



**Χρήση τεχνητής νοημοσύνης για μοντελοποίηση
καταστροφών λόγω βροχοπτώσεων**

**Using Artificial Intelligence to model rainfall induced
damages**

Βαφειάδης Βασίλειος

Διατριβή για λήψη Βασικού Πτυχίου

Τμήμα Μηχανικών Βιομηχανικής Σχεδίασης και Παραγωγής

Πανεπιστήμιο Δυτικής Αττικής

Αθήνα, Σεπτέμβριος 2022



Τα μέλη της τριμελούς επιτροπής:

Γρηγόριος Νικολάου

Σουλτάνα Βασιλειάδου

Χρήστος Δρόσος

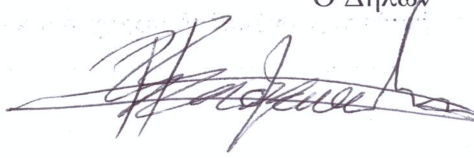
ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Βαφειάδης Βασίλειος του Κων/νου , με αριθμό μητρώου 71441887 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής Μηχανικών του Τμήματος Βιομηχανικής Σχεδίασης και Παραγωγής, δηλώνω υπεύθυνα ότι:

«Είμαι συγγραφέας αυτής της πτυχιακής/διπλωματικής εργασίας και ότι κάθε βοήθεια την οποία είχα για την προετοιμασία της είναι πλήρως αναγνωρισμένη και αναφέρεται στην εργασία. Επίσης, οι όποιες πηγές από τις οποίες έκανα χρήση δεδομένων, ιδεών ή λέξεων, είτε ακριβώς είτε παραφρασμένες, αναφέρονται στο σύνολό τους, με πλήρη αναφορά στους συγγραφείς, τον εκδοτικό οίκο ή το περιοδικό, συμπεριλαμβανομένων και των πηγών που ενδεχομένως χρησιμοποιήθηκαν από το διαδίκτυο. Επίσης, βεβαιώνω ότι αυτή η εργασία έχει συγγραφεί από μένα αποκλειστικά και αποτελεί προϊόν πνευματικής ιδιοκτησίας τόσο δικής μου, όσο και του Ιδρύματος.

Παράβαση της ανωτέρω ακαδημαϊκής μου ευθύνης αποτελεί ουσιώδη λόγο για την ανάκληση του πτυχίου μου».

Ο Δηλών



Βαφειάδης Βασίλειος

Περίληψη

Οι καταστροφές από τα καιρικά φαινόμενα είναι κάτι που κοστίζει υπέρογκα αθροιστικά χρηματικά ποσά, δημιουργεί μεγάλες διαταραχές σε κοινότητες και πολύ συχνά στοιχίζει ακόμα και ανθρώπινες ζωές. Είναι επιτακτική ανάγκη να υπάρξει έγκαιρη ενημέρωση ώστε να μπορούμε να προφυλαχθούμε όσο το δυνατόν περισσότερο.

Στην παρούσα εργασία μοντελοποιήσαμε τις καταστροφές που προήλθαν από βροχοπτώσεις με την χρήση μοντέλων μηχανικής μάθησης. Τα δεδομένα μας είναι ζευγάρια αθροιστικής εικοσιτετράωρης βροχόπτωσης και ύπαρξης ή μη καταστροφής και προέρχονται από το Εθνικό Αστεροσκοπείο Αθηνών. Το γεωχωρικό πλαίσιο αναφοράς είναι η μητροπολιτική περιοχή των Αθηνών, στην οποία λειτουργούν 66 αυτόματοι μετεωρολογικοί σταθμοί.

Χρησιμοποιήσαμε τρία διαφορετικά μοντέλα επιβλεπόμενης μάθησης ώστε να δούμε τα αποτελέσματά τους, αρχικά σε ολόκληρη την περιοχή και έπειτα στους δήμους που την απαρτίζουν. Εκπαιδεύσαμε τα μοντέλα ώστε να είμαστε σε θέση να προβλέψουμε με σχετική ακρίβεια τις περιπτώσεις εκείνες που τα καιρικά φαινόμενα θα είναι αρκετά σφοδρά για να δημιουργήσουν καταστροφές. Όπως κάθε τύπος τεχνητής νοημοσύνης, τα μοντέλα μας λειτουργούν μόνο στις περιπτώσεις που υπάρχουν επαρκή δεδομένα.

Λέξεις - κλειδια: Τεχνητή Νοημοσύνη, T.N., Μηχανική Μάθηση, M.M., επιβλεπόμενη μάθηση, καταστροφές, καιρικά φαινόμενα, Αθήνα

Abstract

Disasters from weather events cost enormous sums of money, cause great disruption to communities and very often even cost human lives. It is imperative that there are timely warnings and notifications so that we can protect ourselves as much as possible. In this thesis we modelled the disasters caused by rainfall using machine learning models.

Our data consists of pairs of cumulative twenty-four-hour rainfall and presence or absence of disaster and come from the National Observatory of Athens. The geospatial reference frame is the metropolitan area of Athens, in which 66 automatic weather stations operate.

We used three different models of supervised learning to see their results, first in the whole region and then in the municipalities that make it up. We trained the models to be able to predict with relative accuracy those cases when weather events will be severe enough for disasters to occur. Like any type of artificial intelligence, our models only work when there is enough data.

Keywords: Artificial Intelligence, A.I., Machine Learning, M.L., supervised learning, damages, weather events, Athens

Περιεχόμενα

Περίληψη	3
Abstract	4
Περιεχόμενα	5
Πίνακας Εικόνων	7
Πίνακας Πινάκων	9
Εισαγωγή	10
Κεφάλαιο 1	
Τεχνητή νοημοσύνη	12
1.1. Εισαγωγή	12
1.2. Μάθηση στην Τεχνητή Νοημοσύνη	12
1.2.1. Επιβλεπόμενη μάθηση	14
1.2.2. Μη Επιβλεπόμενη μάθηση Unsupervised Learning	14
1.2.3. Μάθηση Μερικής Επίβλεψης	15
Κεφάλαιο 2	
Μηχανική μάθηση και καιρός	16
2.1. Καιρός	16
2.2. Καιρός και Κλίμα	17
2.3. Πρόγνωση	17
Κεφάλαιο 3	
Μηχανική μάθηση και καταστροφές	20
3.1. Έγκαιρη προειδοποίηση καταστροφών	22
Κεφάλαιο 5	
Υλικό Μελέτης	23
5.1. Δεδομένα και πηγές	23
5.2. Περιοχή μελέτης	25
5.3. Υπόβαθρο δεδομένων	26
5.4. Ανάλυση συμβάντων βροχής	27
5.5. Περιγραφή των δεδομένων και στατιστικά στοιχεία	28

Κεφάλαιο 6	
Μέθοδοι	34
6.1 Μοντέλα Μηχανικής μάθησης	34
6.1.1 Δέντρο αποφάσεων	34
6.1.2 Τυχαίο δάσος	35
6.1.3 K κοντινότεροι γείτονες	37
6.1. Κριτήρια αξιολόγησης	38
6.1.1. Αληθώς Θετικά	38
6.1.2. Αληθώς Αρνητικά	39
6.1.3. Ψευδώς Θετικά	39
6.1.4. Ψευδώς Αρνητικά	39
6.1.5. Πίνακας Σύγκρισης	39
6.1.6. Εξειδίκευση	41
6.1.7. Ακρίβεια	41
6.1.7. Ανάκληση	41
6.1.8. Μέτρο F1	42
Κεφάλαιο 7	
Αποτελέσματα	43
7.1 Ανάλυση στην Μητροπολιτική Περιοχή των Αθηνών ως σύνολο	43
7.2 Ανάλυση ανά δήμο	46
Κεφάλαιο 8	
Επίλογος	49
Ευχαριστίες	50
Βιβλιογραφία	51
Παράρτημα	61

Πίνακας Εικόνων

Εικόνα 1.1: Διαγραμματική ανάλυση της μηχανικής μάθησης.	13
Εικόνα 1.2: Διαγραμματική απεικόνιση των τύπων μάθησης.	14
Εικόνα 5.1: Αυτόματος μετεωρολογικός σταθμός στο Ε.Α.Α. στην Π. Πεντέλη.	23
Εικόνα 5.2: ΑΜΣ στην μητροπολιτική περιοχή των Αθηνών.	25
Εικόνα 5.3: Κατανομή συχνότητας (histogram) ύψους R24 στα συμβάντα βροχής.	29
Εικόνα 5.4: Στατιστική κατανομή (boxplot) του R24 σε σχέση με την δυναδική μεταβλητή ύπαρξης ζημιάς, DO.	29
Εικόνα 5.5: Στατιστική κατανομή (boxplot) των συμβάντων βροχής, των καταστροφών και του ποσοστού του δεύτερου στο πρώτο, για κάθε δήμο.	32
Εικόνα 6.1: Διάγραμμα δέντρου αποφάσεων.	34
Εικόνα 6.2: Διάγραμμα δέντρου αποφάσεων.	35
Εικόνα 6.3: Κριτήρια αξιολόγησης.	37
Εικόνα 7.1: Πίνακας σύγκρισης μοντέλου DT για την ΜΠΑ.	43

Εικόνα 7.2: Πίνακας σύγκρισης μοντέλου *RF*

για την *ΜΠΑ*.

44

Εικόνα 7.3: Πίνακας σύγκρισης μοντέλου *KNN*

για την *ΜΠΑ*.

45

Πίνακας Πινάκων

Πίνακας 5.1: Βασικά στατιστικά στοιχεία μεταβλητών.	28
Πίνακας 5.2: Πίνακας στατιστικής συσχέτισης μεταβλητών.	28
Πίνακας 5.3 : Πίνακας συμβάντων - καταστροφών ανά δήμο.	31
Πίνακας 6.1: Πίνακας Σύγκρισης 3 τάξεων.	39
Πίνακας 6.2: Ειδική περίπτωση Πίνακα σύγκρισης.	39
Πίνακας 7.1: Μετρικά στοιχεία μοντέλου DT για την ΜΠΑ.	43
Πίνακας 7.2: Μετρικά στοιχεία μοντέλου RF για την ΜΠΑ.	44
Πίνακας 7.3: Μετρικά στοιχεία μοντέλου KNN για την ΜΠΑ.	44
Πίνακας 7.4: Οι 10 δήμοι της ΜΠΑ με τα καλύτερα μετρικά στοιχεία με βάση το μοντέλο DT.	46
Πίνακας 7.5: Οι 10 δήμοι της ΜΠΑ με τα καλύτερα μετρικά στοιχεία με βάση το μοντέλο RF.	46
Πίνακας 7.6: Οι 10 δήμοι της ΜΠΑ με τα καλύτερα μετρικά στοιχεία με βάση το μοντέλο KNN.	47

Εισαγωγή

Οι φυσικές καταστροφές είναι γεγονότα τα οποία είναι δύσκολο να αποφύγει κανείς και αποτελούν όλο και μεγαλύτερο κίνδυνο για τις σημερινές κοινωνίες, δεδομένου ότι τις τελευταίες τέσσερις δεκαετίες υπάρχει μία ολοένα αυξανόμενη τάση στην συχνότητα εμφάνισής τους (Hoeppe et al., 2016). Η σημασία τους είναι τόσο μεγάλη για την ανθρωπότητα, ώστε να είναι απαραίτητη η μελέτη και η από κοινού οργάνωση στην αντιμετώπισή τους (Ghosh & Sar, 2021). Οι υλικές ζημιές, οι εκτεταμένες καταστροφές σε υποδομές αλλά και η απώλεια ανθρώπινων ζωών δεν είναι παρά μερικά από τα αποτελέσματά τους (Petrucci et al., 2018; Petrucci et al., 2019). Ακόμα, δεδομένης της παγκόσμιας οικονομίας και του εμπορικού δικτύου, που δεν περιορίζεται πια στα γεωγραφικά πλαίσια της κάθε χώρας ξεχωριστά, μία καταστροφή σε ένα σημείο του πλανήτη μπορεί να έχει έναν παγκόσμιο οικονομικό απόηχο.

Μέρος της έκθεσης στους κινδύνους μίας καταστροφής, μπορεί να περιοριστεί με κατάλληλες και έγκαιρες δράσεις, όπως η πρόβλεψη των καταστροφών που θα ακολουθήσουν των επικίνδυνων φαινομένων (Paragiannaki et al., 2017). Τα μοντέλα πρόβλεψης του καιρού μπορούν να αποτελέσουν ένα σημαντικό εργαλείο πρόγνωσης ακραίων φαινομένων, όμως έρευνες έχουν δείξει πως ιδιαίτερα οι αστικές περιοχές είναι επιρρεπείς σε ξαφνικές πλημμύρες (flash floods) που προκαλούνται από μικρότερα ποσά νετού (Paragiannaki et al., 2013; Paragiannaki et al., 2015; Faccini et al., 2021). Οι αστικές πλημμύρες είναι, ως επί το πλείστον, επιφανειακά ύδατα τα οποία προκαλούν τοπικά φαινόμενα, ως εκ τούτου η αξιολόγηση των κινδύνων των πλημμυρών θα πρέπει να γίνεται σε τοπικό επίπεδο (Bernet et al., 2019). Η δυσκολία ικανοποιητικής πρόβλεψης, καθιστά απαραίτητη τη χρήση εξελιγμένων μεθόδων, μεταξύ αυτών και της μηχανικής μάθησης/τεχνητής νοημοσύνης.

Μία βασική προϋπόθεση της αντιμετώπισης αυτής της πρόκλησης είναι να βρεθούν αξιόπιστα δεδομένα τόσο κινδύνου, όσο και των επιπτώσεων σε τοπικό επίπεδο. Για παράδειγμα, λόγω της έλλειψης τέτοιων δεδομένων σε οικονομικές απώλειες, κάποιες

μελέτες που αναφέρονται σε Ευρωπαϊκές πόλεις χρησιμοποιούν εναλλακτικές πηγές ως δείκτες των επιπτώσεων, όπως κλήσεις στο πυροσβεστικό σώμα (Paragiannaki et al., 2015), αιτήματα που σχετίζονται με αιτήσεις αποζημίωσης σε ασφαλιστικές εταιρείες από μετεωρολογικές υπηρεσίες (Barbería et al., 2014) ή ακόμη και αναφορές πλημμυρών από πληθοπορισμό (Tian et al., 2019). Τα στοιχεία οικονομικού απολογισμού, όπως αυτά των ασφαλιστικών εταιρειών, είναι σπάνια, όμως όταν είναι διαθέσιμα μπορούν να αποτελέσουν έναν αξιόπιστο δείκτη καταστροφής από πλημμύρες και καταιγίδες (Blumenthal et al., 2019; André et al., 2013; Spekkers et al., 2014; Torgersen et al., 2015). Ο σημαντικός αντίκτυπος των ακραίων φαινομένων βροχόπτωσης σε οικονομικούς όρους έχει καταγραφεί με πολλαπλούς τρόπους ειδικά από ασφαλιστικές εταιρίες (Munich Re, 2022). Επιστήμονες έχουν χρησιμοποιήσει δεδομένα ασφαλιστικών εταιρειών για να εξετάσουν τον ρόλο της κοινωνικο-περιβαλλοντικής τρωτότητας ή της τρωτότητας των δομών από κινδύνους βροχοπτώσεων και αστικών πλημμυρών (Spekkers et al., 2015; Leal et al., 2020), ώστε να αναπτύξουν εξισώσεις κινδύνου για παραθαλάσσιες πλημμύρες και καταιγίδες (Prahl et al., 2015; Moncoulon et al., 2014) και να μοντελοποιήσουν την οικονομική έκθεση σε πλημμύρες για την ασφαλιστική αγορά (Moncoulon et al., 2014). Ταυτόχρονα, μεθοδολογικές δυσκολίες που αφορούν την συλλογή και την επεξεργασία πρωτογενών ασφαλιστικών δεδομένων έχουν προκαλέσει ιδιαίτερη ανησυχία στους ερευνητές (Cortès et al., 2018; Spekkers et al. 2015; Zhou et al. 2013), καθιστώντας την ανάλυση τέτοιων δεδομένων ένα μεθοδολογικό κατόρθωμα. Σε γενικές γραμμές, τα ασφαλιστικά σετ δεδομένων αποτελούν μία πολλά υποσχόμενη πηγή εκτίμησης μετεωρολογικών καταστροφών.

