρούς, ωστόσο υπάρχει και ο φόβος αγνοούμενων που δεν αναζητήθηκαν ποτέ, δεκάδες άνθρωποι είτε τραυματίστηκαν, είτε διακομίστηκαν στα Νοσοκομεία με συμπτώματα υποθερμίας, 1110 κτίρια καταστράφηκαν λόγω των εισρρέομενων υδάτων ενώ φερτά υλικά, όπως λάσπη και αδρομερή υλικά, που παρασύρθηκαν από την ορμή των χειμάρρων συντέλεσαν σε ακόμη μεγαλύτερες καταστροφές. Πλήθος αυτοκινήτων παρασύρθηκαν ενώ στο οδικό δίκτυο καταγράφηκαν φαινόμενα υποσκαφής του δρόμου, καταστρέφοντας και τα δίκτυα ύδρευσης και αποχέτευσης στην περιοχή [14].

1.2 Συστήματα Έγκαιρης Προειδοποίησης

Βάσει των προαναφερθέντων παραγόντων, κρίνεται σκόπιμη η επιστράτευση ειδικών εργαλείων για την αντιμετώπιση των δυσμενών συνεπειών που συνοδεύουν τις φυσικές καταστροφές. Σε επίπεδο πρόβλεψης και προειδοποίησης, τα Συστήματα Έγκαιρης Προειδοποίησης (Early Warning Systems – EWS) κρίνονται πολύτιμα για τη μείωση του ρίσκου και των επακόλουθων των απρόβλεπτων φαινομένων. Υπάρχουν συστήματα έγκαιρης προειδοποίησης που σχετίζονται με φυσικούς, γεωφυσικούς και βιολογικούς κινδύνους, βιομηχανικούς κινδύνους, υγειονομικούς κινδύνους, περίπλοκες κοινωνικοπολιτικές καταστάσεις έκτακτης ανάγκης και πολλούς άλλους σχετικούς κινδύνους [2].

1.2.1 Αναδρομή στη δημιουργία των EWS

Η σημασία του EWS για την προσαρμογή της κλιματικής αλλαγής από παγκόσμιο επίπεδο σε τοπικό επίπεδο, καθώς και οι στρατηγικές μείωσης των κινδύνων που προέρχονται από φυσικές καταστροφές άρχισε να αυξάνει τα τελευταία περίπου 20 χρόνια, με την ένταξη των EWS ως θεματολογία στο πλαίσιο διεθνών συνθηκών, διασκέψεων και σχεδίων δράσης.

Καταστροφές ευρείας κλίμακας, όπως το τσουνάμι του Ινδικού Ωκεανού το 2004, επέβαλαν την ανάπτυξη κατάλληλης νομοθεσίας από τη διεθνή κοινότητα και βελτίωση των EWS για την αποφυγή παρόμοιων δραματικών επιπτώσεων στο μέλλον



Η πρώτη διεθνώς αναγνωρισμένη προσπάθεια για την ανάπτυξη EWS πραγματοποιήθηκε το 1998, στο πρώτο σχετικό διεθνές συνέδριο στο Πότσταμ της Γερμανίας, στο οποίο συζητήθηκαν οι μέχρι τότε γνώσεις για τα EWS, από τη διεθνή κοινότητα. Το 2003, στο δεύτερο συνέδριο, στη Βόννη, το ενδιαφέρον εστιάστηκε στο πώς θα μπορούσαν αυτά τα συστήματα να ενσωματωθούν στη δημόσια πολιτική, καθώς παρά το ότι οι προειδοποιήσεις των συστημάτων αυτών στηρίζονταν σε τεχνικές πληροφορίες και στην παρακολούθηση των κινδύνων, έπρεπε να διασφαλιστεί ότι οι προβλέψεις για τους κινδύνους μεταφράζονταν, από τις κοινωνίες, σε προειδοποίηση και ότι τα μέτρα που επρόκειτο να ληφθούν, θα αποτελούσαν απόκριση σε τέτοιες προειδοποιήσεις. Ο παραπάνω προβληματισμός υπήρξε θεμελιώδης για τη μετέπειτα πρόοδο των EWS καθώς ενδυνάμωσε τις τοπικές κυβερνήσεις, τους τοπικούς θεσμούς και κοινωνίες να συμμετάσχουν σε ολόκληρη τη διαδικασία χάραξης πολιτικής, στην προώθηση της ευαισθητοποίησης και της ετοιμότητας αντιμετώπισης απρόσμενων καταστάσεων, και ουσιαστικά σηματοδότησε τη μετάβαση των EWS από ένα απλό τεχνολογικό προϊόν σε κοινωνικοπολιτικά ενσωματωμένη διαδικασία. Το τρίτο διεθνές συνέδριο έγκαιρης προειδοποίησης, που πραγματοποιήθηκε το 2006 με σύνθημα «από την ιδέα έως τη δράση», ήταν σαφώς προσανατολισμένο στον τομέα της πρόληψης και στόχο είχε να τεθούν πλέον σε εφαρμογή οι γνώσεις που είχαν, έως τότε, αποκομισθεί, ωστόσο εντοπίστηκαν πολλά κενά μεταξύ των ιδεών υλοποίησης και της πρακτικής / εφικτής εφαρμογής του EWS.

Καθώς αυτά τα συνέδρια έγκαιρης προειδοποίησης αύξησαν το διεθνές ενδιαφέρον για τα EWS, η προώθησή τους ενισχύθηκε σταδιακά με τη δημιουργία ενός συντονιστικού οργάνου, της Διεθνούς Στρατηγικής των Ηνωμένων Εθνών για τη μείωση των καταστροφών (United Nations International Strategy on Disaster Reduction - UNISDR), το 2000. Η δημιουργία μίας ειδικής ομάδας εργασίας για τη μείωση του κινδύνου καταστροφών υπογράμμισε τη σημασία του θέματος και έδωσε το έναυσμα για διάλογο σχετικά με την αύξηση της κοινωνικής ανθεκτικότητας. Η άμεση δράση της επικεντρώθηκε σε κοινωνικοοικονομικούς και ανθρωπιστικούς τομείς ενώ μία από τις πιο αξιοσημείωτες συνεισφορές της UNISDR στα EWS ήταν η ανάπτυξη της πλατφόρμας (Platform for the Promotion of Early Warning - PPEW) για την προώθηση έγκαιρης προειδοποίησης. Η PPEW ξεκίνησε να λειτουργεί το 2004 με στόχο τη συμβολή στην ανάπτυξη συστημάτων έγκαιρης προειδοποίησης και ετοιμότητας και η λειτουργία της αφορά κυρίως ενίσχυση της βελτίωσης των EWS, συλλογή και διάδοση πληροφοριών σχετικά με βέλτιστες πρακτικές, ενθάρρυνση της συνεργασίας μεταξύ φορέων έγκαιρης προειδοποίησης και ανάπτυξη νέων τρόπων βελτίωσης του EWS.

Η ανάπτυξη των EWS ενισχύθηκε και από το «Πλαίσιο Δράσης Hyogo 2005 – 2015» σχετικά με την οικοδόμηση της ανθεκτικότητας των εθνών και κοινοτήτων σε καταστροφές. Πιο συγκεκριμένα, το Πλαίσιο Δράσης ενθάρρυνε την ανάπτυξη εκτιμήσεων κινδύνου και χαρτών, την επεξεργασία και διάδοση δεικτών πολλαπλών κινδύνων και παραμέτρων ευπάθειας και την παραγωγή δεδομένων και στατιστικών πληροφοριών απωλειών. Επίσης προτάθηκαν δράσεις σχετικά με τη μοντελοποίηση του κλίματος. Εν καιρώ, τα EWS συζητήθηκαν ως εργαλείο που θα μπορούσε να αποτελέσει μέρος μίας ευρύτερης στρατηγικής μείωσης του κινδύνου καταστροφών.



Από τις εκθέσεις της Διακυβερνητικής Ομάδας για την κλιματική αλλαγή (Inter governmental Panel on Climate Change - IPCC), το 2007 και έπειτα, κατέστη σαφέστερος ο ρόλος που μπορεί να διαδραματίσει η κλιματική αλλαγή στην ενίσχυση της σοβαρότητας και της συχνότητας των φυσικών κινδύνων – και ως εκ τούτου και των καταστροφών, καθώς αναδείχθηκε η υψηλή συσχέτιση της με την αύξηση της θερμοκρασίας της γης κατά 1°C. Υπερτονίζοντας τη συσχέτιση μεταξύ κλιματικής αλλαγής και καταστροφών, οι εκθέσεις της IPCC συνέβαλαν θετικά στην επέκταση της εφαρμογής των EWS ως τρόπου στήριξης των παρεμβάσεων προσαρμογής στην κλιματική αλλαγή: μέσω προειδοποίησης για τυχόν δυνατές επιπτώσεις, οικοδόμησης τεχνικών ικανοτήτων για μοντελοποίηση και κατανόηση της κλιματικής αλλαγής και μέσω παρακολούθησης των συνεχιζόμενων κλιματικών αλλαγών.

Με σκοπό τη δημιουργία ανθεκτικών κοινωνιών και οικονομιών, διαμορφώθηκαν πολιτικές και διεθνείς ατζέντες που αντιμετώπισαν τον κίνδυνο καταστροφών και την κλιματική αλλαγή ως δύο πλευρές του ίδιου νομίσματος, αναγνωρίζοντας τη συσχέτιση ανάμεσα στην προσαρμογή της κλιματικής αλλαγής με τη μείωση του καταστροφικού κινδύνου εξαιτίας αυτής. Οι τρεις εγκεκριμένες ατζέντες, βάσει των οποίων έχει δημιουργηθεί ένα ολοκληρωμένο πλαίσιο για μείωση της ευπάθειας του κλίματος και του αντίστοιχου κινδύνου και αύξηση της ανθεκτικότητας είναι οι: α. Συμφωνία των Παρισίων (Paris Agreement), β. Στόχοι Αειφόρου Ανάπτυξης (Sustainable Development Goals – SDGs) στο πλαίσιο της σχετικής ατζέντας του 2030 και γ. Πλαίσιο Sendai για η Μείωση του κινδύνου των καταστροφών 2015-2030 (Sendai Framework for Disaster Risk Reduction 2015-2030).

Η Συμφωνία των Παρισίων επικεντρώθηκε στο μετριασμό της κλιματικής αλλαγής και στη βελτίωση της ανθεκτικότητας σε αυτή, μέσω διοχέτευσης οικονομικών πόρων και ενέργειας. Κύριος στόχος της εν λόγω Συμφωνίας είναι η διατήρηση της αύξησης της θερμοκρασίας σε παγκόσμιο επίπεδο κάτω από το όριο των 2 °C έως το τέλος του αιώνα.

Με την υιοθέτηση των στόχων αειφόρου ανάπτυξης (SDG), η συσχέτιση μεταξύ κλιματικής αλλαγής και καταστροφών έγινε ακόμη πιο σαφής και επεκτάθηκε στο συμπέρασμα ότι ο συνδυασμός και των δύο υπονομεύει τα πρωταρχικά ανθρώπινα δικαιώματα. Μέσω των SDG, επιχειρείται η δημιουργία μίας κοινωνίας χαμηλών εκπομπών άνθρακα και ανθεκτικής στο κλίμα και τις καταστροφές, προωθώντας την έγκαιρη προειδοποίηση. Στόχος τους είναι η συνεχής αύξηση των κοινωνιών που υιοθετούν και εφαρμόζουν ολοκληρωμένες πολιτικές και σχέδια για την ένταξη, την αποδοτικότητα πόρων, το μετριασμό και προσαρμογή στην κλιματική αλλαγή και την ανθεκτικότητα στις καταστροφές. Οι παραπάνω στόχοι ευθυγραμμίζονται με το Πλαίσιο Sendai, το οποίο πραγματεύεται τη μείωση του κινδύνου που προέρχεται από τις καταστροφές.

Μέσα από το πλαίσιο Sendai, διατυπώθηκε καθαρά η σημασία της έγκαιρης προειδοποίησης όχι μόνο για τη μείωση του κινδύνου καταστροφών, αλλά γενικά για την αειφόρο ανάπτυξη των κοινωνιών με την επισήμανση συγκεκριμένων δράσεων που ενδυναμώνουν την ανθεκτικότητα. Για το σκοπό αυτό, προτείνει διαθεσιμότητα και πρόσβαση σε EWS πολλαπλών κινδύνων και σε πληροφορίες καταστροφών σε όλους τους ανθρώπους έως το 2030 [4].



1.2.2 Ορισμός & Χαρακτηριστικά των EWS

Τα EWS αποτελούν ολοκληρωμένα συστήματα παρακολούθησης κινδύνων, πρόγνωσης και πρόβλεψης, εκτίμησης κινδύνων λόγω καταστροφών, συστημάτων και διαδικασιών επικοινωνίας και ετοιμότητας, και επιτρέπουν σε άτομα, κοινότητες, κυβερνήσεις, επιχειρήσεις και άλλους να λάβουν έγκαιρα μέτρα για τη μείωση του αντίκτυπου των φυσικών κινδύνων λόγω καταστροφών, πριν από επικίνδυνα γεγονότα [15], παρέχοντας έγκαιρη και σχετική πληροφόρηση με συστηματικό τρόπο. Η αποτελεσματική εφαρμογή τους, μπορεί να συμβάλει στην αύξηση της ανθεκτικότητας των πόλεων με αυξημένους κλιματολογικούς κινδύνους και να προσφέρει υποστήριξη στην επίτευξη βιώσιμων αναπτυξιακών στόχων, σχετικά με τη μείωση της απώλειας ζωής και την επιβίωση [4].

Για να είναι πλήρως αποτελεσματικό και ανθρωποκεντρικό, κάθε σύστημα έγκαιρης προειδοποίησης θα πρέπει να περιλαμβάνει τέσσερα αλληλένδετα βασικά στοιχεία [16, 17]:

1. *Γνώση σχετικά με τον κίνδυνο των καταστροφών, βάσει της συστηματικής συλλογής δεδομένων και της εκτίμησης κινδύνου σχετικά με την τρωτότητα των ανθρώπων και των κοινωνιών απέναντι στον κίνδυνο.* Η επικινδυνότητα προκύπτει από το συνδυασμό του κινδύνου και των τρωτών σημείων σε μία συγκεκριμένη τοποθεσία. Για το λόγο αυτό, η γνώση της επικινδυνότητας είναι ένα βασικό στοιχείο ενός επιτυχημένου EWS. Οι πληροφορίες σχετικά με την επικινδυνότητα πηγάζουν από την αλληλεπίδραση μεταξύ της δημιουργίας οργανωτικών ρυθμίσεων, της εκτίμησης της τρωτότητας των κοινωνιών, του καθορισμού των φυσικών κινδύνων και της αποθήκευσης και του διαμοιρασμού πληροφοριών, και η ύπαρξη τους είναι απαραίτητη πριν την ανάληψη οποιασδήποτε δράσης. Πριν τα EWS δοθούν στους τελικούς χρήστες, θα πρέπει η σημασία και οι δυνατότητες τους να έχουν γίνει πλήρως αντιληπτά από τις κοινωνικές αρχές, οι οποίες θα πρέπει να γνωρίζουν ήδη τα κανάλια επικοινωνίας προειδοποίησης κινδύνου καθώς και τα σχέδια εκτάκτων αναγκών. Απαιτείται συνεχής επαγρύπνηση καθώς τα δεδομένα που σχετίζονται με τον κίνδυνο αλλάζουν συνεχώς, λόγω του ότι οι μεταβολές στους πληθυσμούς και τις υποδομές είναι δυναμικές – λόγω αλλαγών στην έκταση / τοποθεσία (έκθεση) και αλλαγών στην ευαισθησία / ικανότητες αντιμετώπισης (ευπάθεια). Η κλιματική αλλαγή, τροποποιεί τις υδρομετεωρολογικές συνθήκες σε όλο τον κόσμο και επηρεάζει τη συχνότητα και την ένταση των κινδύνων, επιφέροντας αβεβαιότητα στις προβλέψεις. Ως εκ τούτου, τα EWS πρέπει να προσαρμόζονται συνεχώς στις μεταβαλλόμενες συνθήκες, ενώ τα γραπτά σχέδια και οι διαδικασίες έκτακτης ανάγκης θα πρέπει να επικαιροποιούνται τακτικά.
2. *Ανίχνευση, τεχνική ικανότητα παρακολούθησης των πρόδρομων κινδύνων, ανάλυση και πρόβλεψη των κινδύνων και πιθανών συνεπειών αυτών.* Η παρακολούθηση και η προειδοποίηση είναι οι πιο συχνά συνδεδεμένες έννοιες με τα EWS. Η παρακολούθηση αφορά στη συλλογή πληροφοριών από ένα σύνολο αντιπροσωπευτικών μεταβλητών που σχετίζονται με τον κίνδυνο, όπως η βροχή (σχετίζεται με πλημμύρες / ξηρασία), ή τα σεισμικά κύματα (σχετίζονται με σεισμούς). Αυτό μπορεί να γίνει, μέσω άμεσων παρατηρήσεων, π.χ. όπως παρατηρώντας



μία επερχόμενη πυρκαγιά ή κατολίσθηση. Ωστόσο, τέτοιες παρατηρήσεις έχουν περιορισμένη χρηστικότητα λόγω του χρονικού ορίου εγγύτητας της απειλής, που αφήνει περιορισμένες δυνατότητες ανάληψης δράσης για τη μείωση του κινδύνου. Αντίστοιχο πρόβλημα προκύπτει με τους σεισμούς, όπου τα σεισμόμετρα μπορούν να δώσουν προειδοποίηση λίγα δευτερόλεπτα πριν τη δόνηση. Για πολλούς άλλους κινδύνους, που τα EWS μπορούν να δώσουν προειδοποιήσεις με πολύ μεγαλύτερο χρόνο αντίδρασης (εβδομάδες έως μήνες) όπως καταιγίδες, πλημμύρες και ξηρασίες, η συνεχής παρακολούθηση σημαντικών αντιπροσωπευτικών μεταβλητών μπορεί να επιφέρει ενέργειες μετριασμού του κινδύνου. Η σύγχρονη τεχνολογία παρέχει τη δυνατότητα συλλογής δεδομένων από πολλές πηγές παρακολούθησης και με υψηλή ταχύτητα, με σκοπό τη βελτίωση της ακρίβειας και της ταχύτητας, με την οποία μπορούν να δημιουργηθούν προβλέψεις για καταστροφές. Μερικές τεχνολογικές λύσεις παρακολούθησης που μπορούν να παρέχουν τέτοια δεδομένα είναι οι δορυφορικές εικόνες, με εκτιμήσεις για τις θερμοκρασίες, την ταχύτητα του ανέμου, και το ύψος της βροχόπτωσης, επίγεια ραντάρ για την εκτίμηση της νεροποντής και επίγεια σημεία μέτρησης της ροής του νερού και της ταχύτητας του ανέμου καθώς και ανάλυση των τάσεων μέσω απευθείας επικοινωνίας για εύρεση ενδείξεων μελλοντικών καταστροφών. Ένα πολλά υποσχόμενο σύστημα παρακολούθησης είναι η νέα πρακτική συλλογής δεδομένων μέσω των κοινωνικών δικτύων. Πολλοί ερευνητές έχουν αποπειραθεί να προεκτείνουν προβλέψεις σχετικά με μελλοντικές καταστροφές, βάσει δεδομένων που ελήφθησαν από τα κοινωνικά δίκτυα είτε από τους ενσωματωμένους αισθητήρες των έξυπνων κινητών τηλεφώνων (smartphones). Αυτός ο τρόπος συλλογής δεδομένων είναι ιδιαίτερα χρήσιμος για τη συλλογή μεγάλου όγκου δεδομένων, από τις περιοχές ενδιαφέροντος, συμπεριλαμβανομένης και πληροφορίας της ακριβούς τοπογραφίας. Σε πολλές χώρες, ιστορικές πληροφορίες καιρού και υποδομών παρακολούθησης υπάρχουν ήδη, αλλά τυχαίνει να είναι διάσπαρτες με αποτέλεσμα τον κατακερματισμό των δυνατοτήτων παρακολούθησης καθώς και τη μειωμένη διαθέσιμη πληροφορία, η οποία είναι απαραίτητη για τις προγνώσεις. Εάν υπάρχει δυνατότητα γεφύρωσης αυτών των χασμάτων, μέσω των κοινωνικών δικτύων και των smartphones, είναι πιθανό να υπάρξει άμεση βελτίωση στις ικανότητες πρόβλεψης.

3. *Διάδοση και επικοινωνία, από επίσημη πηγή, έγκυρων, έγκαιρων, ακριβών και ενεργών προειδοποιήσεων και συναφών πληροφοριών σχετικά με την πιθανότητα και τον αντίκτυπο.* Η διανομή κατανοητών προειδοποιήσεων και πληροφοριών ετοιμότητας σε όσους διατρέχουν κίνδυνο υποστηρίζει τη διαδικασία διάδοσης και επικοινωνίας. Η διασπορά και μετάδοση των προειδοποιήσεων επικεντρώνονται στα συστήματα επικοινωνίας που παρέχουν αναγνωρίσιμα προειδοποιητικά μηνύματα, κατόπιν σχετικών αποφάσεων. Υπό αυτήν την έννοια, αυτό το στοιχείο των EWS θεωρείται ως το σύνολο της υποδομής της επικοινωνίας του κινδύνου (αξιόπιστο και ανθεκτικός εξοπλισμός σε καταστροφές, πληροφορίες και τεχνολογίες επικοινωνίας) και των στρατηγικών (κατάλληλες αλληλεπιδράσεις μεταξύ των βασικών ενδιαφερόμενων, αποτελεσματικά και προσαρμοσμένα προειδοποιητικά μηνύματα). Μη αποτελεσματική επικοινωνία και ανεξέλεγκτη διασπορά



των προειδοποιήσεων συνήθως έχει ως αποτέλεσμα μη αποτελεσματική ή καθόλου απόκριση. Ιστορικά, οι περισσότερες αστοχίες των EWS σημειώθηκαν λόγω εσφαλμένης επικοινωνίας - έλλειψη εξοπλισμού ή αποτυχημένη υποδομή. Ο κύριος στόχος της επικοινωνίας των επερχόμενων καταστροφών είναι αφενός η έγκαιρη προειδοποίηση κινδύνου για μία συγκεκριμένη, γεωγραφικά περιορισμένη καταστροφή, και αφετέρου η ενίσχυση των ανθρώπων να αναλάβουν δράση, να δρομολογήσουν τροποποιήσεις και να λάβουν μέτρα προστασίας πριν από ένα καταστροφικό συμβάν [18]. Για να επιτευχθεί αυτός ο στόχος, πρέπει η επικοινωνία έγκαιρης προειδοποίησης να απευθύνεται σε άτομα και κοινωνίες – δυνητικά απειλούμενους – χρησιμοποιώντας κατάλληλα γλωσσικά και τεχνολογικά εργαλεία και να προτρέπει την ανάληψη δράσης, μέσω σαφών προειδοποιητικών μηνυμάτων, προκειμένου να μην παρερμηνευθούν λόγω του επειγόντως και της πίεσης της κατάστασης. Για το λόγο αυτό, πολλά σύγχρονα συστήματα χρησιμοποιούν μία ορατή κλίμακα χρώματος. Η αποτελεσματική ροή πληροφοριών είναι δυνατό να εξασφαλισθεί μέσω καταγραφής και εφαρμογής σαφών οργανωτικών διαδικασιών και διαδικασιών λήψης αποφάσεων, βάσει συγκεκριμένων προαπαιτούμενων. Οι ξεκάθαροι ρόλοι εξασφαλίζουν ότι οι προειδοποιήσεις ανακοινώνονται άμεσα χωρίς να χρειάζονται εγκρίσεις από πολλά επίπεδα διοίκησης. Επιπλέον, αποτρέπεται η σύγχυση που δημιουργείται όταν διαφορετικές οντότητες παρέχουν αντικρουόμενες ή ελλιπείς πληροφορίες. Όλα αυτά καθορίζονται μέσω πρότυπων πρωτοκόλλων που επιτρέπουν τον ομαλό συντονισμό και την ανταλλαγή δεδομένων μεταξύ των παραγόντων. Πλέον, τα διαθέσιμα τεχνολογικά εργαλεία για την επικοινωνία και τη διάδοση των κινδύνων αυξάνονται ποιοτικά και αριθμητικά. Παράλληλα με τις παραδοσιακές τεχνολογίες για την έγκαιρη πληροφόρηση και επικοινωνία (ακουστικές σειρήνες, ραδιόφωνο, τηλεόραση, SMS, email), η ταχεία εξάπλωση των smartphone έχει αυξήσει τη πρόσβαση σε εργαλεία που βασίζονται στο Διαδίκτυο, όπως εφαρμογές για κινητά και ιστότοπους. Οι εφαρμογές για κινητές συσκευές προσφέρουν όχι μόνο τη δυνατότητα πληροφόρησης για προειδοποίηση επικείμενων κινδύνων, αλλά και πρόσθετες πληροφορίες κατόπιν αιτήματος του χρήστη. Κάποιες πλατφόρμες κοινωνικών μέσων, π.χ. Facebook, διαθέτουν ολοκληρωμένες λειτουργίες προειδοποίησης κινδύνου, ενώ όταν μία πλατφόρμα κοινωνικής δικτύωσης δεν έχει ενσωματωμένη λειτουργία προειδοποίησης, πολλοί πάροχοι (π.χ. Twitter, WhatsApp, Instagram, YouTube, blogs) χρησιμοποιούνται συχνά, ανεπίσημα, προκειμένου να δημιουργήσουν ένα κύμα προειδοποιήσεων μέσω ενεργών αναρτήσεων χρηστών.

4. *γνώση, πλάνα και ικανότητα άμεσης και κατάλληλης απόκρισης από τις αρχές και τις ευάλωτες ομάδες, στις προειδοποιήσεις που λαμβάνονται.* Η ικανότητα απόκρισης αφορά στη συγκεντρωτική γνώση, σε σχέδια και στοιχεία που είναι απαραίτητα για έγκαιρη και κατάλληλη δράση από τις αρχές και εκείνους που κινδυνεύουν. Η ικανότητα απόκρισης μπορεί να βελτιωθεί είτε αυξάνοντας τη δημόσια και θεσμική ετοιμότητα, ή με αυτοματοποιημένες αντιδράσεις έκτακτης ανάγκης. Για τις πιο ξαφνικές καταστροφές όπως εκείνες που προέρχονται από σεισμούς, αισθητήρες παρακολούθησης μπορεί να προκαλέσουν ενέργειες όπως απενεργοποίηση της τροφοδοσίας ρεύματος και φυσικού αερίου ώστε να αποφεύγονται περαιτέρω συνέπειες. Για να επιτευχθεί μία έγκαιρη απόκριση στις προειδοποιήσεις, θα πρέπει οι απειλούμενες κοινότητες να έχουν



εμπιστοσύνη στο EWS. Είναι γεγονός ότι επανειλημμένες ειδοποιήσεις από τα EWS, συνήθως εκτιμώνται ως λάθος συναγερμού με αποτέλεσμα τη στιγμή της πραγματικής ανάγκης να μην υπάρξει η σωστή απόκριση. Για αυτό θα πρέπει τα EWS μέσω ρουτινών αυτό - ελέγχου να ελαχιστοποιούν τις περιπτώσεις ψευδών ή περιττών προειδοποιήσεων. Η ετοιμότητα και η άμεση απόκριση μπορούν να βελτιωθούν μέσω ασκήσεων και εκπαίδευσης ή μέσω καταγραφής, σε λίστες ελέγχου, των ενεργειών που πρέπει να αναληφθούν σε συγκεκριμένες περιπτώσεις. Η ικανότητα απόκρισης επίσης πρέπει να λάβει υπόψη το γεγονός ότι όλοι οι άνθρωποι πρέπει να λάβουν την ίδια προστασία ανεξάρτητα από το φύλο, την ηλικία, την εκπαίδευση ή ενδεχόμενη αναπηρία. Όλα τα παραπάνω πρέπει να υποστηρίζονται από τις κοινωνίες, οι οποίες πρέπει να έχουν θεσμοθετημένες τις ικανότητες απόκρισης τους, όσον αφορά στην οργάνωση. Η ικανότητα αυτή αφορά στο συνδυασμό δυνατοτήτων, χαρακτηριστικών και διαθέσιμων πόρων ώστε η κοινωνία να μπορέσει να αντιδράσει στις προειδοποιήσεις για την ελαχιστοποίηση των επιπτώσεων των καταστροφών. Η θεσμική ικανότητα αντιπροσωπεύει το διαρθρωτικό πλαίσιο που απαιτείται για την εκπλήρωση της εντολής των συστημάτων έγκαιρης προειδοποίησης μέσω της ανάπτυξης και εφαρμογής εσωτερικών πολιτικών, ρυθμίσεων, διαδικασιών και πλαισίων.

Κάθε ένα από τα παραπάνω στοιχεία έχει άμεσες αμφίδρομες συσχετίσεις και αλληλεπιδράσεις με τα υπόλοιπα τρία στοιχεία, και δεν αποτελεί το ένα επακόλουθο του προηγούμενου. Αντίθετα, όλα τα στοιχεία θα πρέπει να συντονίζονται, εντός και μεταξύ τομέων και πολλαπλών επιπέδων, για την αποτελεσματική λειτουργία του συστήματος καθώς και να περιλαμβάνουν έναν μηχανισμό ανατροφοδότησης, για συνεχή βελτίωση. Αν και τα στοιχεία της παρακολούθησης και της πρόβλεψης αποτελούν, όπως αναφέρθηκε, το πιο διαδομένο κομμάτι των EWS, από μόνα τους δεν αρκούν για να περιορίσουν τις επιπτώσεις μίας επερχόμενης καταστροφής. Ως πιο αδύναμα στοιχεία μπορούν να θεωρηθούν τα στοιχεία επικοινωνίας και ετοιμότητας, στα οποία παίζει καθοριστικό ρόλο η ανθρώπινη αντίδραση [19]. Η αποτυχία σε ένα στοιχείο ή η έλλειψη συντονισμού μεταξύ τους θα μπορούσε να οδηγήσει σε αποτυχία ολόκληρου του συστήματος. Η μακροπρόθεσμη αποτελεσματικότητα των EWS απαιτεί ισχυρή πολιτική και θεσμική δέσμευση καθώς και ευαισθητοποίηση των πολιτών.

1.3 Τα EWS για μετεωρολογικές προβλέψεις και προειδοποιήσεις στην Ελλάδα

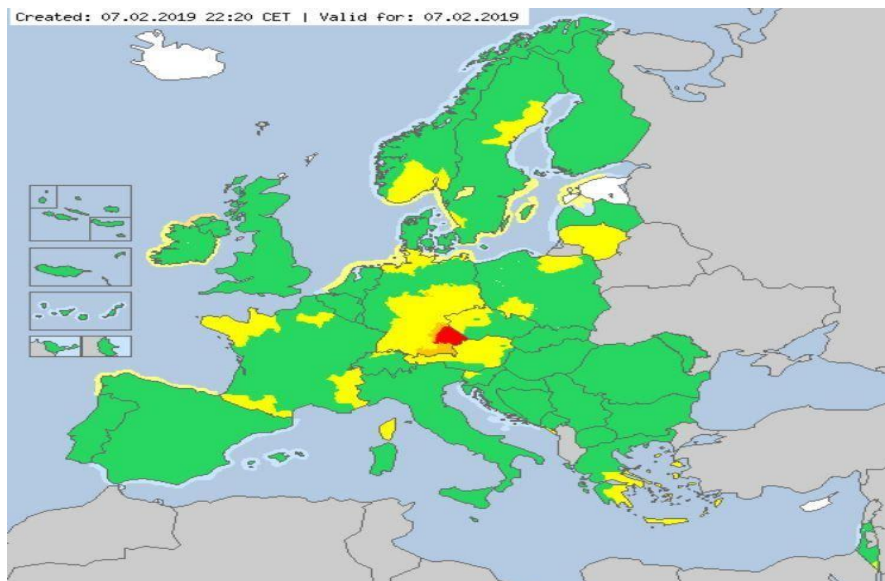
1.3.1 Το σύστημα Meteoalarm

Ένα από τα βασικά συστήματα που χρησιμοποιείται για την πρόβλεψη έντονων καιρικών φαινομένων στην γηραιά Ήπειρο, είναι το Meteoalarm. Το σύστημα αυτό, προειδοποιεί για επικείμενη παρουσία έντονων καιρικών φαινομένων, π.χ. καταιγίδες που μπορεί να οδηγήσουν σε πλημμύρες, χιόνι και πολικό ψύχος αλλά και θυελλώδεις άνεμοι, ενώ μπορεί να προειδοποιήσει ακόμα και για τον κίνδυνο παλιρροιακών παράκτιων κυμάτων αλλά και πυρκαγιών. Είναι εύκολο να διαπιστώσει κανείς σε ποιες περιοχές είναι αρκετά πιθανό να προκύψει κάποιο έντονο καιρικό φαινόμενο, ενώ χρησιμοποιείται το ανάλογο χρώμα για να αναδειχθεί ο βαθμός επικινδυνότητας του



φαινομένου. Έτσι βλέποντας κανείς τον Ευρωπαϊκό χάρτη, μπορεί να καταλάβει το επικινδυνότερο επικείμενο φαινόμενο για κάθε χώρα βάσει του χρώματος που έχει δοθεί σε αυτήν. Ταυτόχρονα ο χρήστης χρησιμοποιώντας υπερσυνδέσεις μπορεί να ενημερωθεί για τις εθνικές αλλά και τοπικές προειδοποιήσεις.

Πιο συγκεκριμένα μπορεί κανείς να ενημερωθεί για τη διάρκεια, την ένταση αλλά και άλλες πληροφορίες που αφορούν το φαινόμενο σε κάθε περιοχή της Ευρώπης που αναζητά. Οι πληροφορίες που παρέχει το www.meteoalarm.eu αφορούν την παρούσα και την αμέσως επόμενη ημέρα, ενώ αν ο χρήστης χρειάζεται περισσότερες πληροφορίες σε εθνικό επίπεδο, μπορεί να συνδεθεί με την Εθνική Μετεωρολογική Υπηρεσία της συγκεκριμένης χώρας.



Εικόνα 1.2: Χάρτης του Meteoalarm για την Ελλάδα [20]

Οι πληροφορίες που παρέχει το Meteoalarm, προέρχονται από Εθνικές Μετεωρολογικές Υπηρεσίες πολλών χωρών της Ευρώπης. Υπάρχει συνεχής ενημέρωση των πληροφοριών, ενώ τη διαχείριση του συστήματος κάνει η Εθνική Μετεωρολογική Υπηρεσία της Αυστρίας (ZAMG), εκπροσωπώντας το Δίκτυο Ευρωπαϊκών Μετεωρολογικών Υπηρεσιών (EUMETNET). Οι χώρες που συμμετέχουν στο σύστημα είναι οι: Αυστρία, Βέλγιο, Ελβετία, Κύπρος, Τσεχία, Γερμανία, Δανία, Εσθονία, Ισπανία, Φινλανδία, Γαλλία, Ελλάδα, Κροατία, Ουγγαρία, Ιρλανδία, Ισλανδία, Ιταλία, Λουξεμβούργο, Λετονία, Μάλτα, Ολλανδία, Νορβηγία, Πολωνία, Πορτογαλία, Ρουμανία, Σερβία, Σουηδία, Σλοβενία, Σλοβακία, Ηνωμένο Βασίλειο.

