ΕΘΝΙΚΟ ΜΕΤΣΟΒΙΟ ΠΟΛΥΤΕΧΝΕΙΟ
ΣΧΟΛΗ ΠΟΛΙΤΙΚΩΝ ΜΗΧΑΝΙΚΩΝ
ΤΟΜΕΑΣ ΜΕΤΑΦΟΡΩΝ ΚΑΙ ΣΥΓΚΟΙΝΩΝΙΑΚΗΣ ΥΠΟΔΟΜΗΣ

Συμπεριφορά χρηστών οδικού δικτύου κατά την εκκένωση οικισμών λόγω πυρκαγιάς

Διπλωματική Εργασία

Αθήνα, Οκτώβριος 2020

Κατζηλιέρης Κωνσταντίνος
Επιβλέπουσα καθηγήτρια: Ελένη Ι. Βλαχογιάννη

ΕΥΧΑΡΙΣΤΙΕΣ

Με την ολοκλήρωση της Διπλωματικής μου εργασίας και συνεπώς των προπτυχιακών μου σπουδών, θα ήθελα να ευχαριστήσω θερμά όλους όσους συνέβαλαν στην ολοκλήρωσή της, αλλά και όλους όσους στήριξαν τη φοιτητική μου πορεία.

Αρχικά θα ήθελα να ευχαριστήσω θερμά την κ. Ελένη Βλαχογιάννη, Αναπληρώτρια Καθηγήτρια της Σχολής Πολιτικών Μηχανικών Ε.Μ.Π., για την ανάθεση και επίβλεψη της παρούσας Διπλωματικής Εργασίας, την υποστήριξη και την πολύτιμη καθοδήγησή της σε όλα τα στάδια εκπόνησής της, καθώς και για την εξαιρετική συνεργασία μας.

Επιπλέον, θα ήθελα να ευχαριστήσω την κ. Ελένη Μαντούκα, υποψήφια Διδάκτορα Ε.Μ.Π., για την συμβολή της στην ολοκλήρωση της εργασίας και τις πολύ χρήσιμες συμβουλές της.

Κλείνοντας, θα ήθελα να ευχαριστήσω την οικογένεια και τους φίλους μου για την υποστήριξη που μου προσέφεραν καθ' όλη τη διάρκεια των σπουδών μου.

Αθήνα, Οκτώβριος 2020
Κατζηλιέρης Κωνσταντίνος

Συμπεριφορά χρηστών οδικού δικτύου κατά την εκκένωση οικισμών λόγω πυρκαγιάς

Κωνσταντίνος Κατζηλιέρης

Επιβλέπουσα Καθηγήτρια: Ελένη Ι. Βλαχογιάννη

ΣΥΝΟΨΗ

Στόχος της παρούσας Διπλωματικής Εργασίας είναι η διερεύνηση της συμπεριφοράς των χρηστών του οδικού δικτύου κατά την εκκένωση οικισμών λόγω πυρκαγιάς. Για την επίτευξη του στόχου αυτού, αξιοποιήθηκαν δεδομένα από έρευνα κατάλληλα διαμορφωμένου ερωτηματολογίου το οποίο συμπληρώθηκε από κατοίκους περιοχών που επλήγησαν από τις πυρκαγιές της Ανατολικής Αττικής το 2018. Τα στοιχεία που συλλέχθηκαν αφορούσαν στην περιγραφή των γεγονότων και των κρίσιμων αποφάσεων που πήραν οι κάτοικοι των πληγεισών αυτών περιοχών κατά τη διάρκεια της πυρκαγιάς, τα δημογραφικά χαρακτηριστικά τους και τις ενδεχόμενες αποφάσεις τους σε μια πιθανή νέα μελλοντική πυρκαγιά. Στη συνέχεια, δημιουργήθηκαν μοντέλα λογιστικής παλινδρόμησης, δέντρων αποφάσεων και τυχαίων δασών που είχαν ως σκοπό τη διερεύνηση των κρισιμότερων παραγόντων που επηρεάζουν την απόφαση για εκκένωση η παραμονή καθώς και την επιλογή του μέσου κατά την εκκένωση. Σχετικά με την απόφαση για εκκένωση/ παραμονή κρισιμότεροι παράγοντες ήταν η αντίληψη του κινδύνου, η ηλικία, τα χρόνια διαμονής στην κατοικία, ο αριθμός ενηλίκων έως 65 ετών στην κατοικία, η προσπάθεια για λήψη πληροφοριών πριν την εκκένωση, το φύλο, η ύπαρξη προηγούμενης προειδοποίησης, ο αριθμός ανηλίκων στην κατοικία, το επίπεδο εκπαίδευσης και το εισόδημα. Αναφορικά με την επιλογή μέσου κατά την εκκένωση, κρισιμότεροι παράγοντες ήταν η ύπαρξη ή μη διαθέσιμου οχήματος, η ηλικία, η προσπάθεια για λήψη πληροφοριών πριν την εκκένωση, η αντίληψη του κινδύνου που είχαν για τη συγκεκριμένη απειλή καθώς και το ετήσιο οικογενειακό τους εισόδημα. Επιπλέον, μέσα από τη στατιστική ανάλυση της βάσης δεδομένων προέκυψαν και ορισμένα γενικά συμπεράσματα ως προς τη συμπεριφορά του πληθυσμού και την επιρροή μιας επίσημης εντολής εκκένωσης από την πολιτεία.

Λέξεις-κλειδιά: Εκκένωση Οικισμών Λόγω Πυρκαγιάς, Συμπεριφορά Χρηστών, Μοντέλα Λογιστικής Παλινδρόμησης , Μοντέλα Δέντρων Απόφασης, Μοντέλα Τυχαίων Δασών

Behaviour of road users during community evacuations due to wildfire

Konstantinos Katzilieris

Supervisor: Eleni I. Vlahogianni

ABSTRACT

The objective of this Diploma Thesis is to study the behaviour of road users during community evacuations due to wildfire. To achieve this objective, data were obtained from a questionnaire survey completed by residents of areas affected by the wildfires of East Attica in 2018. The data collected concerned a description of the event and of critical decisions taken by the residents of these affected areas during the fire, their demographic characteristics and their possible future decisions on a new hypothetical wildfire. The data was processed using simple statistics as well as machine learning algorithms. In particular, models of logistic regression, decision trees and random forests were created to investigate the most critical factors influencing the evacuate/stay decision as well as the mode choice during evacuation. Regarding the decision to evacuate/stay, most critical factors were the perception of risk, age, years in residence, the number of adults up to 65 years of age in residence, the attempt to obtain information before evacuation, gender, the existence of a prior warning, the number of minors in residence, the level of education and income. Regarding the mode choice, the most critical factors were the existence of an available vehicle, age, the attempt to obtain information before the evacuation and the perception of risk. In addition, the statistical analysis of the database also revealed some general conclusions as to the behavior of the population and the influence of an official evacuation order from the state.

Keywords: Community Evacuation Due To Wildfire, User Behavior, Logistic Regression Models, Decision Tree Models, Random Forest Models

ΠΕΡΙΛΗΨΗ

Αντικείμενο της παρούσας διπλωματικής εργασίας αποτέλεσε η έρευνα της συμπεριφοράς των χρηστών οδικού δικτύου κατά την εκκένωση οικισμών λόγω πυρκαγιάς. Εκτενέστερη έρευνα έγινε για τις κρίσιμες αποφάσεις που λαμβάνουν οι άνθρωποι που καλούνται να εκκενώσουν μια περιοχή, δηλαδή την απόφαση για εκκένωση ή παραμονή αλλά και την απόφαση για επιλογή μέσου σε περίπτωση εκκένωσης. Η γνώση των παραπάνω είναι απαραίτητη σύμφωνα με τη βιβλιογραφία για μια ολοκληρωμένη μελέτη εκκένωσης ώστε, σε συνδυασμό και με την τοπολογία και τη χωρική κατανομή του πληθυσμού, να καθοριστεί η ζήτηση των μετακινήσεων για ένα συγκοινωνιακό δίκτυο κατά την εκκένωση μιας περιοχής.

Τα απαραίτητα δεδομένα για την πραγματοποίηση της έρευνας συλλέχθηκαν μέσω ενός κατάλληλα διαμορφωμένου ερωτηματολογίου το οποίο συμπληρώθηκε από κατοίκους περιοχών που επλήγησαν από τις πυρκαγιές της Ανατολικής Αττικής το 2018. Οι ερωτήσεις του ερωτηματολογίου αφορούσαν τον τρόπο με τον οποίο οι ερωτηθέντες είχαν ενημερωθεί για την πυρκαγιά, τις αποφάσεις που πήραν κατά τη διάρκεια της απειλής, τα συναισθήματα και την αντίληψη που είχαν για το ρίσκο, τα μέτρα προστασίας που είχαν λάβει, τις αποφάσεις που θα έπαιρναν σε μία νέα ενδεχόμενη πυρκαγιά έχοντας ζήσει την εμπειρία καθώς και τα μέτρα προστασίας που έλαβαν ύστερα από την πυρκαγιά. Επιπλέον το ερωτηματολόγιο περιελάμβανε ερωτήσεις σχετικές με τα δημογραφικά χαρακτηριστικά των ερωτηθέντων ώστε να εξαχθούν συμπεράσματα ως προς τη συσχέτιση ορισμένων δημογραφικών χαρακτηριστικών με άλλες απαντήσεις του ερωτηματολογίου αλλά και να γίνει έλεγχος της αντιπροσωπευτικότητας του δείγματος.

Από τη στατιστική ανάλυση της βάσης δεδομένων προκύπτουν τα εξής γενικά συμπεράσματα:

Η πλειοψηφία των ερωτηθέντων (86%) προσπάθησε να εκκενώσει την περιοχή κατοικίας του.

Οι άντρες αποφάσισαν να μείνουν και να προστατέψουν τις κατοικίες τους σε μεγαλύτερο ποσοστό σε σχέση με τις γυναίκες αφού μόλις το 80% των ανδρών που συμπλήρωσαν το ερωτηματολόγιο εκκένωσαν την κατοικία τους σε αντίθεση με τις γυναίκες που εκκένωσαν σε ποσοστό 91,4% (Στατιστικά σημαντική διαφορά σε 95% επίπεδο σημαντικότητας)

Υπάρχει συσχέτιση μεταξύ της ηλικίας και της απόφασης για εκκένωση καθώς παρατηρείται σημαντική διαφορά μεταξύ των διαφόρων ηλικιακών ομάδων ως προς την απόφαση για εκκένωση. Πρωτιά στο ποσοστό εκκένωσης σημειώνουν οι ηλικιωμένοι 70+, ακολουθούν οι νέοι μέχρι 29 ετών, και τις τελευταίες θέσεις καταλαμβάνουν οι ηλικιακές ομάδες 30-49 και 50-69

Μια έγκαιρη επίσημη εντολή εκκένωσης από την πολιτεία έχει πολύ σημαντική επιρροή στην απόφαση για εκκένωση. Οι ερωτηθέντες δήλωσαν πως σε μία νέα ενδεχόμενη μελλοντική πυρκαγιά δεν είναι ιδιαίτερα πιθανό (1,86/5) να εκκενώσουν την κατοικία τους χωρίς να έχει δοθεί επίσημη εντολή εκκένωσης. Αντίθετα δήλωσαν πως είναι εξαιρετικά πιθανό (4,45/5) να εκκενώσουν εφόσον έχει προηγηθεί μια εντολή εκκένωσης από την πολιτεία (Στατιστικά σημαντική διαφορά σε 99% επίπεδο σημαντικότητας)

Ενώ το 90,8% των ερωτηθέντων διέθετε όχημα διαθέσιμο κατά τη διάρκεια της πυρκαγιάς μόνο το 76,2% το χρησιμοποίησε ως μέσο διαφυγής. (Στατιστικά σημαντική διαφορά σε 95% επίπεδο

σημαντικότητας). Επομένως η διάθεση οχήματος δεν αποτελεί τον μοναδικό παράγοντα για την επιλογή μέσου κατά την εκκένωση.

Το 74% των ερωτηθέντων που επέλεξαν να εκκενώσουν με τα πόδια κατέληξαν ανάμεσα στην πυρκαγιά ενώ μόνο το 47% εκείνων που διέφυγαν με το αυτοκίνητο βρέθηκαν ανάμεσα στις φλόγες (Στατιστικά σημαντική διαφορά σε 99% επίπεδο σημαντικότητας). Επομένως είναι σαφές πως η εκκένωση με τα πόδια ήταν πιο επικίνδυνη.

Η ηλικιακή ομάδα 18-29 επέλεξε το περπάτημα σε ποσοστό 45%, η ηλικιακή ομάδα 30-49 σε ποσοστό 20%, η ηλικιακή ομάδα 50-69 σε ποσοστό 16% και η ηλικιακή ομάδα 70+ σε ποσοστό 9%. Επομένως γίνεται σαφές πως όσο μεγαλύτερη ήταν η ηλικία, τόσο μεγαλύτερο ήταν και το ποσοστό εκκένωσης με Ι.Χ..

Το 65,6% των ερωτηθέντων δήλωσαν πως σε νέα ενδεχόμενη πυρκαγιά θα επέλεγαν να χρησιμοποιήσουν μόνο 1 όχημα σε αντίθεση με το υπόλοιπο 34,4% που δήλωσε πως θα προσπαθούσε να εκκενώσει με το σύνολο των διαθέσιμων οχημάτων της οικογενείας.

Ακόμα, δημιουργήθηκαν μοντέλα λογιστικής παλινδρόμησης, δέντρων αποφάσεων και τυχαίων δασών που είχαν ως σκοπό τη διερεύνηση των κρισιμότερων παραγόντων που επηρεάζουν την απόφαση για εκκένωση η παραμονή καθώς και την επιλογή του μέσου κατά την εκκένωση. Όλα τα μοντέλα παρουσιάζουν αυξημένη ακρίβεια. Το τυχαίο δάσος δίνει τις καλύτερες προβλέψεις για εκκένωση/παραμονή (f measure =96%) και επιλογής μέσου (f measure =93%) από όλα τα μοντέλα που εξετάστηκαν.

Επιπλέον, από την ανάλυση των μοντέλων μηχανικής μάθησης προκύπτουν οι κρισιμότεροι παράγοντες που επηρεάζουν τις αποφάσεις για εκκένωση/παραμονή και επιλογή μέσου. Όσον αφορά την απόφαση για εκκένωση/παραμονή, κρισιμότεροι παράγοντες σύμφωνα με το μοντέλο λογιστικής παλινδρόμησης προέκυψαν οι εξής:

- Τύπος κυριότητας κατοικίας - Αρνητική για ιδιοκτήτες των ακινήτων
- Ηλικία - Θετική για μεγαλύτερες ηλικίες
- Φύλο - Θετική για τις γυναίκες
- Αριθμός ανηλίκων στην κατοικία - Αρνητική για μεγαλύτερο αριθμό ανηλίκων
- Εισόδημα - Θετική για μεγαλύτερο εισόδημα
- Αντίληψη κινδύνου - Θετική για μεγαλύτερη αντίληψη κινδύνου
- Συμβουλή για εκκένωση από φίλο/συγγενή - Θετική για όσους έλαβαν συμβουλή για εκκένωση από φίλους/συγγενείς
- Προσπάθεια για λήψη επιπλέον πληροφόρησης πριν την απόφαση για εκκένωση - Θετική για όσους δεν προσπάθησαν να λάβουν επιπλέον πληροφορίες πριν εκκενώσουν

Το μοντέλο τυχαίου δάσους που αναπτύχθηκε για τον ίδιο σκοπό ωστόσο έδειξε μια μικρή διαφοροποίηση στους παράγοντες που επιλέχθηκαν. Συγκεκριμένα, σε σειρά προτεραιότητας ως προς τη σημαντικότητα, οι παράγοντες που επιλέχθηκαν ήταν οι εξής:

- Αντίληψη του κινδύνου
- Ηλικία
- Χρόνια διαμονής στην κατοικία
- Αριθμός ενηλίκων έως 65 ετών στην κατοικία
- Προσπάθεια για λήψη πληροφοριών πριν την εκκένωση
- Φύλο

- Ύπαρξη προηγούμενης προειδοποίησης
- Αριθμός ανηλίκων στην κατοικία
- Επίπεδο εκπαίδευσης
- Εισόδημα

Τέλος, για την επιλογή μέσου λήφθηκε υπόψιν μονάχα το μοντέλο τυχαίου δάσους καθώς τα μοντέλα λογιστικής παλινδρόμησης και δέντρου απόφασης κρίθηκαν αναξιόπιστα λόγω μη ικανοποιητικών στατιστικών χαρακτηριστικών. Από την ανάλυση φάνηκε πως υπάρχουν λίγες μόνο μεταβλητές που επηρεάζουν την επιλογή μέσου. Αυτές είναι:

- η ύπαρξη ή μη διαθέσιμου οχήματος
- η ηλικία
- η προσπάθεια για λήψη πληροφοριών πριν την εκκένωση
- η αντίληψη του κινδύνου που είχαν για τη συγκεκριμένη απειλή καθώς και
- το ετήσιο οικογενειακό τους εισόδημα

Από την έρευνα που πραγματοποιήθηκε στην παρούσα διπλωματική εργασία προέκυψαν ορισμένες προτάσεις για περαιτέρω έρευνα.

Αρχικά, η διεξαγωγή παρόμοιας έρευνας σε άλλες περιοχές τις Ελλάδας με ποικίλα κοινωνικά και οικονομικά χαρακτηριστικά θα συνεισέφερε στην επιβεβαίωση ή διόρθωση των αποτελεσμάτων της συγκεκριμένης εργασίας ή στην ανίχνευση επιπλέον παραγόντων επιρροής τη συμπεριφοράς που σχετίζονται με πολιτισμικά και κοινωνικά κριτήρια.

Επιπλέον, σε έρευνες που θα είχαν ένα μεγαλύτερο δείγμα θα μπορούσε να πραγματοποιηθεί και προσομοίωση της κυκλοφορίας κατά τη διάρκεια της πυρκαγιάς, που θα οδηγούσε στον χαρακτηρισμό της ανθεκτικότητας του εξεταζόμενου δικτύου. Στη συγκεκριμένη εργασία συλλέχθηκαν ορισμένα στοιχεία που αφορούσαν την προέλευση και τον προορισμό των μετακινήσεων κατά την εκκένωση, όμως δεν χρησιμοποιήθηκαν. Τα στοιχεία αυτά θα μπορούσαν να συμβάλουν στην ορθή βαθμονόμηση συμπεριφορισιακών μοντέλων κυκλοφορίας που θα αποτελέσουν τη βάση για την ανάπτυξη στρατηγικών διαχείρισης κυκλοφορίας σε κρίσιμές καταστάσεις εκκένωσης.

ΠΕΡΙΕΧΟΜΕΝΑ

1 Εισαγωγή 1
1.1 Εκκένωση οικισμών λόγω φυσικών καταστροφών 1
1.2 Σκοπός διπλωματικής εργασίας 2
1.3 Διάρθρωση διπλωματικής εργασίας 2
2 Βιβλιογραφικη ανασκοπηση 4
2.1 Εισαγωγή 4
2.2 Χρόνος απομάκρυνσης 4
2.3 Ζήτηση Μετακινήσεων 4
2.3.1 Παράγοντες που επηρεάζουν την απόφαση για εκκένωση / παραμονή 5
2.3.2 Επιλογή μέσου κατά την εκκένωση 7
2.4 Προσφορά 8
2.5 Συμπεράσματα 8
3 Μεθοδολογική Προσέγγιση 9
3.1 Περιγραφή της διαδικασίας 9
3.2 Συλλογή δεδομένων 9
3.2.1 Το ερωτηματολόγιο 9
3.2.2 Κωδικοποίηση ερωτηματολογίου 10
3.3 Θεωρητικό Υπόβαθρο 11
3.3.1 Εισαγωγή 11
3.3.2 Μαθηματικά Πρότυπα 11
3.3.3 Κριτήρια αποδοχής μοντέλων 15
4 Ανάλυση και αποτελέσματα 19
4.1 Εισαγωγή 19
4.2 Περιγραφή βάσης δεδομένων 19
4.2.1 Συγκεντρωτικά στοιχεία 19
4.2.2 Αντιπροσωπευτικότητα δείγματος 22
4.3 Ανάλυση βάσης δεδομένων 24
4.3.1 Σημεία με ενδιαφέρον 24
4.4 Πρόβλεψη απόφασης για εκκένωση 31
4.4.1 Μοντέλο λογιστικής παλινδρόμησης για πρόβλεψη εκκένωσης 31
4.4.2 Δέντρο αποφάσεων για πρόβλεψη εκκένωσης 36

4.4.3 Μοντέλο τυχαίου δάσους για πρόβλεψη εκκένωσης 40
4.5 Πρόβλεψη επιλογής μέσου κατά την εκκένωση 43
4.5.1 Μοντέλο λογιστικής παλινδρόμησης για πρόβλεψη επιλογής μέσου 43
4.5.2 Δέντρο αποφάσεων για πρόβλεψη επιλογής μέσου 47
4.5.3 Μοντέλο τυχαίου δάσους για πρόβλεψη επιλογής μέσου 49
5 ΣΥΜΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ 54
5.1 Σύνοψη μεθοδολογίας και αποτελεσμάτων 54
5.2 Βασικά συμπεράσματα 55
5.2.1 Γενικά συμπεράσματα 55
5.2.2 Παράγοντες που επηρεάζουν την απόφαση για εκκένωση/παραμονή 56
5.2.3 Παράγοντες που επηρεάζουν την επιλογή μέσου κατά την εκκένωση 57
5.3 Προτάσεις για περαιτέρω έρευνα 58
Βιβλιογραφία 59
Παράρτημα Ι - Ερωτηματολόγιο 61
Παράρτημα ΙΙ – Κώδικες 64

ΕΥΡΕΤΗΡΙΟ ΠΙΝΑΚΩΝ

Πίνακας 2.1 Παράγοντες που επηρεάζουν την απόφαση για εκκένωση ανά τύπο απειλής σύμφωνα με τη βιβλιογραφία 6
Πίνακας 3.1 Μήτρα Σύγχυσης 16
Πίνακας 4.1 Έλεγχος καλής προσαρμογής του δείγματος ως προς το φύλο 22
Πίνακας 4.2 Έλεγχος καλής προσαρμογής του δείγματος ως προς τον τύπο κυριότητας της κατοικίας 23
Πίνακας 4.3 Έλεγχος καλής προσαρμογής του δείγματος ως προς την οικογενειακή κατάσταση 24
Πίνακας 4.4 Πίνακας μεταβλητών απόφασης για εκκένωση/παραμονή 32
Πίνακας 4.5 Μήτρα σύγχυσης μοντέλου λογιστικής παλινδρόμησης 34
Πίνακας 4.6 Στατιστικά χαρακτηριστικά μοντέλου λογιστικής παλινδρόμησης 34
Πίνακας 4.7 Επίδραση και επίπεδο εμπιστοσύνης για τους διάφορους παράγοντες που χρησιμοποιούνται στο μοντέλο λογιστικής παλινδρόμησης 35
Πίνακας 4.8 Μήτρα σύγχυσης μοντέλου τυχαίου δάσους 41
Πίνακας 4.9 Στατιστικά χαρακτηριστικά μοντέλου τυχαίου δάσους 41
Πίνακας 4.10 Πίνακας μεταβλητών για επιλογή μέσου 44
Πίνακας 4.11 Μήτρα σύγχυσης μοντέλου λογιστικής παλινδρόμησης για επιλογή μέσου 46
Πίνακας 4.12 Στατιστικά χαρακτηριστικά μοντέλου λογιστικής παλινδρόμησης για επιλογή μέσου 46
Πίνακας 4.13 Μήτρα σύγχυσης μοντέλου τυχαίου δάσους 51
Πίνακας 4.14 Στατιστικά χαρακτηριστικά μοντέλου τυχαίου δάσους 51
Πίνακας 5.1 Συγκεντρωτικός πίνακας αξιολόγησης μοντέλων πρόβλεψης της απόφασης για εκκένωση/παραμονή 54
Πίνακας 5.2 Συγκεντρωτικός πίνακας αξιολόγησης μοντέλων πρόβλεψης της επιλογής μέσου 55
Πίνακας 5.3 Σύγκριση αποτελεσμάτων με τη διεθνή βιβλιογραφία 56

ΕΥΡΕΤΗΡΙΟ ΔΙΑΓΡΑΜΜΑΤΩΝ

Διάγραμμα 2.1 Κλίμακες χρόνου και απόστασης κατά την εκκένωση για τους διάφορους τύπους απειλής 7
Διάγραμμα 3.1 Διάγραμμα ροής μεθοδολογικής προσέγγισης 9
Διάγραμμα 3.2 Παράδειγμα δέντρου απόφασης στο dataset "Iris" του UCI Machine Learning Repository. Εξαρτημένη μεταβλητή είναι το είδος του φυτού και ανεξάρτητες μεταβλητές τα χαρακτηριστικά των φύλλων. 14
Διάγραμμα 3.3 Παράδειγμα χαρακτηριστικής καμπύλης λειτουργίας δέκτη (ROC curve)..... 18
Διάγραμμα 4.1 Κατανομή ερωτηθέντων με βάση το φύλο 19
Διάγραμμα 4.2 Κατανομή ερωτηθέντων με βάση την ηλικία 20
Διάγραμμα 4.3 Ποσοστιαία κατανομή ανάλογα με την οικογενειακή κατάσταση των ερωτηθέντων 20
Διάγραμμα 4.4 Ποσοστιαία κατανομή ανάλογα με τον τύπο ιδιοκτησίας των κατοικιών των ερωτηθέντων 21
Διάγραμμα 4.5 Ποσοστιαία κατανομή με βάση το επίπεδο εκπαίδευσης των ερωτηθέντων 21
Διάγραμμα 4.6 Ποσοστιαία κατανομή με βάση το οικογενειακό εισόδημα των ερωτηθέντων 21
Διάγραμμα 4.7 Κατανομή ερωτηθέντων με βάση το αν βρέθηκαν ή όχι ανάμεσα στην πυρκαγιά 25
Διάγραμμα 4.8 Κατανομή ερωτηθέντων με βάση το αν εκκένωσαν ή όχι την περιοχή κατοικίας τους 25
Διάγραμμα 4.9 Ποσοστά ερωτηθέντων που εκκένωσαν ανάλογα με το φύλο 26
Διάγραμμα 4.10 Ποσοστά ερωτηθέντων που εκκένωσαν ανάλογα με την ηλικιακή ομάδα 26
Διάγραμμα 4.11 Διαφορά στην πρόθεση για εκκένωση από την έκδοση επίσημης εντολής εκκένωσης 27
Διάγραμμα 4.12 Κατανομή ερωτηθέντων με βάση την πρώτη πηγή πληροφόρησης για την πυρκαγιά 27
Διάγραμμα 4.13 Ποσοστό ερωτηθέντων όπου είχαν διαθέσιμο όχημα κατά την εκκένωση..... 28
Διάγραμμα 4.14 Κατανομή ερωτηθέντων με βάση τον τρόπο εκκένωσης 28
Διάγραμμα 4.15 Ποσοστά ερωτηθέντων που βρέθηκαν ανάμεσα στην πυρκαγιά ανάλογα με τον τρόπο εκκένωσης 29
Διάγραμμα 4.16 Επιλογή μέσου ανά ηλικιακή ομάδα 29
Διάγραμμα 4.17 Κατανομή ερωτηθέντων με βάση την ύπαρξη σχεδιασμού για μελλοντική εκκένωση 30
Διάγραμμα 4.18 Αριθμός οχημάτων ανά οικογένεια κατά την εκκένωση 30
Διάγραμμα 4.19 Επιλογή μέσου σε πιθανή μελλοντική πυρκαγιά 31
Διάγραμμα 4.20 Κατανομή ερωτηθέντων με βάση τον προγραμματισμό διαδρομής για μελλοντική εκκένωση 31
Διάγραμμα 4.21 Καμπύλη Receiver Operating Characteristic (ROC curve) του μοντέλου λογιστικής παλινδρόμησης 35
Διάγραμμα 4.22 Σχήμα δέντρου απόφασης σχετικά με την απόφαση για εκκένωση 38
Διάγραμμα 4.23 Σημαντικότητα μεταβλητών για τον αλγόριθμο τυχαίου δάσους 40
Διάγραμμα 4.24 Καμπύλη ROC μοντέλου randomForest για την πρόβλεψη της απόφασης για εκκένωση / παραμονή 42
Διάγραμμα 4.25 Σημαντικότητα μεταβλητών για τον τελικό αλγόριθμο τυχαίου δάσους 42

Διάγραμμα 4.26 Καμπύλη Receiver Operating Characteristic μοντέλου λογιστικής παλινδρόμησης για επιλογή μέσου..... 46
Διάγραμμα 4.27 Διάγραμμα δέντρου απόφασης για την επιλογή μέσου κατά την εκκένωση 48
Διάγραμμα 4.28 Σημαντικότητα μεταβλητών για τον αλγόριθμο τυχαίου δάσους 50
Διάγραμμα 4.29 Καμπύλη ROC μοντέλου randomForest για την πρόβλεψη της επιλογής μέσου 51
Διάγραμμα 4.30 Σημαντικότητα μεταβλητών για τον τελικό αλγόριθμο τυχαίου δάσους 52

1 ΕΙΣΑΓΩΓΗ

1.1 Εκκένωση οικισμών λόγω φυσικών καταστροφών

Οι φυσικές καταστροφές και τα ακραία γεγονότα μπορούν να έχουν δραματικές επιπτώσεις στην ευημερία των πολιτών και την οικονομία (Pel et al. 2012). Συχνά μάλιστα μπορεί να οδηγήσουν και στην απώλεια ανθρώπινων ζωών. Οι φυσικές καταστροφές περιλαμβάνουν εκρήξεις ηφαιστείων, σεισμούς, τσουνάμι, τυφώνες, πλημμύρες και πυρκαγιές. Οι ανθρωπογενείς καταστροφές προκαλούνται από βιομηχανικά και πυρηνικά ατυχήματα, χημικές διαρροές και στρατιωτικές και τρομοκρατικές δραστηριότητες. Τα τελευταία χρόνια, ιδιαίτερα λόγω της κλιματικής αλλαγής παρατηρείται μια αύξηση παγκοσμίως στα ακραία καιρικά φαινόμενα που δημιουργούν φυσικές καταστροφές και επομένως την ανάγκη για εκκένωση περιοχών με μεγάλους πληθυσμούς. Οι καταστροφές ποικίλλουν σε μεγάλο βαθμό ως προς τα χαρακτηριστικά τους, όπως οι χρόνοι προειδοποίησης, η διάρκεια του ίδιου του γεγονότος, η αμεσότητα, η χωρική και χρονική έκταση των επιπτώσεων και ούτω καθεξής. Όλα αυτά έχουν σημαντικές επιπτώσεις στον σχεδιασμό της αντιμετώπισης. Ωστόσο, σε όλες τις περιπτώσεις, η αποτελεσματική διαχείριση του συγκοινωνιακού συστήματος είναι ζωτικής σημασίας προκειμένου να διευκολυνθεί η εκκένωση των πληγεισών πληθυσμών και η μετακίνηση του προσωπικού και του εξοπλισμού της πολιτικής προστασίας (Perry and Lindell, 2003).

