



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
& ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

Διερεύνηση μεθόδων μηχανικής μάθησης για την πρόβλεψη δασικών πυρκαγιών με δημόσια δεδομένα

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

της

ΠΑΡΘΕΝΑΣ ΣΤΑΜΠΟΛΙΔΟΥ

Επιβλέπουσα καθηγήτρια: ΛΟΥΤΑ ΜΑΛΑΜΑΤΗ

ΑΝΑΠΛΗΡΩΤΡΙΑ ΚΑΘΗΓΗΤΡΙΑ

ΚΟΖΑΝΗ/ΙΟΥΝΙΟΣ/2024



HELLENIC DEMOCRACY
UNIVERSITY OF WESTERN MACEDONIA
SCHOOL OF ENGINEERING
DEPARTMENT OF ELECTRICAL
& COMPUTER ENGINEERING

Exploring Machine Learning methods for predicting forest fires using public data

THESIS

PARTHENA STAMPOLIDOU

Supervisor: LOUTA MALAMATI

Substitute Teacher

KOZANI/JUNE/2024



ΕΛΛΗΝΙΚΗ ΔΗΜΟΚΡΑΤΙΑ
ΠΑΝΕΠΙΣΤΗΜΙΟ ΔΥΤΙΚΗΣ ΜΑΚΕΔΟΝΙΑΣ
ΠΟΛΥΤΕΧΝΙΚΗ ΣΧΟΛΗ
ΤΜΗΜΑ ΗΛΕΚΤΡΟΛΟΓΩΝ ΜΗΧΑΝΙΚΩΝ
& ΜΗΧΑΝΙΚΩΝ ΥΠΟΛΟΓΙΣΤΩΝ

ΔΗΛΩΣΗ ΜΗ ΛΟΓΟΚΛΟΠΗΣ ΚΑΙ ΑΝΑΛΗΨΗΣ ΠΡΟΣΩΠΙΚΗΣ ΕΥΘΥΝΗΣ

Δηλώνω ρητά ότι, σύμφωνα με το άρθρο 8 του Ν. 1599/1986 και τα άρθρα 2,4,6 παρ. 3 του Ν. 1256/1982, η παρούσα Διπλωματική Εργασία με τίτλο "Διερεύνηση μεθόδων μηχανικής μάθησης για την πρόβλεψη δασικών πυρκαγιών με δημόσια δεδομένα" καθώς και τα ηλεκτρονικά αρχεία και πηγαίοι κώδικες που αναπτύχθηκαν ή τροποποιήθηκαν στα πλαίσια αυτής της εργασίας και αναφέρονται ρητώς μέσα στο κείμενο που συνοδεύουν, και η οποία έχει εκπονηθεί στο Τμήμα Ηλεκτρολόγων Μηχανικών και Μηχανικών Υπολογιστών του Πανεπιστημίου Δυτικής Μακεδονίας, υπό την επίβλεψη του μέλους του Τμήματος κ. Λούτα Μαλαμάτη αποτελεί αποκλειστικά προϊόν προσωπικής εργασίας και δεν προσβάλλει κάθε μορφής πνευματικά δικαιώματα τρίτων και δεν είναι προϊόν μερικής ή ολικής αντιγραφής, οι πηγές δε που χρησιμοποιήθηκαν περιορίζονται στις βιβλιογραφικές αναφορές και μόνον. Τα σημεία όπου έχω χρησιμοποιήσει ιδέες, κείμενο, αρχεία ή / και πηγές άλλων συγγραφέων, αναφέρονται ευδιάκριτα στο κείμενο με την κατάλληλη παραπομπή και η σχετική αναφορά περιλαμβάνεται στο τμήμα των βιβλιογραφικών αναφορών με πλήρη περιγραφή. Απαγορεύεται η αντιγραφή, αποθήκευση και διανομή της παρούσας εργασίας, εξ ολοκλήρου ή τμήματος αυτής, για εμπορικό σκοπό. Επιτρέπεται η ανατύπωση, αποθήκευση και διανομή για σκοπό μη κερδοσκοπικό, εκπαιδευτικής ή ερευνητικής φύσης, υπό την προϋπόθεση να αναφέρεται η πηγή προέλευσης και να διατηρείται το παρόν μήνυμα. Ερωτήματα που αφορούν τη χρήση της εργασίας για κερδοσκοπικό σκοπό πρέπει να απευθύνονται προς τον συγγραφέα. Οι απόψεις και τα συμπεράσματα που περιέχονται σε αυτό το έγγραφο εκφράζουν τον συγγραφέα και μόνο.

Copyright (C) Ονοματεπώνυμο Φοιτητή & Επιβλέποντα, Έτος, Πόλη

Copyright (C) Σταμπολίδου Παρθένα, Λούτα Μαλαμάτη, 2024, Κοζάνη

Υπογραφή Φοιτητή: _____

Περίληψη

Οι δασικές πυρκαγιές θεωρούνται από τα πιο επικίνδυνα και καταστροφικά φαινόμενα παγκόσμιο επίπεδο. Είναι ένα εξαιρετικά κρίσιμο φαινόμενο λόγω της πολυπλοκότητάς του στην πρόβλεψη. Η επιστημονική κοινότητα έχει επιδείξει σημαντική προσπάθεια στην αντιμετώπιση αυτού του προβλήματος, διεξάγοντας πολυάριθμες έρευνες για την πρόβλεψη, την παρακολούθηση και την πρόληψη των δασικών πυρκαγιών. Αυτές οι έρευνες χρησιμοποιούν διάφορες τεχνικές και στρατηγικές τεχνητής νοημοσύνης αλλά και ενημερωμένες δορυφορικές εικόνες για την άμεση παρακολούθηση των πυρκαγιών.

Η παρούσα πτυχιακή εργασία διερευνά την εφαρμογή μεθόδων μηχανικής μάθησης για την πρόβλεψη δασικών πυρκαγιών, αξιοποιώντας δημόσια δεδομένα από το δορυφορικό σύστημα MODIS (Moderate Resolution Imaging Spectroradiometer) για την Ελλάδα από το 2000 έως το 2023. Η μελέτη επικεντρώνεται στη χρήση τριών αλγορίθμων μηχανικής μάθησης: Τυχαίο Δάσος (Random Forest), Λογιστική Παλινδρόμηση (Logistic Regression) και Μηχανών Υποστήριξης Διανυσμάτων (Support Vector Machine). Σκοπός της εργασίας είναι να αξιολογήσει την αποτελεσματικότητα αυτών των αλγορίθμων στην πρόβλεψη εμφάνισης δασικών πυρκαγιών, συγκρίνοντας την ακρίβεια και την απόδοση τους. Τα αποτελέσματα της έρευνας δείχνουν ότι ο αλγόριθμος Τυχαίο Δάσος (Random Forest) επιτυγχάνει υψηλότερη ακρίβεια πρόβλεψης σε σχέση με τους άλλους δύο αλγορίθμους. Η εργασία καταλήγει με προτάσεις για μελλοντική έρευνα, υπογραμμίζοντας την ανάγκη για περαιτέρω βελτιώσεις στα μοντέλα πρόβλεψης, την ενσωμάτωση επιπλέον δεδομένων καθώς και την εφαρμογή μοντέλων deep learning.

Λέξεις Κλειδιά: Πρόβλεψη Δασικών Πυρκαγιών, Μηχανική Μάθηση, MODIS, Αλγόριθμοι Μηχανικής Μάθησης

Abstract

Forest fires are widely regarded as one of the most dangerous and devastating phenomena worldwide. Due to its intricate nature of forecasting, it remains an exceptionally critical challenge. The scientific community has shown considerable efforts towards addressing this issue through extensive research on prediction, monitoring and prevention of forest fires. These studies employ various machine learning techniques and cutting-edge satellite imagery for real-time fire monitoring.

This thesis investigates the application of machine learning methods for predicting forest fires, utilizing publicly available data from the Moderate Resolution Imaging Spectroradiometer (MODIS) satellite system covering Greece from 2000 to 2023. The study specifically examines three machine learning algorithms: Random Forest, Logistic Regression and Support Vector Machine (SVM). The objective is to assess the efficacy of these algorithms in forecasting forest fire occurrences by comparing their accuracy and performance. The research findings indicate that Random Forest outperforms the other two algorithms. The thesis concludes with recommendations for future research, emphasizing the necessity for further improvement in the prediction models, integration of additional datasets and the exploration of deep learning models.

Keywords: Forest Fire Prediction, Machine Learning, MODIS, Machine Learning Algorithms

Ευχαριστίες

Θα ήθελα να ευχαριστήσω θερμά την επιβλέπουσα καθηγήτρια της διπλωματικής μου εργασίας κα Λούτα Μαλαμάτη, καθώς και τον υποψήφιο διδάκτορα του τμήματος κο Ευάγγελο Τσίπη, για την αμέριστη βοήθεια και την καθοδήγηση τους, καθ' όλη τη διάρκεια της εκπόνησης της.

Παρθένα Σταμπολίδου

Περιεχόμενα

Περίληψη.....	vii
Abstract	viii
Ευχαριστίες	ix
Κατάλογος Σχημάτων	2
Κατάλογος Πινάκων.....	3
Κεφάλαιο 1: Εισαγωγή	4
Κεφάλαιο 2: Θεωρητικό υπόβαθρο – Βιβλιογραφική Ανασκόπηση	7
Κεφάλαιο 3: Περιοχή Μελέτης: Ελλάδα	14
3.1 Συλλογή και προ επεξεργασία δεδομένων	14
3.2 Διερευνητική ανάλυση δεδομένων	20
3.3 Επεξεργασία Δεδομένων	30
Κεφάλαιο 4: Μέθοδοι μηχανικής μάθησης	31
4.1 Τυχαίο Δάσος (<i>Random Forest</i>)	31
4.2 <i>Support Vector Machine</i>	32
4.3 <i>Logistic Regression</i>	34
Κεφάλαιο 5: Αποτελέσματα και ανάλυση	35
5.1 <i>Random Forest</i>	35
5.2 <i>Support Vector Machine</i>	37
5.3 <i>Logistic Regression</i>	40
Σύγκριση Μεθόδων Μηχανικής Μάθησης.....	43
Κεφάλαιο 6: Συζήτηση και συμπεράσματα	44
Μελλοντική Έρευνα.....	45
Βιβλιογραφία	46

Κατάλογος Σχημάτων

Εικόνα 1 Τεχνητή Μάθηση	8
Εικόνα 2 Απεικόνιση hotspots.....	15
Εικόνα 3 Συχνότητα Πυρκαγιών στην Ελλάδα ανά Μήνα	21
Εικόνα 4 Κατανομή Πυρκαγιών στην Ελλάδα 2000-2023	22
Εικόνα 5 Μέγιστη Θερμοκρασία ανά έτος.....	22
Εικόνα 6 Κατανομή Δασικών Πυρκαγιών 2000-2023 στην Ελλάδα.....	25
Εικόνα 7 Κατανομή Δασικών Πυρκαγιών στην Ελλάδα το 2007	25
Εικόνα 8 Κατανομή Πυρκαγιών για τον μήνα Αύγουστο	26
Εικόνα 9 Συνολική έκταση που κάηκε ανά Περιοχή [2000 - 2024]	27
Εικόνα 10 Εκτάσεις καμένων περιοχών (2008-2024).....	28
Εικόνα 11 Περιοχές που έχουν πληγεί από Πυρκαγιές στην Ελλάδα	29
Εικόνα 12 Ποσοστό καμένων περιοχών ανά κατηγορία βλάστησης.	30
Εικόνα 13 Confusion Matrix για το μοντέλο Random Forest	35
Εικόνα 14 Σημαντικότητα μεταβλητών στο μοντέλο Random Forest	37
Εικόνα 15 Confusion Matrix για το μοντέλο SVM	39
Εικόνα 16 Σημαντικότητα μεταβλητών για το μοντέλο SVM	42
Εικόνα 17 Confusion Matrix για το μοντέλο Logistic Regression	41
Εικόνα 18 Σημαντικότητα μεταβλητών για το μοντέλο Logistic Regression	42
Εικόνα 19 Correlation matrix.....	43

Κατάλογος Πινάκων

Πίνακας 1 Περιγραφή Δεδομένων	19
Πίνακας 2 Πίνακας Σύγχυσης (Confusion Matrix) - Random Forest.....	36
Πίνακας 3 Πίνακας Σύγχυσης (Confusion Matrix) – SVM.....	39
Πίνακας 4 Πίνακας Σύγχυσης (Confusion Matrix) - Logistic Regression	41

Κεφάλαιο 1: Εισαγωγή

Τα δάση είναι ένα από τα πιο μεγάλα οικοσυστήματα, όπου παρέχει στέγαση σε αμέτρητους οργανισμούς της πανίδας και της χλωρίδας, ενώ παρέχει και οικοσυστημικές υπηρεσίες στον άνθρωπο όπως είναι ο καθαρός αέρας, το νερό, η διατήρηση του κλίματος καθώς και η δέσμευση του διοξειδίου του άνθρακα CO₂, που είναι πολύ σημαντικό για τον παγκόσμιο προϋπολογισμό του άνθρακα. Τα δασικά οικοσυστήματα απειλούνται από πολλούς κινδύνους, όπως την ξηρασία, τους τυφώνες και παθογόνα έντομα. Ωστόσο, τα τελευταία χρόνια, οι πυρκαγιές σε δασικές εκτάσεις απειλούν όσο ποτέ τα δάση και αποτελούν μία από τις πιο σοβαρές φυσικές καταστροφές με επιπτώσεις τόσο στο ίδιο το περιβάλλον αλλά και στην κοινωνία ευρύτερα [1]. Το 2013, η μέση έκταση που κάηκε στον κόσμο κατά τα τελευταία 16 χρόνια υπολογίζεται στα 340 εκατομμύρια εκτάσεις [2]. Πέραν, των καταστροφών αυτών οι δασικές πυρκαγιές επιφέρουν και μελλοντικά προβλήματα σε περιπτώσεις έντονων βροχοπτώσεων όπου αυξάνεται ο κίνδυνος των πλημμυρών από την εδαφική διάβρωση που επιφέρουν οι πυρκαγιές στην φύση του εδάφους.

Η πρόβλεψη, λοιπόν, των δασικών πυρκαγιών είναι όσο ποτέ άλλοτε επιβεβλημένη για τον μετριασμό των κινδύνων που ενέχει αλλά και για την προστασία του περιβάλλοντος καθώς προκαλούνται τεράστιες ζημιές στα οικοσυστήματα, καταστρέφονται σπίτια και υποδομές και θέτουν σε σοβαρό κίνδυνο την ανθρώπινη ζωή. Στο σημερινό συνεχώς μεταβαλλόμενο κόσμο, η σημασία των ακραίων μετεωρολογικών προβλέψεων δεν εκτιμάται τόσο καθώς δεν είναι απόλυτα έγκυρες και έγκαιρες για την καλύτερη δυνατή διαχείριση των επιπτώσεων τους [3] [4]. Ένας τομέας στον οποίο οι μετεωρολογικές προβλέψεις είναι ιδιαίτερα κρίσιμες είναι η πρόληψη και διαχείριση των δασικών πυρκαγιών. Παραδοσιακά, η διαχείριση των δασικών πυρκαγιών βασίζεται σε μεθόδους παρακολούθησης και πρόβλεψης, οι οποίες έχουν περιορισμούς ως προς την ακρίβεια και την αποτελεσματικότητά τους. Η δυνατότητα να προβλέπονται οι πιθανές εστίες φωτιάς διαδραματίζει ουσιαστικό ρόλο στο να μειωθούν οι επιβλαβείς επιπτώσεις στο φυσικό περιβάλλον αλλά και στον άνθρωπο.

Με την εξέλιξη της τεχνολογίας και της πληροφορικής, η πρόβλεψη των δασικών πυρκαγιών μπορεί να γίνει πιο ουσιαστική και αποτελεσματική. Οι πρόσφατες εξελίξεις σε αυτό τον τομέα, έχουν ανοίξει το δρόμο για αποτελεσματικότερες στρατηγικές πρόληψης και διαχείρισης των δασικών πυρκαγιών [5]. Η πρόβλεψη, μπορεί να βασιστεί σε πολλές μεθόδους και τεχνολογίες, όπως είναι η χρήση αισθητήρων, δορυφόρων αλλά και αλγορίθμων μηχανικής μάθησης. Η μηχανική μάθηση αναφέρεται στη χρήση αλγορίθμων και στατιστικών μοντέλων που επιτρέπουν στους υπολογιστές να μαθαίνουν

από δεδομένα και να κάνουν προβλέψεις ή να λαμβάνουν αποφάσεις χωρίς να είναι ρητά προγραμματισμένοι. Ο συνδυασμός αυτών μας επιτρέπει να προσδιορίσουμε με σαφήνεια και ακρίβεια τις περιοχές όπου ενδέχεται να είναι ευάλωτες σε πυρκαγιές. Με αποτέλεσμα, οι αρμόδιες αρχές να είναι έγκαιρα προετοιμασμένες για να αντιμετωπίσουν γρήγορα και αποτελεσματικά τέτοιες καταστάσεις.

Με τη χρήση τεχνικών μηχανικής μάθησης, οι ερευνητές και οι επιστήμονες μπορούν να αναλύσουν δεδομένα από διάφορες πηγές, όπως καιρικά πρότυπα, χαρακτηριστικά εδάφους, πυκνότητα βλάστησης και ιστορικά αρχεία πυρκαγιών για να αναπτύξουν μοντέλα πρόβλεψης που μπορούν να εκτιμήσουν τον κίνδυνο δασικών πυρκαγιών και να παρέχουν έγκαιρα προειδοποιήσεις. Δύο τομείς που έχουν σημειώσει τη μεγαλύτερη επιτυχία στην εφαρμογή μεθόδων μηχανικής μάθησης σε προβλήματα πρόβλεψης είναι η πρόβλεψη χρονοσειρών και η ιεραρχική μοντελοποίηση. Η πρόβλεψη χρονοσειρών είναι ιδιαίτερα σημαντική στο πλαίσιο της πρόβλεψης δασικών πυρκαγιών, καθώς περιλαμβάνει την ανάλυση ιστορικών δεδομένων για τον εντοπισμό μοτίβων και τάσεων. Οι αλγόριθμοι μηχανικής μάθησης μπορούν να εκπαιδευτούν για την ανάλυση μετεωρολογικών δεδομένων, όπως η θερμοκρασία, η υγρασία, η ταχύτητα του ανέμου και η βροχόπτωση μαζί με άλλες σχετικές μεταβλητές, όπως τα χαρακτηριστικά του εδάφους, την πυκνότητα της βλάστησης και τα ιστορικά αρχεία των πυρκαγιών. Αναλύοντας αυτά τα δεδομένα, τα μοντέλα μηχανικής μάθησης μπορούν να εντοπίσουν μοτίβα και σχέσεις που μπορεί να υποδεικνύουν αυξημένο κίνδυνο πυρκαγιάς. Στη συνέχεια, τα μοντέλα αυτά μπορούν να χρησιμοποιηθούν για την πρόβλεψη της πιθανότητας και της έντασης των δασικών πυρκαγιών σε συγκεκριμένες περιοχές, επιτρέποντας τη λήψη στοχευμένων μέτρων πρόληψης και σε ταχύτερους χρόνους αντίδρασης.