1.3.2 Το σύστημα ΖΕΥΣ

Το δίκτυο ανίχνευσης αστραπών ΖΕΥΣ, διαχειρίζεται το Εθνικό Αστεροσκοπείο Αθηνών. Για την καταγραφή των συμβάντων, χρησιμοποιεί έξι δέκτες στην περιοχή της Ευρώπης, στο Μπέρμιγχαμ της Αγγλίας, στο Ρόσκιλντε της Δανίας, στο Ιάσιο της Ρουμανίας, στη Λάρνακα της Κύπρου, στη Λισσαβόνα της Πορτογαλίας καθώς και στην Αθήνα. Το σύστημα καταγραφής λειτουργεί ως εξής, όταν μια ηλεκτρική εκκένωση έρχεται σε επαφή με το έδαφος



γίνεται εκπομπή ηλεκτρομαγνητικής ακτινοβολίας χαμηλής συχνότητας που διαδίδεται με την ταχύτητα του φωτός. Προκειμένου όμως να προκύψει η ακριβής τοποθεσία της εκκένωσης απαιτούνται τέσσερις σταθμοί, οπότε όταν ο κεραυνός έρθει σε επαφή με το έδαφος η κυματομορφή η οποία εκπέμπεται φτάνει σε όλους τους σταθμούς αλλά σε διαφορετικό χρόνο. Ανάλογα λοιπόν με τη χρονική στιγμή στην οποία καταγράφεται η κυματομορφή από το σταθμό, προκύπτει η απόσταση από το σημείο που αυτό συνέβη. Πιο αναλυτικά το κέντρο ελέγχου, θέτει ως «σταθμό αναφοράς» και αποδίδει μηδενική διαφορά χρόνου στον σταθμό που έφτασε πρώτο το σήμα, και παράλληλα υπολογίζει τη διαφορά χρόνου άφιξης για τους άλλους σταθμούς σε σχέση με τον πρώτο σταθμό. Έπειτα υπολογίζεται ο κοινός γεωμετρικός τόπος των σημείων, για τα οποία η διαφορά άφιξης του σήματος του σταθμού αναφοράς σε σχέση με κάποιον άλλο του δικτύου ταυτίζονται, και προκύπτει λοιπόν πως τα σημεία είναι πάνω σε μια υπερβολή. Πραγματοποιείται επανάληψη της διαδικασίας για κάθε σταθμό και το σημείο στο οποίο τελικά σημειώθηκε ο κεραυνός είναι το σημείο τομής όλων των υπερβολών.

1.3.3 Το σύστημα ΠΟΣΕΙΔΩΝ

Το σύστημα παρακολούθησης θαλάσσιου περιβάλλοντος ΠΟΣΕΙΔΩΝ, οφείλει την ύπαρξη του στο Εθνικό Κέντρο θαλάσσιων Ερευνών (ΕΚΘΕ). Η χρηματοδότηση για την ανάπτυξη του συστήματος έγινε κατά 85% από την Ευρωπαϊκή Τράπεζα Επενδύσεων, ενώ το υπόλοιπο 15% χρηματοδότησε το Υπουργείο Οικονομίας. Οι πληροφορίες που παρέχει το σύστημα στον χώρο του Αιγαίου είναι ζωτικής σημασίας, για την πολιτική αλλά και την οικονομία της χώρας. Το σύστημα αποτελείται από έντεκα πλωτούς ωκεανογραφικούς σταθμούς, εννιά από τους οποίους λειτουργούν ενώ δύο είναι σε εφεδρεία, και οι οποίοι έχουν εξοπλιστεί με αισθητήρες για καταγραφή των ατμοσφαιρικών δεδομένων. Καταγράφονται στοιχεία όπως η ατμοσφαιρική πίεση, ο άνεμος, η θερμοκρασία, η ταχύτητα ρευμάτων αλλά και το ύψος και η κατεύθυνση κυμάτων. Πέρα από τους πλωτούς σταθμούς, το σύστημα κάνει χρήση 10 κυματογράφων με διαδικτυακή μετάδοση δεδομένων καθώς και ενός δορυφορικού συστήματος GPS1 που βοηθά στη μέτρηση του ύψους των κυμάτων. Σε περίπτωση που οι πλωτοί σταθμοί εντοπίσουν κάτι, εκπέμπουν προς το επιχειρησιακό κέντρο ΠΟΣΕΙΔΩΝ με τη χρήση των δορυφορικών συστημάτων INMARSAT-C και GSM.

1.3.4 Το σύστημα EFFIS

Το Ευρωπαϊκό Σύστημα Πληροφοριών για τις Πυρκαγιές (European Forest Fire Information System, EFFIS) οφείλει την ύπαρξη του, στη σύμπραξη της Ευρωπαϊκής Επιτροπής και των υπηρεσιών πυρόσβεσης κάθε χώρας. Είναι ένα ολοκληρωμένο σύστημα που διαχειρίζεται ό,τι έχει να κάνει με τις δασικές πυρκαγιές παρέχοντας πληροφορίες σε παραπάνω από τριάντα χώρες, που ανήκουν στις μεσογειακές και ευρωπαϊκές περιοχές. Το σύστημα περιλαμβάνει: Πρόβλεψη Πυρκαγιάς, Ενεργή Ανίχνευση Πυρκαγιάς, Ταχεία Εκτίμηση Ζημιών, αλλά και Ανάλυση των ζημιών μετά το πέρας της πυρκαγιάς. Το σύστημα βασίζεται στα συστήματα τηλεπισκόπησης και γεωγραφικών



πληροφοριών μέσω υπηρεσιών διαδικτυακής λειτουργίας για να πετυχαίνει την καλύτερη κάλυψη σε πραγματικό χρόνο.

1.3.5 Το σύστημα SEVIRI

Το σύστημα SEVIRI αναπτύχθηκε από το Εθνικό Αστεροσκοπείο Αθηνών, και δίνει τη δυνατότητα στο χρήστη να κάνει παρακολούθηση όλων των πυρκαγιών που εκτυλίσσονται στον χώρο της επικράτειας, με πληροφορίες για την κατάσταση στην οποία βρίσκονται. Συνιστά μια από τις βασικές υπηρεσίες δορυφορικής τηλεπισκόπησης, η οποία βρίσκει εφαρμογή σε επίπεδο Ευρωπαϊκής Ένωσης μέσα από το πρόγραμμα GMES (Global Monitoring for Environment and Security).

1.3.6 Το σύστημα Disaster Management Framework

Το σύστημα Disaster Management Framework αποτελεί ένα ελληνικό πληροφοριακό σύστημα, βασισμένο στις προηγμένες τεχνολογίες πληροφορικής και τηλεπικοινωνιών, που σκοπό έχει τη βοήθεια στη διαχείριση εκτάκτων αναγκών σε τοπικό επίπεδο. Το σύστημα μπορεί να είναι ιδιαίτερα χρήσιμο σε όλα τα στάδια διαχείρισης κρίσεων, καθώς συμβάλει στην αποτελεσματική οργάνωση αλλά και το συντονισμό σχεδίων αντιμετώπισης και διαχείρισης κρίσεων, που οφείλονται σε σοβαρά ατυχήματα, σεισμούς, πυρκαγιές και όλες τις πιθανές περιπτώσεις έκτακτων αναγκών. Το σύστημα αναπτύχθηκε από την χρηματοδότηση της Περιφέρειας Ανατολικής Αττικής, αλλά δε χρησιμοποιείται ευρέως καθώς εξαρτάται σε μεγάλο βαθμό από τους χειριστές του, με άλλα λόγια απαιτεί συμμετοχή μεγάλης μερίδας πολιτών.

1.3.7 Η περίπτωση της Μάνδρας Αττικής

Στην περίπτωση της καταστροφής που έπληξε τη Δυτική Αττική και ιδιαίτερα τη Μάνδρα το 2017 φαίνεται πως ένα μέρος από τις συνέπειες θα μπορούσαν να είχαν αποφευχθεί στην περίπτωση που είχαν αξιοποιηθεί τα διαθέσιμα δεδομένα. Δύο ημέρες πριν το συμβάν το κινητό μετεωρολογικό ραντάρ XPOL κατέγραφε ιδιαίτερα ισχυρή βροχόπτωση στην περιοχή που συρρέουν τα ρέματα στη Μάνδρα, με το ύψος βροχής να ξεπερνά τα 200mm μέσα σε έξι ώρες. Ταυτόχρονα, το Εθνικό Καποδιστριακό Πανεπιστήμιο Αθηνών, είχε ανακοινώσει πως το σύστημα «ΠΟΣΕΙΔΩΝ», είχε προβλέψει βροχόπτωση ύψους 50mm στο όρος Πατέρα, επίσης δύο ημέρες πριν. Αντίστοιχα και το Εθνικό Αστεροσκοπείο Αθηνών, επιβεβαίωσε πως για τις δύο προηγούμενες ημέρες του συμβάντος, η γύρω περιοχή είχε δεχτεί 35mm βροχή στην Ελευσίνα και τα Βίλια και 22mm στους Αγίους Θεοδώρους. Γίνεται, λοιπόν, εύκολα αντιληπτό πως εάν βρισκόταν σε λειτουργία ένα Σύστημα Έγκαιρης Προειδοποίησης, του οποίου οι προειδοποιήσεις θα λαμβάνονταν σοβαρά από την πολιτεία, θα είχαν εκτιμηθεί τα ανησυχητικά ύψη υετού των προηγούμενων ημερών και θα υπήρχε δυνατότητα κατάστροφωσης και υλοποίησης ενός ειδικού σχεδίου δράσεων έκτακτης ανάγκης για την αποφυγή των συνεπειών που τελικά ακολούθησαν.

1.4 Κόστη και πλεονεκτήματα των EWS παγκοσμίως



Η ανάλυση κόστους - κέρδους είναι πολύ βασική, για τον υπολογισμό του κόστους για τα μέτρα πρόληψης καταστροφών. Η Παγκόσμια Τράπεζα, σε μια ανάλυση που πραγματοποίησε με θέμα, φυσικοί κίνδυνοι - αφύσικες καταστροφές, αναγνωρίζει 3 βασικούς πυλώνες για την αποφυγή των συνεπειών φυσικών καταστροφών: τα EWS, τα έργα υποδομών και η ενίσχυση του φυσικού περιβάλλοντος. Όπως έχει ειπωθεί, τα Συστήματα Έγκαιρης Προειδοποίησης, είναι ζωτικής σημασίας, καθώς συμβάλουν στην αποφυγή απωλειών ανθρώπινων ζωών αλλά και περιουσιακών στοιχείων. Τα αποτελεσματικά έργα υποδομών συνεισφέρουν επίσης, αλλά μειώνουν επιπλέον και τις καταστροφές που θα προκαλούνταν. Γέφυρες και δρόμοι που χρησιμοποιούνται ως διαδρομές εκκένωσης, σχολεία που παρέχουν καταφύγιο πάνω από τα νερά πλημμύρας, νοσοκομεία που είναι ανθεκτικά στον σεισμό είναι όλα παραδείγματα για έργα υποδομών που σχετίζονται με συγκεκριμένους κινδύνους και συμβάλουν στη μείωση του ρίσκου φυσικών καταστροφών. Η επένδυση στα έργα υποδομών παρέχει τόσα οφέλη που αντισταθμίζουν το κόστος της ίδιας της επένδυσης. Η ενίσχυση του φυσικού περιβάλλοντος προσφέρει προστασία από κινδύνους εντός φυσικών περιορισμών, για παράδειγμα πυκνή βλάστηση για πλημμύρες και έλεγχο της διάβρωσης ή κοραλλιογενείς ύφαλοι και αμμόλοφοι για προστασία από τις καταγίδες. Γενικά είναι φθηνότερη η προστασία απ' την επαναφορά, αν και αυτό που πρέπει να προστατευθεί δεν είναι πάντα προφανές.

Ενώ τα στοιχεία σχετικά με το κόστος και τα οφέλη από τη μείωση του κινδύνου καταστροφών δείχνουν με συνέπεια ότι οι επενδύσεις σε συστήματα έγκαιρης προειδοποίησης μπορούν να σώσουν ζωές και να βοηθήσουν στην προστασία της περιουσίας, τα περισσότερα προληπτικά μέτρα επικεντρώνονται στα έργα υποδομών για την πρόληψη καταστροφών, όπως συστήματα ελέγχου πλημμυρών, ενίσχυση κανόνων δόμησης, κατασκευή καταφυγίων και προστασία του φυσικού περιβάλλοντος. Συχνά αυτά τα προληπτικά μέτρα εξυπηρετούν διαφορετικούς ανθρώπους με διαφορετικούς τρόπους. Οι επενδύσεις σε έργα υποδομών ενδέχεται να μην βελτιώνουν άμεσα την ασφάλεια των φτωχότερων που μπορεί να είναι εκτός των κατασκευαστικών έργων σε πιο περιθωριοποιημένες περιοχές. Η σωστά κατανοημένη επένδυση και στους τρεις τομείς, της έγκαιρης προειδοποίησης, των έργων υποδομών αλλά και της ενίσχυσης του φυσικού περιβάλλοντος, είναι πιθανώς η καλύτερη προσέγγιση, καθώς ούτε οι μεγάλες επενδύσεις σε υποδομές για τον έλεγχο των πλημμυρών, ούτε ένα σύστημα έγκαιρης προειδοποίησης για πλημμύρες μπορεί να αποτρέψει εντελώς μια καταστροφή. Σε συνδυασμό όμως μπορούν να μειώνουν σημαντικά τις απώλειες ζωών αλλά και περιουσιών των ανθρώπων. Επομένως, τα σημαντικά οφέλη της έγκαιρης προειδοποίησης γίνονται συχνά καλύτερα κατανοητά όταν συνδυάζονται με υποδομές και άλλες προληπτικές επενδύσεις. Στο Χονγκ Κονγκ, για παράδειγμα, οι βελτιώσεις στην κατασκευή των σπιτιών υποδεικνύουν ότι η καλύτερη δράση για προστασία του πληθυσμού από έναν τροπικό κυκλώνα, είναι να μείνουν μέσα. Εδώ, η έγκαιρη προειδοποίηση διασφαλίζει ότι οι άνθρωποι θα έχουν το χρόνο να επιστρέψουν με ασφάλεια στα σπίτια τους χρησιμοποιώντας ένα ευπροσάρμοστο σύστημα δημόσιων μεταφορών που ανταποκρίνεται στο προειδοποιήσεις. Έγκαιρες προειδοποιήσεις, αποτελεσματική επικοινωνία και γρήγορη ανταπόκριση επιτρέπουν στην οικονομία του Χονγκ Κονγκ να ανακάμπτει γρήγορα από τέτοιους κινδύνους.



Παρόλα αυτά, σε γενικές γραμμές, τα στοιχεία δείχνουν ότι τείνουμε να δαπανούμε περισσότερα χρήματα για την επαναφορά απ' όσα για την πρόληψη και πως αυτό είναι μία δημόσια προτίμηση, παρότι τα στοιχεία δείχνουν ότι οι επενδύσεις για πρόληψη είναι πολυτιμότερες από δαπάνες για ανάκαμψη [21]. Ακόμα κι όταν όμως πραγματοποιούνται όντως δαπάνες για προληπτικά μέτρα, αυτό που συμβαίνει συχνά είναι να μην αφορούν το κομμάτι του πληθυσμού με μικρή οικονομική επιρροή και δίχως πολιτική φωνή, με αποτέλεσμα μεγάλες επενδύσεις σε υποδομές πρόληψης συχνά να οδηγούν στην αναγκαστική μετεγκατάσταση των φτωχών κατοίκων σε άλλες, επιρρεπείς σε φυσικές καταστροφές, περιοχές μιας πόλης (Παγκόσμια Τράπεζα 2010). Σε αυτήν την περίπτωση, τα συστήματα έγκαιρης προειδοποίησης εξυπηρετούν καλύτερα τους φτωχούς σε σχέση με τις υποδομές πρόληψης.

- **Υπηρεσία προειδοποίησης για τροπικούς κυκλώνες στο Χονγκ Κονγκ**

Η υπηρεσία προειδοποίησης για τροπικούς κυκλώνες του Παρατηρητηρίου του Χονγκ Κονγκ (ΗΚΟ) έχει εξελιχθεί σε ένα σύστημα αρίθμησης: Το σήμα κατάστασης αναμονής με αριθμό 1 εκδίδεται κάθε φορά που ένας τροπικός κυκλώνας βρίσκεται σε απόσταση 800 χιλιομέτρων από το Χονγκ Κονγκ και ενδέχεται να επηρεάσει την περιοχή αργότερα, τα σήματα με αριθμό 3 και 8 προειδοποιούν το κοινό για ισχυρούς ανέμους και θύελλα στην πόλη αντίστοιχα, το σήμα με αριθμό 9 υποδηλώνει αυξανόμενους ανέμους θύελλας ή καταιγίδας, ενώ αυτό με αριθμό 10 προειδοποιεί για ανέμους έντασης τυφώνα. Με αφορμή τους αυστηρούς κανόνες δόμησης στο Χονγκ Κονγκ, ένα σπίτι θεωρείται γενικά το ασφαλέστερο μέρος για τους ανθρώπους να καταφύγουν από ένα τροπικό κυκλώνα. Όταν εκδίδεται το σήμα 8, το ΗΚΟ συμβουλεύει το κοινό να μείνει σπίτι ή να επιστρέψει στο σπίτι. Με την πάροδο του χρόνου, μια παράδοση κοινωνικής ανταπόκρισης έχει εμφανιστεί στην οποία οι επιχειρήσεις και τα σχολεία θα κλείσουν όταν ο τροπικός κυκλώνας εγείρει το προειδοποιητικό σήμα με αριθμό 8. Το σύστημα προειδοποίησης για τροπικούς κυκλώνες χρησιμοποιείται εδώ και πολλά χρόνια και το κοινό είναι ήδη εξοικειωμένο με αυτό. Μαζί με τις καλά συντονισμένες δράσεις αντιμετώπισης που έλαβαν οι υπηρεσίες βοήθειας, το σύστημα αυτό αποδείχθηκε πολύ αποτελεσματικό στη μείωση απωλειών ανθρώπινων ζωών αλλά και περιουσιακών στοιχείων λόγω τροπικών κυκλώνων.

- **Πώς παίρνεται η απόφαση για την κατάλληλη επένδυση σε προληπτικά μέτρα**

Η ανάλυση κόστους - οφέλους επισημαίνει ότι πρέπει να πραγματοποιηθεί μια επένδυση για την οποία τα οφέλη υπερβαίνουν το κόστος, και εάν υπάρχουν ανταγωνιστικές προτάσεις, θα πρέπει να προτιμάται αυτή που έχει το υψηλότερο όφελος από το κόστος. Το δίλημμα που προκύπτει είναι ότι είναι απαραίτητο να λάβουμε υπόψιν τις ζωές που σώθηκαν για να φτάσουμε στη σωστή επενδυτική επιλογή. Μια προσέγγιση, που προτάθηκε από τους Cropper και Sahim [22], υπολογίζει τη μείωση των τραυματισμών και των θανάτων σε σχέση με τα ποιοτικώς σταθμισμένα έτη ζωής (QALY). Οπότε το κόστος της πρόληψης αφαιρώντας τα οφέλη εκτός της υγείας από την πρόληψη, μπορεί να διαιρεθεί με τα QALY που έχουν αποθηκευτεί για να συγκρίνουν το κόστος ανά QALY σε όλες τις πολιτικές και να ενισχύσει τη συνέπεια στη λήψη αποφάσεων [23]. Αυτή η προσέγγιση θα ήταν ιδιαίτερα χρήσιμη για την αξιολόγηση της αποτελεσματικότητας των EWS, δεδομένου ότι έχουν τη μεγαλύτερη επίδραση στη μείωση



των απωλειών ανθρώπινων ζωών αλλά και τραυματισμών. Η μη καταμέτρηση της αξίας των ζωών που σώζονται, δίνει μηδενική αξία στη ζωή και υπονομεύει την πρόληψη. Για παράδειγμα, μόνο με την προσθήκη της αξίας των ζωών που σώθηκαν στην Τουρκία, προληπτικά μέτρα όπως η σεισμική ενίσχυση των διαμερισμάτων και των σχολείων, περνούν την ανάλυση κόστους - οφέλους.

Στις παρακάτω ενότητες περιγράφουμε τα οφέλη της έγκαιρης προειδοποίησης και στη συνέχεια εξετάζουμε με περισσότερες λεπτομέρειες το κόστος και τα οφέλη των προειδοποιήσεων όπου ψευδείς συναγερμοί και κακές αποφάσεις μπορούν να επιφέρουν μεγάλο κόστος .

Δεν υπάρχει αμφιβολία ότι τα αποτελεσματικά EWS έχουν μειώσει σημαντικά τους θανάτους και τους τραυματισμούς από σοβαρά καιρικά φαινόμενα. Οι έγκαιρες προειδοποιήσεις για τον κίνδυνο πλημμύρας έχουν αποδειχθεί αποτελεσματικές στη μείωση των θανάτων που σχετίζονται με τις πλημμύρες [24]. Για παράδειγμα, υπάρχει μια σημαντική διαφορά μεταξύ των πλημμυρών 1992-1994 κατά μήκος του ποταμού Ρήνου και Μόσα και τις πλημμύρες του 1995 κατά μήκος των ίδιων ποταμών [25]. Οι δύο πλημμύρες είχαν παρόμοια χαρακτηριστικά, και οι δύο προκλήθηκαν από μεγάλης διάρκειας δυνατή βροχόπτωση. Δέκα άνθρωποι έχασαν τη ζωή τους και πάνω από 900 εκατομμύρια δολάρια σε ζημιές σημειώθηκαν κατά την πρώτη εκδήλωση, ενώ στη δεύτερη το οικονομικό κόστος ήταν μειωμένο σχεδόν κατά το ήμισυ, δεν χάθηκαν ζωές, και όλα αυτά λόγω της ευαισθητοποίησης και της αλλαγής συμπεριφοράς των ανθρώπων.

Πολλές χώρες χαμηλού εισοδήματος με επαναλαμβανόμενες καταστροφές όπως το Μπαγκλαντές και η Κούβα έχουν μειώσει σημαντικά τη θνησιμότητα αναπτύσσοντας αποτελεσματικά EWS [26]. Το σύστημα έγκαιρης προειδοποίησης κυκλώνων της τροπικής Κούβας, πιστώνεται για τη δραματική μείωση των θανάτων που σχετίζονται με τον καιρό όπως οι τροπικοί κυκλώνες, οι καταιγίδες και οι πλημμύρες. Πέντε διαδοχικοί τυφώνες το 2008 άφησαν πίσω μόνο επτά νεκρούς. Ο Gustav, που θεωρείται ο ισχυρότερος τυφώνας που έπληξε την Κούβα τα τελευταία 50 χρόνια, είχε ως αποτέλεσμα μηδέν θανάτους, αν και καταστράφηκαν περισσότερα από 100.000 σπίτια. Η απουσία θανάτων επιτεύχθηκε χάρη σε ένα σύστημα έγκαιρης προειδοποίησης αλλά και αποτελεσματικής αντίδρασης, επιτρέποντας στους ανθρώπους που κινδυνεύουν να εκκενώσουν γρήγορα και να προσφύγουν στα καταφύγια έκτακτης ανάγκης. Ομοίως στο Μπαγκλαντές, παρά τις μεγάλες οικονομικές απώλειες, το σύστημα έγκαιρης προειδοποίησης και η άμεση ανταπόκριση έχουν μειώσει δραματικά τον αριθμό των ανθρώπων που χάθηκαν λόγω κυκλώνων τα τελευταία χρόνια.

- **Μπαγκλαντές: Οι ζωές που σώθηκαν λόγω πρόληψης (από την Παγκόσμια Τράπεζα 2010)**

Ο κυκλώνας Σιντρ παρατηρήθηκε για πρώτη φορά νοτιοδυτικά των Νήσων Ανταμάν στον Κόλπο της Βεγγάλης έξι ημέρες πριν από τη δημιουργία του στις 15 Νοεμβρίου 2007. Παρακολουθώντας το μονοπάτι και την αυξανόμενη ισχύ του κυκλώνα, οι αρχές του Μπαγκλαντές είχαν χρόνο να προετοιμάσουν ένα ολοκληρωμένο πλάνο: εξέδωσαν προειδοποιήσεις και ενεργοποίησαν 44.000 εθελοντές που βοήθησαν στην αποτομή περίπου 3 εκατομμυρίων



ατόμων από τα σπίτια τους αλλά και τη φιλοξενία 1,5 εκατομμύριου πολιτών σε καταφύγια. Ελάχιστοι δεν το περίμεναν και ήταν απροστάτευτοι όταν ο Σιντρ χτύπησε, παρόλα αυτά η τεράστια δύναμή του ήταν καταστροφική. Ο κυκλώνας κατηγορίας 4 (5 είναι ο πιο σοβαρός) με διάμετρο 1.000 χιλιομέτρων και ένταση ανέμων έως 240 χιλιόμετρα την ώρα, σήκωσε κύματα 5,5 έως 6 μέτρων που πέρασαν επιχώματα σχεδιασμένα να αντέχουν 2,5 μέτρα. Οι δυνάμεις του Σιντρ μετριάσθηκαν κατά τη διέλευση του από το Σούντορμπον, μια μεγάλη έκταση από υγρότοπους με πυκνά δέντρα, αλλά τέτοιοι υγρότοποι έχουν μειωθεί με την πάροδο των ετών και τεράστιες μη προστατευόμενες περιοχές υπέστησαν σοβαρές ζημιές. Οι προσπάθειες διάσωσης ξεκίνησαν αμέσως μετά τον τέλος του κυκλώνα. Οι 12 περιοχές που υπέστησαν τις μεγαλύτερες καταστροφές, αν και λιγότερο πυκνοκατοικημένες και φτωχότερες από τον εθνικό μέσο όρο, απαριθμούσαν 18,7 εκατομμύρια άτομα, από τα οποία προέκυψαν 55.000 τραυματίες και 4.400 νεκροί ή αγνοούμενοι. Η κυβέρνηση εκτιμά ότι περιουσιακά στοιχεία αξίας 1,16 δισεκατομμυρίων δολαρίων υπέστησαν ζημιά, σχεδόν όλα αφορούσαν τη στέγαση και άλλες υποδομές. Αναμένονταν απώλειες 517 εκατομμυρίων δολαρίων, αλλά η κατάσταση θα μπορούσε να ήταν πολύ χειρότερη εάν η χώρα δεν είχε μάθει από προηγούμενες τραγωδίες.

- **Λανθασμένοι συναγερμοί, κόστη και πλεονεκτήματα για έγκαιρη προειδοποίηση**

Για να εκτιμήσουμε πλήρως τη συσχέτιση κόστος - όφελος των EWS, πρέπει να λάβουμε υπόψη το συνολικό λειτουργικό κόστος του συστήματος, τις κοινωνικές και οικονομικές απώλειες λόγω λανθασμένων συναγερμών και την κοινωνική και οικονομική ασφάλεια λόγω έγκαιρης δράσης. Τα παραπάνω δεδομένα χρειάζονται για να εκτιμηθεί σωστά εάν και πού πρέπει να δημιουργηθούν συστήματα έγκαιρης προειδοποίησης. Για να είναι αποτελεσματικά, πρέπει να συνυπάρχουν τα τέσσερα: γνώση κινδύνου, υπηρεσία ελέγχου και προειδοποίησης, διάδοση και επικοινωνία και ικανότητα απόκρισης. Εάν κάποιο από αυτά τα χαρακτηριστικά λείπει ή έχει αναπτυχθεί ελάχιστα, ολόκληρο το σύστημα αποτυγχάνει. Η στρατηγική για αποτελεσματική και έγκαιρη λήψη αποφάσεων πρέπει να είναι γνωστή. Αυτό περιλαμβάνει τον προσδιορισμό των πληροφοριών που χρειάζονται, πώς θα χρησιμοποιηθούν οι προβλέψεις, πόσο αξιόπιστη πρέπει να είναι η πρόβλεψη για να προκαλέσει μία αποτελεσματική απόκριση και πώς να κοινοποιούνται αυτές οι πληροφορίες καθώς και την ανοχή της αβεβαιότητας αυτής της πρόβλεψης.

Η επικοινωνία της έγκαιρης προειδοποίησης και του επιπέδου αβεβαιότητας των πληροφοριών στους χρήστες είναι κρίσιμη, διαφορετικά υπάρχει υψηλός κίνδυνος να αγνοηθούν ή να καταχραστούν οι προειδοποιητικές πληροφορίες που ενδέχεται να οδηγήσουν σε σημαντικό κοινωνικό και οικονομικό κόστος.

- **Λανθασμένοι Συναγερμοί**

Το πλήθος των ψευδών συναγερμών μειώνεται καθώς βελτιώνονται οι επιμέρους διαδικασίες των EWS. Σε γενικές γραμμές, επειδή οι προβλέψεις μικρότερης χρονικής κλίμακας είναι πιο αξιόπιστες, η πιθανότητα ψευδούς συναγερμού μειώνεται από το χρονικό περιθώριο, καθώς από τη προβλεπόμενη μείωση στο σύνολο των κινδύνων. Παρατηρείται μία αντιστάθμιση μεταξύ της επικαιρότητας, της αξιοπιστίας προειδοποίησης, του κόστους μιας



εσφαλμένης ειδοποίησης και της ζημίας που αποφεύγεται ως αποτέλεσμα του χρονικού ορίου, το οποίο πρέπει να συσχετίζεται με την αποδοτικότητα κόστους του αποτελέσματος. Ένας σημαντικός παράγοντας για την επίτευξη της αποδοτικότητας είναι η ικανότητα να επικοινωνείται η πληροφορία στον κατάλληλο χρόνο και τρόπο.

Το κόστος του ίδιου του EWS μπορεί να είναι υψηλό σε σχέση με το όφελος που έχει εάν το σύστημα χρησιμοποιείται αποκλειστικά για σπάνια συμβάντα, όπως μια πλημμύρα που μπορεί να συμβεί μία φορά σε 200 χρόνια. Εάν η ζημία που σχετίζεται με μια καταστροφή είναι 2 δισεκατομμύρια και εάν αυτή η καταστροφή θα μπορούσε να αποφευχθεί με σύστημα προειδοποίησης που κοστίζει 20 εκατομμύρια ετησίως, αυτό θα ήταν μία κακή επένδυση, καθώς το κόστος θα υπερέβαινε τις ετήσιες αναμενόμενες απώλειες κατά 10 εκατομμύρια ετησίως. Η πρόληψη, μέσω έγκαιρης προειδοποίησης, μπορεί να θεωρηθεί οικονομική μόνο εάν το συμβάν είναι συχνό, η ζημία μεγαλύτερη, ή η πρόληψη φθηνότερη (Παγκόσμια Τράπεζα 2010).

Στην πράξη, οι εθνικές μετεωρολογικές και υδρολογικές υπηρεσίες σπάνια κάνουν διάκριση μεταξύ της συνηθισμένης διαδικασίας της πρόβλεψης καθημερινών μετεωρολογικών προγνώσεων και μίας πρόβλεψης εξαιρετικά επικίνδυνων φαινομένων. Το τελευταίο αποτελεί ειδική περίπτωση των συνήθων δραστηριοτήτων των υπηρεσιών που σχετίζονται καιρού. Αυτό σημαίνει ότι το σταδιακά αυξανόμενο κόστος παροχής προειδοποιήσεων για σπάνια ακραία γεγονότα είναι πιθανό να είναι ένα σχετικά μικρό μέρος του κόστους παροχής των συνήθων υπηρεσιών. Σημαίνει επίσης ότι το προσωπικό βελτιώνει συνεχώς τις δεξιότητές του, αντί να τις ασκεί σπάνια. Επιπλέον, είναι δυνατή η περαιτέρω αποδοτικότητα του κόστους με την ανάπτυξη συστημάτων προειδοποίησης που εξυπηρετούν πολλαπλές χρήσεις. Για παράδειγμα, το Μετεωρολογικό Γραφείο της Σαγκάης της Μετρολογικής Διοίκησης της Κίνας, ανέπτυξε ένα πολύ - σύστημα έγκαιρης προειδοποίησης κινδύνου, το οποίο εξυπηρετεί τις ανάγκες πολυάριθμων υπηρεσιών εντός της δημοτικής κυβέρνησης. Σημαντική εξοικονόμηση κόστους πραγματοποιείται σε σχέση με διαφορετικά συστήματα που θα αναπτυχθούν ξεχωριστά για κάθε οργανισμό. Το σύστημα είναι επίσης επεκτάσιμο καθιστώντας δυνατή την προσθήκη νέων υπηρεσιών προειδοποίησης, όπως η πρόβλεψη υγείας. Μια περιφερειακή προσέγγιση που βασίζεται στις ίδιες αρχές με πολλές χώρες να αναπτύσσουν ένα κοινό σύστημα προειδοποίησης πολλαπλών κινδύνων, θα ήταν επίσης πιο οικονομικά αποδοτική σε σχέση με τις μεμονωμένες χώρες που αναπτύσσουν τα δικά τους μοναδικά συστήματα. Τουλάχιστον, στην περίπτωση μετεωρολογικών κινδύνων, η ανταλλαγή δεδομένων μεταξύ χωρών είναι απαραίτητη, καθώς αυτοί οι κίνδυνοι δεν περιορίζονται σε γεωγραφικά όρια και θα πρέπει το βάρος του κόστους για τις παρατηρήσεις αυτών των φαινομένων, να μπορεί να διαμοιραστεί.

- **Εκτίμηση κόστους - οφέλους των συστημάτων έγκαιρης προειδοποίησης – Πλημμύρες**

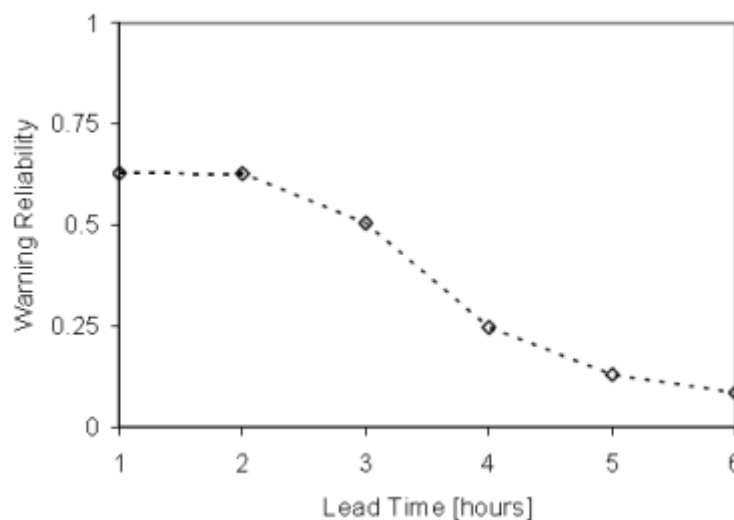
Η ποσοτικοποίηση της αποτελεσματικότητας κόστους των EWS αναγνωρίζεται ότι είναι δύσκολη και ως εκ τούτου δεν γίνεται συχνά. Ένα πρόβλημα είναι ότι πολλοί διαφορετικοί παράγοντες, (πχ. καταστροφές της περιουσίας σε σχέση με τον αριθμό των θανάτων) όπου σπάνια είναι εφικτό να περιληφθούν στις αναλύσεις καθώς δεν είναι εύκολο να ποσοτικοποιηθούν και να συσχετισθούν.