Σε αυτό ακριβώς το σημείο είναι που η Τεχνητή Νοημοσύνη μπορεί να βοηθήσει. Ενώ ένα αριθμητικό μοντέλο βασίζεται στις εξισώσεις και θα δώσει το ίδιο απόλυτο αποτέλεσμα βασισμένο σε αυτές, ένα μοντέλο Μηχανικής Μάθησης μπορεί να ανανεώνεται και να τροφοδοτείται συνέχεια με νέα δεδομένα, αυξάνοντας έτσι την χρονοσειρά ώστε να συνεχίσει να βελτιώνεται. Επίσης, αντιθετα με τα αριθμητικά μοντέλα, σε πολλές περιπτώσεις μπορούν να χρησιμοποιηθούν και εναλλακτικά δεδομένα, όπως πληθυσμιακά. Αυτό καθιστά τα μοντέλα M.M. Μία, ενδεχομένως, καλύτερη λύση για τέτοιες προβλέψεις.

Κεφάλαιο 1

Τεχνητή νοημοσύνη

1.1. Εισαγωγή

Τις τελευταίες δεκαετίες έχουν υπάρξει πολλοί ορισμοί για την Τεχνητή Νοημοσύνη (T.N.; Artificial Intelligence, A.I.). Ο John McCarthy (2004) στην εργασία του “What is A.I.” το έδωσε τον παρακάτω:

“Είναι η επιστήμη και η μηχανική που κάνει τις μηχανές έξυπνες, ειδικότερα έξυπνα προγράμματα ηλεκτρονικών υπολογιστών. Σχετίζεται με την παρόμοια δουλειά του να χρησιμοποιείς Η/Υ για να κατανοήσεις την ανθρώπινη νοημοσύνη, όμως η T.N. δεν είναι αναγκασμένη να περιοριστεί στις μεθόδους οι οποίες είναι βιολογικά παρατηρήσιμες.”

Στο άκουσμα του όρου Τεχνητή Νοημοσύνη ενδεχομένως το μυαλό των περισσότερων ανθρώπων να ανταποκρίνεται με εικόνες από ρομπότ. Η βιομηχανία ταινιών επιστημονικής φαντασίας είναι αυτή που έχει συμβάλλει σε μεγάλο βαθμό σε αυτό το στερεότυπο, μίας ανθρωπομορφης μηχανής που σπέρνει την καταστροφή στο διάβη της. Στην πραγματικότητα τίποτα δεν θα μπορούσε να είναι πιο μακριά από την αλήθεια.

1.2. Μάθηση στην Τεχνητή Νοημοσύνη

Η T.N. βασίζεται στην αρχή, πως η ανθρώπινη νοημοσύνη μπορεί να καθοριστεί με τέτοιο τρόπο που να επιτρέπει την αναπαραγωγή της από ένα πρόγραμμα, από τα πιο απλοϊκά μέχρι και τα πιο σύνθετα προβλήματα. Ένας από τους στόχους είναι να καταφέρει να μιμηθεί την ανθρώπινη νοητική δραστηριότητα. Όσο η τεχνολογία αναπτύσσεται, παλαιότερα ορόσημα που χαρακτήριζαν την T.N. θεωρούνται ξεπερασμένα. Χαρακτηριστικό παράδειγμα αποτελεί η ανάγνωση κειμένου μέσω οπτικής αναγνώρισης χαρακτήρων (Optical Character Recognition, OCR) όπου δεν

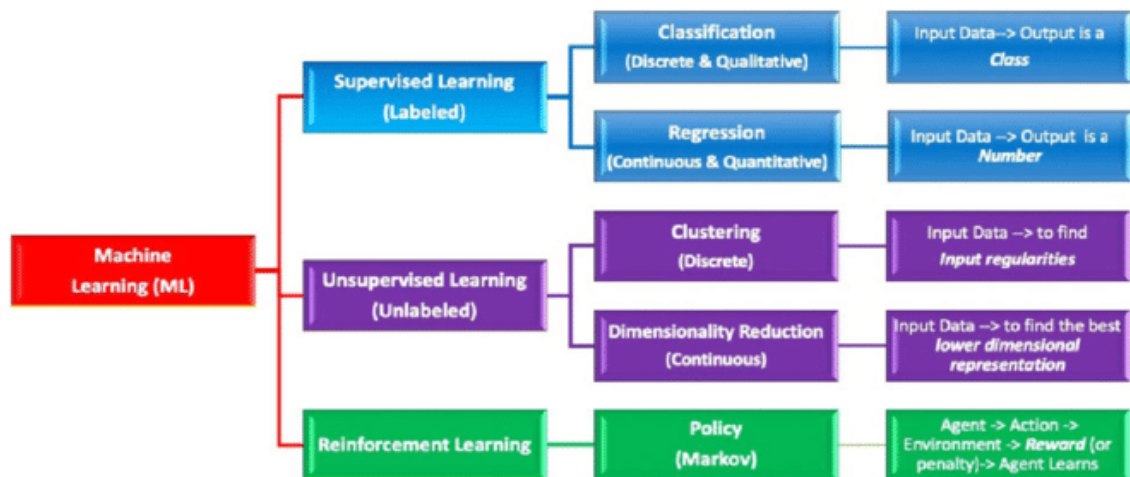
ενσαρκώνει πια την Τ.Ν., αφού αυτό το χαρακτηριστικό θεωρείται πλέον δεδομένο και σε πολλές περιπτώσεις έρχεται προεγκατεστημένο ακόμα και σε έξυπνα κινητά τηλέφωνα (smartphones).

Η Μηχανική Μάθηση (Μ.Μ.; Machine Learning, ML) είναι ένα υποσύνολο της Τ.Ν., υπό την έννοια πως αναφερόμενοι σε Μ.Μ. αναφερόμαστε πάντα σε Τ.Ν., χωρίς όμως το αντίθετο να ισχύει πάντα. Θα μπορούσαμε να ορίσουμε την Μ.Μ. ως το κομμάτι της Τ.Ν. που φροντίζει τα συστήματα να έχουν την δυνατότητα να μαθαίνουν αυτόματα και να βελτιώνονται με βάση την εμπειρία τους και χωρίς να έχουν προγραμματιστεί για κάτι.

Ως επίκεντρο, η Μ.Μ., έχει την ανάπτυξη διεργασιών Η/Υ που μπορούν να κάνουν χρήση δεδομένων ώστε να μάθουν από αυτά, αυτόνομα. Η εκμάθηση, ως διαδικασία, έχει αφετηρία την παρατήρηση των δεδομένων υπό την έννοια παραδειγμάτων ή κάποιου τύπου οδηγίας, με στόχο την εύρεση μοτίβων στα δεδομένα αλλά και την απόκτηση μίας βάσης γνώσης, την οποία θα μπορέσει να χρησιμοποιήσει σε μελλοντικό χρόνο για να πάρει αποφάσεις. Συνοψίζοντας, ο απώτερος σκοπός της Μ.Μ., είναι να επιτρέψει στα υπολογιστικά συστήματα να μαθαίνουν αυτόματα, χωρίς την ανθρώπινη βοήθεια ή εμπλοκή αλλά και να τα καταστήσει ικανά να προσαρμοστούν στις ανάγκες της κάθε περίπτωσης (Leanix, 2022). Όπως μπορούμε να δούμε και στην **Εικόνα 1.1** η μηχανική μάθηση μπορεί να χωριστεί σε 3 κατηγορίες:

1. Επιβλεπόμενη μάθηση (Supervised Learning)
2. Μη Επιβλεπόμενη μάθηση (Unsupervised Learning)
3. Ενισχυτική μάθηση (Reinforcement Learning)

Όπως θα δούμε παρακάτω, κάθε τύπος μάθησης λειτουργεί με διαφορετικό τρόπο και το τι θα χρησιμοποιήσουμε εξαρτάται άμεσα από τα δεδομένα που έχουμε στην διάθεσή μας. Στην **Εικόνα 1.2** μπορούμε να δούμε διαγραμματικά την εξάρτηση αυτή και το πως λειτουργεί για κάθε τύπο.



Εικόνα 1.1: Διαγραμματική ανάλυση της μηχανικής μάθησης. (Hoeppel et al., 2016)

1.2.1. Επιβλεπόμενη μάθηση

Η ικανότητα ενός μοντέλου μηχανικής μάθησης, να μαθαίνει την συνάρτηση που μεταφράζει μία είσοδο σε αποτέλεσμα βασισμένη σε παραδείγματα ζευγαριών εισόδου-εξόδου, ονομάζεται επιβλεπόμενη μάθηση (Supervised Learning, S.L.). Ένα τέτοιο μοντέλο εξάγει την συνάρτηση από ένα σετ δεδομένων με ετικέτα, το οποίο αποτελείται από ένα σύνολο εκπαιδευτικών δειγμάτων. Κάθε δείγμα στην επιβλεπόμενη μάθηση αποτελείται από ένα αντικείμενο εισόδου και μία επιθυμητή τιμή εξόδου. Ένας τέτοιος αλγόριθμος εξετάζει τα εκπαιδευτικά δεδομένα και συμπεραίνει μία συνάρτηση που μπορεί να χρησιμοποιηθεί ώστε να αντιστοιχίσει νέα δείγματα. Στην ιδανική περίπτωση, ο αλγόριθμος θα μπορεί να προβλέψει με ακρίβεια την ετικέτα από άγνωστα, μέχρι τότε, δείγματα (Contributors to Wikimedia projects, 2002).

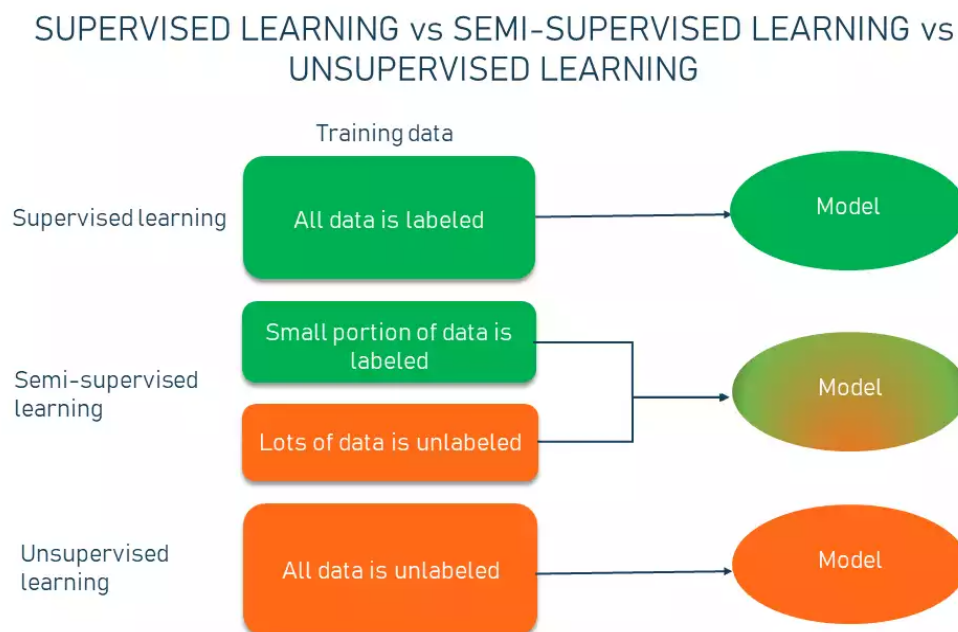
1.2.2. Μη Επιβλεπόμενη μάθηση Unsupervised Learning

Η μη επιβλεπόμενη μάθηση (Unsupervised Learning) είναι ένα είδος αλγορίθμου που χρησιμοποιεί μη κατηγοριοποιημένα δεδομένα για να ανακαλύψει μοτίβα. Ο σκοπός είναι να ωθήσουμε την μηχανή να αναπτύξει μία εσωτερική εικόνα του περιβάλλοντός

της μέσω μίμησης, που είναι ένας πολύ σημαντικός τρόπος εκμάθησης για τους ανθρώπους, και έπειτα να παράξει υλικό από αυτή. Διαφοροποιείται από την επιβλεπόμενη μάθηση ως προς το γεγονός ότι τα δεδομένα δεν έχουν λάβει κάποια ετικέτα από έναν ειδικό αλλά οργανώνονται από την μηχανή η οποία αναγνωρίζει μοτίβα με την μορφή συναρτήσεων πυκνότητας πιθανότητας (Contributors to Wikimedia projects, 2003).

1.2.3. Μάθηση Μερικής Επίβλεψης

Η μάθηση μερικής επίβλεψης (Semi-supervised learning) είναι μία ευρεία κατηγορία τεχνικών Μηχανικής μάθησης η οποία αξιοποιεί ταυτόχρονα δεδομένα με και χωρίς ετικέτα. Κατά αυτή την έννοια αποτελεί μία υβριδική τεχνική ανάμεσα στην επιβλεπόμενη και μη μάθηση. Η ιδέα στον πυρήνα της μάθησης μερικής επίβλεψης είναι να αξιοποιεί διαφορετικά τα δεδομένα ανάλογα με το αν έχουν ή όχι ετικέτα. Για τα δεδομένα με ετικέτα, ο αλγόριθμος χρησιμοποιεί την επιβλεπόμενη μάθηση ώστε να ενημερώσει τα βάρη του μοντέλου, ενώ, στην αντίθετη περίπτωση, ο αλγόριθμος ελαχιστοποιεί την διαφορά των προβλέψεων ανάμεσα σε όμοια δεδομένα (Bewtra, 2022).



Εικόνα 1.2: Διαγραμματική απεικόνιση των τύπων μάθησης. (AltexSoft, 2022)

Κεφάλαιο 2

Μηχανική μάθηση και καιρός

2.1. Καιρός

Ο καιρός αναφέρεται στις βραχυπρόθεσμες συνθήκες της κατώτερης ατμόσφαιρας, όπως είναι η θερμοκρασία, η διεύθυνση και η ταχύτητα του ανέμου, ο υετός και η ατμοσφαιρική πίεση. Ο ήλιος καθορίζει τις διάφορες αλλαγές στον τύπο του καιρού ζεσταίνοντας τον αέρα στα κατώτερα ατμοσφαιρικά στρώματα σε διαφορετικό βαθμό. Οι θερμές αέριες μάζες κινούνται ανοδικά ενώ οι ψυχρές καταρρέουν για να γεμίσουν το κενό, προκαλώντας έτσι τις κινήσεις του ανέμου. Αυτοί οι άνεμοι, μαζί με τους υδρατμούς του αέρα επηρεάζουν την δημιουργία και την κίνηση των νεφών τα οποία με την σειρά τους και σε συσχέτιση με τους προαναφερθέντες παράγοντες προκαλούν υετό (βροχή, χιόνι, χαλάζι) και καταιγίδες. Οι ατμοσφαιρικές συνθήκες που επηρεάζουν τα καιρικά φαινόμενα ταλαντεύονται αέναα προκαλώντας έτσι τις συνεχείς αλλαγές του καιρού (Wallace et al., 2006).