Εκτός από την πρόβλεψη χρονοσειρών, η ιεραρχική μοντελοποίηση έχει επίσης αποδειχθεί αποτελεσματική στην πρόβλεψη δασικών πυρκαγιών. Η ιεραρχική μοντελοποίηση περιλαμβάνει το συνδυασμό πολλαπλών προβλέψεων από διαφορετικές πηγές ή μοντέλα για τη δημιουργία μιας πιο ακριβούς και αξιόπιστης πρόβλεψης. Για παράδειγμα ένα ιεραρχικό μοντέλο μπορεί να ενσωματώσει προβλέψεις από διάφορους αλγόριθμους μηχανικής μάθησης, όπως ο Random Forest, ο Support Vector Machine ή μοντέλα Deep Learning. Αυτοί οι αλγόριθμοι μπορούν να παράγουν μεμονωμένες προβλέψεις από διαφορετικές πηγές δεδομένων και τεχνικές μοντελοποίησης. Σε αυτή την περίπτωση, οι ερευνητές μπορούν να λάβουν μια πιο ολοκληρωμένη και αξιόπιστη πρόβλεψη της συμπεριφοράς των δασικών πυρκαγιών. Τα μοντέλα μηχανικής μάθησης μπορούν να αναλύσουν τεράστιο όγκο δεδομένων σε πραγματικό χρόνο και να παρέχουν έγκαιρες πληροφορίες για την υποστήριξη των λήψεων αποφάσεων. Για παράδειγμα, οι αλγόριθμοι μηχανικής μάθησης μπορούν να αναλύσουν τις τρέχουσες καιρικές συνθήκες, τις δορυφορικές εικόνες και τα δεδομένα αισθητήρων από συσκευές

τηλεσκοπίσης για να εκτιμήσουν τον κίνδυνο εξάπλωσης της πυρκαγιάς και να καθορίσουν τις βέλτιστες στρατηγικές για τον περιορισμό της πυρκαγιάς και την εκκένωση των κατοικημένων περιοχών που απειλούνται, σώζοντας δυνητικά ζωές και ελαχιστοποιώντας τις ζημιές που προκαλούνται από τις δασικές πυρκαγιές.

Επιπλέον, η χρήση της μηχανικής μάθησης στην αξιολόγηση της απειλής δασικών πυρκαγιών μπορεί να βοηθήσει στην αποτελεσματικότερη κατανομή των πόρων. Για παράδειγμα, με τη χρήση των αλγορίθμων μηχανικής μάθησης μπορούν να εντοπιστούν μοτίβα και να δώσουν προτεραιότητα σε περιοχές που διατρέχουν μεγαλύτερο κίνδυνο εξάπλωσης της πυρκαγιάς ή έχουν μεγαλύτερη πιθανότητα σοβαρών ζημιών. Η συνεισφορά αυτή είναι σημαντική και για την ανάπτυξη προληπτικών μέτρων, όπως είναι οι ελεγχόμενες καύσεις, ο καθαρισμός των αντιπυρικών ζωνών ή εφαρμογή αυστηρότερων κανονισμών.

Συμπερασματικά, η εφαρμογή αλγορίθμων μηχανικής μάθησης στην αξιολόγηση και πρόληψη της απειλής δασικών πυρκαγιών έχει τη δυνατότητα να βελτιώσει σημαντικά την ικανότητά μας να μετριάσουμε τους κινδύνους που συνδέονται με τις δασικές πυρκαγιές. Η μηχανική μάθηση μπορεί να φέρει επανάσταση στον τομέα της αξιολόγησης και της πρόληψης των απειλών δασικών πυρκαγιών. Μπορεί να παρέχει έγκαιρες πληροφορίες, να βοηθήσει στην παρακολούθηση και την έγκαιρη ανίχνευση, να καταλείψει αποτελεσματικά τους πόρους και να βοηθήσει στην ανάπτυξη προληπτικών στρατηγικών πρόληψης.

Στόχος λοιπόν της εργασίας αυτής είναι να παρουσιάσει και να αναλύσει τα μοντέλα μηχανικής μάθησης που έχουν αναπτυχθεί για την πρόβλεψη των δασικών πυρκαγιών. Στην συνέχεια η εργασία παρουσιάζει αναλυτικά τους αλγορίθμους που έχουν εφαρμοστεί και καταγράφει τις γενικές λειτουργίες και τις ιδιαιτερότητες τους στον τομέα της πρόβλεψης πυρκαγιών. Η εργασία προχωρά στην υλοποίηση βασικών παραδειγμάτων κάνοντας χρήση τους αλγορίθμους Τυχαίο Δάσος (Random Forest), Λογιστική Παλινδρόμηση (Logistic Regression) και Μηχανών Υποστήριξης Διανυσμάτων (Support Vector Machine – SVM), εφαρμόζοντας δημόσια δεδομένα από το σύστημα MODIS. Τέλος, οι αποδόσεις των μοντέλων αυτών συγκρίνονται και αξιολογούνται ενώ προτείνεται και η βελτιστοποίηση των μοντέλων.

Ένας από τους κύριους στόχους αυτής της εργασίας είναι η αξιολόγηση των μοντέλων μηχανικής μάθησης για την πρόβλεψη δασικών πυρκαγιών, μετά από την ανάλυση και παρουσίαση διαφόρων μεθόδων. Επίσης, πολύ σημαντικό είναι να καθοριστούν οι παράγοντες αυτοί που έχουν ιδιαίτερη σημασία στην έναρξη μια δασικής πυρκαγιάς. Δηλαδή, παράγοντες που έχουν να κάνουν με την θερμοκρασία, την υγρασία του εδάφους αλλά και την κατάσταση της χλωρίδας.

Κεφάλαιο 2: Θεωρητικό υπόβαθρο – Βιβλιογραφική Ανασκόπηση

Στο παρόν κεφάλαιο γίνεται μια εκτενής αναφορά στον τομέα της μηχανικής μάθησης και παρουσιάζεται η ανασκόπηση της βιβλιογραφικής έρευνας σχετικά με την χρήση μεθόδων μηχανικής μάθησης για την πρόβλεψη δασικών πυρκαγιών .

Η μηχανική μάθηση είναι ένα υποσύνολο της τεχνητής νοημοσύνης που επικεντρώνεται στην ανάπτυξη αλγορίθμων και μοντέλων που επιτρέπουν στους υπολογιστές να μαθαίνουν και να κάνουν προβλέψεις ή να λαμβάνουν αποφάσεις χωρίς να προγραμματίζονται ρητά. Η μηχανική μάθηση αποτελείται από διάφορα βασικά στοιχεία. Τα συστατικά αυτά περιλαμβάνουν τα δεδομένα, τους αλγορίθμους, την εκπαίδευση και την πρόβλεψη. Τα δεδομένα παίζουν σημαντικό ρόλο στη μηχανική μάθηση, καθώς παρέχουν τις απαραίτητες πληροφορίες για την εκπαίδευση και την πρόβλεψη. Οι αλγόριθμοι είναι οι μαθηματικές διαδικασίες και οι κανόνες που διέπουν τον τρόπο με τον οποίο τα μοντέλα μηχανικής μάθησης μαθαίνουν από τα δεδομένα και κάνουν προβλέψεις. Τα μοντέλα μηχανικής μάθησης είναι οι αναπαραστάσεις των εκπαιδευμένων προτύπων μεταξύ των δεδομένων εισόδου και της επιθυμητής εξόδου. Η εκπαίδευση είναι η διαδικασία τροφοδοσίας των δεδομένων στο μοντέλο μηχανικής μάθησης και η προσαρμογή των παραμέτρων του για τη βελτιστοποίηση της απόδοσης του. Κατά τη διάρκεια της εκπαίδευσης, το μοντέλο μαθαίνει από τα δεδομένα και προσαρμόζει τις εσωτερικές του παραμέτρους για να βελτιώσει την ικανότητα του να κάνει ακριβείς προβλέψεις. Η πρόβλεψη είναι το τελικό στάδιο της μηχανικής μάθησης, όπου το εκπαιδευμένο μοντέλο χρησιμοποιείται για να κάνει προβλέψεις σε νέα δεδομένα.

Οι αλγόριθμοι μηχανικής μάθησης κατηγοριοποιούνται σε δύο βασικούς τύπους, σε μάθηση με επίβλεψη (supervised learning) και μάθηση χωρίς επίβλεψη (unsupervised learning). Στην μάθηση με επίβλεψη, το μοντέλο μηχανικής μάθησης εκπαιδεύεται χρησιμοποιώντας δεδομένα με ετικέτες, όπου κάθε σημείο δεδομένων συνοδεύεται από τη σωστή ετικέτα. Αντίθετα, στη μάθηση χωρίς επίβλεψη, το μοντέλο εκπαιδεύεται χρησιμοποιώντας μη επισημασμένα δεδομένα και πρέπει να ανακαλύψει μοτίβα και σχέσεις στα δεδομένα μόνο του.

Artificial Intelligence		
<i>Σχεδιασμό Ευφρών πρακτόρων που αντιλαμβάνονται το περιβάλλον και λαμβάνουν ευφρείς αποφάσεις για τη μεγιστοποίηση ενός στόχου.</i>		
Machine Learning		
<i>Δίνει την δυνατότητα στους υπολογιστές να μαθαίνουν χωρίς να προγραμματίζονται άμεσα.</i>		
Supervised Learning	Unsupervised Learning	Reinforcement Learning
<i>Ταξινόμηση, Παλινδρόμηση</i>	<i>Ομαδοποίηση, Μείωση Διαστάσεων</i>	<i>Μεγιστοποίηση ανταμοιβής</i>

Εικόνα 1 Τεχνητή Μάθηση

Οι αλγόριθμοι μηχανικής μάθησης έχουν γίνει όλο και πιο πολύτιμοι για την αντιμετώπιση σύνθετων εργασιών πρόβλεψης της εμφάνισης δασικών πυρκαγιών.

Ένας αλγόριθμος για δυαδική αναζήτηση στο πλαίσιο πρόβλεψης δασικών πυρκαγιών είναι ο αλγόριθμος Τυχαίου Δάσους (Random Forest). Ο αλγόριθμος αυτός είναι μια μέθοδος μάθησης συνόλου που συνδυάζει πολλαπλά δέντρα απόφασης για να βελτιώσει την ακρίβεια της ταξινόμησης. Λειτουργεί με την κατασκευή ενός πλήθους δέντρων απόφασης κατά τη φάση της εκπαίδευσης και την εξαγωγή της κλάσης που αποτελεί τον τρόπο των κλάσεων (ταξινόμησης) των επιμέρους δέντρων. Είναι ένας ισχυρός αλγόριθμος συνόλου (ensemble) με καλή ανθεκτικότητα σε υπερπροσαρμογή (overfitting) και τη δυνατότητα να διαχειρίζεται εύκολα μεγάλα σύνολα δεδομένων. Λόγω της ικανότητας του να λαμβάνει υπόψη τη σημασία των χαρακτηριστικών, είναι κατάλληλος για προβλήματα όπου υπάρχει μεγάλος αριθμός παραμέτρων.

Ένας άλλος αποτελεσματικός αλγόριθμος για δυαδική ταξινόμηση σε αυτόν τον τομέα είναι η Μηχανή Υποστήριξης Διανυσμάτων (Support Vector Machine- SVM). Είναι ένας αλγόριθμος μάθησης με επίβλεψη που προσδιορίζει το βέλτιστο υπερεπίπεδο (hyperparameter) που διαχωρίζει τις δύο κλάσεις με το μέγιστο περιθώριο. Αυτός ο αλγόριθμος έχει εφαρμοστεί με επιτυχία σε διάφορες οικολογικές και περιβαλλοντικές εργασίες πρόβλεψης, συμπεριλαμβανομένης της πρόβλεψης δασικών πυρκαγιών [6].

Επιπλέον, η λογιστική παλινδρόμηση (Logistic Regression) είναι μια δημοφιλής επιλογή για προβλήματα δυαδικής ταξινόμησης. Αυτός ο αλγόριθμος μοντελοποιεί την πιθανότητα ενός δυαδικού αποτελέσματος ως συνάρτηση μιας ή περισσότερων μεταβλητών πρόβλεψης. Η λογιστική

παλινδρόμηση είναι σχετικά απλή στην εφαρμογή και στην ερμηνεία, καθιστώντας την κατάλληλη για την πρόβλεψη δασικών πυρκαγιών. Είναι ένας απλός αλγόριθμος που χρησιμοποιείται ευρέως για προβλήματα ταξινόμησης, ειδικά όταν τα δεδομένα είναι γραμμικά διαχωρίσιμα. Είναι γρήγορος και αποτελεσματικός σε πολλές περιπτώσεις [7].

Κατά την επιλογή του καταλληλότερου αλγορίθμου μηχανικής μάθησης για την πρόβλεψη δασικών πυρκαγιών, είναι σημαντικό να ληφθούν υπόψη τα χαρακτηριστικά του προβλήματος και τα διαθέσιμα δεδομένα. Οι αλγόριθμοι, Random Forest, SVM και Logistic Regression είναι όλες κατάλληλες επιλογές που έχουν επιδείξει πολλά υποσχόμενα αποτελέσματα σε αυτόν τον τομέα. Ωστόσο, η επιλογή εξαρτάται τελικά από παράγοντες όπως το μέγεθος και η ποιότητα του συνόλου δεδομένων, η πολυπλοκότητα των σχέσεων μεταξύ των μεταβλητών και το επιθυμητό επίπεδο ερμηνευσιμότητας (explainability) του μοντέλου [6].

Προκειμένου να αξιολογηθεί η απόδοση και η αποτελεσματικότητα των μοντέλων μηχανικής μάθησης, είναι ζωτικής σημασίας η διεξαγωγή διαδικασιών αξιολόγησης. Στην συγκεκριμένη εργασία καθώς θα χρησιμοποιηθούν οι αλγόριθμοι μηχανικής μάθησης Random Forest, Support Vector Machine και Logistic Regression θα διερευνήσουμε την αξιολόγηση των τριών μοντέλων μηχανικής μάθησης. Η αξιολόγηση αυτή προσφέρει πολύτιμες πληροφορίες για την επιλογή του καταλληλότερου μοντέλου για την πρόβλεψη των δασικών πυρκαγιών.

Για την αξιολόγηση των μοντέλων ML, καταμετρήθηκαν τα ψευδώς θετικά (False Positives) και τα ψευδώς αρνητικά (False Negative) για τον υπολογισμό της ακρίβειας (Accuracy), της ακρίβειας (Precision), της ανάκλησης (Recall) και του F1- score.

Οι όροι ψευδώς θετικά (False Positives) και ψευδώς αρνητικά (False Negatives) είναι πολύ σημαντικοί για την κατανόηση της απόδοσης του μοντέλου. Όταν το μοντέλο μηχανικής μάθησης προβλέπει λανθασμένα ένα στιγμιότυπο ότι ανήκει στην θετική κατηγορία ενώ στην πραγματικότητα ανήκει στην αρνητική κατηγορία, τότε το στιγμιότυπο αυτό χαρακτηρίζεται ως false positive. Στην αντίθετη περίπτωση, δηλαδή όταν το μοντέλο μηχανικής μάθησης προβλέπει ένα στιγμιότυπο ότι ανήκει στην αρνητική κατηγορία ενώ στην πραγματικότητα ανήκει στην θετική, τότε αυτό χαρακτηρίζεται ως false negative.

Accuracy: Ο λόγος των σωστά προβλεπόμενων παρατηρήσεων προς το συνολικό αριθμό των παρατηρήσεων.

$$Accuracy = \frac{TP + TN}{TP + TN + FP + FN}$$

Precision: Ο λόγος των σωστά προβλεπόμενων θετικών παρατηρήσεων προς το σύνολο των προβλεπόμενων θετικών παρατηρήσεων.

$$Precision = \frac{TP}{TP + FP}$$

Recall: Μετρά πόσο καλά ο ταξινομητής μπορεί να ανιχνεύσει θετικές παρατηρήσεις.

$$Recall = \frac{TP}{TP + FN}$$

F1-score: Ο μέσος όρος της ακρίβειας και της ανάκλησης.

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

Η χρήση μεθόδων μηχανικής μάθησης για την πρόβλεψη των δασικών πυρκαγιών με δημόσια δεδομένα έχουν εξεταστεί σε αρκετές μελέτες. Παρακάτω, ακολουθεί μια παρουσίαση των ερευνών αυτών στα πλαίσια της βιβλιογραφικής αναφοράς.

Ομάδα ερευνητών, χρησιμοποίησαν την μέθοδο SVM (Support Vector Machine) και την RF (Random Forest) για να αξιολογήσουν την ευαισθησία των δασών στην εμφάνιση πυρκαγιάς αλλά και να χαρτογραφήσουν την χωρική σχέση μεταξύ της εμφάνισης φωτιάς και την πρόβλεψη των περιβαλλοντικών παραγόντων [8]. Παρόμοια εργασία πραγματοποιήθηκε και από άλλη ομάδα ερευνητών [9], στην οποία χρησιμοποιήθηκαν ετερογενή δεδομένα από διαφορετικές πηγές και επικεντρώθηκαν σε χώρο-χρονικά στοιχεία για την δημιουργία graph-based μοντέλου με στόχο την πρόβλεψη δασικών πυρκαγιών. Επιπλέον, σε άλλη μελέτη [10] ασχολήθηκαν με την υλοποίηση ANN (Artificial Neural Network) και SVM (Support Vector Machine) για να προσδιορίσουν την πιθανότητα δασικής πυρκαγιάς στο Μεσογειακό δασικό οικοσύστημα. Υπογράμμισαν ακόμα ότι η χρήση μετεωρολογικών δεδομένων σε μοντέλα πρόβλεψης δασικής πυρκαγιάς είναι αρκετά σημαντικό για την ακρίβεια των μοντέλων αυτών.

Σε μελέτη που χρησιμοποιήθηκε η μέθοδος Random Forest (RF) είχε ως στόχο τον προσδιορισμό του ρόλου που έχουν οι κλιματικές συνθήκες αλλά και οι ανθρώπινες δραστηριότητες στην πιθανότητα πρόκλησης πυρκαγιάς στο δάσος του Ατλαντικού [11]. Ενώ, μέθοδοι των ANN (Artificial Neural

Network) και SVM (Support Vector Machine) εφαρμόστηκαν για την πρόβλεψη δασικών πυρκαγιών στην Περιφέρεια Quangxi της Κίνας, υπογραμμίζοντας το αυξανόμενο ενδιαφέρον που υπάρχει στην επιστημονική κοινότητα για την χρήση αλγορίθμων μηχανικής μάθησης για την πρόβλεψη δασικών πυρκαγιών [12].

Για την πρόβλεψη και αντιμετώπιση δασικών πυρκαγιών χρησιμοποιώντας δεδομένα δορυφορικής εποπτείας (Remote Sensing) και συγκεκριμένα τον αισθητήρα MODIS, εφαρμόστηκαν επίσης οι μέθοδοι των ANN (Artificial Neural Network) και SVM (Support Vector Machine). Οι παράγοντες στους οποίους εστίασε η έρευνα, είναι Θερμικές Ανωμαλίες (Thermal Anomalies), η θερμοκρασία του εδάφους (LST) και ο δείκτης υγείας των καλλιεργειών (NDVI) όπου προήρθαν από τους δορυφόρους Terra και Aqua για τα δάση του Καναδά την χρονική περίοδο 2013-2014. Σύμφωνα, με τα αποτελέσματα της μελέτης αυτής τα Νευρωνικά Δίκτυα (ANN) είχαν την καλύτερη απόδοση με 98.32% ακρίβεια, σε σχέση με τα SVM με ακρίβεια 97.48%. Ενώ, διαπιστώθηκε ότι οι τρεις παράμετροι που αναφέρθηκαν παραπάνω ήταν πολύ σωστά επιλεγμένες [13].

