

Προγραμματιστική Εργασία Πρόβλεψη κόστους ασφάλισης οχημάτων

Χαρά Τσίρκα, Πρόδρομος Αβραμίδης, and Γεώργιος
Γεροντίδης

{ctsirka, pavramidis, ggerontidis}@e-ce.uth.gr
8 εξάμηνο



Τμήμα Ηλεκτρολόγων Μηχανικών & Μηχανικών
Υπολογιστών
Πανεπιστήμιο Θεσσαλίας, Βόλος

Εξόρυξη Δεδομένων 2023-24
Διδάσκον: Μ.Βασιλακόπουλος

Μάιος 2024

1 Εισαγωγή

Η εργασία μας επικεντρώνεται στην πρόβλεψη του κόστους ασφάλισης μηχανοκίνητων οχημάτων. Η ανάλυση αυτή αποτελεί ένα κρίσιμο ζήτημα στον τομέα της ασφάλισης, καθώς επιτρέπει στους ασφαλιστές να προσδιορίζουν με μεγαλύτερη ακρίβεια τα ασφαλιστικά ασφάλιστρα, λαμβάνοντας υπόψη διάφορους παράγοντες που επηρεάζουν το κόστος. Η διαδικασία της εργασίας ξεκινά με την προ-επεξεργασία των δεδομένων, κατά την οποία πραγματοποιήθηκε εξερευνητική ανάλυση (exploratory analysis) για τον προσδιορισμό των κριτηρίων διαχωρισμού των δεδομένων. Κατά τη διάρκεια αυτής της ανάλυσης, μετρήθηκε ο βαθμός επίδρασης κάθε χαρακτηριστικού (feature) του συνόλου δεδομένων στα αποτελέσματα. Με τη βοήθεια διαγραμμάτων, καταφέραμε να επιλέξουμε τον κατάλληλο διαχωρισμό των δεδομένων για περαιτέρω ανάλυση. Στη συνέχεια, η εργασία θα προχωρήσει στη δημιουργία και αξιολόγηση των μοντέλων πρόβλεψης, λαμβάνοντας υπόψη την είσοδο των χρηστών, ενώ θα ακολουθήσει η οπτικοποίηση και η αξιολόγηση των αποτελεσμάτων.

2 Περιγραφή dataset

Το dataset το οποίο επιλέξαμε αποτελείται από 30 μεταβλητές (columns) και 105555 εγγραφές. Στους παρακάτω πίνακες δίνεται μία σύντομη περιγραφή της κάθε μεταβλητής:

Μεταβλητή	Περιγραφή
ID	Εσωτερικός αριθμός αναγνώρισης που εκχωρείται σε κάθε ετήσια σύμβαση που επισημοποιείται από έναν ασφαλισμένο. Κάθε ασφαλισμένος μπορεί να έχει πολλές σειρές στο σύνολο δεδομένων, που αντιπροσωπεύουν διαφορετικές προσόδους του προϊόντος.
Date_start_contract	Ημερομηνία έναρξης του συμβολαίου (HH/MM/YYYY).
Date_last_renewal	Ημερομηνία τελευταίας ανανέωσης του συμβολαίου (HH/MM/YYYY).
Date_next_renewal	Ημερομηνία επόμενης ανανέωσης του συμβολαίου (HH/MM/YYYY).
Distribution_channel	Κανάλι μέσω του οποίου έγινε το ασφαλιστήριο, 0: για Πράκτορα, 1: για Ασφαλιστικοί μεσίτες.
Date_birth	Ημερομηνία γέννησης του ασφαλισμένου που δηλώνεται στο ασφαλιστήριο (HH/MM/YYYY).
Date_driving_licence	Ημερομηνία έκδοσης της άδειας οδήγησης του ασφαλισμένου (HH/MM/YYYY).