Οι Schröter et al. [27] έχουν αξιολογήσει την αποτελεσματικότητα και την αποδοτικότητα των EWS για πλημμύρες που αφορούν μικρές λεκάνες απορροής ποταμού. Η ανάλυσή τους δίνει πληροφορίες για την ανάπτυξη βέλτιστων προειδοποιήσεων μέσα από την ανάλυση της αντιστάθμισης των πλεονεκτημάτων μίας γρήγορης προειδοποίησης και ταυτόχρονα μίας μείωσης της αξιοπιστίας προειδοποίησης. Η προσέγγισή τους θεωρεί ότι η γρήγορη προειδοποίηση παρέχει πολύτιμο χρόνο για την ολοκλήρωση των προληπτικών μέτρων, ενώ η μείωση της αξιοπιστίας προειδοποίησης θα προκαλέσει οικονομική απώλεια σε περίπτωση εσφαλμένης ειδοποίησης. Η μεθοδολογία έχει εφαρμοστεί σε δύο λεκανοπέδια στην Αυστρία και στην Ισπανία.

Για την αξιολόγηση της αποτελεσματικότητας του EWS, συνέδεσαν την αξιοπιστία της πρόβλεψης με το οικονομικό όφελος ως συνάρτηση μίας γρήγορης προειδοποίησης. Η αξιοπιστία πρόβλεψης καθορίζεται από μία ανάλυση ενός συνολού προβλέψεων για την εμφάνιση πλημμύρας, το οποίο περιγράφει την αβεβαιότητα λόγω των παραμέτρων του μοντέλου πρόβλεψης βροχόπτωσης και τη δομή μοντέλου. Το οικονομικό όφελος προκύπτει από την αποφυγή ζημιών μέσω της χρήσης του EWS.

Η πρόβλεψη πλημμυρών, όπως και οι περισσότερες υδρομετεωρολογικές προβλέψεις, περιλαμβάνει μία σημαντική αβεβαιότητα επειδή η γνώση της μελλοντικής κατάστασης της ατμόσφαιρας και της κατάστασης και της συμπεριφοράς του υδρολογικού συστήματος είναι ακόμη περιορισμένη. Ένα αναπόσπαστο μέτρο της αξιοπιστίας της πρόβλεψης και της προειδοποίησης βασίζεται στην επεξήγηση των σφαλμάτων πρόβλεψης πλημμυρών που προέκυψαν από αναλύσεις προηγούμενων γεγονότων. Οι πληροφορίες για τα σφάλματα πρόβλεψης μεταφέρονται σε ένα μέτρο αξιοπιστίας προειδοποίησης ως συνάρτηση μίας έγκαιρης προειδοποίησης ([Εικόνα 1.3](#)).

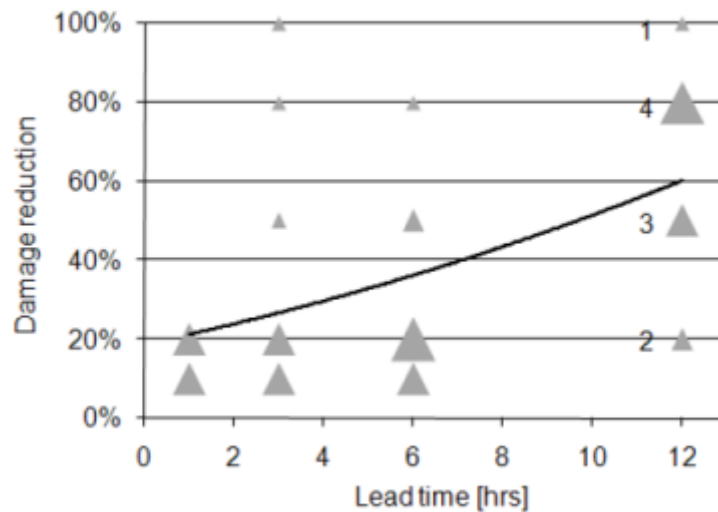


Εικόνα 1.3: Αξιοπιστία προειδοποίησης ως συνάρτηση του χρόνου προειδοποίησης [27]

Ο προσδιορισμός της ικανότητας μείωσης των ζημιών από τις πλημμύρες βασίζεται σε έρευνα των χρηστών. Η πιο σημαντική ερώτηση για την εκτίμηση του οφέλους μιας ειδοποίησης είναι «Υποτίθεται ότι λαμβάνετε μια ειδοποίηση λίγο πριν από μία απρόσμενη πλημμύρα, σε ποιο ποσοστό θα μπορούσατε να μειώσετε καταστροφικές συνέπειες» [27]. Τα αποτελέσματά δείχνουν συσχέτιση μεταξύ της ετοιμότητας και της αποτελεσματικότητας των

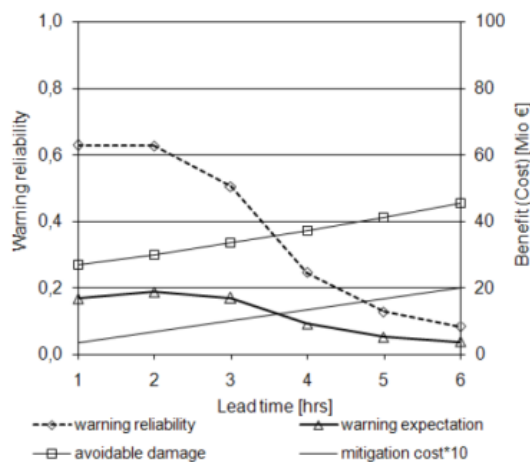


ενεργειών μετριασμού. Σε περίπτωση ξαφνικής πλημμύρας, μία προειδοποίηση που ανακοινώνεται δώδεκα ώρες πριν παρέχει πιθανή μείωση κατά 60% στις ζημιές, ενώ μία προειδοποίηση που ανακοινώνεται μία ώρα πριν προλαμβάνει το 20% των επερχόμενων ζημιών (Εικόνα 1.4). Το μέγεθος των τριγώνων στην εικόνα αντικατοπτρίζει το μέτρο της συχνότητας των απαντήσεων στο ερωτηματολόγιο. Ενώ το μέγεθος του δείγματος είναι σχετικά μικρό, εξυπηρετεί το σκοπό του αποδεικνύοντας ότι υπάρχει μια προφανής συσχέτιση μεταξύ της ετοιμότητας και της αποτελεσματικότητας των μέτρων πρόληψης.



Εικόνα 1.4: Μείωση ζημιών ως συνάρτηση του χρόνου προειδοποίησης [27]

Στην Εικόνα 1.5, παρουσιάζεται η αναμονή προειδοποίησης ως δείκτης για το βέλτιστη ειδοποίηση σε γενική μορφή. Σύμφωνα με αυτή, η πιθανή μείωση ζημιών μειώνεται συνεχώς για μικρότερους χρόνους προειδοποίησης. Η γραμμή στο κάτω μέρος του γραφήματος εισάγει το κόστος για τα μέτρα πρόληψης από την άποψη της χαμένης καθαρής αξίας της παραγωγής (κλίμακα 10). Αυτή η καμπύλη δείχνει το κόστος ανά ώρα που προκύπτει εάν τα ενεργά άτομα διακόψουν την παραγωγική εργασία και στραφούν σε προληπτικά μέτρα.



Εικόνα 1.5: Αναμονή Προειδοποίησης ως ένδειξη ενός βέλτιστου συναγερμού [27]



Λαμβάνοντας υπόψη το σημαντικό κόστος των απωλειών, οι οποίες που σχετίζονται με μία ειδοποίηση, θα πρέπει να προβληματιστούμε σχετικά με την ενεργοποίηση μίας ειδοποίησης για ένα συμβάν πλημμύρας που δεν είναι βέβαιο εάν θα συμβεί, τη στιγμή που θα έπρεπε να εκδοθεί ειδοποίηση. Η αναμονή μιας ειδοποίησης ορίζεται ως το προϊόν της αξιοπιστίας της προειδοποίησης σε συνάρτηση με τη ζημιά.

Η καμπύλη που προκύπτει, με μονάδες € ανά ειδοποίηση, δίνεται ως έντονη γραμμή στην [Εικόνα 1.5](#). Η προσδοκία προειδοποίησης δεν είναι σταθερή αλλά αλλάζει με το χρόνο. Το μέγιστο της καμπύλης προσδοκίας προειδοποίησης ορίζει τη βέλτιστη χρονική στιγμή για την ανακοίνωση μίας ειδοποίησης ανάλογα με την αξιοπιστία και τις συνέπειες αυτής [28].



2 Θεωρητικά Στοιχεία

2.1 Ορισμοί

2.1.1 Ταξινόμηση

Ως ταξινόμηση ορίζεται η τεχνική εξόρυξης δεδομένων κατά την οποία ένα στοιχείο ανατίθεται σε ένα προκαθορισμένο σύνολο κατηγοριών. Και ενώ για τον ανθρώπινο εγκέφαλο είναι μία πράξη της καθημερινότητας, όταν αυτή η πράξη πάει να γίνει από έναν υπολογιστή τα πράγματα γίνονται πιο σύνθετα. Για παράδειγμα, ο χαρακτηρισμός ενός εστιατορίου ως καλό ή κακό δεν σημαίνει κάτι για ένα μηχάνημα. Αν όμως στη θέση του καλού, μπει η βαθμολογία 8 και στη θέση του κακού η βαθμολογία 3, τότε μπορεί να δημιουργηθεί μία ακολουθία αριθμών που θα κατατάσσει τα εστιατόρια με βάση πχ. την ποιότητα τους. Προκειμένου λοιπόν να επιτευχθεί η ταξινόμηση δεδομένων με τη χρήση υπολογιστή, ένας αλγόριθμος ταξινόμησης μεταθέτει τα στοιχεία μιας ακολουθίας ώστε να είναι σε συγκεκριμένη σειρά. Στις διάφορες τεχνικές ταξινόμησης χρησιμοποιούνται ακολουθίες διανυσμάτων, ώστε αν ένας αλγόριθμος εκπαιδευτεί σε ένα σύνολο δεδομένων εκπαίδευσης (training set) να μπορεί να κάνει προβλέψεις και έξω από αυτό. Η διαδικασία για την εξαγωγή των απαραίτητων κανόνων μέσα από τις οποίες θα πραγματοποιηθεί η ταξινόμηση, αποκαλείται διάκριση (discrimination) [29]. Η ταξινόμηση (classification) είναι η διαδικασία κατά την οποία ο αλγόριθμος αποφασίζει αν μία συγκεκριμένη παρατήρηση είναι κομμάτι μίας μεγαλύτερης ομάδας ή κλάσης [30].

2.1.2 Η έννοια του αλγορίθμου

Μία από τις πιο διαδεδομένες έννοιες στον τομέα της πληροφορικής είναι αυτή του αλγορίθμου. Πρόκειται για μία πεπερασμένη σειρά ενεργειών, αυστηρά καθορισμένων και εκτελέσιμων σε πεπερασμένο χρόνο που στόχο έχουν την επίλυση ενός προβλήματος. Η χρήση αλγορίθμων μηχανικής μάθησης και στατιστικής ανάλυσης στην ταξινόμηση παρέχει πολλά πλεονεκτήματα. Δύο από τους βασικούς λόγους για τους οποίους επιλέγονται, είναι οι δυνατότητες που μας παρέχουν για τη διαχείριση μεγάλων συνόλων δεδομένων (big data), μέσω της ικανότητας τους για επανάληψη εντολών, αλλά και τη λύση στο πρόβλημα του υπέρ-επιπέδου. Στο πλαίσιο αυτής της εργασίας θα χρησιμοποιηθούν τρεις αλγόριθμοι, οι οποίοι και θα αναλυθούν στη συνέχεια.

2.1.3 Μηχανική Μάθηση

Ο όρος μηχανική μάθηση (machine learning) αναφέρεται στο πως ο υπολογιστής μαθαίνει να χρησιμοποιεί προγράμματα για την ανάλυση δεδομένων. Συγκεκριμένα στον κλάδο της ταξινόμησης, χρησιμοποιούνται αυτοί οι αλγόριθμοι για να πραγματοποιηθεί ταξινόμηση των δεδομένων σε κλάσεις ανάλογα με τα χαρακτηριστικά τους, έχοντας λάβει υπόψιν προηγουμένως παρόμοια δεδομένα. Με βάση λοιπόν τη δυνατότητα των υπολογιστών για ανάκληση μεγάλου όγκου δεδομένων, δύναται να δημιουργηθούν έξυπνες αυτοματοποιημένες λειτουργίες για την αντιμετώπιση ενός σημαντικού αριθμού των προβλημάτων ταξινόμησης - διάκρισης της καθημερινότητας (στον



κλάδο της τεχνολογίας), ή ακόμη και πιο περίπλοκες διαδικασίες που σχετίζονται με την τεχνητή νοημοσύνη. Ανάλογα, λοιπόν, με τα προς επεξεργασία δεδομένα οι κύριες κατηγορίες αλγορίθμων μηχανικής μάθησης που μπορούν να χρησιμοποιηθούν είναι οι εξής: οι αλγόριθμοι επιτηρούμενης μάθησης, οι αλγόριθμοι μη επιτηρούμενης μάθησης και οι αλγόριθμοι ενισχυτικής μάθησης.

2.1.4 Επιτηρούμενη Μάθηση

Η επιτηρούμενη μάθηση (supervised learning) αποτελεί την πιο αποτελεσματική αλλά και πιο συνηθισμένη χρήση του συνόλου δεδομένων εκπαίδευσης. Στην τεχνική αυτή, ο αλγόριθμος εκπαιδεύεται από ένα σύνολο δεδομένων εκπαίδευσης, για το οποίο γνωρίζει ήδη τις ταμπέλες κλάσης για κάθε παρατήρηση. Συγκεκριμένα, κάθε παρατήρηση διαθέτει κάποιο ονομαστικό χαρακτηριστικό το οποίο δίνει και την κλάση του στο στοιχείο αυτό. Εφόσον γνωρίζει ο αλγόριθμος την κλάση, δύναται να καταλάβει τις ομοιότητες μεταξύ παρατηρήσεων από την ίδια κλάση και επομένως να προκύψει η αποτελεσματικότητα του αλγορίθμου αλλά και των επιμέρους εργαλείων ταξινόμησης [31]. Στην επιτηρούμενη μάθηση λοιπόν, το σύστημα οφείλει να μάθει με επαγωγική μέθοδο μια συνάρτηση, τη «συνάρτηση στόχο», που συνιστά έκφραση του μοντέλου που περιγράφει τα δεδομένα. Σκοπός της συνάρτησης «στόχος» είναι η πρόβλεψη της εξαρτημένης μεταβλητής ή αλλιώς μεταβλητής εξόδου.

Τα δύο βασικά είδη προβλημάτων εξόρυξης γνώσης που καλείται να αντιμετωπίσει η επιτηρούμενη μάθηση, είναι αυτά της ταξινόμησης και της παρεμβολής. Όσον αφορά την ταξινόμηση αυτή αναφέρεται στη δημιουργία μοντέλων πρόβλεψης διακριτών τάξεων, ενώ όσον αφορά την παρεμβολή, αναφέρεται στη δημιουργία μοντέλων πρόβλεψης αριθμητικών τιμών. Τις βασικότερες τεχνικές επιτηρούμενης μάθησης αποτελούν οι:

- Γραμμική Παρεμβολή
- Μάθηση κατά Bayes
- Δένδρα Απόφασης
- Μάθηση κανόνων
- Μάθηση εννοιών
- Νευρωνικά Δίκτυα
- Μηχανές Διανυσμάτων Υποστήριξης
- Μάθηση κατά περίπτωση

2.2 Προγνωστική Μοντελοποίηση (Predictive Modelling)

Η Προγνωστική Ανάλυση (Predictive Analytics) αποτελεί τομέα της στατιστικής ανάλυσης που ασχολείται με την εξαγωγή πληροφοριών από δεδομένα και χρησιμοποιείται για την πρόβλεψη τάσεων και συμπεριφορών. Η Προγνωστική Ανάλυση συχνά ορίζεται ως η πρόβλεψη σε πιο λεπτομερές επίπεδο ευαισθησίας, δηλ. παραγωγή προγνωστικών αποτελεσμάτων (πιθανοτήτων) για κάθε παρεχόμενο στοιχείο, ξεχωριστά, και αυτό διαχωρίζει την προγνωστική ανάλυση από την πρόγνωση (forecasting). Η προστιθέμενη αξία αυτού του είδους ανάλυσης είναι ότι



αποσκοπεί στην πρόβλεψη και αποτροπή σημαντικών ζητημάτων για την ελαχιστοποίηση κάθε είδους καταστροφών και ζημιών και μελλοντικά στην ενσωμάτωση των εξαγόμενων συμπερασμάτων σε προδιαγραφές ανάλυσης για βελτιστοποίηση αποφάσεων.

Στις τεχνικές της προγνωστικής ανάλυσης περιλαμβάνονται διαδικασίες όπως η μοντελοποίηση δεδομένων, η μηχανική μάθηση (machine learning) – η οποία επιστρατεύθηκε στην παρούσα εργασία-, η τεχνητή νοημοσύνη (AI), οι αλγόριθμοι βαθιάς μάθησης (deep learning algorithms) και η εξόρυξη δεδομένων (data mining) [32]. Γενικά, το άγνωστο γεγονός ενδιαφέροντος είναι κάτι μελλοντικό, ωστόσο η προγνωστική ανάλυση μπορεί να εφαρμοσθεί σε οποιοδήποτε είδος άγνωστου γεγονότος, το οποίο μπορεί να ανήκει είτε στο παρελθόν, είτε στο παρόν είτε στο μέλλον. Η Προγνωστική Ανάλυση βασίζεται στη συσχέτιση μεταξύ των επεξηγηματικών μεταβλητών και των προβλεπόμενων μεταβλητών, η γνώση των οποίων βασίζεται σε παρελθοντικά γεγονότα, καθώς και στην αξιοποίησή τους για την πρόβλεψη του άγνωστου αποτελέσματος. Σε κάθε περίπτωση, η ακρίβεια και η χρηστικότητα των αποτελεσμάτων εξαρτάται σε μεγάλο βαθμό από το επίπεδο ανάλυσης των δεδομένων και την ποιότητα των υποθέσεων.

Η αυστηρή χρήση του όρου Προγνωστική Ανάλυση υποδηλώνει την προγνωστική μοντελοποίηση (predictive modeling), δηλ. τη “βαθμολόγηση” δεδομένων με προγνωστικά μοντέλα και την πρόγνωση. Παρόλα αυτά, ο όρος χρησιμοποιείται ευρύτερα όταν πρόκειται για σχετικούς αναλυτικούς κλάδους, όπως περιγραφική μοντελοποίηση (descriptive modeling) και μοντελοποίηση αποφάσεων (decision modeling) ή βελτιστοποίηση (optimization). Πρέπει, ωστόσο, να σημειωθεί ότι οι στατιστικές τεχνικές, στις οποίες βασίζεται κάθε κλάδος, ποικίλλουν.

2.2.1 Προγνωστικά Μοντέλα

Η προγνωστική μοντελοποίηση χρησιμοποιεί μοντέλα πρόβλεψης για την ανάλυση της σχέσης μεταξύ της συγκεκριμένης απόδοσης ενός στοιχείου σε ένα δείγμα και ενός ή περισσότερων γνωστών χαρακτηριστικών του συγκεκριμένου στοιχείου. Ο στόχος των μοντέλων πρόβλεψης είναι να εκτιμήσουν την πιθανότητα να εμφανιστεί η συγκεκριμένη απόδοση, όταν ένα παρόμοιο στοιχείο δοκιμαστεί σε ένα διαφορετικό δείγμα. Ένα από τα μεγαλύτερα πλεονεκτήματα των προγνωστικών μοντέλων είναι η ικανότητά τους να εισάγουν πολλές παραμέτρους [33]. Τα μοντέλα πρόβλεψης εκτελούν συχνά υπολογισμούς κατά τη διάρκεια ζωντανών γεγονότων, προκειμένου να καθοδηγήσουν μία απόφαση.

Οι πιο διαδεδομένοι τύποι προγνωστικών μοντέλων είναι τα μοντέλα πρόγνωσης (forecast models) και τα μοντέλα ταξινόμησης (classification models) [33]. Τα μοντέλα πρόγνωσης είναι πολύ δημοφιλή καθώς είναι πολύ ευπροσάρμοστα και αντιμετωπίζουν την πρόβλεψη μετρικών τιμών εκτιμώντας τις τιμές των νέων δεδομένων με βάση τις γνώσεις που αποκομίζουν από ιστορικά δεδομένα. Συχνά χρησιμοποιούνται για τη δημιουργία αριθμητικών τιμών σε ιστορικά δεδομένα όταν δεν υπάρχει καμία καταγραφή. Αναφορικά με τα μοντέλα ταξινόμησης λειτουργούν κατηγοριοποιώντας πληροφορίες βάσει ιστορικών δεδομένων και έχουν τη δυνατότητα να



επανεκπαιδευτούν εύκολα με νέα δεδομένα παρέχοντας μία ευρεία ανάλυση για την απάντηση ερωτήσεων. Στη συγκεκριμένη εργασία θα χρησιμοποιηθούν μοντέλα ταξινόμησης, για το λόγο αυτό ακολουθεί περαιτέρω ανάπτυξη τους στην παράγραφο §2.2.

Όταν εργαζόμαστε σε προβλήματα ταξινόμησης, κρίνεται απαραίτητη η συλλογή ενός συνόλου δεδομένων εκπαίδευσης. Τα διαθέσιμα σύνολα δεδομένων δειγμάτων με γνωστά χαρακτηριστικά -μετρήσεις (input data) και γνωστές επιδόσεις – ετικέτα κλάσης (output data) αναφέρονται ως «σύνολα δεδομένων εκπαίδευσης» (training datasets). Το σύνολο δεδομένων εκπαίδευσης χρησιμοποιείται για την καλύτερη κατανόηση των δεδομένων εισαγωγής ώστε να είναι καλύτερα προετοιμασμένα για τη μοντελοποίηση. Είναι επίσης χρήσιμο για την αξιολόγηση μίας σειράς διαφορετικών αλγορίθμων μοντελοποίησης ενώ σε κάποιες περιπτώσεις, χρησιμοποιείται για το συντονισμό των υπερπαραμέτρων ενός επιλεγμένου μοντέλου. Τέλος, το σύνολο δεδομένων εκπαίδευσης χρησιμοποιείται για την εκπαίδευση ενός τελικού μοντέλου, σε όλα τα διαθέσιμα δεδομένα που ενδέχεται να χρησιμοποιηθούν στο μέλλον για να γίνουν προβλέψεις σε νέα δεδομένα που πρόκειται να ληφθούν [34]. Τα σύνολα δεδομένων δειγμάτων, με γνωστά χαρακτηριστικά αλλά άγνωστες επιδόσεις, αναφέρονται ως «σύνολα δεδομένων δοκιμής» (testing datasets). Τα σύνολα εκπαίδευσης και δοκιμής δεν φέρουν απαραίτητα χρονολογική συσχέτιση μεταξύ τους.

2.3 Προγνωστική Μοντελοποίηση Ταξινόμησης (Classification Predictive Modeling)

Ένα πρόβλημα προγνωστικής μοντελοποίησης ταξινόμησης μπορεί να έχει δύο (δυναμική ταξινόμηση) ή περισσότερες – μέχρι και εκατοντάδες - ετικέτες τάξεις (πρόβλημα ταξινόμησης πολλαπλών τάξεων). Στη συγκεκριμένη εργασία, το πρόβλημα που αντιμετωπίστηκε ήταν δυναμικής ταξινόμησης, όλα τα δεδομένα δηλ. ανήκουν σε μία από τις δύο τάξεις / κατηγορίες.

2.3.1 Αξιολόγηση δυαδικών ταξινομητών (binary classifiers)

Υπάρχουν διάφορα μεγέθη που μπορούν να χρησιμοποιηθούν για την εκτίμηση της απόδοσης ενός ταξινομητή, καθώς ανάλογα με το πεδίο ενδιαφέροντος υπάρχουν διαφορετικές προτιμήσεις για συγκεκριμένες μετρήσεις λόγω διαφορετικών στόχων. Έστω ένα πρόβλημα πρόβλεψης δύο κατηγοριών (δυναμική ταξινόμηση), στο οποίο η πραγματική κατάσταση μπορεί να είναι «θετική» (1) (P) ή «αρνητική» (0) (N) και τα προβλεπόμενα αποτελέσματα επισημαίνονται είτε ως «θετικά» (P) είτε «αρνητικά» (N). Ο δυαδικός ταξινομητής μπορεί να δώσει τέσσερα πιθανά αποτελέσματα:

- Εάν το αποτέλεσμα από μία πρόβλεψη είναι (P) (1) και η πραγματική κατάσταση είναι (P) (1), τότε το αποτέλεσμα είναι αληθώς θετικό (*true positive* - TP).
- Εάν το αποτέλεσμα από μία πρόβλεψη είναι (N) (0) και η πραγματική κατάσταση είναι (N) (0), τότε το αποτέλεσμα είναι αληθώς αρνητικό (*true negative* - TN).



- Εάν το αποτέλεσμα από μία πρόβλεψη είναι (P) (1) και η πραγματική κατάσταση είναι (N) (0), τότε το αποτέλεσμα είναι ψευδώς θετικό (*false positive* - FP) και ονομάζεται Σφάλμα Τύπου I (*Type I Error* – αντ. σε *false alarm*).
- Εάν το αποτέλεσμα από μία πρόβλεψη είναι (N) (0) και η πραγματική κατάσταση (P) (1), τότε το αποτέλεσμα είναι ψευδώς αρνητικό (*false negative* - FN) και ονομάζεται Σφάλμα Τύπου II (*Type II Error* – αντ. σε *miss*).

Τα παραπάνω μπορούν να αναπαρασταθούν σε έναν πίνακα ενδεχομένων 2×2 (*contingency table/ confusion matrix*) που χρησιμοποιείται για την οπτικοποίηση της απόδοσης ενός αλγορίθμου, με στήλες που αντιστοιχούν στην πραγματική κατάσταση και σειρές που αντιστοιχούν στις παραγόμενες προβλέψεις:

		ΠΡΑΓΜΑΤΙΚΗ ΚΑΤΑΣΤΑΣΗ	
		Κατάσταση αληθής	Κατάσταση ψευδής
ΠΡΟΒΛΕΨΕΙΣ	Πρόβλεψη Θετική	TP	FP
	Πρόβλεψη Αρνητική	FN	TN

Πίνακας 2.1: Πίνακας Ενδεχομένων / Απόδοσης

Από τις στήλες του Πίνα. 2.1 προκύπτουν οι εξής αναλογίες:

ΑΛΗΘΩΣ ΘΕΤΙΚΟΣ ΡΥΘΜΟΣ:

$$(TPR) = \frac{TP}{TP+FN} \quad (2.1)$$

Ο αληθώς θετικός ρυθμός ονομάζεται *ευαισθησία ή ανάκληση (sensitivity / recall)* και αποτελεί το ποσοστό του πληθυσμού με την κατάσταση για την οποία η δοκιμή είναι σωστή.

ΑΛΗΘΩΣ ΑΡΝΗΤΙΚΟΣ ΡΥΘΜΟΣ:

$$(TNR) = \frac{TN}{TN+FP} \quad (2.2)$$

Ο αληθώς αρνητικός ρυθμός είναι γνωστός ως *ειδικότητα / επιλεκτικότητα (specificity / selectivity) (SPC)*.

Συμπληρωματική της (2.1) είναι ο:

ΨΕΥΔΩΣ ΑΡΝΗΤΙΚΟΣ ΡΥΘΜΟΣ:

$$(FNR) = \frac{FN}{TP+FN} = 1 - (TPR) \quad (2.3)$$

Ο ψευδώς αρνητικός ρυθμός ονομάζεται *ποσοστό απώλειας (miss rate)*.



Συμπληρωματική της (2.2) είναι η ο:

ΨΕΥΔΩΣ ΘΕΤΙΚΟΣ ΡΥΘΜΟΣ:

$$(FPR) = \frac{FP}{TN+FP} = 1 - (TNR) \quad (2.4),$$

Ο ψευδώς θετικός ρυθμός ονομάζεται *λάθος εκτίμηση (fall - out)*.

Από τις σειρές του [Πίν. 2.1](#) προκύπτουν οι εξής αναλογίες:

ΘΕΤΙΚΗ ΤΙΜΗ ΠΡΟΓΝΩΣΗΣ:

$$(PPV) = \frac{TP}{TP+FP} \quad (2.5)$$

Η θετική τιμή πρόγνωσης ονομάζεται *ακρίβεια (precision)* και αποτελεί το ποσοστό του πληθυσμού για το οποίο υπάρχει σωστά θετική πρόβλεψη, στο σύνολο όλων των παρατηρήσεων που έχουν θετική πρόβλεψη ή εναλλακτικά το βαθμό στον οποίο οι παραγόμενες μετρήσεις δείχνουν τα ίδια αποτελέσματα, υπό αμετάβλητες συνθήκες.

ΑΡΝΗΤΙΚΗ ΤΙΜΗ ΠΡΟΓΝΩΣΗΣ:

$$(NPV) = \frac{TN}{TN+FN} \quad (2.6)$$

Συμπληρωματική της (2.5) είναι το:

ΠΟΣΟΣΤΟ ΨΕΥΔΟΥΣ ΑΝΑΚΑΛΥΨΗΣ:

$$(FDR) = \frac{FP}{TP+FP} = 1 - (PPV) \quad (2.7)$$

Συμπληρωματική της (2.6) είναι το:

ΠΟΣΟΣΤΟ ΨΕΥΔΟΥΣ ΠΑΡΑΛΕΙΨΗΣ:

$$(FOR) = \frac{FN}{TN+FN} = 1 - (NPV) \quad (2.8)$$

Στη διαγνωστική δοκιμή χρησιμοποιούνται κυρίως οι αναλογίες (2.1) και (2.2), ευαισθησία (sensitivity) και ειδικότητα (specificity) αντίστοιχα. Στην ανάκτηση πληροφοριών, οι κύριοι λόγοι που χρησιμοποιούνται είναι οι αληθώς θετικοί λόγοι, δηλ. οι αναλογίες (2.1) και, (2.5) η ανάκληση (recall) και η ακρίβεια (precision), αντίστοιχα.

Οι αναλογίες των συμπληρωματικών ζευγών αναλογιών, αποδίδουν τέσσερις αναλογίες πιθανότητας:

ΡΥΘΜΟΣ ΘΕΤΙΚΗΣ ΠΙΘΑΝΟΤΗΤΑΣ:

$$(LR^+) = \frac{TPR}{FPR} \quad (2.9)$$

ΡΥΘΜΟΣ ΑΡΝΗΤΙΚΗΣ ΠΙΘΑΝΟΤΗΤΑΣ:



$$(LR^-) = \frac{FNR}{TNR} \quad (2.10)$$

Η αναλογία της (2.9) προς την (2.10) την αναλογία διαγνωστικών αποδόσεων (DOR).

Επιπλέον μεγέθη που απεικονίζουν την απόδοση των δυαδικών κατανομών είναι τα εξής:

ΑΚΡΙΒΕΙΑ

$$(ACC) = \frac{TP+TN}{TP+TN+FP+FN} \quad (2.11)$$

Η ακρίβεια (*accuracy*) αποτελεί το λόγο των σωστά προβλεπόμενων παρατηρήσεων προς το σύνολο των παρατηρήσεων ή εναλλακτικά είναι η εγγύτητα των αποτελεσμάτων μέτρησης στην πραγματική τιμή. Η διαφορά μεταξύ των μεγεθών (ACC) και (PPV) ότι το μεν πρώτο αναφέρεται στην απόσταση των παραγόμενων αποτελεσμάτων από μια συγκεκριμένη τιμή (*systemic errors*), ενώ το δεύτερο αναφέρεται στην απόσταση των παραγόμενων αποτελεσμάτων μεταξύ τους (*random errors*). Στην περίπτωση που η ακρίβεια είναι πολύ κοντά στη μονάδα (υψηλή ακρίβεια) μπορεί να θεωρηθεί ότι το εξεταζόμενο μοντέλο είναι καλό μόνο στην περίπτωση που τα σύνολα δεδομένων είναι συμμετρικά, όπου ουσιαστικά τα σύνολα FP και FN είναι σχεδόν ίδια. Σε κάθε άλλη περίπτωση θα πρέπει να επιστρατευτούν και άλλες παράμετροι προκειμένου να αξιολογηθεί η ποιότητα του μοντέλου.

F1 score

$$F1 \text{ score} = \frac{2*(TPR*PPV)}{TPR+PPV} \quad (2.11)$$

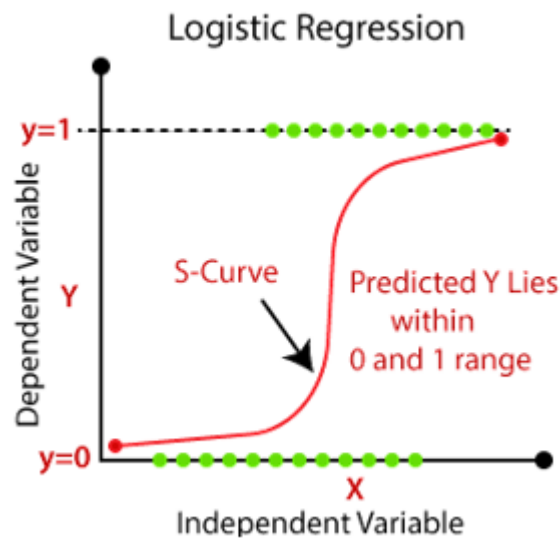
Το F1 score αποτελεί ένα μέτρο της ακρίβειας μιας δοκιμής και θεωρείται ο σταθμισμένος μέσος όρος της ακρίβειας και της ανάκλησης. Λαμβάνει τιμές από 0 έως 1, όπου ως 1 θεωρείται η τέλεια ακρίβεια και ανάκληση. Θεωρείται πιο ενδεικτικό μέγεθος σε σχέση με την ακρίβεια, ειδικά στην περίπτωση των άνισων κατανομών τάξης.