Μελέτη που εστίασε στην περιοχή της νοτιοδυτικής Ισπανίας και συγκεκριμένα δυτικά της Ανδαλουσίας όπου είναι μια πολύπαθη περιοχή από πυρκαγιές, χρησιμοποιήθηκαν 20 μεταβλητές που αφορούν μετεωρολογικά και περιβαλλοντικά δεδομένα για την πρόβλεψη εμφάνισης πυρκαγιάς. Οι αλγόριθμοι που εξετάστηκαν στην συγκεκριμένη περίπτωση ήταν το Random Forest (RF), το Logistic Regression (LR), το Support Vector Machines (SVM) και το Multi-Layer Perceptron. Ενώ, χρησιμοποιήθηκε και η μέθοδος του GridSearchCV για την εύρεση των βέλτιστων υπερπαραμέτρων (hyper-parameters) για την βελτιστοποίηση των μοντέλων. Σύμφωνα με τα αποτελέσματα της έρευνας οι αλγόριθμοι των Logistic Regression και Multi-Layer Perceptron είχαν την καλύτερη απόδοση ενώ καλύτερη ακρίβεια παρείχαν οι μεταβλητές που αφορούσαν την ταχύτητα του ανέμου, την μέση θερμοκρασία, την υγρασία αλλά και ο δείκτης υγείας των καλλιεργειών [14].

Σε μια ακόμα έρευνα που αφορά την περιοχή της Ισπανίας για την περίοδο 1988-2007, παρουσιάζονται οι αποδόσεις των Random Forest, Boosted Regression Trees και των Support Vector Machines με τα αποτελέσματα τους να συγκρίνονται με αυτά της Logistic Regression, όπου είναι το πιο ευρέως διαδεδομένο μοντέλο για την πιθανότητα εμφάνισης πυρκαγιάς. Η τελική εξαρτημένη μεταβλητή κατηγοριοποιήθηκε σε περιοχές με υψηλή εμφάνιση πυρκαγιών, δηλαδή σε αυτές τις περιοχές έχουν σημειωθεί κατά το παρελθόν τουλάχιστον 2 πυρκαγιές και σε περιοχές με χαμηλή εμφάνιση πυρκαγιών. Τα αποτελέσματα της έρευνας έδειξαν ότι η χρήση των Random Forests είχε καλύτερη απόδοση με AUC τιμές 0.746, ενώ λιγότερο αποδοτική μέθοδος ήταν η χρήση SVM με AUC τιμές 0.709. Η έρευνα αυτή ανέδειξε ως κύριους παράγοντες για την εμφάνιση πυρκαγιάς σε δάση, την

αμέλεια και τα ατυχήματα από μηχανές που υπάρχουν κοντά ή μέσα στην δασική έκταση, από την χρήση φωτιάς για τις γεωργικές εργασίες αλλά και την αύξηση και παρουσία της ανθρώπινης δραστηριότητας στις περιοχές αυτές [15].

Εργασία όπου επικεντρώθηκε και πάλι στην περιοχή της Ιβηρικής χερσονήσου και συγκεκριμένα στην περιοχή Mostesinos στην Πορτογαλία, εστίασε στην πρόβλεψη δασικών πυρκαγιών με μοντέλα μηχανικής μάθησης. Χρησιμοποιήθηκαν Linear Regression, Ridge Regression και Lasso Regression. Η εφαρμογή αυτών ανέδειξε ως πιο αποδοτική μέθοδο αυτή της Linear Regression χωρίς όμως και οι άλλες δύο μέθοδοι να μην είναι ιδιαίτερα ικανοποιητικές [16].

Μοντέλα μηχανικής μάθησης, όπως Random Forest, XGBoost, Multi-Layer Perceptron, Logistic και Linear Regression εφαρμόστηκαν για την πρόβλεψη εμφάνισης και μεγέθους της πυρκαγιάς. Τα μοντέλα αυτά εκπαιδεύτηκαν σε ένα μεγάλο σύνολο δεδομένων που περιείχε παρατηρήσεις δασικών πυρκαγιών σε όλη την παγκόσμια κλίμακα. Τα ευρήματα της έρευνας αυτής όσον αφορά το αποδοτικότερο μοντέλο μηχανικής μάθησης, ανέδειξε το XGBoost ενώ οι παράμετροι που αναδείχθηκαν ως οι πιο σημαντικοί για την πρόβλεψη των πυρκαγιών ήταν το εμβαδόν της περιοχής όπου εμφανίστηκε η πυρκαγιά, οι μετεωρολογικοί παράγοντες με ιδιαίτερη σημασία στις βροχοπτώσεις, στον δείκτη υγείας των καλλιεργειών, στα χαρακτηριστικά της πυρκαγιάς αλλά και στην περιοχή όπου εκτυλίχθηκε, με ιδιαίτερη έμφαση σε παραμέτρους που αφορούσαν τον πληθυσμό και την κλίση του εδάφους [17].

Έρευνα [18] εστίασε σε αλγόριθμους Μηχανικής Μάθησης για την πρόβλεψη και ανίχνευση δασικών πυρκαγιών και συγκεκριμένα σε CNN (Convolutional Neural Network) και σε Logistic Regression, με την χρήση των CNN να αναδεικνύεται ως πιο αποδοτική μέθοδος. Και σε αυτή την έρευνα οι παράγοντες που ενσωματώθηκαν στα μοντέλα αφορούσαν μετεωρολογικές, κλιματικές και τοπολογικές συνθήκες, ενώ και εδώ παρατηρήθηκε ως σημαντικός παράγοντας η ανθρώπινη δραστηριότητα.

Μια ακόμη μελέτη όπου χρησιμοποιήθηκαν έξι διαφορετικά μοντέλα μηχανικής μάθησης για την αξιολόγηση της ευαισθησίας των τροπικών δασών στην Ινδία στην εμφάνιση πυρκαγιών χρησιμοποιώντας δεδομένα της περιόδου 2001-2020 από τον αισθητήρα MODIS υλοποιήθηκε το 2022 από ερευνητές [19]. Η αξιολόγηση των μοντέλων έγινε με μέθοδο ROC/AUC και τα μοντέλα SVM και ANN σημείωσαν εξαιρετική απόδοση με 0.908 και 0.903 αντίστοιχα. Η μελέτη αυτή επεσήμανε την σημασία των κλιματικών παραμέτρων και της τοπογραφίας για τον εντοπισμό περιοχών με υψηλό κίνδυνο πυρκαγιάς.

Μέθοδοι μηχανικής μάθησης εξετάστηκαν για την πρόβλεψη εμφάνισης φωτιάς σε βαλτώδεις περιοχές της Ινδονησίας [20]. Οι μέθοδοι που χρησιμοποιήθηκαν είναι SVM , KNN (k-Nearest Neighborhood), Logistic Regression, Decision Trees, Naïve Bayes, και AdaBoost. Τα δεδομένα που συλλέχθηκαν αφορούσαν τις μετεωρολογικές παραμέτρους για την επαρχία της Νότιας Καλιμαντάν της Ινδονησίας για το έτος 2018 και πιο συγκεκριμένα την θερμοκρασία του εδάφους, την ταχύτητα του ανέμου, την υγρασία και τον δείκτη υγείας των καλλιεργειών. Τα συμπεράσματα της μελέτης ήταν ότι η μέθοδος AdaBoost είχε την καλύτερη απόδοση έναντι των άλλων.

Εργασία ερευνητών, επικεντρώθηκε στην χρήση ασύρματων αισθητήρων σε δασικές εκτάσεις για την ανίχνευση πυρκαγιών χρησιμοποιώντας μεθόδους μηχανικής μάθησης [21]. Οι μέθοδοι που χρησιμοποιήθηκαν ήταν SVM, Linear Regression, Decision Trees και ANN. Το dataset που χρησιμοποιήθηκε περιλάμβανε πραγματικές τιμές για την θερμοκρασία, την υγρασία, τον άνεμο καθώς και πιο στοχευμένες παραμέτρους που συνδέονται με τις συνθήκες που επηρεάζουν την πυρκαγιά σε δασώδης περιοχές. Υποστηρίχθηκε ότι η μέθοδος της Linear Regression είχε την υψηλότερη ακρίβεια στην ανίχνευση πυρκαγιών σε αντίθεση με τις υπόλοιπες, ενώ σημειώθηκε ότι απαιτεί και λιγότερο επεξεργαστικό χρόνο.

Μια άλλη πρόταση για την πρόβλεψη πυρκαγιών σε δασώδης περιοχές ήταν η χρήση του μοντέλου Perceptron και της τεχνολογίας LoRa (Long Range) [22]. Με τη χρήση αισθητήρων συλλέχθηκαν μετεωρολογικά δεδομένα όπως η θερμοκρασία και η υγρασία για την εκπαίδευση του μοντέλου όπου επέτρεψε την πρόβλεψη του επιπέδου κινδύνου πυρκαγιάς στα δάση. Ωστόσο, η συλλογή των δεδομένων λόγω του μεγάλου όγκου επέφερε περιορισμούς και απαιτούσε περισσότερο χρόνο.

Κεφάλαιο 3: Περιοχή Μελέτης: Ελλάδα

Το παρόν κεφάλαιο παρουσιάζει την περιοχή μελέτης και την ανάλυση των δεδομένων που συλλέχθηκαν. Ακόμα περιγράφεται η επεξεργασία που χρειάστηκε να υλοποιηθεί στα δεδομένα, πριν χρησιμοποιηθούν για την εκπαίδευση των αλγορίθμων.

Για την παρούσα εργασία, η περιοχή μελέτης είναι η Ελλάδα (N: 41 45' 01, S: 34 48' 11, E: 36 06' 17, W: 39 51' 11). Η Ελλάδα βρίσκεται ανάμεσα σε Ευρώπη, Ασία και Αφρική με συνολική έκταση περίπου 131.957 τ.μ.. Χαρακτηρίζεται από ποικίλα γεωγραφικά χαρακτηριστικά, όπως βουνά, νησιά και μια μεγάλη ακτογραμμή. Η τοπολογία της χώρας είναι ποικίλη, με ορεινές περιοχές να κυριαρχούν στην ηπειρωτική Ελλάδα και πολυάριθμα νησιά σε Αιγαίο και Ιόνιο Πέλαγος. Η Ελλάδα είναι γνωστή για το ορεινό της ανάγλυφο και κυριαρχείται από οροσειρές, με υψηλότερη κορυφή τον Όλυμπο, που φτάνει τα 2.917 μέτρα πάνω από την επιφάνεια της θάλασσας. Άλλες οροσειρές της Ελλάδας είναι η Πίνδος που βρίσκεται στα βορειοδυτικά της χώρας και η οροσειρά της Ροδόπης που βρίσκεται στα βορειοανατολικά της χώρας. Η τοπολογία της χώρας χαρακτηρίζεται περαιτέρω και από την παρουσία πολυάριθμων νησιών. Συγκεκριμένα αποτελείται από περισσότερα από 6.000 νησιά, με τα μεγαλύτερα να είναι η Κρήτη, η Εύβοια και η Ρόδος.

Το κλίμα της Ελλάδας χαρακτηρίζεται κυρίως ως μεσογειακό, παρουσιάζοντας τοπικές διαφοροποιήσεις λόγω της γεωγραφικής ποικιλότητας της χώρας. Οι περισσότερες περιοχές της Ελλάδας, ιδιαίτερα οι παράκτιες περιοχές και τα νησιά έχουν μεσογειακό κλίμα. Αυτό συνεπάγεται ότι η θερμοκρασία το καλοκαίρι κυμαίνεται από 25° C έως 35° C, ενώ οι χειμώνες είναι πιο ήπιοι με θερμοκρασίες από 5° C έως 15° C. Τα ορεινά της χώρας χαρακτηρίζονται από ψυχρότερους χειμώνες με συχνές χιονοπτώσεις και πιο δροσερά καλοκαίρια σε σχέση με τις παράκτιες περιοχές. Σε περιοχές όπως η Ανατολική Στερεά Ελλάδα και η Θεσσαλία, παρουσιάζονται λιγότερες βροχοπτώσεις και πιο υψηλές θερμοκρασίες κατά τη διάρκεια του καλοκαιριού.

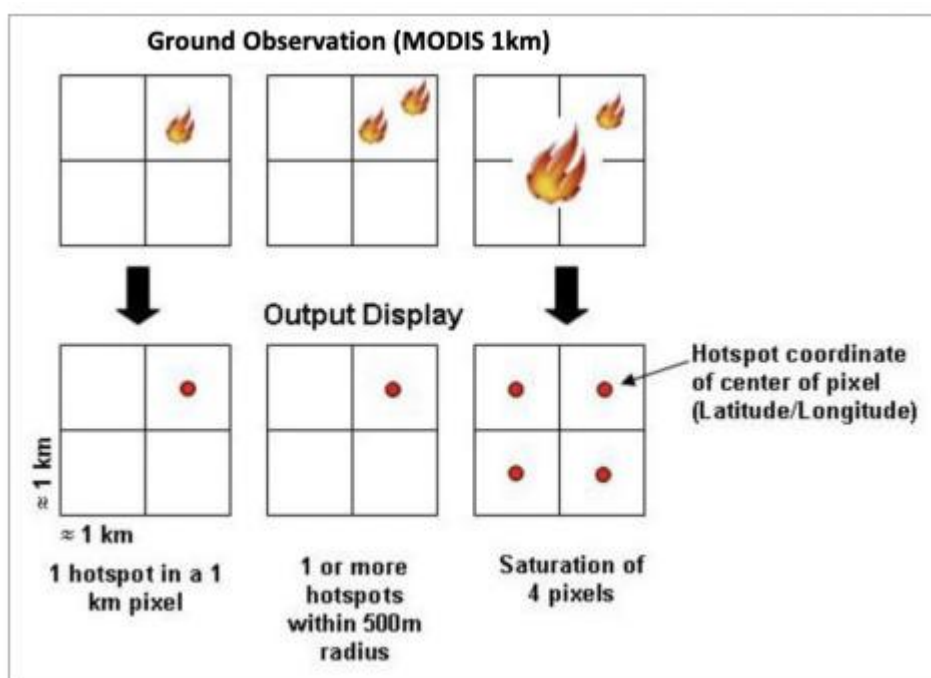
Τέλος, η περίοδος μελέτης της παρούσας εργασίας είναι από τον Οκτώβριο του 2000 έως και τον Ιανουάριο του 2023.

3.1 Συλλογή και προ επεξεργασία δεδομένων

Η εκδήλωση πυρκαγιάς μπορεί να ανιχνευθεί και να παρακολουθηθεί σε ένα από τα προϊόντα του αισθητήρα MODIS που βρίσκεται στους δορυφόρους Terra και Aqua. Οι δορυφόροι αυτοί ξεκίνησαν

την λειτουργία τους το 1999 και το 2002 αντίστοιχα. Ο αισθητήρας έχει πλάτος προβολής 2.330 χιλιόμετρα και έχει εικόνα από όλη την επιφάνεια της γης. Οι ανιχνευτές του αισθητήρα MODIS μετρούν 36 φασματικές ζώνες μεταξύ 0,405 και 14,385 μm και έχουν τη δυνατότητα απόκτησης δεδομένων σε τρεις διαφορετικές χωρικές αναλύσεις, σε 250m, 500m και 1000m.

Οι δορυφόροι λαμβάνουν στιγμιότυπα γεγονότων καθώς περνούν πάνω από τη Γη, εντοπίζοντας θερμικά σημεία ή ενεργές πυρκαγιές. Επιτρέπουν την παρακολούθηση των πυρκαγιών παγκοσμίως έως και τέσσερις φορές ημερησίως, δύο κατά την διάρκεια της μέρας και δύο κατά τη διάρκεια της νύχτας. Κάθε εντοπισμός πυρκαγιάς αντιστοιχεί στο κέντρο ενός pixel που περιέχει τη φωτιά. Ο αισθητήρας MODIS ανιχνεύει ενεργές πυρκαγιές σε εικονοστοιχεία 1 km υπό συνθήκες χωρίς σύννεφα. Η τοποθεσία όπου παρατηρείται το φαινόμενο είναι το κέντρο του εικονοστοιχείου και όχι απαραίτητα οι συντεταγμένες της περιοχής. Το ακριβές μέγεθος της φωτιάς δεν μπορούμε να το γνωρίζουμε αλλά γνωρίζουμε ότι βρίσκεται τουλάχιστον μια φωτιά εντός του εικονοστοιχείου.



Εικόνα 2 Απεικόνιση hotspots

Η ανίχνευση πυρκαγιάς πραγματοποιείται με την εφαρμογή ενός αλγορίθμου που βασίζεται στην έντονη εκπομπή μεσαίας υπέρυθρης ακτινοβολίας από την πυρκαγιά. Ο αλγόριθμος εξετάζει κάθε εικονοστοιχείο της σάρωσης του αισθητήρα MODIS και αποδίδει σε κάθε ένα εικονοστοιχείο τις κατηγορίες, ελλιπή δεδομένα, σύννεφα, νερό, όχι πυρκαγιά, πυρκαγιά ή άγνωστο [23].

Τα δεδομένα MODIS Terra Collection 6.1 για εστίες πυρκαγιάς από το 2000 έως το 2013 λήφθηκαν από το Fire Information for Resource Management System (FIRMS) (Σύστημα Πληροφοριών Πυρκαγιών για τη Διαχείριση Πόρων) (<https://earthdata.nasa.gov/firms>). Πρόκειται για ένα σύστημα που παρέχει πληροφορίες και παρακολούθηση σχετικά με τις πυρκαγιές σε παγκόσμιο επίπεδο, χρησιμοποιώντας δορυφορικές εικόνες και δεδομένα υπέρυθρων. Το σύστημα αυτό αναπτύχθηκε από τη NASA για την παροχή πληροφοριών σχετικά με τις τρέχουσες πυρκαγιές συμπεριλαμβανομένων των θερμικών ανωμαλιών που αντιμετωπίζονται από τις δορυφορικές αναλύσεις.

Επιπλέον, αυτά τα αρχεία δεδομένων περιέχουν το γεωγραφικό πλάτος και μήκος, την ημερομηνία και την ώρα λήψης. Επίσης, περιέχουν την εμπιστοσύνη (confidence) με εύρος από 0-100%.

Χρησιμοποιήθηκαν οι μέσες τιμές για κλιματικά δεδομένα από το 2000 έως το 2013. Αυτές οι κλιματικές μεταβλητές είναι η μέγιστη και ελάχιστη θερμοκρασία, το είδος βροχόπτωσης, η ταχύτητα του ανέμου.

Τα δεδομένα τα οποία έχουν συλλεχθεί από τους δορυφόρους Terra και Aqua περιέχουν εγγραφές για τον τύπο της περιοχής, τις επικρατούσες καιρικές συνθήκες, τις συντεταγμένες του συμβάντος καθώς και άλλες μεταβλητές που παρουσιάζονται στον παρακάτω πίνακα.