Μπορεί να ορίζουμε τον καιρό ως κάτι το οποίο συμβαίνει σε μία δεδομένη περιοχή όμως τα φαινόμενα με τα οποία εκφράζεται μπορεί να έχουν πολύ διαφορετική τοπικότητα. Η θερμοκρασία επηρεάζεται πολύ από την υψομετρική διαφορά, ενώ ο άνεμος μπορεί να διατηρήσει την διεύθυνση και την ταχύτητά του για πολύ μεγάλες αποστάσεις αρκεί να μην υπάρχει κάποιο εμπόδιο καθ' ύψος. Κατά συνέπεια, το πιο τοπικό καιρικό φαινόμενο με το οποίο ερχόμαστε αντιμέτωποι είναι ο υετός, υπό την έννοια ό,τι μπορεί να επηρεαστεί τόσο από τα εμπόδια καθ' ύψος, στα οποία ο άνεμος μπορεί να αυξήσει τα επίπεδα της υγρασίας άρα να έχουμε κάποιο φαινόμενο, όσο και από δύο συγκρουόμενες αέριες μάζες με μεγάλη θερμοκρασιακή διαφορά, όπου μπορεί να δημιουργηθεί μία καταιγίδα (δημιουργία ψυχρού μετώπου).

2.2. Καιρός και Κλίμα

Ενώσω ο καιρός αναφέρεται σε βραχυπρόθεσμες αλλαγές της ατμόσφαιρας, το κλίμα αναφέρεται σε ατμοσφαιρικές αλλαγές που αφορούν μεγαλύτερες χρονικές περιόδους που ορίζονται συνήθως σε 30 ή περισσότερα έτη. Αυτό εξηγεί το πως είναι δυνατόν να έχουμε ιδιαίτερα κρύες περιόδους, ενώ κατά μέσο όρο η θερμοκρασία του πλανήτη αυξάνεται. Το πρώτο αναφέρεται σε ένα καιρικό γεγονός το οποίο αφορά μερικές ημέρες, ενώ στην δεύτερη περίπτωση μιλάμε για μία ευρύτερη αλλαγή στο κλίμα η οποία λαμβάνει χώρα μέσα σε δεκαετίες. Θα μπορούσαμε να πούμε πως ένας κρύος χειμώνας είναι μία συγκριτικά μικρή ατμοσφαιρική διακύμανση μέσα σε μία, κατά τα άλλα, μακροχρόνια τάση προς υψηλότερες θερμοκρασίες. Παρά τις διαφορές τους, ο καιρός και το κλίμα είναι δύο αλληλένδετες έννοιες. Όπως και ο καιρός, έτσι και το κλίμα, λαμβάνει υπόψη τις μετεωρολογικές παραμέτρους, όπως η θερμοκρασία, ο υετός και ο άνεμος. Θα μπορούσαμε να πούμε πως το κλίμα αφορά τις μεσοποιημένες τιμές των μετεωρολογικών παραμέτρων ανά τα έτη. Κατά συνέπεια, μία αλλαγή του κλίματος μίας περιοχής είναι σίγουρο πως θα προκαλέσει αλλαγές στο μοτίβο του καιρού. Οι κλιματικές συνθήκες διαφοροποιούνται σε μεγάλο βαθμό ανάλογα με την γεωγραφική περιοχή και έχουν άμεση επίδραση πάνω στην χλωρίδα και στην πανίδα ενός τόπου. Αυτό γίνεται φανερό αν συγκρίνουμε δύο ακραία διαφορετικές περιοχές του πλανήτη, όπως αυτή της Ανταρκτικής, που έχει πολικό κλίμα, και μίας περιοχής του Αμαζονίου, που έχει τροπικό. Στην πρώτη περίπτωση έχουμε θερμοκρασίες υπό του μηδενός, ισχυρούς ανέμους και εξαιρετικά ξηρό κλίμα. Σε αυτή την περίπτωση η χλωρίδα παρουσιάζει κοινά χαρακτηριστικά, εξειδικευμένα στην επιβίωση ενός τόσο αφιλόξενου περιβάλλοντος. Στον αντίποδα, τα τροπικά δάση βιώνουν ήπια θερμές θερμοκρασίες με υψηλή σχετική υγρασία, καθώς και καθορισμένες εποχικές αλλαγές. Αυτές είναι συνθήκες που έχουν χαρίσει στις περιοχές αυτές, μία τόσο μεγάλη βιοποικιλότητα (Wallace et al., 2006).

2.3. Πρόγνωση

Όπως είδαμε λοιπόν, η πρόληψη κατά των έντονων και μη καιρικών φαινομένων είναι κάτι το οποίο αφορά το σύνολο του πληθυσμού καθώς μπορεί να μας προστατέψει από

διάφορες δυσάρεστες καταστάσεις. Σε αυτό ακριβώς το πρόβλημα έρχεται να δώσει απάντηση η πρόβλεψη του καιρού. Οι μετεωρολόγοι αναλύουν δεδομένα από πολλαπλές πηγές, όπως δορυφόροι, μετεωρολογικοί σταθμοί τόσο επίγειοι όσο και επιφάνειας θάλασσας, δεδομένα από πλοία και αεροσκάφοι κλπ. ώστε να είναι σε θέση να προβλέψουν τον καιρό των επερχόμενων ημερών ή την τάση καιρού των επόμενων εβδομάδων. Αυτές οι προγνώσεις είναι απαραίτητες για την ανθρώπινη καθημερινότητα και επηρεάζουν ολοκληρω το φάσμα της. Οι πιλότοι και οι ναυτικοί πρέπει να γνωρίζουν το πότε να περιμένουν μία μεγάλη καταιγίδα και οι αγρότες πρέπει να προγραμματίσουν τη φύτευση και την συγκομιδή. Αντίστοιχα, οι το πυροσβεστικό σώμα πρέπει να γνωρίζει εάν οι συνθήκες των επόμενων ημερών ευνοούν την ανάπτυξη μίας δασικής πυρκαγιάς ώστε να είναι σε ετοιμότητα. Το εμπόριο και οι μεταφορές επίσης εξαρτώνται άμεσα από τον καιρό. Ακόμη και ο γενικός πληθυσμός καλό είναι να είναι ενήμερος ώστε να αποφεύγονται καταστροφές από έντονα καιρικά φαινόμενα (Wallace et al., 2006).

Η ανάπτυξη των ατμοσφαιρικών επιστημών, ειδικά κατά την διάρκεια του 20ου αιώνα, βρήκε κινητήριο δύναμη στην ανάγκη για πιο ακριβή πρόβλεψη του καιρού. Στην καθομιλουμένη ο όρος “μετεωρολόγος”, ένα συνώνυμο του επιστήμονα της ατμόσφαιρας, έχει φτάσει να σημαίνει προγνώστης καιρού. Κατά τον τελευταίο αιώνα, η πρόγνωση καιρού, έχει εξελιχθεί από κάτι το οποίο βασιζόταν καθαρά στην εμπειρία και το ένστικτο, σε μία επιστήμη η οποία βασίζεται σε αριθμητικά μοντέλα που χρησιμοποιούν την διατήρηση της μάζας, της ορμής και της ενέργειας. Η αυξανόμενη πολυπλοκότητα των μοντέλων, έχει οδηγήσει σε δραματικές βελτιώσεις της ικανότητας πρόγνωσης. Η σημερινή πρόγνωση καιρού είναι σε θέση να αντιμετωπίσει, όχι μόνο την αιτιοκρατική (ντετερμινιστική), καθημερινή εξέλιξη των καιρικών μοτίβων για μία με δύο εβδομάδες, αλλά και την πιθανότητα επικίνδυνων καιρικών συμβάντων όπως μία σφοδρή καταιγίδα, σε ωριαία βάση, κάτι που αποκαλείται “nowcasting” (σε ελεύθερη μετάφραση: άμεση πρόβλεψη). Ακόμη, υπάρχει πια η δυνατότητα να υπολογίζονται παρεκκλίσεις του κλίματος από τις εποχικές προσαρμοσμένες τιμές μέχρι και για ένα έτος μετά. Η πρόβλεψη του καιρού, πέρα από το νοητικό φάσμα για την ανάπτυξη της ατμοσφαιρικής επιστήμης, έχει βοηθήσει πολύ και στην υποδομή.

Αυτό που ξεκίνησε στα τέλη του 19ου αιώνα ως ένα σύνολο περιφερειακών κέντρων συγκέντρωσης για την εκπομπή των παρατηρήσεων των επίγειων καιρικών μεταβλητών σε πραγματικό χρόνο, εξελίχθηκε σε ένα εξεζητημένο σύστημα παρατήρησης στο οποίο δορυφορικές και μετρήσεις πεδίου αλλά και της ανώτερης ατμόσφαιρας, συγχωνεύονται με ένα δυναμικό και συνεπή τρόπο ώστε να παράγουν τις βέλτιστες εκτιμήσεις του αντίστοιχου τρισδιάστατου (3D) πεδίου πάνω από ολόκληρο τον πλανήτη. Το σημερινό παγκόσμιο σύστημα παρακολούθησης του καιρού είναι μέρος ενός ευρύτερου συστήματος παρακολούθησης του πλανήτη, ζωτικής σημασίας, το οποίο υποστηρίζει μία ευρεία γκάμα επιστημονικών εγχειρημάτων, μία εκ των οποίων είναι και η κλιματολογία σε παγκόσμια κλίμακα (Wallace et al., 2006).

Ωστόσο, η ατμόσφαιρα είναι ένα ιδιαιτέρως χαοτικό σύστημα. Έτσι, η τέλεια πρόβλεψη είναι αδύνατη, κατά συνέπεια, κάθε πρόβλεψη έρχεται με ένα βαθμό αβεβαιότητας (Lorenz, 1963). Με αφορμή αυτό, κατάλληλα μοντέλα μηχανικής μάθησης θα μπορούσαν να εκπαιδευτούν ώστε να αναγνωρίζουν τις πιθανότητες να πληγεί μία περιοχή και έτσι να λαμβάνονται τα απαραίτητα μέτρα πρόληψης. Κάνοντας χρήση των δεδομένων που έχουμε συλλέξει από προηγούμενες καταστροφές στην ίδια ή σε περιοχές με παρόμοια χαρακτηριστικά οι αλγόριθμοι της μηχανικής μάθησης θα μπορούσαν να αναγνωρίσουν μοτίβα σε “τυφλά” σημεία και να φέρουν αποτελέσματα τα οποία ένα αριθμητικό μοντέλο μόνο του δεν είναι σε θέση επιτύχει.

Κεφάλαιο 3

Μηχανική μάθηση και καταστροφές

Οι φυσικοί κίνδυνοι (natural hazards) αποτελούσαν από την αρχαιότητα μία σημαντική απειλή για την επιβίωση και την ευημερία της ανθρωπότητας, ανεξάρτητα από την γεωγραφική τοποθεσία στην οποία οι άνθρωποι επέλεξαν να κατοικήσουν. Η ετοιμότητα μίας περιοχής να τους αντιμετωπίσει και η ενίσχυση που θα λάβει έπειτα από κάποιο καταστροφικό γεγονός αποτελούν εξαιρετικά σημαντικούς παράγοντες. Οι φυσικές καταστροφές (natural disasters), δεδομένου ότι απορρέουν των φυσικών κινδύνων είναι κάτι με το οποίο ερχόμαστε αντιμέτωποι ως κοινωνία, επίσης, από αρχαιοτάτων χρόνων. Η συχνότητα και η έντασή τους, συχνά οδήγησαν ακόμα και στην δημιουργία θεοτήτων οι οποίες τις κυβερνούσαν. Στον αντίποδα, εδώ και αρκετούς αιώνες οι άνθρωποι, μέσω των επιστημών, τις μελετούν και προσπαθούν να τις προβλέψουν, στην προσπάθειά τους να τις αντιμετωπίσουν. Με δεδομένο πως η κλιματική αλλαγή έχει συμβάλλει στην αύξηση της έντασης και της συχνότητάς τους, έχουμε σαν αποτέλεσμα ακόμη και στα πιο “ασφαλή” σημεία όπως οι πόλεις, η συμβολή του ανθρώπινου παράγοντα, να έχει δημιουργήσει καταστροφές “νέου τύπου”, παραδείγματος χάρη, ξαφνικές πλημμύρες (flash floods) εντός πόλεων λόγω κακής ρυμοτομίας. Γίνεται, επομένως, αντιληπτό πως περισσότερο από ποτέ, η μελέτη και η πρόγνωση είναι απαραίτητες (Cruz et al., 2006).

Για να δούμε το ζήτημα υπό τις ορθές διαστάσεις αρκεί μόνο να αναλογιστούμε τα στοιχεία που έφερε στο φως το πλαίσιο Sendai για την μείωση του ρίσκου των καταστροφών (Ghosh & Sar, 2021). Κατά την περίοδο ισχύος του πλαισίου Hyogo (Basabe, 2013) η ασφάλεια και ευημερία τόσο ανθρώπων όσο και ολόκληρων χωρών διαταράχθηκε σφοδρά. Πάνω από 700 χιλιάδες άνθρωποι έχασαν την ζωή του και περισσότεροι από 1,4 εκατομμύρια τραυματίστηκαν. Επίσης, περίπου 23 εκατομμύρια άνθρωποι βρέθηκαν στο δρόμο. Σε τελική απογραφή φαίνεται πως οι ζωές περισσότερων από 1,5 δισεκατομμυρίου ανθρώπων επηρεάστηκαν από καταστροφικά

γεγονότα, με ευαίσθητες κοινωνικές ομάδες να έχουν βρεθεί δυσανάλογα στο επίκεντρο των καταστροφών. Ο τελικός οικονομικός απολογισμός ξεπέρασε τα 1,3 τρισεκατομμύρια δολάρια.

Η ανάλυση δεδομένων των καταστροφών μπορεί να βοηθήσει στην χαρτογράφηση των χαρακτηριστικών που τις διέπουν, από το επίπεδο πρόγνωσης και έγκαιρης ενημέρωσης έως την ταχύτερη αποκατάσταση των πληγέντων περιοχών (Antonio Javier et al., 2022). Η διαχείριση ρίσκου φυσικών καταστροφών μαρτυρά την ύπαρξη στόχων με επίκεντρο την μείωση των επιδράσεων τους (Department of Regional Development and Environment Executive Secretariat for Economic and Social Affairs, 1991). Οι στόχοι αυτοί ασχολούνται με προβλήματα τα οποία χρήζουν μηχανικής βοήθειας μίας και ο καθαρός όγκος των δεδομένων και οι πολύπλοκες μεταξύ τους σχέσεις είναι παράγοντες που οι παραδοσιακές μέθοδοι αδυνατούν να αποσαφηνίσουν (Burkon, 2019). Μάλιστα, στο σύγγραμμά τους οι Aginta R. και Emanuel A. (2019) καταλήγουν, πως με βάση την πρότερη βιβλιογραφία, υπάρχουν 6 χαρακτηριστικές περιοχές, στις οποίες η μηχανική μάθηση μπορεί να βοηθήσει με την διαχείριση καταστροφών. Οι περιοχές αυτές είναι:

- έγκαιρη προειδοποίηση καταστροφών
- αξιολόγηση της καταστροφής
- ανίχνευση και παρακολούθηση
- πρόγνωση καταστροφής και αναγκών συντονισμού έπειτα από την καταστροφή
- ανταπόκριση
- μακροχρόνια ανάλυση και μείωση κινδύνου

Λαμβάνοντας υπόψη τα παραπάνω, είναι λογικό η χρήση της τεχνητής νοημοσύνης και κυρίως της μηχανικής μάθησης να χαίρει μίας ολοένα αυξανόμενης απάντησης στον χώρο των γεωφυσικών επιστημών, π.χ. Badr et al. (2014); Rivero-Calle (2015); Tripathi et al. (2006).

3.1. Έγκαιρη προειδοποίηση καταστροφών

Τα συστήματα έγκαιρης προειδοποίησης αποτελούν βασικά και αναπόσπαστα κομμάτια της μείωσης των καταστροφών από φυσικούς κινδύνους. Με μία σωστή διαχείριση μειώνουμε αποδοτικά τις υλικές και οικονομικές απώλειες και φυσικά ελαχιστοποιούμε τους θανάτους.