Η αδυναμία πρόβλεψης των ακριβών συνθηκών μιας καταστροφής σημαίνει ότι τα προτεινόμενα σχέδια εκκένωσης και μεταφοράς πρέπει να προσαρμοστούν και να αξιολογηθούν σε ένα ευρύ φάσμα σεναρίων όσον αφορά τη φύση του γεγονότος και τις επιπτώσεις του στο συγκοινωνιακό σύστημα (Urata and Hato, 2012). Η μοντελοποίηση των συγκοινωνιακών συστημάτων είναι απαραίτητη για την υποστήριξη της ανάπτυξης τέτοιων σχεδίων, με τα μοντέλα ενδεχομένως να μπορούν να χρησιμοποιηθούν και στα κέντρα ελέγχου και λήψης αποφάσεων της πολιτικής προστασίας κατά τη διάρκεια μιας καταστροφής (Brachman and Church, 2009).

Η πλευρά της προσφοράς του συγκοινωνιακού συστήματος μπορεί να επηρεαστεί κατά τη διάρκεια ενός συμβάντος μέσω του κλεισίματος μιας η περισσότερων οδών και της διακοπής της λειτουργίας των ΜΜΜ (Taylor and Freeman, 2010). Μπορούν ωστόσο να εφαρμοστούν και ορισμένες στρατηγικές διαχείρισης της κυκλοφορίας που αυξάνουν την προσφορά του δικτύου όπως είναι η παραχώρηση μίας η περισσότερων λωρίδων κυκλοφορίας από το αντίθετο ρεύμα (contraflow).

Οι επιπτώσεις της ζήτησης μπορεί να είναι πιο πολύπλοκες και δύσκολο να προβλεφθούν. Ως εκ τούτου, η κατανόηση της ζήτησης για μετακινήσεις κατά τη διάρκεια καταστροφών, ιδίως όταν πραγματοποιούνται εκκενώσεις, είναι ζωτικής σημασίας για την αποτελεσματική διαχείριση των κινδύνων (van der Gun et al., 2016). Η συνήθης κυκλοφοριακή δραστηριότητα και οι οδηγικές συνήθειες των χρηστών του δικτύου μπορεί να είναι εντελώς διαφορετικές κατά τη διάρκεια μιας καταστροφής, γι' αυτό και στις μελέτες εκκένωσης οικισμών χρησιμοποιούνται διαφορετικές εκτιμήσεις από τις συνήθεις για τα διάφορα κυκλοφοριακά μεγέθη (McLennan et al., 2012). Αυτές εξαρτώνται από τη φύση και την έκταση του γεγονότος, από τις πληροφορίες και τις οδηγίες που παρέχονται στον πληθυσμό και από τις τοποθεσίες και τις δραστηριότητες που πραγματοποιούν τα μέλη του νοικοκυριού κατά την έναρξη του συμβάντος. Οι αλληλεπιδράσεις και οι εξαρτήσεις μεταξύ των συμπεριφορών των διαφορετικών μελών του νοικοκυριού, καθώς και με άλλους στο εκτεταμένο οικογενειακό και κοινωνικό δίκτυο, είναι επίσης εξαιρετικά σημαντικές και μπορεί να

υπαγορεύουν μοτίβα ως προς την κινητικότητα (Urata and Hato, 2012). Για παράδειγμα, σε μια κατάσταση εκκένωσης, οι ενήλικες μπορεί πρώτα να χρειαστεί να παραλάβουν εξαρτώμενα παιδιά ή άλλα άτομα που χρειάζονται βοήθεια, πριν εκκενώσουν. Τα καθήκοντα αυτά μπορούν να ανατεθούν σε συγκεκριμένα μέλη του νοικοκυριού, με βάση τη φυσική ή κοινωνική εγγύτητά τους με τα άτομα που χρειάζονται βοήθεια, τη διαθεσιμότητα των οχημάτων και ούτω καθεξής (Liu et al., 2014).

1.2 Σκοπός διπλωματικής εργασίας

Δεδομένου των συχνών καταστροφικών πυρκαγιών που αντιμετωπίζει η χώρα μας κάθε χρόνο, είναι απαραίτητη η πραγματοποίηση ερευνών για τον προσδιορισμό των διαδικασιών βέλτιστης εκκένωσης περιοχών κοντά σε δασικές εκτάσεις στις οποίες υπάρχει κίνδυνος πυρκαγιάς. Μέσω των ερευνών αυτών, θα μπορεί να εξετάζεται η ανθεκτικότητα και η αντοχή των συγκοινωνιακών συστημάτων των περιοχών αυτών σε φορτίσεις που προκύπτουν από εκκένωση και να προτείνονται βελτιώσεις που ενδεχομένως να απαιτούνται στην υπάρχουσα υποδομή.

Η πιο σύνθετη και πολύπλοκη διαδικασία μιας τέτοιας έρευνας είναι η μοντελοποίηση της ζήτησης των μετακινήσεων, αφού αυτή δεν εξαρτάται μόνο από το μέγεθος του πληθυσμού και τη χωρική κατανομή του, αλλά και από τη συμπεριφορά των ανθρώπων και τις κρίσιμες αποφάσεις που παίρνουν κατά τη διάρκεια της απειλής. Σημαντικότερες είναι οι αποφάσεις για εκκένωση ή παραμονή στην περιοχή και η απόφαση για την επιλογή του μέσου κατά την εκκένωση. Για τη διερεύνηση των αποφάσεων αυτών, έχουν πραγματοποιηθεί πολλές έρευνες στο εξωτερικό, ωστόσο κρίνεται απαραίτητη η διερεύνηση των αντίστοιχων συμπεριφορών των Ελλήνων και η σύγκριση της με τα αποτελέσματα της διεθνούς βιβλιογραφίας.

Η παρούσα διπλωματική εργασία έχει ως σκοπό λοιπόν τη διερεύνηση της συμπεριφοράς των χρηστών του οδικού δικτύου κατά την εκκένωση οικισμών λόγω πυρκαγιάς. Για τον σκοπό αυτό συλλέχθηκαν στοιχεία μέσω ενός κατάλληλα σχεδιασμένου ερωτηματολογίου το οποίο συμπληρώθηκε από κατοίκους περιοχών που επλήγησαν από τις πυρκαγιές της Ανατολικής Αττικής το 2018. Τα δεδομένα που αντλήθηκαν από τις ερωτήσεις του ερωτηματολογίου υποβλήθηκαν σε στατιστική ανάλυση, ώστε να εξαχθούν συμπεράσματα και να συγκριθούν με τα αντίστοιχα της διεθνούς βιβλιογραφίας.

1.3 Διάρθρωση διπλωματικής εργασίας

Στο παρόν υποκεφάλαιο παρουσιάζεται η δομή της Διπλωματικής Εργασίας, καθώς επίσης και μια συνοπτική αναφορά στο περιεχόμενο των κεφαλαίων της.

Στο πρώτο κεφάλαιο παρουσιάζεται η εισαγωγή της παρούσας διπλωματικής εργασίας. Εκεί γίνεται αναφορά στη σημασία της μελέτης των εκκενώσεων περιοχών λόγω φυσικών καταστροφών και των παραγόντων που τις επηρεάζουν. Επιπλέον παρουσιάζεται ο σκοπός της διπλωματικής εργασίας καθώς και η δομή της.

Στο δεύτερο κεφάλαιο παρουσιάζονται τα ευρήματα της βιβλιογραφικής ανασκόπησης που προέκυψαν από την αναζήτηση και καταγραφή ερευνών σχετικών με την εκκένωση οικισμών λόγω φυσικών καταστροφών και τις κρίσιμες αποφάσεις των χρηστών του οδικού δικτύου κατά τη διάρκεια των εκκενώσεων αυτών. Επιπλέον γίνεται αναφορά στη σχέση προσφοράς και ζήτησης των μετακινήσεων κατά την εκκένωση με τον απαιτούμενο χρόνο απομάκρυνσης.

Στο τρίτο κεφάλαιο παρουσιάζεται η μεθοδολογική προσέγγιση που ακολουθήθηκε. Αναλύεται η ροή των εργασιών, περιγράφεται η διαδικασία της συλλογής και κωδικοποίησης των δεδομένων και παρουσιάζεται το θεωρητικό υπόβαθρο που απαιτείται για την ανάλυση των δεδομένων αυτών στα επόμενα κεφάλαια. Συγκεκριμένα περιγράφεται η μέθοδος λογιστικής παλινδρόμησης, η μέθοδος δέντρων αποφάσεων και η μέθοδος τυχαίων δασών.

Στο τέταρτο κεφάλαιο παρουσιάζεται η ανάλυση της βάσης δεδομένων και τα αποτελέσματα αυτής. Αρχικά παρουσιάζονται ορισμένα ενδιαφέροντα στοιχεία που προέκυψαν από την έρευνα με τη βοήθεια της περιγραφικής στατιστικής και στη συνέχεια πραγματοποιείται η εκπαίδευση και αξιολόγηση των διαφόρων μοντέλων όπου έχουν σκοπό την πρόβλεψη της απόφασης για εκκένωση/ παραμονή καθώς και της απόφασης για επιλογή μέσου κατά την εκκένωση.

Τέλος, στο πέμπτο κεφάλαιο παρουσιάζονται τα συμπεράσματα που προέκυψαν από την ανάλυση που πραγματοποιήθηκε στο κεφάλαιο 4 σχετικά με την πρόβλεψη της απόφασης για εκκένωση/ παραμονή και την απόφαση για την επιλογή μέσου κατά την εκκένωση. Επιπλέον παρουσιάζονται ορισμένες προτάσεις για περεταίρω μελλοντική έρευνα.

2 ΒΙΒΛΙΟΓΡΑΦΙΚΗ ΑΝΑΣΚΟΠΗΣΗ

2.1 Εισαγωγή

Στο κεφάλαιο αυτό παρουσιάζονται τα αποτελέσματα ερευνών, σχετικών με το αντικείμενο που πραγματεύεται η παρούσα διπλωματική εργασία, τη μελέτη δηλαδή της συμπεριφοράς χρηστών οδικού δικτύου κατά την εκκένωση οικισμών λόγω πυρκαγιάς. Η αναζήτηση επικεντρώνεται στη διερεύνηση των παραγόντων που οδηγούν στην απόφαση για εκκένωση ή την παραμονή, καθώς και στην επιλογή του μέσου εκκένωσης. Στο τέλος της ενότητας θα παρουσιαστούν τα συμπεράσματα από την ανάλυση της βιβλιογραφίας καθώς και οι ελλείψεις που παρατηρήθηκαν, πάνω στις οποίες θα βασιστεί ο προσδιορισμός του στόχου της διπλωματικής εργασίας.

2.2 Χρόνος απομάκρυνσης

Σύμφωνα με τους Zhang et al. (2019), στόχος μιας εκκένωσης είναι η αποφυγή τραυματισμών, απωλειών ανθρωπίνων ζωών και, ενδεχομένως, φθορών και απωλειών περιουσίας. Επομένως, πρωταρχικός σκοπός τέτοιων ενεργειών είναι η μεταφορά των απομακρυνόμενων εκτός περιοχής απειλής κατά τον ασφαλέστερο και συντομότερο τρόπο.

Ο χρόνος που απαιτείται μέχρι και τη φυγή του τελευταίου ατόμου από τη ζώνη κινδύνου μετά την αναγνώριση μιας απειλής αναφέρεται συνήθως ως χρόνος απομάκρυνσης (Tu 2010, Hsu and Peeta 2015). Οι χρόνοι απομάκρυνσης διαφέρουν αρκετά ανάλογα με τα χαρακτηριστικά της απειλής, το μέγεθος και την αντίδραση του πληθυσμού που εκκενώνει, την τοπολογία του δικτύου, τις συνθήκες ταξιδιού και άλλων παραγόντων. Για απειλές που δίνουν αρκετό χρόνο για προειδοποίηση, ο χρόνος απομάκρυνσης μπορεί να χρησιμοποιηθεί και για τον σχεδιασμό, την οργάνωση και την προετοιμασία της εκκένωσης (Zhang et al. 2019).

2.3 Ζήτηση Μετακινήσεων

Σύμφωνα με τους Murray-Tuite and Wolshon (2013) μια έρευνα για τη ζήτηση των μετακινήσεων κατά τη διάρκεια της εκκένωσης μπορεί να χωριστεί σε 4 στάδια:

- Προσδιορισμός του «ποιοι» εκκενώνουν και των παραγόντων που επηρεάζουν την απόφαση για εκκένωση/παραμονή.
- Προσδιορισμός της γένεσης των μετακινήσεων λαμβάνοντας υπόψιν τον αριθμό των νοικοκυριών που εκκενώνουν, την χρονική κατανομή των μετακινήσεων, την ήδη υπάρχουσα κυκλοφορία, τις πιθανές ενδιάμεσες στάσεις των ανθρώπων που εκκενώνουν πριν τον τελικό προορισμό αλλά και κυκλοφορία λόγω εκκένωσης από άτομα που δεν είναι μέρος της υποχρεωτικής εκκένωσης (shadow evacuation).
- Προσδιορισμός της επιλογής προορισμού και της κατανομής στο δίκτυο για κάθε μετακίνηση
- Προσδιορισμός της επιλογής μέσου κατά την εκκένωση

2.3.1 Παράγοντες που επηρεάζουν την απόφαση για εκκένωση / παραμονή

Τα ερωτήματα σχετικά με το ποιος εκκενώνει και ποιοι παράγοντες επηρεάζουν την απόφαση εκκένωσης / παραμονής διερευνώνται συχνά, κυρίως στις κοινωνικές επιστήμες, αλλά πιο πρόσφατα και από μηχανικούς. Ο Πίνακας 2.1 παρουσιάζει μια περίληψη των παραγόντων που επηρεάζουν αυτήν την απόφαση για τους διάφορους τύπους απειλής.

Για την εκκένωση λόγω πυρκαγιάς, οι σημαντικότεροι παράγοντες που φαίνεται πως επηρεάζουν την απόφαση για εκκένωση είναι:

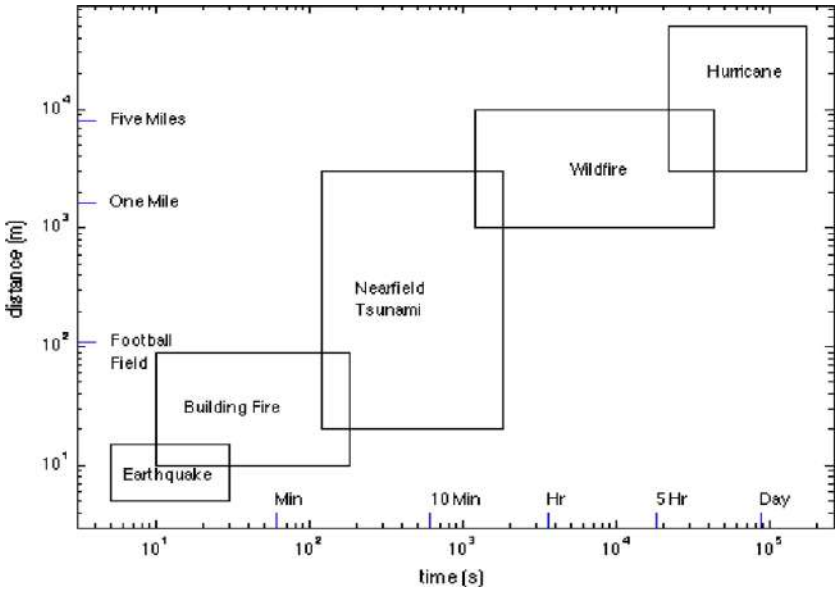
- η προηγούμενη/ πρόσφατη εμπειρία απειλής (αύξηση)
- η λήψη προειδοποίησης (αύξηση),
- η αντίληψη υψηλού κινδύνου (αύξηση),
- η ηλικία (αύξηση στις μεγαλύτερες ηλικίες), το φύλο (αύξηση στις γυναίκες),
- η ύπαρξη ανηλίκων στην κατοικία (αύξηση) και
- τα χρόνια διαμονής στην κατοικία (μείωση για περισσότερα χρόνια στην κατοικία).

Πίνακας 2.1 Παράγοντες που επηρεάζουν την απόφαση για εκκένωση ανά τύπο απειλής σύμφωνα με τη βιβλιογραφία

Παράγοντας	Επίδραση	Τυφώνας/ Καταιγίδα	Πλημμύρα	Σεισμός	Ηφαίστειο	Πυρκαγιά	Επικίνδυνα Υλικά	Πυρηνικό εργοστάσιο	Τρομοκρατική επίθεση
Προηγούμενη ή πρόσφατη εμπειρία απειλής	↑ και –	Arlikatti et al. 2006, Dow and Cutter 1998	Gruntfest (1977)			Mozumder et al. (2008)			
Συχνή εμπειρία απειλής	↓	Anderson (1968)	Anderson (1968)		Anderson (1968)				
Προηγούμενη εκκένωση	↑	Murray-Tuite et al., 2012, Riad et al., 1999							
Λήψη προειδοποίησης	↑	Baker, 1979, Baker,1991, Gladwin and Peacock, 1997	Perry (1981)	Bourque et al. (1971)	Perry (1981)	Fischer III et al. (1995)		Perry (1981)	
Αντίληψη υψηλού κινδύνου/ αντίκτυπου	↑	Arlikatti et al., 2006, Baker, 1979, Baker, 1991, Bateman and Edwards, 2002	Perry (1981)		Perry (1981)	Alsnih et al. (2015), Mozumder et al. (2008)		Houts et al., 1984, Perry, 1981	
Παρουσία τραυματισμών	↑								Aguirre et al. (1998)
Απόσταση από την απειλή (κοντινότερα)	↑	Carnegie and Deka (2010)		Bourque et al., 1971, Bourque and Russell, 1994				Cutter and Barnes, 1982, Houts et al., 1984	
Δομική ζημιά	↑			Bourque et al., 1971, Bourque and Russell, 1994					
Πρόσβαση σε καταφύγιο	↑	Baker, (1979)							
Ισχυρό κοινωνικό δίκτυο	↑	Baker (1979)	Anderson et al. (1984)						Aguirre et al. (1998)
Μεγαλύτερη ηλικία	Ανάμεικτη	Gladwin and Peacock, 1997, Riad et al., 1999				Alsnih et al. (2015)		Cutter and Barnes (1982)	Aguirre et al. (1998)
Φύλο (Γυναίκα)	↑	Bateman and Edwards, 2002, Riad et al., 1999				Alsnih et al. (2015), Mozumder et al. (2008)			Aguirre et al. (1998)
Παιδιά στην κατοικία	↑ και –	Baker, 1991, Gladwin and Peacock, 1997, Riad et al., 1999				Fischer III et al. (1995)		Houts et al., 1984, Perry, 1985	
Φυλή (Καυκάσιοι)	↑ και –	Bateman and Edwards, 2002, Gladwin and Peacock, 1997, Riad et al., 1999	Perry and Lindell, 1991, Perry et al., 1982				Perry and Lindell (1991)		
Εισόδημα (Υψηλότερο)	↑ και –	Elliott and Pais, 2006, Gladwin and Peacock, 1997	Perry and Lindell (1991)				Perry and Lindell (1991)		
Χρόνια διαμονής στην κατοικία (περισσότερα)	↓	Baker, 1979, Gladwin and Peacock, 1997				Alsnih et al. (2015)			
Ύπαρξη κατοικιδίων	↓ και ↑	Baker, 1991, Solís et al., 2010							
Απαιτήσεις εργασίας	↓	Baker (1979)						Perry (1983)	
Φόβος λεηλασίας	↓							Perry (1983)	

2.3.2 Επιλογή μέσου κατά την εκκένωση

Ο χρόνος και η διανυόμενη απόσταση κατά την εκκένωση διαφέρουν σημαντικά για τους διαφορετικούς τύπους απειλής (Wang et al. 2016). Για παράδειγμα, σε περίπτωση σεισμού, κάποιος μπορεί να καταφύγει κάτω από ένα γραφείο μέσα σε λίγα δευτερόλεπτα από την αίσθηση της έντονης ανατάραξης του κτιρίου. Από την άλλη πλευρά, οι εκκενώσεις τυφώνων έχουν συχνά προειδοποιήσεις αρκετές ώρες έως μέρες πριν το συμβάν και οι εκκενωτές βασίζονται σε οχήματα για να βρουν καταφύγιο αρκετά χιλιόμετρα μακριά από τη ζώνη κινδύνου (Wang et al. 2016).
Στο παρακάτω διάγραμμα (Διάγραμμα 2.1) οι διαφορές αυτές φαίνονται αναλυτικά.



Διάγραμμα 2.1 Κλίμακες χρόνου και απόστασης κατά την εκκένωση για τους διάφορους τύπους απειλής

Σημαντικότεροι παράγοντες λοιπόν για τον καθορισμό της επιλογής μέσου είναι το είδος της απειλής και η απαιτούμενη απόσταση μέχρι μια ασφαλή περιοχή. Εκτός ωστόσο από το είδος και τα χαρακτηριστικά της απειλής, σημαντικό ρόλο στην επιλογή μέσου κατά την εκκένωση έχει το μέρος στο οποίο βρίσκεται κάποιος τη στιγμή της εντολής εκκένωσης (π.χ. στον χώρο εργασίας, ή στην κατοικία), οι διαθέσιμες επιλογές του (π.χ. η ύπαρξη διαθέσιμου αυτοκινήτου) καθώς και η προθυμία του να οδηγήσει κάτω από τις συγκεκριμένες συνθήκες (Zhang, et al., 2019).

Σε έρευνα των Wu et al. (2012) για την εκκένωση λόγω των τυφώνων Katrina και Rita, βρέθηκε ότι μόνο το 11% όσων διέθετε αυτοκίνητο αποφάσισε να εκκενώσει με διαφορετικό τρόπο εκ των οποίων το 71% εκκένωσε με όχημα φίλου ή συγγενή και το 28% με άλλο τρόπο. Σύμφωνα με μία περίληψη των ευρημάτων από παλαιότερες έρευνες τους, αυτό το ποσοστό δεν ξεπερνάει το 13%.

Για νοικοκυριά με πρόσβαση σε τουλάχιστον ένα προσωπικό όχημα, προκύπτει το ερώτημα του πόσα οχήματα θα πάρει το κάθε νοικοκυριό και τι είδους οχήματα θα είναι αυτά. Σύμφωνα με έρευνες των Lindell and Prater (2007) στις Ηνωμένες Πολιτείες, ο αριθμός αυτός κυμαίνεται από 1.1 μέχρι 2.15.

2.4 Προσφορά

Σύμφωνα με τους Zhang, et al., (2019), η πλευρά της προσφοράς των διαδικασιών εκκένωσης περιλαμβάνει τους παράγοντες και τις μεταβλητές που σχετίζονται με τα μέσα και τη διαθεσιμότητα της κίνησης κατά τη διάρκεια έκτακτης ανάγκης. Αυτό περιλαμβάνει τον τρόπο ταξιδιού (με τα πόδια, όχημα κ.λπ.), αλλά είναι επίσης συνάρτηση της αντοχής και της τοπολογίας ενός δικτύου εκκένωσης. Οι βασικοί τοπολογικοί παράγοντες περιλαμβάνουν κυρίως τον αριθμό, τη θέση και την χωρητικότητα των σημείων εξόδου από την περιοχή απειλών καθώς και την εσωτερική δομή του δικτύου και τις δυνατότητες διαδρομής.

Παρόμοια με τις παραμέτρους ζήτησης που συζητήθηκαν προηγουμένως, οι παράμετροι χωρικής χωρητικότητας δεν είναι πάντα στατικές και μπορεί να διαφέρουν με την πάροδο του χρόνου. Για παράδειγμα, η χωρητικότητα σημείων εξόδου ή τμημάτων ταξιδιού εντός ενός συστήματος εκκένωσης μπορεί να μειωθεί λόγω αποκλεισμένων πορτών ή διαδρόμων σε ένα κτίριο ή λόγω τροχαίων ατυχημάτων ή πλημμυρών που εμποδίζουν ή κλείνουν λωρίδες αυτοκινητόδρομου. Οι συνεχώς μεταβαλλόμενες συνθήκες απειλής και οι τεχνικές διαχείρισης της κυκλοφορίας, μπορούν να επηρεάσουν την ώρα αναχώρησης, την τοποθεσία και τη φόρτωση σε ένα σύστημα. Η δυναμική φύση της προσφοράς και της ζήτησης εκκένωσης προσθέτει επιπλέον επίπεδα πολυπλοκότητας στον σχεδιασμό και τη διαχείριση εκκένωσης.

Κατά τους Murray-Tuite and Wolshon (2013), η συνηθέστερη στρατηγική διαχείρισης της κυκλοφορίας στο οδικό δίκτυο κατά τη διάρκεια μιας εκκένωσης είναι η επίτρεψη κυκλοφορίας σε μία ή περισσότερες λωρίδες του αντίθετου ρεύματος (Contraflow). Είναι μια πολύ αποτελεσματική στρατηγική, διότι μπορεί να αυξήσει άμεσα και σημαντικά την φέρουσα ικανότητα μιας οδού χωρίς τον χρόνο ή το κόστος που απαιτείται για τον σχεδιασμό και την κατασκευή πρόσθετων λωρίδων. Είναι επίσης δημοφιλής στο κοινό επειδή θεωρείται ως λογική αξιοποίηση της αχρησιμοποίητης χωρητικότητας των γειτονικών λωρίδων αντίθετης κατεύθυνσης.

2.5 Συμπεράσματα

Από τη σύνοψη των παραπάνω, προκύπτει το πλαίσιο στο οποίο θα πρέπει να πραγματοποιηθεί μία πλήρης διερεύνηση της εκκένωσης για μια περιοχή, καθώς και ο υπολογισμός του κρίσιμου χρόνου που απαιτείται για την ολοκλήρωσή της. Από τη διερεύνηση της προσφοράς προκύπτει ο ρυθμός εξυπηρέτησης του συστήματος ενώ από τη διερεύνηση της ζήτησης προκύπτει ο αντίστοιχος ρυθμός αφίξεων. Έτσι είναι δυνατός ο υπολογισμός του απαιτούμενου χρόνου εκκένωσης του πληθυσμού.

Ενώ η προσφορά εξαρτάται απόλυτα από το υπάρχον συγκοινωνιακό δίκτυο της εκάστοτε περιοχής και τη χωρητικότητά του, η ζήτηση των μετακινήσεων, εκτός από το μέγεθος και την χωρική κατανομή του πληθυσμού, εξαρτάται άμεσα και από τη συμπεριφορά των ανθρώπων και τις κρίσιμες αποφάσεις τους κατά τη διάρκεια της απειλής. Σημαντικότερες αποφάσεις είναι η απόφαση για εκκένωση ή παραμονή στην περιοχή και η απόφαση για την επιλογή του μέσου κατά την εκκένωση. Για τη μελέτη των αποφάσεων αυτών, έχουν πραγματοποιηθεί αρκετές έρευνες στο εξωτερικό, με τις περισσότερες αυτών να ασχολούνται με εκκενώσεις λόγω τυφώνα.

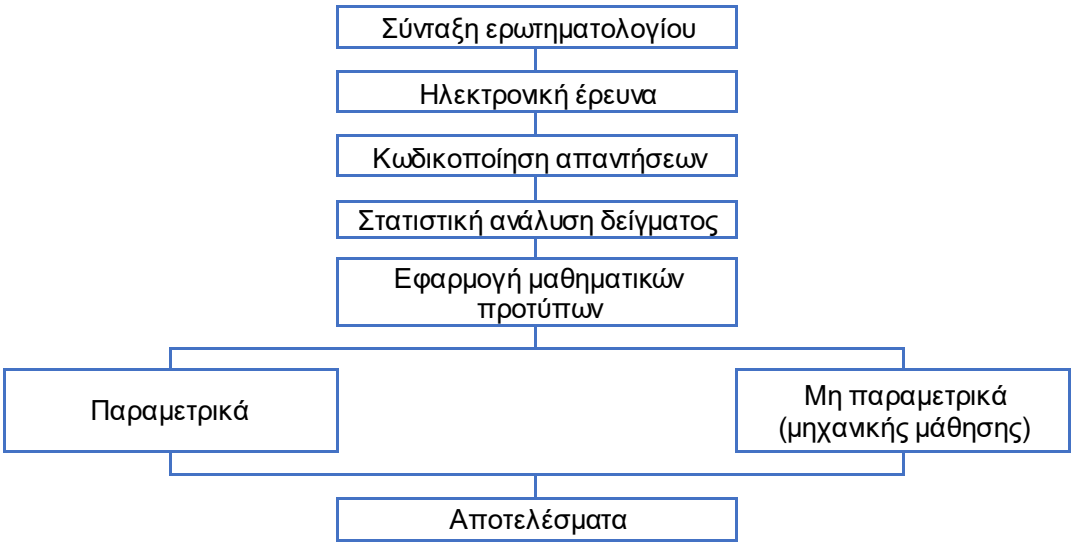
Τα ζητήματα επομένως που καλείται να καλύψει η παρούσα διπλωματική εργασία, είναι αφενός η διερεύνηση της συμπεριφοράς των Ελλήνων και η σύγκριση της με τα αποτελέσματα της διεθνούς βιβλιογραφίας και αφετέρου η εκτενέστερη διερεύνηση της επιλογής μέσου κατά την εκκένωση λόγω πυρκαγιάς.

3 ΜΕΘΟΔΟΛΟΓΙΚΗ ΠΡΟΣΕΓΓΙΣΗ

3.1 Περιγραφή της διαδικασίας

Στόχος της παρούσας Διπλωματικής Εργασίας είναι η διερεύνηση της **συμπεριφοράς των χρηστών του οδικού δικτύου κατά την εκκένωση οικισμών λόγω πυρκαγιάς**. Συγκεκριμένα, θα διερευνηθούν οι παράγοντες που επηρεάζουν την απόφαση για εκκένωση ή παραμονή και θα εξετασθούν τα κριτήρια της επιλογής μέσου κατά την εκκένωση.

Για τον σκοπό αυτό συλλέχθηκαν στοιχεία μέσω ενός κατάλληλα σχεδιασμένου **ερωτηματολογίου** το οποίο συμπληρώθηκε από κατοίκους περιοχών που επλήγησαν από τις πυρκαγιές της Ανατολικής Αττικής το 2018. Τα δεδομένα που αντλήθηκαν από τις ερωτήσεις του ερωτηματολογίου υποβλήθηκαν σε στατιστική ανάλυση, ώστε να εξαχθούν συμπεράσματα αλλά και να αξιολογηθεί η σημαντικότητά τους. Παρακάτω φαίνεται τη προτεινόμενη μεθοδολογική προσέγγιση (Διάγραμμα 3.1).