Όνομα μεταβλητών	Περιγραφή Δεδομένων
Latitude	Γεωγραφικό πλάτος όπου καταγράφεται η φωτιά
Longitude	Γεωγραφικό μήκος όπου καταγράφεται η φωτιά
Scan	Αντικατοπτρίζει το πραγματικό μέγεθος σάρωσης του εικονοστοιχείου
Track	Αντικατοπτρίζει το πραγματικό μέγεθος διαδρομής του εικονοστοιχείου
Acq_date	Ημερομηνία απόκτησης των δεδομένων
Acq_time	Ώρα απόκτησης των δεδομένων

Satellite	Όνομα δορυφόρου – Terra ή Aqua
Instrument	Όνομα του οργάνου μέτρησης – MODIS
Confidence	Εμπιστοσύνη ανίχνευσης % (εύρος 0-100)
Version	Συλλογή και πηγή επεξεργασίας των δεδομένων URT (Ultra Real-Time) RT (Real Time) Near Real-Time (NRT) SP (Standard Processing)- μόνο συλλογή
Brightness	Θερμοκρασία φωτεινότητας του καναλιού 21/22 του εικονοστοιχείου φωτιάς σε βαθμούς Kelvin
Bright_t31	Θερμοκρασία φωτεινότητας του εικονοστοιχείου φωτιάς σε βαθμούς Kelvin (Κανάλι 31)
Frp	Ρυθμός εκπομπής της ακτινοβολίας ανά μονάδα χρόνου από όλες τις φωτιές εντός ενός pixel. Σχετίζεται με τον ρυθμό καύσης βιομάζας και τον ρυθμό εκπομπών.
Daynight	Εάν η φωτιά καταγράφηκε ημέρα ή νύχτα
Type	Τύπος θερμού σημείου. 0: Πυρκαγιά σε βλάστηση 1: Ενεργό ηφαίστειο 2: Άλλη πηγή στερεάς γης 3: Ανοικτή θάλασσα
Temp max	Μέγιστη θερμοκρασία

Temp min	Ελάχιστη θερμοκρασία
Temp	Θερμοκρασία
Feels like max	Μέγιστη αίσθηση θερμοκρασίας
Feels like min	Ελάχιστη αίσθηση θερμοκρασίας
Dew	Σημείο δρόσου
Humidity	Υγρασία
Precip prob	Πιθανότητα βροχής
Precip cover	Κάλυψη βροχής
Precip type	Τύπος καιρού (βροχή, χιόνι, κ.λ.π.)
Snow	Ποσότητα χιονιού
Snow depth	Βάθος χιονιού
Wind gust	Ριπές ανέμου
Wind speed	Ταχύτητα ανέμου
Wind speed max	Μέγιστη ταχύτητα ανέμου
Wind speed mean	Μέση ταχύτητα ανέμου
Wind speed min	Ελάχιστη ταχύτητα ανέμου

Wind direction	Διεύθυνση του ανέμου
Sea level pressure	Πίεση στην επιφάνεια της θάλασσας
Cloud cover	Κάλυψη συννεφιάς
Visibility	Ορατότητα
Solar radiation	Ηλιακή ακτινοβολία
Solar energy	Ηλιακή ενέργεια
UV index	Δείκτης υπεριώδους ακτινοβολίας

Πίνακας 1 Περιγραφή Δεδομένων

Η τιμή εμπιστοσύνης των δεδομένων (confidence) χρησιμοποιείται για να αξιολογηθεί η ποιότητα κάθε εικονοστοιχείου της πυρκαγιάς. Για τον αισθητήρα MODIS κυμαίνεται από 0% έως και 100% και αναθέτει κάθε εικονοστοιχείο σε τρεις κλάσεις πυρκαγιών, σε χαμηλής, κανονικής και υψηλής εμπιστοσύνης. Η τάξη εμπιστοσύνης κυμαίνεται μεταξύ 0-30% για χαμηλή πιθανότητα πυρκαγιάς, 30-80% για μέτρια πιθανότητα πυρκαγιάς, ενώ για υψηλή ανιχνευσιμότητα πυρκαγιάς κυμαίνεται από 80-100% [24].

Το μέγεθος της πυρκαγιάς που μπορεί να ανιχνευθεί από τον αισθητήρα εξαρτάται από πολλούς παράγοντες. Πιο συγκεκριμένα, οι παράγοντες αυτοί είναι η γωνία σάρωσης του δορυφόρου, ο τύπος του βιότοπου, η θέση του ήλιου, η θερμοκρασία της επιφάνειας της γης, η νεφοκάλυψη, η ποσότητα του παραγόμενου καπνού και η κατεύθυνση του ανέμου. Ο αισθητήρας μπορεί να ανιχνεύσει πυρκαγιές είτε φλεγόμενες είτε σιγοκαμένες, μεγέθους 1000τ.μ.. Βέβαια, σε εξαιρετικά ιδανικές συνθήκες παρατήρησης μπορεί να ανιχνεύσει και πυρκαγιές μικρότερου μεγέθους [23].

Η θερμοκρασία φωτεινότητας (bright_t31) είναι στην πραγματικότητα μέτρηση των φωτονίων σε ένα συγκεκριμένο μήκος κύματος που δέχεται ο δορυφόρος αλλά παρουσιάζεται σε μονάδες θερμοκρασίας και συγκεκριμένα στην κλίμακα μέτρηση Kelvin.

Η μεταβλητή Fire Radiative Power (FRP) αναπαριστά το ρυθμό εκπομπής ακτινοβολίας της ενέργειας που προέρχεται από πυρκαγιές εντός ενός εικονοστοιχείου. Χαρακτηρίζει την ένταση της πυρκαγιάς και σχετίζεται με τον ρυθμό καύσης της βιομάζας και τον ρυθμό εκπομπών. Παράγοντες όπως η

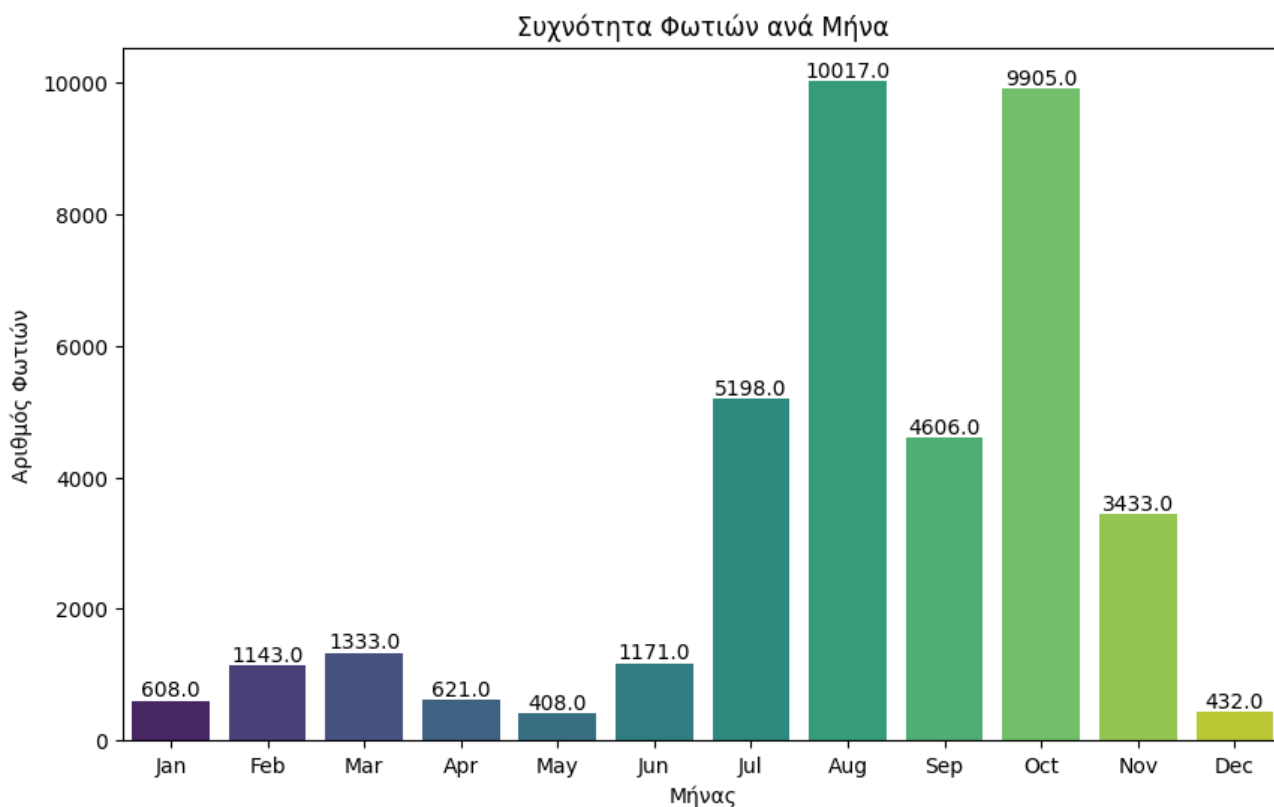
χωρική ανάλυση, η θερμοκρασία, η γωνία θέασης του δορυφόρου και ο χρόνος παρατήρησης επηρεάζουν την ακρίβεια της μεταβλητής.

Οι τύποι από 0 έως και 2 απεικονίζουν αποκλειστικά την ξηρά, ενώ ο τύπος 3 που αναφέρεται σε «ανοικτή θάλασσα» πιθανώς απεικονίζει υπεράκτιες φωτοβολίδες αερίου. Ο τύπος 2, δηλαδή, η κατηγορία «άλλη στατική πηγή εδάφους» περιλαμβάνει στατικά σημεία που εκπέμπουν θερμότητα και παρατηρούνται επανειλημμένα για πάνω από 16 ημέρες κατά την διάρκεια του έτους. Επίσης περιλαμβάνει σημεία που εντοπίζονται σε αστική περιοχή [24].

Οι τιμές των scan και track περιλαμβάνονται στα πεδία των δεδομένων των ενεργών πυρκαγιών. Η τιμή σάρωσης (scan) αντιπροσωπεύει τη χωρική ανάλυση στην κατεύθυνση Ανατολή-Δύση και η τιμή διαδρομής (track) αντιπροσωπεύει τη χωρική ανάλυση Βορρά-Νότου της σάρωσης αντίστοιχα. Τα εικονοστοιχεία δεν είναι πάντα 1 km όπως αναφέρθηκε παραπάνω, αλλά οι πραγματικές τιμές του μεγέθους του εικονοστοιχείου είναι αυτές που εμφανίζονται στις τιμές των scan και track.

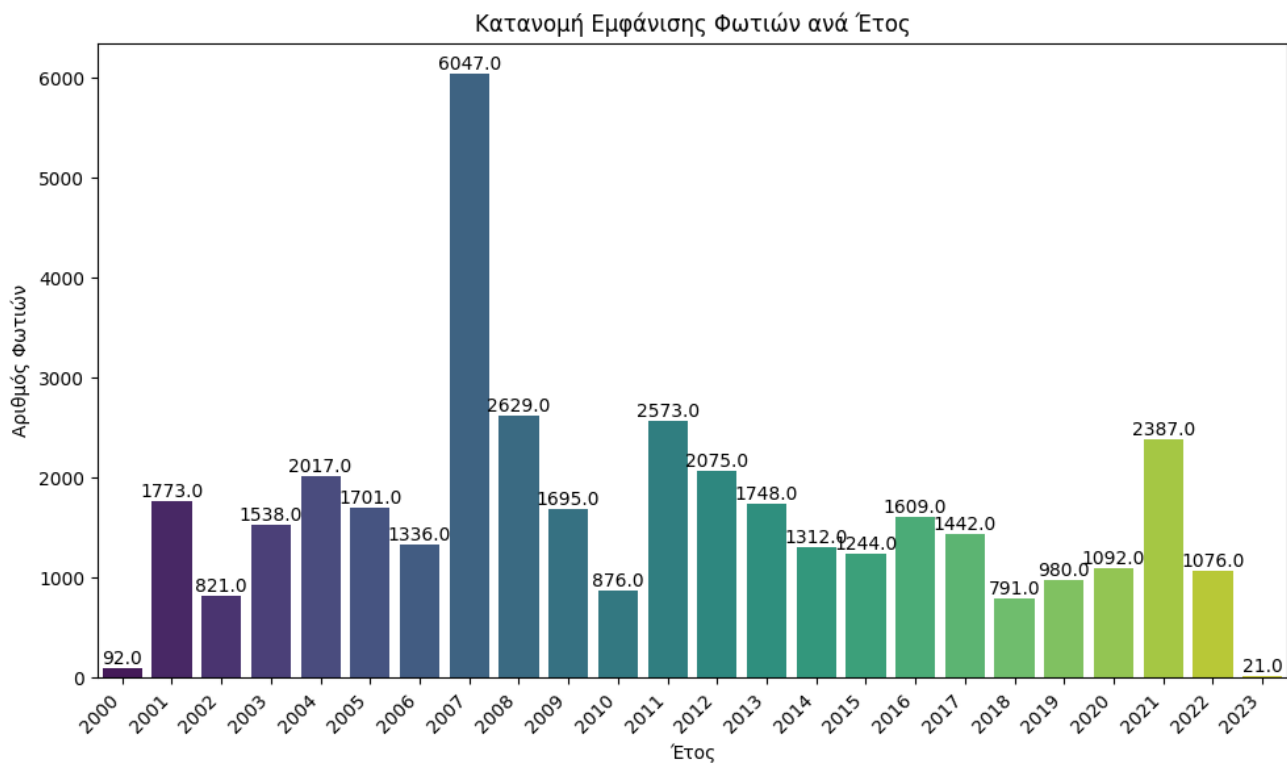
3.2 Διερευνητική ανάλυση δεδομένων

Το παρακάτω γράφημα παρουσιάζει τον αριθμό των φωτιών ανά μήνα, προσφέροντας μια οπτική αναπαράσταση της συχνότητας εμφάνισης των πυρκαγιών καθ' όλη την διάρκεια του χρονικού εύρους της εργασίας. Από το γράφημα παρατηρούμε ότι οι μήνες από Ιούλιο έως και Οκτώβριο καταγράφουν υψηλότερο αριθμό φωτιών, γεγονός που πιθανόν να οφείλεται στις υψηλότερες θερμοκρασίες και τις ξηρασίες της θερινής περιόδου.



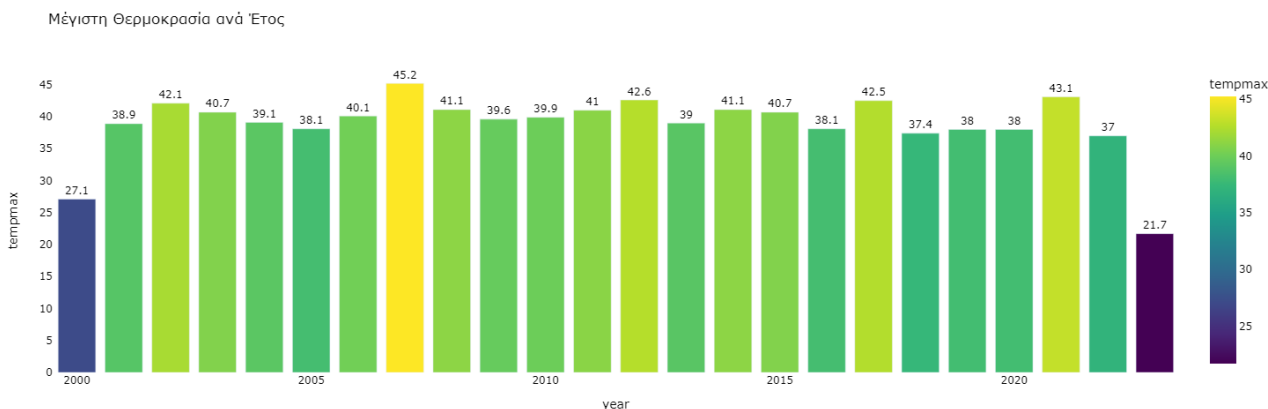
Εικόνα 3 Συχνότητα Πυρκαγιών στην Ελλάδα ανά Μήνα

Στη συνέχεια παρουσιάζεται η κατανομή των πυρκαγιών ανά έτος από το 2000 – 2023, επιτρέποντας την παρατήρηση των τάσεων και των μεταβολών στον αριθμό των πυρκαγιών κάθε χρόνο. Το έτος 2007 ξεχωρίζει με τον μεγαλύτερο αριθμό φωτιών, φτάνοντας τις 6.047, πιθανόν λόγω ακραίων καιρικών συνθηκών ή άλλων εξαιρετικά κρίσιμων γεγονότων. Μετά το 2007, παρατηρείται μια γενική τάση μείωσης των φωτιών μέχρι και το 2020.



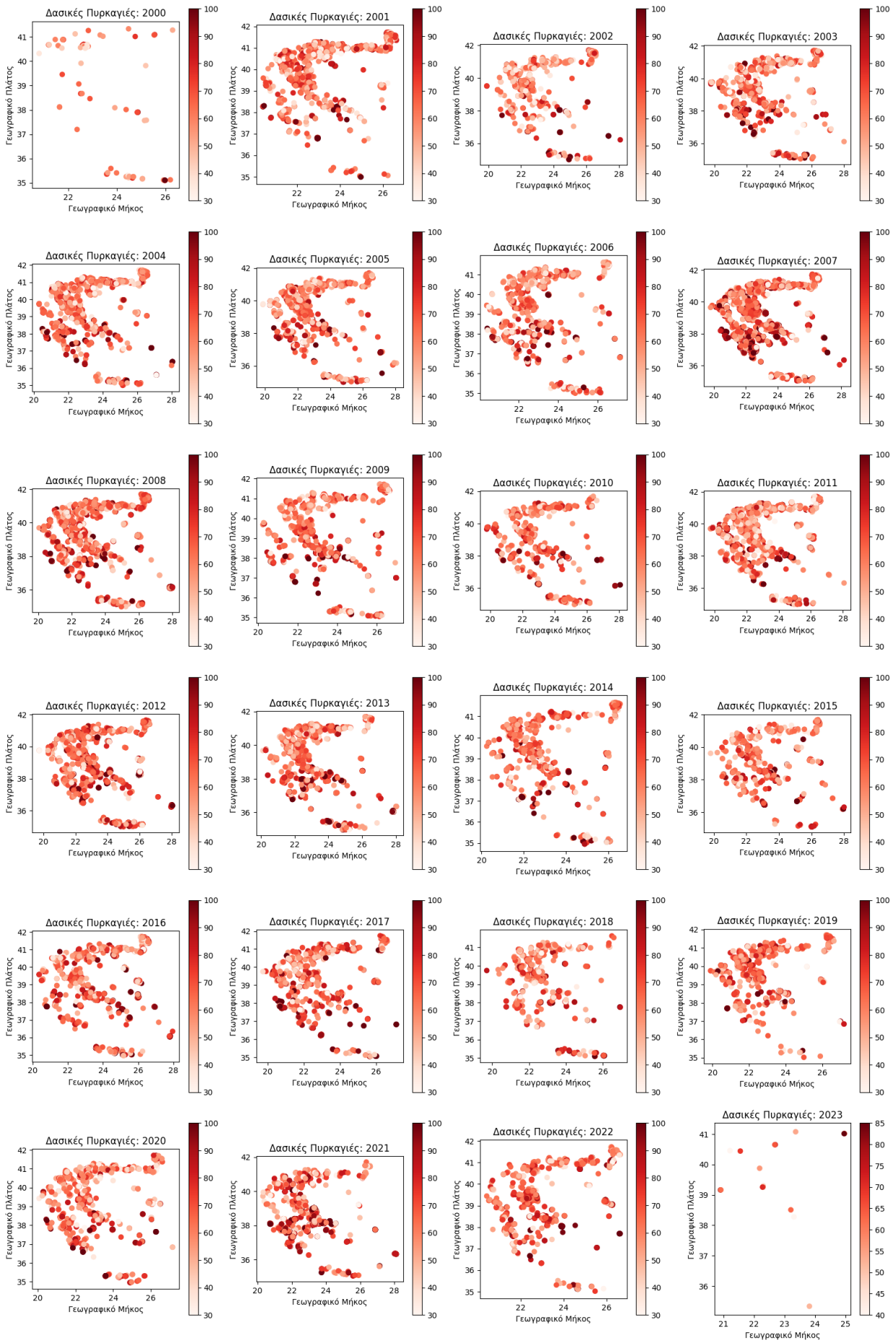
Εικόνα 4 Κατανομή Πυρκαγιών στην Ελλάδα 2000-2023

Στο παρακάτω γράφημα παρουσιάζεται η μέγιστη θερμοκρασία κάθε έτους από το 2000 – 2023. Οι θερμοκρασίες αναγράφονται πάνω από τις ράβδους και η χρωματική κλίμακα προσδιορίζει τις τιμές της μέγιστης θερμοκρασίας, από το κίτρινο (υψηλότερες θερμοκρασίες) έως το μωβ (χαμηλότερες θερμοκρασίες). Παρατηρείται γενικά μια τάση προς υψηλότερες μέγιστες θερμοκρασίες σε πολλά έτη και μάλιστα το έτος 2007 που σημειώθηκαν και οι περισσότερες πυρκαγιές στην Ελλάδα.