2.4 Ταξινομητές

2.4.1 Λογιστική Παλινδρόμηση

Μία μορφή Ανάλυσης Παλινδρόμησης αποτελεί η Λογιστική Παλινδρόμηση (Logistic Regression), με τη διαφορά ότι σε αντίθεση με τις άλλες μεθόδους, το αποτέλεσμα είναι διακριτό, στη μορφή μίας κατηγορικής μεταβλητής που αντλεί τις τιμές από ένα διακριτό σύνολο. Εάν μάλιστα αφορά δυαδική ταξινόμηση, τότε το σύνολο περιορίζεται σε δύο πιθανά αποτελέσματα. Η Λογιστική Παλινδρόμηση λοιπόν, είναι ικανή να πραγματοποιήσει δυαδική ταξινόμηση καθώς χρησιμοποιεί λογαριθμική συνάρτηση, με την οποία οι εισερχόμενες μεταβλητές που μπορεί να έχουν συνεχείς τιμές σε όλο το σύνολο αριθμών, ανάγονται σε τιμές στο πεδίο [0, 1], και αποδίδουν τιμές πάνω του 0.5 για τη μία κλάση και κάτω του 0.5 για την άλλη.

Στην [Εικόνα 2.1](#), παρουσιάζεται η γραφική αναπαράσταση της μεθόδου Logistic Regression:



Εικόνα 2.1: Γραφική αναπαράσταση του Logistic Regression

Για να υπολογιστούν οι πιθανότητες της κάθε κλάσης, η Logistic Regression κάνει αρχικά χρήση μίας γραμμικής συνάρτησης πρόβλεψης, η οποία συνίσταται από ένα γραμμικός συνδυασμό όλων των χαρακτηριστικών των παραδειγμάτων πολλαπλασιασμένα με τους κατάλληλους συντελεστές. Έστω λοιπόν ότι υπάρχουν τα χαρακτηριστικά $x_1, x_2, x_3, \dots, x_n$, η συνάρτηση θα είναι της μορφής:

$$f(x) = b_0 + b_1x_1 + b_2x_2 + \dots \quad (2.11)$$

$$f(x) = b$$

με τα b να είναι οι συντελεστές των χαρακτηριστικών (ανεξάρτητες μεταβλητές). Επόμενο βήμα είναι να εξισωθεί η $f(x)$ με το λογάριθμο του μέτρου της απόδοσης:

$$\ln(odds) = f(x) \quad (2.12)$$

Η απόδοση είναι συνδεδεμένη με την σχετική πιθανότητα και ορίζεται ως $p/(1-p)$, με p να είναι η πιθανότητα (εξαρτημένη μεταβλητή) η παρατήρηση να ανήκει στην κλάση 1. Αν θεωρηθούν τα δύο μέλη της εξίσωσης ως δυνάμεις του e , προκύπτει λοιπόν η σχέση:

$$odds = e^{f(x)} \quad (2.13)$$

Προκειμένου να εμφανιστεί η πιθανότητα, πραγματοποιείται διαίρεση με το $(1+odds)$ και προκύπτει ο τύπος που μας δίνει την πιθανότητα:

$$p = \frac{e^{f(x)}}{1+e^{f(x)}} \quad (2.14)$$



Η τιμή που μας επιστρέφει αυτός ο τύπος είναι στο πεδίο τιμών $[0, 1]$, ανεξάρτητα από την τιμή που παίρνει η $f(x)$, που αποτελεί γραμμικό συνδυασμό των χαρακτηριστικών. Ωστόσο οι συντελεστές αρχικά είναι τυχαίοι αριθμοί και δεν αναπαριστούν τα πραγματικά βάρη που οφείλει να έχει το κάθε χαρακτηριστικό. Με χρήση όμως της εκτίμησης μέγιστης πιθανότητας, υπάρχει ανανέωση των συντελεστών μέχρι να προσεγγίσουν κάποιο ιδανικό επίπεδο.

Γίνεται μία εφαρμογή γραμμικής παλινδρόμησης στα χαρακτηριστικά, και στη συνέχεια μέσω της σιγμοειδή λογαριθμικής συνάρτησης προκύπτει το αποτέλεσμα της ταξινόμησης. Για κάθε παρατήρηση που προστίθεται, τα χαρακτηριστικά ρυθμίζουν την $f(x)$ και ανάλογα αν το αποτέλεσμα είναι πάνω ή κάτω του 0.5, αντιστοιχεί σε μία από τις δύο κλάσεις.

Σε γενικές γραμμές, η Logistic Regression με την Linear Discriminant Analysis μοιάζουν αρκετά, όμως η Logistic Regression λόγω του ότι παράγει λογικές τιμές πιθανότητας αλλά και δεν προχωρά στις υποθέσεις της Linear Discriminant Analysis, αναφέρεται σαν καλύτερη μέθοδος. Παρόλα αυτά αν είναι δεδομένα πολλών κλάσεων, η Linear Discriminant Analysis είναι καταλληλότερη, καθώς για να λειτουργήσει η Logistic Regression χωρίς προβλήματα για πολλές κλάσεις, απαιτείται η εισαγωγή περισσότερων εξαρτημένων μεταβλητών, όπου όποια έχει την τιμή 1 είναι και η κλάση που ανήκει το παράδειγμα.

Προκείμενου να επιτευχθούν ακόμη καλύτερα αποτελέσματα με την Logistic Regression, δύναται να χρησιμοποιηθούν οι υποθέσεις ανεξαρτησίας του αλγορίθμου NB [35]. Η πραγματοποίηση αυτής της ιδέας, γίνεται με την μέθοδο χωριζόμενης Λογιστικής Παλινδρόμησης, με την κατασκευή πολλών μοντέλων Logistic Regression όπου το καθένα εκπαιδεύεται σε διαφορετικά μέρη του feature space, όμως με τις ίδιες παρατηρήσεις. Ο συνδυασμός των δύο μεθόδων, έχει δείξει ότι δίνει μεγαλύτερη ακρίβεια και από τις δύο μεθόδους που τον συνιστούν. Ο Logistic Regression λοιπόν, σε μία έρευνα που τη συνέκρινε με τον NB, φανέρωσε ότι ασυμπτωτικά είναι πιθανό να οδηγήσει σε ακριβέστερα αποτελέσματα. Παρόλα αυτά τα μοντέλα διάκρισης (βλέπε Logistic Regression, Linear Discriminant Analysis), απαιτούν τουλάχιστον ένα πλήθος δεδομένων εκπαίδευσης, για να είναι ακριβέστερα σε σχέση με τα γενετικά, οπότε δεν προτείνονται σε όλες τις περιπτώσεις.

2.4.2 Δένδρα Απόφασης

Από τα πιο καθιερωμένα γραφικά εργαλεία στον τομέα της πληροφορικής είναι τα δένδρα απόφασης (decision trees). Όπως φανερώνει και το όνομα της, η τεχνική αυτή θυμίζει ένα δένδρο, απλά ανεστραμμένο. Από τα υψηλότερα προς τα χαμηλότερα στρώματα, το δένδρο διακρίνεται σε k διακλαδώσεις με βάση την τιμή κάποιου χαρακτηριστικού. Τα φύλλα του δένδρου, αποτελούν τις τελικές αποφάσεις που αφορούν μία συγκεκριμένη παρατήρηση αυτού του δένδρου. Δύο διακλαδώσεις δύναται να μην έχουν το ίδιο μήκος, ανάλογα με την εξέταση όλων ή ενός μόνο χαρακτηριστικού, όμως το συνολικό βάθος του δένδρου είναι συγκεκριμένο ανάλογα με τα χαρακτηριστικά που χρησιμοποιήθηκαν για την κατασκευή του. Οι αποφάσεις όμως, που βγαίνουν από το δένδρο



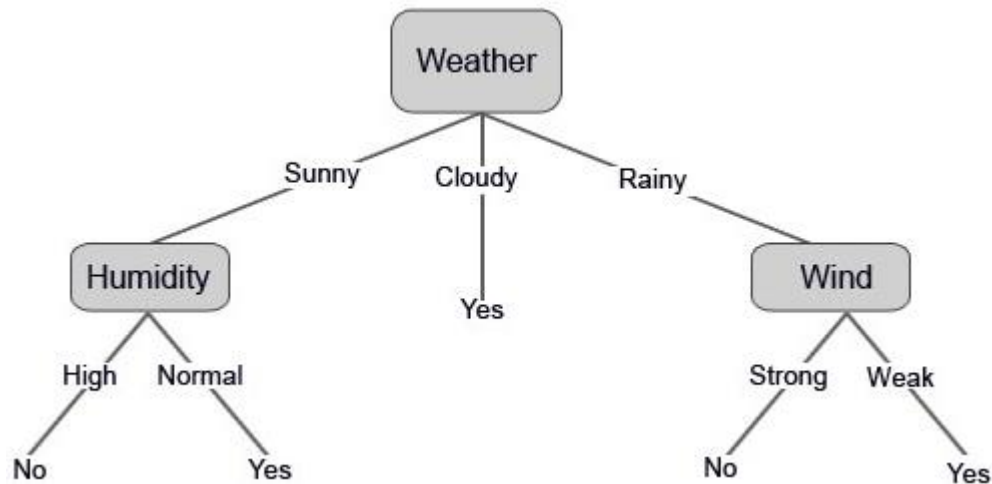
δεν είναι πάντα ονομαστικές. Ας θεωρηθεί για παράδειγμα ένα δένδρο να χρησιμοποιείται για να προβλέψει τη ζημιά μίας επιχείρησης παιχνιδιών βάσει του προϋπολογισμού, του πλήθους εργατών, του κόστους των μισθών κ.α. Στη συγκεκριμένη περίπτωση το δένδρο αυτό αποτελεί ένα Regression tree. Στην περίπτωση όμως που το δένδρο καταλήγει σε ονομαστική ταξινόμηση, και όπως στη δική μας περίπτωση προβλέπει το αν θα βρέξει και θα κάνει κάποιο έκτακτο φαινόμενο ή όχι, με βάση κάποιες κλιματικές παραμέτρους, τότε πρόκειται για ένα Classification tree.

Αρκετές φορές τα δένδρα απόφασης ενισχύονται από την τεχνική boosting. Στην περίπτωση αυτή χρησιμοποιείται η μέθοδος Boosted Decision Trees ή Boosted Decision Stumps, και πρόκειται για δένδρα τα οποία αποφασίζουν με βάση ένα χαρακτηριστικό. Οι μέθοδοι αυτές ακολουθούν πλήρως τον αλγόριθμο boosting χωρίς ιδιαίτερες τροποποιήσεις, οδήγησαν όμως στην ανάπτυξη πολυπλοκότερων αλγορίθμων, όπως τα τυχαία δάση.

Για να κατασκευαστεί το δένδρο, πρέπει να οριστεί πρώτα η σειρά με την οποία θα τοποθετηθούν τα χαρακτηριστικά σε κάθε κόμβο. Αν τα χαρακτηριστικά έχουν διακριτές τιμές [36], η συνηθέστερη τεχνική που χρησιμοποιείται βασίζεται στην Θεωρία Πληροφορίας και στην εντροπία, δηλαδή στην ποσότητα πληροφορίας που κατέχει το κάθε χαρακτηριστικό. Παρότι η τεχνική αυτή είναι βολική, αφού η εντροπία μπορεί να υπολογιστεί από τύπο με βάση τις πιθανότητες, στην περίπτωση των χαρακτηριστικών που έχουν συνεχείς τιμές αυτό το καθιστά αδύνατον. Λόγω ακριβώς αυτής της ιδιαιτερότητας, αποφασίζεται κάποιο σημείο διαχωρισμού για τα πεδία των χαρακτηριστικών, με τη χρήση των τεχνικών που ακολουθούν:

- Διαχωρίζοντας στη μέση τα πεδία, μία τεχνική που έχει αποδειχθεί ότι δεν λειτουργεί καθόλου καλά.
- Χωρίζοντας τα πεδία σε υποδιαστήματα και υπολογίζοντας το απόλυτο σφάλμα, έχει συνήθως σαν αποτέλεσμα ένα καλό σημείο διαχωρισμού.
- Έχοντας τις δύο κλάσεις των χαρακτηριστικών, πχ. ημέρες με καιρικά φαινόμενα και μέρες χωρίς, μπορεί να μετρηθεί ο μέσος όρος του κάθε χαρακτηριστικού, και στη συνέχεια ο μέσος όρος όλων των μέσων όρων. Η μέθοδος αυτή εντοπίζει συνήθως ένα αρκετά καλό σημείο, αλλά όχι το ιδανικό.
- Η τεχνική όπου μετριέται η εντροπία είναι αρκετά δύσκολο να λειτουργήσει, καθώς σε συνεχές πεδίο δεν είναι δυνατό να υπολογιστεί για κάθε πιθανό σημείο. Αυτό που εφαρμόζεται όμως είναι να επιλεγθούν σημεία, το καθένα από τα οποία έχει κάθε φορά διαφορετικό αριθμό παρατηρήσεων αριστερά και δεξιά σε σχέση με το προηγούμενο που επιλέχθηκε, οδηγώντας έτσι σε μικρότερο αριθμό πιθανών θέσεων. Εφόσον είναι γνωστό ότι η θέση με την μεγαλύτερη εντροπία βρίσκεται μεταξύ σημείων που θα ανήκαν σε διαφορετική κλάση, επιτυγχάνεται να μειωθεί ακόμη περισσότερο ο αριθμός των πιθανών θέσεων.

Στην [Εικόνα 2.2](#) παρουσιάζεται ένα παράδειγμα δένδρου απόφασης.



Εικόνα 2.2: Παράδειγμα δένδρου απόφασης, με βάση καιρικά δεδομένα

Οι πιο ευρέως χρησιμοποιούμενοι αλγόριθμοι που ανήκουν στην κατηγορία των δένδρων απόφασης είναι τα τυχαία δάση (random forests) και τα εναλλακτικά δένδρα απόφασης (alternating decision trees).

2.4.2.1 Τυχαία Δάση (Random Forests)

Η ιδέα για τον συγκεκριμένο αλγόριθμο ανήκει στον Tin Kam Ho (1995), αλλά αναπτύχθηκε και πήρε το όνομα της από τους Leo Breiman και Adele Cutler [37]. Μία από τις αποτελεσματικότερες μεθόδους, ανήκει στην κατηγορία των μεθόδων που επαναλαμβάνουν μία κλασική τεχνική ταξινόμησης πολλές φορές για πετύχουν μεγαλύτερη αποτελεσματικότητα της συνολικής μεθόδου. Οι “ασθενείς” επιμέρους κατηγοριοποιητές δύναται να είναι οποιοδήποτε εργαλείο ή μέθοδος ταξινόμησης, στη συγκεκριμένη περίπτωση τα δένδρα αποφάσεων. Μίας και η συγκεκριμένη μέθοδος χρησιμοποιεί κάποιο τύπο boosting, λειτουργεί όπως όλες οι συναφείς της, δηλαδή αφού κατασκευαστούν οι επιμέρους αδύναμοι κατηγοριοποιητές, ο καθένας αποφασίζει για την κλάση που ανήκει η παρατήρηση προς τον ταξινόμηση. Η κλάση που τελικά επικρατεί, είναι αυτή που ψηφίστηκε από τους περισσότερους. Έτσι λοιπόν, ενώ ο κάθε ταξινομητής μεμονωμένα έχει ανακριβή συνήθως αποτελέσματα και αντίστοιχα χαμηλή αξιοπιστία, ως σύνολο καταφέρνουν πολύ ακριβείς κατηγοριοποιήσεις.

Η μέθοδος των τυχαίων δασών λειτουργεί ως εξής:, επιλέγεται για αρχή ένα μικρό τυχαίο δείγμα όλων των χαρακτηριστικών της βάσης, τα οποία εν συνέχεια χρησιμοποιούνται τυχαία σε κάθε κόμβο μέχρι να κατασκευαστεί το δένδρο. Στη συνέχεια επιλέγεται ένα τυχαίο δείγμα παρατηρήσεων από την βάση στο οποίο θα γίνει η εκπαίδευση, ενώ το υπόλοιπο χρησιμοποιείται για να εκτιμήσει το σφάλμα του δένδρου. Η μέθοδος επαναλαμβάνεται τόσες φορές όσος και ο αριθμός των δένδρων που έχουν καθορίσει. Τα δύο βασικά στοιχεία που καθορίζουν την ακρίβεια του ταξινομητή σύμφωνα με τους δημιουργούς των Random Forests, είναι η επιρροή του κάθε δένδρου καθώς και η συσχέτιση ανάμεσα σε δύο δένδρα. Ένα δένδρο με χαμηλό σφάλμα έχει περισσότερη δύναμη και επομένως επηρεάζει θετικά τον ρυθμό σφάλματος του δάσους. Όσον αφορά στη συσχέτιση, αυτή έχει να κάνει τις ομοιότητες



δύο δένδρων, πιο συγκεκριμένα με το αν χρησιμοποιούν κοινά χαρακτηριστικά στην κατασκευή τους, και πάει αναλογικά με το ρυθμό σφάλματος.

Το πόσα χαρακτηριστικά θα επιλεγούν για κάθε δένδρο επηρεάζει προφανώς τους παράγοντες που αναφέραμε. Όσο λιγότερα είναι τα χαρακτηριστικά που θα επιλεγούν, τόσο λιγότερο συσχετισμένα θα είναι τα δένδρα, οδηγώντας σε μεγαλύτερο συνολικό σφάλμα.

Συνοψίζοντας παρότι τα τυχαία δάση έχουν πολλά προτερήματα, υψηλή ακρίβεια, μεγάλη αξιοπιστία καθώς και θεώρηση όλων των χαρακτηριστικών τους, έχει φανεί ότι είναι ευάλωτα σε overfitting, σε συγκεκριμένες βέβαια βάσεις. Ακόμη, λόγω του τυχαίου τρόπου δημιουργίας, δεν μπορούν να δημιουργηθούν ίδια δάση, πράγμα που σημαίνει πως αποτελεσματικότερα δένδρα δεν μπορούν να επαναληφθούν παρά την καλύτερη τους ακρίβεια.

2.4.2.2 Εναλλακτικά Δένδρα Απόφασης (Alternating Decision Trees)

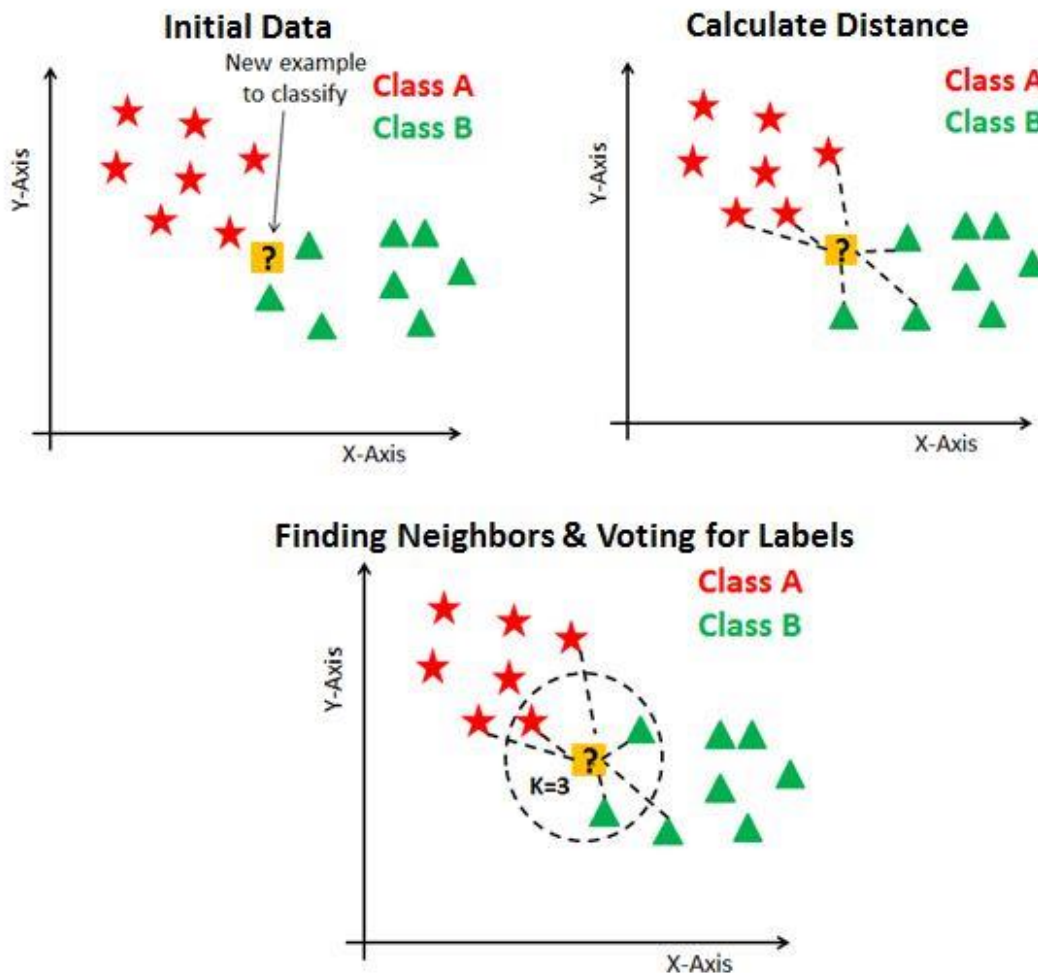
Η μέθοδος alternating decision trees (ADTrees) επίσης κατασκευάζει δένδρα, όμως αυτά δεν καταλήγουν σε φύλλα τελικής ταξινόμησης, αλλά σε τιμές που αντιπροσωπεύουν βάρη. Η ιδέα ανήκει στους Yoav Freund και Llew Mason [38]. Θα μπορούσε να ειπωθεί πως η τεχνική αυτή θυμίζει αλγορίθμους που χρησιμοποιούν boosting, καθώς αποδίδει βάρη και στα δεδομένα εκπαίδευσης, παρόμοια με τον αλγόριθμο Adaboost. Η διαφοροποίηση τους σε σχέση με τα απλά και boosted δένδρα ταξινόμησης, είναι το ότι επιλύουν το θέμα της έλλειψης συσχέτισης δύο παρόμοιων χαρακτηριστικών των παρατηρήσεων. Προκειμένου να επιτευχθεί αυτό σε κάθε επανάληψη του αλγορίθμου, όσο διαρκεί η κατασκευή των κόμβων, χρειάζεται να υπάρχει συσχέτιση ανάμεσα στο χαρακτηριστικού του κόμβου και αυτό του parent κόμβου του. Με τον τρόπο αυτό, το δένδρο αποκτά δομή και λαμβάνει υπόψη σημαντικές σχέσεις μεταξύ των χαρακτηριστικών. Επίσης αξίζει να σημειωθεί ότι οι κόμβοι και τα φύλλα του δύναται να είναι και τύπου πρόβλεψης (prediction nodes), όπου ο κόμβος περιέχει μία τιμή βάρους, εκτός από κόμβους απόφασης (decision nodes). Ας αναλυθεί όμως η μέθοδος με την οποία λαμβάνει μία απόφαση, για να γίνει κατανοητό πως μας ωφελεί. Στη μέθοδο αυτή λοιπόν ακολουθούνται ταυτόχρονα όλα τα μονοπάτια για τα οποία ισχύουν οι υποθέσεις των κόμβων απόφασης, εν αντιθέσει με τα περισσότερα δένδρα απόφασης που διασχίζονται μέσω ορειβασίας (hill climbing). Εφόσον προσπελαστεί και ο τελευταίος κόμβος-φύλλο που ισχύει, αθροίζονται οι τιμές όλων των κόμβων πρόβλεψης που ισχύουν, και ανάλογα με το κατώφλι απόφασης καταλήγει σε ένα αποτέλεσμα.

2.4.3 k - Nearest Neighbors

Η μέθοδος k κοντινότερων γειτόνων (k - NN), αποτελεί έναν από τους απλούστερους σε εφαρμογή αλγορίθμους, κατά την οποία κατηγοριοποιούνται παραδείγματα ανάλογα με τις γειτονικές παρατηρήσεις της. Πρόκειται για μία κλασική μέθοδο επιβλεπόμενης μάθησης, υπάρχει όμως μία σημαντική διαφορά, πως δεν υφίσταται μία ουσιαστική περίοδος κατά την οποία ο αλγόριθμος εκπαιδεύεται. Το σύνολο δεδομένων εκπαίδευσης απλά αποθηκεύεται, ενώ

η ταξινόμηση, η εξαγωγή κανόνων και η εύρεση ενός ορίου αποφάσεων (decision boundary) πραγματοποιείται κατά την εφαρμογή του αλγορίθμου. Η μέθοδος αυτή εκτελείται ως εξής:

- Αποθηκεύονται στη μνήμη οι εκπαιδευτικές παρατηρήσεις.
- Μία νέα παρατήρηση προβάλλεται στο επίπεδο.
- Οι k κοντινότερες παρατηρήσεις ψηφίζουν για την κλάση της νέας.
- Η κλάση με τις περισσότερες ψήφους, αποφασίζει για το αποτέλεσμα.



Εικόνα 2.3: Βήματα k -NN [39]

Το θετικό στοιχείο με τον k -NN είναι ότι δύναται να χρησιμοποιηθεί και για παλινδρόμηση, δηλαδή ακόμα κι αν πρόκειται για δεδομένα με συνεχείς τιμές. Σε τέτοια περίπτωση, συλλέγονται οι τιμές απόφασης των k κοντινότερων γειτόνων και προκύπτει ο μέσος όρος τους, που συνιστά και το αποτέλεσμα που αποδίδεται στη νέα παρατήρηση. Εννοείται πως για μεγαλύτερη ακρίβεια, αποδίδονται βάρη αναλόγως της απόστασης από τους γείτονες.

Οι παράμετροι που είναι απαραίτητοι για τη ρύθμιση σε έναν αλγόριθμο k -NN, είναι λίγοι. Καταρχάς, πρέπει να οριστεί ο αριθμός των k γειτόνων που θα ληφθούν υπόψιν. Αν επιλεγεί μικρός αριθμός γειτόνων, αυτό



μπορεί να οδηγήσει σε μεγάλο σφάλμα, ενώ αν επιλεγεί μεγάλος αριθμός αυτό όπως είναι λογικό, δίνει πλεονέκτημα στην κλάση με τις περισσότερες παρατηρήσεις στο σύνολο δεδομένων εκπαίδευσης. Η επιλογή του k , σε γενικές γραμμές δεν ακολουθεί κάποιο κανόνα. Όπως γίνεται κατανοητό, όσο μεγαλύτερος ο αριθμός του k τόσο καλύτερη ταξινόμηση μπορεί να επιτευχθεί, όμως κάτι τέτοιο θα προϋπέθετε δείγμα εκπαίδευσης με πλήθος παρατηρήσεων που να τείνει στο άπειρο. Αφού λοιπόν σε γενικές γραμμές τα δείγματα είναι περιορισμένα, είναι αναγκαία η επιλογή του ιδανικού για κάθε δείγμα αριθμού γειτόνων, που θα αποφεύγει “θορυβώδη” όρια απόφασης ελέω λίγων γειτόνων αλλά και δεν θα υπερ - μοντελοποιεί λόγω πολλών. Για να αποφασιστεί λοιπόν ο αριθμός k , θα μπορούσε να χρησιμοποιηθεί cross-validation ή bootstrapping [40]. Συγκεκριμένα αν πρόκειται για προβλήματα δυαδικής ταξινόμησης, καλό είναι να επιλέγεται περιττός αριθμός γειτόνων, ώστε να αποφεύγονται οι ισοπαλίες στις ψήφους.

Προκειμένου να γίνει ευκολότερα κατανοητό, με ποιο τρόπο επιλέγονται οι κοντινότεροι γείτονες, όσον αφορά το δισδιάστατο μπορούν εύκολα να βρεθούν μετρώντας την Ευκλείδεια απόσταση όλων από τη νέα παρατήρηση. Στην περίπτωση όμως που πρόκειται για επίπεδο περισσότερων διαστάσεων, ορισμένες διαστάσεις ενδέχεται να έχουν περισσότερη επιρροή όσον αφορά την απόσταση δύο παρατηρήσεων, όμως η Ευκλείδεια απόσταση θεωρεί πάντοτε ότι όλες έχουν την ίδια επιρροή. Αν το πρόβλημα, αφορά σε ταξινόμηση σε κείμενο, μπορεί να χρησιμοποιηθεί η απόσταση Hamming δίχως προβλήματα. Η απόσταση Hamming λειτουργεί μετρώντας πόσες διακριτές τιμές στις μεταβλητές έχουν τα παραδείγματα, και θεωρεί αυτό σαν απόσταση μεταξύ τους. Υπάρχουν όμως και άλλες μέθοδοι, όπως η Neighborhood Component Analysis [41] η οποία χρησιμοποιεί την απόσταση Mahalanobis προκειμένου να υπολογίσει τις αποστάσεις μεταξύ των παραδειγμάτων. Υπάρχει ακόμη, η μέθοδος μέγιστου περιθωρίου κοντινότερων γειτόνων (LMNN) [42]. Με τη χρήση των μεθόδων αυτών διαπιστώθηκε σημαντική αύξηση της ακρίβειας και κατέστη σαφές το πόσο σημαντικό είναι να επιλεγεί το κατάλληλο μετρικό σύστημα.

Από την αρχική επινόηση των Cover και Hart (1967), ο αλγόριθμος k - nn έχει δοκιμαστεί σε διάφορες παραλλαγές, ήδη την επόμενη χρονιά ο Hart είχε την ιδέα για μείωση του συνόλου δεδομένων εκπαίδευσης με τη μέθοδο k - condensed nearest neighbors (k - CNN). Σκοπός της μεθόδου αυτής είναι να καταλήξει σε ένα υποσύνολο του συνόλου δεδομένων εκπαίδευσης, που να είναι όμως αρκετό για την ταξινόμηση. Για να το πετύχει αυτό χρειάζεται ο υπολογισμός του decision boundary, παρόμοιο με τη λειτουργία των μηχανών διανυσματικής υποστήριξης (SVM), με την ακόλουθη μεθοδολογία:

- Το υποσύνολο αρχικοποιείται με ένα τυχαίο παράδειγμα ή ομάδα παραδειγμάτων.
- Εφόσον υπάρχει το υποσύνολο, μπορεί να γίνει ταξινόμηση και όλων των υπολοίπων παραδειγμάτων.
- Εν συνεχεία επιλέγεται μία τυχαία ταξινόμηση από τις λανθασμένες, συνήθως κοντά στο όριο απόφασης, και προστίθεται στο υποσύνολο.
- Η διαδικασία αυτή επαναλαμβάνεται, μέχρι να εξαλειφθεί και η τελευταία λάθος ταξινόμηση.



Με τον τρόπο αυτό δημιουργείται ένα μικρότερο σύνολο παρατηρήσεων, το οποίο παρόλα αυτά παρέχει εξίσου αποτελεσματική ταξινόμηση στα δεδομένα εκπαίδευσης. Η μέθοδος αυτή βοηθάει στην εξοικονόμηση ταχύτητας, όμως δεν είναι βέβαιο πως θα καταλήξει στο ιδανικό αντιπροσωπευτικό σύνολο, αφού η πρώτη τυχαία παρατήρηση ή ομάδα που θα επιλεγεί, καθίσταται πλέον περιττή εφόσον δεν θα βρίσκεται κοντά στο όριο [43]. Αξίζει να σημειωθεί ακόμη, πως ο αλγόριθμος εκτελείται καλύτερα όταν τα δεδομένα είναι διαχωριζόμενα, είτε με γραμμικό είτε με μη γραμμικό τρόπο. Ο λόγος για αυτό, είναι πως ο αλγόριθμος στην αρχική του μορφή είχε μικρή ευαισθησία σε μακρινές αποστάσεις, ελέω πολλών παρατηρήσεων, ενώ ο k - CNN μεταφέρει τη σημασία πολλών σε λιγότερα σημεία, και με αυτό τον τρόπο αυξάνει τον παράγοντα ευαισθησίας. Η λύση για αυτό βρίσκεται στη χρήση ενός άλλου τύπου του k - NN, τη γνωστή και ως μέθοδος επεξεργασμένων συμπυκνωμένων κοντινότερων γειτόνων (ECNN) (Hart, 1978).

Προκειμένου να μειώσει περαιτέρω τον αριθμό των παραδειγμάτων εκπαίδευσης, η μέθοδος αυτή προχωρά σε μία μορφή προ-επεξεργασίας των δεδομένων [44]. Στην ουσία με τη μέθοδο αυτή αντιμετωπίζεται το πρόβλημα της ευαισθησίας σε μακρινές παρατηρήσεις, που συναντούσαν οι προηγούμενες εκδόσεις, ενώ προσαρμόζονται τα όρια απόφασης μεταξύ των κλάσεων, αφαιρώντας παρατηρήσεις που προσθέτουν θόρυβο στα δεδομένα. Για τη χρήση της μεθόδου αυτής χρησιμοποιούνται δύο συναρτήσεις πυκνότητας πιθανότητας (pdfs), μία για κάθε κλάση. Έστω δύο συναρτήσεις f και h , αντίστοιχα για τις δύο κλάσεις, και προκύπτει μία νέα παρατήρηση c , αν $f(c) > h(c)$, τότε κατά πάσα πιθανότητα η νέα παρατήρηση ανήκει στην πρώτη κλάση. Σε περίπτωση όμως που η παρατήρηση c ανήκει τελικά στην δεύτερη κλάση, ενώ ισχύει η προαναφερθείσα σχέση των pdfs, τότε εύκολα κατανοούμε ότι η συγκεκριμένη παρατήρηση είναι περικυκλωμένη από παρατηρήσεις άλλης κλάσης, και άρα είναι πολύ πιθανό να προκαλέσει σφάλματα στην ταξινόμηση. Αν θεωρηθεί δηλαδή το $k=1$, τότε στην περιοχή κοντά σε αυτή την παρατήρηση υπάρχει μεγάλη πιθανότητα λάθος ταξινόμησης. Αυτό το σφάλμα λοιπόν, αποφεύγεται μέσω μίας μετατροπής του συνόλου δεδομένων εκπαίδευσης, ακολουθώντας τα παρακάτω βήματα:

- Υπολογίζονται όλα των pdfs για όλες τις παρατηρήσεις του συνόλου δεδομένων εκπαίδευσης
- Ελέγχονται όλες οι πληροφορίες και αφαιρούνται αυτές για τις οποίες ισχύει ένας από τους δύο κανόνες, είτε το c να ανήκει στην 1^η κλάση και να ισχύει $f(c) < h(c)$, είτε το c να ανήκει στη 2^η κλάση αλλά και να ισχύει $f(c) > h(c)$.

Ακολουθώντας τη μεθοδολογία αυτή αφαιρούνται όλες τις παρατηρήσεις τα οποία δύναται να προκαλέσουν σφάλματα. Να σημειωθεί όμως πως η μέθοδος αυτή, δεν αφαιρεί παρατηρήσεις με λίγη σημασία, δηλαδή παρατηρήσεις σε πυκνοκατοικημένες περιοχές με κοινή κλάση, οπότε χρησιμοποιείται αποτελεσματικότερα σε συνδυασμό με τον αλγόριθμο k - CNN.