Διάγραμμα 3.1 Διάγραμμα ροής μεθοδολογικής προσέγγισης

3.2 Συλλογή δεδομένων

3.2.1 Το ερωτηματολόγιο

Το ερωτηματολόγιο χωρίζεται σε τρεις βασικές ενότητες καλύπτοντας συνολικά 40 ερωτήσεις και ο χρόνος συμπλήρωσης είναι κατά μέσο όρο 15 λεπτά. Το ερωτηματολόγιο παρατίθεται στο σύνολό του στο Παράρτημα Ι.

Ο πληθυσμός ο οποίος επιλέχθηκε για την έρευνα είναι οι κάτοικοι των περιοχών της Ανατολικής Αττικής που επλήγησαν από τις πυρκαγιές του 2018. Δεδομένου του μικρού μεγέθους του

πληθυσμού αυτού θεωρήθηκε πως το δείγμα 151 έγκυρων ερωτηματολογίων που συμπληρώθηκαν είναι επαρκές.

Η συλλογή του συνόλου των ερωτηματολογίων έγινε υπό μορφή διαδικτυακής έρευνας. Για την αντιπροσωπευτικότητα του δείγματος έγινε σύγκριση των δημογραφικών χαρακτηριστικών του δείγματος με τα αντίστοιχα χαρακτηριστικά του συνόλου του πληθυσμού με βάση τα στοιχεία της ΕΛΣΤΑΤ για την περιοχή. Ακολουθήθηκε η μέθοδος του χ^2-τεστ του Pearson για τον έλεγχο καλής προσαρμογής του δείγματος, με βάση τα προσδοκώμενα και τα παρατηρούμενα στο δείγμα χαρακτηριστικά.

Λόγω έλλειψης λεπτομερών δεδομένων από την ΕΛΣΤΑΤ δεν κατέστη δυνατός ο έλεγχος καλής προσαρμογής του δείγματος με βάση την ηλικία και το επίπεδο σπουδών.

Η πρώτη ενότητα του ερωτηματολογίου αποτελείται από ερωτήσεις που αφορούν περιγραφή των γεγονότων πριν και κατά τη διάρκεια της πυρκαγιάς. Πιο συγκεκριμένα υπάρχουν ερωτήσεις που αφορούν τον τρόπο με τον οποίο ο ερωτώμενος έλαβε ενημέρωση για την πυρκαγιά, τις κρίσιμες αποφάσεις που πήρε κατά τη διάρκεια της εκκένωσης αλλά και τα συναισθήματα και τις προσδοκίες που είχε τη στιγμή εκείνη.

Η δεύτερη ενότητα του ερωτηματολογίου αποτελείται από ερωτήσεις σχετικές με τις αποφάσεις του ερωτώμενου σε μια νέα ενδεχόμενη απειλή πυρκαγιάς ύστερα από την εμπειρία του, τα συναισθήματα αλλά και τις ενέργειες που έχει αναλάβει για την πρόληψη απέναντι στο ενδεχόμενο αυτό.

Τέλος, η τρίτη και τελευταία ενότητα του ερωτηματολογίου περιλαμβάνει ερωτήσεις σχετικά με τα δημογραφικά χαρακτηριστικά των ερωτηθέντων. Ζητούνται στοιχεία όπως είναι το φύλο, η ηλικία, το μορφωτικό επίπεδο, η οικογενειακή κατάσταση και το ετήσιο οικογενειακό εισόδημα. Η καταγραφή των παραπάνω κοινωνικοοικονομικών χαρακτηριστικών χρησιμεύει:

- Στον έλεγχο της αντιπροσωπευτικότητας του δείγματος.
- Στην εξαγωγή συμπερασμάτων ως προς τη συσχέτιση των δημογραφικών χαρακτηριστικών με τις υπόλοιπες απαντήσεις του ερωτηματολογίου.

3.2.2 Κωδικοποίηση ερωτηματολογίου

Οι ερωτήσεις του ερωτηματολογίου με διακριτές απαντήσεις κωδικοποιήθηκαν σε κατηγορικές μεταβλητές με διακριτές τιμές 0,1,2,…,k σε περιβάλλον Excel για ευκολότερη επεξεργασία κατά την ανάλυση. Όποιες απαντήσεις είχαν μηδενική ή πολύ μικρή συχνότητα, συγχωνεύτηκαν.

Επιπλέον, ερωτήσεις στις οποίες ο ερωτώμενος είχε δυνατότητα πολλαπλής επιλογής απαντήσεων (checkboxes) χωρίστηκαν σε επιμέρους dummy μεταβλητές με πιθανές απαντήσεις 0 και 1 που αντιστοιχούν στη συμπλήρωση ή μη του κάθε checkbox ξεχωριστά. Αυτό είχε ως αποτέλεσμα να μεγαλώσει το μέγεθος της βάσης δεδομένων από 151x40 σε 151x111.

Λόγω του μεγάλου αριθμού των μεταβλητών θεωρήθηκε σκόπιμο η ονομασία τους να γίνει με βάση τον αριθμό της αντίστοιχης ερώτησης του ερωτηματολογίου (π.χ. Q1, Q2a, Q2b κ.λπ.) εκτός των δημογραφικών χαρακτηριστικών όπου δόθηκε στις μεταβλητές ονομασία ανάλογη με την έννοια που εκπροσωπούν (π.χ. age, income, gender κ.λπ.).

3.3 Θεωρητικό Υπόβαθρο

3.3.1 Εισαγωγή

Σε αυτό το κεφάλαιο θα αναλυθεί το θεωρητικό υπόβαθρο που χρησιμοποιήθηκε για τη στατιστική ανάλυση της παρούσας διπλωματικής εργασίας. Χρησιμοποιήθηκαν μοντέλα λογιστικής παλινδρόμησης, δέντρων αποφάσεων και τυχαίων δασών, με σκοπό την πρόβλεψη της επιλογής μέσου από τους χρήστες. Εκτός από το θεωρητικό υπόβαθρο των μαθηματικών προτύπων γίνεται και αναφορά στον τρόπο αξιολόγησης και αποδοχής τους.

3.3.2 Μαθηματικά Πρότυπα

<u>Μοντέλο Λογιστικής Παλινδρόμησης</u>

Το μοντέλο της λογιστικής παλινδρόμησης (logistic regression) χρησιμοποιείται συχνά σε συγκοινωνιακές έρευνες, στις οποίες ζητείται η πρόβλεψη της επιρροής ορισμένων χαρακτηριστικών στην επιλογή κάποιου γεγονότος. Μέσω του μοντέλου αυτού αναπτύσσεται ένα μαθηματικό πρότυπο που δίνει μια γραμμική συνάρτηση χρησιμότητας του εν λόγω γεγονότος σε σχέση με τα χαρακτηριστικά που το επηρεάζουν. Έπειτα, μέσω κατάλληλου μετασχηματισμού υπολογίζεται η πιθανότητα πραγματοποίησης αυτού του γεγονότος. Η συνάρτηση χρησιμότητας της λογιστικής παλινδρόμησης δίνεται από τη σχέση:

$$\boldsymbol{Ui = a0 + a1x1 + a2x2 + \cdots + anxn} \qquad (3.1)$$

όπου:

- Ui, η συνάρτηση χρησιμότητας του γεγονότος i
- $x_1 \ldots x_n$, οι μεταβλητές του προβλήματος
- α_0, η σταθερά που αντιπροσωπεύει την επιρροή των παραγόντων που δεν έχουν συμπεριληφθεί ως μεταβλητές στο μαθηματικό μοντέλο
- $\alpha_1..\alpha_n$, οι συντελεστές των μεταβλητών

Η πιθανότητα να πραγματοποιηθεί το γεγονός i δίνεται από τη σχέση:

$$P_i = e^{Ui} / (1 + e^{Ui}) \qquad (3.2)$$

Εύκολα προκύπτει ότι η πιθανότητα να μην πραγματοποιηθεί το γεγονός i δίνεται από τη σχέση $1-P_i$.

Το μοντέλο της λογιστικής παλινδρόμησης μπορεί να χρησιμοποιηθεί τόσο στην ανάπτυξη του διωνυμικού προτύπου πρόβλεψης (binary model), όσο και του πολυωνυμικού προτύπου πρόβλεψης (multinomial model). Όταν οι πιθανές επιλογές είναι δύο, τότε επιλέγεται το πρώτο μοντέλο, ενώ όταν οι επιλογές είναι περισσότερες επιλέγεται το δεύτερο. Στην παρούσα Διπλωματική Εργασία χρησιμοποιείται το διωνυμικό μοντέλο.

Μια άλλη έννοια που αξίζει να αναλυθεί, είναι αυτή του λόγου συμπληρωματικών πιθανοτήτων (odds ratio). Πρόκειται για ένα κλάσμα στον αριθμητή του οποίου βρίσκεται η πιθανότητα να συμβεί το γεγονός και στον παρονομαστή η πιθανότητα να μην συμβεί. Αν, λοιπόν, P ορίσουμε τη πιθανότητα να συμβεί το γεγονός και 1-P την πιθανότητα να μην συμβεί, τότε η αναλογία είναι P/(1-P). Αυτός ο λόγος χρησιμοποιείται κυρίως σε λογαριθμική μορφή ως εξής:

$$logit(P) = ln\frac{P}{1-P} = \beta 0 + \beta 1\chi 1 + \cdots + \beta \nu \chi \nu \qquad (3.3)$$

Για παράδειγμα, κατά τη ρίψη ενός νομίσματος, η πιθανότητα το αποτέλεσμα να είναι «κορώνα» είναι 50%. Επομένως, ο λόγος συμπληρωματικών πιθανοτήτων να συμβεί το ενδεχόμενο «κορώνα» ισούται με 0.5/(1-0.5)=0.5/0.5=1. Γενικά:

- όταν ο λόγος γίνεται μεγαλύτερος της μονάδας η πιθανότητα να συμβεί το γεγονός αυξάνεται
- όταν ο λόγος γίνεται μικρότερος της μονάδας η πιθανότητα να συμβεί το γεγονός μειώνεται

Μοντέλο Δέντρων Αποφάσεων

Μια από τις πιο διαδεδομένες μεθόδους αυτόματης εξαγωγής λογικών συμπερασμάτων και κανόνων μέσα από δεδομένα αποτελεί αυτή των δένδρων αποφάσεων. Σε μαθηματικούς όρους ένα δένδρο αποφάσεων είναι μια ιεραρχημένη συλλογή σύνθετων διαζευκτικών προτάσεων οι οποίες αποτελούνται από ένα σύνολο λογικών συζεύξεων που αναφέρονται σε τιμές χαρακτηριστικών συγκεκριμένων παραδειγμάτων. Τα δέντρα απόφασης είναι προϊόντα της μη καθοδηγούμενης εκμάθησης. Ένα δέντρο απόφασης επάγεται από ένα σύνολο εκμάθησης, που αποτελείται από αντικείμενα. Κάθε αντικείμενο περιγράφεται πλήρως από ένα σύνολο ιδιοτήτων και από μια ετικέτα που ως έχουμε αναφέρει είναι το χαρακτηριστικό εξόδου. Οι ιδιότητες μπορούν να έχουν συνεχόμενες τιμές ή να είναι διακριτές. Ο στόχος ενός αλγόριθμου μηχανικής μάθησης αναφορικά με την κατασκευή ενός μοντέλου από ένα σύνολο δεδομένων, είναι η εύρεση ή η μεγαλύτερη δυνατή προσέγγιση της πραγματικής αντιστοίχισης ανάμεσα στο σύνολο των ιδιοτήτων και στο χαρακτηριστικό εξόδου.

Ένα δένδρο απόφασης περιέχει έναν ή περισσότερους εσωτερικούς κόμβους και έναν ή περισσότερους κόμβους - φύλλα. Όλοι οι εσωτερικοί κόμβοι έχουν δύο περισσότερους κόμβους–παιδιά. Όλοι οι μη τερματικοί κόμβοι περιέχουν τεμαχισμούς στους οποίους ελέγχεται η τιμή μιας μαθηματικής ή μιας λογικής έκφρασης των ιδιοτήτων. Μόλις χτιστεί το δέντρο, εφαρμόζεται σε κάθε μια εγγραφή στη βάση δεδομένων και καταλήγει σε μια κατηγοριοποίηση για κάθε πλειάδα . Υπάρχουν δύο βασικά βήματα σε αυτή την τεχνική: η κατασκευή του δέντρου και η εφαρμογή του στη βάση δεδομένων. Η περισσότερη έρευνα έχει επικεντρωθεί στον τρόπο κτίσης των δέντρων, καθώς η διαδικασία εφαρμογής είναι αρκετά προφανής. Ένας ορισμός για ένα δένδρο απόφασης που χρησιμοποιείται στην κατηγοριοποίηση είναι ο παρακάτω:

Δίνεται μία βάση δεδομένων D={t1,t2, ., tn} και το σχήμα αυτής περιέχει τα ακόλουθα γνωρίσματα: {A1, A2… Ah}. Επίσης δίνεται το σύνολο από τα πεδία C= {C1, C2,…, Cn}. Ένα δέντρο απόφασης είναι ένα δέντρο που σχετίζεται με το D και έχει τις ακόλουθες ιδιότητες:

- Κάθε εσωτερικός κόμβος παίρνει το όνομά του από ένα γνώρισμα, Ai.
- Κάθε τόξο παίρνει το όνομά του από ένα κατηγόρημα το οποίο μπορεί να εφαρμοστεί στο γνώρισμα που συνδέεται με τον πατέρα κόμβο.
- Κάθε φύλλο έχει ως όνομα μια κατηγορία Cj.

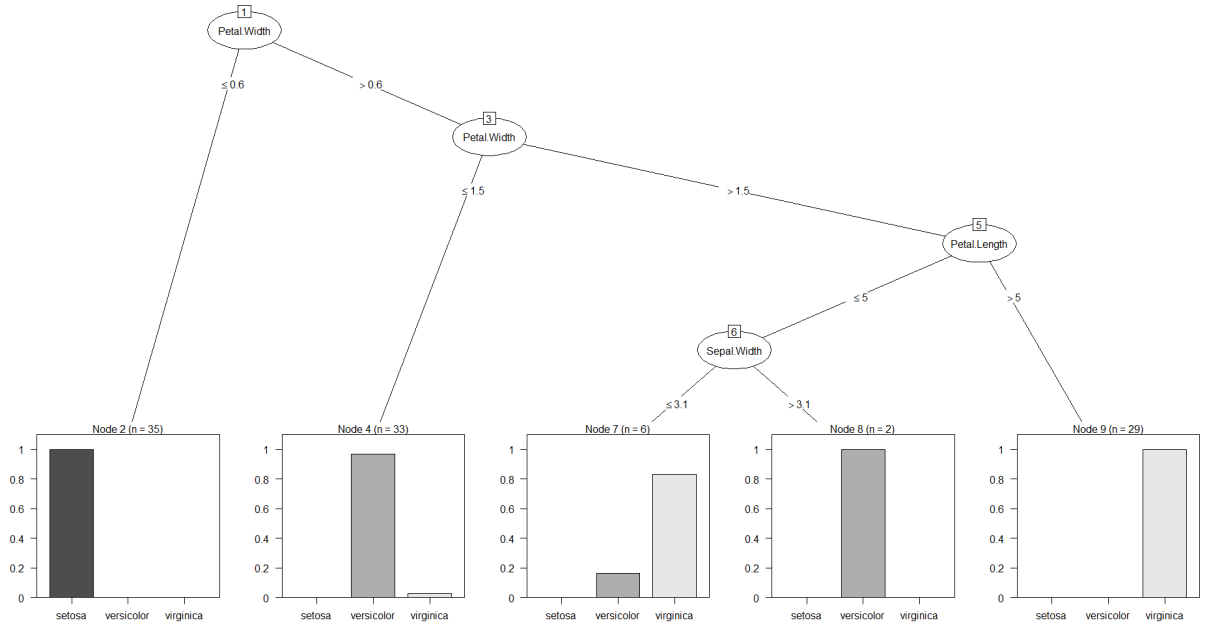
Η λύση του προβλήματος της κατηγοριοποίησης με αποφάσεων, είναι μια διαδικασία δύο βημάτων:

1. Επαγωγή δέντρου απόφασης: Η δημιουργία ενός δέντρου απόφασης χρησιμοποιώντας δεδομένα εκπαίδευσης

2. Για κάθε ti που ανήκει στο D, εφαρμογή του δέντρου απόφασης για τον προσδιορισμό της κατηγορίας του.

Το δέντρο απόφασης ορίζει έμμεσα την απεικόνιση της κατηγοριοποίησης. Διαφορετικό δέντρο απόφασης αποδίδει διαφορετική κατηγοριοποίηση. Υπάρχουν πολλά πλεονεκτήματα από την χρήση δέντρων αποφάσεων για την κατηγοριοποίηση . Τα δέντρα είναι εύκολα στην χρήση και αποτελεσματικά. Μπορούν να δημιουργηθούν κανόνες οι οποίοι είναι εύκολοι στο να κατανοηθούν και να ερμηνευτούν. Τα δέντρα αποφάσεων αποδίδουν καλά για μεγάλες βάσεις δεδομένων. Κάθε πλειάδα της βάσης δεδομένων πρέπει να φιλτραριστεί μέσα από το δέντρο. Αυτό παίρνει χρόνο ανάλογο με το ύψος του δέντρου, το οποίο είναι συγκεκριμένο. Μπορούμε να κατασκευάσουμε δέντρα για δεδομένα με κόλλα γνωρίσματα.

Υπάρχουν όμως και μειονεκτήματα για τους αλγορίθμους δέντρων αποφάσεων. Πρώτον, δεν χειρίζονται εύκολα συνεχή δεδομένα. Αυτά τα πεδία των γνωρισμάτων θα πρέπει να χωριστούν σε κατηγορίες για να μπορέσει να τα χειριστεί το δέντρο. Η προσέγγιση που χρησιμοποιείται είναι ότι ο χώρος του πεδίου διαιρείται σε ορθογώνιες περιοχές. Βέβαια δεν είναι όλα τα προβλήματα κατηγοριοποίησης αυτού του τύπου. Επίσης, ο χειρισμός που γίνεται στα ελλιπή δεδομένα είναι δύσκολος, γιατί δεν μπορούν να βρεθούν σωστές διακλαδώσεις του δέντρου για να ακολουθηθούν. Επειδή το δέντρο απόφασης δημιουργείται από τα δεδομένα εκπαίδευσης, μπορεί να εμφανιστεί υπερπροσαρμογή (overfitting). Αυτό θα μπορούσαμε να το προσπεράσουμε με την φάση περικοπής του δέντρου. Τέλος, οι συσχετίσεις μεταξύ των γνωρισμάτων της βάσης δεδομένων αγνοούνται από η διαδικασία του δέντρου απόφασης. Μια από τις πρώτες εργασίες στον τομέα αποτέλεσε το σύστημα CLS (Concept Learning System) του Hunt (1962). Σήμερα ένα πλήθος αλγορίθμων υλοποιούν τα δέντρα αποφάσεων, αντιπροσωπευτικότεροι από τους οποίους είναι οι ID3 (Quinlan, 1979), C4.5 (Quinlan, 1993), CART (Friedman, 1977), (Brieman et al. 1984), CHAID, C5.0 και Quest.



Διάγραμμα 3.2 Παράδειγμα δέντρου απόφασης στο dataset "Iris" του UCI Machine Learning Repository. Εξαρτημένη μεταβλητή είναι το είδος του φυτού και ανεξάρτητες μεταβλητές τα χαρακτηριστικά των φύλλων.

<u>Μοντέλο Τυχαίων Δασών</u>

Το μοντέλο τυχαίων δασών (Breiman 2001) δημιουργεί τυχαία δάση που αποτελούνται από πολλά δέντρα αποφάσεων. Ο αλγόριθμος που εφαρμόζεται για τη δημιουργία ενός τυχαίου δάσους είναι ο εξής:

- Δημιουργία κάθε δέντρου από ένα ανεξάρτητο δείγμα εκκίνησης (bootstrap sample), το οποίο επιλέγεται από τα δεδομένα εκπαίδευσης (training data) με τη μέθοδο της αντικατάστασης. Σε κάθε κόμβο γίνεται:
    - Επιλογή m μεταβλητών τυχαία από όλες τις δυνατές M μεταβλητές
    - Εύρεση βέλτιστου διαχωρισμού των m μεταβλητών
- Ανάπτυξη των δέντρων κατηγοριοποιώντας (ταξινομώντας) τα δεδομένα στο μέγιστο βαθμό.
- Κατάταξη των δέντρων ώστε να προβλεφθεί η συμπεριφορά των καινούργιων δέντρων.
- Κατάταξη των εναπομεινάντων (M-m) δεδομένων, τα οποία ονομάζονται " out of bag" ή "oob", σε κλάσεις των δέντρων του δάσους.
- Εξετάζεται αν η κλάση που υπερέχει ως επιλογή από τα περισσότερα δέντρα απόφασης είναι η πραγματική κλάση του κάθε δεδομένου εισαγωγής. Ο ρυθμός σφάλματος αυτής της κατηγοριοποίησης (oob error rate) συνιστά το ρυθμό σφάλματος πρόγνωσης του δάσους.

Τα πλεονεκτήματα που προσφέρει η χρήση του μοντέλου τυχαίων δασών είναι πολλά:

- Μπορεί να χρησιμοποιηθεί τόσο για ταξινομήσεις όσο και για διαδικασίες παλινδρόμησης.
- Λειτουργεί καλά τόσο με κατηγορικές όσο και με συνεχείς μεταβλητές.
- Παρέχει υψηλή ακρίβεια
- Το σφάλμα γενίκευσης είναι αρκετά περιορισμένο από τη στιγμή που αναπτύσσεται ένας πολύ μεγάλος αριθμός δέντρων με αποτέλεσμα να είναι απίθανο να εμφανιστεί πρόβλημα υπέρ-εκπαίδευσης (over fitting).

- Η τυχαία επιλογή των μεταβλητών πρόβλεψης μειώνει τη σχέση των μεγάλων και un-pruned δέντρων, κάτι που κάνει τη μέθοδο αρκετά αμερόληπτη.
- Δεν απαιτεί κλιμάκωση χαρακτηριστικών (τυποποίηση και κανονικοποίηση) καθώς χρησιμοποιεί προσέγγιση κατά κανόνα για τον υπολογισμό της απόστασης.
- Αντιμετωπίζει αποτελεσματικά τις γραμμικές παραμέτρους.
- Είναι ανθεκτικό στα ακραία σημεία και μπορεί να τα χειριστεί αυτόματα.
- Έχει τη δυνατότητα να χειριστεί τιμές που ενδεχομένως να λείπουν και να διατηρήσει την ακρίβεια ενός μεγάλου ποσοστού δεδομένων.
- Έχει τη δυνατότητα να χειρίζεται μεγάλο σύνολο δεδομένων.
- Είναι πολύ σταθερό. Ακόμα και αν εισαχθεί ένα νέο στοιχείο στη βάση δεδομένων, ο συνολικός αλγόριθμος δεν επηρεάζεται πολύ, καθώς το νέο δεδομένο ενδέχεται να επηρεάσει ένα δέντρο, αλλά είναι πολύ δύσκολο να επηρεάσει όλα τα δέντρα.

3.3.3 Κριτήρια αποδοχής μοντέλων

<u>Βασικά κριτήρια ελέγχου λογιστικού μοντέλου</u>

Παρακάτω αναφέρονται τα βασικά κριτήρια ελέγχου για την αξιολόγηση και την αποδοχή των μοντέλων. Απαραίτητη προϋπόθεση είναι ο έλεγχος της συσχέτισης μεταξύ των μεταβλητών, δηλαδή οι ανεξάρτητες μεταβλητές πρέπει να είναι γραμμικώς ανεξάρτητες μεταξύ τους.

Λογική εξήγηση συντελεστών μοντέλου

Στην εξίσωση που θα προκύψει από τα μοντέλα εξετάζεται αν τα πρόσημα των συντελεστών παλινδρόμησης (βi) έχουν λογική ερμηνεία. Γίνεται, δηλαδή, έλεγχος βάσει του πρόσημου των για το αν η εξαρτημένη μεταβλητή αναμένεται να αυξηθεί ή να μειωθεί αν το πρόσημο των συντελεστών είναι θετικό ή αρνητικό αντίστοιχα. Σε περίπτωση που τα πρόσημα αυτά δεν έχουν λογική ερμηνεία, η αντίστοιχη μεταβλητή θα απορριφθεί.

Στατιστική σημαντικότητα:

Για την επιλογή ενός μοντέλου προσδιορίζεται το επίπεδο εμπιστοσύνης, το οποίο πρέπει να έχει υψηλή τιμή. Για τα λογιστικά μοντέλα γίνεται ο έλεγχος Wald test (z-test), με τον εξής τύπο:

$$zi = \frac{\beta i}{s\beta i} \qquad (3.4)$$

όπου:

βi: οι συντελεστές παλινδρόμησης των ανεξάρτητων μεταβλητών xi

sβi: το τυπικό σφάλμα των συντελεστών παλινδρόμησης βi

Ενδεικτικές τιμές του συντελεστή z είναι για 95% επίπεδο εμπιστοσύνης 1.7 και για 90% επίπεδο εμπιστοσύνης 1.3.

Μήτρα σύγχυσης (Confusion Matrix)

Η μήτρα σύγχυσης χρησιμοποιείται για τη μέτρηση της απόδοσης ενός συστήματος για δύο ή περισσότερες κλάσεις. Η αξιολόγηση ενός μοντέλου κατηγοριοποίησης περιλαμβάνει την εξέταση της απόδοσής του σε ένα σύνολο δεδομένων που συνήθως είναι διαφορετικό από αυτό της εκπαίδευσής του. Για καθένα από τα στιγμιότυπα της βάσης εξέτασης (test), το μοντέλο κατηγοριοποίησης θα είναι είτε σωστό (θα του αναθέσει την προβλεπόμενη κλάση) είτε εσφαλμένο. Βάσει αυτού, συμπεραίνεται πως διακρίνοντας το σύνολο των στιγμιότυπων που ο ταξινομητής κατηγοριοποίησε σωστά ή εσφαλμένα, μπορεί να γίνει μια πρώτη εκτίμηση σχετικά με την απόδοση του εξεταζόμενου ταξινομητή. Σε μια βάση δεδομένων με δύο μόνο κλάσεις (θετική/αρνητική) ορίζονται τέσσερις περιπτώσεις κατηγοριοποίησης των πλειάδων της βάσης οι οποίες είναι οι εξής:

1. **Αληθώς Θετικά (True Positives – TP):** Το πλήθος των στιγμιότυπων της βάσης (+) που κατηγοριοποιήθηκαν ως (+) από τον ταξινομητή.
2. **Αληθώς Αρνητικά (True Negative – TN):** Το πλήθος των στιγμιότυπων που ανήκουν στην κλάση (-) και ο ταξινομητής κατηγοριοποίησε ως (-).
3. **Ψευδώς Θετικά (False Positive – FP):** Είναι το πλήθος των παραδειγμάτων της κλάσης (-) που εσφαλμένα ο ταξινομητής κατηγοριοποίησε ως (+).
4. **Ψευδώς Αρνητικά (False Negative – FN):** Είναι το πλήθος των παραδειγμάτων της κλάσης (+) που εσφαλμένα κατηγοριοποιήθηκαν από τον ταξινομητή ως (-).

Πίνακας 3.1 Μήτρα Σύγχυσης

Confusion matrix	Actual Positive	Actual Negative
Predicted Positive	TP	FN
Predicted Negative	FP	TN

Ορθότητα (accuracy)

Ορίζεται ως η συνολική ακρίβεια ή το ποσοστό των σωστών προβλέψεων του μοντέλου:

$$accuracy = \frac{TP + TN}{TP + TN + FP + FN} \tag{3.5}$$

Αντί της ορθότητας μπορεί να χρησιμοποιηθεί το μέτρο του λόγου σφάλματος (error rate) ή λόγος εσφαλμένων κατηγοριοποιήσεων (misclassification rate), το οποίο κατ' αντιστοιχία εκφράζει το βαθμό εσφαλμένων κατηγοριοποιήσεων του ταξινομητή:

$$error\ rate\ = 1 - \ accuracy \tag{3.6}$$

Στατιστικός Συντελεστής Κάππα (Kappa Statistic)

Ο στατιστικός συντελεστής Κάππα αποτελεί μέτρο αξιολόγησης εξεταζόμενου μοντέλου κατηγοριοποίησης:

$$Kappa\ Statistic = \frac{P(A) - P(E)}{1 - P(E)} \tag{3.7}$$

όπου:

P(A): η παρατηρούμενη σχετική συμφωνία μεταξύ των μοντέλων κατηγοριοποίησης

P(E): η πιθανότητα η συμφωνία αυτή να οφείλεται σε τυχαίο παράγοντα

Ευαισθησία και Εξειδικευτικότητα (Sensitivity and Specificity)

Η ευαισθησία ή ανάκληση (sensitivity ή recall) εκτιμά την ικανότητα του ταξινομητή να κατηγοριοποιήσει σωστά τα θετικά στιγμιότυπα της βάσης εξέτασης, ενώ η εξειδικευτικότητα (specificity) αξιολογεί την απόδοση του μοντέλου για την κατηγοριοποίηση αρνητικών στιγμιότυπων:

$$sensitivity \text{ ή } recall = \frac{TP}{TP + FN} \tag{3.8}$$

$$specificity = \frac{TN}{TN + FP} \tag{3.9}$$

Ακρίβεια (Precision)

Το μέτρο της ακρίβειας (precision) εκφράζει το βαθμό πιστότητας της διαδικασίας κατηγοριοποίησης:

$$precision = \frac{TP}{TP + FP} \tag{3.10}$$

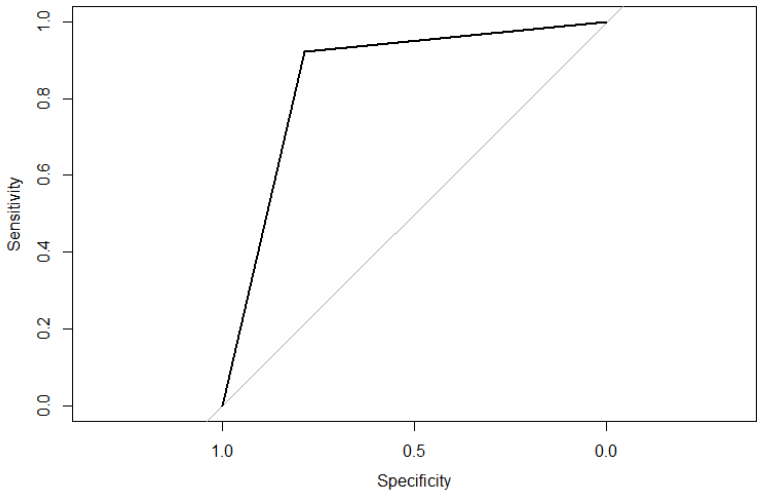
Μέτρο F (F-measure)

Το μέτρο F εκφράζει τον αρμονικό μέσο της ακρίβειας και της ανάκλησης συνδυάζοντας τα δύο αυτά μέτρα με τη σχέση:

$$F\text{-}measure = \frac{2 * \text{precision} * \text{recall}}{\text{precision} + \text{recall}} \tag{3.11}$$

Καμπύλη Receiver Operating Characteristic (ROC Curve)

Η καμπύλη ROC είναι μια γραφική παράσταση που απεικονίζει τη διαγνωστική ικανότητα ενός δυαδικού συστήματος ταξινόμησης καθώς το όριο διάκρισής του ποικίλει. Η καμπύλη δημιουργείται από την απεικόνιση του True Positive Rate (sensitivity) προς το False Positive Rate (1 - specificity) σε διάφορες τιμές στιγμιότυπων (Διάγραμμα 3.2). Η ιδανική περίπτωση θα ήταν μια καμπύλη ROC που θα ταυτίζεται με τον κατακόρυφο άξονα.