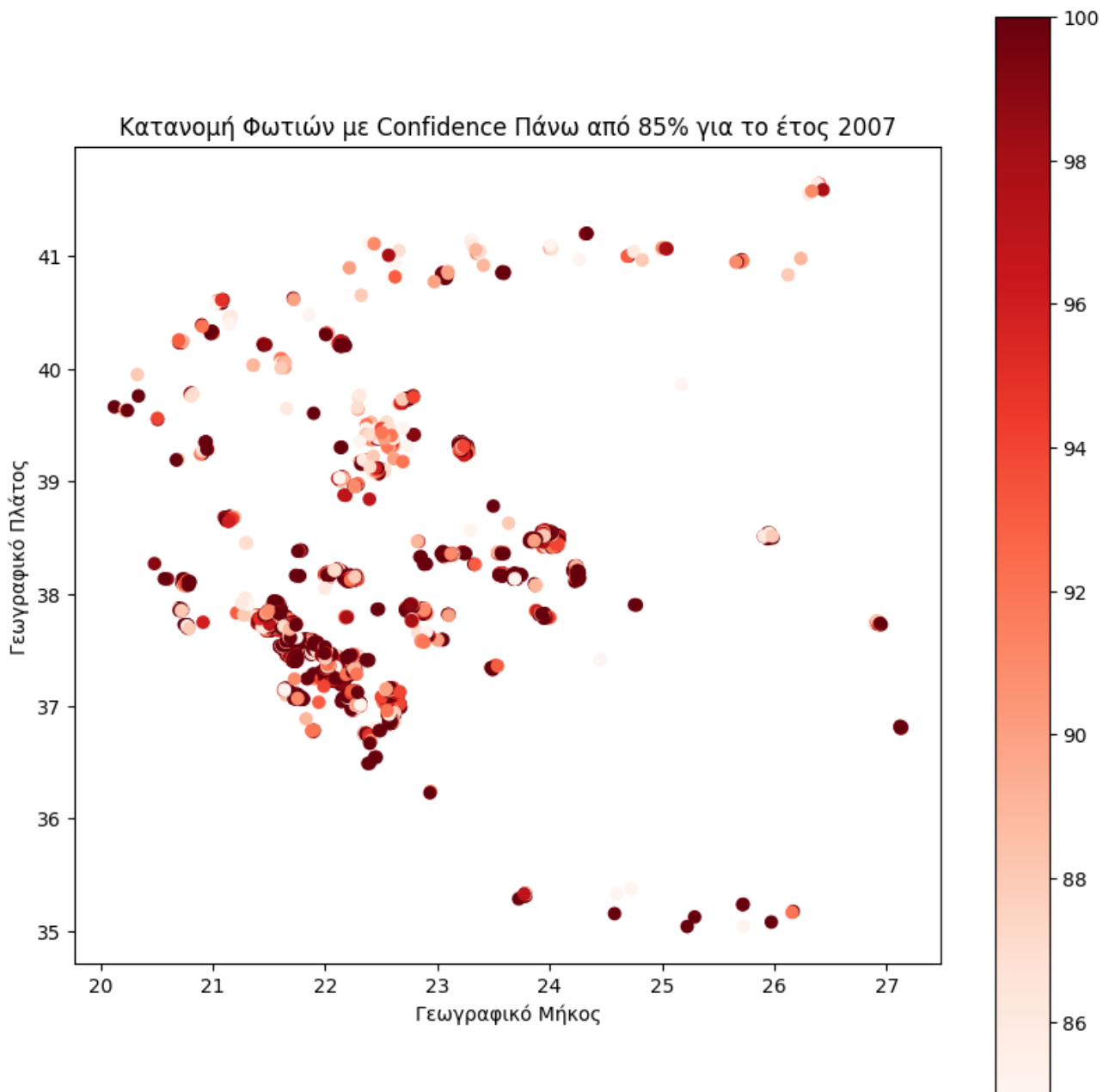
Εικόνα 5 Μέγιστη Θερμοκρασία ανά έτος

Το παρακάτω γράφημα παρουσιάζει την κατανομή των δασικών πυρκαγιών από το 2000 έως το 2023. Οι πυρκαγιές απεικονίζονται με σημεία που αντιπροσωπεύουν τις συντεταγμένες, ενώ η ένταση της χρωματικής κλίμακας υποδεικνύει τον αριθμό των πυρκαγιών σε κάθε περιοχή.



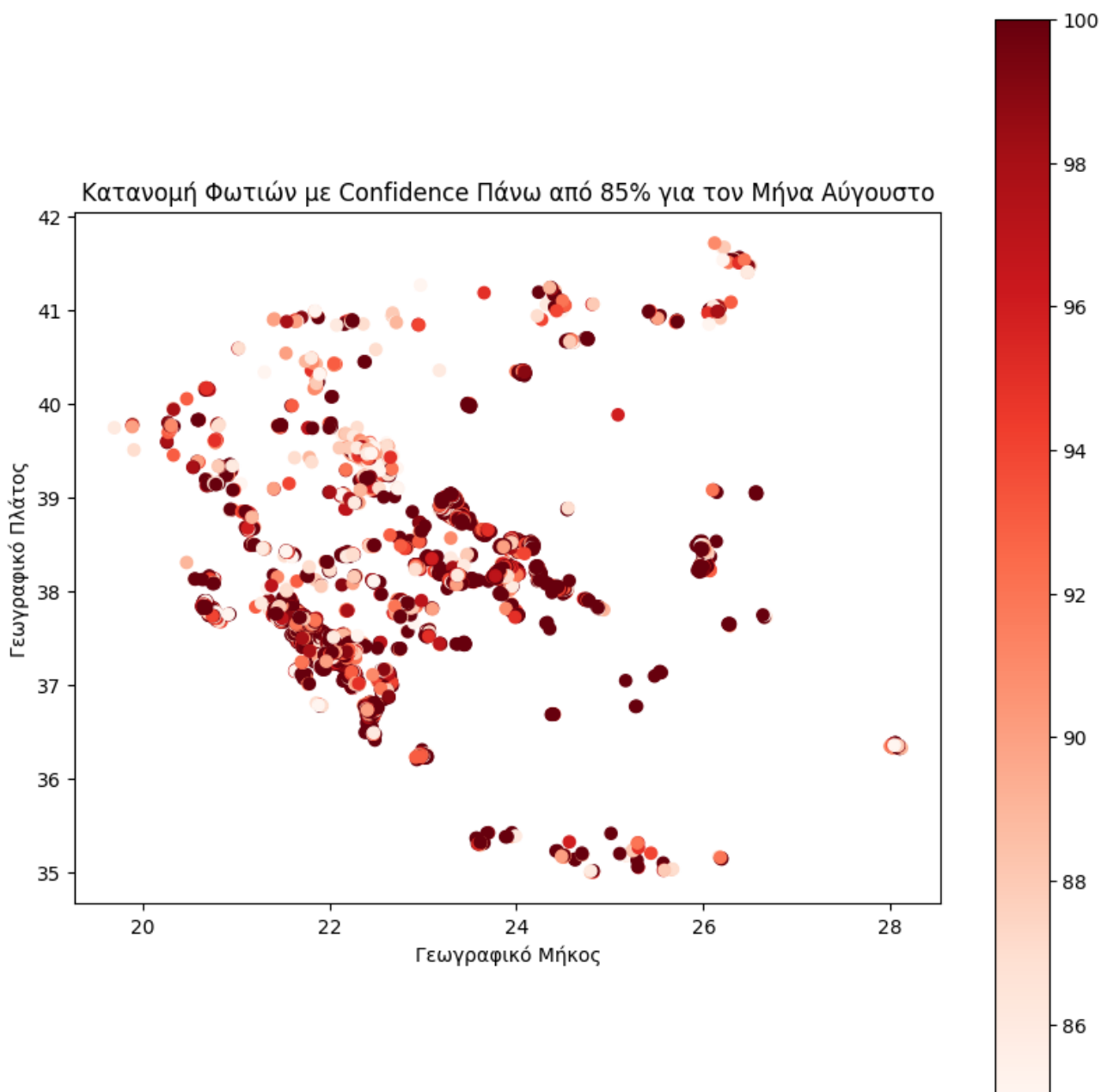
Εικόνα 6 Κατανομή Δασικών Πυρκαγιών 2000-2023 στην Ελλάδα

Οι πυρκαγιές παρουσιάζουν διαφοροποιήσεις ως προς την τοποθεσία με κάποιες περιοχές να εμφανίζουν συχνότερα πυρκαγιές. Κάποιες χρονιές, όπως το 2007 παρατηρείται μεγαλύτερη πυκνότητα πυρκαγιών και όπως παρατηρούμε και στο επόμενο γράφημα, οι πυρκαγιές μεγάλης επικινδυνότητας που εκδηλώθηκαν ήταν εξίσου πολυάριθμες και διάσπαρτες στον ελλαδικό χώρο και κυρίως στην Πελοπόννησο, Εύβοια και Αττική.



Εικόνα 7 Κατανομή Δασικών Πυρκαγιών στην Ελλάδα το 2007

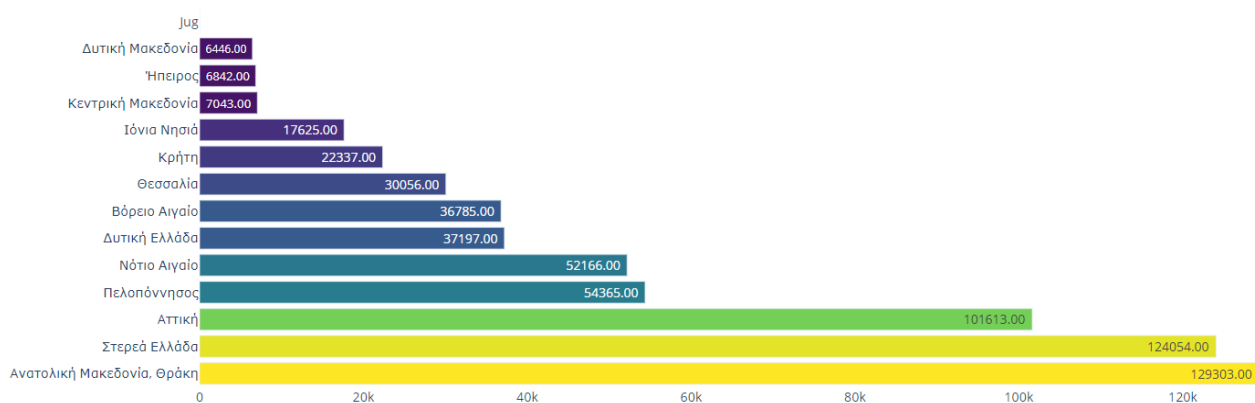
Εν συνεχεία, αφού εξετάστηκε η χωρική κατανομή των πυρκαγιών για όλα τα έτη, καθώς και για το έτος 2007, εστίασαμε περαιτέρω την κατανομή των πυρκαγιών για τον μήνα Αύγουστο, ο οποίος φαίνεται ότι είναι ο μήνας με τις περισσότερες πυρκαγιές. Το παρακάτω γράφημα παρουσιάζει την χωρική κατανομή των πυρκαγιών με ποσοστό εμπιστοσύνης άνω του 85%. Η ανάλυση αυτή μας επιτρέπει να δούμε με ακρίβεια τις περιοχές που επηρεάστηκαν περισσότερο από τις πυρκαγιές κατά τον μήνα Αύγουστο. Παρατηρήθηκε λοιπόν ότι κατά τον μήνα Αύγουστο παρουσιάστηκε υψηλή συγκέντρωση πυρκαγιών στις περιοχές με γεωγραφικό πλάτος μεταξύ 37 και 41 βαθμών και γεωγραφικού μήκους μεταξύ 22 και 26 βαθμών.



Εικόνα 8 Κατανομή Πυρκαγιών για τον μήνα Αύγουστο

Επίσης ιδιαίτερο ενδιαφέρον έχουν και οι περιοχές της Ελλάδας που έχουν υποστεί μεγάλη απώλεια εδαφών καθώς η ύπαρξη φωτιάς έχει μειώσει κατά πολύ την επιφάνεια του εδάφους. Στη συνέχεια παρουσιάζεται η συνολική έκταση που κάηκε ανά Περιφέρεια από το 2000 έως το 2023, η συνολική έκταση ανά έτος καθώς και η συνολική καμένη έκταση ανά Περιφέρεια. Οι Περιφέρειες είναι ταξινομημένες κατά αύξουσα σειρά καμένης έκτασης, ξεκινώντας από την Περιφέρεια με την μικρότερη ζημιά και καταλήγοντας σε εκείνη με την μεγαλύτερη.

Συνολική έκταση που κάηκε ανά Περιοχή



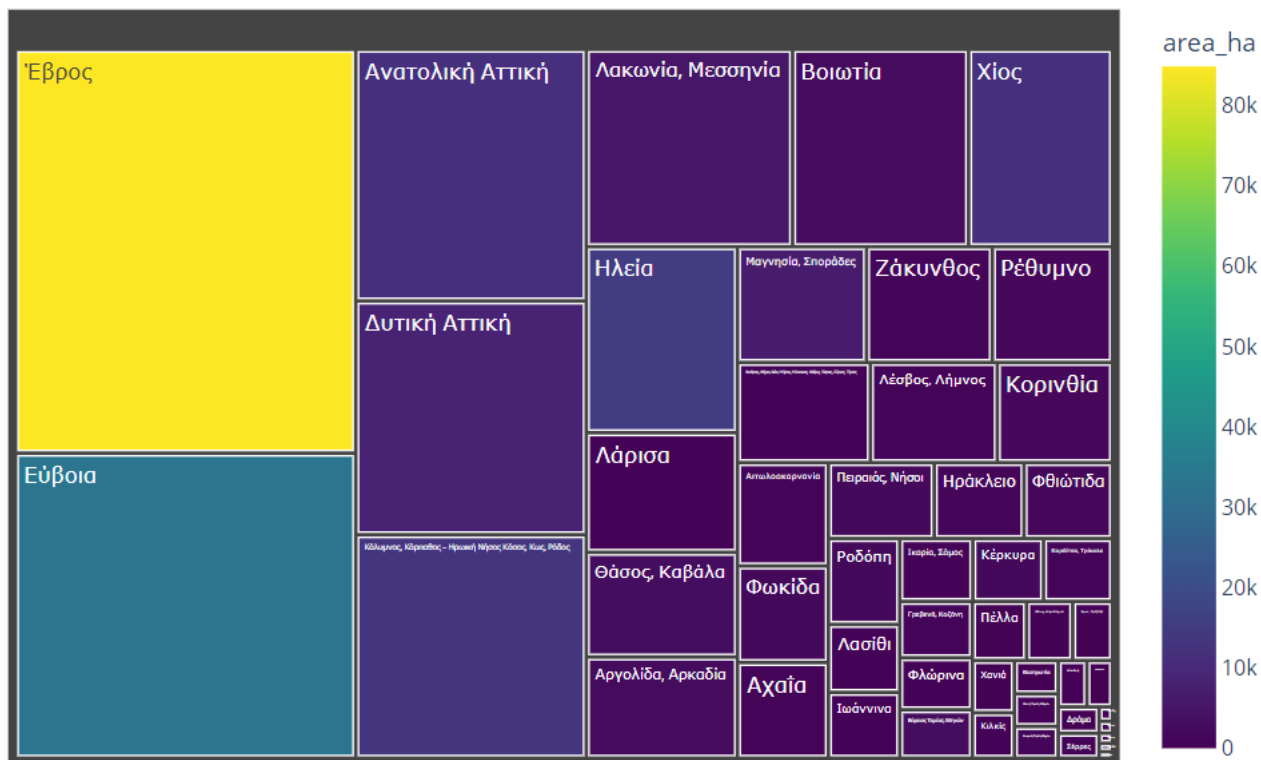
Εικόνα 9 Συνολική έκταση που κάηκε ανά Περιοχή [2000 - 2024]



Εικόνα 10 Εκτάσεις καμένων περιοχών (2008-2024)

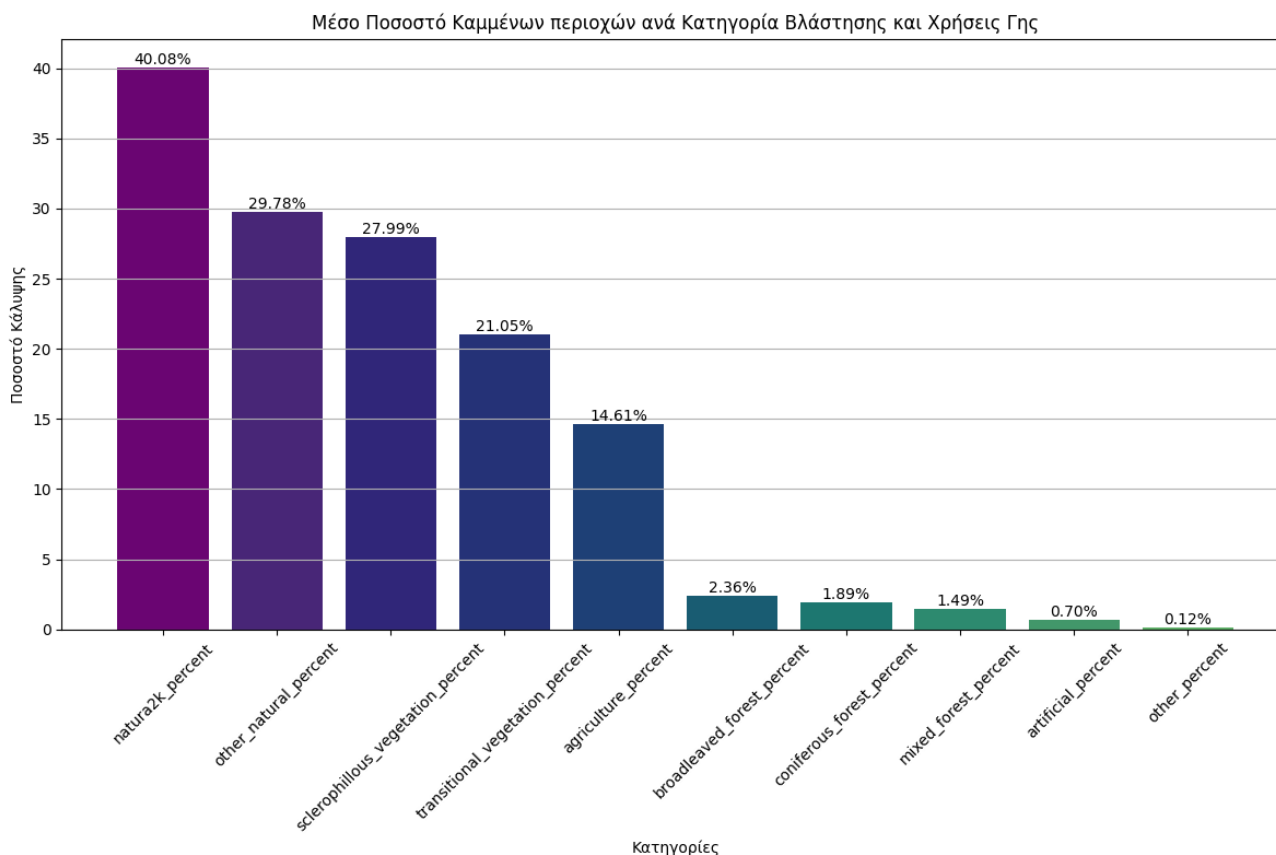
Στο επόμενο γράφημα παρουσιάζονται περιοχές της χώρας που έχουν πληγεί από φωτιές καθώς και η έκταση των καμένων γαιών σε κάθε περιοχή. Το μέγεθος και το χρώμα του κάθε μπλοκ

αντιπροσωπεύουν την έκταση της καμένης γης, με τις πιο φωτεινές αποχρώσεις να δείχνουν μεγαλύτερες εκτάσεις. Από το γράφημα προκύπτει ότι ο Έβρος και η Εύβοια είναι οι περιοχές με την μεγαλύτερη καμένη έκταση γης, ξεπερνώντας τα 80.0000 και 70.0000 εκτάρια αντίστοιχα. Άλλες περιοχές με σημαντικά καμένες εκτάσεις περιλαμβάνουν την Ανατολική και Δυτική Αττική καθώς και την Ηλεία.