Μία νεότερη τροποποίηση του k - NN, για να μειωθεί ακόμα περισσότερο το σύνολο δεδομένων εκπαίδευσης, επινόησε ο Gates [45], και πιο συγκεκριμένα τη μέθοδο μειωμένων κοντινότερων γειτόνων (k -RNN). Αν εφαρμοστεί πρώτα ο k -CNN, ο k -RNN σε πρώτη φάση εξαλείφει το πρώτο δείγμα ή ομάδα δειγμάτων που προστέθηκε στο



υποσύνολο που κατασκεύασε ο k -CNN, αφού κρίνεται συνήθως περιττό. Στη συνέχεια, κατηγοριοποιούνται όλες οι παρατηρήσεις του συνόλου δεδομένων εκπαίδευσης και αν αυτό γίνει σωστά, ο αλγόριθμος αφαιρεί το επόμενο στοιχείο του υποσυνόλου και η διαδικασία επαναλαμβάνεται. Ακόμη και σε περίπτωση που προκύψουν λάθος κατηγοριοποιήσεις, το στοιχείο που εξαλείφθηκε επιστρέφει στο υποσύνολο και γίνεται επιλογή κάποιου άλλου για έλεγχο. Παρότι η μέθοδος αυτή έχει αυξημένες υπολογιστικές απαιτήσεις, αν συνδυαστεί με τον k -CNN πετυχαίνει μία πολύ ικανοποιητική εκτίμηση των αντιπροσωπευτικών παρατηρήσεων του συνόλου δεδομένων εκπαίδευσης.

Αρκετά χρόνια μετέπειτα, οι Zhang, Yao και Zhung [46] απέδειξαν το ότι ο k -NN δεν είναι τόσο ακριβής όπως άλλοι αλγόριθμοι, βλέπε SVM, AdaBoost, NB.

Συνοψίζοντας, ο k -NN θεωρείται ένας έξυπνος αλγόριθμος ταξινόμησης, λόγω και των προτερημάτων που εμφανίζει. Μερικά από τα σημαντικότερα:

- Εφαρμόζεται εύκολα, χωρίς δυσκολίες στη λειτουργία του.
- Δεν έχει περιορισμούς να εφαρμοστεί σε data οποιασδήποτε κατανομής.
- Είναι δυνατή η επίτευξη πολύ αποτελεσματικής ταξινόμησης, αν υπάρχει φυσικά επαρκής αριθμός παρατηρήσεων.

Έχει όμως και σημαντικά μειονεκτήματα, τα οποία στοχεύουν να βελτιώσουν οι διάφορες μορφές του, δηλαδή:

- Η επιστημονική κοινότητα, δεν έχει καταλήξει ακόμα στην επιλογή του ιδανικού k .
- Για να επιτευχθεί μεγάλη ακρίβεια, αρκετές φορές το σύνολο δεδομένων εκπαίδευσης μεγαλώνει ανάλογα.
- Απαιτείται αποθηκευτικός χώρος για να διατηρηθούν όλες οι παρατηρήσεις.
- Υπάρχει ευαισθησία αν τα δεδομένα έχουν πολλά χαρακτηριστικά, καθώς επηρεάζεται από τις πολλές διαστάσεις που πρέπει να ληφθούν υπόψη στις μετρήσεις των αποστάσεων.
- Απαιτείται αυξημένη υπολογιστική ισχύς και χρόνος για να υπολογιστούν οι κοντινότερες αποστάσεις.

2.4.4 Τεχνική σύνθεσης υπερ – δειγματοληψίας μειοτήτων SMOTE

Στο σημείο αυτό έχει αξία να συζητηθεί επιγραμματικά η τεχνική SMOTE.

Όπως θα φανεί και στις ενότητες που ακολουθούν στα σύνολα δεδομένων που μας διατέθηκαν, δεν ήταν ισόποσες όλες οι καταγραφές. Οπότε ήταν ξεκάθαρη η ανάγκη ισοστάθμισης των δεδομένων και συγκεκριμένα υπερ – δειγματοληψίας των δεδομένων με τις λιγότερες καταγραφές. Όπως θα φανεί και παρακάτω, στη συγκεκριμένη μελέτη τα μετεωρολογικά δεδομένα παρουσιάζουν μεγάλη ανισορροπία μεταξύ των καταγραφών έντονων φαινομένων και ηλιοφάνειας.

Προκειμένου να γίνει αυτό, χρησιμοποιήθηκε μία συνήθης τεχνική σύνθεσης που χρησιμοποιείται στη βιβλιογραφία και ονομάζεται Synthetic Minority Oversampling Technique (SMOTE) και η οποία εφαρμόζεται επάνω σε μη ισορροπημένα σύνολα δεδομένων προκειμένου να τα ισορροπήσει. Συγκεκριμένα η τεχνική SMOTE

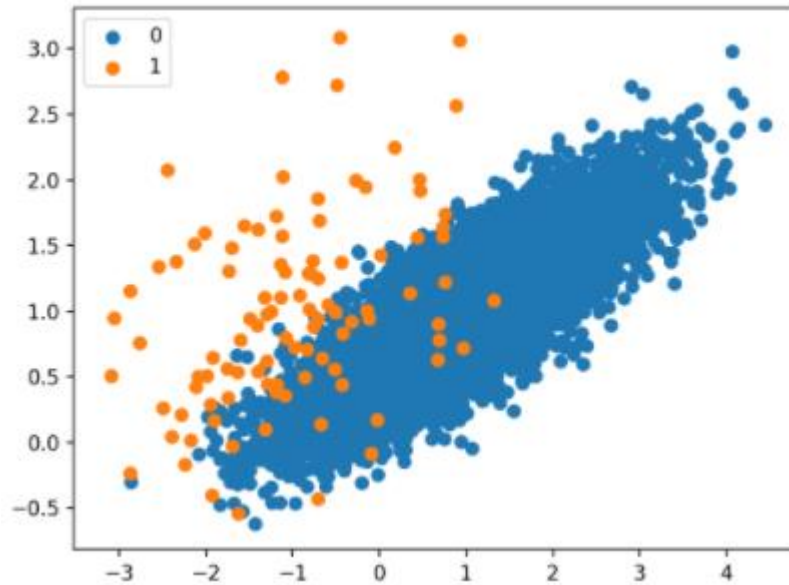


χρησιμοποιεί τη μέθοδο της υπερ - δειγματοληψίας της μειοψηφικής κλάσης “oversampling of the minority” ή/και της υπο - δειγματοληψίας της πλειοψηφικής κλάσης “undersampling of the majority”, προκειμένου να εξισορροπήσει τα δεδομένα του υπό εξέταση συνόλου δεδομένων.

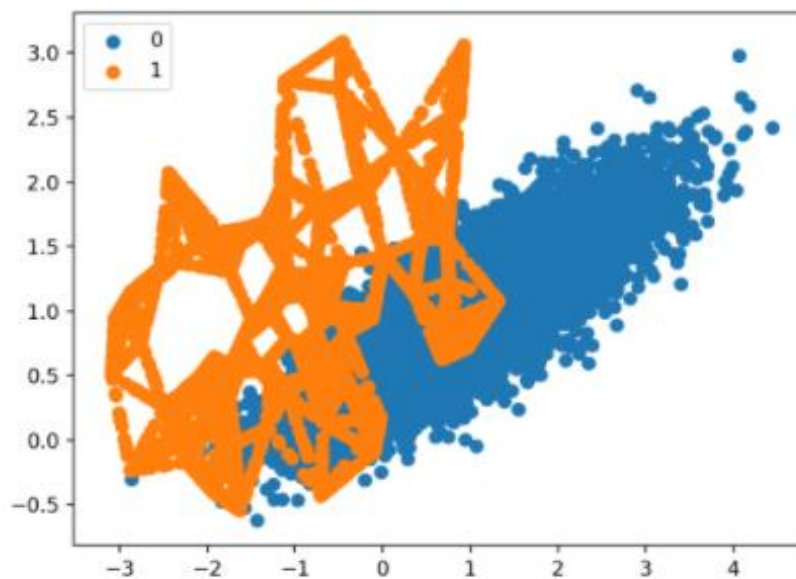
Ο πιο απλός τρόπος για την αύξηση των δεδομένων είναι η αναπαραγωγή καταγραφών στην κατηγορία των μειονοτήτων, κάτι το οποίο γενικά ενώ εξισορροπεί την κατανομή τάξης δεν προσδίδει νέα πληροφορία για την εκπαίδευση του αλγόριθμου. Εναλλακτικός τρόπος για την αντιμετώπιση της δυσκολίας των μην ισορροπημένων συνόλων δεδομένων και για την αύξηση της ευαισθησίας των αλγορίθμων στην κατηγορία των μειονοτήτων είναι η υπο - δειγματοληψία της πλειοψηφικής τάξης, δηλ. του υποσυνόλου των δεδομένων με τις περισσότερες καταγραφές. Ωστόσο έχει αποδειχθεί ότι ένας συνδυασμός της μεθόδου υπο - δειγματοληψίας της πλειοψηφικής κλάσης και της μεθόδου υπερ - δειγματοληψίας της μειοψηφικής κλάσης μπορεί να επιτύχει καλύτερη απόδοση ταξινόμησης, συγκριτικά με την εφαρμογή μόνο της υπο - δειγματοληψίας της πλειοψηφικής κλάσης [47]. Έτσι, η περίπτωση της υπερ-δειγματοληψίας της μειοψηφικής κλάσης υλοποιείται μέσω της σύνθεσης νέων καταγραφών από τις ήδη υπάρχουσες, βάσει της τεχνικής συνθετικής υπερπαραβολής δειγμάτων μειοψηφίας (Synthetic Minority Oversampling Technique - SMOTE) [48].

Η SMOTE λειτουργεί επιλέγοντας παραδείγματα που βρίσκονται κοντά στο χαρακτηριστικό χώρο ενδιαφέροντος, σχεδιάζοντας μία γραμμή μεταξύ των παραδειγμάτων στο χαρακτηριστικό χώρο και σχεδιάζοντας ένα νέο δείγμα σε ένα σημείο κατά μήκος αυτής της γραμμής.

Χαρακτηριστική απεικόνιση μίας τυχαίας εφαρμογής SMOTE για εξισορρόπηση δεδομένων σε ένα μη ισορροπημένο πρόβλημα δυαδικής ταξινόμησης είναι η ακόλουθη (Εικ. 2.4 – 2.5). Στην [Εικ. 2.4](#) φαίνεται το διάγραμμα διασποράς των δεδομένων, όπου με μπλε χρώμα απεικονίζεται η τάξη της πλειοψηφίας, ενώ με πορτοκαλί χρώμα απεικονίζεται η τάξη της μειοψηφίας. Στην [Εικ. 2.5](#) παρουσιάζεται το ίδιο σύνολο δεδομένων, που μετά την εφαρμογή του SMOTE, έχει μετασχηματισθεί σε ένα ισορροπημένο σύνολο δεδομένων, έχοντας συντεθεί επιπλέον δεδομένα για την κλάση της μειοψηφίας.



Εικόνα 2.4: Διάγραμμα Διασποράς μη ισορροπημένου δυαδικού προβλήματος ταξινόμησης



Εικόνα 2.5: Διάγραμμα Διασποράς μη ισορροπημένου δυαδικού προβλήματος ταξινόμησης μετασχηματισμένου με SMOTE

Η SMOTE είναι πολύ χρήσιμη έως απαραίτητη για τη σωστή υλοποίηση των ταξινομητών που χρησιμοποιούν δεδομένα όπως αυτά της παρούσας εργασίας. Στα δεδομένα της εργασίας αυτής, εφαρμόστηκε “oversampling of the minority”, προκειμένου να συνθέσουμε ισορροπημένα σύνολα δεδομένων φαινομένων και ηλιοφαινιών.

2.5 Προ – επεξεργασία Δεδομένων

Η προ-επεξεργασία (pre - processing) των δεδομένων, αποτελεί ένα από τα σημαντικότερα βήματα της εξόρυξης δεδομένων, και συνιστά το 60 έως και 80% ολόκληρης της διαδικασίας. Γενικά, τα δεδομένα που διατίθενται, μπορεί



να είναι ελλιπή ή να έχουν λάθη ή υπάρχουν ανομοιόμορφα δεδομένα από διαφορετικές πηγές ενώ είναι και διαφορετικού τύπου. [49]. Τα λεγόμενα «μη καθαρά» δεδομένα εντάσσονται στις εξής κατηγορίες:

- Ελλιπή: εντοπίζονται ελλείψεις στις τιμές των χαρακτηριστικών αλλά και δεδομένα που περιέχουν μόνο συναθροιστικά δεδομένα.
- Θορυβώδη: εντοπίζονται λάθη στα δεδομένα ή ενδεχομένως ακραίες τιμές (outliers).
- Ασυνεπή: εντοπίζονται ασυνέπειες στις ονομασίες ή τους κωδικούς, και επομένως απαιτείται οργάνωση των δεδομένων, καθώς αυτά δεν ακολουθούν κάποιου είδους συστηματοποίηση.

Σκοπός λοιπόν του σταδίου αυτού, είναι να απορριφθούν τα λανθασμένα δεδομένα και να συλλεχθούν και εξεταστούν τα ελλιπή δεδομένα προκειμένου να μειωθεί ο κίνδυνος τέτοιου είδους θέματα να κοστίζουν στην ακρίβεια των αποτελεσμάτων.



3 Μεθοδολογία

Για την πρόβλεψη εμφάνισης έντονων καιρικών φαινομένων, η διαδικασία που ακολουθήθηκε στην παρούσα εργασία είναι η ακόλουθη. Αφού ελήφθησαν μετεωρολογικά δεδομένα και καταγραφές έντονων καιρικών φαινομένων από την Εθνική Μετεωρολογική Υπηρεσία (Ε.Μ.Υ) (§3.1), προχωρήσαμε σε επεξεργασία τους κάνοντας τα σύνολα δεδομένων ομοιογενή. Παρατηρήθηκε σαφής ανισορροπία μεταξύ των δεδομένων καθώς οι καταγραφές των καιρικών φαινομένων ήταν σημαντικά λιγότερες από εκείνες των μετεωρολογικών δεδομένων. Έτσι, για πιο σωστή διαχείριση των ανισόρροπων δεδομένων κρίθηκε σκόπιμη η εφαρμογή της τεχνικής SMOTE. Συγκριμένα έγινε oversampling of the minority για την εξισορρόπηση των μετεωρολογικών δεδομένων με τα καιρικά φαινόμενα, δημιουργώντας περισσότερα τεχνητά δεδομένα φαινομένων προκειμένου να βοηθήσουμε στην δημιουργία των ταξινομητών.

Εξετάσαμε 9 πόλεις και για κάθε πόλη χρησιμοποιήσαμε τους τρεις αλγόριθμους Logistic Regression, Random Forests και $k - NN$ για να φτιάξουμε τους ταξινομητές κάθε πόλης. Στη συνέχεια για να αξιολογήσουμε τις δυνατότητες κάθε ταξινομητή κάθε πόλης δημιουργήσαμε ένα σύνολο δεδομένων δοκιμής με ίσο αριθμό δεδομένων, έντονα φαινόμενα (1) και ηλιοφάνειες (0). Τα δεδομένα δεν είναι τεχνητά αλλά προέρχονται από τις πραγματικές καταγραφές φαινομένων που δόθηκαν από την Ε.Μ.Υ.

Ακολούθως, πραγματοποιήθηκε και παρουσιάζεται αξιολόγηση και συγκριτική ανάλυση των αποτελεσμάτων κάθε ταξινομητή για κάθε πόλη.

Προκειμένου να ελεγχθεί περαιτέρω η ικανότητα προβλέψεων κάθε ταξινομητή κάθε πόλης πραγματοποιήθηκε συγκριτική ανάλυση ευαισθησίας για όλους τους ταξινομητές με βάση τις μεταβολές σε τέσσερις παραμέτρους:

1. Ύψος υετού, το οποίο αυξήθηκε κατά 2εκ.
2. Ελάχιστη θερμοκρασία η οποία μειώθηκε κατά 3°C και μέγιστη θερμοκρασία η οποία αυξήθηκε κατά 3°C.
3. Ταχύτητα ανέμου η οποία αυξήθηκε κατά 2 κόμβους

Τα αποτελέσματα της ανάλυσης ευαισθησίας παρουσιάζονται και εξετάζονται στο πλαίσιο αυτής της εργασίας.

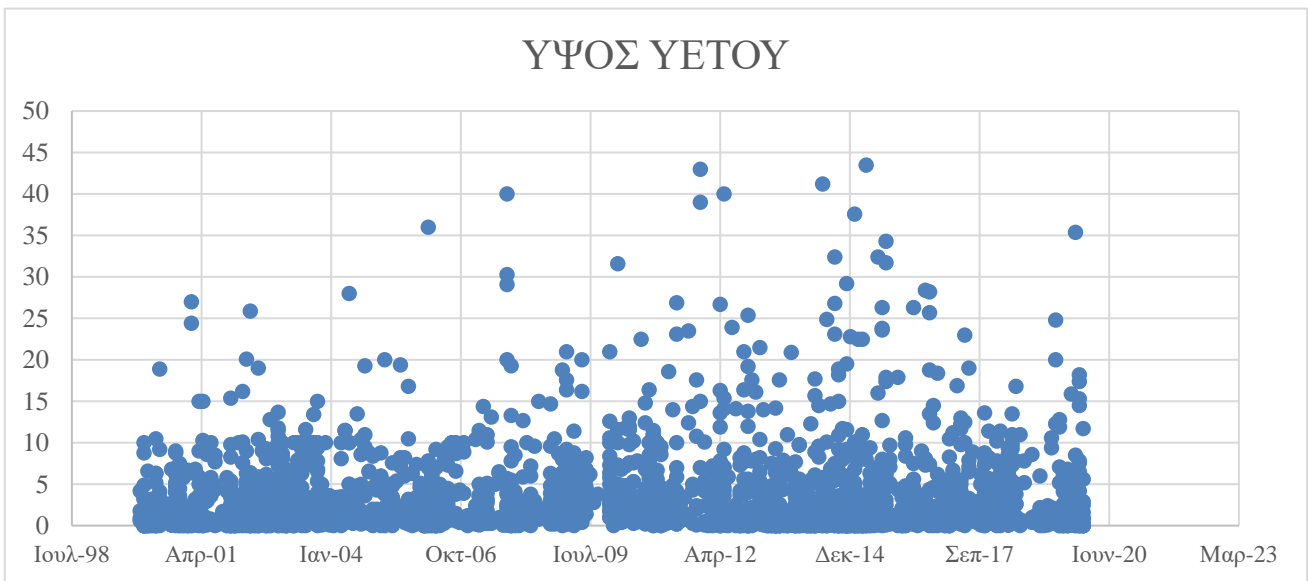
3.1 Πειραματικά Δεδομένα

Τα σύνολα δεδομένων που χρησιμοποιήθηκαν σε αυτή την εργασία, παραχωρήθηκαν σε εμάς κατόπιν επικοινωνίας με τη Διεύθυνση Κλιματολογίας και Εφαρμογών της Ε.Μ.Υ. Πιο συγκεκριμένα, μας παραχωρήθηκαν μετεωρολογικά δεδομένα, δηλ. δεδομένα βροχόπτωσης (ύψους υετού), θερμοκρασίας (μέγιστης και ελάχιστης θερμοκρασίας) και ανέμου (ξηρή θερμοκρασία (BDT), ταχύτητα ανέμου, κόμβοι δύναμης ανέμου, μποφόρ, μέγιστοι κόμβοι δύναμης ανέμου, σχετική υγρασία, σημείο δρόσου) για 15 πόλεις της Ελλάδας, για την εικοσαετία από το 2000 έως το 2019 σε μορφή csv. files. Επιπλέον, μας παραχωρήθηκαν απλές καταγραφές ειδικών γεγονότων, που εν προκειμένω αφορούσαν σε έντονα καιρικά φαινόμενα (έντονες βροχοπτώσεις, χαλάζι, χιονοπτώσεις, υψηλής

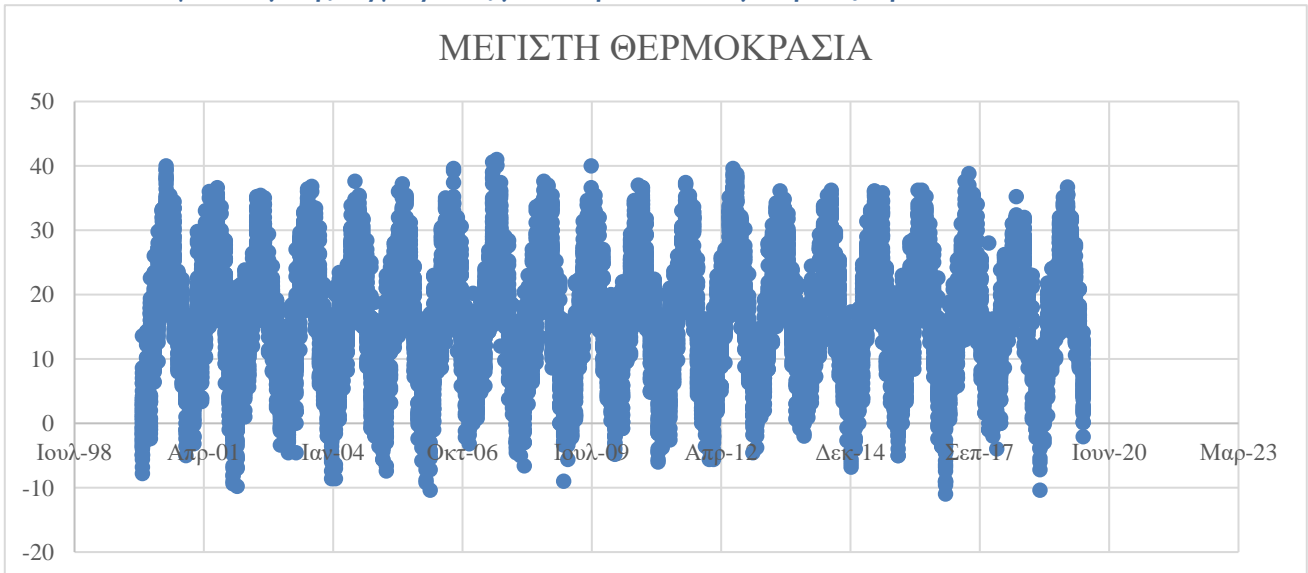


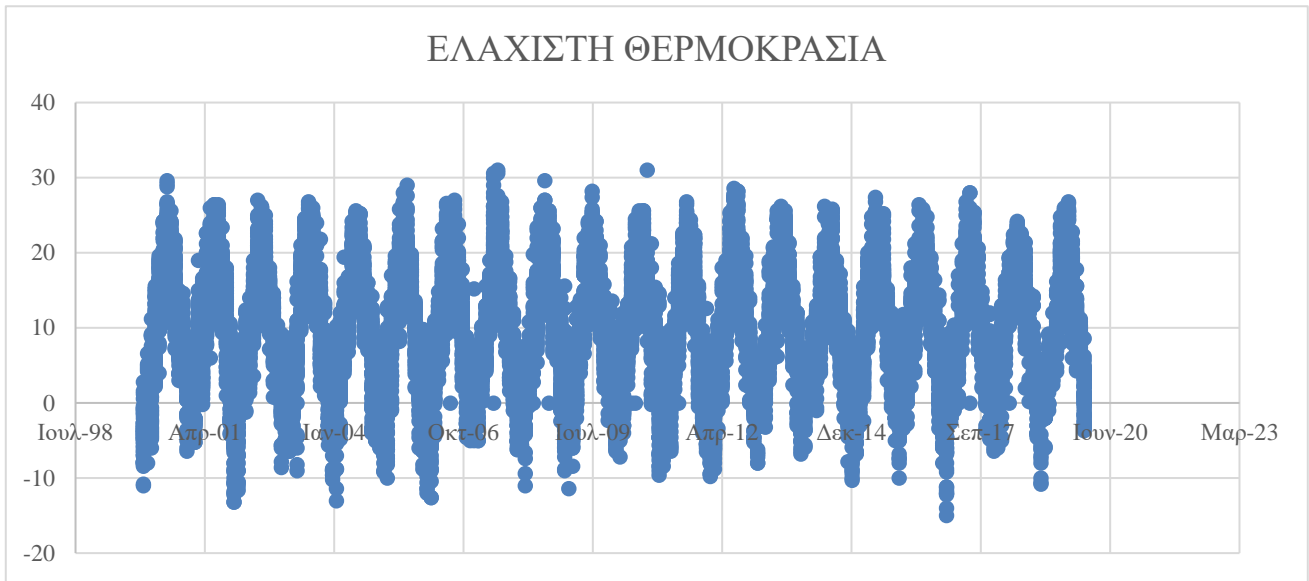
έντασης άνεμοι) και έλαβαν χώρα σε διάφορες πόλεις της Ελλάδας, την πενταετία 2014 έως 2019. Τα έντονα καιρικά φαινόμενα αντιστοιγήθηκαν με τα μετεωρολογικά φαινόμενα, βάσει της κοινής χρονοσφραγίδας, δηλ. της κοινής ημερομηνίας και ώρας. Η διαδικασία αυτή έγινε χειροκίνητα στη εφαρμογή Office Excel. Στη συνέχεια δημιουργήθηκαν ξεχωριστά .csv files, ένα για τα δεδομένα ύψους νετού και φαινόμενα, ένα για τα δεδομένα θερμοκρασιών και φαινόμενα και ένα για τα δεδομένα ανέμου και φαινόμενα. Παρακάτω φαίνονται ενδεικτικές χρονοσειρές για τα δεδομένα ύψους νετού (Εικ. 3.1) και μέγιστης (Εικ. 3.2) και ελάχιστης θερμοκρασίας (Εικ. 3.3) για την πόλη της Κοζάνης για τη δεκαετία 2000 έως 2019.

Εικόνα 3.1. Δεδομένα Ύψους Υετού για τα έτη 2000 – 2019 για την Κοζάνη



Εικόνα 3.2. Δεδομένα Μέγιστης Θερμοκρασίας για τα έτη 2000 – 2019 για την Κοζάνη





Εικόνα 3.3. Δεδομένα Ελάχιστης Θερμοκρασίας για τα έτη 2000 – 2019 για την Κοζάνη

Από το πλήθος των δεδομένων που είχαμε στη διάθεσή μας τέθηκαν σε επεξεργασία δεδομένα για 9 περιοχές της Ελλάδας – Αλεξανδρούπολη, Δυτικά Προάστια Αττικής, Ηράκλειο, Θεσσαλονίκη, Ιωάννινα, Καλαμάτα, Κοζάνη, Λάρισα και Τρίπολη – για τις οποίες υπήρχαν περισσότερες καταγραφές έντονων φαινομένων σε σχέση με τις υπόλοιπες περιοχές. Στον [Πίνακα 3.1](#) που ακολουθεί φαίνεται πόσες καταγραφές δεδομένων υπήρχαν για κάθε πόλη που εξετάστηκε:

ΠΟΛΕΙΣ	ΔΙΑΘΕΣΙΜΕΣ ΚΑΤΑΓΡΑΦΕΣ
Αλεξανδρούπολη	13
Δυτική Αττική	70
Ηράκλειο	4
Θεσσαλονίκη	14
Ιωάννινα	16
Καλαμάτα	14
Κοζάνη	21
Λάρισα	9
Τρίπολη	36

Πίνακας 3.1: Διαθέσιμες καταγραφές έντονων καιρικών φαινομένων στις πόλεις της Ελλάδας που εξετάστηκαν.



3.2 Προ - επεξεργασία των δεδομένων

Τα μετεωρολογικά δεδομένα που μας παραδόθηκαν από την Ε.Μ.Υ, παρουσίαζαν ανομοιομορφίες. Αρχικά τα δεδομένα ύψους υετού και μέγιστης και ελάχιστης θερμοκρασίας ήταν λιγότερα σε πλήθος σε σχέση με τα δεδομένα του ανέμου, καθώς τα μεν πρώτα λαμβάνονταν κάθε 8 ώρες ενώ τα μεν δεύτερα κάθε 3 ώρες. Επομένως, στο στάδιο αυτό έγινε επιλογή των κοινών καταγραφών (ίδια ημερομηνία και ίδια ώρα) προκειμένου όλα τα σύνολα δεδομένων να έχουν ίσο αριθμό δεδομένων. Επιπλέον έγινε επεξεργασία των ονομάτων κάθε στήλης δεδομένων προκειμένου όλες οι στήλες να έχουν ένα μοναδικό όνομα, κοινό για ίδια καταγραφή κάθε πόλης. Ακόμη, αντιστοιχήθηκαν οι καταγραφές των έντονων καιρικών φαινομένων με τα μετεωρολογικά δεδομένα στις κοινές τους ημερομηνίες. Στο σημείο αυτό αντιμετωπίστηκε το πρόβλημα της ανισορροπίας των καταγραφών μεταξύ των φαινομένων και των μετεωρολογικών δεδομένων.

3.2.1 Μη ισορροπημένα σύνολα δεδομένων

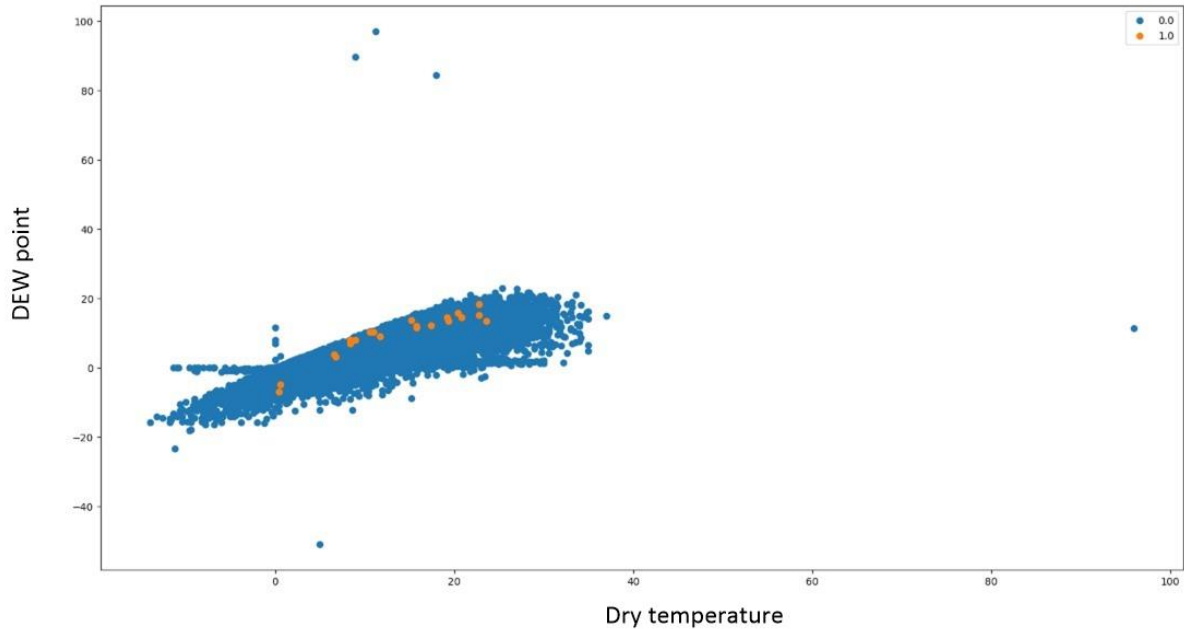
Λόγω της δυνατότητας της Ε.Μ.Υ. να παραχωρήσει δεδομένα και καταγραφές έντονων καιρικών φαινομένων μόνο για την πενταετία από το 2014 έως το 2019 και λόγω των αντικειμενικά μειωμένων καταγραφών ανά πόλη της Ελλάδος, δεδομένου ότι δεν εμφανίζονται έντονα φαινόμενα, που προκαλούν μεγάλες καταστροφές, πολλές φορές μέσα στο έτος, τα σύνολα δεδομένων που προέκυψαν από την αντιστοίχιση των μετεωρολογικών στοιχείων με τα καιρικά φαινόμενα ήταν μη ισορροπημένα. Τα μη ισορροπημένα σύνολα δεδομένων (imbalanced datasets) απαιτούν ειδική επεξεργασία και παράλληλα οι μη ισορροπημένες ταξινομήσεις αποτελούν πρόκληση για την προγνωστική μοντελοποίηση, καθώς οι περισσότεροι από τους αλγόριθμους μηχανικής εκμάθησης που χρησιμοποιούνται για την ταξινόμηση (classification), ανάμεσα τους και οι τρεις αλγόριθμοι που χρησιμοποιήθηκαν στην παρούσα εργασία, έχουν σχεδιαστεί γύρω από την υπόθεση ίσου αριθμού δειγμάτων για κάθε τάξη. Αυτό έχει ως αποτέλεσμα τα μοντέλα να έχουν χαμηλή προγνωστική απόδοση, ειδικά για την τάξη των μειονοτήτων (minority class), εν προκειμένω για το πλήθος των μετεωρολογικών δεδομένων για τα οποία έχει καταγραφεί φαινόμενο. Αυτό αποτελεί πρόβλημα γιατί συνήθως, η τάξη των μειονοτήτων είναι πιο σημαντική και επομένως το πρόβλημα είναι πιο ευαίσθητο σε σφάλματα ταξινόμησης για την τάξη των μειονοτήτων από την τάξη της πλειοψηφίας [34]. Ειδικότερα, τα μη ισορροπημένα σύνολα δεδομένων (imbalanced datasets), αντιμετωπίζονται από τους περισσότερους αλγόριθμους ως εξής: θα αγνοήσουν, και στη συνέχεια θα έχουν χαμηλή απόδοση στα σημεία που δεν υπάρχει αντιστοίχιση των μεν δεδομένων με τα δε, και από την άλλη πλευρά η απόδοση που θα προκύψει ως είναι η απόδοση στην κατηγορία μειονοτήτων που είναι πιο σημαντική.