Η αύξηση της ατμοσφαιρικής θερμοκρασίας και της επιφάνειας της θάλασσας επηρεάζει τα μοτίβα βροχοπτώσεων. Ο θερμότερος αέρας μπορεί να απορροφήσει περισσότερη υγρασία, γεγονός που αυξάνει την πιθανότητα έντονων βροχοπτώσεων. Εκτεταμένες επιστημονικές μελέτες που αναλύθηκαν από τη Διακυβερνητική Ομάδα για την Κλιματική Αλλαγή (IPCC) υποδηλώνουν ότι οι ακραίες εκρήξεις και οι πλημμύρες ποταμών μπορεί να αυξηθούν σε πολλές περιοχές. Οι επιστήμονες πιστεύουν ότι η κλιματική αλλαγή είναι επίσης τουλάχιστον εν μέρει υπεύθυνη για την κλίμακα των τελευταίων καταστροφικών πλημμυρών στην Ευρώπη, μετά τις ακραίες βροχοπτώσεις που έπληξαν κυρίως τη Γερμανία τον Ιούλιο του 2021 (MunichRe, 2022).

Κεφάλαιο 5

Υλικό Μελέτης

5.1. Δεδομένα και πηγές

Για να προχωρήσουμε στην μοντελοποίηση των καταστροφών λόγω των βροχοπτώσεων όπως είναι προφανές θα πρέπει να έχουμε δεδομένα τόσο για την βροχή όσο και για τις ζημιές. Όπως αναφέραμε και νωρίτερα είναι ιδιαίτερα δύσκολο να βρούμε ποσοτικά δεδομένα καταστροφών οπότε θα ήταν πιο απλό να χρησιμοποιήσουμε ποιοτικά, εάν είχαμε ή δεν είχαμε καταστροφή, στην δική μας περίπτωση. Σε ό,τι αφορά τις βροχοπτώσεις, θα πρέπει να έχουμε έγκυρα δεδομένα υψηλής χωρικής ανάλυσης, μίας και η βροχή είναι πολύ τοπικό φαινόμενο όπως είδαμε νωρίτερα.

Το Εθνικό Αστεροσκοπείο Αθηνών (ΕΑΑ; National Observatory of Athens, NOA) πήρε, το 2006, την πρωτοβουλία να ξεκινήσει ένα δίκτυο αυτόματων μετεωρολογικών σταθμών (ΑΜΣ) (Papagiannaki et al., 2013). Στους λόγους για την έναρξη ενός τέτοιου δικτύου εμπίπτουν, η μελέτη του μικροκλίματος κάθε περιοχής, όπως επίσης και η επικύρωση μοντέλων πρόγνωσης, διότι και στις δύο περιπτώσεις είναι απαραίτητο να υπάρχουν ιστορικά δεδομένα των μετεωρολογικών συνθηκών. Ένας άλλος λόγος που συμβάλλει σε μία τέτοια ενέργεια είναι η ενημέρωση του πληθυσμού, αλλά και των αρμόδιων υπηρεσιών για τις ακριβείς συνθήκες που επικρατούν σε κάθε τόπο, μία πληροφορία που σε έκτακτες περιπτώσεις θα μπορούσε να σώσει ζωές. Στηριζόμενοι στο τελευταίο, στην παρούσα εργασία χρησιμοποιήθηκαν δεδομένα από τους ΑΜΣ ώστε να έχουμε ποσοτικά δεδομένα που αφορούν το ύψος της βροχής. Στην συνέχεια από τα δεδομένα αυτά παράχθηκαν τα συμβάντα βροχής (rainfall events) τα οποία συσχετίστηκαν με την ύπαρξη ή μη καταστροφών σε κάθε Ταχυδρομικό Κώδικα (Τ. Κ.; Postal Code, PC) (Papagiannaki et al., 2022).



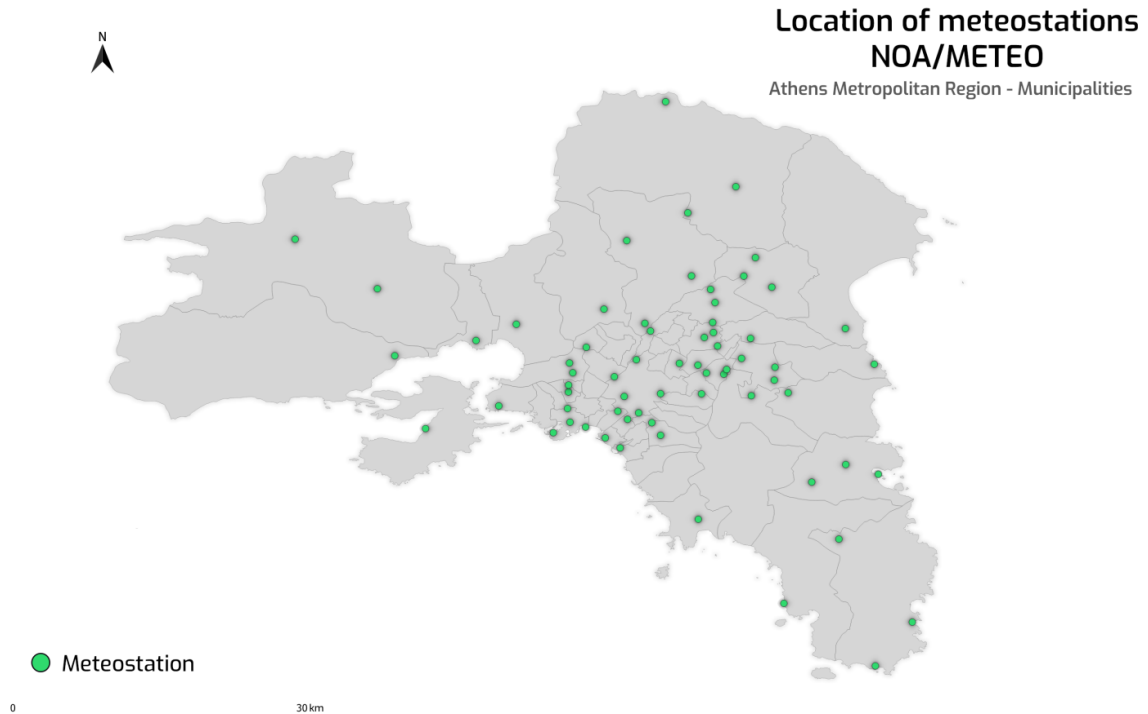
Εικόνα 5.1: *Αυτόματος μετεωρολογικός σταθμός στο Ε.Α.Α. στην Π. Πεντέλη.*

5.2. Περιοχή μελέτης

Η Μητροπολιτική Περιοχή των Αθηνών (ΜΠΑ; Athens Metropolitan Area, AMA) του νομού της Αττικής είναι η πιο πυκνοκατοικημένη περιοχή της χώρας, με πληθυσμό που ανέρχεται στους περίπου 3,8 Εκ. κατοίκους, δεδομένου πως περιέχει την πόλη των Αθηνών που είναι η πρωτεύουσα. Το κλίμα είναι εύκρατο, τυπικό Μεσογειακό, και η μέση αθροιστική βροχόπτωση ανέρχεται περίπου στα 450mm, με τις μέγιστες να έχουν καταγραφεί από τα τέλη του Φθινοπώρου μέχρι και τις αρχές του χειμώνα (Paragiannaki et al., 2015; Lagouvardos et al., 1996). Επίσης, πρόκειται για την πιο πληγείσα περιοχή σε ό,τι αφορά κοινωνικές επιπτώσεις από καιρικά φαινόμενα. Κατά την βάση δεδομένων καιρικών φαινομένων υψηλών επιπτώσεων (High-Impact Weather Events, HIWE) που έχει αναπτυχθεί από την ομάδα ΜΕΤΕΟ του Εθνικού Αστεροσκοπείου Αθηνών (Paragiannaki et al., 2013), η εν λόγω περιοχή υποφέρει ιδιαίτερος από ξαφνικές πλημμύρες λόγω βροχοπτώσεων.

Η AMA καλύπτει περίπου 3200 km² και συμπεριλαμβάνει 284 υποπεριοχές με βάση τον διαχωρισμό σε Ταχυδρομικούς Κώδικες. Οι Τ.Κ. καλύπτουν κατά μέσο όρο 11km² (με εύρος από 0.05 μέχρι 348 km²) και μέσο πληθυσμό 12100 κατοίκους (με εύρος από 990 μέχρι 59000). Τα πληθυσμιακά δεδομένα προέρχονται από την Ελληνική Στατιστική Αρχή (ΕΛΣΤΑΤ) και την απογραφή του 2011 (HSA et al., 2021).

5.3. Υπόβαθρο δεδομένων



Εικόνα 5.2: ΑΜΣ στην μητροπολιτική περιοχή των Αθηνών. (Paragiannaki et al., 2022)

Το σετ δεδομένων που χρησιμοποιήθηκε στην εργασία αφορά ζεύγη δεδομένων ύψους βροχής και ύπαρξη ή μη καταστροφής στην ΑΜΑ. Πρόκειται για το ίδιο σετ δεδομένων που παράχθηκε και χρησιμοποιήθηκε στην εργασία των Paragiannaki et al. το 2022.

Τα δεδομένα βροχοπτώσεων για κάθε επεισόδιο και ΤΚ προέκυψαν από το επιφανειακό δίκτυο των 450 μετεωρολογικών σταθμών που λειτουργεί η Μονάδα ΜΕΤΕΟ του Ινστιτούτου Περιβαλλοντικών Ερευνών και Βιώσιμης Ανάπτυξης του Εθνικού Αστεροσκοπείου Αθηνών, ΙΕΠΒΑ/ΕΑΑ, (Lagouvardos et al., 2017). Πραγματοποιήθηκε ταυτοποίηση των μετεωρολογικών επεισοδίων που εκδηλώθηκαν στη χώρα την περίοδο 2012-2019 και των ΤΚ που το κάθε επεισόδιο έπληξε, βάσει των ασφαλιστικών ζημιών, και υπολογίστηκαν οι σχετικές μετεωρολογικές παράμετροι από αντιπροσωπευτικούς μετεωρολογικούς σταθμούς κάθε ΤΚ.

Τα δεδομένα καταστροφών προήλθαν από αιτήσεις αποζημίωσης από μία από τις μεγαλύτερες Ελληνικές ασφαλιστικές εταιρείες με αναφορά στον Τ.Κ. Η εν λόγω εταιρεία εκπροσωπεί περίπου το 11% της εγχώριας αγοράς που αφορούν ασφαλίσεις ζημιών και κατατάσσεται πρώτη με το μερίδιό της (Paragiannaki et al., 2022).

5.4. Ανάλυση συμβάντων βροχής

Πιο αναλυτικά η μεθοδολογία που ακολουθήθηκε κατά την συγγραφή της εργασίας των Paragiannaki et al. το 2022 έχει ως εξής,

- Τα επεισόδια βροχόπτωσης καταγράφηκαν από κάθε σταθμό από την έναρξη της λειτουργίας του, σύμφωνα με διαπιστευμένη μεθοδολογία που αναπτύχθηκε από τη μονάδα ΜΕΤΕΟ του ΕΑΑ (Paragiannaki et al., 2015). Τα γεγονότα βροχής έχουν διαφορά ώρας έναρξης τουλάχιστον 24 ωρών και ως έναρξη καταγραφής ορίζουμε το πρώτο δεκάλεπτο με καταγραφή βροχόπτωσης $>0.2\text{mm}$. Μόνο τα επεισόδια με αθροιστική συνολική βροχόπτωση άνω των 20 mm και αθροιστική βροχόπτωση 60 λεπτών άνω των 5 mm επιλέχθηκαν για την ανάλυση, ώστε να ληφθούν υπόψη αντικειμενικά δυνητικά επικίνδυνες συνθήκες. Αυτά τα κριτήρια βασίστηκαν σε προηγούμενες μελέτες σχετικά με την εμφάνιση ξαφνικών πλημμυρών στην περιοχή μελέτης (Paragiannaki et al., 2015).
- Υπολογισμός της απόστασης κάθε μετεωρολογικού σταθμού από τα γεωχωρικά κέντρα (centroids) των ΤΚ.
- Επιλογή αντιπροσωπευτικού σταθμού για κάθε ΤΚ και επεισόδιο από το σύνολο των σταθμών που υπάρχουν σε απόσταση έως 5 km από το κέντρο του ΤΚ. Στις λίγες περιπτώσεις που σημειώθηκε υπέρβαση αυτού του ορίου, επιλέχθηκε ο πλησιέστερος σταθμός στο κέντρο σε απόσταση έως και 20 km.
- Αντιστοίχιση των επεισοδίων με τις ασφαλισμένες ζημιές σε επίπεδο ΤΚ, λαμβάνοντας υπόψη μία χρονική απόκλιση ± 24 ωρών για την κάλυψη όλων

των περιπτώσεων με καθυστέρηση υποβολής αιτήματος για ασφαλιστική αποζημίωση καθώς και τη χνότητα αντιστοίχισης του κάθε σταθμού που πληροί τα παραπάνω κριτήρια με αναγγελίες ζημιάς.

- Βελτιστοποίηση της διαδικασίας ώστε να ταυτοποιηθούν οι κοντινότεροι σταθμοί χρησιμοποιώντας αλγόριθμο εύρεσης της διεύθυνσης των φαινομένων σε σχέση με το γεωχωρικό κέντρο (centroid) του Τ.Κ.
- Αξιολόγηση των αποτελεσμάτων ως προς την πιο αντιπροσωπευτική αντιστοίχιση επεισοδίων-ζημιών, με βάση στατιστική ανάλυση των παραγόμενων επεισοδίων σε επίπεδο Τ.Κ.

5.5. Περιγραφή των δεδομένων και στατιστικά στοιχεία

Κατα την περίοδο 2012-2019 είχαμε 228 συμβάντα βροχής στην μητροπολιτική περιοχή των Αθηνών. Από αυτά τα 228 συμβάντα, τα 115 ήταν αυτά που προκάλεσαν πλημμυρικά φαινόμενα σε διάφορα μέρη της περιοχής. Κάθε συμβάν επηρέασε από 1 έως και 243 από τους Τ.Κ., οι οποίοι είναι 284 στο σύνολό τους. Αυτά μας δίνουν ένα Μ.Ο. στα 38,3 με μία τυπική απόκλιση της τάξεως του 56,3. Κατά συνέπεια διαθέτουμε 8726 ζευγάρια Τ.Κ. - ύψους βροχής ώστε να προχωρήσουμε την ανάλυσή μας. Η μέγιστη συνολική εικοσιτετράωρη βροχόπτωση (R24) κυμαίνεται σε ύψος από τα 20 μέχρι τα 179,6 mm, με μία μέση τιμή στα 38,2 mm.

Στον **Πίνακα 5.1** μπορούμε να δούμε κάποια βασικά στατιστικά στοιχεία για όλες τις μεταβλητές.