Εικόνα 11 Περιοχές που έχουν πληγεί από Πυρκαγιές στην Ελλάδα

Τέλος, με βάση τις καμένες περιοχές παρουσιάζεται στο παρακάτω γράφημα το μέσο ποσοστό καμένων περιοχών ανά κατηγορία βλάστησης και χρήσης γης. Αυτό το γράφημα μας βοηθά να κατανοήσουμε ποιοι τύποι βλάστησης και χρήσεων γης επηρεάστηκαν περισσότερο από τις πυρκαγιές. Από την ακόλουθη ανάλυση προκύπτει ότι οι περιοχές Natura 2000 είναι αυτές που έχουν επηρεαστεί περισσότερο με μέσο ποσοστό καμένης γης 40,08%.



Εικόνα 12 Ποσοστό καμμένων περιοχών ανά κατηγορία βλάστησης.

3.3 Επεξεργασία Δεδομένων

Η προεπεξεργασία των δεδομένων αποτελεί κρίσιμο βήμα στη διαδικασία της ανάπτυξης μοντέλων μηχανικής μάθησης καθώς επιτρέπει την απομάκρυνση ανωμαλιών, την αντιμετώπιση ελλειπουσών τιμών και τη μετατροπή δεδομένων σε μια μορφή που είναι κατάλληλη για την εκπαίδευση των μοντέλων μηχανικής μάθησης.

Αρχικά, πραγματοποιήθηκε η αντικατάσταση των τιμών NaN (Not a Number) με κατάλληλες τιμές σε συγκεκριμένες στήλες. Για παράδειγμα, οι στήλες “winddir” και “snowdepth” αντικαταστάθηκαν με την τιμή 0, ενώ η στήλη “preciptype” αντικαταστάθηκε με την τιμή “unknown” για τις απουσιάζουσες τιμές.

Στη συνέχεια, πραγματοποιήθηκε ο υπολογισμός της μέσης τιμής των στηλών “tempmax” και “tempmin”, καθώς και “windspeedmax” και “windspeedmin”. Οι τιμές NaN στις στήλες “temp” και “windspeed” αντικαταστάθηκαν αντίστοιχα με τις υπολογισμένες μέσες τιμές.

Επιπλέον, πραγματοποιήθηκε ο υπολογισμός του μέσου όρου των στηλών “visibility”, “dew” και “cloudcover” και οι απουσιάζουσες τιμές αντικαταστάθηκαν με τον υπολογισμένο μέσο όρο.

Επίσης, διαγράφηκαν οι γραμμές που περιείχαν NaN (Not a Number) στις στήλες “humidity”, “dew”, “precipcover” και “winddir”.

Τέλος, πραγματοποιήθηκε η μετατροπή των στηλών “daynight”, και “satellite” σε αριθμητικές τιμές, καθώς και η μετατροπή της στήλης “preciprype” σε αριθμητικές τιμές μέσω λεξικού αντιστοίχισης. Ενώ δημιουργήθηκαν νέες μεταβλητές για την αναπαράσταση δεδομένων χρονικής σειράς και αφαιρέθηκαν οι εστίες πυρκαγιάς από μη δασικές περιοχές, όπως υδάτινες περιοχές και ηφαίστεια.

Κεφάλαιο 4: Μέθοδοι μηχανικής μάθησης

Στο παρόν κεφάλαιο οι αλγόριθμοι που αξιολογούνται είναι το τυχαίο δάσος (Random Forest), η λογιστική παλινδρόμηση (Logistic Redression) καθώς και ο αλγόριθμος SVM (Support Vector Machine). Στη συνέχεια, παρουσιάζονται αναλυτικά τα βήματα που ακολουθήθηκαν για την επεξεργασία των δεδομένων, την επιλογή των κατάλληλων μεταβλητών για την εκπαίδευση του μοντέλου και η αξιολόγηση των αποτελεσμάτων των επιλεγμένων μοντέλων.

4.1 Τυχαίο Δάσος (Random Forest)

Ο αλγόριθμος μηχανικής μάθησης, Random Forest, είναι ένα μοντέλο συνόλου που αποτελείται από πολλά ξεχωριστά εκπαιδευμένα Δέντρα Αποφάσεων (Decision Trees) και είναι από τα πιο δημοφιλή bagged Decision Trees. Κάθε Δέντρο Απόφασης σε ένα Τυχαίο Δάσος λαμβάνει μια απόφαση ταξινόμησης, σύμφωνα με την οποία η κλάση με τον μέγιστο αριθμό ψήφων προσδιορίζεται ως η τελική ταξινόμηση για τα δεδομένα εισόδου. Η βασική αρχή του αλγορίθμου είναι ότι ένα τυχαίο υποσύνολο χαρακτηριστικών επιλέγεται σε κάθε κόμβο. Τα δείγματα για την εκπαίδευση των επιμέρους δέντρων επιλέγονται με τη χρήση bagging, το οποίο επαναλαμβάνει τα δείγματα με αντικατάσταση του αρχικού συνόλου των σημείων. Η υψηλή απόδοση αυτού του αλγορίθμου επιτυγχάνεται με την ελαχιστοποίηση της συσχέτισης μεταξύ των δέντρων, μειώνοντας τη διακύμανση

του μοντέλου, έτσι ώστε ένας μεγάλος αριθμός διαφορετικών δέντρων να παρέχει μεγαλύτερη ακρίβεια από ό,τι τα μεμονωμένα δέντρα. Ωστόσο, η βελτίωση του μοντέλου ως προς αυτό τον τρόπο αυξάνει τη μεροληψία και την απώλεια της ερμηνευσιμότητας (explainability) [25].

Το πρώτο παράδειγμα για την πρόβλεψη δασικής πυρκαγιάς πραγματοποιήθηκε για τον ταξινομητή Random Forest, χρησιμοποιώντας την βιβλιοθήκη scikit-learn.

Διαχωρισμός Δεδομένων

Αρχικά, διαχωρίστηκαν τα δεδομένα σε τρία σύνολα, εκπαίδευσης, επικύρωσης και ελέγχου χρησιμοποιώντας τη συνάρτηση `'train_test_split'`. Ορίστηκε το μέγεθος του συνόλου δοκιμής στο 20% του συνόλου και το μέγεθος επικύρωσης στο 20% του συνόλου εκπαίδευσης.

Εκπαίδευση μοντέλου

Εκπαιδεύτηκε ο ταξινομητής `'RandomForestClassifier'` στο σύνολο εκπαίδευσης χρησιμοποιώντας τη μέθοδο `fit`.

Εκτίμηση με Cross-Validation

Αξιολογήθηκε η απόδοση του μοντέλου χρησιμοποιώντας `cross-validation` στο σύνολο επικύρωσης. Χρησιμοποιήθηκε η συνάρτηση `'cross_val_score'` με `'5-fold cross-validation'` και υπολογίστηκε το μέσο ποσοστό επιτυχίας των εκτιμήσεων.

Πρόβλεψη και Αξιολόγηση

Προβλέφθηκαν οι ετικέτες συνόλου δοκιμής χρησιμοποιώντας το εκπαιδευμένο μοντέλο και αξιολογήθηκε η ακρίβεια του μοντέλου χρησιμοποιώντας τη μέθοδο `'accuracy_score'`. Επιπλέον, παρουσιάστηκε η αναφορά ταξινόμησης χρησιμοποιώντας τη συνάρτηση `'classification_report'`, παρέχοντας λεπτομερείς μετρικές όπως `precision`, `recall` και `F1-score` για κάθε κλάση.

4.2. Support Vector Machine

Μια άλλη κατηγορία `supervised learning` περιλαμβάνει τον αλγόριθμο μηχανικής μάθησης Support Vector Machine (SVM), Μηχανές Διανυσμάτων Υποστήριξης και τις συναφείς μεθόδους `kernel-based` [26]. Ο SVM είναι ένας ταξινομητής που προσδιορίζει το όριο απόφασης σε έναν χώρο n διαστάσεων που διαχωρίζει το όριο κάθε κλάσης, για δεδομένα σε n διαστάσεις. Η μέθοδος αυτή βρίσκει το βέλτιστο όριο με τέτοιο τρόπο ώστε η απόφαση μεταξύ των πλησιέστερων σημείων κάθε κλάσης και του ορίου απόφασης να είναι το μέγιστο. Εάν τα δεδομένα μπορούν να διαχωριστούν σε μια γραμμή, τότε το όριο απόφασης ορίζεται ως εξής:

$$w^T x + b = 0$$

Όπου, w : διάνυσμα βάρους, x : διάνυσμα εισόδου, b : μεροληψία.

Η απόσταση του ορίου απόφασης από το πλησιέστερο σημείο δεδομένων d , ονομάζεται διάνυσμα υποστήριξης και ορίζεται ως περιθώριο διαχωρισμού. Ο στόχος, είναι να βρεθεί το βέλτιστο όριο απόφασης που ελαχιστοποιεί το περιθώριο διαχωρισμού. Εάν δεν είναι γραμμικά διαχωρισμένα εφαρμόζεται ένα σύνολο μετασχηματισμών στα δεδομένα σε έναν υψηλότερης διάστασης χώρο, όπου η εύρεση αυτού του ορίου απόφασης είναι ευκολότερο να υπολογιστεί.

Εισαγωγή Βιβλιοθηκών

Το παρόν παράδειγμα ξεκίνησε εισάγοντας τις απαραίτητες βιβλιοθήκες, συμπεριλαμβανομένης της `sklearn.svm` για την εκπαίδευση του μοντέλου SVM. Ο διαχωρισμός των δεδομένων είναι ακριβώς ο ίδιος με το προηγούμενο παράδειγμα.

Αρχικοποίηση SVM

Δημιουργήθηκε ένα αντικείμενο ταξινομητή SVM χρησιμοποιώντας την κλάση `SCV()`.

Εκπαίδευση Ταξινομητή

Εκπαιδεύτηκε ο ταξινομητής SVM στο σύνολο δεδομένων εκπαίδευσης (`X_train`, `y_train`) χρησιμοποιώντας τη μέθοδο `fit()`.

Cross-Validation

Εκτελέστηκε 5-fold cross-validation στα δεδομένα επικύρωσης (`X_val`, `y_val`) για να εκτιμηθεί η απόδοση του μοντέλου μας. Αυτό επιτεύχθηκε με τη χρήση της μεθόδου `cross_val_score()`.

Αξιολόγηση στο Test Set

Αξιολογήθηκε η επίδοση του μοντέλου στα ανεξάρτητα δεδομένα ελέγχου (`X_test`, `y_test`) με τη μέθοδο `score`.

Πρόβλεψη & Αξιολόγηση

Τέλος, προβλέφθηκαν οι ετικέτες για τα δεδομένα δοκιμής και υπολογίστηκε και παρουσιάστηκε η αναφορά ταξινόμησης για την αξιολόγηση της απόδοσης του μοντέλου.

4.3 Logistic Regression

Τέλος, εφαρμόστηκε η Logistic Regression για την κατηγοριοποίηση των δεδομένων μας.

Εκπαίδευση Μοντέλου

Αρχικοποιήθηκε ο ταξινομητής Logistic Regression και εκπαιδεύτηκε στο σύνολο δεδομένων εκπαίδευσης που έχουν εκπαιδευτεί και τα δύο προηγούμενα μοντέλα.

Cross-Validation

Για να αξιολογηθεί η απόδοση του ταξινομητή και να αποφευχθεί η υπερεκπαίδευση, χρησιμοποιήσαμε την μέθοδο του Cross-Validation. Συγκεκριμένα, και εδώ εκτελέστηκε 5-fold cross-validation στα δεδομένα επικύρωσης με τη χρήση της συνάρτησης ‘cross_val_score’, η οποία επιστρέφει μια λίστα με τις βαθμολογίες ακρίβειας σε κάθε διαίρεση.

Αξιολόγηση Απόδοσης

Για την αξιολόγηση της απόδοσης του μοντέλου μας στα δεδομένα ελέγχου, χρησιμοποιήθηκε η μέθοδος ‘score’, η οποία επέστρεψε το ποσοστό των σωστών προβλέψεων.

Ανάλυση Αποτελεσμάτων

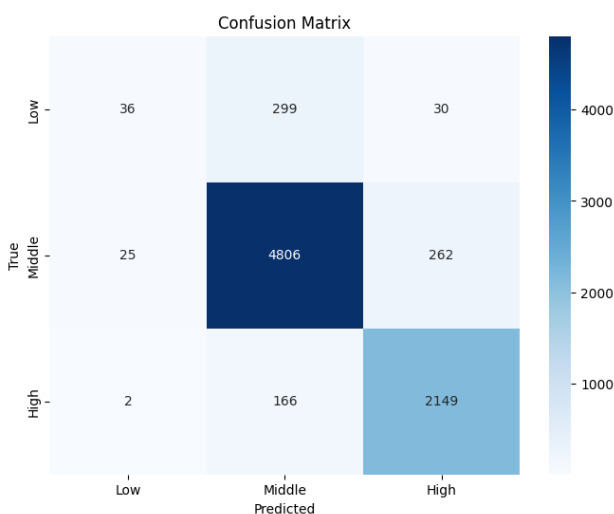
Τέλος, για μια λεπτομερή ανάλυση των αποτελεσμάτων χρησιμοποιήθηκε ο πίνακας σύγχυσης (confusion matrix) και η αναφορά ταξινόμησης (classification report).

Κεφάλαιο 5: Αποτελέσματα και ανάλυση

Στο παρόν κεφάλαιο παρουσιάζονται, αναλύονται και συγκρίνονται τα αποτελέσματα των αποδόσεων των αλγορίθμων που χρησιμοποιούνται.

5.1 Random Forest

Το μοντέλο παρουσιάζει υψηλή ακρίβεια στην πρόβλεψη των τριών κλάσεων (χαμηλή, μεσαία και υψηλή πιθανότητα πυρκαγιάς).



Εικόνα 13 Confusion Matrix για το μοντέλο Random Forest

Ακρίβεια (Accuracy)

Το μοντέλο παρουσιάζει υψηλή ακρίβεια στην πρόβλεψη των κλάσεων. Συγκεκριμένα, οι περισσότερες περιπτώσεις που ανήκουν στην μεσαία και υψηλή κλάση κατηγοριοποιούνται σωστά, όπως φαίνεται από τις υψηλές τιμές των διαγώνιων στοιχείων στον παραπάνω πίνακα (4806 για μεσαία πιθανότητα, 2149 για υψηλή πιθανότητα).

Ευαισθησία (Recall)

Για την κλάση “Χαμηλή Πιθανότητα πυρκαγιάς” η ευαισθησία είναι χαμηλή καθώς μόνο 36 από τις 365 περιπτώσεις προβλέπονται σωστά. Για την κλάση “Μέτρια Πιθανότητα Πυρκαγιάς” η ευαισθησία είναι πολύ υψηλή, με 4.806 σωστές προβλέψεις από τις 5.093. Ενώ για την κλάση “Υψηλή Πιθανότητα Πυρκαγιάς” η ευαισθησία είναι επίσης υψηλή, με 2.149 σωστές προβλέψεις από τις 2.317.

Ακρίβεια Κατηγοριών (Precision)

Για την κλάση με “Χαμηλή Πιθανότητα πυρκαγιάς” η ακρίβεια είναι χαμηλή καθώς από τις 63 συνολικές προβλέψεις μόνο οι 36 είναι σωστές. Για την κλάση “Μέτρια Πιθανότητα Πυρκαγιάς” η ακρίβεια είναι πολύ υψηλή, καθώς από τις 5.217 συνολικές προβλέψεις, οι 4.806 είναι σωστές. Ενώ για την κλάση “Υψηλή Πιθανότητα Πυρκαγιάς” η ακρίβεια είναι επίσης υψηλή, με 2.149 σωστές προβλέψεις από τις 2.441.

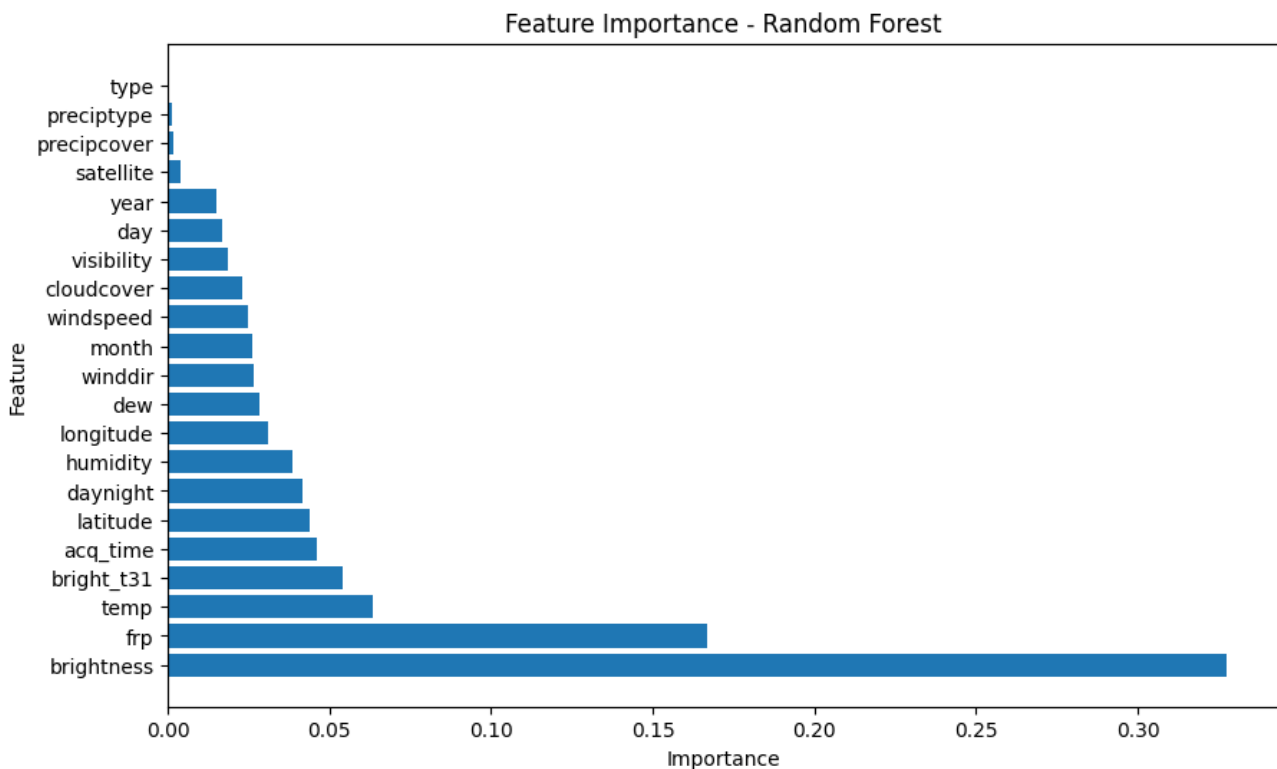
Το μοντέλο δείχνει εξαιρετική απόδοση στις προβλέψεις μεσαίας και υψηλής πιθανότητας ύπαρξης δασικών πυρκαγιών, με υψηλές τιμές ευαισθησίας και ακρίβειας σε αυτές τις κατηγορίες. Ωστόσο, υπάρχει σημαντικό περιθώριο βελτίωσης στην πρόβλεψη της χαμηλής πιθανότητας πυρκαγιάς, καθώς τόσο η ευαισθησία όσο και η ακρίβεια είναι χαμηλές σε αυτήν την κατηγορία.