3.2.2 Τεχνική Σύνθεσης Πειραματικών Συνόλων Δεδομένων με λίγες καταγραφές

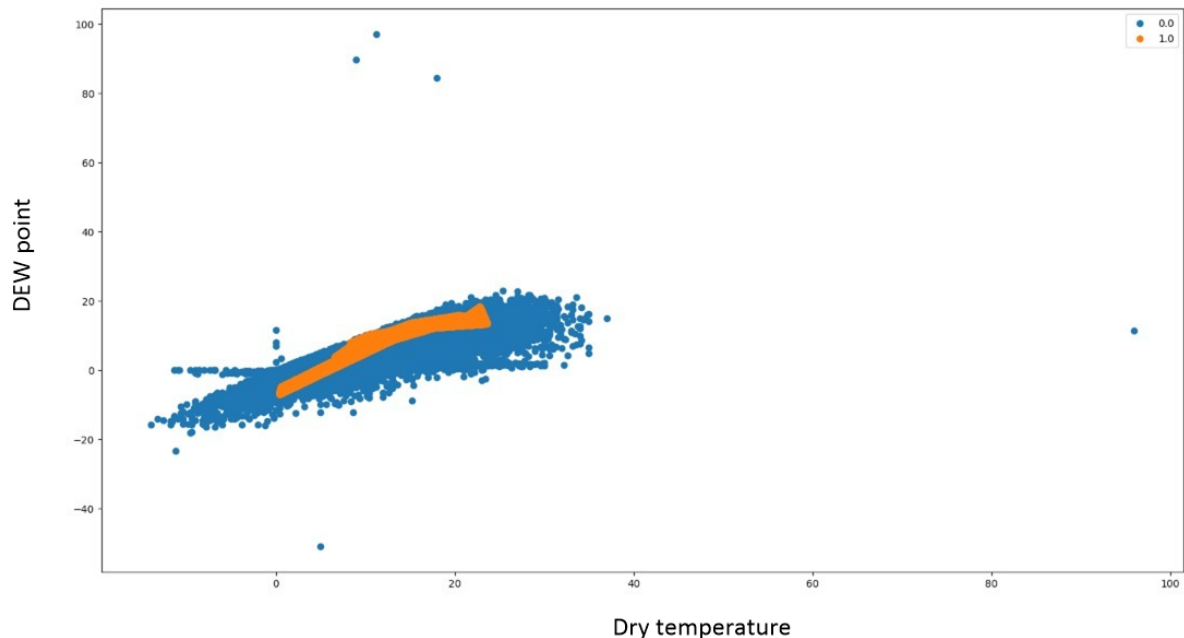
Στα δεδομένα που επεξεργαστήκαμε και λόγω της έντονης ανισορροπίας που προκλήθηκε από το γεγονός ότι οι διαθέσιμες καταγραφές φαινομένων ήταν σημαντικά λιγότερες από τα μετεωρολογικά δεδομένα, εφαρμόσαμε την τεχνική SMOTE και συγκεκριμένα κάναμε oversampling of the minority. Στις [Εικ. 3.4](#) και [3.5](#) που ακολουθούν



παρουσιάζεται ενδεικτικά η εφαρμογή της τεχνικής του oversampling για δεδομένα ανέμου (dry temp vs. dew point), όπως καταγράφηκαν για την πόλη των Ιωαννίνων.



Εικόνα 3.4. Διάγραμμα Διασποράς μη ισορροπημένων καταγραφών για την πόλη των Ιωαννίνων



Εικόνα 3.5. Διάγραμμα Διασποράς μετά την εφαρμογή oversampling για την πόλη των Ιωαννίνων

Στην Εικ. 3.4 είναι εμφανής η ανισορροπία μεταξύ των καιρικών δεδομένων και των ηλιοφαιεών. Το ζήτημα αυτό επηρεάζει σε μεγάλο βαθμό την υλοποίηση των ταξινομητών, για το λόγο αυτό επιχειρήθηκε η σύνθεση



επιπλέον δεδομένων έντονων καιρικών φαινομένων, όπως φαίνεται στην Εικ. 3.5.

3.3 Σύνολα Εκπαίδευσης & Σύνολα Δοκιμής

Τα σύνολα των δεδομένων για κάθε πόλη, που είχαμε στη διάθεση μας, περιέχουν κατά μέσο όρο 14.000 καταγραφές για τα μεγέθη ύψους νετού, μέγιστη και ελάχιστη θερμοκρασία και περίπου 50.000 καταγραφές για τα μεγέθη του ανέμου, δεδομένου του ότι οι καταγραφές για τις δύο πρώτες περιπτώσεις γίνονταν ανά 12 ώρες και στην περίπτωση του ανέμου ανά 3 ώρες. Τελικά, κατόπιν κατάλληλης επεξεργασίας των δεδομένων του ανέμου, όλα τα σύνολα δεδομένων είχαν ίσο αριθμό καταγραφών.

Ως σύνολα δεδομένων εκπαίδευσης (training datasets), με τα οποία εκπαιδεύτηκαν οι τρεις αλγόριθμοι που χρησιμοποιήθηκαν, ορίστηκαν τα σύνολα δεδομένων των πόλεων, δηλ. τα δεδομένα κάθε πόλης χρησιμοποιήθηκαν προκειμένου να εκπαιδευτούν οι τρεις αλγόριθμοι για κάθε πόλη. Σε αυτά τα σύνολα δεδομένων περιελήφθησαν και οι δεκατρείς παράμετροι (ύψους νετού, δύο παράμετροι θερμοκρασίας, επτά παράμετροι ανέμου και ημερομηνία και ώρα).

Ως σύνολο δεδομένων δοκιμής (testing dataset) ορίστηκε ένα κοινό ισορροπημένο σύνολο δεδομένων, το οποίο περιείχε 362 καταγραφές και δημιουργήθηκε από τα 181 συνολικά φαινόμενα που εμφανίστηκαν σε όλες τις πόλεις και 181 ηλιοφάνειες, που επιλέχθηκαν τυχαία.

Τέλος, να σημειωθεί ότι η ύπαρξη έκτακτου καιρικού φαινομένου έλαβε την τιμή 1 και η μη ύπαρξη φαινομένου (ηλιοφάνεια) έλαβε την τιμή 0. Στο εξής τα έκτακτα καιρικά φαινόμενα θα αναφέρονται ως events.

4 Αποτελέσματα

4.1 Απόδοση Ταξινομητών ανά Πόλη

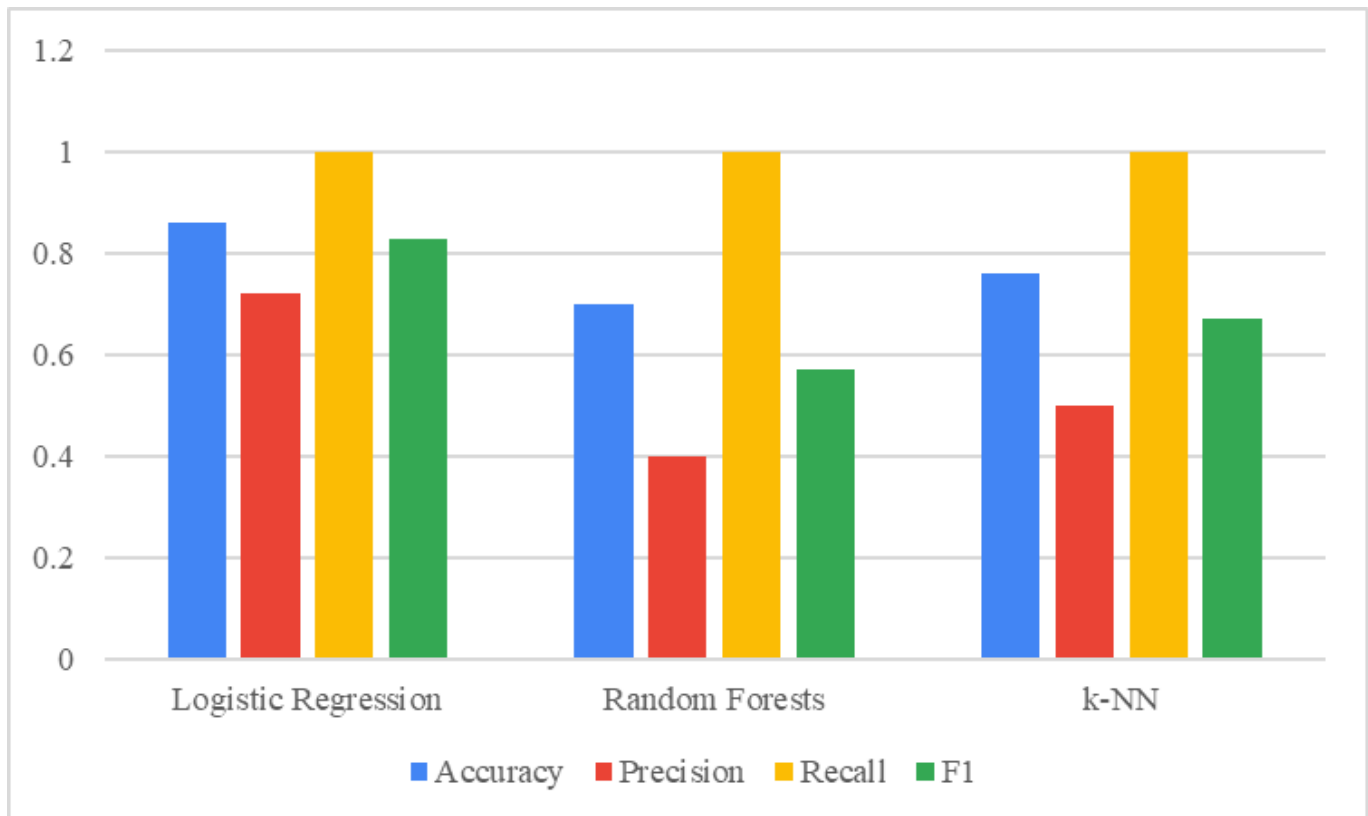
Στο κεφάλαιο αυτό παρουσιάζονται τα αποτελέσματα που έδωσαν οι ταξινομητές Logistic Regression, Random Forests και k -NN, για κάθε πόλη της Ελλάδας που εξετάστηκε.

4.1.1 Δυτική Αττική

Για τη Δυτικά προάστια της Αττικής οι συνολικές μετεωρολογικές καταγραφές είναι 11.825 και τα events είναι 70. Στον [Πίνακα 4.1.1](#) παρατίθενται οι δείκτες απόδοσης των τριών αλγορίθμων και στην [Εικ. 4.1.1](#) ακολουθεί το σχετικό γράφημα:

Classifier:	Δυτική Αττική		
	Logistic Regression	Random Forests	k -NN
Accuracy	0.86	0.7	0.76
Precision	0.72	0.4	0.5
Recall	1	1	1
F1	0.83	0.57	0.67

Πίνακας 4.1.1: Πίνακας Δεικτών Απόδοσης των Ταξινομητών για τη Δυτική Αττική



Εικόνα 4.1.1: Γραφική Απεικόνιση της απόδοσης των ταξινομητών για τη Δυτική Αττική

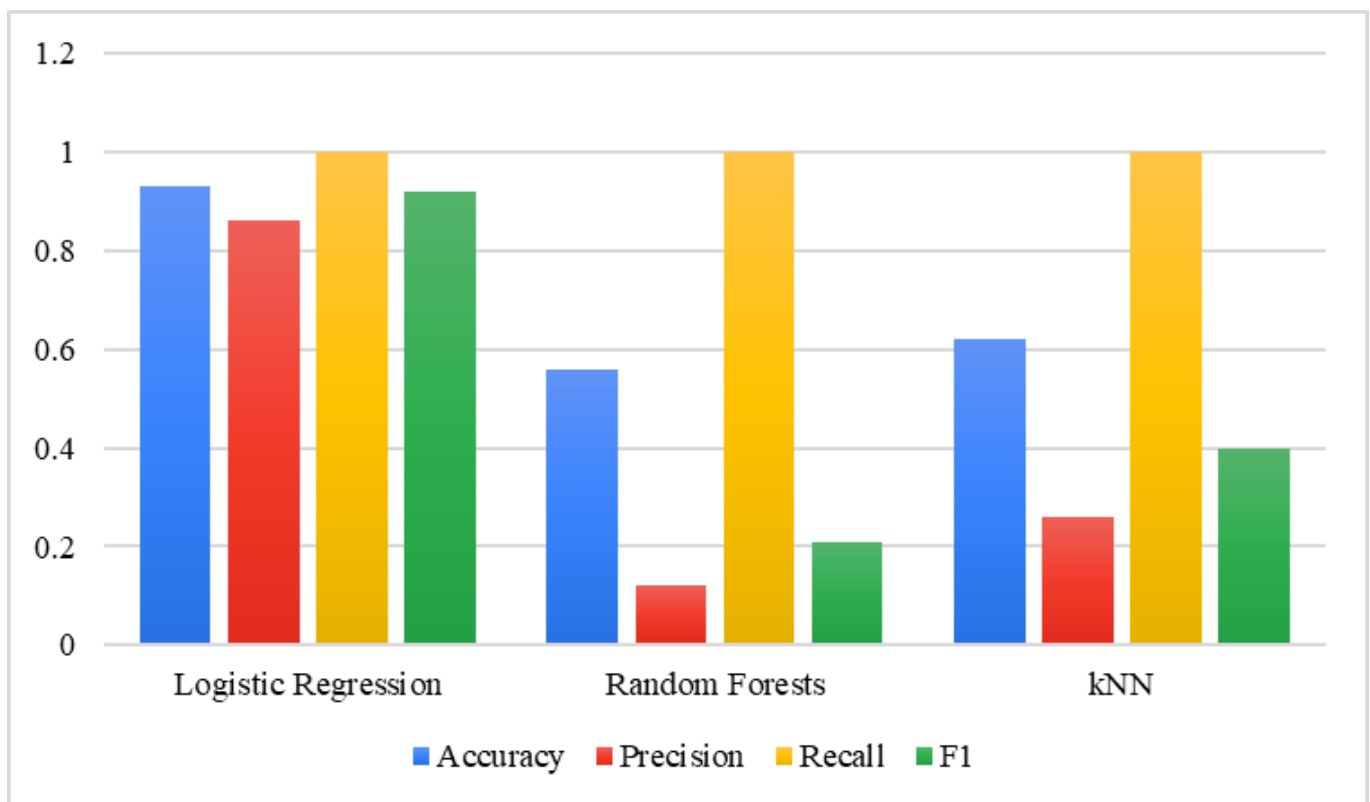
Παρατηρείται ότι για τα Δυτικά Προάστια της Αττικής ο Logistic Regression είναι πολύ πιο αποδοτικός από τον k - NN και τον Random Forests, αποδίδοντας precision 72%, έναντι του k - NN (50%) και του Random Forests (40%) και accuracy 86% έναντι του k - NN (76%) και του Random Forests (70%).

4.1.2 Κοζάνη

Για την Κοζάνη οι συνολικές μετεωρολογικές καταγραφές είναι 13.520 και τα events είναι 21. Στον [Πίνακα 4.1.2](#) παρατίθενται οι δείκτες απόδοσης των τριών αλγορίθμων και στην [Εικ. 4.1.2](#) ακολουθεί το σχετικό γράφημα:

Classifier:	Κοζάνη		
	Logistic Regression	Random Forests	k - NN
Accuracy	0.93	0.56	0.62
Precision	0.86	0.12	0.26
Recall	1	1	1
F1	0.92	0.21	0.4

Πίνακας 4.1.2: Πίνακας Δεικτών Απόδοσης των Ταξινομητών για την Κοζάνη



Εικόνα 4.1.2: Απεικόνιση της απόδοσης των ταξινομητών για την Κοζάνη

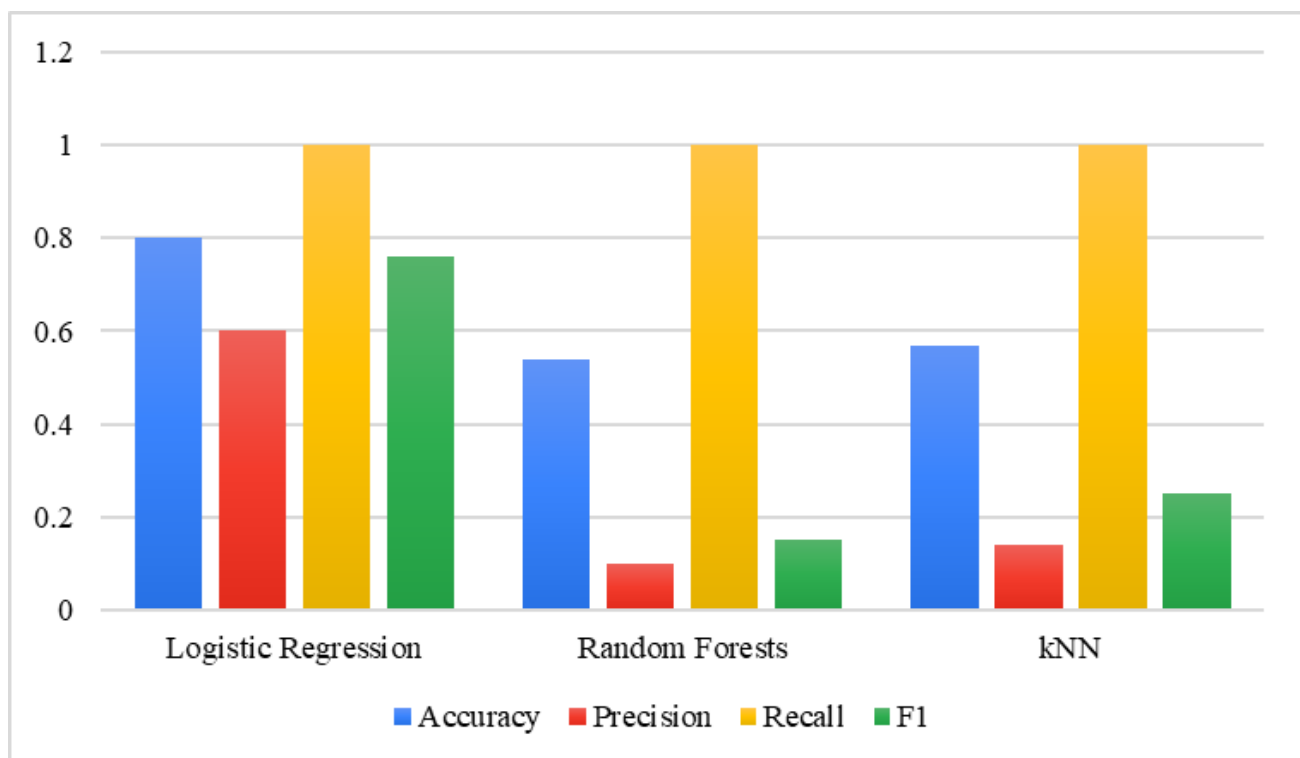
Παρατηρείται ότι για την Κοζάνη ο Logistic Regression είναι χαρακτηριστικά πιο αποδοτικός από τον k - NN και τον Random Forests, αποδίδοντας precision 82%, έναντι του k - NN (26%) και του Random Forests (12%) και accuracy 93% έναντι του k - NN (62%) και του Random Forests (56%).

4.1.3 Αλεξανδρούπολη

Για την Αλεξανδρούπολη οι συνολικές μετεωρολογικές καταγραφές είναι 14.473 και τα events είναι 13. Στον [Πίνακα 4.1.3](#) παρατίθενται οι δείκτες απόδοσης των τριών αλγορίθμων και στην [Εικ. 4.1.3](#) ακολουθεί το σχετικό γράφημα:

Αλεξανδρούπολη			
Classifier:	Logistic Regression	Random Forests	k - NN
Accuracy	0.8	0.54	0.57
Precision	0.6	0.1	0.14
Recall	1	1	1
F1	0.76	0.15	0.25

Πίνακας 4.1.3: Πίνακας Δεικτών Απόδοσης των Ταξινομητών για την Αλεξανδρούπολη



Εικόνα 4.1.3: Απεικόνιση της απόδοσης των ταξινομητών για την Αλεξανδρούπολη

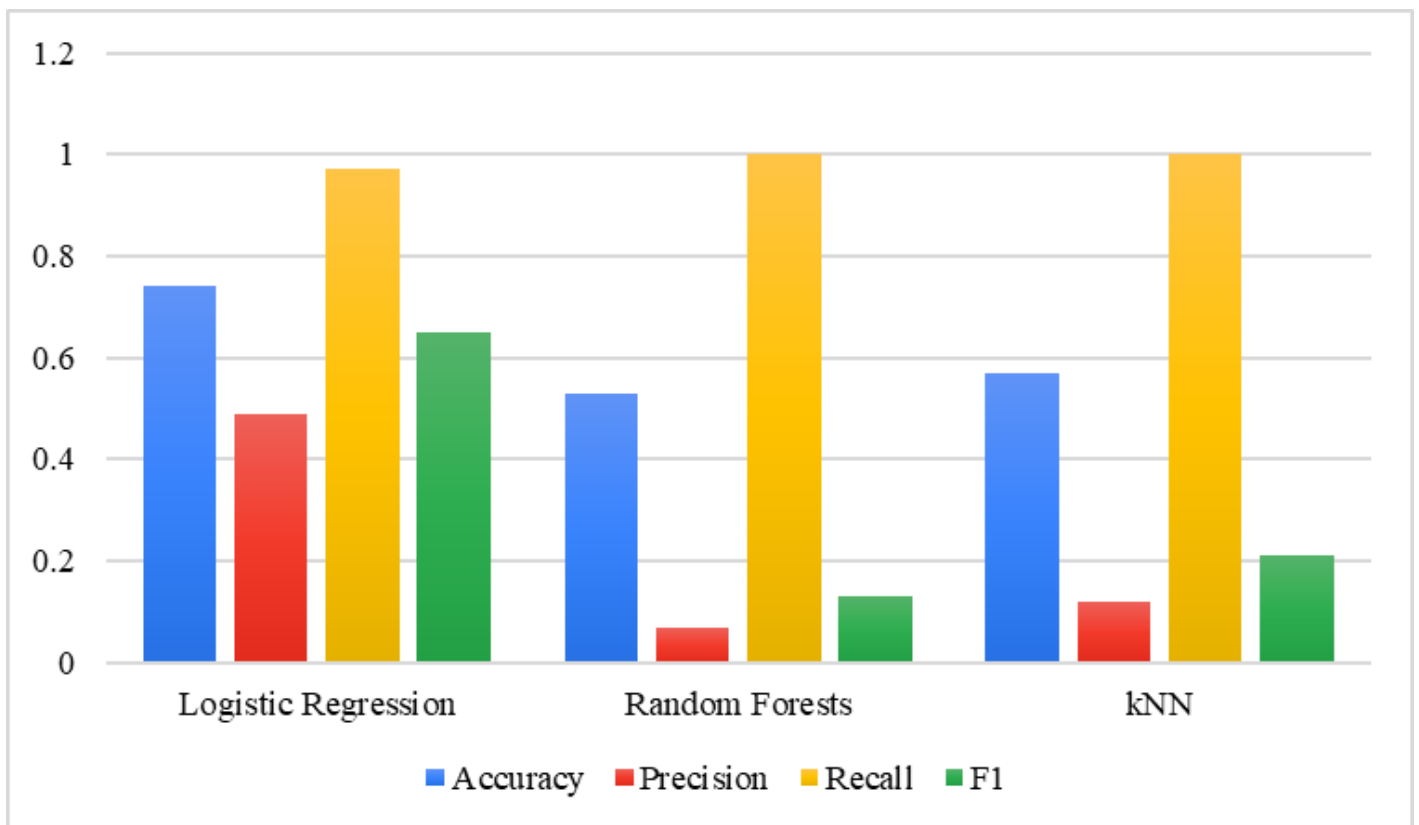
Παρατηρείται ότι για την Αλεξανδρούπολη ο Logistic Regression είναι αποδοτικός σε σχέση με τον k - NN και τον Random Forests, αποδίδοντας precision 60%, έναντι του k - NN (14%) και του Random Forests (10%). Το accuracy για τον Logistic Regression είναι στο 80%, αρκετά καλύτερο από εκείνο του k - NN (57%) και του Random Forests (54%).

4.1.4 Λάρισα

Για τη Λάρισα οι συνολικές μετεωρολογικές καταγραφές είναι 14.370 και τα events είναι 9. Στον [Πίνακα 4.1.4](#) παρατίθενται οι δείκτες απόδοσης των τριών αλγορίθμων και στην [Εικ. 4.1.4](#) ακολουθεί το σχετικό γράφημα:

Classifier:	Λάρισα		
	Logistic Regression	Random Forests	k - NN
Accuracy	0.74	0.53	0.57
Precision	0.49	0.07	0.12
Recall	0.97	1	1
F1	0.65	0.13	0.21

Πίνακας 4.1.4: Πίνακας Δεικτών Απόδοσης των Ταξινομητών για τη Λάρισα



Εικόνα 4.1.4. Διάγραμμα Διασποράς μετά την εφαρμογή oversampling για την πόλη της Λάρισας

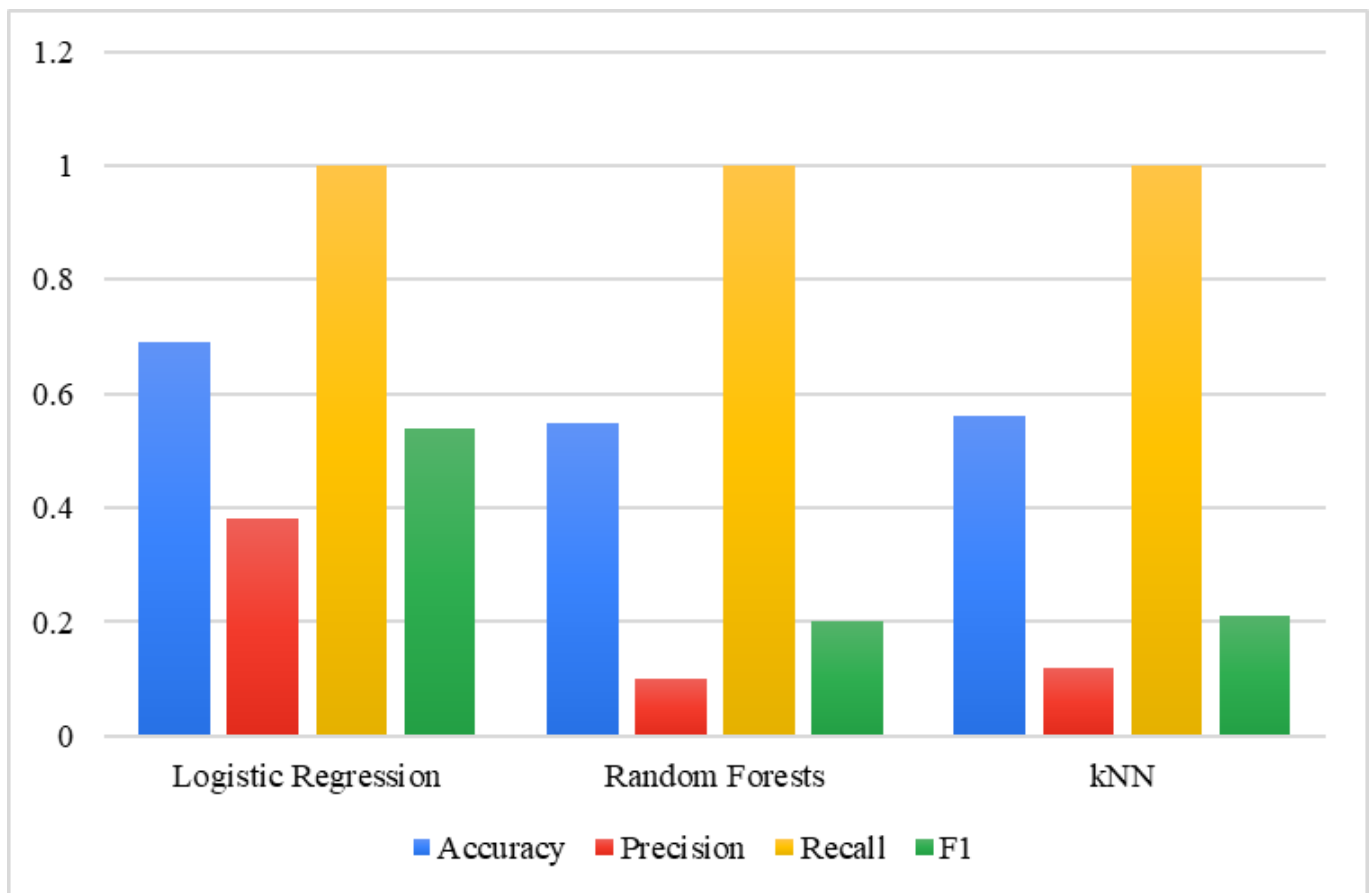
Παρατηρείται ότι για τη Λάρισα ο Logistic Regression είναι πιο αποδοτικός από τον k - NN και τον Random Forests, ωστόσο αποδίδει χαμηλό precision 49%, και ακόμη χαμηλότερα βρίσκονται οι άλλοι δύο αλγόριθμοι. Το accuracy για τον Logistic Regression είναι στο 74%, αρκετά καλύτερο από εκείνο του k - NN (57%) και του Random Forests (53%).

4.1.5 Ιωάννινα

Για τα Ιωάννινα οι συνολικές μετεωρολογικές καταγραφές είναι 14.267 και τα events είναι 16. Στον [Πίνακα 4.1.5](#) παρατίθενται οι δείκτες απόδοσης των τριών αλγορίθμων και στην [Εικ. 4.1.5](#) ακολουθεί το σχετικό γράφημα:

Classifier:	Ιωάννινα		
	Logistic Regression	Random Forests	k - NN
Accuracy	0.69	0.55	0.56
Precision	0.38	0.10	0.12
Recall	1	1	1
F1	0.54	0.20	0.21

Πίνακας 4.1.5: Πίνακας Δεικτών Απόδοσης των Ταξινομητών για τα Ιωάννινα



Εικόνα 4.1.5: Απεικόνιση της απόδοσης των ταξινομητών για τα Ιωάννινα

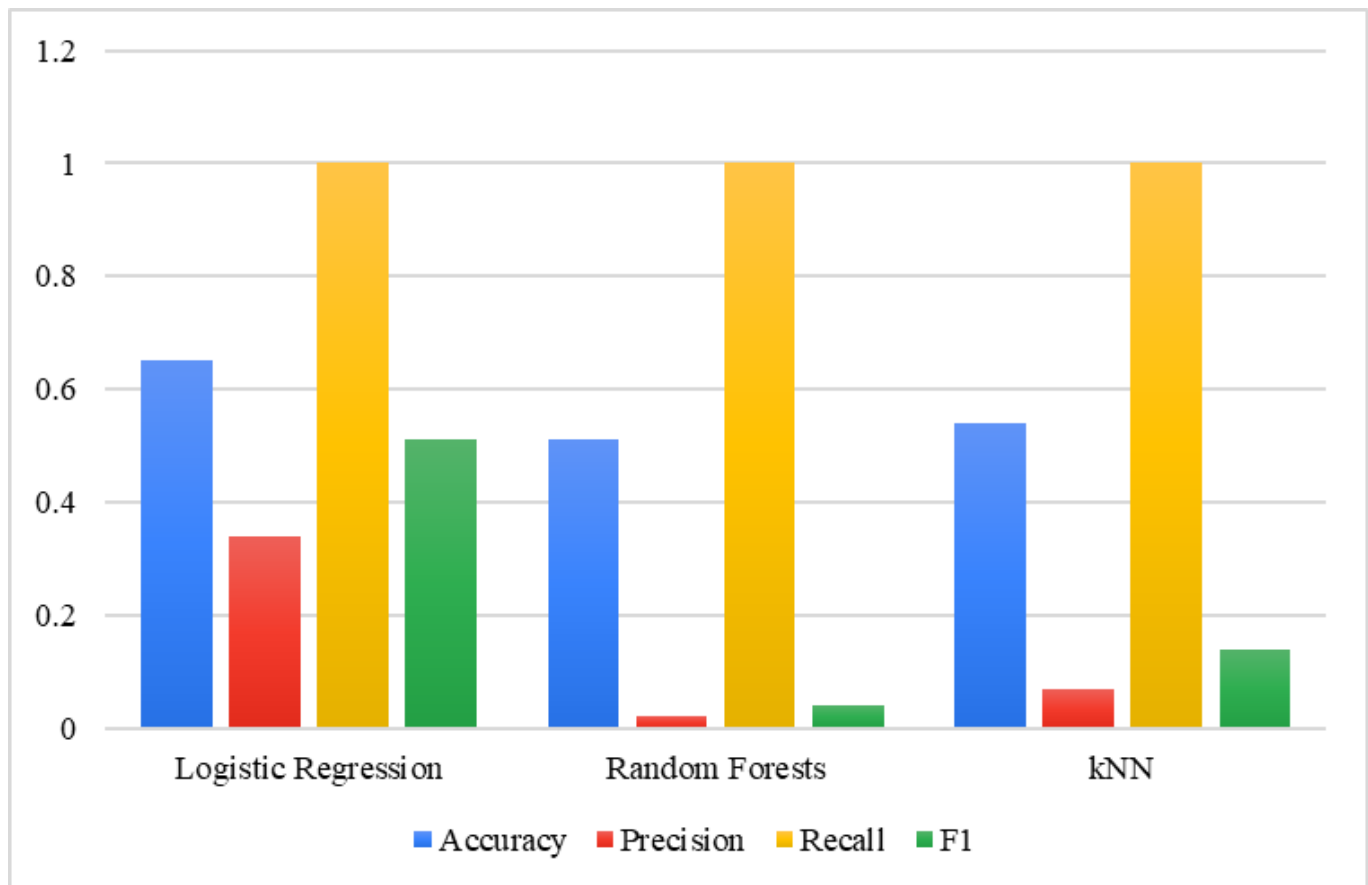
Παρατηρείται ότι για τα Ιωάννινα και οι τρεις αλγόριθμοι δεν δίνουν σημαντικά ποσοστά precision, με καλύτερο τον Logistic Regression που δίνει precision 38%. Ως προς το accuracy παρατηρείται ένα ποσοστό 69% για τον Logistic Regression και ακολουθεί ο k - NN (56%) και ο Random Forests (55%).