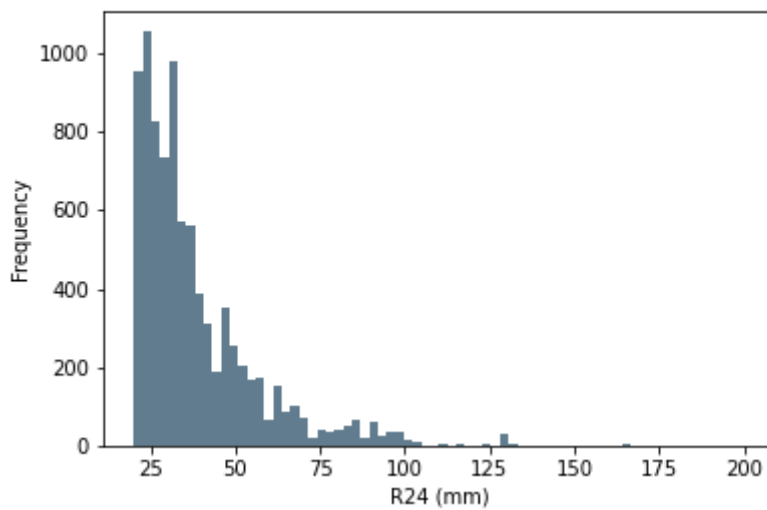
index	dist	dur	R_tot	R24	R24_day	R30	pop	pop_dens
count	8726.000	8726.000	8726.000	8726.000	8726.000	8726.000	8726.000	8726.000
mean	1978.630	150736.214	47.104	38.198	27.348	13.238	12144.333	12.718
std	1903.663	120445.799	27.285	18.338	12.750	7.368	8849.715	8.177
min	50.000	1800.000	20.000	20.000	4.816	5.000	990.155	0.027
25%	1038.000	55800.000	28.000	25.400	19.310	7.400	5211.735	4.961
50%	1561.000	120600.000	37.600	32.400	25.200	10.800	10806.373	13.212
75%	2129.000	227400.000	55.600	45.200	32.600	17.600	17368.602	21.909
max	16052.000	968400.000	297.000	179.600	109.405	57.000	58978.982	27.820

Πίνακας 5.1: Βασικά στατιστικά στοιχεία μεταβλητών.

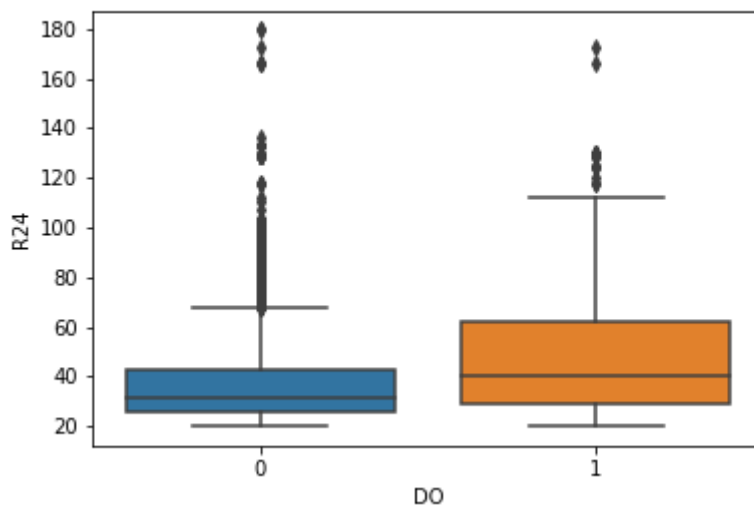
Όπως είναι εμφανές στον **Πίνακα 5.2**, πετυχαίνουμε την βέλτιστη στατιστική συσχέτιση για ζεύγος τιμών R24-D.O. Επίσης το κριτήριο Pearson p -value < 0,00.

index	DO	dist	dur	R_tot	R24	R24_day	R30	pop	pop_dens
DO	1.000	0.040	0.085	0.204	0.206	0.122	0.166	0.158	-0.121
dist	0.040	1.000	0.077	0.063	0.051	0.010	-0.018	0.201	-0.444
dur	0.085	0.077	1.000	0.600	0.319	-0.378	-0.129	0.052	-0.127
R_tot	0.204	0.063	0.600	1.000	0.893	0.377	0.256	0.034	-0.097
R24	0.206	0.051	0.319	0.893	1.000	0.630	0.361	0.029	-0.076
R24_day	0.122	0.010	-0.378	0.377	0.630	1.000	0.412	-0.025	0.009
R30	0.166	-0.018	-0.129	0.256	0.361	0.412	1.000	-0.016	-0.014
pop	0.158	0.201	0.052	0.034	0.029	-0.025	-0.016	1.000	-0.306
pop_dens	-0.121	-0.444	-0.127	-0.097	-0.076	0.009	-0.014	-0.306	1.000

Πίνακας 5.2: Πίνακας στατιστικής συσχέτισης μεταβλητών.



Εικόνα 5.3: Κατανομή συχνότητας (histogram) ύψους R24 στα συμβάντα βροχής.



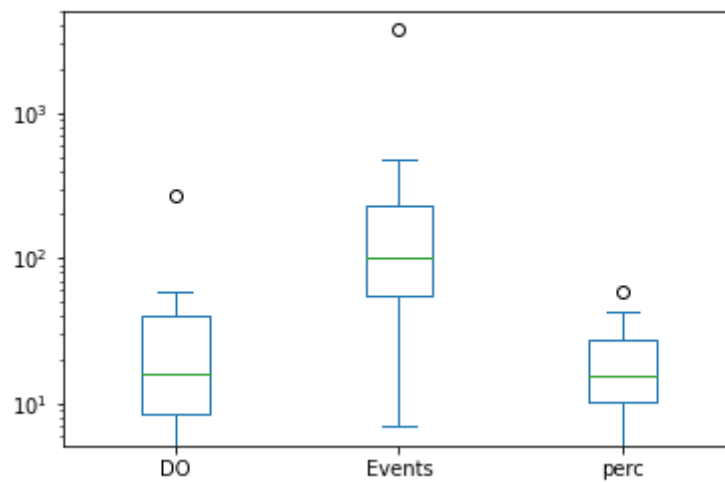
Εικόνα 5.4: Στατιστική κατανομή (boxplot) του R24 σε σχέση με την δυαδική μεταβλητή ύπαρξης ζημιάς, DO.

Municipality	DO	Events	Do/Event Percentage
AGIAS PARASKEYIS	13	86	15
AGIAS VARVARAS	0	7	0
AGION ANARGYRON - KAMATEROY	15	79	19
AGIOY DHMHTRIOY	8	64	12

ΑΗΑΡΝΟΝ	34	152	22
ΑΙΓΑΛΕΟ	9	83	11
ΑΛΙΜΟΥ	8	32	25
ΑΜΑΡΟΥΣΙΟΥ	31	138	22
ΑΣΠΡΟΠΥΡΓΟΥ	6	14	43
ΑΘΗΝΑΙΟΝ	266	3695	7
ΔΑΦΝΗΣ - ΥΜΗΤ ΤΟΥ	6	98	6
ΔΙΟΝΥΣΟΣ	32	278	12
ΕΛΕΦΣΙΝΑΣ	5	16	31
ΕΛΛΗΝΙΚΟΥ - ΑΡΓΥΡΟΠΟΛΗΣ	7	46	15
ΦΙΛΑΔΕΛΦΕΙΑΣ - ΗΛΚΗΔΟΝΟΣ	12	79	15
ΦΙΛΟΘΕΙ - ΨΥΧΙΚΟ	14	123	11
ΦΥΛΙΣ	44	236	19
ΓΑΛΑΤΣΙΟΥ	8	74	11
ΓΛΥΦΑΔΑΣ	28	94	30
ΗΛΑΝΔΡΙΟΥ	40	146	27
ΗΑΪΔΑΡΙΟΥ	6	24	25
ΗΛΙΟΥΠΟΛΗΣ	8	84	10
ΗΡΑΚΛΕΙΟΥ	15	59	25
ΙΛΙΟΥ	20	65	31
ΚΑΙΣΑΡΙΑΝΙΣ	9	73	12
ΚΑΛΛΙΘΕΑΣ	20	129	16
ΚΕΡΑΤΣΙΝΙΟΥ - ΔΡΑΠΕΤΣΟΝΑΣ	10	87	11
ΚΗΦΙΣΙΑΣ	43	209	21
ΚΟΡΥΔΑΛΛΟΥ	8	65	12
ΚΡΟΠΙΑΣ	4	27	15
ΛΑΥΡΕΟΤΙΚΙΣ	5	41	12
ΛΥΚΟΒΡΥΣΗΣ - ΡΕΥΚΗΣ	12	113	11
ΜΑΝΔΡΑΣ - ΕΙΔΥΛΛΙΑΣ	7	76	9
ΜΑΡΑΘΙΝΟΣ	11	86	13
ΜΑΡΚΟΠΟΥΛΟΥ ΜΕΣΟΓΑΙΑΣ	3	25	12
ΜΕΓΑΡΕΟΝ	15	69	22
ΜΕΤΑΜΟΡΦΟΣΕΟΣ	4	26	15
ΜΟΣΧΑΤΟΥ - ΤΑΥΡΟΥ	7	86	8
ΝΕΑΣ ΙΟΝΙΑΣ	17	169	10

ΝΕΑΣ ΣΜΥΡΝΗΣ	13	136	10
ΝΙΚΑΙΑΣ - ΑΓΙΟΥ Ι. ΡΕΝΤΗ	9	138	7
ΟΡΟΡΟΥ	7	81	9
ΡΑΙΑΝΙΑΣ	12	41	29
ΠΑΛΑΙΟΥ ΦΑΛΙΡΟΥ	22	133	17
PALLINHS	9	30	30
PAPAGOY - HOLARGOY	5	44	11
ΠΕΙΡΑΙΟΣ	17	278	6
PANTELHS	4	48	8
PERAMATOS	5	20	25
PERISTERIOY	50	232	22
PETROPOLIS	6	16	38
ΡΑΦΙΝΑΣ - ΠΙΚΕΡΜΙ	7	40	18
SALAMINAS	9	39	23
SARONIKOY	4	33	12
SPATON - ARTEMIDOS	9	66	14
VARIS - VOULAS - VOULIAGMENIS	10	81	12
VRILISSION	7	23	30
VYRONOS	5	52	10
ZOGRAFOY	15	142	11

Πίνακας 5.3 : Πίνακας συμβάντων - καταστροφών ανά δήμο.



Εικόνα 5.5: Στατιστική κατανομή (boxplot) των συμβάντων βροχής, των καταστροφών και του ποσοστού του δεύτερου στο πρώτο, για κάθε δήμο.

Κεφάλαιο 6

Μέθοδοι

6.1 Μοντέλα Μηχανικής μάθησης

Στην παρούσα εργασία θα ασχοληθούμε με μοντέλα επιβλεπόμενης μηχανικής μάθησης. Όπως είδαμε παραπάνω σε αυτά ανήκουν τα μοντέλα ταξινόμησης (classification). Στην Μ.Μ. η ταξινόμηση είναι μία διαδικασία δύο βημάτων, το πρώτο αφορά την μάθηση και το δεύτερο με την πρόβλεψη. Στο πρώτο βήμα τροφοδοτούμε το μοντέλο με ένα σετ δεδομένων από το οποίο θα μάθει πως πρέπει να τα ταξινομήι έτσι ώστε στο δεύτερο να μπορέσει να κάνει προβλέψεις (Chauhan, 2022).

6.1.1 Δέντρο αποφάσεων

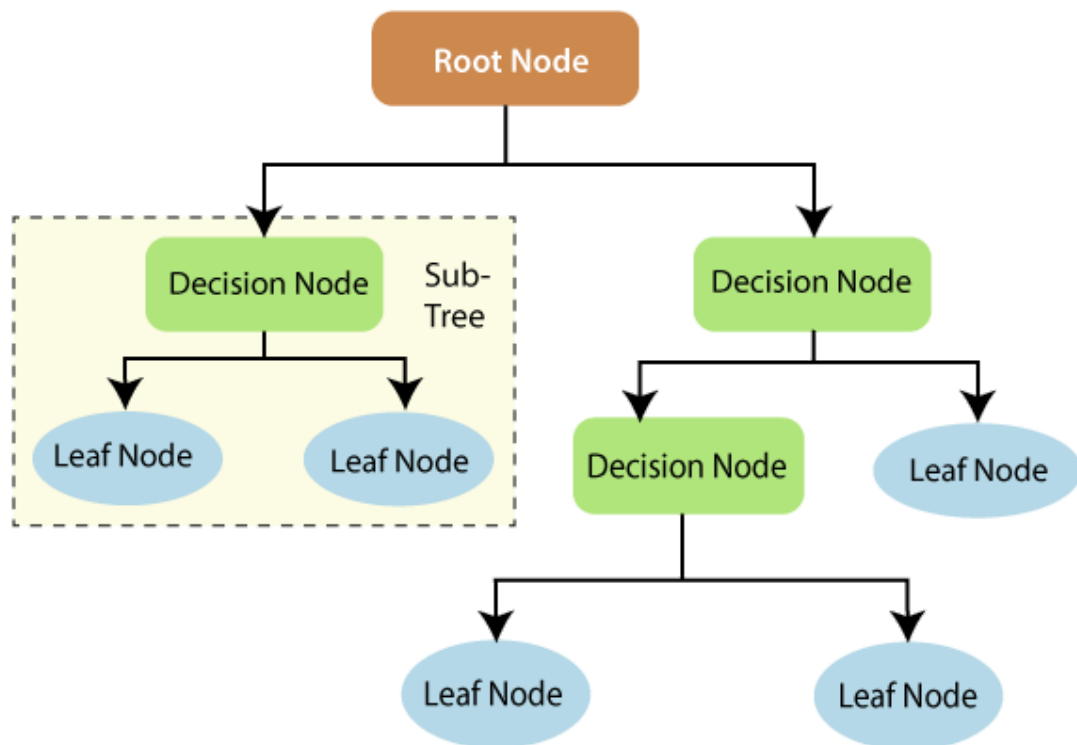
Ο αλγόριθμος δέντρου αποφάσεων (Decision Tree) ανήκει στην οικογένεια των αλγορίθμων επιβλεπόμενης μάθησης. Το δέντρο αποφάσεων μπορεί να χρησιμοποιηθεί για την επίλυση προβλημάτων τόσο ταξινόμησης όσο και παλινδρόμησης (regression). Ο στόχος αυτού του αλγορίθμου είναι να δημιουργήσει ένα εκπαιδευτικό μοντέλο (training model) το οποίο μπορεί να χρησιμοποιηθεί για να προβλέψει την κατηγορία (ή κλάση) ή την τιμή της στοχευμένης μεταβλητή, μαθαίνοντας απλούς κανόνες απόφασης, τους οποίους έχει συμπεράνει από προηγούμενα (ή εκπαιδευτικά) δεδομένα. Ένα δέντρο αποφάσεων αποτελείται από τρία διακριτά στοιχεία:

- Τον κύριο κόμβο ή κόμβο ρίζα (root node)
- Κόμβους απόφασης (decision nodes)
- Κόμβος φύλλου (leaf nodes)

Ένα δέντρο αποφάσεων χωρίζει τα εκπαιδευτικά δεδομένα σε “κλαδιά”, από τα οποία διαχωρίζει αργότερα περισσότερα “κλαδιά”. Αυτός ο αλγόριθμος για να προβλέψει την

ετικέτα της κατηγορίας για κάθε στοιχείο, ξεκινάει από την ρίζα του δέντρου. Σε αυτό το σημείο συγκρίνει την τιμή μεταξύ ρίζας και στοιχείου και επαναλαμβάνει την διαδικασία, σε κάθε κόμβο, μέχρι να φτάσει σε έναν από τους τερματικούς κόμβους, τα φύλλα.

Έχουμε δύο τύπους δέντρων αποφάσεων ανάλογα με τον τύπο της μεταβλητής που μας ενδιαφέρει, Δέντρα κατηγορικής, ή ποιοτικής, μεταβλητής και δέντρα ποσοτικής μεταβλητής (Chauhan, 2022).