Διάγραμμα 3.3 Παράδειγμα χαρακτηριστικής καμπύλης λειτουργίας δέκτη (ROC curve)

Ένα ακόμα κριτήριο επιλογής κατάλληλου ταξινομητή με βάση τις χαρακτηριστικές καμπύλες λειτουργίας είναι το εμβαδόν κάτω από την καμπύλη ROC (Area Under the ROC Curve – AUC). Στην ιδανική περίπτωση AUC=1 και στην περίπτωση τυχαίας κατηγοριοποίησης (καμπύλη τυχαίας επίδοσης του σχήματος 3.1) ισχύει ότι AUC=0.5.

4 ΑΝΑΛΥΣΗ ΚΑΙ ΑΠΟΤΕΛΕΣΜΑΤΑ

4.1 Εισαγωγή

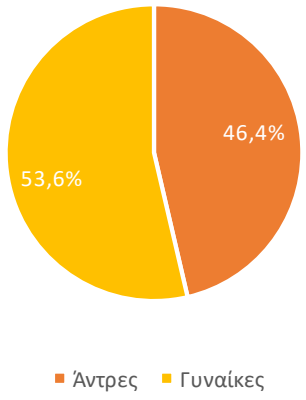
Στο κεφάλαιο αυτό παρουσιάζεται η περιγραφή της βάσης δεδομένων και η ανάλυση της μέσω των μαθηματικών προτύπων, όπως αυτά περιγράφονται στο κεφάλαιο 3. Στόχος της ανάλυσης είναι η μελέτη της συμπεριφοράς των χρηστών του οδικού δικτύου κατά την εκκένωση οικισμών λόγω πυρκαγιάς και συγκεκριμένα η μελέτη της απόφασης για εκκένωση η μη καθώς και της απόφασης για την επιλογή μέσου κατά την εκκένωση.

Για το σκοπό αυτό αξιοποιήθηκαν τα δεδομένα που συλλέχθηκαν από το ερωτηματολόγιο πάνω στα οποία εφαρμόστηκαν μοντέλα διωνυμικής λογιστικής παλινδρόμησης, μοντέλα δέντρων αποφάσεων και μοντέλα τυχαίων δασών σε περιβάλλον R Studio. Τέλος, παρατίθενται τα αποτελέσματα που προκύπτουν από την εφαρμογή των μεθοδολογιών, η περιγραφή και η ερμηνεία τους. Η παρουσίαση των αποτελεσμάτων περιλαμβάνει τόσο τη μαθηματική σχέση του κάθε προτύπου, όσο και σχετικά διαγράμματα που επιτρέπουν τη γραφική απεικόνιση των αποτελεσμάτων, όπου αυτό είναι εφικτό.

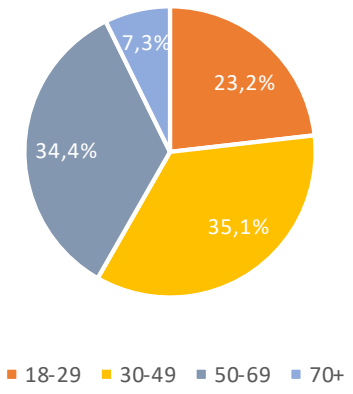
4.2 Περιγραφή βάσης δεδομένων

4.2.1 Συγκεντρωτικά στοιχεία

Σε αυτό το κεφάλαιο παρατίθενται συγκεντρωτικά με τη μορφή διαγραμμάτων τα δημογραφικά χαρακτηριστικά του δείγματος της έρευνας. Συγκεκριμένα, στα Διαγράμματα 4.1 και 4.2 φαίνεται η κατανομή του δείγματος ανά φύλο και ηλικία.

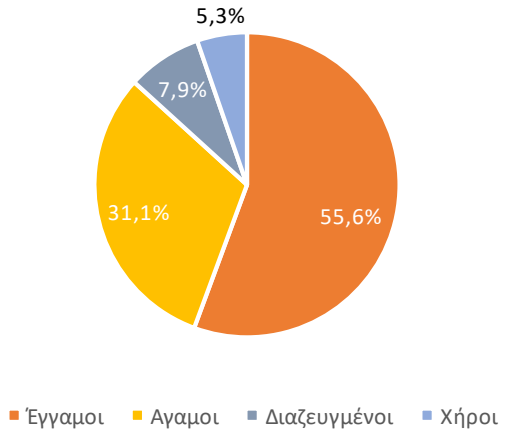


Διάγραμμα 4.1 Κατανομή ερωτηθέντων με βάση το φύλο

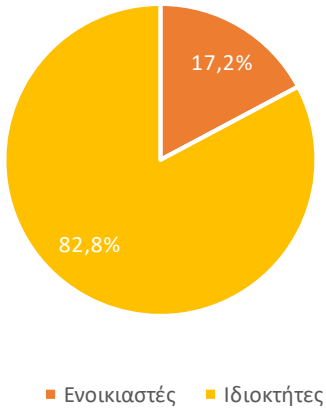


Διάγραμμα 4.2 Κατανομή ερωτηθέντων με βάση την ηλικία

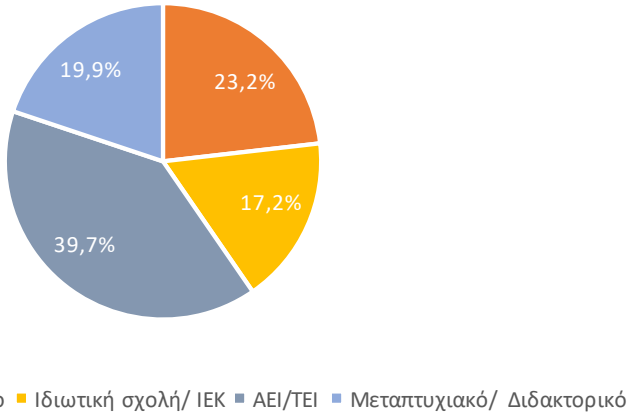
Στα Διαγράμματα 4.3, 4.4, 4.5 και 4.6 φαίνεται η κατανομή του δείγματος ανά οικογενειακή κατάσταση, τύπο ιδιοκτησίας κατοικίας, επίπεδο εκπαίδευσης και εισόδημα αντίστοιχα.



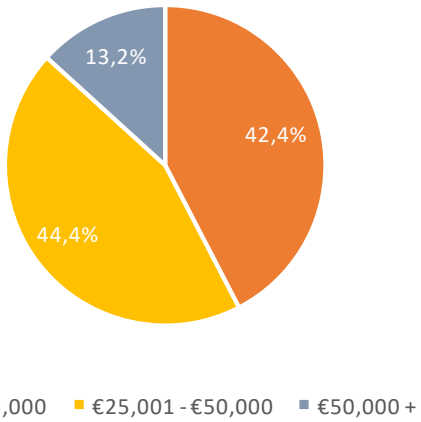
Διάγραμμα 4.3 Ποσοστιαία κατανομή ανάλογα με την οικογενειακή κατάσταση των ερωτηθέντων



Διάγραμμα 4.4 Ποσοστιαία κατανομή ανάλογα με τον τύπο ιδιοκτησίας των κατοικιών των ερωτηθέντων



Διάγραμμα 4.5 Ποσοστιαία κατανομή με βάση το επίπεδο εκπαίδευσης των ερωτηθέντων



Διάγραμμα 4.6 Ποσοστιαία κατανομή με βάση το οικογενειακό εισόδημα των ερωτηθέντων

4.2.2 Αντιπροσωπευτικότητα δείγματος

Για τη διερεύνηση της αντιπροσωπευτικότητας του δείγματος επιλέχθηκε η μέθοδος του ελέγχου καλής προσαρμογής χ^2 του Pearson. Από τα στοιχεία της ΕΛΣΤΑΤ προέκυψαν οι αναλογίες του πληθυσμού για τα επιμέρους δημογραφικά χαρακτηριστικά που εξετάστηκαν. Στη συνέχεια από τις αναλογίες αυτές προβλέφθηκαν οι συχνότητες των αντίστοιχων χαρακτηριστικών για δεδομένο πληθυσμό δείγματος και έγινε σύγκριση με τις παρατηρούμενες συχνότητες. Για τον υπολογισμό της στατιστικής συνάρτησης ελέγχου χρησιμοποιείται ο παρακάτω τύπος:

$$x^2 = \sum_{i=1}^{\kappa} \frac{(Oi - Ei)^2}{Ei} \qquad (4.1)$$

Όπου:

Oi: Οι παρατηρηθείσες συχνότητες (observed frequencies)

Ei: Οι προβλεπόμενες συχνότητες (expected frequencies)

Η τιμή αυτή συγκρίνεται με την κρίσιμη τιμή ελέγχου για δεδομένο επίπεδο σημαντικότητας α και βαθμούς ελευθερίας df από τον στατιστικό πίνακα κατανομής χ^2 (βλ. Παράρτημα). Όταν η τιμή χ^2 υπερβεί την κρίσιμη τιμή ελέγχου τότε βρίσκεται στην απορριπτική περιοχή ελέγχου που σημαίνει πως η μηδενική υπόθεση απορρίπτεται. Αν η τιμή είναι μικρότερη από την κρίσιμη τιμή τότε δεν υπάρχουν στατιστικά σημαντικές αποδείξεις εναντίον της μηδενικής υπόθεσης.

Σύμφωνα με τα στοιχεία της ΕΛΣΤΑΤ για τους δήμους Ραφήνας και Μαραθώνα ο πληθυσμός της περιοχής αποτελείται κατά 53.1% από άντρες και κατά 46.9% από γυναίκες. Έτσι, σε τυχαίο δείγμα 151 ατόμων η προβλεπόμενη κατανομή συχνοτήτων είναι 80 άντρες και 71 γυναίκες. Οι παρατηρούμενες όμως συχνότητες είναι 70 άντρες και 81 γυναίκες επομένως ήταν αναγκαίο να πραγματοποιηθεί έλεγχος καλής προσαρμογής του δείγματος όπως φαίνεται στον Πίνακα 4.1.

Πίνακας 4.1 Έλεγχος καλής προσαρμογής του δείγματος ως προς το φύλο

χ^2 test

Φύλο	Άντρας	Γυναίκα	Σύνολο
Παρατηρήσεις	70	81	151
Προβλέψεις	80	71	

α	0,05
p.value	0,098
df	1
χ^2	2,743
critical Value	3,841
	accept null

Μηδενική Υπόθεση:
Δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των παρατηρούμενων και των προβλεπόμενων τιμών.

Σύμφωνα με τα αποτελέσματα του ελέγχου, δεν υπάρχουν αποδείξεις που να υποστηρίζουν στατιστικά σημαντική απόκλιση του δείγματος από τον πληθυσμό ως προς το φύλο, επομένως το δείγμα γίνεται δεκτό.

Σύμφωνα με τα στοιχεία της ΕΛΣΤΑΤ για τους δήμους Ραφήνας και Μαραθώνα ο τύπος κυριότητας των ακινήτων της περιοχής είναι κατά 78.6% ιδιοκατοικούμενες κατοικίες και κατά 21.4% ενοικιαζόμενες ή άλλου τύπου κατοικίες. Έτσι, σε τυχαίο δείγμα 151 ατόμων η προβλεπόμενη κατανομή συχνοτήτων είναι 119 ιδιόκτητες και 32 ενοικιαζόμενες κατοικίες. Οι παρατηρούμενες όμως συχνότητες είναι 125 ιδιόκτητες και 26 ενοικιαζόμενες κατοικίες επομένως ήταν αναγκαίο να πραγματοποιηθεί έλεγχος καλής προσαρμογής του δείγματος όπως φαίνεται στον Πίνακα 4.2.

Πίνακας 4.2 Έλεγχος καλής προσαρμογής του δείγματος ως προς τον τύπο κυριότητας της κατοικίας

χ^2 test

Τύπος κυριότητας	Ιδιοκτήτης	Ενοικιαστής	Σύνολο
Παρατηρήσεις	125	26	151
Προβλέψεις	119	32	

α	0,05
p.value	0,211
df	1
x^2	1,567
critical Value	3,841
	accept null

Μηδενική Υπόθεση:
Δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των παρατηρούμενων και των προβλεπόμενων τιμών.

Σύμφωνα με τα αποτελέσματα του ελέγχου, δεν υπάρχουν αποδείξεις που να υποστηρίζουν στατιστικά σημαντική απόκλιση του δείγματος από τον πληθυσμό ως προς τον τύπο κυριότητας ακινήτων, επομένως το δείγμα γίνεται δεκτό.

Σύμφωνα με τα στοιχεία της ΕΛΣΤΑΤ για τους δήμους Ραφήνας και Μαραθώνα η οικογενειακή κατάσταση των κατοίκων της περιοχής είναι κατά 39.4% άγαμοι, κατά 52.1% έγγαμοι, κατά 5.3% διαζευγμένοι και κατά 3.2% χήροι. Έτσι, σε τυχαίο δείγμα 151 ατόμων η προβλεπόμενη κατανομή συχνοτήτων είναι 59 άγαμοι, 78 έγγαμοι, 8 διαζευγμένοι και 5 χήροι. Οι παρατηρούμενες όμως συχνότητες είναι 46 άγαμοι ,84 έγγαμοι, 12 διαζευγμένοι και 8 χήροι επομένως ήταν αναγκαίο να πραγματοποιηθεί έλεγχος καλής προσαρμογής του δείγματος όπως φαίνεται στον Πίνακα 4.3.

Πίνακας 4.3 Έλεγχος καλής προσαρμογής του δείγματος ως προς την οικογενειακή κατάσταση

χ^2 test

Οικογενειακή Κατάσταση	Άγαμοι	Έγγαμοι	Διαζευγμένοι	Χήροι	Σύνολο
Παρατηρήσεις	47	84	12	8	151
Προβλέψεις	59	78	8	5	

α	0,05
p.value	0,075
df	3
x^2	6,918
critical Value	7,815
	accept null

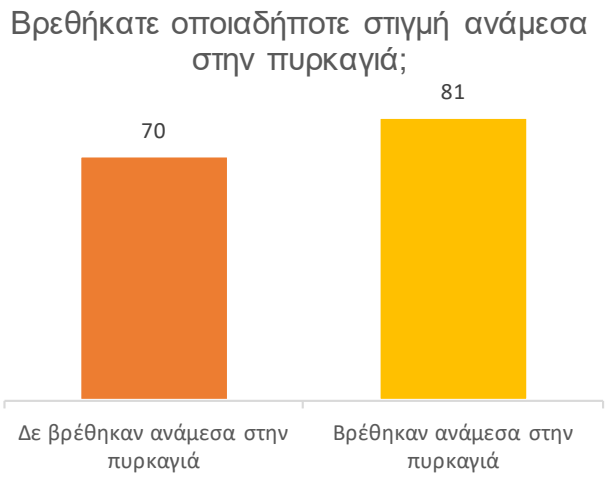
Μηδενική Υπόθεση:
Δεν υπάρχει στατιστικά σημαντική διαφορά μεταξύ των παρατηρούμενων και των προβλεπόμενων τιμών.

Σύμφωνα με τα αποτελέσματα του ελέγχου, δεν υπάρχουν αποδείξεις που να υποστηρίζουν στατιστικά σημαντική απόκλιση του δείγματος από τον πληθυσμό ως προς την οικογενειακή κατάσταση επομένως το δείγμα γίνεται δεκτό.

4.3 Ανάλυση βάσης δεδομένων

4.3.1 Σημεία με ενδιαφέρον

Όπως φαίνεται στο Διάγραμμα 4.7, η πλειοψηφία των ερωτηθέντων βρέθηκε κάποια στιγμή ανάμεσα στην πυρκαγιά. Η απόφαση για εκκένωση δε φαίνεται να έπαιξε σημαντικό ρόλο αφού το 54% των ερωτηθέντων που εκκένωσαν την κατοικία τους κατέληξαν τελικά ανάμεσα στην πυρκαγιά ενώ το 52% όσων δεν εκκένωσαν και επέλεξαν να προστατέψουν τις περιουσίες τους βρέθηκαν ανάμεσα στις φλόγες. Αυτό οφείλεται κατά πάσα πιθανότητα στο γεγονός ότι το οδικό δίκτυο ήταν κορεσμένο κατά τη διάρκεια της πυρκαγιάς με αποτέλεσμα πολλοί οδηγοί να εγκλωβιστούν στα οχήματά τους.



Διάγραμμα 4.7 Κατανομή ερωτηθέντων με βάση το αν βρέθηκαν ή όχι ανάμεσα στην πυρκαγιά

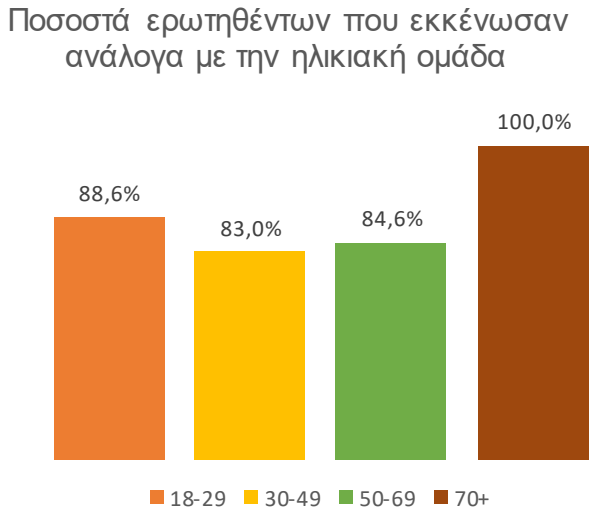
Από το Διάγραμμα 4.8, φαίνεται πως η πλειοψηφία των ερωτηθέντων (86%) προσπάθησε να εκκενώσει την περιοχή κατοικίας του.

Διάγραμμα 4.8 Κατανομή ερωτηθέντων με βάση το αν εκκένωσαν ή όχι την περιοχή κατοικίας τους

Στο Διάγραμμα 4.9 παρατηρείται πως οι άντρες έμειναν να προστατέψουν τις κατοικίες τους σε μεγαλύτερο ποσοστό σε σχέση με τις γυναίκες αφού μόλις το 80% των ανδρών που συμπλήρωσαν το ερωτηματολόγιο εκκένωσαν την κατοικία τους σε αντίθεση με τις γυναίκες που εκκένωσαν σε ποσοστό 91,4%. (Στατιστικά σημαντική διαφορά σε 95% επίπεδο εμπιστοσύνης)

Διάγραμμα 4.9 Ποσοστά ερωτηθέντων που εκκένωσαν ανάλογα με το φύλο

Φαίνεται επίσης πως υπάρχει συσχέτιση μεταξύ της ηλικίας και της απόφασης για εκκένωση καθώς στο Διάγραμμα 4.10 παρατηρείται σημαντική διαφορά μεταξύ των διαφόρων ηλικιακών ομάδων ως προς την απόφαση για εκκένωση. Πρωτιά στο ποσοστό εκκένωσης σημειώνουν οι ηλικιωμένοι 70+, ακολουθούν οι νέοι μέχρι 29 ετών, και τις τελευταίες θέσεις καταλαμβάνουν οι ηλικιακές ομάδες 30-49 και 50-69.



Διάγραμμα 4.10 Ποσοστά ερωτηθέντων που εκκένωσαν ανάλογα με την ηλικιακή ομάδα

Στο Διάγραμμα 4.11 αποτυπώνεται η επιρροή που έχει μια έγκαιρη επίσημη εντολή εκκένωσης από την πολιτεία στην απόφαση για εκκένωση. Ακόμη και ύστερα από το συμβάν οι ερωτηθέντες δήλωσαν πως σε μία νέα ενδεχόμενη μελλοντική πυρκαγιά δεν είναι ιδιαίτερα πιθανό (1,86/5) να εκκενώσουν την κατοικία τους χωρίς να έχει δοθεί επίσημη εντολή εκκένωσης. Αντίθετα δήλωσαν

πως είναι εξαιρετικά πιθανό (4,45/5) να εκκενώσουν εφόσον έχει προηγηθεί μια εντολή εκκένωσης από την πολιτεία (Στατιστικά σημαντική διαφορά σε 99% επίπεδο εμπιστοσύνης).

Διάγραμμα 4.11 Διαφορά στην πρόθεση για εκκένωση από την έκδοση επίσημης εντολής εκκένωσης

Οι περισσότεροι άνθρωποι δεν ενημερώθηκαν εκ των προτέρων για τον κίνδυνο της πυρκαγιάς. Η συντριπτική πλειοψηφία απάντησε ότι η πρώτη πηγή πληροφόρησης που έλαβε για την πυρκαγιά ήταν πως είδε τη φωτιά να πλησιάζει, όπως παρουσιάζεται στο παρακάτω διάγραμμα.

Διάγραμμα 4.12 Κατανομή ερωτηθέντων με βάση την πρώτη πηγή πληροφόρησης για την πυρκαγιά

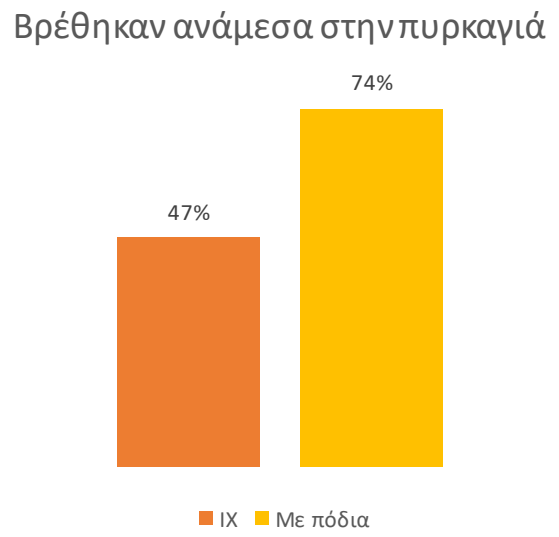
Σύμφωνα με τα διαγράμματα 4.13 και 4.14, ενώ το 90,8% των ερωτηθέντων διέθετε όχημα διαθέσιμο κατά τη διάρκεια της πυρκαγιάς μόνο το 76,2% το χρησιμοποίησε ως μέσο διαφυγής. .

(Στατιστικά σημαντική διαφορά σε 95% επίπεδο εμπιστοσύνης). Το υπόλοιπο 23,8% προσπάθησε να εκκενώσει την περιοχή με τα πόδια.

Διάγραμμα 4.13 Ποσοστό ερωτηθέντων όπου είχαν διαθέσιμο όχημα κατά την εκκένωση

Διάγραμμα 4.14 Κατανομή ερωτηθέντων με βάση τον τρόπο εκκένωσης

Στο Διάγραμμα 4.15 φαίνεται ωστόσο πως το περπάτημα ήταν ο πιο επικίνδυνος τρόπος διαφυγής καθώς το 74% των ερωτηθέντων που επέλεξαν να εκκενώσουν με τα πόδια κατέληξαν ανάμεσα στην πυρκαγιά ενώ μόνο το 47% εκείνων που διέφυγαν με το αυτοκίνητο βρέθηκαν ανάμεσα στις φλόγες. . (Στατιστικά σημαντική διαφορά σε 99% επίπεδο εμπιστοσύνης)

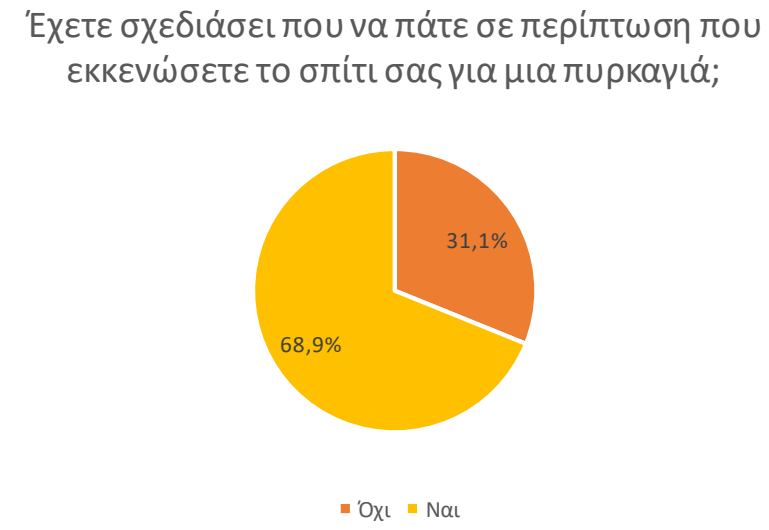


Διάγραμμα 4.15 Ποσοστά ερωτηθέντων που βρέθηκαν ανάμεσα στην πυρκαγιά ανάλογα με τον τρόπο εκκένωσης

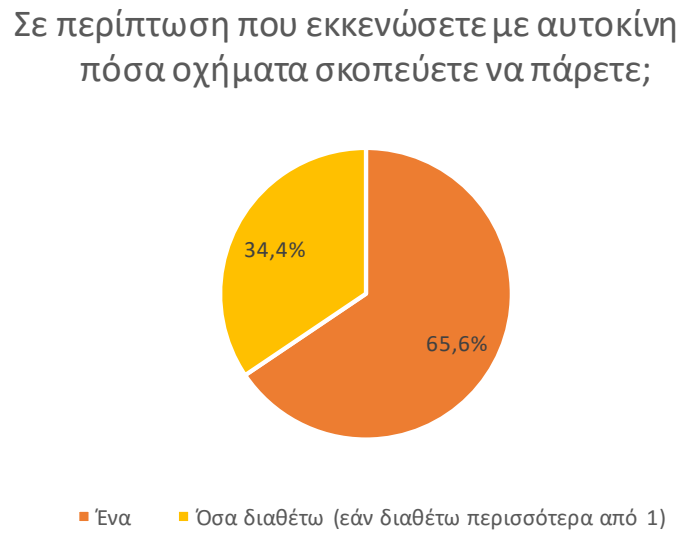
Σύμφωνα με το Διάγραμμα 4.16**Error! Reference source not found.**, όσο μικρότεροι σε ηλικία ήταν οι κάτοικοι, τόσο μεγαλύτερα ήταν τα ποσοστά όσων επέλεξαν να εκκενώσουν με τα πόδια. Όπως παρατηρείται και στο διπλανό διάγραμμα, η ηλικιακή ομάδα 18-29 επέλεξε το περπάτημα σε ποσοστό 45%, η ηλικιακή ομάδα 30-49 σε ποσοστό 20%, η ηλικιακή ομάδα 50-69 σε ποσοστό 16% και η ηλικιακή ομάδα 70+ σε ποσοστό 9%.

Διάγραμμα 4.16 Επιλογή μέσου ανά ηλικιακή ομάδα

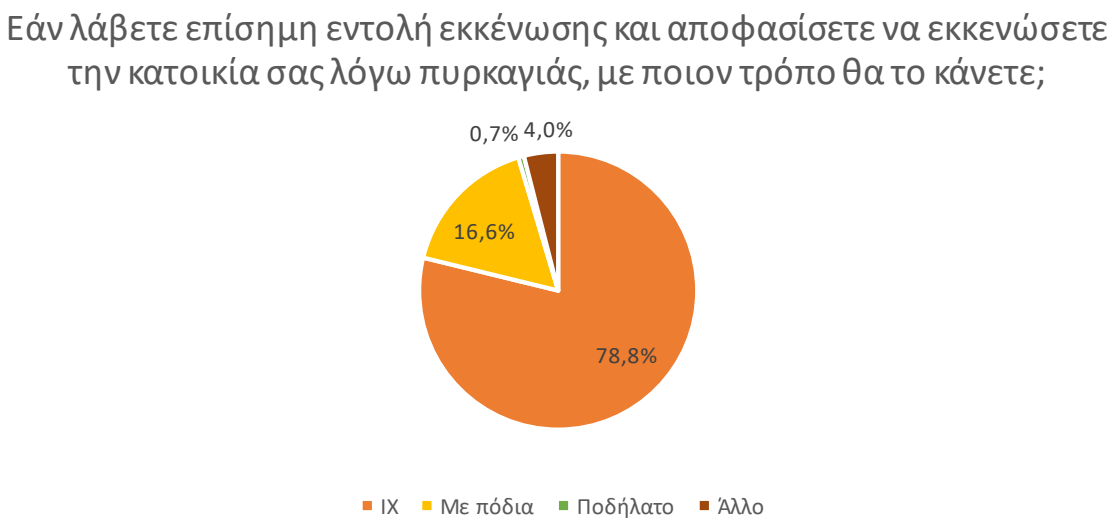
Στα διαγράμματα 4.17 έως 4.20 παρουσιάζονται ορισμένες κρίσιμες αποφάσεις που οι ερωτηθέντες δήλωσαν ότι θα λάβουν σε μία νέα ενδεχόμενη απειλή πυρκαγιάς



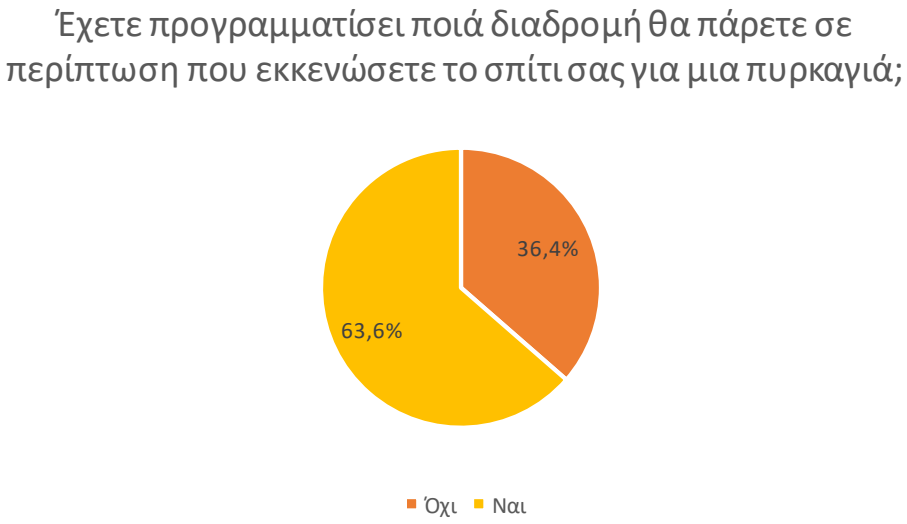
Διάγραμμα 4.17 Κατανομή ερωτηθέντων με βάση την ύπαρξη σχεδιασμού για μελλοντική εκκένωση



Διάγραμμα 4.18 Αριθμός οχημάτων ανά οικογένεια κατά την εκκένωση



Διάγραμμα 4.19 Επιλογή μέσου σε πιθανή μελλοντική πυρκαγιά



Διάγραμμα 4.20 Κατανομή ερωτηθέντων με βάση τον προγραμματισμό διαδρομής για μελλοντική εκκένωση

4.4 Πρόβλεψη απόφασης για εκκένωση

4.4.1 Μοντέλο λογιστικής παλινδρόμησης για πρόβλεψη εκκένωσης

<u>Προεπεξεργασία της βάσης δεδομένων</u>

Για το μοντέλο πρόβλεψης της απόφασης για εκκένωση η μη χρησιμοποιήθηκε ο πίνακας που προέκυψε κατά την κωδικοποίηση των δεδομένων του ερωτηματολογίου όπως αυτή περιγράφεται στο 3.2.2. Ως εξαρτημένη μεταβλητή επιλέχθηκε η μεταβλητή Q9 που υποδηλώνει

την απόφαση των ερωτηθέντων σχετικά με το αν εκκένωσαν ή όχι το μέρος στο οποίο βρίσκονταν. Για τον καθορισμό των ανεξάρτητων μεταβλητών, αφαιρέθηκαν από τη βάση δεδομένων οι μεταβλητές που αντιστοιχούν στις ερωτήσεις 10 έως 31 καθώς πρόκειται για μεταβλητές που αφορούν συμβάντα μεταγενέστερα της απόφασης για εκκένωση, επομένως οι πληροφορίες που περιλαμβάνουν περιέχουν «προκατάληψη» (bias). Στη συνέχεια αφαιρέθηκαν από τη βάση δεδομένων οι μεταβλητές που παρουσίαζαν μηδενική ή πολύ μικρή διακύμανση με χρήση της εντολής nearZeroVar {caret} καθώς και οι μεταβλητές που παρουσίαζαν μεταξύ τους correlation άνω του 0,5 μέσω των εντολών cor, findCorrelation {caret}. Ύστερα από τα παραπάνω βήματα το μέγεθος της βάσης δεδομένων μειώθηκε από 151x111 σε 151x24. Οι μεταβλητές που αποτέλεσαν τη βάση δεδομένων, μαζί με την επεξήγηση και τις τιμές που λαμβάνουν, φαίνονται στον Πίνακα 4.4.