4.1.6 Ηράκλειο Κρήτης

Για το Ηράκλειο οι συνολικές μετεωρολογικές καταγραφές είναι 14.409 και τα events είναι 4. Στον [Πίνακα 4.1.6](#) παρατίθενται οι δείκτες απόδοσης των τριών αλγορίθμων και στην [Εικ. 4.1.6](#) ακολουθεί το σχετικό γράφημα:

Classifier:	Ηράκλειο		
	Logistic Regression	Random Forests	k - NN
Accuracy	0.65	0.51	0.54
Precision	0.34	0.02	0.07
Recall	1	1	1
F1	0.51	0.04	0.14

Πίνακας 4.1.6: Πίνακας Δεικτών Απόδοσης των Ταξινομητών για το Ηράκλειο



Εικόνα 4.1.6: Απεικόνιση της απόδοσης των ταξινομητών για το Ηράκλειο

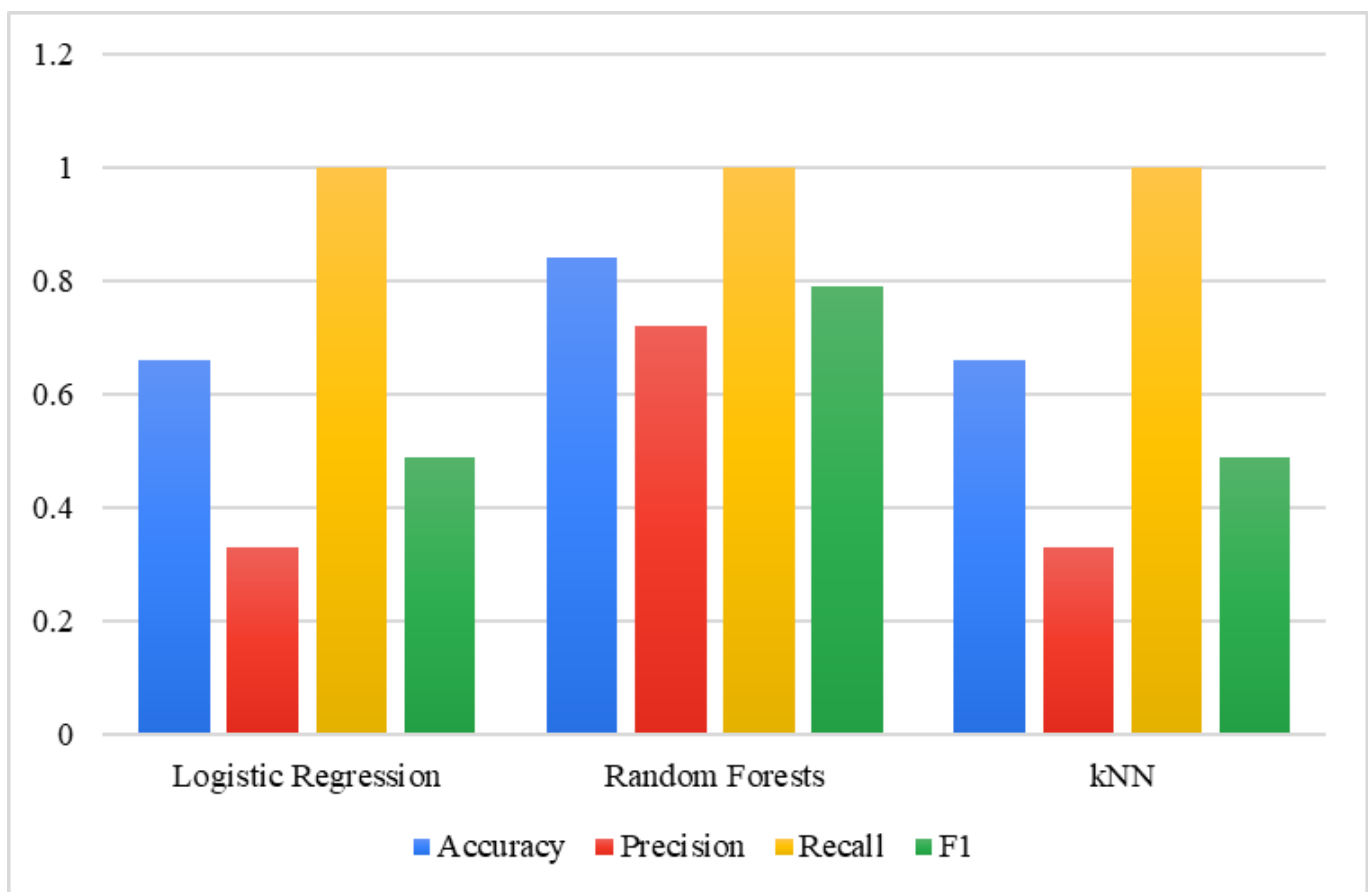
Η εικόνα των προβλέψεων στο Ηράκλειο μοιάζει σε μεγάλο βαθμό με εκείνη των Ιωάννινων όπου το precision είναι αρκετά χαμηλά, με καλύτερο τον Logistic Regression που δίνει precision 34%, ενώ για το accuracy εμφανίζει ένα ποσοστό 65% για τον Logistic Regression και ακολουθεί ο k - NN (54%) και ο Random Forests (51%).

4.1.7 Τρίπολη

Για την Τρίπολη οι συνολικές μετεωρολογικές καταγραφές είναι 14.370 και τα events είναι 36. Στον [Πίνακα 4.1.7](#) παρατίθενται οι δείκτες απόδοσης των τριών αλγορίθμων και στην [Εικ. 4.1.7](#) ακολουθεί το σχετικό γράφημα:

Classifier:	Τρίπολη		
	Logistic Regression	Random Forests	k - NN
Accuracy	0.66	0.84	0.66
Precision	0.33	0.72	0.33
Recall	1	1	1
F1	0.49	0.79	0.49

Πίνακας 4.1.7: Πίνακας Δεικτών Απόδοσης των Ταξινομητών για την Τρίπολη



Εικόνα 4.1.7: Γραφική Απεικόνιση της απόδοσης των ταξινομητών για την Τρίπολη

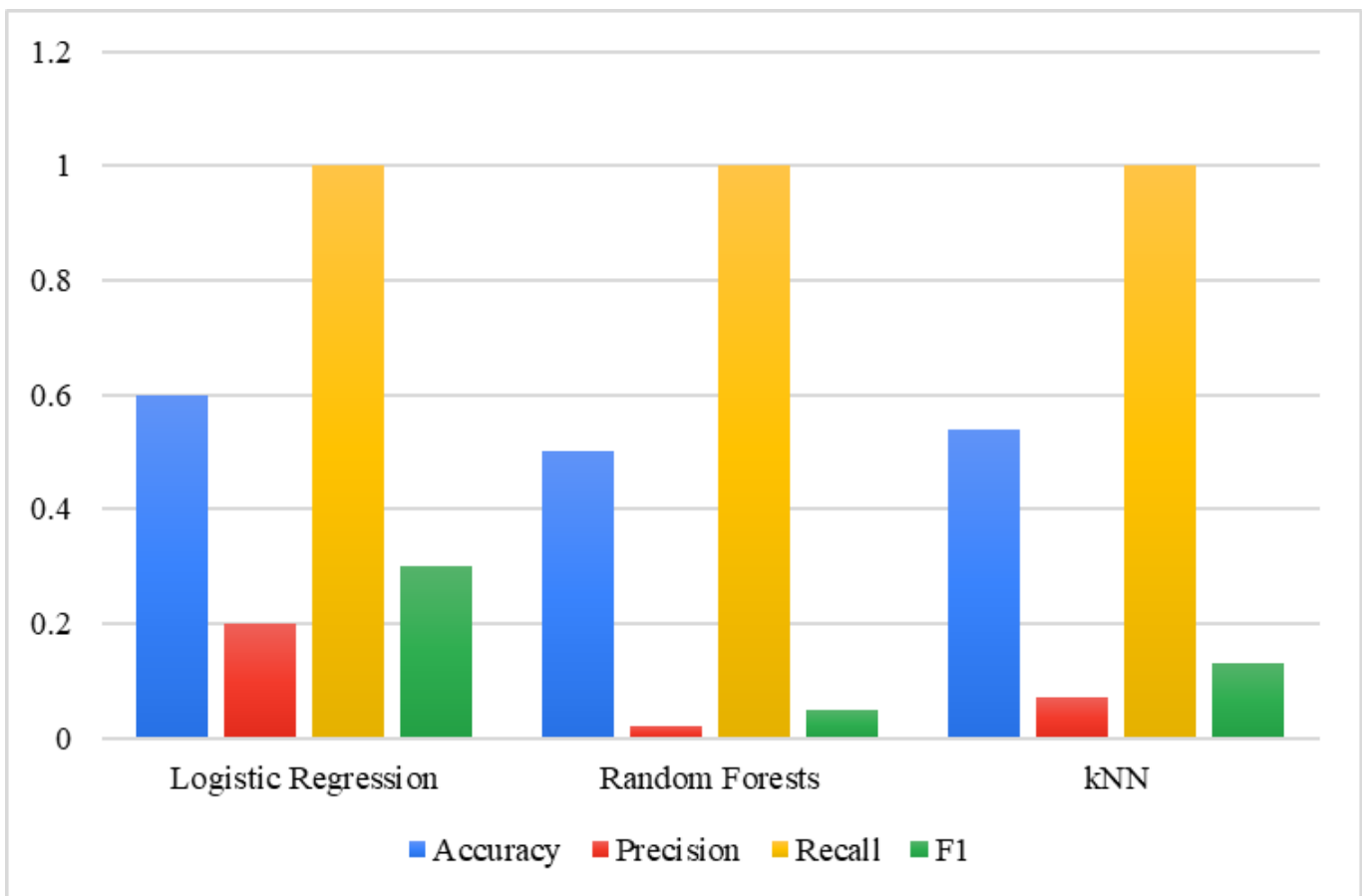
Η περίπτωση της Τρίπολης φαίνεται να διαφοροποιείται σε σχέση με τις προαναφερθείσες πόλεις καθώς φαίνεται ότι ο αλγόριθμος που δίνει καλύτερα αποτελέσματα είναι ο Random Forests. Πιο συγκεκριμένα δίνει εμφανώς καλύτερο precision (72%), σε σχέση με τους δύο άλλους αλγόριθμους (33%). Καλύτερο είναι το ποσοστό του Random Forests και για το accuracy (84%) έναντι 66% των άλλων δύο αλγορίθμων.

4.1.8 Καλαμάτα

Για την Καλαμάτα οι συνολικές μετεωρολογικές καταγραφές είναι 14.395 και τα events είναι 14. Στον [Πίνακα 4.1.8](#) παρατίθενται οι δείκτες απόδοσης των τριών αλγορίθμων και στην [Εικ. 4.1.8](#) ακολουθεί το σχετικό γράφημα:

Καλαμάτα			
Classifier:	Logistic Regression	Random Forests	k - NN
Accuracy	0.6	0.5	0.54
Precision	0.2	0.02	0.07
Recall	1	1	1
F1	0.3	0.05	0.13

Πίνακας 4.1.8: Πίνακας Δεικτών Απόδοσης των Ταξινομητών για την Καλαμάτα



Εικόνα 4.1.8: Γραφική Απεικόνιση της απόδοσης των ταξινομητών για την Καλαμάτα

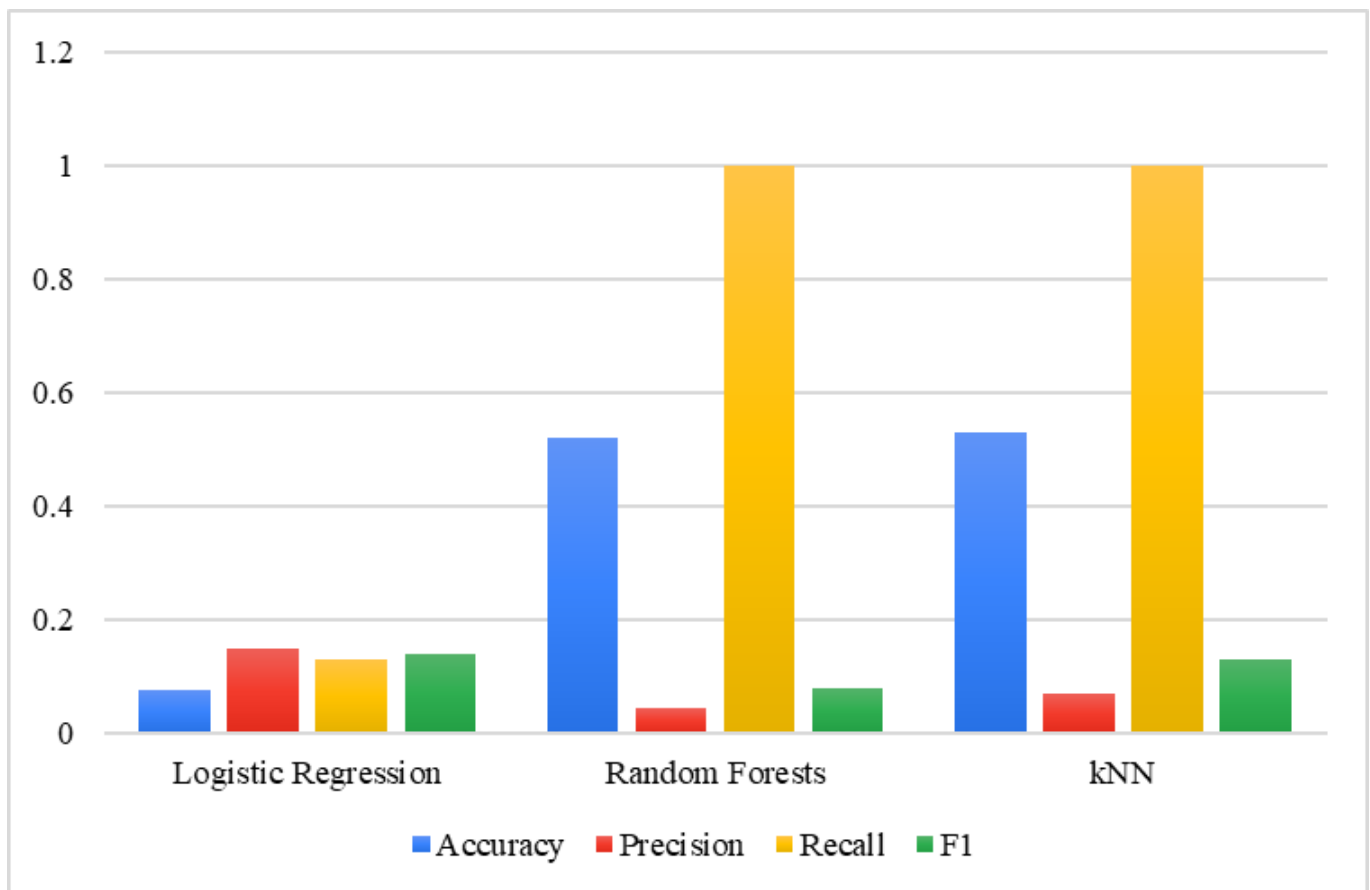
Στην περίπτωση της Καλαμάτα το επίπεδο ποιότητας των προβλέψεων δεν είναι πολύ υψηλό ωστόσο επαναλαμβάνεται το μοτίβου της καλύτερης απόδοσης του Logistic Regression σχετικά με τους δύο άλλους αλγόριθμους. Βέβαια, ακόμη και σε αυτή την περίπτωση το precision είναι της τάξης του 20%, με πολύ χαμηλότερα του k - NN (7%) ακόμη χαμηλότερα του Random Forest (2%). Το ποσοστό του accuracy φαίνεται σχετικά καλύτερο με ποσοστό 60% για τον Logistic Regression.

4.1.9 Θεσσαλονίκη

Για τη Θεσσαλονίκη οι συνολικές μετεωρολογικές καταγραφές είναι 14.445 και τα events είναι 14. Στον Πίνακα 4.1.9 παρατίθενται οι δείκτες απόδοσης των τριών αλγορίθμων και στην [Εικ. 4.1.9](#) ακολουθεί το σχετικό γράφημα:

Classifier:	Θεσσαλονίκη		
	Logistic Regression	Random Forests	k - NN
Accuracy	0.08	0.52	0.53
Precision	0.15	0.05	0.07
Recall	0.13	1	1
F1	0.14	0.08	0.13

Πίνακας 4.1.9: Πίνακας Δεικτών Απόδοσης των Ταξινομητών για τη Θεσσαλονίκη



Εικόνα 4.1.9: Γραφική Απεικόνιση της απόδοσης των ταξινομητών για τη Θεσσαλονίκη

Στην περίπτωση της Θεσσαλονίκης εμφανίζονται κακής ποιότητας αποτελέσματα για τον Logistic Regression. Το precision είναι εξαιρετικά χαμηλό και για τους τρεις αλγόριθμους, με καλύτερο το 15% του Logistic Regression ενώ το ποσοστό του accuracy για τους άλλους δύο αλγόριθμους φαίνεται να είναι σχετικά καλύτερο σε σχέση με τον Logistic Regression (53% έναντι 7,7%). Κάτι που επίσης εμφανίζεται σε αυτά τα δεδομένα για πρώτη φορά είναι το πολύ χαμηλό ποσοστό Recall του Logistic Regression (13%).



4.2 Συγκριτική Απεικόνιση των Ταξινομητών

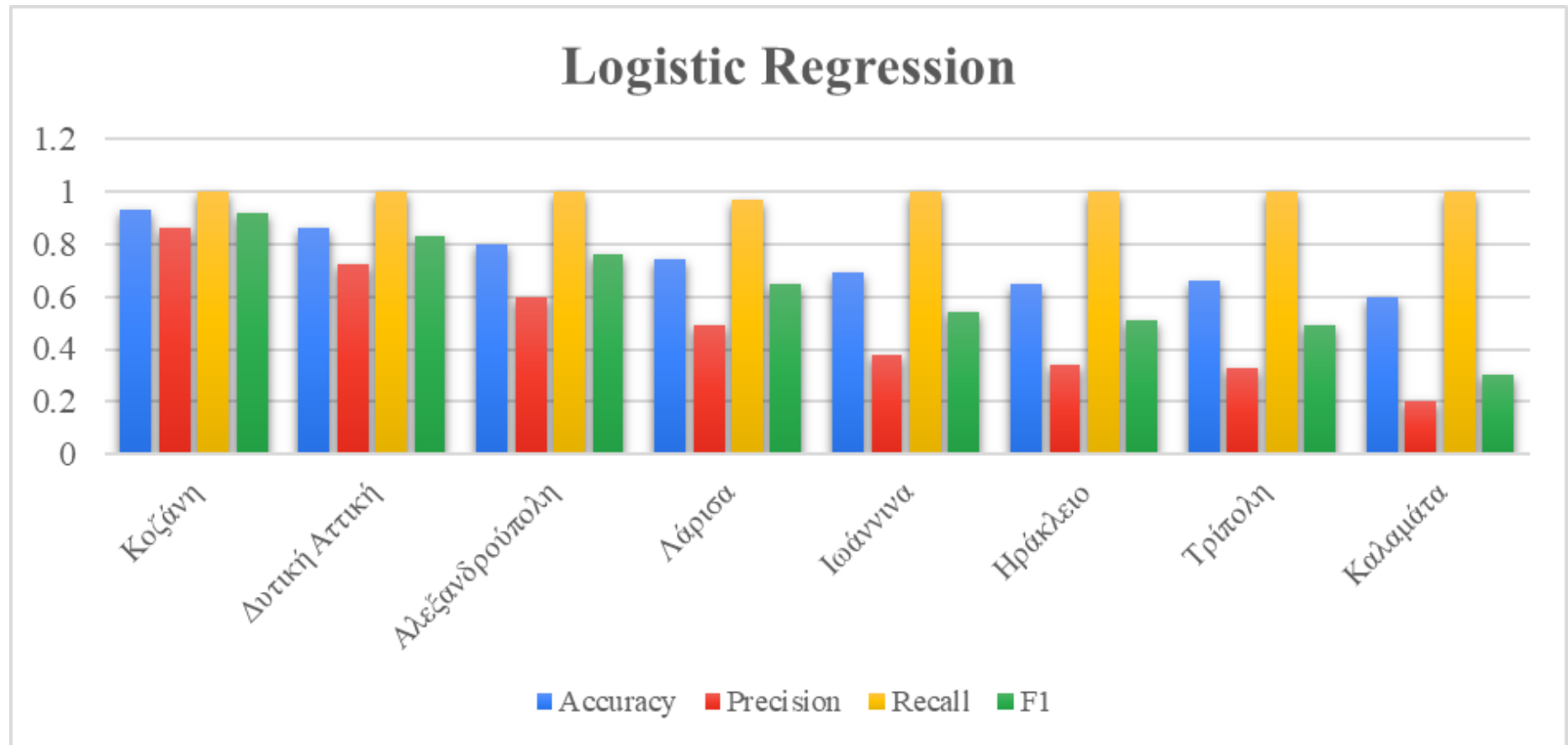
Στα γραφήματα που ακολουθούν παρουσιάζεται συγκεντρωτικά για όλες τις πόλεις που εξετάστηκαν, πλην της Θεσσαλονίκης, η απόδοση των τριών αλγορίθμων. Συγκεκριμένα, στην [Εικ. 4.2.1](#) απεικονίζεται η απόδοση του Logistic Regression, στην [Εικ. 4.2.2](#) η απόδοση του Random Forest και στην [Εικ. 4.2.3](#) η απόδοση του k - NN. Για κάθε αλγόριθμο, η παράθεση των πόλεων έχει γίνει με βάση την απόδοση των δεικτών Accuracy και Precision, ξεκινώντας από την πόλη με τους καλύτερους δείκτες

Όπως φαίνεται και από τα συγκριτικά διαγράμματα όλων των πόλεων και ισχύει και για κάθε πόλη, ο k - NN, ως πιο απλός αλγόριθμος, είναι ο λιγότερο αξιόπιστος ταξινομητής, καθώς είναι αυτός που δίνει τις χαμηλότερες τιμές για accuracy και precision, σχεδόν σε όλες τις περιπτώσεις, πέρα από την περίπτωση της Τρίπολης, που για κάποιο λόγο που χρήζει περαιτέρω διερεύνησης, δίνει ίδιες τιμές με τον Logistic Regression.

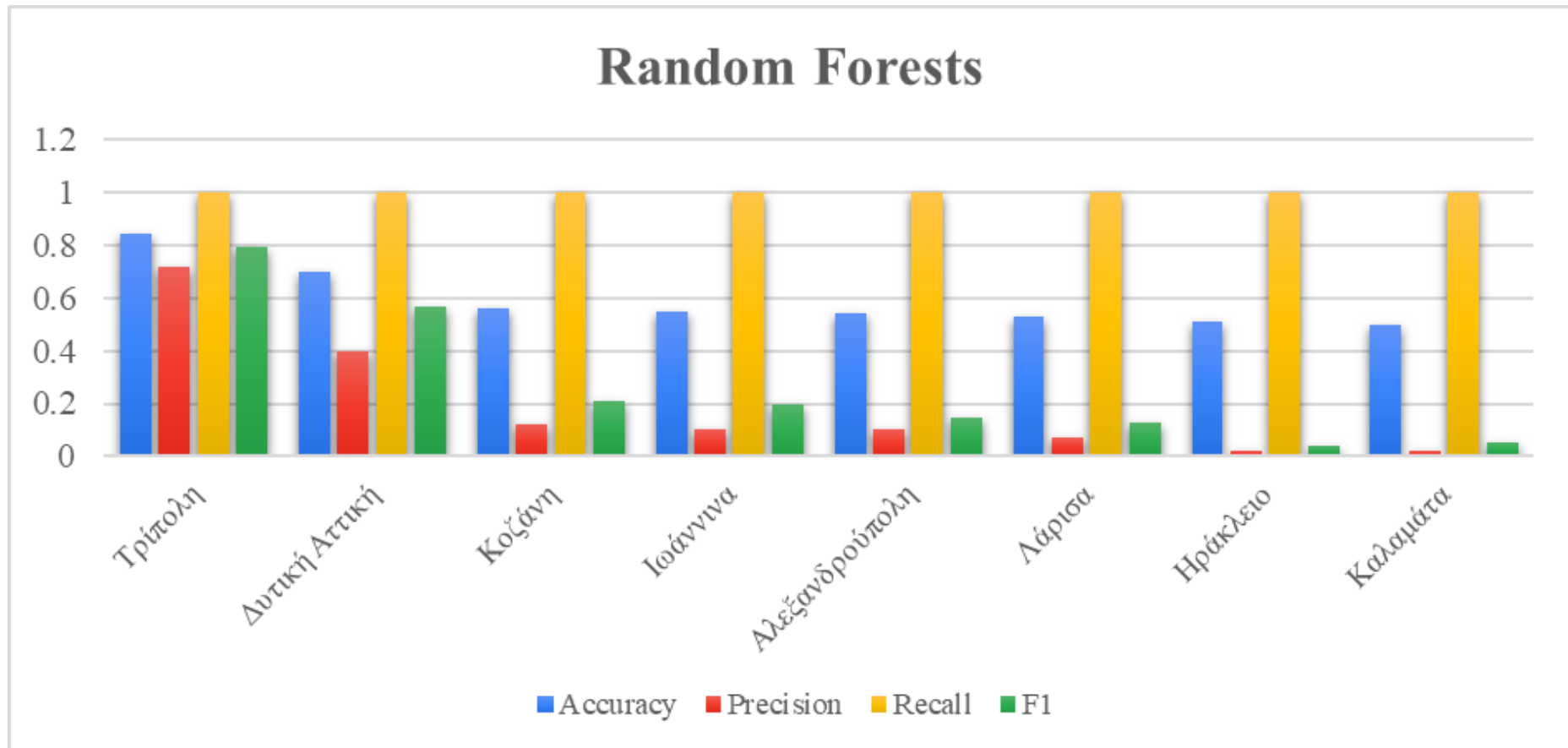
Παρατηρείται επίσης ότι σε γενικές γραμμές ο Logistic Regression είναι ο καλύτερος ταξινομητής με βάση τα δεδομένα, αποδίδοντας πολύ καλά σε όλες τις πόλεις, ωστόσο υπάρχουν περιπτώσεις που και ο Random Forests δίνει πολύ υψηλή ικανότητα πρόβλεψης.

Οι καλύτερες προβλέψεις έχουν δοθεί από τον Logistic Regression για τη Δυτική Αττική και την Κοζάνη. Στην περίπτωση της Δυτικής Αττικής, που υπάρχουν οι περισσότερες καταγραφές (70) το accuracy είναι 86% και το precision 72%, ενώ στην περίπτωση της Κοζάνης, που εντοπίστηκαν 21 φαινόμενα το accuracy είναι 93% και το precision 82%. Είναι εμφανές από τα παραπάνω, ότι η αξιοπιστία του συγκεκριμένου αλγόριθμου ως προς την πρόβλεψη έντονων φαινομένων ειδικά για τις δύο αυτές πόλεις, είναι υψηλή.

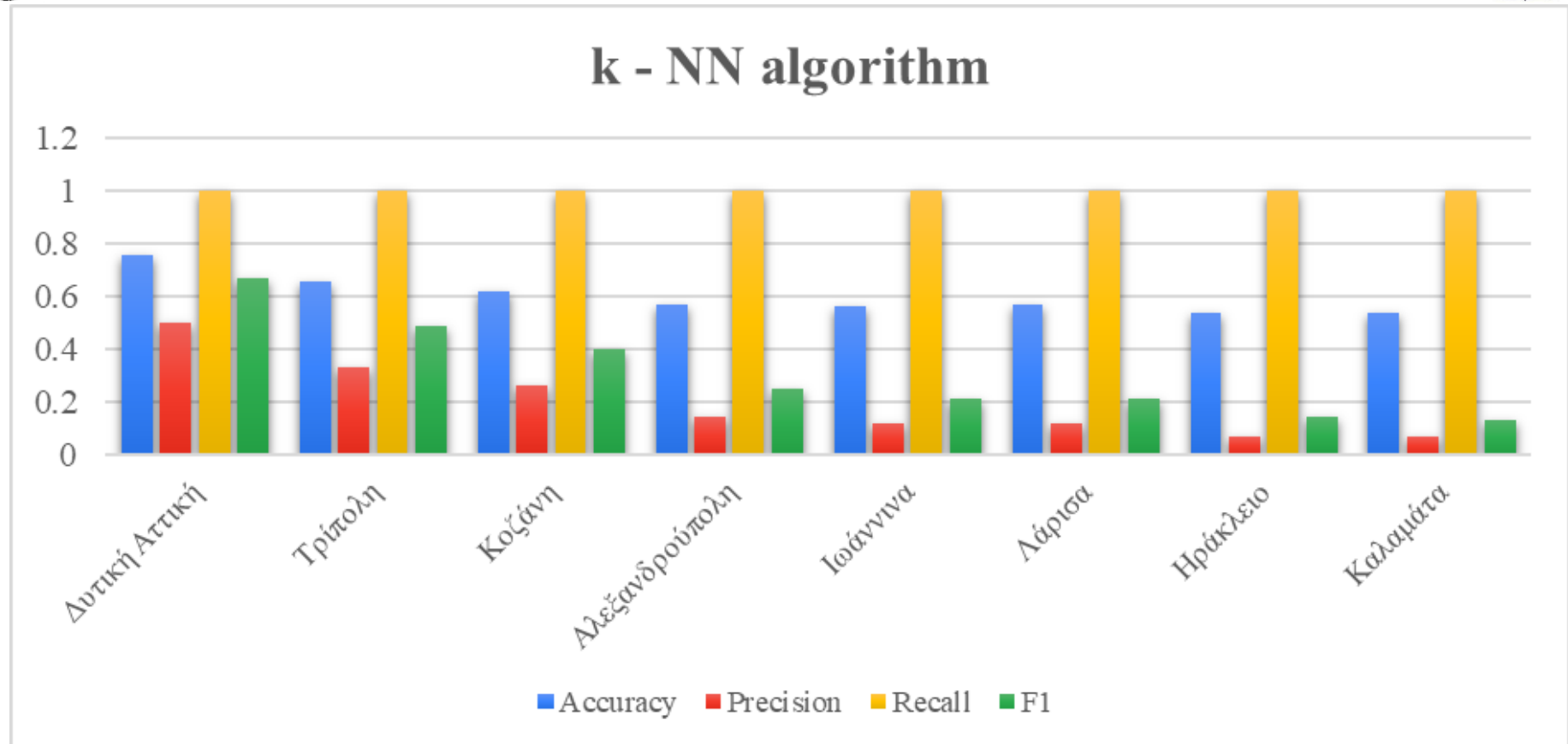
Συγκεντρωτικά, τα αποτελέσματα που παρουσιάζονται σε αυτή τη ενότητα και κατόπιν της μεταξύ τους σύγκρισης ο Logistic Regression αποτελεί την καταλληλότερη επιλογή ταξινομητή.



Εικόνα 4.2.1: Δείκτες απόδοσης Logistic Regression για όλες τις πόλεις που εξετάστηκαν



Εικόνα 4.2.2: Δείκτες απόδοσης Random Forests για όλες τις πόλεις που εξετάστηκαν



Εικόνα 4.2.3: Δείκτες Απόδοσης k-NN για όλες τις πόλεις που εξετάστηκαν



4.3 Ανάλυση Ευαισθησίας

Προκειμένου να εξετασθεί περαιτέρω η απόδοση των αλγορίθμων, εξετάστηκαν περιπτώσεις στις οποίες τα μετεωρολογικά δεδομένα δέχθηκαν συγκεκριμένες μεταβολές. Συγκεκριμένα, αυξήθηκε το ύψους υετού κατά 3 cm, μειώθηκε και αυξήθηκε η ελάχιστη και η μέγιστη θερμοκρασία κατά 3 °C και προστέθηκαν 2 κόμβοι στην ταχύτητα του ανέμου. Οι μεταβολές υετού, θερμοκρασιών και ανέμου εξετάστηκαν ξεχωριστά και στο τέλος εξετάστηκαν όλες μαζί.

4.3.1 Ανάλυση Ευαισθησίας στη μεταβολή του Ύψους Υετού

Στην περίπτωση της μεταβολής του ύψους της βροχόπτωσης κατά 3cm, οι αποδόσεις των τριών αλγορίθμων για κάθε πόλη παρουσιάζονται στον [Πίνακα 4.3.1](#) στο τέλος της ενότητας. Στην περίπτωση της μεταβολής τους ύψους υετού λιγότερο ανεπηρέαστος φάνηκε, γενικά, να μένει ο k - NN σε σχέση με τους δύο άλλους αλγόριθμους. Ειδικά για τη Δυτική Αττική και την Κοζάνη, οι τιμές των precision και accuracy για τον k - NN δεν μεταβλήθηκαν καθόλου σε σχέση με τα αρχικά δεδομένα, ενώ για τον Logistic Regression, παρατηρήθηκε μικρή μεταβολή για τις τιμές του precision. Ο Random Forests παρουσίασε μικρή μεταβολή προς τα κάτω στις τιμές των precision και accuracy, για τις συγκεκριμένες πόλεις.

4.3.2 Ανάλυση Ευαισθησίας στη μεταβολή της Μέγιστης και της Ελάχιστης Θερμοκρασίας

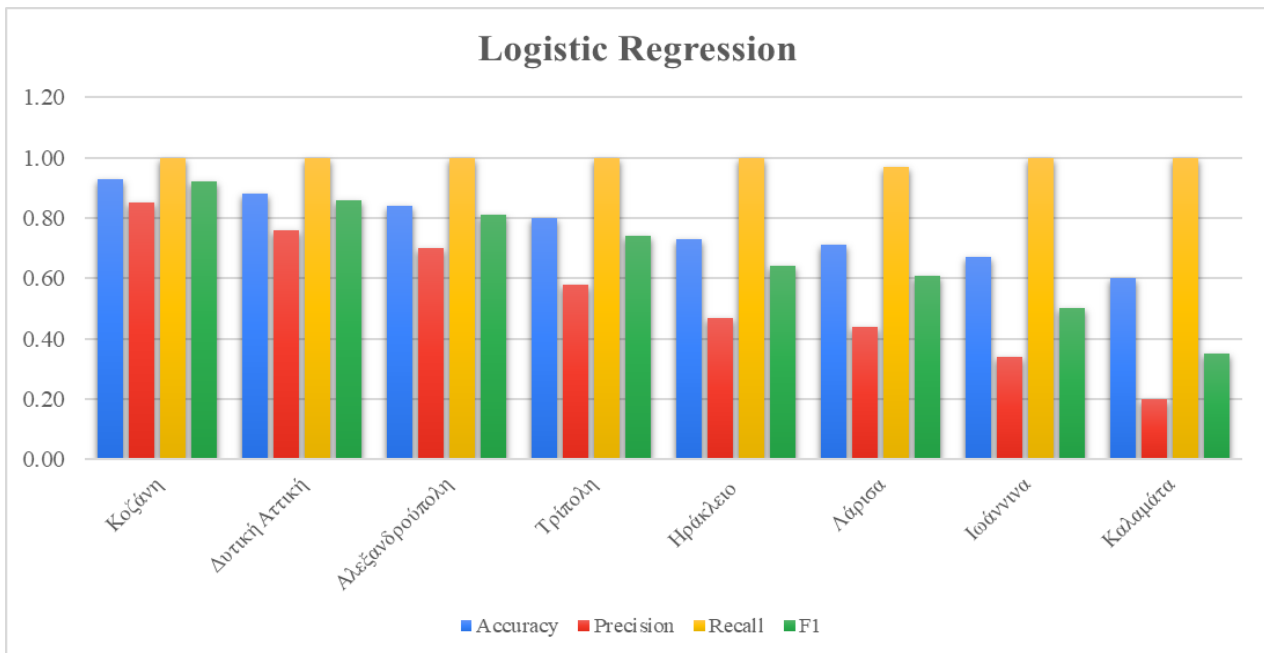
Στην περίπτωση της μεταβολής της μέγιστης και της ελάχιστης θερμοκρασίας, δηλ. πρόσθεση 3 °C στη μέγιστη θερμοκρασία και αφαίρεση 3 °C από την ελάχιστη θερμοκρασία, οι αποδόσεις των τριών αλγορίθμων για κάθε πόλη παρουσιάζονται στον [Πίνακα 4.3.2](#) στο τέλος της ενότητας. Στην περίπτωση αυτή, ως λιγότερο ανεπηρέαστος φαίνεται ο k - NN, παρουσιάζοντας τις λιγότερες μεταβολές είτε προς τα κάτω είτε τα επάνω, στις αποδόσεις των προβλέψεων μεταξύ των πόλεων. Για τη Δυτική Αττική και την Κοζάνη, ο k - NN εμφανίζει σταθερό accuracy σε σχέση με τα μη μεταβαλλόμενα δεδομένα, όμως παρουσιάζει μικρή μεταβολή προς τα κάτω για το precision. Για τις συγκεκριμένες περιοχές, οι αλγόριθμοι Logistic Regression και Random Forests φαίνεται να επηρεάζονται από τις θερμοκρασιακές μεταβολές, καθώς παρουσιάζουν προς τα κάτω μεταβολή του accuracy και του precision.