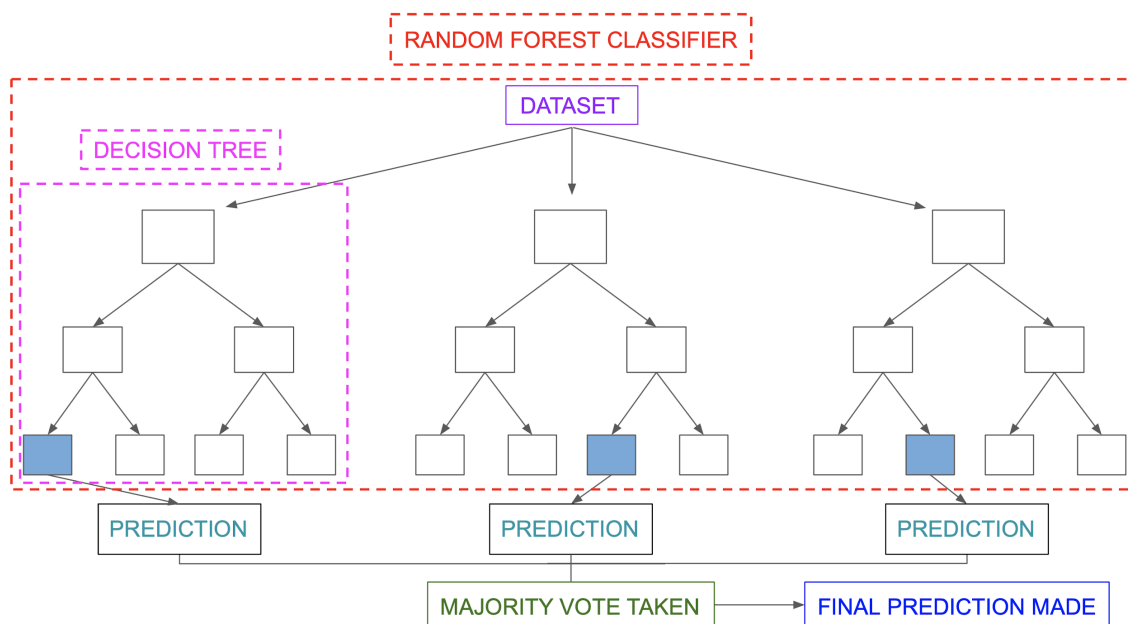


Εικόνα 6.1: Διάγραμμα δέντρου αποφάσεων. (Maudoux & Boumerdassi, 2021)

6.1.2 Τυχαίο δάσος

Το τυχαίο δάσος (random forest) είναι ένας αλγόριθμος επιβλεπόμενης μηχανικής μάθησης ο οποίος σχηματίζεται από πολλαπλούς αλγορίθμους δέντρου αποφάσεων. Ο αλγόριθμος χρησιμοποιείται σε ποικίλες βιομηχανίες όπως σε τράπεζες και

ηλεκτρονικές αγορές (e-commerce) ώστε να προβλέψουν συμπεριφορές και αποτελέσματα. Το RF, όντας ουσιαστικά μία επέκταση του DT, είναι μία τεχνική μηχανικής μάθησης που μπορεί να χρησιμοποιηθεί τόσο για επίλυση προβλημάτων ταξινόμησης όσο και παλινδρόμησης. Εφαρμόζει συνολική μάθηση (ensemble learning), η οποία είναι μία τεχνική που συνδιάζει πολλούς ταξινομητές έτσι ώστε να παρέχει λύσεις σε περίπλοκα προβλήματα. Το αποτέλεσμα καθορίζεται με βάση τις προβλέψεις των δέντρων αποφάσεων. Η πρόβλεψη γίνεται παίρνοντας τον μέσο όρο των αποτελεσμάτων των δέντρων. Κατά συνέπεια, αυξάνοντας τον αριθμό των δέντρων, αυξάνεται και η ακρίβεια του αποτελέσματος. Το αλγόριθμος RF εξουδετερώνει τους περιορισμούς του DT, καθώς μειώνει την υπερπροσαρμογή (overfitting) του μοντέλου στο σετ των δεδομένων (Mbaabu, 2020).



Εικόνα 6.2: Διάγραμμα δέντρου αποφάσεων. (Mbaabu, 2020)

6.1.3 K κοντινότεροι γείτονες

Ο αλγόριθμος K κοντινότεροι γείτονες (K Nearest Neighbors, KNN) ανήκει στους αλγορίθμους επιβλεπόμενης μάθησης της Μ.Μ. και χρησιμοποιείται για επίλυση προβλημάτων τόσο ταξινόμησης όσο και παλινδρόμησης.

Χρησιμοποιούμε τον αλγόριθμο KNN για να κάνουμε προβλέψεις πάνω σε δοκιμαστικά σετ δεδομένων βασιζόμενοι στα χαρακτηριστικά (ετικέτες) των εκπαιδευτικών δεδομένων.

Η μέθοδος που χρησιμοποιείται για να κάνουμε αυτές τις προβλέψεις λειτουργεί υπολογίζοντας την απόσταση ανάμεσα στα δοκιμαστικά και εκπαιδευτικά δεδομένα, θεωρώντας ως δεδομένο πως η απόσταση ανάμεσα σε δεδομένα με παρόμοια χαρακτηριστικά είναι μικρή. Επιτρέπει την αναγνώριση και κατηγοριοποίηση των νέων δεδομένων, ενώ ταυτόχρονα λαμβάνει υπόψη τα χαρακτηριστικά τους, με βάση τις ήδη καταχωρημένες θέσεις από τα εκπαιδευτικά δεδομένα. Αυτά τα χαρακτηριστικά των νέων θέσεων, θα καταχωρηθούν με την σειρά τους από τον αλγόριθμο και με βάση την εγγύτητα σε άλλες θέσεις, θα κατηγοριοποιηθούν.

Το “K” στο KNN είναι η παράμετρος που αναφέρεται στο πλήθος των κοντινότερων γειτόνων, είναι δηλαδή ο χώρος που δημιουργείται από τα εγγύτερα αυτά σημεία. Χρησιμοποιώντας την τιμή του K, υπολογίζουμε την απόσταση ανάμεσα στα νέα δεδομένα και τα σημεία των εκπαιδευτικών δεδομένων. Η τιμή του K είναι ένας θετικός ακέραιος αριθμός, τυπικά μικρός και συνιστάται να είναι περιττός (Nisha, 2022).



Εικόνα 6.3: Κριτήρια αξιολόγησης. (Agrawal, 2021)

6.1. Κριτήρια αξιολόγησης

Η αξιολόγηση της απόδοσης ενός μοντέλου μηχανική μάθησης είναι ένα από τα πιο σημαντικά βήματα όταν χτίζει κανείς ένα αποτελεσματικό μοντέλο. Για να αξιολογηθεί η απόδοση ή η ποιότητα του μοντέλου χρησιμοποιούνται διάφορες μέθοδοι μέτρησης οι οποίες είναι γνωστές ως μετρήσεις απόδοσης ή μετρήσεις αξιολόγησης. Αυτές οι μετρήσεις μας βοηθούν να κατανοήσουμε πόσο καλά έχει αποδώσει το μοντέλο μας στα δεδομένα που του παρέχουμε. Με αυτό τον τρόπο, μπορούμε να βελτιώσουμε την απόδοση ρυθμίζοντας τις παραμέτρους (hyper-parameters) του. Κάθε μοντέλο μηχανικής μάθησης στοχεύει στο να γενικεύσει επαρκώς νέα δεδομένα και οι μέθοδοι αξιολόγησης κρίνουν το πόσο σωστά ανταποκρίνεται κάθε μοντέλο σε αυτό (Performance Metrics in Machine Learning - Javatpoint, 2022).

6.1.1. Αληθώς Θετικά

Ως αληθώς θετικά (True Positive, TP) ορίζουμε τα θετικά παραδείγματα που ταξινομούνται σωστά από ένα μοντέλο ταξινόμησης (Sammut & Webb, 2010).

6.1.2. Αληθώς Αρνητικά

Ως αληθώς αρνητικά (True Negative, TN) ορίζουμε τα αρνητικά παραδείγματα που ταξινομούνται σωστά από ένα μοντέλο ταξινόμησης (Sammut & Webb, 2010).

6.1.3. Ψευδώς Θετικά

Σε ένα πρόβλημα δύο κλάσεων, ένα μοντέλο ταξινόμησης μπορεί να κάνει δύο τύπους λάθους: αρνητικώς θετικά και αρνητικώς αρνητικά. Ψευδώς αρνητικό είναι ένα παράδειγμα θετικής κλάσης που έχει λανθασμένα κατηγοριοποιηθεί ως αρνητικό (Sammut & Webb, 2010).

6.1.4. Ψευδώς Αρνητικά

Στο πλαίσιο μίας ιατρικής διάγνωσης, αρνητικώς αρνητικό λάθος προκύπτει όταν ένα άτομο με μία ασθένεια αναγνωρίζεται από ένα διαγνωστικό τεστ ως υγιές (Sammut & Webb, 2010).

Στην επιστήμη, αν θεωρήσουμε κάτι ως αληθές ενώ είναι ψευδές, χρησιμοποιούμε τους όρους “λάθος τύπου Ι”, “λάθος άλφα” ή “ψευδώς θετικό”. Αν κάτι θεωρήσουμε πως είναι ψευδές ενώ είναι αληθές τότε χρησιμοποιούμε τους όρους, “λάθος τύπου ΙΙ”, “λάθος βήτα” ή “ψευδώς αρνητικό” (Jekel et al., 2007).

6.1.5. Πίνακας Σύγχυσης

Ο Πίνακας Σύγχυσης (Confusion Matrix) συνοψίζει την απόδοση της ταξινόμησης με βάση τα εξεταστικά δεδομένα. Είναι ένας δισδιάστατος (2D) πίνακας όπου στη μία του διάσταση έχουμε την αληθή ταξινόμηση, ενώ στην άλλη έχουμε τα αποτελέσματα ταξινόμησης του μοντέλου. Ο Πίνακας 6.1 παρουσιάζει ένα παράδειγμα πίνακα σύγχυσης για μία περίπτωση ταξινόμησης σε τρεις τάξεις, A, B και C. (Sammut & Webb, 2010)

Confusion Matrix. Table 1 An example of three-class confusion matrix

		Assigned Class		
		A	B	C
Actual Class	A	10	2	1
	B	0	6	1
	C	0	3	8

Πίνακας 6.1: Πίνακας Σύγκρισης 3 τάξεων. (Sammut & Webb, 2010)

Η πρώτη γραμμή του πίνακα μας δείχνει πως 13 αντικείμενα ανήκουν στην κλάση A και πως 10 έχουν κατηγοριοποιηθεί ορθά πως ανήκουν στην A, ενώ 3 έχουν κατηγοριοποιηθεί λανθασμένα, 2 σαν B και 1 ως C.

Μία ειδική περίπτωση του πίνακα σύγκρισης χρησιμοποιείται συχνά όταν έχουμε μόνο δύο κλάσεις. Σε αυτή την περίπτωση, η μία κλάση ορίζεται ως θετική και η άλλη ως αρνητική. Σε αυτό το πλαίσιο, τα 4 κελιά του πίνακα ορίζουν τα αληθώς θετικά, ψευδώς θετικά, αληθώς αρνητικά και ψευδώς αρνητικά, όπως φαίνεται στον **Πίνακα 6.2**

Confusion Matrix. Table 2 The outcomes of classification into positive and negative classes

		Assigned Class	
		Positive	Negative
Actual Class	Positive	TP	FN
	Negative	FP	TN

Πίνακας 6.2: Ειδική περίπτωση Πίνακα σύγκρισης. (Sammut & Webb, 2010)

6.1.6. Εξειδίκευση

Η Εξειδίκευση (Accuracy, specificity) αποτελεί ένα μέτρο του βαθμού στον οποίο η προβλέψεις ενός μοντέλου ανταποκρίνονται στην πραγματικότητα που θέλουμε να μοντελοποιήσουμε. Ο όρος Εξειδίκευση συχνά χρησιμοποιείται στο γενικό πλαίσιο των μοντέλων ταξινόμησης. Υπό αυτό το πρίσμα έχουμε $accuracy = P(\lambda(X) = Y)$, όπου X και Y αποτελούν την από κοινού κατανομή και το μοντέλο ταξινόμησης λ είναι συνάρτηση $X \rightarrow Y$. Κάποιες φορές, μπορεί να αναφερθούμε σε αυτό το μέτρο με ποσοστό, αντίθετα από μία τιμή ανάμεσα στο 0,0 και το 1,0. Η ακρίβεια ενός μοντέλου μπορεί να υπολογιστεί αν εφαρμόσουμε το μοντέλο σε ένα σετ δοκιμαστικών δεδομένων για τα οποία γνωρίζουμε την ετικέτα που τους έχει αποδοθεί. Η ακρίβεια ενός ταξινομητή στα δοκιμαστικά δεδομένα μπορεί να υπολογιστεί ως ο αριθμός των ορθά ταξινομημένων αντικειμένων προς το σύνολο των αντικειμένων. Η ακρίβεια είναι ευθέως συνδεδεμένη με τον δείκτη σφάλματος (error rate), που θα δούμε παρακάτω, έτσι ώστε $accuracy = 1.0 - error\ rate$ (Sammut & Webb, 2010).

6.1.7. Ακρίβεια

Η ακρίβεια (Precision) ορίζεται ως ο λόγος των αληθώς θετικών προς το συνολικό αριθμό των θετικών αποτελεσμάτων που προέβλεψε το μοντέλο. Η ακρίβεια μπορεί έπειτα να οριστεί και με βάση τις τιμές των αληθώς θετικών και ψευδώς θετικών (Sammut & Webb, 2010).

$$Precision = TP / (TP + FN)$$

6.1.7. Ανάκληση

Η ανάκληση, όπως και η ακρίβεια, είναι μεγέθη που μας δείχνουν πόσο καλά λειτουργεί ένα σύστημα ανεύρεσης. Η ανάκληση ορίζεται ως ο λόγος των αληθώς θετικών ως προς το σύνολο των θετικών αποτελεσμάτων.

$$Recall = TP / (TP + FP)$$

6.1.8. Μέτρο F1

Το μέτρο F1 (F1-measure) χρησιμοποιείται για να αξιολογήσει την ακρίβεια της πρόβλεψης σε ένα πρόβλημα δυαδικής ταξινόμησης. Το F1 είναι ένα πιο σημαντικά κριτήρια αξιολόγησης στην τεχνητή νοημοσύνη. Συνοψίζει την προγνωστική απόδοση του μοντέλου συνδυάζοντας δύο κατά τα άλλα αντίπαλα κριτήρια, την ακρίβεια και την ανάκληση. Εξ ορισμού, το μέτρο F1 είναι ο αρμονικός μέσος ακρίβειας και ανάκλησης. Συνδυάζει τα δύο αυτά κριτήρια σε έναν αριθμό ως εξής:

$$\text{F1-score} = 2 \times \frac{\text{Precision} \times \text{Recall}}{\text{Precision} + \text{Recall}}$$

Όπως παρατηρούμε, λαμβάνοντας υπόψη και τα δύο άλλα κριτήρια, σημαίνει πως λαμβάνει υπόψη ταυτόχρονα και τα ψευδώς θετικά αλλά και τα ψευδώς αρνητικά. Όσο πιο καλή ακρίβεια και ανάκληση έχουμε, τόσο καλύτερο θα είναι και το F1. Όσο πιο κοντά στο 1 είναι, τόσο καλύτερο και το μοντέλο, δεδομένου ότι οι τιμές του είναι ανάμεσα στο 0 και το 1 (Zeya, 2021).

Κεφάλαιο 7

Αποτελέσματα

Με βάση τα δεδομένα που έχουμε στην διάθεσή μας για την παρούσα εργασία, επιλέξαμε να προχωρήσουμε με επιβλεπόμενη μηχανική μάθηση. Πιο συγκεκριμένα, θα τρέξουμε τα 3 μοντέλα που αναλύσαμε σε προηγούμενο κεφάλαιο έτσι ώστε να εξετάσουμε, κατά πόσον είναι εφικτή η μοντελοποίηση καταστροφών λόγω βροχοπτώσεων.