Πίνακας 4.4 Πίνακας μεταβλητών απόφασης για εκκένωση/παραμονή

Μεταβλητή	**Επεξήγηση**	**Τιμές**
Q9	Εκκενώσατε το μέρος που βρισκόσασταν όταν απειλήθηκε από την πυρκαγιά;	Ναι, Όχι
years in community	Πόσα χρόνια μένετε στην κοινότητα που ζείτε τώρα;	num
years in residence	Πόσα χρόνια μένετε στην τωρινή κατοικία σας;	num
ownership_status	Νοικιάζετε ή σας ανήκει η κατοικία στην οποία διαμένετε;	num
age	Πόσο χρονών είστε;	num
gender	Ποιό είναι το φύλο σας;	Άντρας, Γυναίκα
marital status	Ποιά είναι η οικογενειακή σας κατάσταση;	Άγαμος, Έγγαμος, Διαζευγμένος, Χήρος
children.in.residence	Ύπαρξη ανηλίκων στην κατοικία	Ναι, Όχι
household under18	Πόσα άτομα στην κατοικία σας είναι κάτω των 18;	num
household 18to65	Πόσα άτομα στην κατοικία σας είναι 18-65 χρονών;	num
household over65	Πόσα άτομα στην κατοικία σας είναι πάνω από 65 χρονών;	num
income	Ποιές από τις παρακάτω κατηγορίες περιγράφουν καλύτερα το ετήσιο εισόδημα του νοικοκυριού σας προ φόρων;	<25.000, 25.000-50.000, >50.000
education	Ποιό από παρακάτω αντικατοπτρίζει το επίπεδο της εκπαίδευσης που έχετε ολοκληρώσει;	Δημοτικό, Γυμνάσιο, Λύκειο, Ιδιωτικό κολλέγιο, Πανεπιστήμιο, Μεταπτυχιακό/Διδακτορικό
prior warning	Ύπαρξη προηγούμενης προειδοποίησης	Ναι, Όχι
perceived risk	Αντίληψη του κινδύνου της πυρκαγιάς	1 έως 5
Q7: Λάβατε από κάποιον συμβουλή να καταφύγετε σε μια ασφαλέστερη τοποθεσία;		
Q7a	Όχι	Ναι, Όχι
Q7b	Ναι, από κάποια αρχή (π.χ. αστυνομία)	Ναι, Όχι
Q7c	Ναι, από κάποιον φίλο, συγγενή ή γείτονα	Ναι, Όχι
Q7d	Ναι, από τα Μέσα Μαζικής Ενημέρωσης	Ναι, Όχι
Q7e	Άλλο	Ναι, Όχι

Q8: Προσπαθήσατε να λάβετε περισσότερες πληροφορίες από κάποιον πριν αποφασίσετε να εκκενώσετε;		
Q8a	Όχι	Ναι, Όχι
Q8b	Ναι, από φίλους, συγγενείς, γείτονες	Ναι, Όχι
Q8c	Ναι, από τις αρχές (π.χ. αστυνομία)	Ναι, Όχι
Q8d	Ναι, από τα Μέσα Μαζικής Επικοινωνίας	Ναι, Όχι

Τέλος, λόγω της ανισοκατανομής των απαντήσεων της εξαρτημένης μεταβλητής Q9 (130 εκκένωσαν, 21 όχι) ακολουθήθηκε στρατηγική oversampling κατά την οποία επαναλήφθηκαν απαντήσεις που οι ερωτηθέντες είχαν δηλώσει πως δεν εκκένωσαν την περιοχή. Μετά το oversampling προέκυψαν 130 θετικές και 50 αρνητικές απαντήσεις. Η στρατηγική αυτή κρίθηκε απαραίτητη για τον ευκολότερο εντοπισμό των αρνητικών απαντήσεων από το μοντέλο αφού πριν την εφαρμογή της το μοντέλο έτεινε να προβλέπει τη συντριπτική πλειοψηφία των απαντήσεων στη θετική κλάση. Η διαδικασία αυτή πραγματοποιήθηκε μέσω της εντολής ovun.sample {ROSE}.

Εκπαίδευση μοντέλου λογιστικής παλινδρόμησης

Με τη βοήθεια του πακέτου glmulti και έπειτα από δοκιμές, εισάγοντας κάθε φορά διαφορετικούς συνδυασμούς και ελέγχοντας το πόσο επηρεάζουν την πρόβλεψη, προέκυψαν οι ανεξάρτητες μεταβλητές που χρησιμοποιούνται στο μοντέλο το οποίο περιγράφει πιο αποτελεσματικά την εξαρτημένη μεταβλητή.

Εκπαιδεύτηκε λοιπόν το μοντέλο και προέκυψε η εξής συνάρτηση χρησιμότητας:

$$Q9 = -6.54887 - 1.09905 * ownership.status + 0.03288 * age + 0.65752 * gender - 0.28471 * household.under18 + 1.48271 * income + 0.93230 * perceived.risk + 2.24657 * Q7c + 1.41085 * Q8a \quad (4.2)$$

```
Deviance Residuals:
    Min       1Q   Median       3Q      Max
-1.8671  -0.3914   0.3299   0.5662   2.4570

Coefficients:
                  Estimate Std. Error z value Pr(>|z|)
(Intercept)       -6.54887    1.55223  -4.219 2.45e-05 ***
Ownership.status  -1.09905    0.67185  -1.636 0.101868
age                0.03288    0.01592   2.065 0.038922 *
gender             0.65752    0.41660   1.578 0.114496
household.under18 -0.28471    0.16262  -1.751 0.079977 .
income             1.48271    0.39585   3.746 0.000180 ***
perceived.risk     0.93230    0.20312   4.590 4.43e-06 ***
Q7c                2.24657    0.65772   3.416 0.000636 ***
Q8a                1.41085    0.47762   2.954 0.003138 **
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 212.70  on 179  degrees of freedom
Residual deviance: 147.13  on 171  degrees of freedom
AIC: 165.13

Number of Fisher Scoring iterations: 5
```

Όπου:

- ownership.status: Τύπος κυριότητας κατοικίας (0 για ενοίκιο, 1 για ιδιοκτησία)
- age: Ηλικία (num)

- gender: Φύλο ερωτηθέντων (male 0, female 1)
- household_under18: Αριθμός ανηλίκων που διαμένουν στην κατοικία των ερωτηθέντων (num)
- income: Οικογενειακό εισόδημα (1,2,3 για εισόδημα μικρότερο από 25.000, ανάμεσα σε 25.000 και 50.000 και άνω των 50.000 αντίστοιχα)
- perceived.risk: Μεταβλητή που προέκυψε από τον μέσο όρο των ερωτήσεων 6α με 6δ και δείχνει το πόσο πιθανά θεωρούσαν οι ερωτηθέντες τα ενδεχόμενα να τραυματιστούν εκείνοι ή οι γείτονές τους καθώς και να δημιουργηθούν σοβαρές ζημιές στην κατοικία τους ή των γειτόνων τους (1 έως 5 για «καθόλου πιθανό» έως «σχεδόν σίγουρο»)
- Q7c: Λάβατε από κάποιον συμβουλή να καταφύγετε σε μια ασφαλέστερη τοποθεσία; - Ναι, από κάποιον φίλο, συγγενή ή γείτονα (0,1)
- Q8a: Προσπαθήσατε να λάβετε περισσότερες πληροφορίες από κάποιον πριν αποφασίσετε να εκκενώσετε; - Όχι (0,1)

Έλεγχος μοντέλου λογιστικής παλινδρόμησης

Ο έλεγχος του μοντέλου έγινε με 10-fold cross validation. Η μήτρα σύγχυσης και τα στατιστικά χαρακτηριστικά που προέκυψαν βρίσκονται στον Πίνακα 4.5 και τον Πίνακα 4.6.

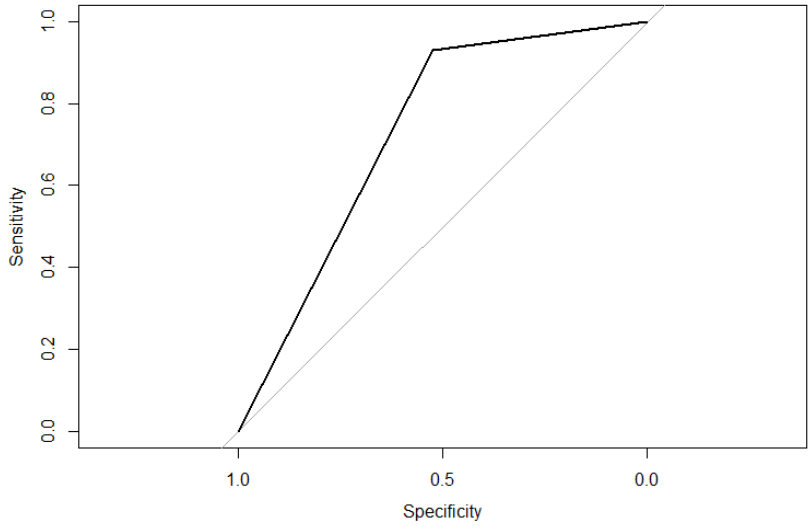
Πίνακας 4.5 Μήτρα σύγχυσης μοντέλου λογιστικής παλινδρόμησης

Confusion matrix	Predicted Negative	Predicted Positive
Actual Negative	27	23
Actual Positive	12	118

Πίνακας 4.6 Στατιστικά χαρακτηριστικά μοντέλου λογιστικής παλινδρόμησης

Ορθότητα (accuracy)	80,56%
Συντελεστής Κάππα	48,02%
Ευαισθησία (sensitivity)	90,77%
Εξειδικευτικότητα (specificity)	54,00%
Ακρίβεια (precision)	83,69%
Μέτρο F	87,08%

Η καμπύλη ROC του μοντέλου φαίνεται στο Διάγραμμα 4.21. Η επιφάνεια που βρίσκεται κάτω από την καμπύλη ισούται με AUC= 0.7238.



Διάγραμμα 4.21 Καμπύλη Receiver Operating Characteristic (ROC curve) του μοντέλου λογιστικής παλινδρόμησης

Αξιολόγηση μοντέλου λογιστικής παλινδρόμησης

Στον Πίνακα 4.7 φαίνονται αναλυτικά οι παράγοντες που χρησιμοποιούνται από το μοντέλο λογιστικής παλινδρόμησης, η επίδραση που έχουν στην απόφαση για εκκένωση ή παραμονή καθώς και το επίπεδο εμπιστοσύνης τους. Η επίδραση των παραγόντων εξαρτάται από το πρόσημο των συντελεστών των αντίστοιχων παραγόντων στη συνάρτηση χρησιμότητας (4.2). Ως θετική θεωρείται η επίδραση των παραγόντων που λειτουργούν υπέρ της της εκκένωσης.

Πίνακας 4.7 Επίδραση και επίπεδο εμπιστοσύνης για τους διάφορους παράγοντες που χρησιμοποιούνται στο μοντέλο λογιστικής παλινδρόμησης

Παράγοντας	**Επίδραση**	**Επίπεδο εμπιστοσύνης**	
Τύπος κυριότητας κατοικίας	Αρνητική για ιδιοκτήτες των ακινήτων	89,81%	<90%
Ηλικία	Θετική για μεγαλύτερες ηλικίες	96,11%	
Φύλο	Θετική για τις γυναίκες	88,55%	<90%
Αριθμός ανηλίκων στην κατοικία	Αρνητική για μεγαλύτερο αριθμό ανηλίκων	92,00%	
Εισόδημα	Θετική για μεγαλύτερο εισόδημα	99,98%	
Αντίληψη κινδύνου	Θετική για μεγαλύτερη αντίληψη κινδύνου	99,99%	

Συμβουλή για εκκένωση από φίλο/συγγενή	Θετική για όσους έλαβαν συμβουλή για εκκένωση από φίλους/συγγενείς	99,94%
Προσπάθεια για λήψη επιπλέον πληροφόρησης πριν την απόφαση για εκκένωση	Θετική για όσους δεν προσπάθησαν να λάβουν επιπλέον πληροφορίες πριν εκκενώσουν	99,69%

Από τα αποτελέσματα προκύπτει ότι:

- Η συνολική ορθότητα του μοντέλου κρίνεται ικανοποιητική
- Ο στατιστικός συντελεστής Κάππα κρίνεται μη ικανοποιητικός
- Το μοντέλο προβλέπει εξαιρετικά τη θετική κλάση όμως δεν προβλέπει ικανοποιητικά την αρνητική κλάση
- Το εμβαδόν κάτω από την καμπύλη ROC κρίνεται ικανοποιητικό

4.4.2 Δέντρο αποφάσεων για πρόβλεψη εκκένωσης

Προεπεξεργασία της βάσης δεδομένων

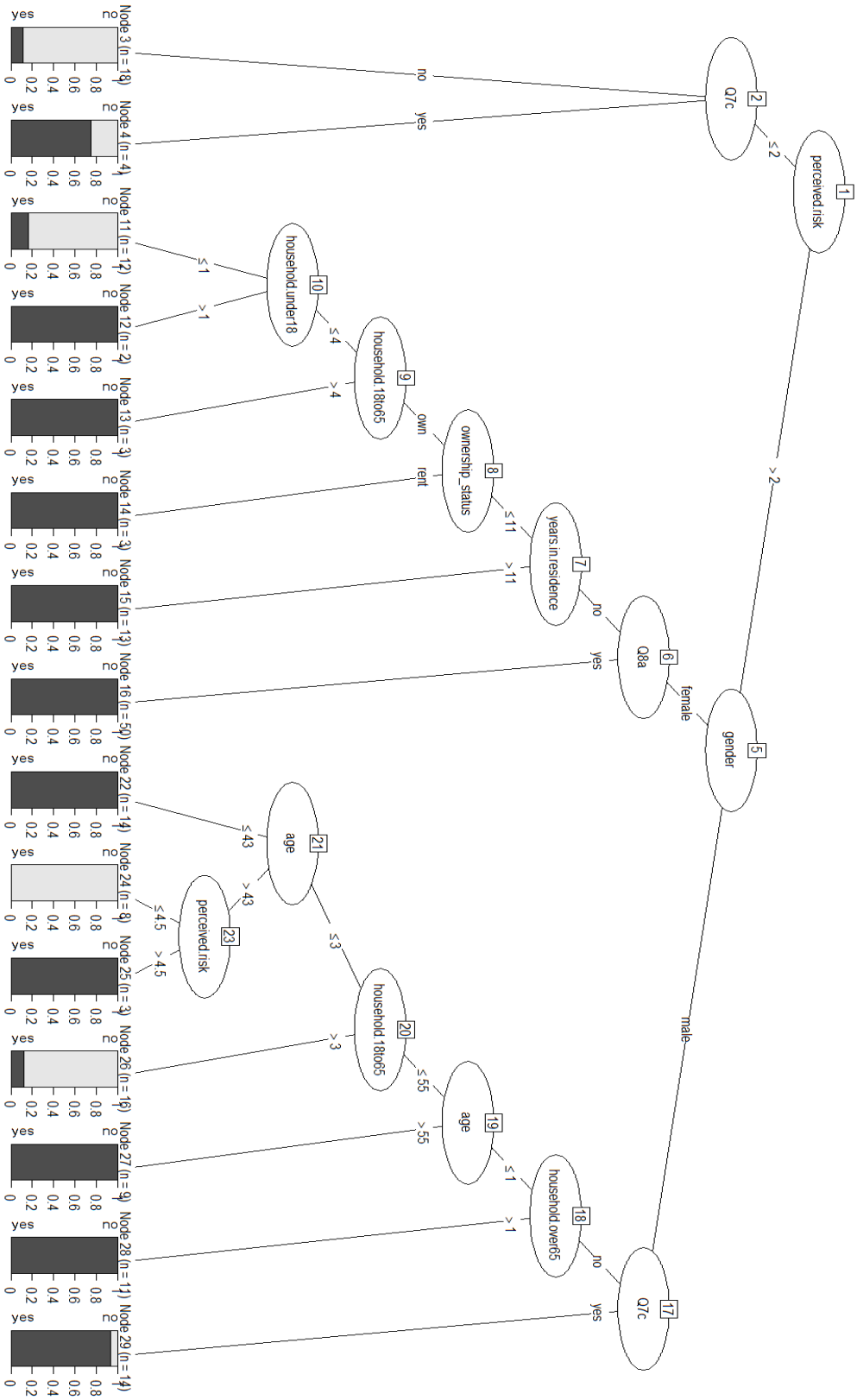
Για την προεπεξεργασία της βάσης δεδομένων ακολουθήθηκε σχεδόν η ίδια διαδικασία με την αντίστοιχη του μοντέλου λογιστικής παλινδρόμησης που παρουσιάζεται στο 4.4.1, με εξαίρεση τη μη αφαίρεση των μεταβλητών που παρουσιάζουν μεταξύ τους υψηλή συσχέτιση καθώς οι αλγόριθμοι δέντρων αποφάσεων δεν επηρεάζονται από τη συσχέτιση των ανεξάρτητων μεταβλητών. Επομένως, επιπλέον των μεταβλητών του Πίνακας 4.4 στη βάση δεδομένων συμπεριλήφθηκε και η μεταβλητή Q5 (Αφού λάβατε τις πρώτες πληροφορίες σχετικά με την πυρκαγιά, από ποιες άλλες πηγές λάβατε πληροφόρηση;)

Εκπαίδευση μοντέλου δέντρου απόφασης

Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε η εντολή J48 {Rweka}. Το διάγραμμα του δέντρου απόφασης που δημιουργήθηκε φαίνεται στην παρακάτω σελίδα (Διάγραμμα 4.22). Οι μεταβλητές που χρησιμοποιηθήκαν είναι οι εξής:

- perceived.risk: Μεταβλητή που προέκυψε από τον μέσο όρο των ερωτήσεων 6α με 6δ και δείχνει το πόσο πιθανά θεωρούσαν οι ερωτηθέντες τα ενδεχόμενα να τραυματιστούν εκείνοι ή οι γείτονές τους καθώς και να δημιουργηθούν σοβαρές ζημιές στην κατοικία τους ή των γειτόνων τους (1 έως 5 για «καθόλου πιθανό» έως «σχεδόν σίγουρο»)
- Q7c: Λάβατε από κάποιον συμβουλή να καταφύγετε σε μια ασφαλέστερη τοποθεσία; - Ναι, από κάποιον φίλο, συγγενή ή γείτονα (0,1)
- Q8a: Προσπαθήσατε να λάβετε περισσότερες πληροφορίες από κάποιον πριν αποφασίσετε να εκκενώσετε; - Όχι (0,1)
- age: Ηλικία ερωτηθέντων (num)
- ownership.status: Τύπος κυριότητας κατοικίας (rent, own)
- gender: Φύλο ερωτηθέντων (male, female)
- years.in.residence: Έτη που οι ερωτηθέντες διέμεναν στην κατοικία τους (num)

- household_under18: Αριθμός ανηλίκων που διαμένουν στην κατοικία των ερωτηθέντων (num)
- household.18to65: Αριθμός ενηλίκων κάτω των 65 που διαμένουν στην κατοικία των ερωτηθέντων (num)
- household.over65: Αριθμός ενηλίκων άνω των 65 που διαμένουν στην κατοικία των ερωτηθέντων (num)



Διάγραμμα 4.22 Σχήμα δέντρου απόφασης σχετικά με την απόφαση για εκκένωση

Έλεγχος μοντέλου δέντρου απόφασης

Ο έλεγχος του μοντέλου έγινε με 10-fold cross validation. Η μήτρα σύγχυσης και τα στατιστικά χαρακτηριστικά που προέκυψαν βρίσκονται παρακάτω:

```
=== 10-Fold Cross Validation ===

=== Summary ===

Correctly Classified Instances         155               86.1111 %
Incorrectly Classified Instances        25               13.8889 %
Kappa statistic                          0.6642
Mean absolute error                      0.1665
Root mean squared error                  0.3563
Relative absolute error                 41.3689 %
Root relative squared error             79.5427 %
Total Number of Instances              180

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0,800    0,115    0,727      0,800    0,762      0,666    0,868     0,658     no
               0,885    0,200    0,920      0,885    0,902      0,666    0,868     0,929     yes
Weighted Avg.  0,861    0,176    0,866      0,861    0,863      0,666    0,868     0,854

=== Confusion Matrix ===

   a   b   <-- classified as
  40  10 |   a = no
  15 115 |   b = yes
```

Αξιολόγηση μοντέλου δέντρου απόφασης

Παρατηρείται μια μικρή διαφοροποίηση των παραγόντων που επιλέχθηκαν από το μοντέλο δέντρου αποφάσεων σε σχέση με το αντίστοιχο της λογιστικής παλινδρόμησης. Επιπλέον της μεταβλητής household.under18 επιλέχθηκαν και οι μεταβλητές household.18to65 και household.over65 ενώ δεν επιλέχθηκε το εισόδημα.

Η επίδραση των παραγόντων αυτών μπορεί να επεξηγηθεί από το Διάγραμμα 4.22. Δεν είναι δυνατή η εξαγωγή αντίστοιχων συμπερασμάτων με τον πίνακα 4.7 του μοντέλου λογιστικής παλινδρόμησης, αφού το μοντέλο δέντρου απόφασης δεν προϋποθέτει γραμμική σχέση μεταξύ των ανεξάρτητων μεταβλητών και της εξαρτημένης μεταβλητής.

Η συνολική ορθότητα του μοντέλου είναι μεγαλύτερη από αυτή του μοντέλου λογιστικής παλινδρόμησης και επιπλέον, προβλέπει εξαιρετικά και τις δύο κλάσεις. Ωστόσο, είναι πιθανή η υπερεκπαίδευση (overfitting) αφού τα μοντέλα δέντρου απόφασης είναι επιρρεπή (M. Kubat, 1998) και ο αριθμός των δεδομένων από το ερωτηματολόγιο είναι μικρός.

Προκύπτει ότι:

- Η συνολική ορθότητα του μοντέλου κρίνεται εξαιρετική
- Το μοντέλο προβλέπει εξαιρετικά τόσο τη θετική όσο και την αρνητική κλάση
- Ο συντελεστής Κάππα είναι ικανοποιητικός
- Το μέτρο F είναι αρκετά ικανοποιητικό
- Το εμβαδό κάτω από την επιφάνεια της καμπύλης ROC κρίνεται εξαιρετικό

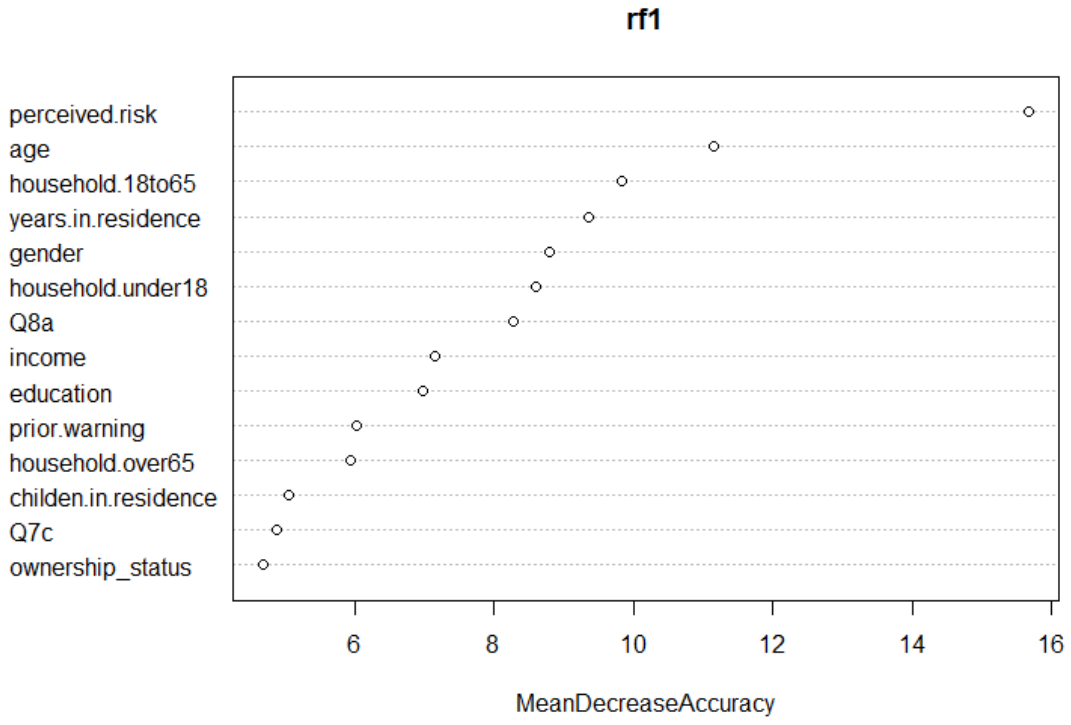
4.4.3 Μοντέλο τυχαίου δάσους για πρόβλεψη εκκένωσης

Προεπεξεργασία βάσης δεδομένων

Για την προεπεξεργασία της βάσης δεδομένων ακολουθήθηκε σχεδόν η ίδια διαδικασία με την αντίστοιχη του μοντέλου λογιστικής παλινδρόμησης που παρουσιάζεται στο 4.4.1, με εξαίρεση τη μη αφαίρεση των μεταβλητών που παρουσιάζουν μεταξύ τους υψηλή συσχέτιση καθώς οι αλγόριθμοι τυχαίων δασών δεν επηρεάζονται από τη συσχέτιση των ανεξάρτητων μεταβλητών. Επομένως, επιπλέον των μεταβλητών του Πίνακας 4.4 στη βάση δεδομένων συμπεριλήφθηκε και η μεταβλητή Q5 (Αφού λάβατε τις πρώτες πληροφορίες σχετικά με την πυρκαγιά, από ποιες άλλες πηγές λάβατε πληροφόρηση;)

Εκπαίδευση μοντέλου τυχαίων δασών

Για την εκπαίδευση του αλγορίθμου επιλέχθηκαν οι 10 σημαντικότερες μεταβλητές. Για την επιλογή των μεταβλητών αρχικά έτρεξε ένας αλγόριθμος randomforest με το σύνολο των μεταβλητών και στη συνέχεια χρησιμοποιήθηκε η εντολή importance {randomForest}. Στη συνέχεια έτρεξε νέος αλγόριθμος με τις 10 σημαντικότερες μεταβλητές, τα αποτελέσματα του οποίου φαίνονται στη συνέχεια (Διάγραμμα 4.23). Η κατάταξη των μεταβλητών έγινε μέσω του μεγέθους mean decrease accuracy. Το μέγεθος αυτό ποσοτικοποιεί τη σημασία μιας μεταβλητής μετρώντας την αλλαγή στην ακρίβεια πρόβλεψης, όταν οι τιμές της μεταβλητής μετατρέπονται τυχαία σε σύγκριση με τις αρχικές παρατηρήσεις.