Η παραπάνω ανάλυση υποδεικνύει ότι το μοντέλο είναι ιδιαίτερα χρήσιμο για την πρόβλεψη μεσαίων και υψηλών κινδύνων πυρκαγιάς, παρέχοντας αξιόπιστες εκτιμήσεις για τις κρίσιμες περιπτώσεις που απαιτούν άμεση προσοχή και δράση. Ωστόσο, απαιτείται περαιτέρω βελτιστοποίηση για να βελτιωθεί η ακρίβεια στην πρόβλεψη των χαμηλών κινδύνων, διασφαλίζοντας έτσι μια πιο ολοκληρωμένη και αποτελεσματική πρόγνωση.

	Ακρίβεια (Accuracy)	Ευαισθησία (Recall)	F1-score	Πλήθος Περιπτώσεων
Χαμηλή Πιθανότητα	0.57	0.10	0.17	365
Μέτρια Πιθανότητα	0.91	0.94	0.93	5093
Υψηλή Πιθανότητα	0.88	0.93	0.90	2317

Πίνακας 2 Πίνακας Σύγχυσης (Confusion Matrix) - Random Forest

Παρακάτω παρουσιάζεται η σημασία των χαρακτηριστικών όπως υπολογίστηκε από το μοντέλο Random Forest. Η σημασία των χαρακτηριστικών δείχνει πόσο συμβάλλει κάθε χαρακτηριστικό στην πρόβλεψη του στόχου του μοντέλου. Πιο αναλυτικά, παρατηρούμε ότι το χαρακτηριστικό με τη μεγαλύτερη σημασία είναι το brightness και ακολουθεί το fir. Αυτά τα δύο χαρακτηριστικά έχουν σημαντικά μεγαλύτερη σημασία σε σχέση με τα υπόλοιπα, κάτι που υποδηλώνει ότι είναι κρίσιμα για την απόδοση του μοντέλου, καθώς σχετίζεται με την ένταση και τη θερμότητα των πυρκαγιών.



Εικόνα 14 Σημαντικότητα μεταβλητών στο μοντέλο Random Forest

Μετά τα δύο πρώτα χαρακτηριστικά η σημαντικότητα μειώνεται σταδιακά. Τα χαρακτηριστικά temp και bright_t31 έχουν επίσης κάποια σημαντικότητα αλλά σαφώς μικρότερη. Τέλος, τα χαρακτηριστικά type, preciptype και precipcover έχουν πολύ μικρή έως και μηδαμινή σημαντικότητα. Αυτό υποδεικνύει ότι τα χαρακτηριστικά αυτά δεν συνεισφέρουν στην πρόβλεψη στόχου του μοντέλου.

5.2 Support Vector Machine

Ακρίβεια (Accuracy)

Το μοντέλο παρουσιάζει υψηλή ακρίβεια στην πρόβλεψη των κλάσεων. Συγκεκριμένα, οι περισσότερες περιπτώσεις που ανήκουν στην μεσαία και υψηλή κλάση κατηγοριοποιούνται σωστά, όπως φαίνεται από τις υψηλές τιμές των διαγώνιων στοιχείων στον παρακάτω πίνακα (4675 για μεσαία πιθανότητα, 1624 για υψηλή πιθανότητα).

Ευαισθησία (Recall)

Για την κλάση “Χαμηλή Πιθανότητας πυρκαγιάς” δεν υπάρχουν σωστές προβλέψεις με αποτέλεσμα να είναι μηδέν το ποσοστό των πραγματικών θετικών περιπτώσεων. Για την κλάση “Μέτρια Πιθανότητα Πυρκαγιάς” η ευαισθησία είναι πολύ υψηλή, με 4.675 σωστές προβλέψεις από τις 5.093.

Ενώ για την κλάση “Υψηλή Πιθανότητα Πυρκαγιάς” η ευαισθησία είναι επίσης υψηλή, με 1.624 σωστές προβλέψεις από τις 2.317.

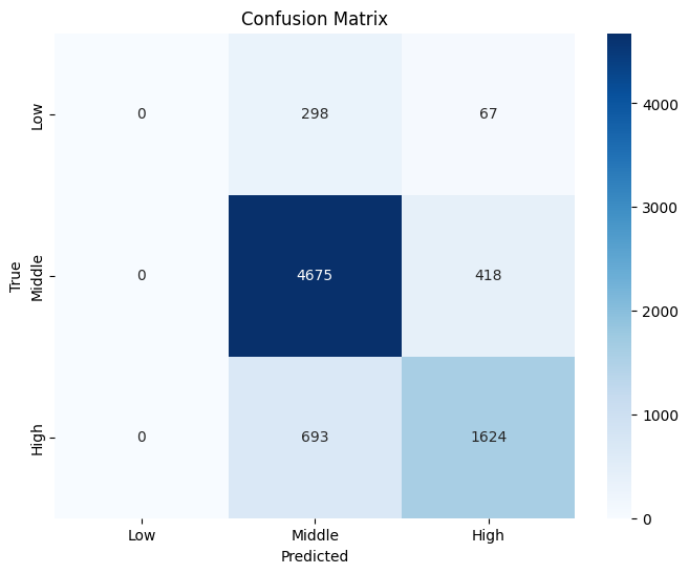
Ακρίβεια Κατηγοριών (Precision)

Για την κλάση με “Χαμηλή Πιθανότητα πυρκαγιάς” δεν υπάρχουν σωστές προβλέψεις. Για την κλάση “Μέτρια Πιθανότητα Πυρκαγιάς” η ακρίβεια είναι πολύ υψηλή, καθώς από τις 5.093 συνολικές προβλέψεις, οι 4.675 είναι σωστές. Ενώ για την κλάση “Υψηλή Πιθανότητα Πυρκαγιάς” η ακρίβεια είναι επίσης υψηλή, με 1624 σωστές προβλέψεις από τις 2.317.

Το μοντέλο εμφανίζει υψηλή ακρίβεια στην κατηγορία της μέτριας πιθανότητας ύπαρξης πυρκαγιάς, με 4.675 σωστές προβλέψεις. Ωστόσο, υπάρχει σημαντική απόκλιση στις κατηγορίες χαμηλής και υψηλής πιθανότητας. Ακόμα, δεν υπάρχει καμία σωστή πρόβλεψη για την χαμηλή πιθανότητα, ενώ αρκετές περιπτώσεις υψηλής πιθανότητας προβλέπονται ως μέτρια πιθανότητα πυρκαγιάς.

Η μετατόπιση των προβλέψεων από χαμηλή προς μεσαία και από υψηλή προς μεσαία πιθανότητα υποδηλώνει πιθανή δυσκολία του μοντέλου μας να διακρίνει τις ακραίες κατηγορίες, καταλήγοντας να ταξινομεί πολλές περιπτώσεις στην πιο πολυάριθμη κατηγορία την μεσαία.

Το μοντέλο μας χρειάζεται βελτίωση στη διάκριση των χαμηλών και υψηλών κατηγοριών πιθανοτήτων πυρκαγιάς. Οι προσπάθειες αυτές θα μπορούσαν να επικεντρωθούν στην επανεκπαίδευση του μοντέλου μας χρησιμοποιώντας περισσότερα χαρακτηριστικά ή στην εξισορρόπηση των κατηγοριών για την αντιμετώπιση της υπάρχουσας μεροληψίας προς τη μεσαία κατηγορία.

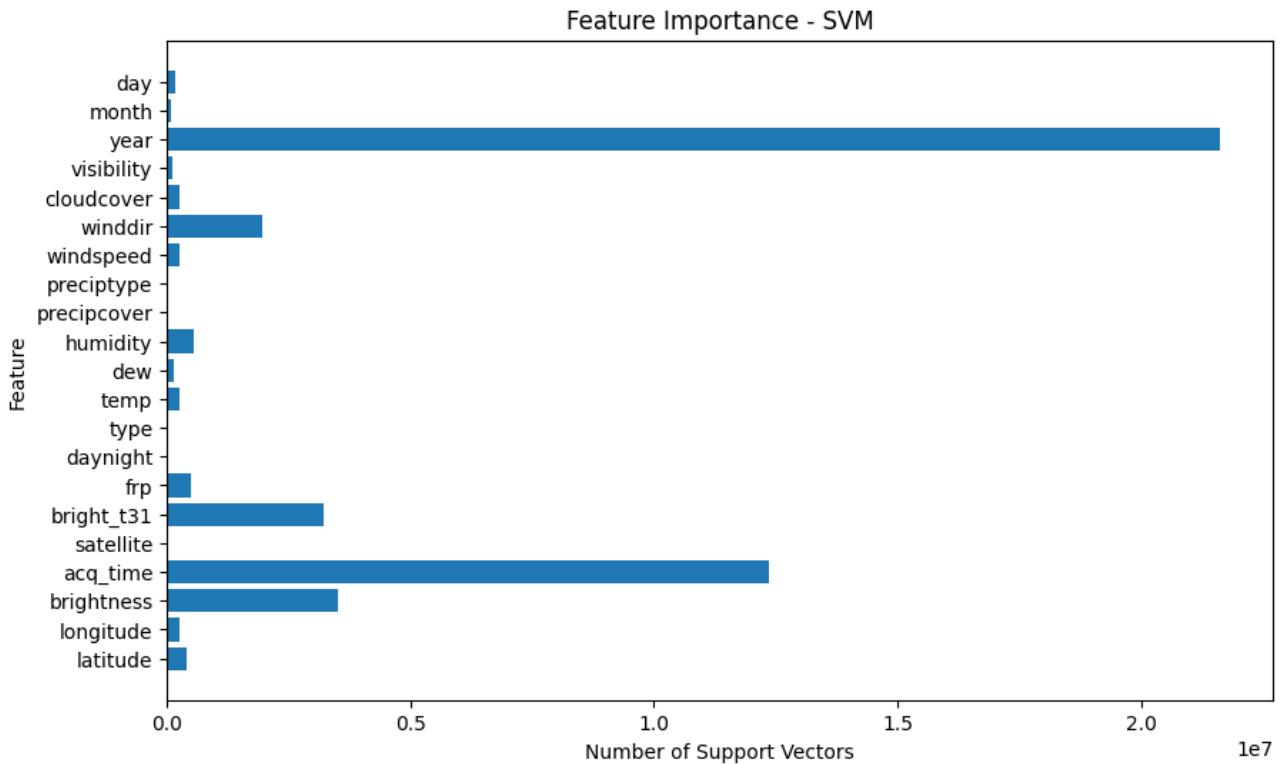


Εικόνα 15 Confusion Matrix για το μοντέλο SVM

	Ακρίβεια (Accuracy)	Ευαισθησία (Recall)	F1-score	Πλήθος Περιπτώσεων
Χαμηλή Πιθανότητα	0.00	0.00	0.00	365
Μέτρια Πιθανότητα	0.83	0.92	0.87	5093
Υψηλή Πιθανότητα	0.77	0.70	0.73	2317

Πίνακας 3 Πίνακας Σύγχυσης (Confusion Matrix) – SVM

Το παρακάτω γράφημα απεικονίζει τη σημαντικότητα των χαρακτηριστικών στο SVM μοντέλο μας βάσει των διανυσμάτων υποστήριξης (support vectors) που σχετίζεται με καθένα από αυτά. Από την ανάλυση του παρακάτω γραφήματος προκύπτει ότι η μεταβλητή *year* έχει την υψηλότερη σημαντικότητα, καθιστώντας την το πιο κρίσιμο στοιχείο για την απόδοση του μοντέλου. Ακολουθεί, η μεταβλητή *acq_time* με υψηλή σχετικά σημαντικότητα, ενώ οι μεταβλητές *brightness* και *bright_t31* συμβάλλουν σημαντικά. Αντίθετα, οι μεταβλητές *day*, *month*, *visibility* και άλλα εμφανίζουν πολύ μικρή έως και αμελητέα σημαντικότητα.



Εικόνα 16 Feature Importance – SVM

5.3 Logistic Regression

Ακρίβεια (Accuracy)

Το μοντέλο δεν παρουσιάζει υψηλή ακρίβεια στην πρόβλεψη όλων των κλάσεων. Συγκεκριμένα, οι περισσότερες περιπτώσεις που ανήκουν στην μεσαία και υψηλή κλάση κατηγοριοποιούνται σωστά, όπως φαίνεται από τις υψηλές τιμές των διαγώνιων στοιχείων στον παρακάτω πίνακα (4.772 για μεσαία πιθανότητα, 1.670 για υψηλή πιθανότητα).

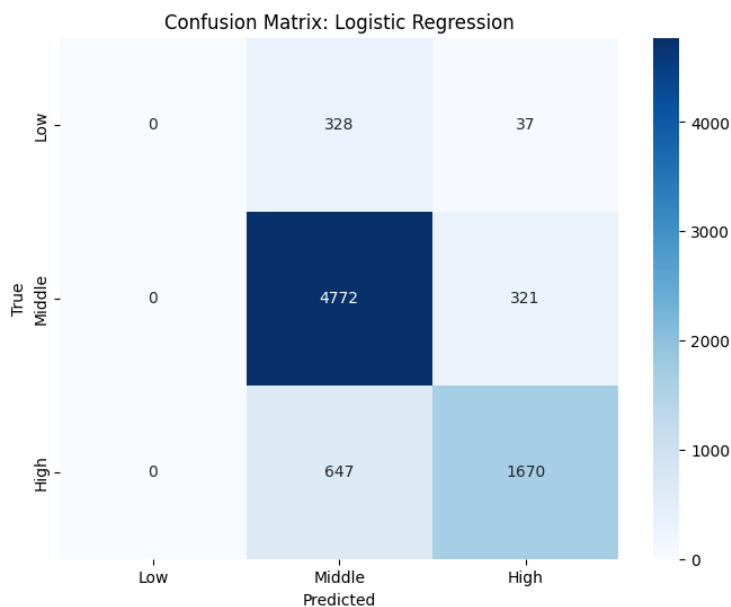
Ευαισθησία (Recall)

Για την κλάση “Χαμηλή Πιθανότητας πυρκαγιάς” δεν υπάρχουν σωστές προβλέψεις με αποτέλεσμα να είναι μηδέν το ποσοστό των πραγματικών θετικών περιπτώσεων. Για την κλάση “Μέτρια Πιθανότητα Πυρκαγιάς” η ευαισθησία είναι πολύ υψηλή, με 4.772 σωστές προβλέψεις από τις 5.093. Ενώ για την κλάση “Υψηλή Πιθανότητα Πυρκαγιάς” η ευαισθησία είναι επίσης υψηλή, με 1.670 σωστές προβλέψεις από τις 2.317.

Ακρίβεια Κατηγοριών (Precision)

Για την κλάση με “Χαμηλή Πιθανότητα πυρκαγιάς” δεν υπάρχουν σωστές προβλέψεις. Για την κλάση “Μέτρια Πιθανότητα Πυρκαγιάς” η ακρίβεια είναι πολύ υψηλή, καθώς από τις 5.093 συνολικές

προβλέψεις, οι 4.772 είναι σωστές. Ενώ για την κλάση “Υψηλή Πιθανότητα Πυρκαγιάς” η ακρίβεια είναι επίσης υψηλή, με 1.670 σωστές προβλέψεις από τις 2.317.



Εικόνα 17 Confusion Matrix για το μοντέλο Logistic Regression

	Ακρίβεια (Accuracy)	Ευαισθησία (Recall)	F1-score	Πλήθος Περιπτώσεων
Χαμηλή Πιθανότητα	0.00	0.00	0.00	365
Μέτρια Πιθανότητα	0.83	0.94	0.88	5093
Υψηλή Πιθανότητα	0.82	0.72	0.77	2317

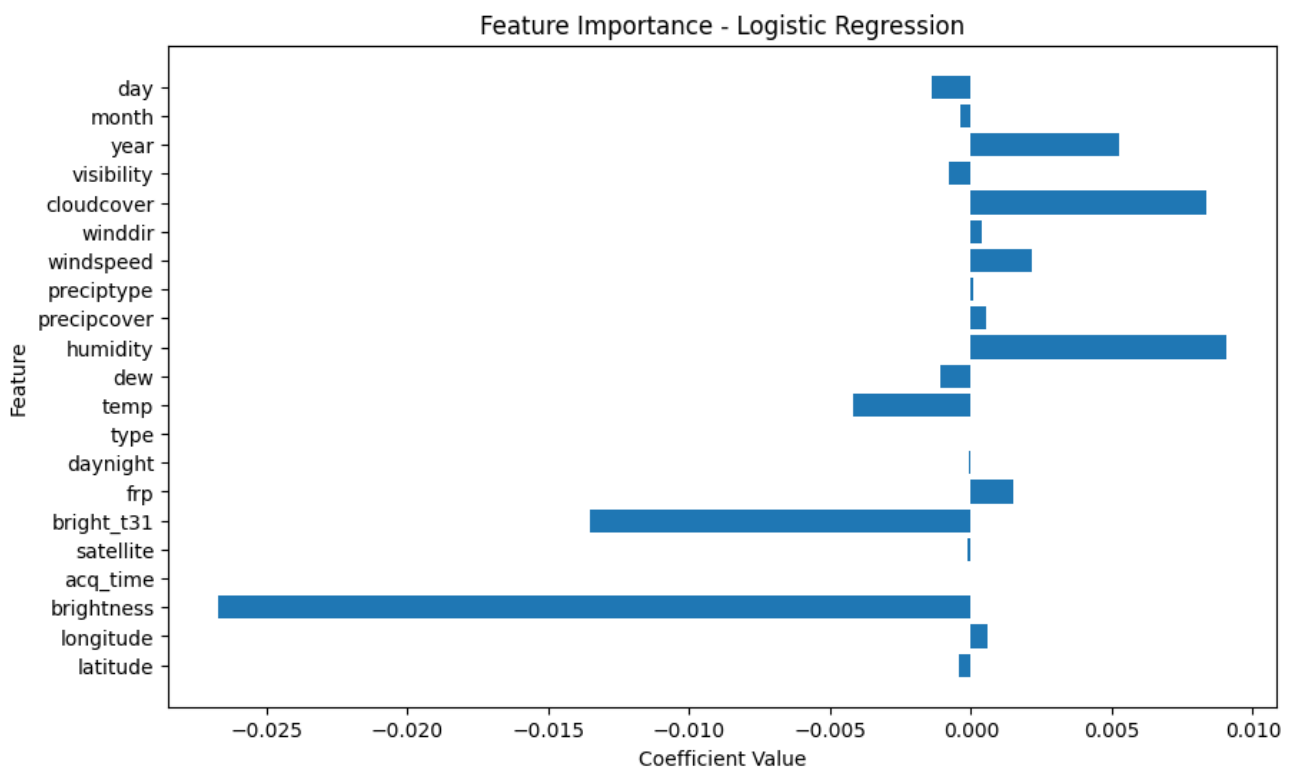
Πίνακας 4 Πίνακας Σύγχυσης (Confusion Matrix) - Logistic Regression

Το μοντέλο εμφανίζει υψηλή ακρίβεια στην κατηγορία της μέτριας πιθανότητας ύπαρξης πυρκαγιάς, με 4.772 σωστές προβλέψεις. Ωστόσο, υπάρχει σημαντική απόκλιση στις κατηγορίες χαμηλής και υψηλής πιθανότητας. Ακόμα, δεν υπάρχει καμία σωστή πρόβλεψη για την χαμηλή πιθανότητα, ενώ αρκετές περιπτώσεις υψηλής πιθανότητας προβλέπονται ως μέτρια πιθανότητα πυρκαγιάς.