Για την μοντελοποίηση αυτή θα χρησιμοποιήσουμε μία βασική παράμετρο, την μέγιστη 24ωρη αθροιστική βροχόπτωση (R24). Άλλες παράμετροι για την βροχόπτωση, όπως η μέγιστη αθροιστική βροχόπτωση για μικρότερη χρονική περίοδο θα μπορούσαν να σηματοδοτούν την ένταση της βροχής, κάτι που παίζει σημαντικό ρόλο στην δημιουργία σοβαρών καταστροφών. Ωστόσο, θεωρούμε την R24 την πιο κατάλληλη μεταβλητή για την παρούσα εργασία, δεδομένου ότι με βάση τους Kotroni & Lagouvardos (2001, 2004) έχει καλύτερη προγνωστική ακρίβεια από αυτή που πετυχαίνουμε σε μικρότερες χρονικές περιόδους και θα μπορούσε να χρησιμοποιηθεί στο μέλλον ως βάση για την δημιουργία ενός Συστήματος Έγκαιρης Προειδοποίησης (ΣΕΠ; Early Warning System, EWS).

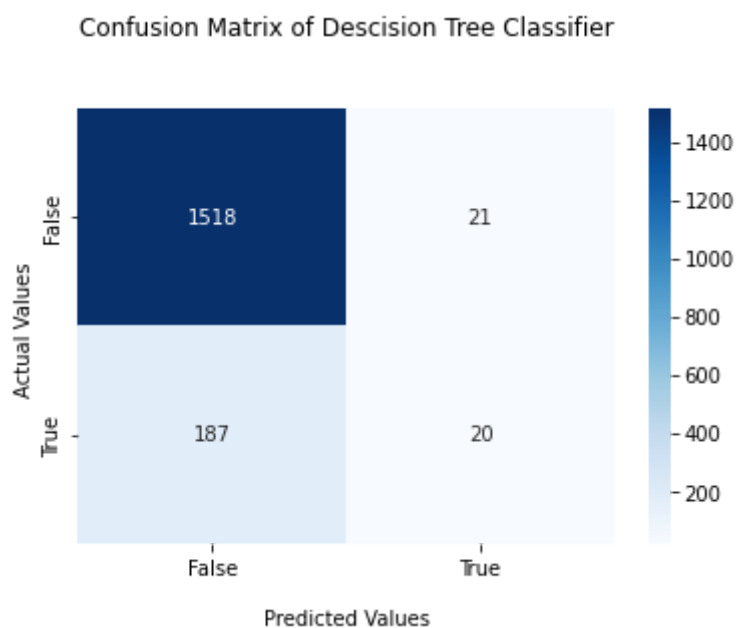
7.1 Ανάλυση στην Μητροπολιτική Περιοχή των Αθηνών ως σύνολο

Στην ανάλυση των δεδομένων είδαμε πως έχουμε έναν πίνακα 8726 γραμμών με 20 παραμέτρους. Η παράμετρος του DO θα χρησιμοποιηθεί ως εξαρτημένη μεταβλητή ενώ ως ανεξάρτητη θα χρησιμοποιήσουμε μόνο το R24, που όπως είδαμε είχε την πιο υψηλή συσχέτιση. Παρατηρώντας τα στοιχεία στους **Πίνακα 7.1**, **Πίνακα 7.2** και **Πίνακα 7.3** θα δούμε πως τα 3 μοντέλα δεν παρουσιάζουν μεγάλη διαφοροποίηση στα

αποτελέσματα τους. Στις εικόνες **Εικόνα 7.1**, **Εικόνα 7.2** και **Εικόνα 7.3** μπορούμε να δούμε και τους ακριβείς αριθμούς για τον πίνακα σύγκρισης του κάθε μοντέλου. Δεδομένων των γεωμορφολογικών στοιχείων που διέπουν την ΜΠΑ αλλά και την τοπικότητα των φαινομένων βροχής, είναι ασφαλές να πούμε πως τα χαμηλά σκορ στα αποτελέσματα ήταν αναμενόμενα.

Precision Score	0.488
Recall Score	0.097
Accuracy Score	0.881
F1 Score	0.161

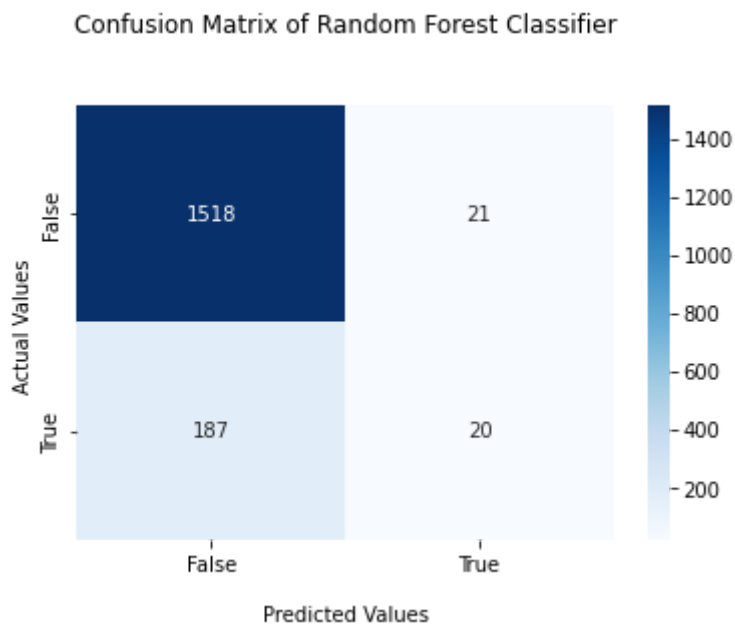
Πίνακας 7.1: Μετρικά στοιχεία μοντέλου DT για την ΜΠΑ.



Εικόνα 7.1: Πίνακας σύγκρισης μοντέλου DT για την ΜΠΑ.

Precision Score	0.471
Recall Score	0.116
Accuracy Score	0.880
F1 Score	0.186

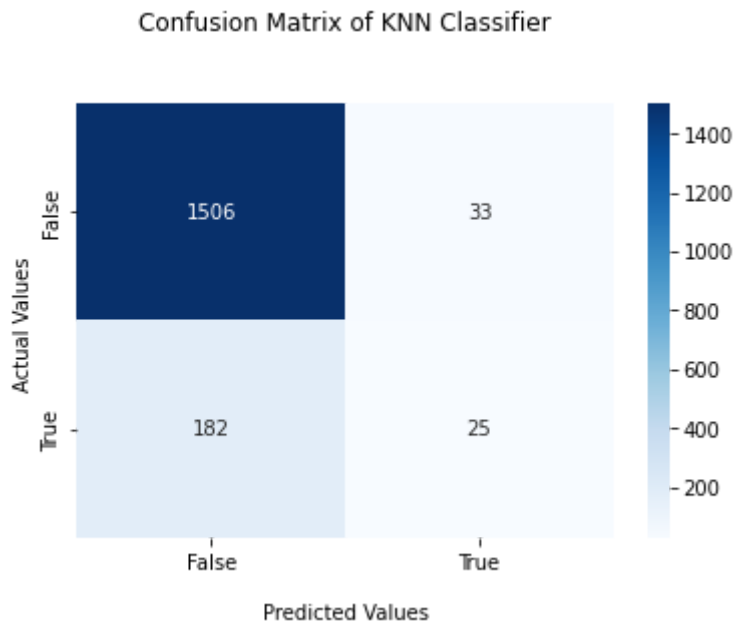
Πίνακας 7.2: Μετρικά στοιχεία μοντέλου RF για την ΜΠΑ.



Εικόνα 7.2: Πίνακας σύγκρισης μοντέλου RF για την ΜΠΑ.

Precision Score	0.431
Recall Score	0.121
Accuracy Score	0.877
F1 Score	0.189

Πίνακας 7.3: Μετρικά στοιχεία μοντέλου KNN για την ΜΠΑ.



Εικόνα 7.3: Πίνακας σύγκρισης μοντέλου KNN για την ΜΠΑ.

7.2 Ανάλυση ανά δήμο

Εφόσον δεν λάβαμε ικανοποιητικά αποτελέσματα από τα μοντέλα στο σύνολο της ΜΠΑ, δοκιμάσαμε να τρέξουμε τα 3 μοντέλα με τις ίδιες μεταβλητές για κάθε δήμο ξεχωριστά. Στους παρακάτω πίνακες (**Πίνακας 7.4**, **Πίνακας 7.5** και **Πίνακας 7.6**), μπορούμε να δούμε τα αποτελέσματα των κριτηρίων αξιολόγησης των μοντέλων για τους 10 δήμους με το καλύτερο F1- score κάθε φορά. Παρατηρούμε πως σε κάθε πίνακα έχουμε διαφορετική σειρά στους δήμους, όμως υπάρχουν κάποιοι δήμοι οι οποίοι εμφανίζονται και στα 3 μοντέλα (π.χ. FILOTHEI - PSYCHIKO). Συμπεραίνουμε, πως είμαστε σε θέση να μοντελοποιήσουμε τις καταστροφές σε πολύ καλό επίπεδο, μιας και το F1 φτάνει ακόμα και την μέγιστη τιμή του, αρκεί να έχουμε επαρκή δεδομένα για την περιοχή.

Municipality	Precision	Recall	Accuracy	F1	DO	Events	perc
AGION ANARGYRON - KAMATEROY	0.75	0.75	0.875	0.75	15	79	19
ALIMOY	0.5	1	0.714	0.667	8	32	25
VRILISSION	0.5	1	0.8	0.667	7	23	30
KHFISIAS	0.625	0.714	0.881	0.667	43	209	21
FILOTHEI - PSYCHIKO	1	0.5	0.96	0.667	14	123	11
PETROPOLIS	1	0.5	0.75	0.667	6	16	38
AHARNON	0.5	0.571	0.774	0.533	34	152	22
MEGAREON	0.333	1	0.857	0.5	15	69	22
GLYFADAS	0.429	0.6	0.684	0.5	28	94	30
PERISTERIOU	0.625	0.357	0.745	0.455	50	232	22

Πίνακας 7.4: Οι 10 δήμοι της ΜΠΑ με τα καλύτερα μετρικά στοιχεία με βάση το μοντέλο DT.

Municipality	Precision	Recall	Accuracy	F1	DO	Events	perc
AGION ANARGYRON - KAMATEROY	0.75	0.75	0.875	0.75	15	79	19
FILOTHEI - PSYCHIKO	1	0.5	0.96	0.667	14	123	11
KHFISIAS	0.625	0.714	0.881	0.667	43	209	21
VRILISSION	0.5	1	0.8	0.667	7	23	30
PETROPOLIS	1	0.5	0.75	0.667	6	16	38
ALIMOY	0.5	1	0.714	0.667	8	32	25
AHARNON	0.5	0.571	0.774	0.533	34	152	22
MEGAREON	0.333	1	0.857	0.5	15	69	22
AMAROYSIOY	0.5	0.5	0.786	0.5	31	138	22
GLYFADAS	0.429	0.6	0.684	0.5	28	94	30

Πίνακας 7.5: Οι 10 δήμοι της ΜΠΑ με τα καλύτερα μετρικά στοιχεία με βάση το μοντέλο RF.

Municipality	Precision	Recall	Accuracy	F1	DO	Events	perc
PERAMATOS	1	1	1	1	5	20	25
SPATON - ARTEMIDOS	0.75	1	0.929	0.857	9	66	14
FILOTHEI - PSYCHIKO	1	0.5	0.96	0.667	14	123	11
MARATHINOS	1	0.5	0.944	0.667	11	86	13
MEGAREON	0.5	1	0.929	0.667	15	69	22
AMAROYSIOY	0.5	0.5	0.786	0.5	31	138	22
PALLINHS	0.5	0.5	0.667	0.5	9	30	30
GALATSIΟΥ	0.5	0.333	0.8	0.4	8	74	11
AHARNON	0.5	0.286	0.774	0.364	34	152	22
NEAS SMYRNHS	0.5	0.25	0.857	0.333	13	136	10

Πίνακας 7.6: Οι 10 δήμοι της ΜΠΑ με τα καλύτερα μετρικά στοιχεία με βάση το μοντέλο KNN.

Κεφάλαιο 8

Επίλογος

Ο πρωταρχικός σκοπός της παρούσας εργασίας ήταν να αποδείξει πως μπορούμε να μοντελοποιήσουμε τις καταστροφές που προέρχονται από καιρικά φαινόμενα και πιο συγκεκριμένα, τις βροχοπτώσεις. Ένα τέτοιο εγχείρημα βασίζεται πρωτίστως στα δεδομένα και την διαθεσιμότητά τους. Έτσι, τα μοντέλα είναι ιδιαιτέρως αδύναμα στις περιπτώσεις όπου τα ποιοτικά δεδομένα δεν είναι ευρέως διαθέσιμα. Επίσης, η γεωγραφική κάλυψη παίζει έναν καθοριστικό ρόλο τόσο στην σύνδεση όσο και στην ανάλυση των δεδομένων. Ως εκ τούτου, ένας χώρος όπως η μητροπολιτική περιοχή των Αθηνών απαιτεί ένα πολύ μεγαλύτερο πλήθος μετεωρολογικών σταθμών ώστε να είμαστε σε θέση να καταγράψουμε πλήρως ένα καιρικό γεγονός, όπως και τα όρια επίδρασής του.

Στοχεύοντας στην βελτιστοποίηση του κάθε μοντέλου, μελλοντικά θα μπορούσαμε να ασχοληθούμε με μεταβλητές παραμετροποίησης ή ακόμη και εμπλουτισμό των εκπαιδευτικών δεδομένων από εναλλακτικές πηγές και διαφορετικούς τύπους. Επί παραδείγματι, οι κλησεις σε μονάδες ασφαλείας όπως και οι κρατικές αποζημιώσεις θα μπορούσαν να αποτελέσουν μία σημαντική πηγή δεδομένων, όταν βέβαια είναι διαθέσιμα και έπειτα από κατάλληλη και επιμελή μορφοποίηση. Τέλος, θα μπορούσαν να εξεταστούν οι γεωμορφολογικοί παράγοντες της κάθε περιοχής όπως και οι χρήσεις γης ή ακόμη και κατηγοριοποιήσεις όπως οι Τοπικές Κλιματικές Ζώνες (Local Climate Zones) με σκοπό να λάβουμε υπόψη παράγοντες που ευθύνονται για την τρωτότητα μίας περιοχής.

Ευχαριστίες

Σε αυτό το σημείο θα ήθελα να ευχαριστήσω θερμά τον καθηγητή μου, για την υπομονή και την καθοδήγησή του. Επίσης, θα ήθελα να ευχαριστήσω τους συνεργάτες μου για την κατανόηση και την πολύτιμη βοήθειά τους. Τέλος, δεν θα μπορούσα να ξεχάσω τους αγαπημένους μου συγγενείς και φίλους που τόσα χρόνια με στηρίζουν με κάθε δυνατό τρόπο.