Διάγραμμα 4.23 Σημαντικότητα μεταβλητών για τον αλγόριθμο τυχαίου δάσους

Οι μεταβλητές που επιλέχτηκαν λοιπόν είναι οι:

- perceived.risk: Μεταβλητή που προέκυψε από τον μέσο όρο των ερωτήσεων 6α με 6δ και δείχνει το πόσο πιθανά θεωρούσαν οι ερωτηθέντες τα ενδεχόμενα να τραυματιστούν εκείνοι ή οι γείτονές τους καθώς και να δημιουργηθούν σοβαρές ζημιές στην κατοικία τους ή των γειτόνων τους (1 έως 5 για «καθόλου πιθανό» έως «σχεδόν σίγουρο»)
- Q8a: Προσπαθήσατε να λάβετε περισσότερες πληροφορίες από κάποιον πριν αποφασίσετε να εκκενώσετε; - Όχι (0,1)
- age: Ηλικία ερωτηθέντων (num)
- ownership.status: Τύπος κυριότητας κατοικίας (rent, own)
- gender: Φύλο ερωτηθέντων (male, female)
- years.in.residence: Έτη που οι ερωτηθέντες διέμεναν στην κατοικία τους (num)
- household_under18: Αριθμός ανηλίκων που διαμένουν στην κατοικία των ερωτηθέντων (num)
- household.18to65: Αριθμός ενηλίκων κάτω των 65 που διαμένουν στην κατοικία των ερωτηθέντων (num)
- income: Οικογενειακό εισόδημα (1,2,3 για εισόδημα μικρότερο από 25.000, ανάμεσα σε 25.000 και 50.000 και άνω των 50.000 αντίστοιχα)
- education: Επίπεδο εκπαίδευσης ερωτηθέντων (1,2,3 για low,medium,high)
- prior.warning: Ενημέρωση για τον κίνδυνο από αρχές, ΜΜΕ, φίλους και συγγενείς (no, yes)

Έλεγχος μοντέλου τυχαίων δασών

Ο έλεγχος του μοντέλου έγινε με 10-fold cross validation. Η μήτρα σύγχυσης και τα στατιστικά χαρακτηριστικά που προέκυψαν βρίσκονται στους παρακάτω πίνακες:

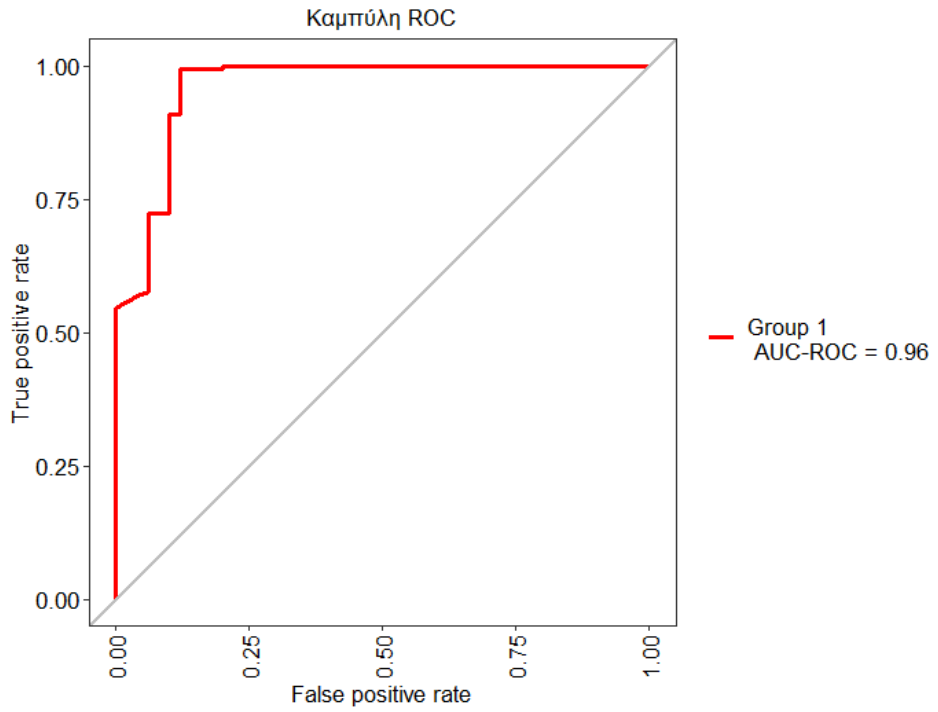
Πίνακας 4.8 Μήτρα σύγχυσης μοντέλου τυχαίου δάσους

Αριθμός δέντρων: 100		
Αριθμός μεταβλητών που δοκιμάστηκαν σε κάθε διαχωρισμό: 2		
	Predicted negative	Predicted positive
Actual negative	44	6
Actual positive	4	126

Πίνακας 4.9 Στατιστικά χαρακτηριστικά μοντέλου τυχαίου δάσους

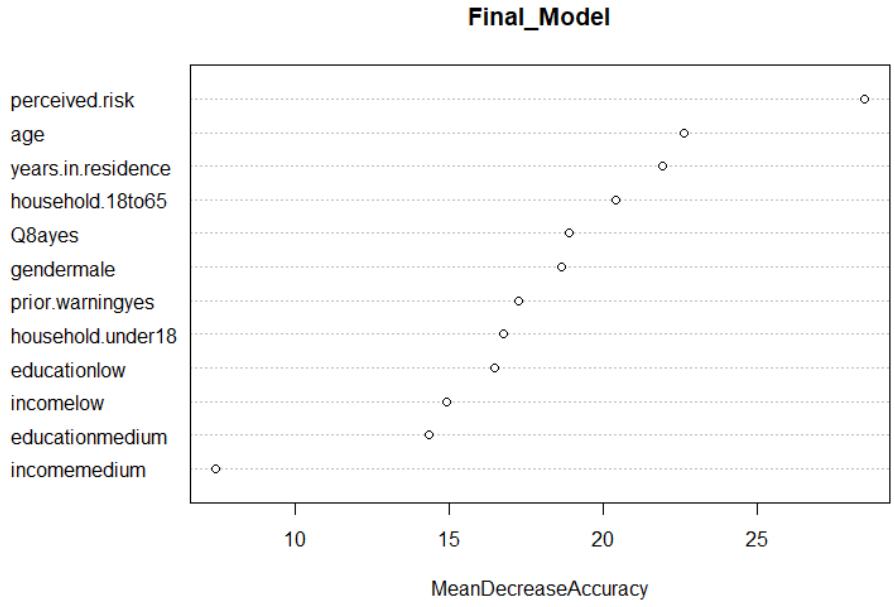
Ορθότητα (accuracy)	94,44%
Συντελεστής Κάππα	85,98%
Ευαισθησία (sensitivity)	96,92%
Εξειδικευτικότητα (specificity)	88,00%
Ακρίβεια (precision)	95,45%
Μέτρο F	96,18%

Η καμπύλη ROC του μοντέλου φαίνεται στο Διάγραμμα 4.24. Η επιφάνεια που βρίσκεται κάτω από την καμπύλη ισούται με AUC= 0.96



Διάγραμμα 4.24 Καμπύλη ROC μοντέλου randomForest για την πρόβλεψη της απόφασης για εκκένωση / παραμονή

Επιπλέον, η σημαντικότητα των μεταβλητών φαίνεται στο παρακάτω διάγραμμα:



Διάγραμμα 4.25 Σημαντικότητα μεταβλητών για τον τελικό αλγόριθμο τυχαίου δάσους

	no	yes	MeanDecreaseAccuracy	MeanDecreaseGini
perceived.risk	31.64106	13.666351	28.465887	14.510441
age	27.36412	8.646497	22.632462	11.163421
household.18to65	23.83060	6.505241	20.399736	6.099634
years.in.residence	25.07536	2.652524	21.906519	8.714000
Q8ayes	21.49268	6.776802	18.912883	3.940146
education low	16.83018	9.393343	16.471330	2.664046
education medium	14.04173	4.845589	14.330362	1.837571
household.under18	20.12060	2.972517	16.765684	5.045202
gender male	21.02890	6.733516	18.666734	3.000806
prior.warning yes	18.66934	6.978248	17.275837	3.076176
income low	18.36622	2.304866	14.920011	2.373681
income medium	13.08445	-4.434751	7.440753	1.448276

Αξιολόγηση μοντέλου τυχαίων δασών

Τα στατιστικά χαρακτηριστικά του μοντέλου είναι εξαιρετικά αφού αυτό προβλέπει σχεδόν τέλεια τόσο τη θετική, όσο και την αρνητική κλάση.

Παρατηρούνται μεγάλες διαφορές σε σχέση με τα προηγούμενα μοντέλα και ειδικά με το μοντέλο λογιστικής παλινδρόμησης όσον αφορά την επιλογή μεταβλητών και ιδιαίτερα τη σημαντικότητα τους. Η σημαντικότητα των παραγόντων που προκύπτουν από τον αλγόριθμο randomForest είναι αρκετά πιο αξιόπιστη από την αντίστοιχη της λογιστικής παλινδρόμησης αφού εκείνη στηρίζεται στην παραδοχή ότι οι μεταβλητές είναι μεταξύ τους ανεξάρτητες. Υπάρχει πάντα το ρίσκο να μην ανιχνευθεί κάποια αλληλεπίδραση μεταξύ των μεταβλητών και επομένως να δημιουργηθεί κάποια προκατάληψη στο μοντέλο που θα οδηγήσει σε λανθασμένους συντελεστές.

Επιπλέον, ο αλγόριθμος randomForest πλεονεκτεί εκ φύσεως έναντι των decision trees λόγω του μικρότερου σφάλματος γενίκευσης που προκύπτει από την τυχαιότητα της επιλογής των μεταβλητών πρόβλεψης.

Το μειονέκτημα ωστόσο του αλγορίθμου randomForest είναι πως είναι δύσκολη η επεξήγηση της επίδρασης των επιμέρους παραγόντων.

4.5 Πρόβλεψη επιλογής μέσου κατά την εκκένωση

4.5.1 Μοντέλο λογιστικής παλινδρόμησης για πρόβλεψη επιλογής μέσου

Προεπεξεργασία της βάσης δεδομένων

Για το μοντέλο πρόβλεψης της επιλογής μέσου κατά την εκκένωση χρησιμοποιήθηκε ο πίνακας που προέκυψε κατά την κωδικοποίηση των δεδομένων του ερωτηματολογίου όπως αυτή περιγράφεται στο 3.2.2. Ως εξαρτημένη μεταβλητή επιλέχθηκε η μεταβλητή Q13 που υποδηλώνει την απόφαση των ερωτηθέντων σχετικά με την επιλογή του μέσου εκκένωσης. Για τον καθορισμό των ανεξάρτητων μεταβλητών, αφαιρέθηκαν από τη βάση δεδομένων οι μεταβλητές που αντιστοιχούν στις ερωτήσεις 14 έως 31 καθώς πρόκειται για μεταβλητές που αφορούν συμβάντα μεταγενέστερα της απόφασης για επιλογή μέσου, επομένως οι πληροφορίες που περιλαμβάνουν περιέχουν «προκατάληψη» (bias).Στη συνέχεια αφαιρέθηκαν από τη βάση δεδομένων οι μεταβλητές που παρουσίαζαν μηδενική ή πολύ μικρή διακύμανση με χρήση της εντολής nearZeroVar {caret} καθώς και οι μεταβλητές που παρουσίαζαν μεταξύ τους correlation άνω του

0,5 μέσω των εντολών cor, findCorrelation {caret}. Ύστερα από τα παραπάνω βήματα το μέγεθος της βάσης δεδομένων μειώθηκε από 151x111 σε 151x31. Οι μεταβλητές που αποτέλεσαν τη βάση δεδομένων, μαζί με την επεξήγηση και τις τιμές που λαμβάνουν, φαίνονται στον Πίνακα 10.

Πίνακας 4.10 Πίνακας μεταβλητών για επιλογή μέσου

Μεταβλητή	**Επεξήγηση**	**Τιμές**
Q13	Πώς εκκενώσατε την περιοχή σας;	Όχημα, Με πόδια
years in community	Πόσα χρόνια μένετε στην κοινότητα που ζείτε τώρα;	num
years in residence	Πόσα χρόνια μένετε στην τωρινή κατοικία σας;	num
ownership_status	Νοικιάζετε ή σας ανήκει η κατοικία στην οποία διαμένετε;	num
Age	Πόσο χρονών είστε;	num
Gender	Ποιό είναι το φύλο σας;	Άντρας, Γυναίκα
marital status	Ποιά είναι η οικογενειακή σας κατάσταση;	Άγαμος, Έγγαμος, Διαζευγμένος, Χήρος
children.in.residence	Ύπαρξη ανηλίκων στην κατοικία	Ναι, Όχι
household under18	Πόσα άτομα στην κατοικία σας είναι κάτω των 18;	num
household 18to65	Πόσα άτομα στην κατοικία σας είναι 18-65 χρονών;	num
household over65	Πόσα άτομα στην κατοικία σας είναι πάνω από 65 χρονών;	num
Income	Ποιές από τις παρακάτω κατηγορίες περιγράφουν καλύτερα το ετήσιο εισόδημα του νοικοκυριού σας προ φόρων;	<25.000, 25.000-50.000, >50.000
Education	Ποιό από παρακάτω αντικατοπτρίζει το επίπεδο της εκπαίδευσης που έχετε ολοκληρώσει;	Δημοτικό, Γυμνάσιο, Λύκειο, Ιδιωτικό κολλέγιο, Πανεπιστήμιο, Μεταπτυχιακό/Διδακτορικό
prior warning	Ύπαρξη προηγούμενης προειδοποίησης	Ναι, Όχι
perceived risk	Αντίληψη του κινδύνου της πυρκαγιάς	1 έως 5
Q7: Λάβατε από κάποιον συμβουλή να καταφύγετε σε μια ασφαλέστερη τοποθεσία;		
Q7a	Όχι	Ναι, Όχι
Q7b	Ναι, από κάποια αρχή (π.χ. αστυνομία)	Ναι, Όχι
Q7c	Ναι, από κάποιον φίλο, συγγενή ή γείτονα	Ναι, Όχι
Q7d	Ναι, από από τα Μέσα Μαζικής Ενημέρωσης	Ναι, Όχι
Q7e	Άλλο	Ναι, Όχι
Q8: Προσπαθήσατε να λάβετε περισσότερες πληροφορίες από κάποιον πριν αποφασίσετε να εκκενώσετε;		
Q8a	Όχι	Ναι, Όχι
Q8b	Ναι, από φίλους, συγγενείς, γείτονες	Ναι, Όχι
Q8c	Ναι, από τις αρχές (π.χ. αστυνομία)	Ναι, Όχι
Q8d	Ναι, από τα Μέσα Μαζικής Επικοινωνίας	Ναι, Όχι
Q9	Εκκενώσατε το μέρος που βρισκόσασταν όταν απειλήθηκε από την πυρκαγιά;	Ναι, Όχι
Q11: Κάνατε κάποια άλλη ενέργεια πριν εκκενώσετε;		

Q11a	Μάζεψα είδη πρώτης ανάγκης	Ναι, Όχι
Q11b	Προσπάθησα να προστατέψω την περιουσία μου	Ναι, Όχι
Q11c	Προσπάθησα να εντοπίσω μέλη της οικογένειάς μου που δε βρίσκονταν μαζί μου	Ναι, Όχι
Q11d	Προσπάθησα να προειδοποιήσω άλλους	Ναι, Όχι
Q11e	Όχι	Ναι, Όχι
Q11f	Προσπάθησα να μεταφέρω σε ασφαλή περιοχή μέλη της οικογένειάς μου σε πάνω από μία διαδρομή	Ναι, Όχι
Q12	Είχατε κάποιο όχημα (π.χ. αυτοκίνητο, μοτοσυκλέτα κ.λ.π.) διαθέσιμο όταν αποφασίζατε εάν έπρεπε να εκκενώσετε;	Ναι, Όχι

Εκπαίδευση μοντέλου λογιστικής παλινδρόμησης

Με τη βοήθεια του πακέτου glmulti και έπειτα από δοκιμές, εισάγοντας κάθε φορά διαφορετικούς συνδυασμούς και ελέγχοντας το πόσο επηρεάζουν την πρόβλεψη, προέκυψαν οι ανεξάρτητες μεταβλητές που χρησιμοποιούνται στο μοντέλο το οποίο περιγράφει πιο αποτελεσματικά την εξαρτημένη μεταβλητή.

Εκπαιδεύτηκε λοιπόν το μοντέλο και προέκυψε η εξής συνάρτηση χρησιμότητας:

$$Q13 = 24.25710 + +age * (-0.08411) + household.18to65 * (-0.42518) + gender * 1.02578 - Q8a * (-1.44114) + Q12 * (-21.22314) \quad (4.3)$$

```
Deviance Residuals:
     Min        1Q    Median        3Q       Max
-1.70444  -0.54193  -0.28372  -0.08745   2.54523

Coefficients:
                   Estimate Std. Error z value Pr(>|z|)
(Intercept)        24.25710 1686.79146   0.014 0.988526
age                -0.08411    0.02492  -3.376 0.000737 ***
gender              1.02578    0.61767   1.661 0.096769 .
household.18to65   -0.42518    0.25628  -1.659 0.097099 .
Q8a                -1.44114    0.61887  -2.329 0.019877 *
Q12               -21.22314 1686.79045  -0.013 0.989961
---
Signif. codes:  0 ‘***’ 0.001 ‘**’ 0.01 ‘*’ 0.05 ‘.’ 0.1 ‘ ’ 1

(Dispersion parameter for binomial family taken to be 1)

    Null deviance: 142.818  on 129  degrees of freedom
Residual deviance:  79.707  on 124  degrees of freedom
AIC: 91.707

Number of Fisher Scoring iterations: 17
```

Όπου:

- age: Ηλικία ερωτηθέντων (num)
- gender: Φύλο ερωτηθέντων (0 άντρας,1 γυναίκα)
- household.18to65: Αριθμός ενηλίκων κάτω των 65 που διαμένουν στην κατοικία των ερωτηθέντων (num)
- Q8a: Προσπαθήσατε να λάβετε περισσότερες πληροφορίες από κάποιον πριν αποφασίσετε να εκκενώσετε; - Όχι (0 για μη συμπλήρωση checkbox ,1 για συμπλήρωση)

- Q12: Είχατε κάποιο όχημα (π.χ. αυτοκίνητο, μοτοσυκλέτα κ.λπ.) διαθέσιμο όταν αποφασίζατε εάν έπρεπε να εκκενώσετε; (0 όχι, 1 ναι)

Έλεγχος μοντέλου λογιστικής παλινδρόμησης

Ο έλεγχος του μοντέλου έγινε με 10-fold cross validation. Η μήτρα σύγχυσης και τα στατιστικά χαρακτηριστικά που προέκυψαν βρίσκονται στους παρακάτω πίνακες:

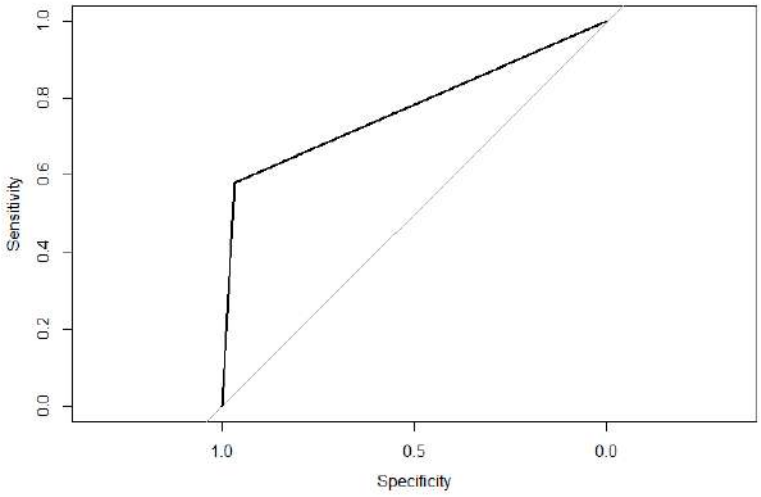
Πίνακας 4.11 Μήτρα σύγχυσης μοντέλου λογιστικής παλινδρόμησης για επιλογή μέσου

Confusion matrix	Actual Negative	Actual Positive
Predicted Negative	14	2
Predicted Positive	17	97

Πίνακας 4.12 Στατιστικά χαρακτηριστικά μοντέλου λογιστικής παλινδρόμησης για επιλογή μέσου

Ορθότητα (accuracy)	79,23%
Συντελεστής Κάππα	38,01%
Ευαισθησία (sensitivity)	89,90%
Εξειδικευτικότητα (specificity)	45,16%
Ακρίβεια (precision)	83,96%
Μέτρο F	86,83%

Η καμπύλη ROC του μοντέλου φαίνεται στο Διάγραμμα 4.26. Η επιφάνεια που βρίσκεται κάτω από την καμπύλη ισούται με AUC= 0.7641



Διάγραμμα 4.26 Καμπύλη Receiver Operating Characteristic μοντέλου λογιστικής παλινδρόμησης για επιλογή μέσου

Αξιολόγηση μοντέλου λογιστικής παλινδρόμησης

Το μοντέλο βασίζεται κυρίως στη μεταβλητή Q12, δηλαδή την ύπαρξη ή μη διαθέσιμου αυτοκινήτου. Έτσι θεωρεί πως όσοι δεν είχαν διαθέσιμο αυτοκίνητο εκκένωσαν με άλλο τρόπο και για όσους διέθεταν όχημα εξετάζει την επιλογή τους μέσα από τις μεταβλητές age, gender, household.18to65 και Q8a. Η κατηγοριοποίηση αυτή δεν κρίνεται ικανοποιητική αφού πέρα από τους 12 ερωτηθέντες που δήλωσαν πως δεν διέθεταν όχημα, το μοντέλο κατηγοριοποιεί στην αρνητική κλάση άλλους 2 από τους 19 υπολειπόμενους χρήστες που επέλεξαν να εκκενώσουν με τα πόδια ενώ διέθεταν όχημα. Τα παραπάνω, σε συνδυασμό με τη μικρή στατιστική σημαντικότητα του μοντέλου οδηγούν στο συμπέρασμα πως το μοντέλο είναι αναξιόπιστο.

4.5.2 Δέντρο αποφάσεων για πρόβλεψη επιλογής μέσου

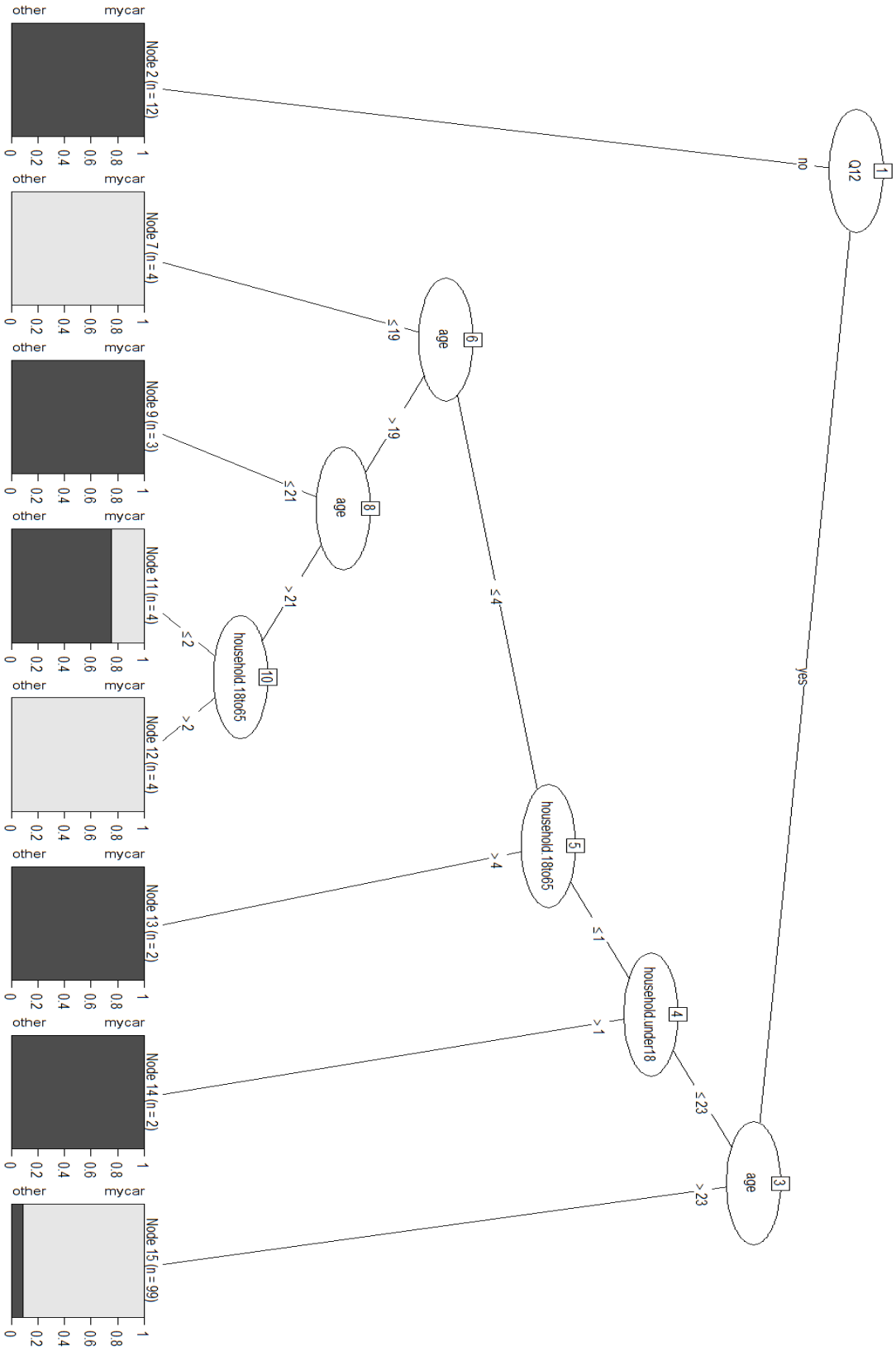
Προεπεξεργασία της βάσης δεδομένων

Για την προεπεξεργασία της βάσης δεδομένων ακολουθήθηκε σχεδόν η ίδια διαδικασία με την αντίστοιχη του μοντέλου λογιστικής παλινδρόμησης που παρουσιάζεται στο 4.5.1, με εξαίρεση τη μη αφαίρεση των μεταβλητών που παρουσιάζουν μεταξύ τους υψηλή συσχέτιση καθώς οι αλγόριθμοι δέντρων αποφάσεων δεν επηρεάζονται από τη συσχέτιση των ανεξάρτητων μεταβλητών. Η βάση δεδομένων που προέκυψε ωστόσο ήταν η ίδια με τον Πίνακας 4.10.

Εκπαίδευση μοντέλου δέντρου απόφασης

Για την εκπαίδευση του μοντέλου χρησιμοποιήθηκε η εντολή J48 {Rweka}. Το διάγραμμα του δέντρου απόφασης που δημιουργήθηκε φαίνεται παρακάτω (Διάγραμμα 4.27). Οι μεταβλητές που χρησιμοποιηθήκαν είναι οι εξής:

- Q12: Είχατε κάποιο όχημα (π.χ. αυτοκίνητο, μοτοσυκλέτα κ.λπ.) διαθέσιμο όταν αποφασίζατε εάν έπρεπε να εκκενώσετε;
- age: Ηλικία ερωτηθέντων
- household.18to65: Αριθμός ενηλίκων κάτω των 65 που διαμένουν στην κατοικία των ερωτηθέντων (num)
- household.under18: Αριθμός ανηλίκων που διαμένουν στην κατοικία των ερωτηθέντων (num)



Διάγραμμα 4.27 Διάγραμμα δέντρου απόφασης για την επιλογή μέσου κατά την εκκένωση

Έλεγχος μοντέλου δέντρου απόφασης

Ο έλεγχος του μοντέλου έγινε με 10-fold cross validation. Η μήτρα σύγχυσης και τα στατιστικά χαρακτηριστικά που προέκυψαν βρίσκονται παρακάτω:

```
=== 10 Fold Cross Validation ===

=== Summary ===

Correctly Classified Instances         105               80.7692 %
Incorrectly Classified Instances        25               19.2308 %
Kappa statistic                          0.4523
Mean absolute error                      0.2284
Root mean squared error                  0.4104
Relative absolute error                 62.4817 %
Root relative squared error             96.2632 %
Total Number of Instances              130

=== Detailed Accuracy By Class ===

               TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
               0,889    0,452    0,863      0,889    0,876      0,453    0,797     0,912     mycar
               0,548    0,111    0,607      0,548    0,576      0,453    0,797     0,534     other
Weighted Avg.  0,808    0,370    0,802      0,808    0,804      0,453    0,797     0,822

=== Confusion Matrix ===

  a  b   <-- classified as
 88 11 |  a = mycar
 14 17 |  b = other
```

Αξιολόγηση μοντέλου δέντρου απόφασης

Το μοντέλο βασίζεται αρχικά στη μεταβλητή Q12 που υποδηλώνει την ύπαρξη ή μη διαθέσιμου οχήματος. Κατηγοριοποιεί όπως είναι λογικό, όσους δεν διέθεταν όχημα στην κλάση other και στη συνέχεια εξετάζει την επιλογή μέσου ανάμεσα σε όσους διέθεταν όχημα με τη χρήση των μεταβλητών age, household.18to65 και household.under18 δηλαδή με βάση την ηλικία και τον αριθμό ανηλίκων και ενηλίκων μέχρι 65 ετών στην κατοικία. Η επίδραση της επιρροής των παραπάνω μεταβλητών στην απόφαση για την επιλογή μέσου φαίνεται στο Διάγραμμα 4.27. Το μοντέλο προβλέπει την επιλογή μέσου ελαφρώς καλύτερα από το αντίστοιχο μοντέλο λογιστικής παλινδρόμησης ωστόσο εξακολουθεί να μην έχει ικανοποιητικά στατιστικά χαρακτηριστικά αφού προβλέπει μέτρια την κλάση «other» και ο συντελεστής Κάππα κρίνεται ανεπαρκής.

4.5.3 Μοντέλο τυχαίου δάσους για πρόβλεψη επιλογής μέσου

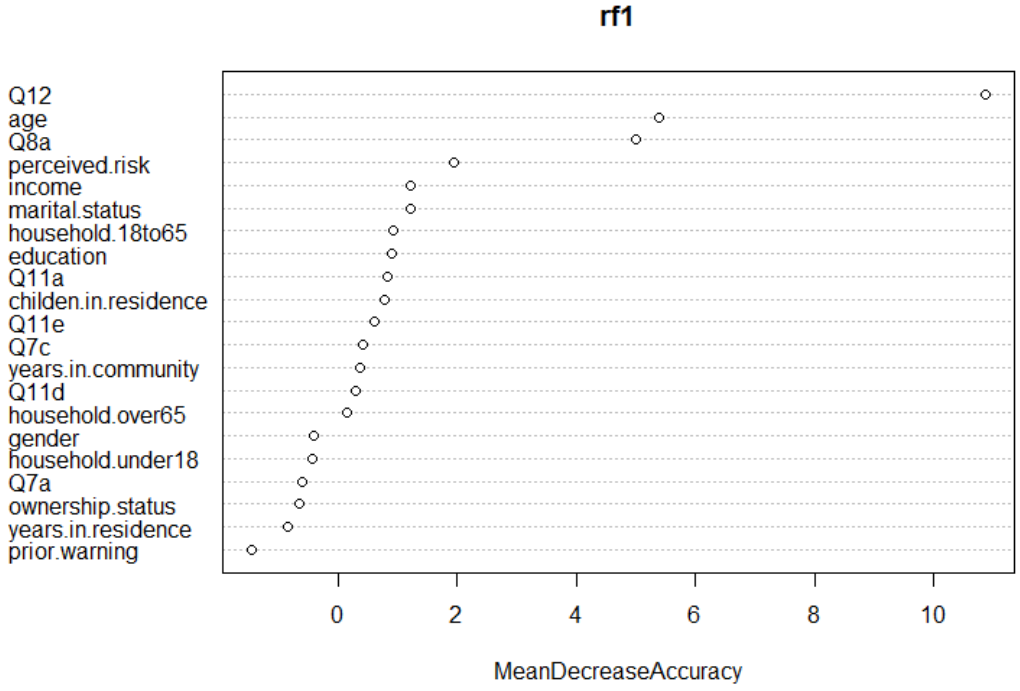
Προεπεξεργασία της βάσης δεδομένων

Για την προεπεξεργασία της βάσης δεδομένων ακολουθήθηκε σχεδόν η ίδια διαδικασία με την αντίστοιχη του μοντέλου λογιστικής παλινδρόμησης που παρουσιάζεται στο 4.5.1, με εξαίρεση τη μη αφαίρεση των μεταβλητών που παρουσιάζουν μεταξύ τους υψηλή συσχέτιση καθώς οι αλγόριθμοι δέντρων αποφάσεων δεν επηρεάζονται από τη συσχέτιση των ανεξάρτητων μεταβλητών. Η βάση δεδομένων που προέκυψε ωστόσο ήταν η ίδια με τον Πίνακας 4.10.