Το μοντέλο μας καταλήγει να ταξινομεί πολλές περιπτώσεις στην πιο πολυάριθμη κατηγορία την μεσαία. Για αυτό το λόγο χρειάζεται βελτίωση στη διάκριση των χαμηλών και υψηλών κατηγοριών

πιθανοτήτων πυρκαγιάς, όπως συνέβη και στο μοντέλο του SVM. Οι προσπάθειες αυτές θα μπορούσαν να επικεντρωθούν στην επανεκπαίδευση του μοντέλου μας χρησιμοποιώντας περισσότερα χαρακτηριστικά ή στην εξισορρόπηση των κατηγοριών για την αντιμετώπιση της υπάρχουσας μεροληψίας προς τη μεσαία κατηγορία.

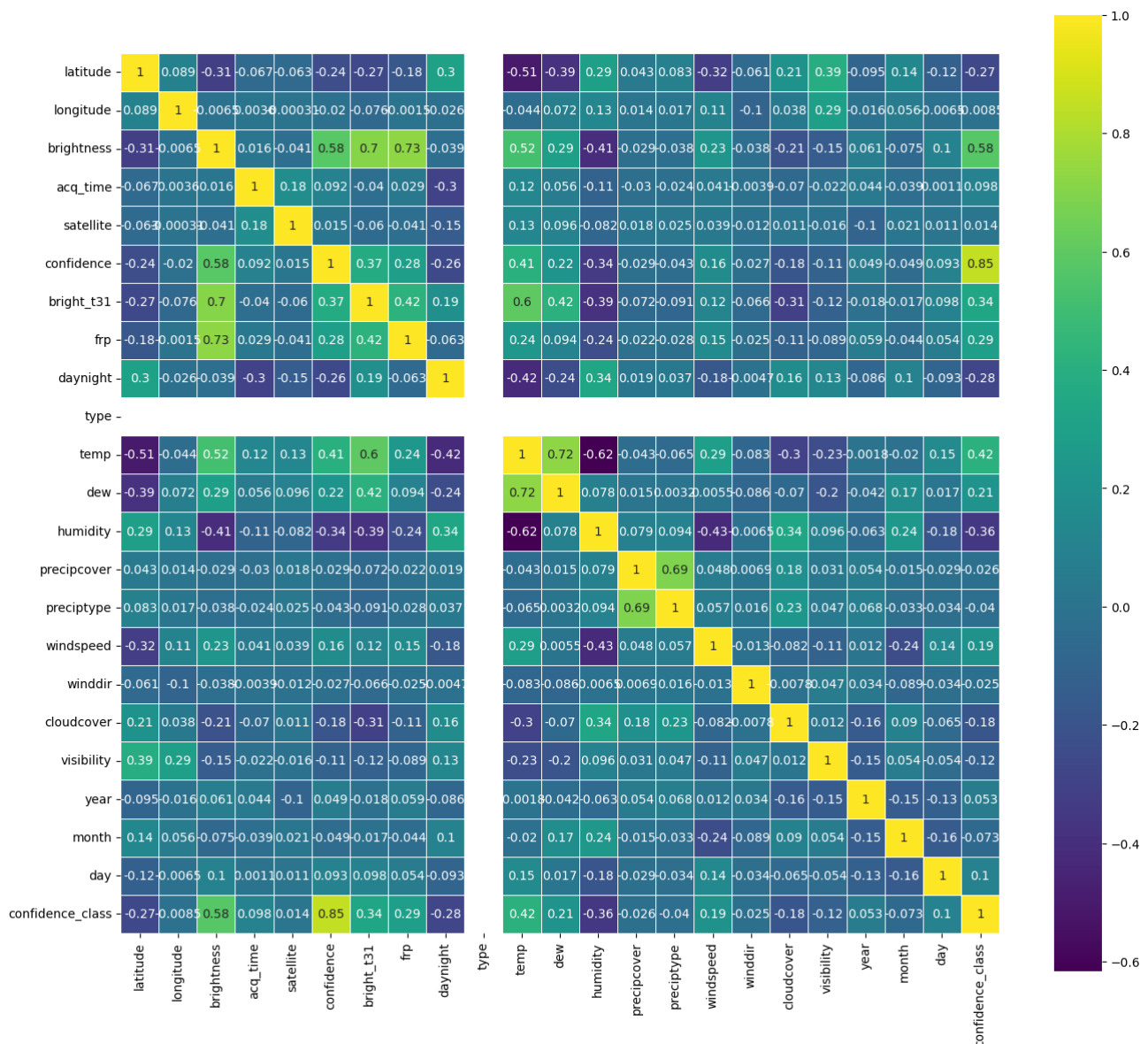
Στο παρακάτω γράφημα παρουσιάζεται η σημασία των χαρακτηριστικών του μοντέλου Logistic Regression. Το μοντέλο Logistic Regression χρησιμοποιεί τους συντελεστές των χαρακτηριστικών για να αναδείξει την επίδραση τους στην εξαρτημένη μεταβλητή. Οι θετικές τιμές συντελεστών υποδεικνύουν ότι η αύξηση του χαρακτηριστικού συνδέεται με την αύξηση της πιθανότητας θετικού αποτελέσματος, ενώ οι αρνητικές τιμές υποδεικνύουν μείωση αυτής της πιθανότητας. Με βάση την ανάλυση αυτή, τα χαρακτηριστικά με τη μεγαλύτερη επίδραση στο μοντέλο μας είναι το brightness, το humidity και το cloudcover, η επίδραση τους βέβαια εξαρτάται από το πρόσημο του συντελεστή.



Εικόνα 18 Σημαντικότητα μεταβλητών για το μοντέλο Logistic Regression

Σύγκριση Μεθόδων Μηχανικής Μάθησης

Η ανάλυση του πίνακα συσχέτισης αποκαλύπτει σημαντικές σχέσεις μεταξύ των χαρακτηριστικών και του στόχου μας, που είναι το confidence_class. Οι υψηλές θετικές συσχετίσεις των χαρακτηριστικών brightness (0.58) και frp (0.34) επιβεβαιώνουν τη σημαντικότητά τους, όπως αυτή αναδείχθηκε και από το διάγραμμα σημαντικότητας του Random Forest. Η μέτρια θετική συσχέτιση του bright_t31 (0.34) και η αρνητική συσχέτιση του humidity (-0.36) παρέχουν περαιτέρω ενδείξεις για την επίδραση τους στον στόχο.



Εικόνα 19 Correlation matrix

Αντίθετα χαρακτηριστικά όπως type και daynight, που εμφανίζουν αδύναμες συσχετίσεις με το confidence class επιβεβαιώνουν την χαμηλή σημασία τους που φαίνεται και στο διάγραμμα του Random Forest.

Μετά από την ανάλυση του correlation matrix, παρατηρήσαμε ότι τα χαρακτηριστικά brightness και fir παρουσίασαν υψηλή θετική συσχέτιση με το στόχο μας. Αυτή η παρατήρηση επιβεβαιώθηκε από τα διαγράμματα σημαντικότητας των χαρακτηριστικών του μοντέλου Random Forest αλλά όχι από το μοντέλο Logistic Regression τόσο έντονα. Το μοντέλο Logistic Regression, ανέδειξε την αρνητική σημασία του brightness και την θετική συσχέτιση του humidity κάτι που δεν ήταν τόσο εμφανές στο correlation matrix. Αυτό μπορεί να οφείλεται στην ικανότητα του Random Forest να ανιχνεύει μη γραμμικές σχέσεις και αλληλεπιδράσεις που δεν αποτυπώνονται πλήρως στις απλές συσχετίσεις.

Τέλος, χαρακτηριστικά όπως type και preciptype έδειξαν χαμηλή ή και μηδαμινή συσχέτιση με το στόχο στο correlation matrix, που επιβεβαιώθηκαν ως λιγότερο σημαντικά στο μοντέλο του Random Forest. Αυτή η σύμπτωση ενισχύει την αξιοπιστία των συμπερασμάτων και την ορθότητα του μοντέλου του Random Forest που χρησιμοποιήσαμε.

Κεφάλαιο 6: Συζήτηση και συμπεράσματα

Οι αλγόριθμοι μηχανικής μάθησης έχουν φέρει επανάσταση στον τομέα της αξιολόγησης και της πρόληψης των απειλών από δασικές πυρκαγιές, προσφέροντας ένα ευρύ φάσμα δυνατοτήτων για τον μετριασμό των κινδύνων που συνδέονται με τις δασικές πυρκαγιές. Αναλύοντας ιστορικά δεδομένα καιρού και άλλες σχετικές μεταβλητές, τα μοντέλα μηχανικής μάθησης μπορούν να εντοπίσουν μοτίβα και σχέσεις που υποδεικνύουν αυξημένο κίνδυνο πυρκαγιάς. Αυτά τα μοντέλα στη συνέχεια μπορούν να χρησιμοποιηθούν για την πρόβλεψη της πιθανότητας και της έντασης των δασικών πυρκαγιών σε συγκεκριμένες περιοχές, επιτρέποντας τη λήψη στοχευμένων μέτρων πρόληψης και ταχύτερων χρόνων αντίδρασης.

Στην παρούσα εργασία, εφαρμόσαμε τρεις διαφορετικούς αλγόριθμους μηχανικής μάθησης αξιοποιώντας τα δεδομένα που συλλέξαμε από το MODIS. Αρχικά, τα δεδομένα υποβλήθηκαν σε προ-επεξεργασία με στόχο την εξαγωγή πληροφοριών. Έπειτα, τα δεδομένα χρησιμοποιήθηκαν για την εκπαίδευση των αλγορίθμων Random Forest, Support Vector Machine και Logistic Regression με στόχο την πρόβλεψη τριών κατηγοριών πιθανότητας εμφάνισης πυρκαγιάς: χαμηλής, μεσαίας και

υψηλής. Η περιοχή μελέτης ήταν η Ελλάδα για το χρονικό διάστημα από το 2000 έως και το 2023. Το μοντέλο που είχε την καλύτερη απόδοση ήταν το Random Forest και η απόδοση των μοντέλων αξιολογήθηκε με βάση της μετρικές ταξινόμησης (classifications metrics).

Μελλοντική Έρευνα

Η πρόβλεψη και η πρόληψη των δασικών πυρκαγιών αποτελούν ένα σημαντικό πρόβλημα παγκοσμίως λόγω της σοβαρότητας των επιπτώσεων τους στο περιβάλλον και στην κοινωνία. Η συνεχής έρευνα στον τομέα αυτό είναι ζωτικής σημασίας. Είναι χρήσιμο λοιπόν να αναφερθούν κάποιες μελλοντικές προτάσεις για την βελτίωση των μοντέλων που παρουσιάστηκαν με στόχο την καλύτερη και πιο ακριβή απόδοση:

1. Βελτίωση Υπαρχόντων Μοντέλων: Η μελέτη και δοκιμή νέων παραμετροποιήσεων στα ήδη υπάρχοντα μοντέλα που παρουσιάστηκαν στην παρούσα εργασία ενδέχεται να προσφέρουν ακριβέστερες προβλέψεις και βελτίωση των αποδόσεων τους.
2. Ενσωμάτωση Δεδομένων: Η προσθήκη δεδομένων που σχετίζονται με την υγεία της βλάστησης είναι πολύ σημαντική για την εκπαίδευση και απόδοση των μοντέλων, όπως διαπιστώθηκε, βάση της βιβλιογραφικής αναφοράς που μελετήθηκε στην παρούσα εργασία.
3. Εφαρμογή Μοντέλων Deep Learning: Καθώς η πρόβλεψη των δασικών πυρκαγιών είναι ένα αρκετά πολύπλοκο πρόβλημα διότι συντελούν πολλοί παράγοντες, η χρήση Νευρωνικών Δικτύων θα μπορούσε να αποτυπώσει πιο αποτελεσματικά τις σχέσεις αυτές παρουσιάζοντας πιο αποτελεσματικά μοντέλα πρόβλεψης.

Βιβλιογραφία

- [1] D. Foster, D. Knight και J. Franklin, «Landscape patterns and legacies resulting from large, infrequent forest disturbances.,» *Ecosystems*, αρ. 1, pp. 497-510, 1998.
- [2] L. Giglio, J. Randerson και G. van der Werf, «Analysis of daily, monthly, and annual burned area using the fourth-generation global fire emissions database (GFED4).,» *J. Geophys. Res. Biogeosci.*, τόμ. 118, αρ. 1, pp. 317-328, 2013.
- [3] Z. Gao, W. Yu, H. Wang, Z. Li, T. E. Long και Q. Wang, «Exploring the Effectiveness of Case Teaching Model in Graduate Forest Fire Management Course.,» *Frontiers in Humanities and Social Sciences*, τόμ. 3, αρ. 5, pp. 86-92, 2023.
- [4] M. Biswas, T. Dhoom και S. Barua, «Weather Forecast Prediction: An Integrated Approach for Analyzing and Measuring Weather Data.,» *International Journal of Computer Applications*, τόμ. 182, αρ. 34, pp. 20-14, 2018.
- [5] E. K. Faison, S. A. Masino και W. R. & Moomaw, «The importance of natural forest stewardship in adaptation planning in the United States.,» *Society for Conservation Biology*, τόμ. 5, αρ. 6, 2023.
- [6] S.-U.-D. M και A. Alam, «Performance Evaluation of Machine Learning Algorithms in Ecological Dataset.,» pp. 60-95, 2019.
- [7] D. Kleinbaum και M. Klein, «Logistic regression: a self-learning text.,» *2nd Edition Springer*, pp. 1-513, 2010.
- [8] L. Gigović, H. Pourghasemi, S. Drobnyak και S. Bai, «Testing a new ensemble model based on svm and random forest in forest fire susceptibility assessment and its mapping in serbia's tara national park.,» *Forests*, τόμ. 10, αρ. 5, p. 408, 2019.
- [9] X. Ge, Y. Yang, L. Peng, L. Chen, W. Li, W. Zhang, ... και J. Chen, «Spatio-temporal knowledge graph based forest fire prediction with multi source heterogeneous data.,» *Remote Sensing*, τόμ. 14(14), αρ. 14, p. 3496, 2022.
- [10] O. Satir, S. Berberoglu και C. Donmez, «Mapping regional forest fire probability using artificial neural network model in a mediterranean forest ecosystem.,» *Geomatics Natural Hazards and Risk*, τόμ. 7, αρ. 5, pp. 1645-1658, 2015.
- [11] M. Singh και Z. Huang, «Analysis of forest fire dynamics, distribution and main drivers in the atlantic forest.,» *Sustainability*, τόμ. 14, αρ. 2, p. 992, 2022.
- [12] Y. Li, F. Zhong-ke, S. Chen, Z. Zhao και F. Wang, «Application of the artificial neural network and support vector machines in forest fire prediction in the guangxi autonomous region, china.,» *Discrete Dynamics in Nature and Society*, τόμ. 2020, pp. 1-14, 2020.
- [13] Y. O. Sayad, H. Mousannif και H. Al Moatassime, «Predictive modeling of wildfires: A new dataset and machine learning approach.,» *Fire Safety Journal*, τόμ. 104, pp. 130-146.

- [14] F.-J. Pérez-Porras, P. Triviño-Tarradas, C. Cima-Rodríguez, J.-E. Meroño-de-Larriva, A. García-Ferrer και F.-J. Mesas-Carrascosa, «Machine Learning Methods and Synthetic Data Generation to Predict Large Wildfires.,» *Sensors*, τόμ. 21, αρ. 11, p. 3694, 2021.
- [15] M. Rodrigues και J. De La Riva, «An insight into machine-learning algorithms to model human-caused wildfire occurrence.,» *Environmental Modelling & Software*, τόμ. 57, pp. 192-201, 2014.
- [16] A. M. Elshewey και A. A. Elsonbaty, «Forest Fires Detection Using Machine Learning Techniques,» 2020.
- [17] A. Shmuel και E. Heifetz, «Global Wildfire Susceptibility Mapping Based on Machine Learning Models.,» *Forests*, τόμ. 13, αρ. 7, p. 1050, 2022.
- [18] F. Abid, «A Survey of Machine Learning Algorithms Based Forest Fires Prediction and Detection Systems.,» *Fire Technology*, τόμ. 57, αρ. 2, pp. 559-590, 2021.
- [19] L. K. Sharma, R. Gupta και N. Fatima, «Assessing the predictive efficacy of six machine learning algorithms for the susceptibility of Indian forests to fire.,» *International Journal of Wildland Fire*, τόμ. 31, αρ. 8, p. 735–758, 2022.
- [20] D. Rosadi, W. Andriyani, D. Arisanty και D. Agustina, «Prediction of Forest Fire Occurrence in Peatlands using Machine Learning Approaches.,» *2020 3rd International Seminar on Research of Information Technology and Intelligent Systems (ISRITI)*, p. 48–51, 2020.
- [21] A. Kansal, Y. Singh, N. Kumar και V. Mohindru, «Detection of forest fires using machine learning technique: A perspective,» *2015 Third International Conference on Image Information Processing (ICIIP)*, pp. 241-245, 2015.
- [22] H. Zhu, D. Gao και S. Zhang, «A Perceptron Algorithm for Forest Fire Prediction Based on Wireless Sensor Networks.,» *Journal on Internet of Things.*, τόμ. 1, αρ. 1, pp. 25-31, 2019.
- [23] L. Giglio, J. Descloitres, C. Justice και Y. Kaufman, «An enhanced contextual fire detection algorithm for MODIS.,» *Remote Sensing of Environment*, τόμ. 87, pp. 273-282.
- [24] L. Giglio, W. Schroeder, J. Hall και C. Justice, «MODIS Collection 6 and Collection 6.1 Active Fire Product User’s Guide,» *NASA*, 2021.
- [25] L. Breiman, «Statistical Modeling: The Two Cultures.,» *Statistical Science*, τόμ. 16, αρ. 3, pp. 199-231, 2001.
- [26] M. Hearst, S. Dumais, E. Osuna, J. Platt και B. Scholkopf, «Support vector machines.,» *IEEE Intelligent Systems and their Applications*, τόμ. 13, αρ. 4, pp. 18-28, 1998.
- [27] A. M. Elshewey και A. A. Elsonbaty, «Forest Fires Detection Using Machine Learning Techniques».