Βιβλιογραφία

- Agrawal, S. K. (2021, July 20). *Evaluation Metrics For Classification Model | Classification Model Metrics*. Analytics Vidhya; <https://www.facebook.com/AnalyticsVidhya/>.
<https://www.analyticsvidhya.com/blog/2021/07/metrics-to-evaluate-your-classification-model-to-take-the-right-decisions/>
- AltexSoft. (2022, March 18). *Semi-Supervised Learning, Explained | AltexSoft*. AltexSoft; AltexSoft. <https://www.altexsoft.com/blog/semi-supervised-learning/>
- André, C., Monfort, D., Bouzit, M., & Vinchon, C. (2013). Contribution of insurance data to cost assessment of coastal flood damage to residential buildings: insights gained from Johanna (2008) and Xynthia (2010) storm events. *Natural Hazards and Earth System Sciences*, 8, 2003–2012. <https://doi.org/10.5194/nhess-13-2003-2013>
- Antonio Javier , Hovstad Janna, Ruth Mathilde, Parmeggiani Stefano, Patriarca Riccardo, Paltrinieri Nicola, & Paltrinieri, N. (2022). A Machine Learning Approach to Analyze Natural Hazards Accidents Scenarios. *Chemical Engineering Transactions*, 397–402. <https://doi.org/10.3303/CET2291067>
- Arinta, R. R., & Andi W.R., E. (2019). Natural Disaster Application on Big Data and Machine Learning: A Review. *2019 4th International Conference on Information Technology, Information Systems and Electrical Engineering (ICITISEE)*. <https://doi.org/10.1109/icitisee48480.2019.9003984>
- Arlot, S., & Celisse, A. (2010). A survey of cross-validation procedures for model selection. *Statistics Surveys*, none. <https://doi.org/10.1214/09-ss054>

-
- Badr, H. S., Zaitchik, B. F., & Guikema, S. D. (2014). Application of Statistical Models to the Prediction of Seasonal Rainfall Anomalies over the Sahel. *Journal of Applied Meteorology and Climatology*, 3, 614–636. <https://doi.org/10.1175/jamc-d-13-0181.1>
- Basabe, P. (2013). Hyogo Framework for Action 2005–2015. In *Encyclopedia of Natural Hazards* (pp. 508–516). Springer Netherlands. http://dx.doi.org/10.1007/978-1-4020-4399-4_180
- Bewtra, A. (2022). *The Ultimate Guide to Semi-Supervised Learning [2022]*. The Ultimate Guide to Semi-Supervised Learning. <https://www.v7labs.com/blog/semi-supervised-learning-guide#h1>
- Blumenthal, B., & Nyberg, L. (2018). The impact of intense rainfall on insurance losses in two Swedish cities. *Journal of Flood Risk Management*, S2. <https://doi.org/10.1111/jfr3.12504>
- Burkov, A. (2020). *Machine Learning Engineering*. True Positive Incorporated.
- Chauhan, N. S. (2022, February 9). *Decision Tree Algorithm, Explained - KDnuggets*. KDnuggets; <https://www.facebook.com/kdnuggets>. <https://www.kdnuggets.com/2020/01/decision-tree-algorithm-explained.html>
- Chikalamo, E. E., Mavrouli, O. C., Ettema, J., van Westen, C. J., Muntohar, A. S., & Mustofa, A. (2020). Satellite-derived rainfall thresholds for landslide early warning in Bogowonto Catchment, Central Java, Indonesia. *International Journal of Applied Earth Observation and Geoinformation*, 102093. <https://doi.org/10.1016/j.jag.2020.102093>
- Contributors to Wikimedia projects. (2002, January 8). *Supervised learning - Wikipedia*. Wikipedia, the Free Encyclopedia; Wikimedia Foundation, Inc. https://en.wikipedia.org/wiki/Supervised_learning
-

-
- Contributors to Wikimedia projects. (2003, May 25). *Unsupervised learning* - *Wikipedia*. Wikipedia, the Free Encyclopedia; Wikimedia Foundation, Inc. https://en.wikipedia.org/wiki/Unsupervised_learning
- Cortès, M., Turco, M., Llasat-Botija, M., & Llasat, M. C. (2018). The relationship between precipitation and insurance data for floods in a Mediterranean region (northeast Spain). *Natural Hazards and Earth System Sciences*, 3, 857–868. <https://doi.org/10.5194/nhess-18-857-2018>
- Cruz, A. M., Steinberg, L. J., & Vetere-Arellano, A. L. (2006). Emerging Issues for Natech Disaster Risk Management in Europe. *Journal of Risk Research*, 5, 483–501. <https://doi.org/10.1080/13669870600717657>
- De Luca, D. L., & Versace, P. (2017). Diversity of Rainfall Thresholds for early warning of hydro-geological disasters. *Advances in Geosciences*, 53–60. <https://doi.org/10.5194/adgeo-44-53-2017>
- Delua, J. (2021). *Supervised vs. Unsupervised Learning: What's the Difference?* | IBM. IBM - United States. <https://www.ibm.com/cloud/blog/supervised-vs-unsupervised-learning>
- Department of Regional Development and Environment Executive Secretariat for Economic and Social Affairs. (1991). *CHAPTER 2 - NATURAL HAZARD RISK REDUCTION IN PROJECT FORMULATION AND EVALUATION*. OAS - Organization of American States: Democracy for Peace, Security, and Development. <http://www.oas.org/dsd/publications/unit/oea66e/ch02.htm>
- EAAE. (2022). Hellenic Association of Insurance Companies. <http://www1.eaee.gr/>
- Economic losses from climate-related extremes in Europe*. (2020). European Environment Agency's Home Page — European Environment Agency. <https://www.eea.europa.eu/ims/economic-losses-from-climate-related>

-
- England, W. L. (1988). An Exponential Model Used for optimal Threshold selection on ROC Curves. *Medical Decision Making*, 2, 120–131. <https://doi.org/10.1177/0272989x8800800208>
- Faccini, F., Luino, F., Paliaga, G., Roccati, A., & Turconi, L. (2021). Flash Flood Events along the West Mediterranean Coasts: Inundations of Urbanized Areas Conditioned by Anthropic Impacts. *Land*, 6, 620. <https://doi.org/10.3390/land10060620>
- Fernandez, M. A. L. (2019). *migariane/cvAUROC: Cross-validated AUC for Stata users*. Zenodo. <https://zenodo.org/record/2592814>
- Flood risk: Underestimated natural hazards | Munich Re.* (n.d.). Home | Munich Re. Retrieved October 6, 2022, from <https://www.munichre.com/en/risks/natural-disasters-losses-are-trending-upwards/floods-and-flash-floods-underestimated-natural-hazards.html>
- Ghosh, S. K., & Sar, A. K. (2021). The Sendai Framework for Disaster Risk Reduction and Disaster Management Plan of Indian Railways. *Disaster Advances*, 79–89. <https://doi.org/10.25303/148da7921>
- Hoeppe, P. (2016). Trends in weather related disasters – Consequences for insurers and society. *Weather and Climate Extremes*, 70–79. <https://doi.org/10.1016/j.wace.2015.10.002>
- Jongman, B., Ward, P. J., & Aerts, J. C. J. H. (2012). Global exposure to river and coastal flooding: Long term trends and changes. *Global Environmental Change*, 4, 823–835. <https://doi.org/10.1016/j.gloenvcha.2012.07.004>
- Kotroni, V., & Lagouvardos, K. (2001). Precipitation forecast skill of different convective parameterization and microphysical schemes: Application for the

cold season over Greece. *Geophysical Research Letters*, 10, 1977–1980.
<https://doi.org/10.1029/2000gl012705>

Kotroni, V., & Lagouvardos, K. (2004). Evaluation of MM5 High-Resolution Real-Time Forecasts over the Urban Area of Athens, Greece. *Journal of Applied Meteorology*, 11, 1666–1678. <https://doi.org/10.1175/jam2170.1>

Kotsiantis, S. B., Zaharakis, I. D., & Pintelas, P. E. (2006). Machine learning: a review of classification and combining techniques. *Artificial Intelligence Review*, 3, 159–190. <https://doi.org/10.1007/s10462-007-9052-3>

Leal, M., Fragoso, M., Lopes, S., & Reis, E. (2020). Material damage caused by high-magnitude rainfall based on insurance data: Comparing two flooding events in the Lisbon Metropolitan Area and Madeira Island, Portugal. *International Journal of Disaster Risk Reduction*, 101806. <https://doi.org/10.1016/j.ijdrr.2020.101806>

Lee, W. Y., Park, S. K., & Sung, H. H. (2021). The optimal rainfall thresholds and probabilistic rainfall conditions for a landslide early warning system for Chuncheon, Republic of Korea. *Landslides*, 5, 1721–1739. <https://doi.org/10.1007/s10346-020-01603-3>

Lorenz, E. (2017). Deterministic Nonperiodic Flow. In *Universality in Chaos* (pp. 367–378). CRC Press. <http://dx.doi.org/10.1201/9780203734636-38>

Maudoux, C., & Boumerdassi, S. (2021). Smart and Sustainable Agriculture. In *Smart and Sustainable Agriculture*. Springer International Publishing. http://dx.doi.org/10.1007/978-3-030-88259-4_8

Mbaabu, O. (2020). *Introduction to Random Forest in Machine Learning | Engineering Education (EngEd) Program | Section*. Engineering Education (EngEd) Program | Section.

<https://www.section.io/engineering-education/introduction-to-random-forest-in-machine-learning/#:~:text=A%20random%20forest%20is%20a,algorithm%20and%20how%20it%20works>.

Medlock, C., & Oppenheim, A. (2020). Optimal ROC Curves from Score Variable Threshold Tests. *ArXiv*. <https://doi.org/10.48550/ARXIV.2012.08391>

Menon, V., & Kolathayar, S. (2021). Review on Landslide Early Warning System: A Brief History, Evolution, and Controlling Parameters. In *Springer Tracts in Civil Engineering* (pp. 129–145). Springer Singapore. http://dx.doi.org/10.1007/978-981-16-5312-4_10

Moftakhari, H. R., AghaKouchak, A., Sanders, B. F., & Matthew, R. A. (2017). Cumulative hazard: The case of nuisance flooding. *Earth's Future*, 2, 214–223. <https://doi.org/10.1002/2016ef000494>

Moncoulon, D., Labat, D., Ardon, J., Leblois, E., Onfroy, T., Poulard, C., Aji, S., Rémy, A., & Quantin, A. (2014). Analysis of the French insurance market exposure to floods: a stochastic model combining river overflow and surface runoff. *Natural Hazards and Earth System Sciences*, 9, 2469–2485. <https://doi.org/10.5194/nhess-14-2469-2014>

Nisha, A. (2022, July 28). *K-nearest Neighbors in Scikit-learn - KDnuggets*. KDnuggets; <https://www.facebook.com/kdnuggets>. <https://www.kdnuggets.com/2022/07/knearest-neighbors-scikitlearn.html>

Papagiannaki, K., Lagouvardos, K., & Kotroni, V. (2013). A database of high-impact weather events in Greece: a descriptive impact analysis for the period 2001–2011. *Natural Hazards and Earth System Sciences*, 3, 727–736. <https://doi.org/10.5194/nhess-13-727-2013>

Papagiannaki, Katerina, Kotroni, V., Lagouvardos, K., Bezes, A., Vafeiadis, V., Messini, I., Kroustallis, E., & Totos, I. (2022). Identification of Rainfall Thresholds Likely to Trigger Flood Damages across a Mediterranean Region, Based on Insurance Data and Rainfall Observations. *Water*, 6, 994. <https://doi.org/10.3390/w14060994>

Performance Metrics in Machine Learning - Javatpoint. (2022). www.javatpoint.com. <https://www.javatpoint.com/performance-metrics-in-machine-learning>

Prahl, B. F., Rybski, D., Burghoff, O., & Kropp, J. P. (2015). Comparison of storm damage functions and their performance. *Natural Hazards and Earth System Sciences*, 4, 769–788. <https://doi.org/10.5194/nhess-15-769-2015>

Rashidi, H. H., Tran, N. K., Betts, E. V., Howell, L. P., & Green, R. (2019). Artificial Intelligence and Machine Learning in Pathology: The Present Landscape of Supervised Methods. *Academic Pathology*, 2374289519873088. <https://doi.org/10.1177/2374289519873088>

Rivero-Calle, S., Gnanadesikan, A., Del Castillo, C. E., Balch, W. M., & Guikema, S. D. (2015). Multidecadal increase in North Atlantic coccolithophores and the potential role of rising CO₂. *Science*, 348, 1533–1537. <https://doi.org/10.1126/science.aaa8026>

Sammut, C., & Webb, G. (2010). *Encyclopedia of Machine Learning*. <https://doi.org/10.1007/978-0-387-30164-8>

Scher, S., & Messori, G. (2018). Predicting weather forecast uncertainty with machine learning. *Quarterly Journal of the Royal Meteorological Society*, 144, 2830–2841. <https://doi.org/10.1002/qj.3410>

Segoni, S., Battistini, A., Rossi, G., Rosi, A., Lagomarsino, D., Catani, F., Moretti, S., & Casagli, N. (2015). Technical Note: An operational landslide early warning

-
- system at regional scale based on space–time-variable rainfall thresholds. *Natural Hazards and Earth System Sciences*, 4, 853–861. <https://doi.org/10.5194/nhess-15-853-2015>
- Spekkers, M. H., Clemens, F. H. L. R., & ten Veldhuis, J. A. E. (2015). On the occurrence of rainstorm damage based on home insurance and weather data. *Natural Hazards and Earth System Sciences*, 2, 261–272. <https://doi.org/10.5194/nhess-15-261-2015>
- Spekkers, M. H., Kok, M., Clemens, F. H. L. R., & ten Veldhuis, J. A. E. (2014). Decision-tree analysis of factors influencing rainfall-related building structure and content damage. *Natural Hazards and Earth System Sciences*, 9, 2531–2547. <https://doi.org/10.5194/nhess-14-2531-2014>
- Thiese, M. S., Ronna, B., & Ott, U. (2016). P value interpretations and considerations. *Journal of Thoracic Disease*, 9, E928–E931. <https://doi.org/10.21037/jtd.2016.08.16>
- Tripathi, S., Srinivas, V. V., & Nanjundiah, R. S. (2006). Downscaling of precipitation for climate change scenarios: A support vector machine approach. *Journal of Hydrology*, 3–4, 621–640. <https://doi.org/10.1016/j.jhydrol.2006.04.030>
- Verbakel, J. Y., Steyerberg, E. W., Uno, H., De Cock, B., Wynants, L., Collins, G. S., & Van Calster, B. (2020). ROC curves for clinical prediction models part 1. ROC plots showed no added value above the AUC when evaluating the performance of clinical prediction models. *Journal of Clinical Epidemiology*, 207–216. <https://doi.org/10.1016/j.jclinepi.2020.01.028>
- Wallace, J. M., & Hobbs, P. V. (2006). *Atmospheric Science*. Elsevier.

-
- Wang, Y. (Victor), & Sebastian, A. (2021). Community flood vulnerability and risk assessment: An empirical predictive modeling approach. *Journal of Flood Risk Management*, 3. <https://doi.org/10.1111/jfr3.12739>
- Zeya, L. (2021, November 23). *Essential Things You Need to Know About F1-Score*. Towardsdatascience; <https://towardsdatascience.com/essential-things-you-need-to-know-about-f1-score-dbd973bf1a3#:~:text=F1%2Dscore%20is%20one%20of,competing%20metrics%20%E2%80%94%20precision%20and%20recall>.
- Zhou, Q., Panduro, T. E., Thorsen, B. J., & Arnbjerg-Nielsen, K. (2013). Verification of flood damage modelling using insurance data. *Water Science and Technology*, 2, 425–432. <https://doi.org/10.2166/wst.2013.268>
- Maes, S., Tuyls, K., Vanschoenwinkel, B., & Manderick, B. (2002). Credit Card Fraud Detection Using Bayesian and Neural Networks. In *Proceedings of the First International NAISO Congress on NEURO FUZZY TECHNOLOGIES January 16 - 19, 2002 (Havana, Cuba)* Proceedings of the First International NAISO Congress on NEURO FUZZY TECHNOLOGIES January 16 - 19, 2002 (Havana, Cuba).
- Chan, Philip & Stolfo, Salvatore. (1998). Toward Scalable Learning with Non-uniform Class and Cost Distributions: A Case Study in Credit Card Fraud Detection.
- Allison, P. D. (2014). Measures of fit for logistic regression, Paper No. 1485-2014. Paper presented at the SAS Global Forum 2014 Conference, Washington D.C. 23-26 March, 2014.
- Guha-Sapir, Debarati. 2020. EM-DAT, the Emergency Events Database. Université catholique de Louvain (UCL), Brussels, Belgium. Center for Research on the Epidemiology of Disasters (CRED). www.emdat.be. Date accessed 23 December 2020

Torgersen, Geir & Bjerkholt, Jarle & Kvaal, Knut & Lindholm, Oddvar. (2015). Correlation between extreme rainfall and insurance claims due to urban flooding – Case study fredrikstad, Norway. Journal of Urban and Environmental Engineering. 9. 127-138. 10.4090/juee.2015.v9n2.127138.