Εκπαίδευση μοντέλου τυχαίων δασών

Για την εκπαίδευση του αλγορίθμου επιλέχθηκαν οι 5 σημαντικότερες μεταβλητές. Για την επιλογή των μεταβλητών αρχικά έτρεξε ένας αλγόριθμος randomforest με το σύνολο των μεταβλητών και

στη συνέχεια χρησιμοποιήθηκε η εντολή importance {randomForest}. Στη συνέχεια έτρεξε νέος αλγόριθμος με τις 5 σημαντικότερες μεταβλητές, τα αποτελέσματα του οποίου φαίνονται στη συνέχεια (Διάγραμμα 4.28). Η κατάταξη των μεταβλητών έγινε μέσω του μεγέθους mean decrease accuracy. Το μέγεθος αυτό ποσοτικοποιεί τη σημασία μιας μεταβλητής μετρώντας την αλλαγή στην ακρίβεια πρόβλεψης, όταν οι τιμές της μεταβλητής μετατρέπονται τυχαία σε σύγκριση με τις αρχικές παρατηρήσεις.



Διάγραμμα 4.28 Σημαντικότητα μεταβλητών για τον αλγόριθμο τυχαίου δάσους

Οι μεταβλητές που επιλέχτηκαν λοιπόν είναι οι:

- Q12: Είχατε κάποιο όχημα (π.χ. αυτοκίνητο, μοτοσυκλέτα κ.λπ.) διαθέσιμο όταν αποφασίζατε εάν έπρεπε να εκκενώσετε;
- age: Ηλικία ερωτηθέντων (num)
- Q8a: Προσπαθήσατε να λάβετε περισσότερες πληροφορίες από κάποιον πριν αποφασίσετε να εκκενώσετε; - Όχι (0,1)
- perceived.risk: Μεταβλητή που προέκυψε από τον μέσο όρο των ερωτήσεων 6α με 6δ και δείχνει το πόσο πιθανά θεωρούσαν οι ερωτηθέντες τα ενδεχόμενα να τραυματιστούν εκείνοι ή οι γείτονές τους καθώς και να δημιουργηθούν σοβαρές ζημιές στην κατοικία τους ή των γειτόνων τους (1 έως 5 για «καθόλου πιθανό» έως «σχεδόν σίγουρο»)
- income: Οικογενειακό εισόδημα (1,2,3 για εισόδημα μικρότερο από 25.000, ανάμεσα σε 25.000 και 50.000 και άνω των 50.000 αντίστοιχα)

Έλεγχος μοντέλου τυχαίων δασών

Ο έλεγχος του μοντέλου έγινε με 10-fold cross validation. Η μήτρα σύγχυσης και τα στατιστικά χαρακτηριστικά που προέκυψαν βρίσκονται στους παρακάτω πίνακες:

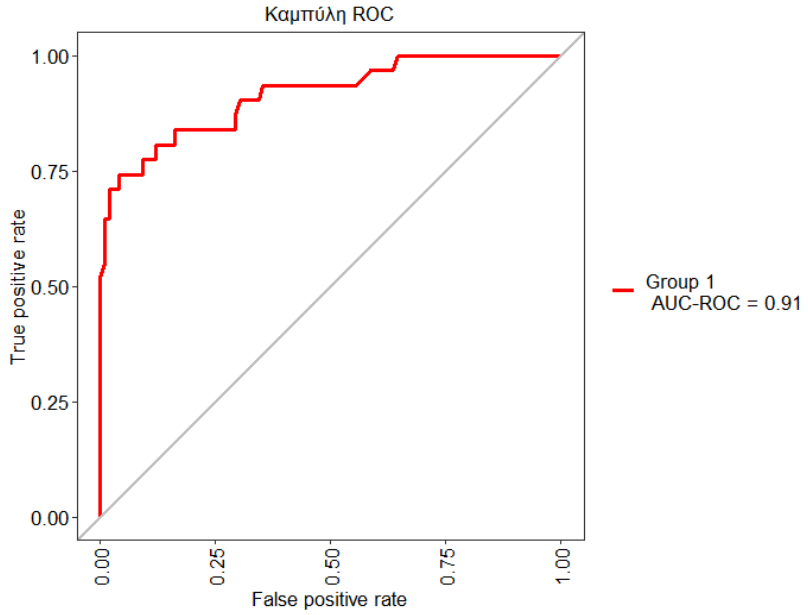
Πίνακας 4.13 Μήτρα σύγχυσης μοντέλου τυχαίου δάσους

Αριθμός δέντρων: 100		
Αριθμός μεταβλητών που δοκιμάστηκαν σε κάθε διαχωρισμό: 4		
	Actual negative	Actual positive
Predicted negative	23	6
Predicted positive	8	93

Πίνακας 4.14 Στατιστικά χαρακτηριστικά μοντέλου τυχαίου δάσους

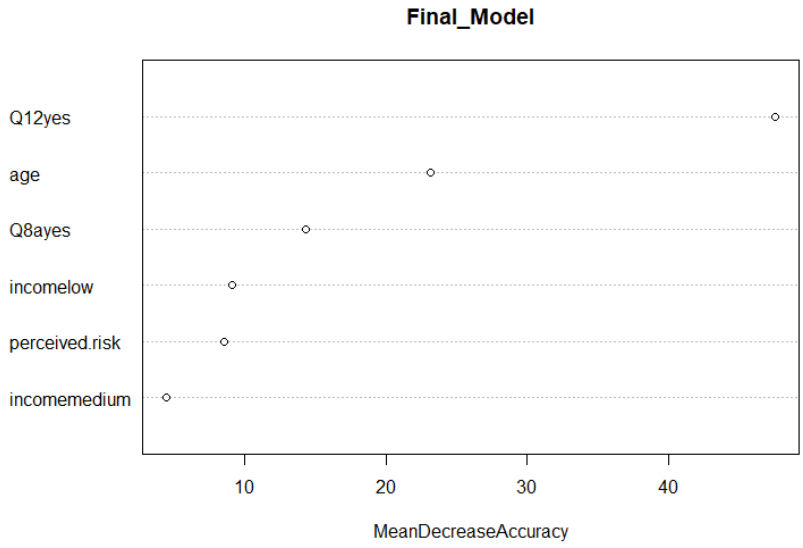
Ορθότητα (accuracy)	89,23%
Συντελεστής Κάππα	69,68%
Ευαισθησία (sensitivity)	93,94%
Εξειδικευτικότητα (specificity)	74,19%
Ακρίβεια (precision)	92,08%
Μέτρο F	93,00%

Η καμπύλη ROC του μοντέλου φαίνεται στο Διάγραμμα 4.29 . Η επιφάνεια που βρίσκεται κάτω από την καμπύλη ισούται με AUC= 0.91



Διάγραμμα 4.29 Καμπύλη ROC μοντέλου randomForest για την πρόβλεψη της επιλογής μέσου

Η σημαντικότητα των μεταβλητών φαίνεται στο Διάγραμμα 4.30



Διάγραμμα 4.30 Σημαντικότητα μεταβλητών για τον τελικό αλγόριθμο τυχαίου δάσους

	mycar	*other*	*MeanDecreaseAccuracy*	*MeanDecreaseGini*
Q12yes	42.994327	44.9435550	47.490458	14.186785
age	16.865443	19.4318209	23.125404	17.760715
Q8ayes	9.604920	12.6287969	14.349232	3.451152
perceived.risk	4.283669	8.4181777	8.551059	6.667898
incomelow	7.556254	5.4464965	9.121730	1.527708
incomemedium	5.385715	0.1024086	4.506400	1.644018

<u>Αξιολόγηση μοντέλου λογιστικής παλινδρόμησης</u>

Από το <u>Διάγραμμα 4.28</u> φαίνεται πως υπάρχουν λίγες μόνο μεταβλητές όπου φαίνεται να επηρεάζουν την επιλογή μέσου, γι' αυτό και αποφασίστηκε το τελικό μοντέλο να περιλαμβάνει μόνο τις 5 σημαντικότερες. Αυτές ήταν η Q12, age, Q8a, perceived.risk και income που αντιπροσωπεύουν αντίστοιχα την ύπαρξη ή μη διαθέσιμου οχήματος, την ηλικία των ερωτηθέντων, το αν προσπάθησαν ή όχι να λάβουν περισσότερες πληροφορίες πριν να εκκενώσουν, την αντίληψη του κινδύνου που είχαν για τη συγκεκριμένη απειλή καθώς και το ετήσιο οικογενειακό τους εισόδημα.

Το μοντέλο διαθέτει σημαντικά καλύτερα στατιστικά χαρακτηριστικά και από τα δύο προηγούμενα αφού είναι το μόνο που καταφέρνει να προβλέψει και τις δύο κλάσεις με ικανοποιητική ακρίβεια. Αναλυτικότερα:

- Έχει μεγάλη συνολική ακρίβεια, οπότε και υψηλό βαθμό πιστότητας για τη διαδικασία κατηγοριοποίησης
- Προβλέπει πολύ ικανοποιητικά τη θετική κλάση, δηλαδή την επιλογή IX

- Προβλέπει ικανοποιητικά και την αρνητική κλάση, δηλαδή την εκκένωση με άλλο τρόπο
- Έχει μεγάλο μέτρο F και ικανοποιητικό συντελεστή Κάππα
- Έχει πολύ μεγάλο εμβαδόν κάτω από την καμπύλη (AUC).

5 ΣΥΜΕΡΑΣΜΑΤΑ ΚΑΙ ΠΡΟΤΑΣΕΙΣ

5.1 Σύνοψη μεθοδολογίας και αποτελεσμάτων

Αντικείμενο της παρούσας διπλωματικής εργασίας αποτέλεσε η έρευνα της συμπεριφοράς των χρηστών οδικού δικτύου κατά την εκκένωση οικισμών λόγω πυρκαγιάς. Εκτενέστερη έρευνα έγινε για τις κρίσιμες αποφάσεις που λαμβάνουν οι άνθρωποι που καλούνται να εκκενώσουν μια περιοχή, δηλαδή την απόφαση για εκκένωση ή παραμονή αλλά και την απόφαση για επιλογή μέσου σε περίπτωση εκκένωσης. Η γνώση των παραπάνω είναι απαραίτητη για μια ολοκληρωμένη μελέτη εκκένωσης ώστε, σε συνδυασμό και με την τοπολογία και τη χωρική κατανομή του πληθυσμού, να καθοριστεί η ζήτηση των μετακινήσεων για ένα συγκοινωνιακό δίκτυο κατά την εκκένωση μιας περιοχής.

Τα απαραίτητα δεδομένα για την πραγματοποίηση της έρευνας συλλέχθηκαν μέσω ενός κατάλληλα διαμορφωμένου ερωτηματολογίου το οποίο συμπληρώθηκε από κατοίκους περιοχών που επλήγησαν από τις πυρκαγιές της Ανατολικής Αττικής το 2018. Οι ερωτήσεις του ερωτηματολογίου αφορούσαν τον τρόπο με τον οποίο οι ερωτηθέντες είχαν ενημερωθεί για την πυρκαγιά, τις αποφάσεις που πήραν κατά τη διάρκεια της απειλής, τα συναισθήματα και την αντίληψη που είχαν για το ρίσκο, τα μέτρα προστασίας που είχαν λάβει, τις αποφάσεις που θα έπαιρναν σε μία νέα ενδεχόμενη πυρκαγιά έχοντας ζήσει την εμπειρία καθώς και τα μέτρα προστασίας που έλαβαν ύστερα από την πυρκαγιά. Επιπλέον το ερωτηματολόγιο περιελάμβανε ερωτήσεις σχετικές με τα δημογραφικά χαρακτηριστικά των ερωτηθέντων ώστε να εξαχθούν συμπεράσματα ως προς τη συσχέτιση ορισμένων δημογραφικών χαρακτηριστικών με άλλες απαντήσεις του ερωτηματολογίου αλλά και να γίνει έλεγχος της αντιπροσωπευτικότητας του δείγματος.

Τα δεδομένα επεξεργάστηκαν με τη χρήση απλής στατιστικής αλλά και αλγορίθμων μηχανικής μάθησης. Συγκεκριμένα δημιουργήθηκαν μοντέλα λογιστικής παλινδρόμησης, δέντρων αποφάσεων και τυχαίων δασών που είχαν ως σκοπό τη διερεύνηση των κρισιμότερων παραγόντων που επηρεάζουν την απόφαση για εκκένωση η παραμονή καθώς και την επιλογή του μέσου κατά την εκκένωση. Στους πίνακες 5.1 και 5.2 παρουσιάζονται συγκεντρωτικά τα στατιστικά χαρακτηριστικά τους.

Πίνακας 5.1 Συγκεντρωτικός πίνακας αξιολόγησης μοντέλων πρόβλεψης της απόφασης για εκκένωση/παραμονή

Απόφαση για Εκκένωση/Παραμονή	Λογιστική Παλινδρόμηση	Δέντρο Απόφασης	Τυχαίο Δάσος
Ορθότητα (accuracy)	81%	86%	94%
Συντελεστής Κάππα	48%	66%	86%
Ευαισθησία (sensitivity)	91%	88%	97%
Εξειδικευτικότητα (specificity)	54%	80%	88%
Ακρίβεια (precision)	84%	92%	95%
Μέτρο F	87%	90%	96%

Πίνακας 5.2 Συγκεντρωτικός πίνακας αξιολόγησης μοντέλων πρόβλεψης της επιλογής μέσου

Απόφαση για Επιλογή Μέσου	Λογιστική Παλινδρόμηση	Δέντρο Απόφασης	Τυχαίο Δάσος
Ορθότητα (accuracy)	84%	81%	89%
Συντελεστής Κάππα	48%	45%	70%
Ευαισθησία (sensitivity)	96%	89%	94%
Εξειδικευτικότητα (specificity)	45%	55%	74%
Ακρίβεια (precision)	85%	86%	92%
Μέτρο F	90%	88%	93%

5.2 Βασικά συμπεράσματα

5.2.1 Γενικά συμπεράσματα

Από τη στατιστική ανάλυση της βάσης δεδομένων προκύπτουν τα εξής γενικά συμπεράσματα:

Η πλειοψηφία των ερωτηθέντων (86%) προσπάθησε να εκκενώσει την περιοχή κατοικίας του.

Οι άντρες έμειναν να προστατέψουν τις κατοικίες τους σε μεγαλύτερο ποσοστό σε σχέση με τις γυναίκες αφού μόλις το 80% των ανδρών που συμπλήρωσαν το ερωτηματολόγιο εκκένωσαν την κατοικία τους σε αντίθεση με τις γυναίκες που εκκένωσαν σε ποσοστό 91,4% (Στατιστικά σημαντική διαφορά σε 95% επίπεδο σημαντικότητας)

Υπάρχει συσχέτιση μεταξύ της ηλικίας και της απόφασης για εκκένωση καθώς παρατηρείται σημαντική διαφορά μεταξύ των διαφόρων ηλικιακών ομάδων ως προς την απόφαση για εκκένωση. Πρωτιά στο ποσοστό εκκένωσης σημειώνουν οι ηλικιωμένοι 70+, ακολουθούν οι νέοι μέχρι 29 ετών, και τις τελευταίες θέσεις καταλαμβάνουν οι ηλικιακές ομάδες 30-49 και 50-69

Μια έγκαιρη επίσημη εντολή εκκένωσης από την πολιτεία έχει πολύ σημαντική επιρροή στην απόφαση για εκκένωση. Οι ερωτηθέντες δήλωσαν πως σε μία νέα ενδεχόμενη μελλοντική πυρκαγιά δεν είναι ιδιαίτερα πιθανό (1,86/5) να εκκενώσουν την κατοικία τους χωρίς να έχει δοθεί επίσημη εντολή εκκένωσης. Αντίθετα δήλωσαν πως είναι εξαιρετικά πιθανό (4,45/5) να εκκενώσουν εφόσον έχει προηγηθεί μια εντολή εκκένωσης από την πολιτεία (Στατιστικά σημαντική διαφορά σε 99% επίπεδο σημαντικότητας)

Ενώ το 90,8% των ερωτηθέντων διέθετε όχημα διαθέσιμο κατά τη διάρκεια της πυρκαγιάς μόνο το 76,2% το χρησιμοποίησε ως μέσο διαφυγής. (Στατιστικά σημαντική διαφορά σε 95% επίπεδο σημαντικότητας). Επομένως η διάθεση οχήματος δεν αποτελεί τον μοναδικό παράγοντα για την επιλογή μέσου κατά την εκκένωση.

Το 74% των ερωτηθέντων που επέλεξαν να εκκενώσουν με τα πόδια κατέληξαν ανάμεσα στην πυρκαγιά ενώ μόνο το 47% εκείνων που διέφυγαν με το αυτοκίνητο βρέθηκαν ανάμεσα στις φλόγες (Στατιστικά σημαντική διαφορά σε 99% επίπεδο σημαντικότητας). Επομένως είναι σαφές πως η εκκένωση με τα πόδια ήταν πιο επικίνδυνη.

Η ηλικιακή ομάδα 18-29 επέλεξε το περπάτημα σε ποσοστό 45%, η ηλικιακή ομάδα 30-49 σε ποσοστό 20%, η ηλικιακή ομάδα 50-69 σε ποσοστό 16% και η ηλικιακή ομάδα 70+ σε ποσοστό 9%. Επομένως γίνεται σαφές πως όσο μεγαλύτερη ήταν η ηλικία, τόσο μεγαλύτερο ήταν και το ποσοστό εκκένωσης με Ι.Χ..

Το 65,6% των ερωτηθέντων δήλωσαν πως σε νέα ενδεχόμενη πυρκαγιά θα επέλεγαν να χρησιμοποιήσουν μόνο 1 όχημα σε αντίθεση με το υπόλοιπο 34,4% που δήλωσε πως θα προσπαθούσε να εκκενώσει με το σύνολο των διαθέσιμων οχημάτων της οικογενείας.

5.2.2 Παράγοντες που επηρεάζουν την απόφαση για εκκένωση/παραμονή

Όσον αφορά την απόφαση για εκκένωση/παραμονή, λήφθηκαν υπόψιν τα 3 μοντέλα μηχανικής μάθησης. Μεγαλύτερη βαρύτητα δόθηκε ωστόσο στα μοντέλα λογιστικής παλινδρόμησης, λόγω της επεξηγηματικότητάς του, και τυχαίου δάσους, λόγω της ακρίβειάς του. Από το μοντέλο λογιστικής παλινδρόμησης προέκυψε ο Πίνακας 4.7 όπου παρουσιάζονται οι σημαντικότεροι παράγοντες που εντόπισε το μοντέλο και η επίδραση αυτών. Στον Πίνακα 5.3 τα αποτελέσματα του παραπάνω μοντέλου συγκρίνονται με αυτά της διεθνούς βιβλιογραφίας όπως αυτά προέκυψαν στο κεφάλαιο 2.3.1.

Πίνακας 5.3 Σύγκριση αποτελεσμάτων με τη διεθνή βιβλιογραφία

Παράγοντας	**Επίδραση**	**Διεθνής Βιβλιογραφία**	**Επίδραση κατά τη διεθνή βιβλιογραφία**
Τύπος κυριότητας κατοικίας	Αρνητική για ιδιοκτήτες των ακινήτων	-	-
Ηλικία	Θετική για μεγαλύτερες ηλικίες	Alsnih et al. (2015)	Θετική για μεγαλύτερες ηλικίες
Φύλο	Θετική για τις γυναίκες	Alsnih et al. (2015), Mozumder et al. (2008)	Θετική για τις γυναίκες
Αριθμός ανηλίκων στην κατοικία	Αρνητική για μεγαλύτερο αριθμό ανηλίκων	Fischer III et al. (1995)	Αρνητική για μεγαλύτερο αριθμό ανηλίκων
Εισόδημα	Θετική για μεγαλύτερο εισόδημα	-	-
Αντίληψη κινδύνου	Θετική για μεγαλύτερη αντίληψη κινδύνου	Alsnih et al. (2015), Mozumder et al. (2008)	Θετική για μεγαλύτερη αντίληψη κινδύνου
Συμβουλή για εκκένωση από φίλο/συγγενή	Θετική για όσους έλαβαν συμβουλή για εκκένωση από φίλους/συγγενείς	-	-
Προσπάθεια για λήψη επιπλέον πληροφόρησης πριν την απόφαση για εκκένωση	Θετική για όσους δεν προσπάθησαν να λάβουν επιπλέον πληροφορίες πριν εκκενώσουν	-	-

Τα αποτελέσματα του μοντέλου δεν αντικρούονται από τη διεθνή βιβλιογραφία. Φαίνεται ωστόσο να λαμβάνονται υπόψιν και ορισμένοι νέοι παράγοντες που δεν αναφέρονται από τη διεθνή βιβλιογραφία. Επιπλέον σύμφωνα με τη διεθνή βιβλιογραφία υπάρχουν ακόμα 3 σημαντικοί παράγοντες που δεν λαμβάνονται υπόψιν από το μοντέλο. Ο πρώτος είναι τα χρόνια διαμονής στην κατοικία όπου παρόλο που υπήρχε σχετική ερώτηση, δεν επιλέχθηκε από το μοντέλο. Ο δεύτερος παράγοντας αφορά την ύπαρξη προειδοποίησης και επίσης δεν επιλέχθηκε από το μοντέλο ενώ υπήρχε σχετική ερώτηση. Ο τρίτος παράγοντας αφορά την ύπαρξη εμπειρίας σε αντίστοιχα περιστατικά στο παρελθόν, κάτι που δεν ρωτήθηκε στο ερωτηματολόγιο και επομένως είναι αδύνατο να εξετασθεί.

Ωστόσο, τα αποτελέσματα από το μοντέλο τυχαίου δάσους είχαν μια μικρή διαφοροποίηση σχετικά με την κατάταξη της σημαντικότητα των μεταβλητών, αφού ο παράγοντας «χρόνια διαμονής στην κατοικία» ενώ δεν συμπεριλήφθηκε στο μοντέλο λογιστικής παλινδρόμησης φαίνεται πως είναι εξαιρετικά σημαντικός στην πρόβλεψη της τελικής απόφασης με το μοντέλο τυχαίου δάσους. Επιπλέον σημαντική επίδραση στο μοντέλο φαίνεται πως έχει και η μεταβλητή prior.warning που δηλώνει την ύπαρξη προειδοποίησης που επίσης δεν συμπεριλήφθηκε στο μοντέλο λογιστικής παλινδρόμησης.

Φαίνεται λοιπόν πως το μοντέλο τυχαίου δάσους λαμβάνει υπόψιν του το σύνολο των παραγόντων που προκύπτουν από τη διεθνή βιβλιογραφία. Σε συνδυασμό με τα εξαιρετικά στατιστικά χαρακτηριστικά του και το γεγονός πως η σημαντικότητα των παραγόντων που προκύπτουν από τον αλγόριθμο randomForest είναι αρκετά πιο αξιόπιστη από την αντίστοιχη της λογιστικής παλινδρόμησης, αφού εκείνη στηρίζεται στην παραδοχή ότι οι μεταβλητές είναι μεταξύ τους ανεξάρτητες, οδηγεί στην επιλογή του μοντέλου randomForest ως το καταλληλότερο για την πρόβλεψη της απόφασης για εκκένωση/ παραμονή.

5.2.3 Παράγοντες που επηρεάζουν την επιλογή μέσου κατά την εκκένωση

Για την επιλογή μέσου λήφθηκε υπόψιν μονάχα το μοντέλο τυχαίου δάσους καθώς τα μοντέλα λογιστικής παλινδρόμησης και δέντρου απόφασης κρίθηκαν αναξιόπιστα λόγω μη ικανοποιητικών στατιστικών χαρακτηριστικών. Από το Διάγραμμα 4.30 φαίνεται πως υπάρχουν λίγες μόνο μεταβλητές που επηρεάζουν την επιλογή μέσου. Αυτές είναι κυρίως η Q12 και η age και στη συνέχεια οι Q8a, perceived.risk και income που αντιπροσωπεύουν αντίστοιχα :

- την ύπαρξη ή μη διαθέσιμου οχήματος
- την ηλικία
- την προσπάθεια για λήψη πληροφοριών πριν την εκκένωση
- την αντίληψη του κινδύνου που είχαν για τη συγκεκριμένη απειλή καθώς και
- το ετήσιο οικογενειακό τους εισόδημα

Φαίνεται λοιπόν πως ο σημαντικότερος παράγοντας για την επιλογή μέσου είναι η ύπαρξη διαθέσιμου οχήματος και στη συνέχεια η ηλικία. Συνδυάζοντας τα παραπάνω εύκολα μπορεί κανείς να συμπεράνει πως η πλειοψηφία όσων διέθεταν όχημα κατά την εκκένωση, το χρησιμοποίησε. Επίσης είναι ξεκάθαρη η επιρροή της ηλικίας στην επιλογή του μέσου καθώς φαίνεται πως όσο μικρότερη ήταν η ηλικία των ερωτηθέντων, τόσο πιθανότερο ήταν να επιλέξουν την εκκένωση με τα πόδια.

5.3 Προτάσεις για περαιτέρω έρευνα

Από την έρευνα που πραγματοποιήθηκε στην παρούσα διπλωματική εργασία προέκυψαν ορισμένες προτάσεις για περαιτέρω έρευνα.

Αρχικά, η διεξαγωγή παρόμοιας έρευνας σε άλλες περιοχές τις Ελλάδας με ποικίλα κοινωνικά και οικονομικά χαρακτηριστικά θα συνεισέφερε στην επιβεβαίωση ή διόρθωση των αποτελεσμάτων της συγκεκριμένης εργασίας ή στην ανίχνευση επιπλέον παραγόντων επιρροής τη συμπεριφοράς που σχετίζονται με πολιτισμικά και κοινωνικά κριτήρια.

Επιπλέον, σε έρευνες που θα είχαν ένα μεγαλύτερο δείγμα θα μπορούσε να πραγματοποιηθεί και προσομοίωση της κυκλοφορίας κατά τη διάρκεια της πυρκαγιάς, που θα οδηγούσε στον χαρακτηρισμό της ανθεκτικότητας του εξεταζόμενου δικτύου. Στη συγκεκριμένη εργασία συλλέχθηκαν ορισμένα στοιχεία που αφορούσαν την προέλευση και τον προορισμό των μετακινήσεων κατά την εκκένωση, όμως δεν χρησιμοποιήθηκαν. Τα στοιχεία αυτά θα μπορούσαν να συμβάλουν στην ορθή βαθμονόμηση συμπεριφορισιακών μοντέλων κυκλοφορίας που θα αποτελέσουν τη βάση για την ανάπτυξη στρατηγικών διαχείρισης κυκλοφορίας σε κρίσιμές καταστάσεις εκκένωσης.

ΒΙΒΛΙΟΓΡΑΦΙΑ

A.J. Pel, M.C. Bliemer, S.P. Hoogendoorn, A review on travel behavior modelling in dynamic traffic simulation models for evacuations, Transportation, 39 (1) (2012), pp. 97-123

H.W. Fischer III, G.F. Stine, B.L. Stoker, M.L. Trowbridge, E.M. Drain, Evacuation behavior: why do some evacuate while others do not? A case study of the Ephrata, Pennsylvania (USA) evacuation, Disaster Prevention and Management, 4 (4) (1995), pp. 30-36

H.W. Fischer III, G.F. Stine, B.L. Stoker, M.L. Trowbridge, E.M. Drain, Evacuation behavior: why do some evacuate while others do not? A case study of the Ephrata, Pennsylvania (USA) evacuation, Disaster Prevention and Management, 4 (4) (1995), pp. 30-36

Haizhong Wang, Alireza Mostafizi, Lori A. Cramer, Dan Cox, Hyoungsu Park, An agent-based model of a multimodal near-field tsunami evacuation: Decision-making and life safety, Transportation Research Part C: Emerging Technologies, Volume 64, 2016, Pages 86-100, ISSN 0968-090X

Hao-Che Wu, Michael K. Lindell, Carla S. Prater, Logistics of hurricane evacuation in Hurricanes Katrina and Rita, Transportation Research Part F: Traffic Psychology and Behaviour, Volume 15, Issue 4, 2012, Pages 445-461, ISSN 1369-8478

Hsu, Y. -., & Peeta, S. (2015). Clearance time estimation for incorporating evacuation risk in routing strategies for evacuation operations. Networks and Spatial Economics, 15(3), 743-764.

Huizhao Tu, Guus Tamminga, Hans Drolenga, Jeroen de Wit, Wouter van der Berg, Evacuation plan of the city of almere: simulating the impact of driving behavior on evacuation clearance time, Procedia Engineering, Volume 3, 2010, Pages 67-75, ISSN 1877-7058

Hunt, E.B. (1962). Concept learning: An information processing problem. New York: Wiley

J. Ross Quinlan , C4.5: Programs for Machine Learning ,Morgan Kaufmann Publishers, (1993)

J. McLennan, G. Elliott, M. Omodei, Householder decision-making under imminent wildfire threat: stay and defend or leave? Int. J. Wildland Fire, 21 (2012), pp. 915-925

J. Urata, E. Hato, Modeling the cooperation network formation process for evacuation systems design in disaster areas with a focus on Japanese megadisasters, Leadersh. Manag. Eng., 12 (2012), pp. 231-246

J.H. Friedman, (1977) A recursive partitioning decision rule for nonparametric classification. IEEE Trans. Computers, C-26, 404. Also available as Stanford Linear Accelerator Center Rep. SLAC-PUB-1373 (Rev. 1975).

J.P. van der Gun, A.J. Pel, B. van Arem, A general activity-based methodology for simulating multimodal transportation networks during emergencies, Eur. J. Transp. Infrastruct. Res., 16 (3) (2016), pp. 490-511

J.R. Quinlan, Discovering rules by induction from large collections of examples. In D. Michie (Ed.), *Expert systems in the micro electronic age*. Edinburgh University Press. (1979).