4.3.3 Ανάλυση Ευαισθησίας στη μεταβολή της ταχύτητας του ανέμου

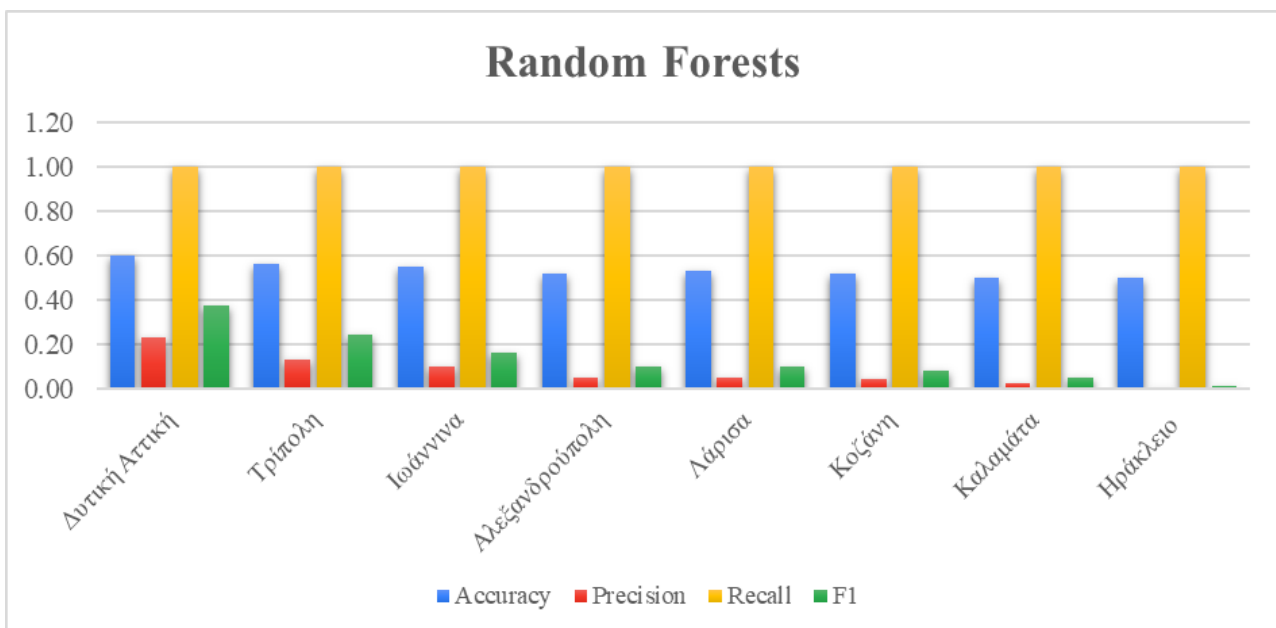
Στην περίπτωση της προσθήκης 2 κόμβων στον άνεμο, οι αποδόσεις των τριών αλγορίθμων για κάθε πόλη παρουσιάζονται στον [Πίνακα 4.3.3](#) στο τέλος της ενότητας. Στην περίπτωση της μεταβολής της ταχύτητας του ανέμου λιγότερο ανεπηρέαστος φάνηκε, γενικά, να μένει ο Random Forests σε σχέση με τους δύο άλλους αλγόριθμους. Ειδικά για τη Δυτική Αττική, οι τιμές των precision και accuracy για τον Random Forests δεν μεταβλήθηκαν καθόλου σε σχέση με τα αρχικά δεδομένα, ενώ για τον Logistic Regression, παρατηρήθηκε μικρή μεταβολή προς τα επάνω για τις τιμές των precision και accuracy. Από την άλλη πλευρά για την περίπτωση της Κοζάνης, και οι τρεις ταξινομητές παρέμειναν ανεπηρέαστοι από τη μεταβολή του ανέμου.

4.3.4 Ανάλυση Ευαισθησίας στη ταυτόχρονη μεταβολή όλων των παραμέτρων

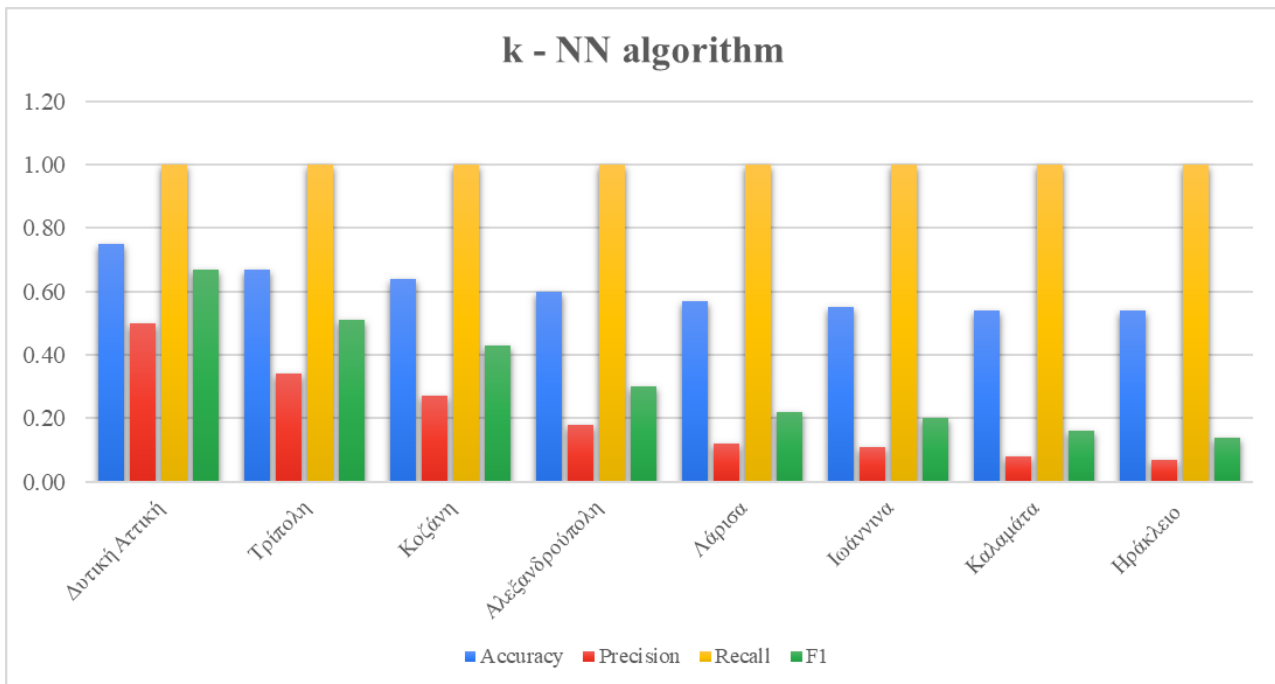
Τέλος, εξετάστηκε η περίπτωση της ταυτόχρονης μεταβολής όλων των παραμέτρων, κατά τις τιμές που παρουσιάστηκαν στις τρεις ενότητες που προηγήθηκαν. Σε αυτή την περίπτωση, οι αποδόσεις των τριών αλγορίθμων για κάθε πόλη παρουσιάζονται στον [Πίνακα 4.3.4](#) στο τέλος της ενότητας, στον οποίο οι πόλεις εμφανίζονται με αλφαβητική σειρά. Στις [Εικ. 4.3.4.1](#), [4.3.4.2](#) και [4.3.4.3](#) οι αποδόσεις των αλγορίθμων έχουν τοποθετηθεί με βάση την απόδοση των δεικτών Accuracy και Precision, ξεκινώντας από την πόλη με τους καλύτερους δείκτες.



Εικόνα 4.3.4.1: Δείκτες Απόδοσης *Logistic Regression* για όλες τις πόλεις με μεταβαλλόμενες όλες τις παραμέτρους



Εικόνα 4.3.4.2: Δείκτες Απόδοσης *Random Forests* για όλες τις πόλεις με μεταβαλλόμενες όλες τις παραμέτρους



Εικόνα 4.3.4.3: Δείκτες Απόδοσης k -NN για όλες τις πόλεις με μεταβαλλόμενες όλες τις παραμέτρους

Συνολικά, φαίνεται ότι τόσο για την περίπτωση μεταβολής των μεμονωμένων παραμέτρων, όσο για την περίπτωση που όλες οι παράμετροι μεταβάλλονται ταυτόχρονα τις πιο ακριβείς προβλέψεις δίνει και πάλι ο Logistic Regression σε σχέση με τους δύο άλλους ταξινομητές. Ωστόσο, στην περίπτωση της ταυτόχρονης μεταβολής δεν προκύπτει σαφές συμπέρασμα για το ποιος αλγόριθμος επηρεάζεται λιγότερο από τις αλλαγές, καθώς οι τρεις ταξινομητές παρουσιάζουν μεταβολές, είτε προς τα επάνω είτε προς τα κάτω για τις τιμές των accuracy και precision, σε όλες τις πόλεις που εξετάστηκαν. Επίσης παρατηρείται ότι σε όλους τους ταξινομητές και σε κάθε περίπτωση που εξετάζεται το recall παραμένει ίσο με τη μονάδα.



ΜΕΤΑΒΟΛΗ ΥΨΟΥΣ ΥΕΤΟΥ

	Logistic Regression				Random Forests				k – NN			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Αλεξανδρούπολη	0.80	0.6	1	0.76	0.52	0.05	1	0.1	0.59	0.17	1	0.30
Δυτική Αττική	0.86	0.75	1	0.86	0.64	0.30	1	0.45	0.76	0.50	1	0.67
Ηράκλειο	0.71	0.42	1	0.6	0.51	0.02	1	0.04	0.54	0.07	1	0.14
Ιωάννινα	0.70	0.40	1	0.57	0.55	0.10	1	0.20	0.56	0.12	1	0.21
Καλαμάτα	0.60	0.2	1	0.35	0.5	0.04	1	0.08	0.54	0.08	1	0.15
Κοζάνη	0.93	0.85	1	0.91	0.53	0.07	1	0.14	0.62	0.26	1	0.40
Λάρισα	0.76	0.55	0.96	0.70	0.53	0.07	1	0.13	0.57	0.14	1	0.25
Τρίπολη	0.87	0.75	1	0.86	0.58	0.17	1	0.30	0.67	0.34	1	0.50

Πίνακας 4.3.1: Πίνακας Δεικτών Απόδοσης των Ταξινομητών μετά τη μεταβολή στο Ύψος Υετού

ΜΕΤΑΒΟΛΗ ΤΙΜΩΝ ΜΕΓΙΣΤΗΣ ΚΑΙ ΕΛΑΧΙΣΤΗΣ ΘΕΡΜΟΚΡΑΣΙΑΣ

	Logistic Regression				Random Forests				k – NN			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Αλεξανδρούπολη	0.84	0.7	1	0.81	0.52	0.05	1	0.1	0.57	0.16	1	0.26
Δυτική Αττική	0.84	0.7	1	0.82	0.62	0.25	1	0.41	0.74	0.48	1	0.65
Ηράκλειο	0.70	0.41	1	0.58	0.5	0.01	1	0.02	0.54	0.08	1	0.15
Ιωάννινα	0.64	0.3	1	0.45	0.53	0.07	1	0.12	0.56	0.12	1	0.21
Καλαμάτα	0.6	0.2	1	0.30	0.5	0.02	1	0.05	0.54	0.07	1	0.13
Κοζάνη	0.91	0.82	1	0.90	0.53	0.07	1	0.13	0.62	0.24	1	0.4
Λάρισα	0.65	0.32	0.96	0.50	0.52	0.04	1	0.08	0.55	0.10	1	0.18
Τρίπολη	0.76	0.52	1	0.69	0.56	0.12	1	0.21	0.66	0.32	1	0.49

Πίνακας 4.3.2: Πίνακας Δεικτών Απόδοσης των Ταξινομητών μετά τις μεταβολές των θερμοκρασιών



ΜΕΤΑΒΟΛΗ ΚΟΜΒΩΝ ΑΝΕΜΟΥ

	Logistic Regression				Random Forests				k – NN			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Αλεξανδρούπολη	0.80	0.60	1.0	0.75	0.54	0.07	1.0	0.14	0.57	0.14	1.0	0.25
Δυτική Αττική	0.88	0.75	1.0	0.85	0.70	0.40	1.0	0.54	0.75	0.50	1.0	0.67
Ηράκλειο	0.68	0.36	1.0	0.53	0.51	0.02	1.0	0.04	0.54	0.07	1.0	0.14
Ιωάννινα	0.68	0.37	1.0	0.54	0.54	0.08	1.0	0.16	0.56	0.11	1.0	0.20
Καλαμάτα	0.60	0.18	1.0	0.30	0.50	0.02	1.0	0.04	0.54	0.07	1.0	0.13
Κοζάνη	0.93	0.86	1.0	0.92	0.56	0.12	1.0	0.21	0.62	0.26	1.0	0.40
Λάρισα	0.76	0.54	1.0	0.70	0.53	0.07	1.0	0.13	0.57	0.13	1.0	0.24
Τρίπολη	0.87	0.74	1.0	0.85	0.60	0.19	1.0	0.33	0.67	0.34	1.0	0.50

Πίνακας 4.3.3: Πίνακας Δεικτών Απόδοσης των Ταξινομητών μετά τη μεταβολή στους κόμβους του ανέμου

ΜΕΤΑΒΟΛΗ ΟΛΩΝ ΤΩΝ ΠΑΡΑΜΕΤΡΩΝ

	Logistic Regression				Random Forests				k – NN			
	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1	Accuracy	Precision	Recall	F1
Αλεξανδρούπολη	0.84	0.70	1	0.81	0.52	0.05	1	0.10	0.60	0.18	1	0.30
Δυτική Αττική	0.88	0.76	1	0.86	0.60	0.23	1	0.38	0.75	0.50	1	0.67
Ηράκλειο	0.73	0.47	1	0.64	0.50	0.01	1	0.01	0.54	0.07	1	0.14
Ιωάννινα	0.67	0.34	1	0.50	0.55	0.10	1	0.16	0.55	0.11	1	0.20
Καλαμάτα	0.60	0.20	1	0.35	0.50	0.02	1	0.05	0.54	0.08	1	0.16
Κοζάνη	0.93	0.85	1	0.92	0.52	0.04	1	0.08	0.64	0.27	1	0.43
Λάρισα	0.71	0.44	0.97	0.61	0.53	0.05	1	0.10	0.57	0.12	1	0.22
Τρίπολη	0.80	0.58	1	0.74	0.56	0.13	1	0.24	0.67	0.34	1	0.51

Πίνακας 4.3.4: Πίνακας Δεικτών Απόδοσης των Ταξινομητών μετά τις μεταβολές όλων των παραμέτρων ταυτόχρονα



5 Συμπεράσματα- Συζήτηση

5.1 Σύγκριση των ταξινομητών

Συγκρίνοντας τα αποτελέσματα που προέκυψαν από τους ταξινομητές Logistic Regression, Random Forests και k -NN (§ 4.2), φαίνεται ότι γενικά ο επικρατέστερος είναι ο Logistic Regression καθώς για τις περισσότερες πόλεις οι δείκτες του δίνουν τα καλύτερα αποτελέσματα. Η μοναδική περίπτωση που φαίνεται να υπερτερεί ο ταξινομητής Random Forests είναι στην περίπτωση της Τρίπολης (§ 4.1.7). Η περίπτωση της Θεσσαλονίκης (§ 4.1.9) εάν και παρουσιάζεται στα αποτελέσματα, εφόσον εξετάσθηκε, δεν φαίνεται να έχει προβλεφθεί σωστά από κανέναν ταξινομητή, για αυτό και δεν συμμετέχει στις συγκρίσεις δεδομένων. Σε σχέση με τους τρεις ταξινομητές, ο k -NN, με εξαίρεση την Τρίπολη που κατέληξε στα ίδια αποτελέσματα με τον Logistic Regression αλλά και οι δύο μειονεκτούσαν έναντι του Random Forests, παρέμενε σταθερά τρίτος, αποδίδοντας χαμηλότερες τιμές για τους δείκτες.

Ο ταξινομητής Logistic Regression έχει τη δυνατότητα να προβλέψει καλύτερα τα events για τα Δυτικά Προάστια της Αττικής και για την Κοζάνη, ενώ λιγότερο καλές είναι οι προβλέψεις για την Τρίπολη και την Καλαμάτα. Οι δύο πρώτες πόλεις έχουν περισσότερα καταγεγραμμένα events, συγκριτικά με τις υπόλοιπες πόλεις, εξαιρουμένης της Τρίπολης η οποία γενικά παρουσιάζει μία διαφορετική συμπεριφορά. Οι αποδόσεις για τις υπόλοιπες πόλεις βρίσκονται ενδιάμεσα. Στο ίδιο μοτίβο, αλλά με χαμηλότερης ποιότητας απόδοση, βρίσκονται και οι άλλοι δύο ταξινομητές, οι οποίοι δίνουν τις καλύτερες τους προβλέψεις για τα Δυτικά Προάστια της Αττικής και για την Κοζάνη και τις λιγότερο καλές προβλέψεις για το Ηράκλειο και την Καλαμάτα.

Ειδικά για την περίπτωση των Δυτικών Προαστίων της Αττικής, παρατηρείται ότι τα πραγματικά καιρικά φαινόμενα και οι πραγματικές ηλιοφάνειες έχουν προβλεφθεί σε ποσοστό 86% και το ποσοστό των “false alarm” είναι αρκετά μικρό. Ακόμη καλύτερα το ποσοστό πρόβλεψης των πραγματικών events και πραγματικών ηλιοφανειών είναι για την Κοζάνη 93% και ακόμη μικρότερο το ποσοστό των “false alarm” σε σχέση με τη Δυτική Αττική. Αντίστοιχα καλά αποτελέσματα με την περίπτωση της Δυτικής Αττικής φαίνεται να δίνει ο Random Forests για την Τρίπολη, εμφανίζοντας ποσοστό πρόβλεψης της πραγματικής κατάστασης 84% και μικρό ποσοστό “false alarm”. Κοινό χαρακτηριστικό και των Δυτικών Προαστίων Αττικής και της Τρίπολης είναι ότι έχουν παραπάνω καταγεγραμμένα events σε σχέση με τις άλλες πόλεις. Από την άλλη πλευρά βέβαια τα καταγεγραμμένα events της Κοζάνης είναι αρκετά λίγα και οι δείκτες της εξαιρετικά καλοί.

Παρατηρείται επίσης στα αποτελέσματα σχεδόν παντού και στους τρεις αλγόριθμους ότι το Recall είναι ίσο με τη μονάδα. Αυτό σημαίνει ότι η περίπτωση του να μην επικρατεί ηλιοφάνεια και να την έχω προβλέψει είναι μηδενική.



Συμπερασματικά, ταξινομητές γεωγραφικών περιοχών πρωτίστως της Κοζάνης και εν συνεχεία της Δυτικής Αττικής εμφανίζουν πολύ υψηλή απόδοση στην πρόβλεψη των μετεωρολογικών συνθηκών και των 9 περιοχών της Ελλάδας που εξετάστηκαν. Επομένως, μία ευρεία εφαρμογή και επαλήθευση τους θα μπορούσε να τα καταστήσει μέρος ενός EWS.

5.2 Η περίπτωση της Θεσσαλονίκης

Όπως φαίνεται από τα αποτελέσματα που προέκυψαν από τη Θεσσαλονίκη (§ 4.1.9), η ποιότητα των δεικτών και των τριών αλγορίθμων δεν είναι ικανοποιητική. Ο Logistic Regression δίνει υπερβολικά πολλά FP, γεγονός που δεν είναι επιθυμητό. Ωστόσο, θα μπορούσε κανείς να πει ότι είναι προτιμότερο να υπάρχουν περισσότερες λάθος προειδοποιήσεις παρά αιφνιδιαστικά και έντονα συμβάντα. Ο μικρός αριθμός των καταγραφών κυρίως καθώς και μία περαιτέρω ενδεχομένως προεπεξεργασία των συνολικών δεδομένων της Θεσσαλονίκης που χρησιμοποιούνται για τις προβλέψεις είναι κάτι που χρήζει περαιτέρω ανάλυσης προκειμένου να δοθούν σωστές προβλέψεις για τη συγκεκριμένη περιοχή.

5.3 Ανάλυση Ευαισθησίας

Οι τιμές που προκύπτουν από την ανάλυση ευαισθησίας φαίνονται στους πίνακες [4.3.1](#), [4.3.2](#), [4.3.3](#) και [4.3.4](#). Συγκρίνοντας τις τιμές των δεικτών για κάθε μία από τις παραπάνω περιπτώσεις με τα αποτελέσματα των πρωταρχικών δεδομένων παρατηρούμε στις περιπτώσεις των μεταβαλλόμενων τιμών ύψους υετού και ταχύτητας ανέμου περισσότερο ανεπηρέαστος είναι ο αλγόριθμος k -NN, ενώ στην περίπτωση των μεταβαλλόμενων θερμοκρασιών περισσότερο ανεπηρέαστος μένει ο Random Forests. Στην περίπτωση που μεταβάλλονται και οι τέσσερις παράμετροι ταυτόχρονα, και οι τρεις ταξινομητές επηρεάζονται σχεδόν το ίδιο. Αυτό έχει ως αποτέλεσμα να μην μπορεί να εξαχθεί ασφαλές συμπέρασμα για το ποιος ταξινομητής θα μπορούσε να δώσει πιο ακριβή αποτελέσματα, καθώς οι δείκτες απόδοσης των ταξινομητών είτε αυξάνονται είτε μειώνονται, με την επίδραση των εσφαλμένων δεδομένων.

Παρατηρούμε λοιπόν ότι το accuracy και το precision σε κάποιες περιπτώσεις αυξάνουν με τη χρήση των εσφαλμένων δεδομένων. Ωστόσο, σε όλες τις περιπτώσεις, το Recall παραμένει ίσο με τη μονάδα. Αυτό ενδεχομένως να συμβαίνει εξαιτίας του ότι οι αλγόριθμοι αντιλαμβάνονται ότι υπάρχουν περισσότερα καιρικά φαινόμενα, εξαιτίας της αύξησης των μετεωρολογικών δεδομένων που είναι μετατοπισμένα προς πιο ακραίες τιμές.

Περαιτέρω διερεύνηση της επίδρασης των εσφαλμένων δεδομένων στον Logistic Regression και τον k -NN σε σχέση με τον Random Forests θα ήταν χρήσιμη, ωστόσο ξεπερνά το πλαίσιο της συγκεκριμένης εργασίας.



5.4 Συμπεράσματα

Ο αλγόριθμος k -NN θεωρείται ο πιο απλός αλγόριθμος και αυτός που έχει τα περισσότερα σφάλματα λόγω της απλότητας του, βάσει της θεωρίας στην οποία βασίζεται. Για το λόγο αυτό και παρατηρήσαμε ότι έδωσε τα λιγότερο ακριβή αποτελέσματα για όλες τις πόλεις.

Από την άλλη πλευρά φάνηκε ότι ο αλγόριθμος Logistic Regression δίνει πιο ακριβή αποτελέσματα με βάση τα δεδομένα που έχουμε. Βέβαια, υπάρχουν περιπτώσεις που ο Random Forests εμφανίζει πολύ υψηλή ικανότητα πρόβλεψης, γεγονός που επαληθεύεται και από τη βιβλιογραφία όπου τα συμπεράσματα περί των αποδόσεων των Logistic Regression και Random Forests είναι συγκρίσιμα και τελικά, η επιλογή του καταλληλότερου αλγόριθμου εξαρτάται σε μεγάλο βαθμό από τα υπό επεξεργασία δεδομένα.

Κρίνοντας από την υψηλή απόδοση του Logistic Regression για την περίπτωση της Δυτικής Αττικής και της Κοζάνης είναι εμφανές ότι η ικανότητα των ταξινομητών δεν εξαρτάται από το πλήθος των καταγραφών που χρησιμοποιήθηκαν για να υλοποιηθούν. Επομένως, φαίνεται ότι ο τρόπος της διαχείρισης και της επιπλέον επεξεργασίας των δεδομένων είναι ένα θέμα που πρέπει να τεθεί προς περαιτέρω διερεύνηση.

Συμπερασματικά, θα μπορούσαμε να πούμε ότι ο ταξινομητής Logistic Regression, βάσει των αποδόσεων του που παρουσιάστηκαν στην εργασία, θα μπορούσε να αποτελέσει μέρος ενός Συστήματος Έγκαιρης Προειδοποίησης, προβλέποντας την ύπαρξη μελλοντικών έντονων καιρικών φαινομένων και δίνοντας το έναυσμα για τη σωστή προετοιμασία και την έγκαιρη πρόληψη.

5.5 Περιορισμοί των μοντέλων πρόβλεψης

Τα προγνωστικά μοντέλα, αν και παρέχουν μεγάλα οφέλη, δεν θεωρούνται απόλυτα ασφαλή και παρουσιάζουν συγκεκριμένα μειονεκτήματα. Τα μοντέλα πρόβλεψης χρειάζονται ένα συγκεκριμένο σύνολο συνθηκών για να λειτουργήσουν, οι οποίες εάν δεν πληρούνται τότε δεν μπορούν να αποδώσουν σωστά [33], [50].

5.5.1 Η ανάγκη για μαζικά εκπαιδευτικά σύνολα δεδομένων

Για να είναι επιτυχημένα τα μοντέλα προγνωστικής ανάλυσης στην πρόβλεψη των αποτελεσμάτων, πρέπει να υπάρχει ένα τεράστιο μέγεθος δείγματος, αντιπροσωπευτικό του πληθυσμού. Στην ιδανική περίπτωση, το μέγεθος του δείγματος θα πρέπει να είναι κάποιες χιλιάδες δεδομένα. Εάν οι βάσεις δεδομένων είναι μικρότερες από τα προγνωστικά μοντέλα ανάλυσης, οι αλγόριθμοι θα επηρεαστούν αδικαιολόγητα από ανωμαλίες στα δεδομένα, τα οποία θα παραμορφώσουν τα ευρήματα.

5.5.2 Σωστή ταξινόμηση δεδομένων

Τα προγνωστικά μοντέλα ανάλυσης βασίζονται στους αλγόριθμους μηχανικής μάθησης και αυτοί οι αλγόριθμοι μπορούν να αξιολογήσουν σωστά τα δεδομένα, στην περίπτωση που αυτά έχουν επισημανθεί σωστά. Η επισημάνση



δεδομένων είναι μία ιδιαίτερα απαιτητική και σχολαστική διαδικασία, επειδή πρέπει να είναι ακριβής. Η εσφαλμένη ταξινόμηση και επισήμανση προκαλεί πολλά προβλήματα, όπως κακή απόδοση και ακρίβεια στα ευρήματα

5.5.3 Εφαρμογή εκπαιδύσεων σε διαφορετικές περιπτώσεις

Τα μοντέλα δεδομένων παρουσιάζουν πρόβλημα γενίκευσης, δηλ. μη επαρκή ικανότητα μεταφοράς ευρημάτων από τη μία περίπτωση στην άλλη. Ενώ τα προγνωστικά μοντέλα είναι αποτελεσματικά στα ευρήματά τους για μία περίπτωση, συχνά παλεύουν να μεταφέρουν τα ευρήματά τους σε μια διαφορετική περίπτωση. Ως εκ τούτου, υπάρχουν ορισμένα ζητήματα εφαρμογής όσον αφορά στα ευρήματα που προέρχονται από ένα μοντέλο προγνωστικής ανάλυσης. Ωστόσο, υπάρχει μια λύση σε ορισμένες μεθόδους, όπως η μεταφορά μάθησης που θα μπορούσε να βοηθήσει στο μετριασμό κάποιων από αυτά τα ζητήματα.

5.6 Προγνωστικά μοντέλα στο μέλλον

Μελλοντικά, τα μοντέλα προγνωστικής ανάλυσης θα διαδραματίζουν αναπόσπαστο ρόλο σε όλες τις διαδικασίες λόγω της τεράστιας οικονομικής αξίας που δημιουργούν. Αν και δεν είναι τέλεια, η προστιθέμενη αξία που προσφέρουν είναι τεράστια τόσο σε οικονομικό όσο και σε κοινωνικό επίπεδο. Με τα προγνωστικά μοντέλα, τόσο οι κοινωνίες όσο και οι οργανισμοί έχουν την ευκαιρία να αναλάβουν δράση προληπτικά οργανώνοντας και εκτελώντας ένα μεγάλο πλήθος λειτουργιών. Η πρόληψη σε όλους τους τομείς αλλά και στο τομέα της μετεωρολογίας για την προστασία των ανθρώπινων ζωών πρωτίστως, και των υποδομών δευτερευόντως, είναι μία από τις απτές δυνατότητες που μπορούν να παρέχουν τα προγνωστικά μοντέλα πρόβλεψης και για το λόγο αυτό αποτελούν ένα άυλο περιουσιακό στοιχείο κάθε ευημερούσας κοινωνίας.



- M. Cropper and S. Sahin, «Valuing Mortality in the Context of Disaster Risks,» σε *Background paper for the World Bank*, 2009.
- 22] World_Bank, «Natural Hazards, Unnatural Disasters: Effective prevention through an economic lens,» 2010.
- 23] J. Malilay, «Public health consequences of disasters,» 1997.
- 24] M. M. M. D. a. o. T. Estrela, «Sustainable water use in Europe,» 2001.
- 25] M. Golnaraghi, «Institutional Partnerships in Multi-Hazard Early Warning Systems,» 2010.
- 26] K. S. e. al, «Effectiveness and Efficiency of Early Warning Systems for Flash Floods,» 2008.
- 27] T. V. Rogers D., «Costs and Benefits of Early Warning Systems,» Global Assessment Report on Disaster Risk Reduction, 2010.
- 28] L. H. D.J. Hand, «Discrimination and Classification,» 1985. [Ηλεκτρονικό].
- 29] W. Klecka, «Discriminant Analysis, Sage Publications,» 1980. .
- 30] C. A. Breiman L., «Random Forests, http://www.stat.berkeley.edu/~breiman/RandomForests/cc_home.htm.»
- 31] «https://en.wikipedia.org/wiki/Predictive_analytics,» [Ηλεκτρονικό].
- 32] «<https://seleritysas.com/blog/2019/12/12/types-of-predictive-analytics-models-and-how-they-work/>,» [Ηλεκτρονικό].
- 33] J. Brownlee, «<https://machinelearningmastery.com/what-is-imbalanced-classification/>,» [Ηλεκτρονικό].
- 34] .. C. M. M.-w. C. W.-t. Y., «Partitioned logistic regression for spam filtering,» 2008.
- 35] H. M. G. Bhaya, «An Evaluation Study of Machine Learning Techniques for Identifying Spam,» 2004.
- 36] C. A. Breiman L., «"Random Forests"».
- 37] F. Y. a. M. L., «"The alternating decision tree learning algorithm. In Machine Learning,"» 1999. .
- 38] «https://www.datacamp.com/?utm_source=adwords_ppc&utm_campaignid=805200711&utm_adgroupid=39268379982&utm_device=c&utm_keyword=datacamp&utm_matchtype=e&utm_network=g&utm_adpostion=&utm_creative=340731356548&utm_targetid=kwd-297372810188&utm_loc_interest_ms=,» [Ηλεκτρονικό].
- 39] P. B. a. S. R. H. P., «Choice of neighbor order in nearest-neighbor classification,» 2004.
- 40] S. R. G. H. R. S. J. Goldberger, «Neighbourhood Component Analysis. Advances in Neural Information Processing Systems,» 2005.
- 41] B. J. S. L. K. Weinberger K. Q., «Distance Metric Learning for Large Margin Nearest Neighbor Classification,» 2005.
- 42] V. Almeida, «Ship Noise Classification: A contribution to prototype based classifier design,» 2002.
- 43] D. H. a. B. Batchelor, «Experiments on the Edited Condensed Nearest Neighbor Rule,» 1978..
- 44] G. W., «The reduced nearest neighbor rule,» 1972.
- 45]



- 46] J. Z. , T. Y. Le Zhang, «An evaluation of statistical spam filtering techniques,» 2004.
N. V. Chawla, K. W. Bowyer, L. O. Hall και W. P. Kegelmeyer, «SMOTE: Synthetic Minority Over-sampling
47] Technique,» *Journal Of Artificial Intelligence Research*, τόμ. 16, pp. 321-357, 2002.
- 48] <<https://machinelearningmastery.com/smote-oversampling-for-imbalanced-classification/>,» [Ηλεκτρονικό].
- 49] Β. Χαρπαντίδης, «Ομαδοποιημένη οπτικοποίηση γεωγραφικών δεδομένων με χρήση web τεχνολογιών,» 2014.
<<https://www.riverlogic.com/blog/the-limitations-of-predictive-analytics-tools-and-why-execs-should-care>,»
50] [Ηλεκτρονικό].
- 51] Breiman, «Heuristics of instability in model selection, Technical Report,» 1994.
T. M. A. Blum, «Combining labeled and unlabeled data with co-training, Proceedings of the eleventh annual conference
52] on Computational learning theory,» 1998.
- 53] W. H. D. D, «Spam Detection using Clustering, Random Forests, and Active Learning,» 2009.
- 54] T. Hofmann, «Unsupervised Learning by Probabilistic Latent Semantic Analysis, Machine Learning,» 2001.
- 55] H. M. G.R Bhaya, «An Evaluation Study of Machine Learning Techniques for Identifying Spam,» 2004.
- 56] F. Y. a. M. L., «The alternating decision tree learning algorithm. In Machine Learning,» 1999.
- 57] J. R. Quinlan, «Bagging, boosting, and C4.5,» 1996.
- 58] P. B. a. S. R. Hall P., «Choice of neighbor order in nearest-neighbor classification,» 2004.
S. R. G. H. R. S. J. Goldberger, Neighbourhood Component Analysis. Advances in Neural Information Processing
59] Systems, 2005.
- 60] B. J. S. L. K. Weinberger K. Q., Distance Metric Learning for Large Margin Nearest Neighbor Classification, 2005.
- 61] V. Almeida, Ship Noise Classification: A contribution to prototype based classifier design, 2002.
- 62] D. H. a. B. Batchelor, «Experiments on the Edited Condensed Nearest Neighbor Rule,» 1978.
- 63] G. W., «The reduced nearest neighbor rule,» 1972.
- 64] J. R. Quinlan, «Bagging, boosting, and C4.5,» 1996.
- 65] <<https://www.meteoalarm.eu/>,» [Ηλεκτρονικό].