L. Breiman, J.H. Friedman, R.A. Olshen, and C.J. Stone, Classification and Regression Trees, Wadsworth, Belmont, CA. Republished by CRC Press, (1984)

Lindell, M.K., Prater, C.S. , Critical behavioral assumptions in evacuation time estimate analysis for private vehicles: Examples from hurricane research and planning (2007) Journal of Urban Planning and Development, 133 (1), pp. 18-29

M. Kubat, "Decision trees can initialize radial-basis function networks," in IEEE Transactions on Neural Networks, vol. 9, no. 5, pp. 813-821, Sept. 1998

M.A. Taylor, S.K. Freeman, A review of planning and operational models used for emergency evacuation situations in Australia, Procedia Eng., 3 (2010), pp. 3-14

M.L. Brachman, R.L. Church,Planning for a Disaster: a Review of the Literature with a Focus on Transportation Related Issues, University of California, Santa Barbara (2009)

Pamela Murray-Tuite, Brian Wolshon, Evacuation transportation modeling: An overview of research, development, and practice, Transportation Research Part C: Emerging Technologies, Volume 27, 2013, Pages 25-45, ISSN 0968-090X

R.W. Perry, M.K. Lindell, Preparedness for emergency response: guidelines for the emergency planning process, Disasters, 27 (4) (2003), pp. 336-350

S. Liu, P. Murray-Tuite, L. Schweitzer, Incorporating household gathering and mode decisions in large-scale no-notice evacuation modeling, Comput.-Aided Civil. Infrastruct. Eng., 29 (2) (2014), pp. 107-122

Zhao Zhang, Brian Wolshon, Pamela Murray-Tuite, A conceptual framework for illustrating and assessing risk, resilience, and investment in evacuation transportation systems, Transportation Research Part D: Transport and Environment, Volume 77, 2019, Pages 525-534, ISSN 1361-9209

ΠΑΡΑΡΤΗΜΑ Ι - ΕΡΩΤΗΜΑΤΟΛΟΓΙΟ

Ερωτηματολόγιο Εκκένωσης Οικισμών λόγω Πυρκαγιάς

1. Σε ποια κοινότητα/χωριό της Αττικής που απειλήθηκε από την πυρκαγιά του 2018 βρισκόσασταν;

_____

2. Ποια ήταν η πρώτη πηγή πληροφόρησης σας σχετικά με την πυρκαγιά; (Συμπληρώστε ένα από τα παρακάτω)

- ❑ Είδα ανθρώπους κοντά μου να εκκενώνουν την περιοχή (Πηγαίνετε στην ερώτηση 5)
- ❑ Είδα τη φωτιά να πλησιάζει (Πηγαίνετε στην ερώτηση 5)
- ❑ Με προειδοποίησαν αυτοπροσώπως (Πηγαίνετε στην ερώτηση 3)
- ❑ Ενημερώθηκα από τηλεφωνική κλήση ή γραπτό μήνυμα (Πηγαίνετε στην ερώτηση 3)
- ❑ Ενημερώθηκα από το ραδιόφωνο ή την τηλεόραση (Ποιον σταθμό)_____ (Πηγαίνετε στην ερώτηση 3)
- ❑ Άλλο (παρακαλώ εξηγήστε) _____ (Πηγαίνετε στην ερώτηση 3)

3. Ποιός σας ενημέρωσε πρώτα σχετικά με την πυρκαγιά; (Συμπληρώστε ένα από τα παρακάτω)

- ❑ Οι αρχές (π.χ. κάποιος αστυνομικός)
- ❑ Φίλος, συγγενής, γείτονας ή συνάδελφος
- ❑ Κάποιος ραδιοφωνικός ή τηλεοπτικός παρουσιαστής
- ❑ Κάποιος άλλος (παρακαλώ εξηγήστε) _____

4. Τι πληροφορίες λάβατε από την πρώτη αυτή πηγή; (Μπορείτε να συμπληρώσετε παραπάνω από μία επιλογή)

- ❑ Ποιά ήταν η απειλή (π.χ. η π
- ❑ Ποιές περιοχές θα επηρεάζονταν
- ❑ Τι δράση να λάβετε για την προστασία σας (π.χ. εκκένωση)
- ❑ Ποιές περιοχές θα ήταν ασφαλείς/Που έπρεπε να πάτε
- ❑ Που να λάβετε βοήθεια στην εκκένωση
- ❑ Που να λάβετε επιπλέον πληροφορίες
- ❑ Άλλο (παρακαλώ εξηγήστε) _____

5. Αφού λάβατε τις πρώτες πληροφορίες σχετικά με την πυρκαγιά, από ποιές άλλες πηγές λάβατε πληροφόρηση;

- ❑ Είδα την πυρκαγιά να πλησιάζει
- ❑ Είδα ανθρώπους κοντά μου να εκκενώνουν την περιοχή
- ❑ Με προειδοποίησαν αυτοπροσώπως
- ❑ Έλαβα πληροφορίες από τηλεφωνική κλήση
- ❑ Έλαβα πληροφορίες από το ραδιόφωνο
- ❑ Έλαβα πληροφορίες από την τηλεόραση
- ❑ Έλαβα πληροφορίες από γραπτό μήνυμα
- ❑ Έλαβα πληροφορίες μέσω ηλεκτρονικού ταχυδρομείου
- ❑ Έλαβα πληροφορίες μέσω ίντερνετ
- ❑ Έλαβα πληροφορίες από τα μέσα κοινωνικής δικτύωσης
- ❑ Άλλο (παρακαλώ εξηγήστε) _____

6. Αφού λάβατε οποιαδήποτε προειδοποίηση για την πυρκαγιά, πόσο πιθανό πιστεύατε πως ήταν η πυρκαγιά να...

		Καθόλου πιθανό				Σχεδόν σίγουρο
a.	δημιουργήσει σοβαρές ζημιές ή να καταστρέψει πολλά σπίτια στην κοινότητα;	1	2	3	4	5
b.	τραυματίσει ή να σκοτώσει πολλούς ανθρώπους στον οικισμό σας εάν δεν εκκένωναν;	1	2	3	4	5
c.	δημιουργήσει σοβαρές ζημιές ή να καταστρέψει το σπίτι σας	1	2	3	4	5
d	τραυματίσει ή να σκοτώσει εσάς ή την οικογένειά σας εάν δεν εκκενώνατε;	1	2	3	4	5

7. Λάβατε από κάποιον συμβουλή να καταφύγετε σε μια ασφαλέστερη τοποθεσία; (Μπορείτε νασυμπληρώσετε παραπάνω από μία επιλογή)

- ❑ Όχι
- ❑ Ναι, από κάποια αρχή (π.χ. αστυνομία)
- ❑ Ναι, από άλλους (παρακαλώ εξηγήστε) _____
- ❑ Ναι, από κάποιον φίλο, συγγενή ή γείτονα
- ❑ Ναι, από από τα Μέσα Μαζικής Ενημέρωσης

8. Προσπαθήσατε να λάβετε περισσότερες πληροφορίες από κάποιον πριν αποφασίσετε να εκκενώσετε; (Μπορείτε να συμπληρώσετε παραπάνω από μία επιλογή)

- ❑ Όχι
- ❑ Ναι, από τις αρχές (π.χ. αστυνομία)
- ❑ Ναι, από φίλους, συγγενείς, γείτονες
- ❑ Ναι, από τα ΜΜΕ
- ❑ Ναι, από άλλες πηγές (παρακαλώ εξηγήστε) _____

9. Εκκενώσατε το μέρος που βρισκόσασταν όταν απειλήθηκε από την πυρκαγιά; (Συμπληρώστε ένα από τα παρακάτω)

- ❑ Ναι, εκκένωσα πριν δοθεί επίσημη εντολή εκκένωσης (πηγαίνετε στην ερώτηση 11)
- ❑ Ναι, εκκένωσα αφού έλαβα επίσημη εντολή εκκένωσης (πηγαίνετε στην ερώτηση 10)
- ❑ Όχι, έμεινα εκεί που ήμουν και συνέχισα ό,τι έκανα (πηγαίνετε στην ερώτηση 17)
- ❑ Όχι, έμεινα εκεί που ήμουν και περίμενα επιπλέον πληροφόρηση (πηγαίνετε στην ερώτηση 17)
- ❑ Όχι, έμεινα να προστατεύσω την ιδιοκτησία μου (πηγαίνετε στην ερώτηση 17)
- ❑ Όχι, έλαβα κάποια άλλη δράση (παρακαλώ εξηγήστε και μετά πηγαίνετε στην ερώτηση 17)

_____

10. Πόσο χρόνο περιμένατε αφού προειδοποιηθήκατε μέχρι να εκκενώσετε; _____ Λεπτά

11. Κάνατε κάποια άλλη ενέργεια πριν εκκενώσετε; (Μπορείτε να συμπληρώσετε παραπάνω από μία επιλογή)

❑ Μάζεψα είδη πρώτης ανάγκης ❑ Προσπάθησα να εντοπίσω μέλη της οικογένειάς μου που δε βρίσκονταν μαζί μου
❑ Προσπάθησα να προστατέψω την περιουσία μου ❑ Προσπάθησα να προειδοποιήσω άλλους
❑ Άλλο (παρακαλώ εξηγήστε) _____

12. Είχατε κάποιο όχημα (π.χ. αυτοκίνητο, μοτοσυκλέτα κ.λ.π.) διαθέσιμο όταν αποφασίζατε εάν έπρεπε να εκκενώσετε;

❑ Όχι ❑ Ναι

13. Πώς εκκενώσατε την περιοχή σας; (Μπορείτε να συμπληρώσετε παραπάνω από μία επιλογή)

❑ Με τα πόδια ❑ Με όχημα φίλου/γείτονα ❑ Με δικό μου όχημα ❑ Με όχημα έκτακτης ανάγκης (π.χ. περιπολικό)
❑ Με Μέσα Μαζικής Μεταφοράς ❑ Άλλο (παρακαλώ εξηγήστε) _____

14. Όταν εκκενώσατε, πήγατε σε κάποια διαφορετική περιοχή από αυτή που ήσασταν όταν πληροφορηθήκατε για την πυρκαγιά; ❑ Όχι (πηγαίνετε στην ερώτηση 17) ❑ Ναι (πηγαίνετε στην ερώτηση 15)

15. Σε ποιά περιοχή/οικισμό πήγατε ; _____

16. Σε ποιό από τα παρακάτω μέρη πήγατε όταν εκκενώσατε; (Μπορείτε να συμπληρώσετε παραπάνω από μία επιλογή)

❑ Σε σπίτι συγγενή ❑ Σε σπίτι φίλου ❑ Σε επίσημο καταφύγιο ❑ Σε δημόσια τοποθεσία (π.χ. πάρκο)
❑ Σε εκκλησία ❑ Άλλο (παρακαλώ εξηγήστε) _____

17. Βρεθήκατε σε οποιαδήποτε στιγμή ανάμεσα στην πυρκαγιά; ❑ Ναι ❑ Όχι

18. Ποιά από τα παρακάτω είχατε έτοιμα πριν την πυρκαγιά;

❑ Απόθεμα νερού 3 ημερών ❑ Απόθεμα φαγητού 3 ημερών ❑ Προκαθορισμένο μέρος για εκκένωση
❑ Οικογενειακό σχέδιο έκτακτης ανάγκης ❑ Ραδιόφωνο με μπαταρίες ❑ Κουτί πρώτης ανάγκης με προμήθειες

19. Πόσο πιθανό πιστεύετε πως είναι το ενδεχόμενο στα επόμενα 10 χρόνια μία πυρκαγιά να...

a. ...προκαλέσει σοβαρή ζημιά σε ιδιοκτησίες στην κοινότητά σας
b. ...τραυματίσει ή να σκοτώσει κατοίκους εάν δεν εκκενώσουν;
c. ...προκαλέσει σοβαρές ζημιές στην κατοικία σας;
d. ...τραυματίσει ή σκοτώσει εσάς η μέλη της οικογένειάς σας εάν δεν εκκενώσετε;

20. Πόσο συχνά...	Καθημερινά	Εβδομαδιαία	Μηνιαία	Ετήσια	Ποτέ
a. σκέφτεσαι τις πυρκαγιές	1	2	3	4	5
b. μιλάς σε άλλους ανθρώπους σχετικά με τις πυρκαγιές	1	2	3	4	5

21. Σε ποιό βαθμό το ρίσκο πιθανής μελλοντικής πυρκαγιάς σας κάνει να νιώθετε…	Καθόλου				Σε μεγάλο βαθμό
a. ...ανησυχία	1	2	3	4	5
b. ...νευρικότητα	1	2	3	4	5
c. ...φόβο	1	2	3	4	5
d. ...θυμό	1	2	3	4	5
e. ...απογοήτευση	1	2	3	4	5
f. ...ενόχληση	1	2	3	4	5

22. Πόσο πιθανό πιστεύετε πως είναι να συμβεί κάθε ένα από τα παρακάτω ύστερα από μια πυρκαγιά αλλά πριν απειληθεί το σπίτι σας...	Καθόλου πιθανό				Εξαιρετικά πιθανό
a. ...να λάβετε μία ακριβή πρόβλεψη του χρόνου άφιξης της πυρκαγιάς στην κατοικία σας	1	2	3	4	5
b. ...να λάβετε μία επίσημη προειδοποίηση εκκένωσης	1	2	3	4	5
c. ...να προετοιμαστείτε για την εκκένωση	1	2	3	4	5
d. ...να μπορείτε να καταφύγετε σε μίασφαλή τοποθεσία	1	2	3	4	5

23. Σε τι βαθμό έχεις λάβει πληροφορίες σχετικές με την απειλή πυρκαγιάς στην περιοχή αυτή από...	Καθόλου				Σε μεγάλο βαθμό
a ..έντυπα μέσα (εφημερίδες, περιοδικά, φυλλάδια);	1	2	3	4	5
b. ..ηλεκτρονικά μέσα (τηλεόραση, ραδιόφωνο);	1	2	3	4	5
c. ...ίντερνετ (π.χ. ειδησεογραφικές ιστοσελίδες, ιστοσελίδες έκτακτων αναγκών);	1	2	3	4	5
d. ...μέσα κοινωνικής δικτύωσης	1	2	3	4	5
e. ...συζητήσεις με φίλους, συγγενείς, γείτονες, συναδέλφους;					
f. ...συναντήσεις με ομάδες της κοινότητας;					
g. ...συναντήσεις με κρατικές και τοπικές αρχές εκτάκτων αναγκών;					

24. Πόσο πιθανό είναι να λάβετε κάποια από τις παρακάτω δράσεις εάν είστε στην κατοικία σας όταν μια πυρκαγιά πλησιάζει την περιοχή σας;	Καθόλου πιθανό				Σε μεγάλο βαθμό
a. Εκκένωση πριν δοθεί επίσημη εντολή εκκένωσης για την περιοχή μου	1	2	3	4	5
b. Εκκένωση μόλις μάθω για επίσημη εντολή εκκένωσης για την περιοχή μου	1	2	3	4	5
c. Μένω να προστατέψω την ιδιοκτησία μου ακόμα κι αν έχει εκδοθεί επίσημη εντολή εκκένωσης	1	2	3	4	5
d. Μένω να δω εάν το ρίσκο είναι μεγάλο ακόμα κι αν έχει εκδοθεί επίσημη εντολή εκκέννωσης	1	2	3	4	5

25. Εάν λάβετε επίσημη εντολή εκκένωσης και αποφασίσετε να εκκενώσετε την κατοικία σας λόγω πυρκαγιάς, με ποιον τρόπο θα το κάνετε; Αυτοκίνητο Περπάτημα Ποδήλατο Άλλο _____

26. Σε περίπτωση που εκκενώσετε με αυτοκίνητο, πόσα οχήματα σκοπεύετε να πάρετε; _____ αυτοκίνητα

27. Έχετε σχεδιάσει που να πάτε σε περίπτωση που εκκενώσετε το σπίτι σας για μια πυρκαγιά; ❑ Ναι ❑ Όχι

28. Έχετε προγραμματίσει ποιά διαδρομή θα πάρετε σε περίπτωση που εκκενώσετε το σπίτι σας για μια πυρκαγιά; ❑ Ναι ❑ Όχι

29. Πόσο χρόνο θα σας πάρει να προετοιμαστείτε για να εκκενώσετε (π.χ. να μαζευτείτε όλα τα μέλη της οικογένειας, να πάρετε το κουτί πρώτης ανάγκης κ.λ.π.) για μια πυρκαγιά; _____ λεπτά

30. Πόσο χρόνο θα σας πάρει να φτάσετε σε ασφαλή περιοχή, αφού αφήσετε την κατοικία σας χρησιμοποιώντας την προτιμούμενη διαδρομή εκκένωσης από μια πυρκαγιά; _____ λεπτά

31. Έχετε κάνει κάτι από τα παρακάτω για να προετοιμαστείτε για μία πυρκαγιά;	Όχι	Όχι αλλά σκοπεύω	Ναι
a. Έχω κάνει διαχείρηση της βλάστησης	1	2	3
b. Έχω κάνει την κατοικία μου πιο ανθεκτική στη φωτιά	1	2	3
c. Έχω προετοιμάσει ένα οικογενειακό σχέδιο έκτακτης ανάγκης	1	2	3
d. Έχω μαζέψει πληροφορίες σχετικές με το τι να κάνω κατά τη διάρκεια απειλής πυρκαγιάς	1	2	3
e. Έχω μαζέψει πληροφορίες σχετικές με το πού να πάω κατά την εκκένωση λόγω πυργαγιάς	1	2	3
f. Έχω φτιάξει ένα κουτί πρώτης ανάγκης για απειλές πυρκαγιάς	1	2	3

32. Πόσα χρόνια μένετε: Στην κοινότητα που ζείτε τώρα _____ Στην τωρινή κατοικία σας _____

33. Νοικιάζετε ή σας ανήκει η κατοικία στην οποία διαμένετε; ❑ Νοικιάζω ❑ Μου ανήκει

34. Πόσο χρονών είστε; _____ χρονών

35. Ποιό είναι το φύλο σας; ❑ Άντρας ❑ Γυναίκα ❑ Προτιμώ να αυτοπροσδιορίζομαι

36. Ποιά είναι η οικογενειακή σας κατάσταση; ❑ Παντρεμένος/η ❑ Άγαμος/η ❑ Διαζευγμένος/η ❑ Χήρος/α

37. Πόσα άτομα στην κατοικία σας είναι: ___ Κάτω των 18 ___ 18 65 χρονών ___ πάνω από 65 χρονών

38. Ποιές από τις παρακάτω κατηγορίες περιγράφουν καλύτερα το ετήσιο εισόδημα του νοικοκυριού σας προ φόρων;
❑ Λιγότερο από 25.000€ ❑ 25.001€-50.000€ ❑ 50.001€-75.000€ ❑ 75.001€-100.000€ ❑ πάνω από100.000€

39. Ποιό από παρακάτω αντικατοπτρίζει το επίπεδο της εκπαίδευσης που έχετε ολοκληρώσει;
❑ Δημοτικό ❑ Γυμνάσιο ❑ Λύκειο ❑ Ιδιωτική/Δημόσια σχολη Ι.Ε.Κ ❑ Πτυχίο ΑΕΙ/ΤΕΙ ❑ Μεταπτυχιακό/Διδακτορικό

40. Έχετε κάποιο άλλο σχόλιο σχετικά με την ετοιμότητα εκκένωσης νοικοκυριών;

_____

Σας ευχαριστούμε πολύ που συμμετείχατε σε αυτή την έρευνα.

ΠΑΡΑΡΤΗΜΑ ΙΙ – ΚΩΔΙΚΕΣ

Μοντέλο Λογιστικής Παλινδρόμησης για πρόβλεψη απόφασης για εκκένωση/παραμονή

```
#_____________________DATA PREPROCESSING_____________________
set.seed(123)
#Import dataset
data <- read.csv2("evacuation_data_logit.csv")
# Remove Zero and Near Zero-Variance Predictors
library(caret)
nzv <- nearZeroVar(data)
data <- data[, -nzv]
# Calculate correlation matrix
descrCor <- cor(data)
# Print correlation matrix and look at max correlation
print(descrCor)
summary(descrCor[upper.tri(descrCor)])
# find attributes that are highly corrected
highlyCorrelated <- findCorrelation(descrCor, cutoff=0.5)
# Indentifying Variable Names of Highly Correlated Variables
highlyCorCol <- colnames(data)[highlyCorrelated]
highlyCorCol
 # Remove highly correlated variables and create a new dataset
 data <- data[, -which(colnames(data) %in% highlyCorCol)]
 dim(data)
#Oversampling
table(data$Q9)
library(ROSE)
data <- ovun.sample(Q9 ~ ., data = data, method = "over",N = 180)$data
table(data$Q9)
```

```
#_____________________________________LINEAR CLASSIFIER_____________________________
library(glmulti)
res <-glmulti(Q9 ~ ., data = data,
        level = 1,              # No interaction considered
        method = "g",           # "g" for genetic algorithm and 'h" for Exhaustive approach
        crit = "aicc",          # c-AIC as criterion
        confsetsize = 10,       # Keep 10 best models
        plotty = F, report = F, # No plot or interim reports
        fitfunction = "glm")    # glm function
print(res)
plot(res)
library(caret)
# define training control
train_control <- trainControl(method = "cv", number = 10)
# train the model
mylogit <- train(as.factor(Q9) ~1 + ownership_status + age + gender + household.under18 +
income + perceived.risk + Q7c + Q8a,
              data = data,
              trControl = train_control,
              method = "glm",
              family=binomial())
# Cross Validation scores
summary(mylogit)
probs <-predict(mylogit, data)
class <- data.frame(response1 = data$Q9,
                 predicted1 = probs)
confusionMatrix(as.factor(class$predicted1),as.factor(class$response1),
             positive = "1")
#ROC
library(pROC)
mroc<-roc(class$response1,factor(class$predicted1, ordered = TRUE), plot=TRUE)
auc(mroc)
```

Μοντέλο δέντρου απόφασης για πρόβλεψη απόφασης για εκκένωση/παραμονή

```
#_____________________________DATA PREPROCESSING_____________________________
#set seed
set.seed(123)
#Import dataset
data <- read.csv2("evacuation_data.csv")
# Remove Zero and Near Zero-Variance Predictors
library(caret)
nzv <- nearZeroVar(data)
data <- data[, -nzv]
#Oversampling
table(data$Q9)
library(ROSE)
data <- ovun.sample(Q9 ~ ., data = data, method = "over",N = 180)$data
table(data$Q9)
#_____________________________DECISION TREES_____________________________
#Categorical variables as factors
col_names <- names(data[,1:9])
data[,col_names] <- lapply(data[,col_names] , factor)
data$years.in.community<- as.numeric(data$years.in.community)
data$years.in.residence<- as.numeric(data$years.in.residence)
data$age <- as.numeric(data$age)
data$household.18to65 <- as.numeric(data$household.18to65)
data$household.under18 <- as.numeric(data$household.under18)
data$household.over65 <- as.numeric(data$household.over65)
data$perceived.risk <- as.numeric(data$perceived.risk)
#Train a Decision Tree
library(RWeka)
dt <- J48(Q9 ~ . , data = data)
```

```
summary(dt)
library(partykit)
plot(as.party(dt))
#Evaluate a Decision Tree
eval_dt <- evaluate_Weka_classifier(dt, numFolds = 10, complexity = FALSE,
                                    class = TRUE)
eval_dt
```

Μοντέλο τυχαίου δάσους για πρόβλεψη απόφασης για εκκένωση/παραμονή

```
#_____________________________DATA PREPROCESSING_____________________________
#set seed
set.seed(123)
#Import dataset
data <- read.csv2("evacuation_data.csv")
# Remove Zero and Near Zero-Variance Predictors
library(caret)
nzv <- nearZeroVar(data)
data <- data[, -nzv]
#Oversampling
table(data$Q9)
library(ROSE)
data <- ovun.sample(Q9 ~ ., data = data, method = "over",N = 180)$data
table(data$Q9)
#_____________________________RANDOM FOREST_____________________________
#Categorical variables as factors
col_names <- names(data[,1:9])
data[,col_names] <- lapply(data[,col_names] , factor)
data$years.in.community<- as.numeric(data$years.in.community)
```

```
data$years.in.residence<- as.numeric(data$years.in.residence)
data$age <- as.numeric(data$age)
data$household.18to65 <- as.numeric(data$household.18to65)
data$household.under18 <- as.numeric(data$household.under18)
data$household.over65 <- as.numeric(data$household.over65)
data$perceived.risk <- as.numeric(data$perceived.risk)
#Train Random Forest
library(randomForest)
rf1 <-randomForest(Q9~.,data=data, importance=TRUE,ntree=100)
#Evaluate variable importance
imp = importance(rf1, type=1)
imp <- data.frame(predictors=rownames(imp),imp)
# Order the predictor levels by importance
library(plyr)
imp.sort <- arrange(imp,desc(MeanDecreaseAccuracy))
imp.sort$predictors <- factor(imp.sort$predictors,levels=imp.sort$predictors)
# Select the top 10 predictors
important<- imp.sort[1:10,]
print(important)
# Plot Important Variables
varImpPlot(rf1, type=1)
myvars<- as.vector(important$predictors)
myvars
# Subset data with 10 independent and 1 dependent variables
data =  cbind(Q9=data$Q9, data[,myvars])
rf  <- train(as.factor(Q9)~., tuneLength = 3, data = data, method ="rf",
             importance = TRUE,
             trControl = trainControl(method = "cv",
                                      number = 10,
                                      savePredictions = "final",
```

```
                                classProbs = T))
print(rf)
summary(rf)
#confusion matrix
rf$pred
rf$pred[order(rf$pred$rowIndex),2]
confusionMatrix(rf$pred[order(rf$pred$rowIndex),2], data$Q9,positive = "yes")
# #Evaluate variable importance
Final_Model <- rf$finalModel
importance(Final_Model)
varImpPlot(Final_Model,type=1)
library(MLeval)
x <- evalm(rf, plots = "r", title= "Καμπύλη ROC")
## get AUC and other metrics
x$stdres
```

Μοντέλο Λογιστικής Παλινδρόμησης για πρόβλεψη επιλογής μέσου

```
#_____________________________DATA PREPROCESSING_____________________________
set.seed(123)
#Import dataset
data <- read.csv2("modechoice_data_logit.csv")
# Remove Zero and Near Zero-Variance Predictors
library(caret)
nzv <- nearZeroVar(data)
data <- data[, -nzv]
#fill N/A with value -1 (NA only categorical data)
index <- is.na(data)
data[index] <- -1
```

```
# Calculate correlation matrix
descrCor <- cor(data)
# Print correlation matrix and look at max correlation
print(descrCor)
summary(descrCor[upper.tri(descrCor)])
# find attributes that are highly corrected
highlyCorrelated <- findCorrelation(descrCor, cutoff=0.5)
# Indentifying Variable Names of Highly Correlated Variables
highlyCorCol <- colnames(data)[highlyCorrelated]
highlyCorCol
# Remove highly correlated variables and create a new dataset
data <- data[, -which(colnames(data) %in% highlyCorCol)]
dim(data)
#remove rows with N/A in dependent variable Q13
data1 <- subset(data, Q13==-1)
data <- data[!row.names(data) %in% row.names(data1),]
#_____________________________LINEAR CLASSIFIER_____________________________
# library(glmulti)
 res <-glmulti(Q13 ~ ., data = data,
        level = 1,              # No interaction considered
        method = "g",           # "g" for genetic algorithm and 'h" for Exhaustive approach
        crit = "aicc",          # c-AIC as criterion
        confsetsize = 10,       # Keep 10 best models
        plotty = F, report = F, # No plot or interim reports
        fitfunction = "glm")    # glm function
 print(res)
 plot(res)
library(caret)
# define training control
train_control <- trainControl(method = "cv", number = 10)
```

```
# train the model on training set
mylogit <- train(as.factor(Q13) ~ 1  + age + gender + household.18to65 + Q8a + Q12,
                 data = data,
                 trControl = train_control,
                 method = "glm",
                 family=binomial())
# Cross Validation scores
summary(mylogit)
probs <-predict(mylogit, data)
class <- data.frame(response1 = data$Q13,
                    predicted1 = probs)
confusionMatrix(as.factor(class$predicted1),as.factor(class$response1),positive = "0")
#ROC
library(pROC)
mroc<-roc(class$response1,factor(class$predicted1, ordered = TRUE), plot=TRUE)
auc(mroc)
```

<u>Μοντέλο δέντρου απόφασης για πρόβλεψη επιλογής μέσου</u>

```
#_____________________________DATA PREPROCESSING_____________________________
#set seed
set.seed(123)
#Import dataset
data <- read.csv2("modechoice_data.csv")
#remove rows with N/A in dependent variable Q13
data<- data[which(data$Q13 != ""), ]
#Categorical variables as factors
col_names <- names(data[,1:25])
data[,col_names] <- lapply(data[,col_names] , factor)
```

```
data$years.in.community<- as.numeric(data$years.in.community)
data$years.in.residence<- as.numeric(data$years.in.residence)
data$age <- as.numeric(data$age)
data$household.18to65 <- as.numeric(data$household.18to65)
data$household.under18 <- as.numeric(data$household.under18)
data$household.over65 <- as.numeric(data$household.over65)
data$perceived.risk <- as.numeric(data$perceived.risk)
# Remove Zero and Near Zero-Variance Predictors
library(caret)
nzv <- nearZeroVar(data)
data <- data[, -nzv]
#_____________________DECISION TREES_____________________
#Train a Decision Tree
library(RWeka)
dt <- J48(Q13 ~ . , data = data)
summary(dt)
library(partykit)
plot(as.party(dt))
#Evaluate a Decision Tree
eval_dt <- evaluate_Weka_classifier(dt, numFolds = 10, complexity = FALSE,
                                    class = TRUE)
eval_dt
```

<u>Μοντέλο τυχαίου δάσους για πρόβλεψη επιλογής μέσου</u>

```
#_____________________DATA PREPROCESSING_____________________
#set seed
set.seed(123)
```

```
#Import dataset
data <- read.csv2("modechoice_data.csv")
#remove rows with N/A in dependent variable Q13
data<- data[which(data$Q13 != ""), ]
#Categorical variables as factors
col_names <- names(data[,1:25])
data[,col_names] <- lapply(data[,col_names] , factor)
data$years.in.community<- as.numeric(data$years.in.community)
data$years.in.residence<- as.numeric(data$years.in.residence)
data$age <- as.numeric(data$age)
data$household.18to65 <- as.numeric(data$household.18to65)
data$household.under18 <- as.numeric(data$household.under18)
data$household.over65 <- as.numeric(data$household.over65)
data$perceived.risk <- as.numeric(data$perceived.risk)
# Remove Zero and Near Zero-Variance Predictors
library(caret)
nzv <- nearZeroVar(data)
data <- data[, -nzv]
#_________________________________RANDOM FOREST_________________________
#Train Random Forest
library(randomForest)
rf1 <-randomForest(as.factor(Q13)~.,data=data, importance=TRUE,ntree=100)
#Evaluate variable importance
imp = importance(rf1, type=1)
imp <- data.frame(predictors=rownames(imp),imp)
# Order the predictor levels by importance
library(plyr)
imp.sort <- arrange(imp,desc(MeanDecreaseAccuracy))
imp.sort$predictors <- factor(imp.sort$predictors,levels=imp.sort$predictors)
# Select the top 10 predictors
```

```
important<- imp.sort[1:5,]
print(important)
# Plot Important Variables
varImpPlot(rf1, type=1)
myvars<- as.vector(important$predictors)
myvars
# Subset data with 10 independent and 1 dependent variables
data =  cbind(Q13=data$Q13, data[,myvars])
#train model
rf  <- train(as.factor(Q13)~., tuneLength = 3, data = data, method ="rf",
            importance = TRUE,
            metric = "Kappa",
            trControl = trainControl(method = "cv",
                                     number = 10,
                                     savePredictions = "final",
                                     classProbs = T))
print(rf)
summary(rf)
#confusion matrix
rf$pred
rf$pred[order(rf$pred$rowIndex),2]
confusionMatrix(rf$pred[order(rf$pred$rowIndex),2], data$Q13,positive = "mycar")
# #Evaluate variable importance
Final_Model <- rf$finalModel
importance(Final_Model)
varImpPlot(Final_Model,type=1)
library(MLeval)
x <- evalm(rf, plots = "r", title= "Καμπύλη ROC")
```