Μεταβλητή	Περιγραφή
Seniority	Συνολικός αριθμός ετών που ο ασφαλισμένος έχει συνδεθεί με την ασφαλιστική οντότητα, υποδεικνύοντας το επίπεδο αρχαιότητάς του.
Policies_in_force	Συνολικός αριθμός συμβολαίων που κατείχε ο ασφαλισμένος στην ασφαλιστική οντότητα κατά την περίοδο αναφοράς.
Max_policies	Μέγιστος αριθμός συμβολαίων που είχε ποτέ σε ισχύ ο ασφαλισμένος με τον ασφαλιστικό φορέα.
Max_products	Μέγιστος αριθμός προϊόντων που κατέχει ο ασφαλισμένος ταυτόχρονα σε οποιαδήποτε δεδομένη χρονική στιγμή.
Lapse	Αριθμός πολιτικών που ο πελάτης έχει ακυρώσει ή έχει ακυρωθεί λόγω μη πληρωμής κατά το τρέχον έτος λήξης, εξαιρουμένων αυτών που έχουν αντικατασταθεί από άλλο συμβόλαιο.
Date_Lapse	Ημερομηνία ακύρωσης της σύμβασης (HH/MM/YYYY).
Payment	Τελευταία μέθοδος πληρωμής της πολιτικής 1: εξαμηνιαία πληρωμή, 0: ετήσια πληρωμή
Premium	Καθαρό ποσό ασφαλιστρού που σχετίζεται με το ασφαλιστήριο συμβόλαιο κατά τη διάρκεια του τρέχοντος έτους.
Cost_claims_year	Συνολικό κόστος ζημιών που πραγματοποιήθηκαν για το ασφαλιστήριο συμβόλαιο κατά τη διάρκεια του τρέχοντος έτους.
N_claims_year	Συνολικός αριθμός ζημιών που πραγματοποιήθηκαν για το ασφαλιστήριο συμβόλαιο κατά τη διάρκεια του τρέχοντος έτους.
N_claims_history	Συνολικός αριθμός απαιτήσεων που υποβλήθηκαν καθ' όλη τη διάρκεια του ασφαλιστηρίου συμβολαίου.
R_Claims_history	Παρέχει μια ένδειξη του ιστορικού συχνότητας αξιώσεων του ασφαλιστηρίου.
Type_risk	Τύπος κινδύνου για κάθε όχημα 1: μοτοσικλέτα, 2: μικρά φορτηγά, 3: επιβατικά οχήματα, 4: αγροτικά οχήματα
Area	0: αγροτική περιοχή, 1: αστική περιοχή (>30.000 κάτοικοι όσον αφορά τις κυκλοφοριακές συνθήκες)
Second_driver	1: περισσότεροι από ένας δηλωμένοι οδηγοί 0: μόνο ένας δηλωμένος οδηγός
Year_matriculation	Έτος καταχώρησης οχήματος (EEEE)
Power	Ίπποι δύναμης οχήματος
Cylinder_capacity	Χωρητικότητα κυλίνδρων του οχήματος
Value_vehicle	Αξία αγοράς οχήματος στις 31/12/2019
N_doors	Αριθμός θυρών οχήματος
Type_fuel	Τύπος καυσίμου, P: πετρέλαιο, D: ντίζελ
Length	Μήκος του οχήματος σε m
Weight	Βάρος του οχήματος σε kg

3 Data preprocessing

Το πρώτο βήμα για την προεπεξεργασία των δεδομένων ήταν να κρατήσουμε μία γραμμή για κάθε 'ID'. Σε ένα 'ID' μπορεί να αντιστοιχούν περισσότερες από μία γραμμές που αντιπροσωπεύουν το ίδιο συμβόλαιο του ίδιου πελάτη για διαφορετική χρονική περίοδο. Έτσι, για κάθε 'ID' κρατάμε την τελευταία ανανέωση του συμβολαίου, δηλαδή την γραμμή με το μεγαλύτερο χρονολογικά 'last_renewal_date'. Έπειτα, στη θέση του 'premium' υπολογίζουμε και τοποθετούμε τον μέσο όρο των 'premium' όλων των γραμμών με κοινό 'ID'.

Το δεύτερο βήμα ήταν η επεξεργασία όλων των ημερομηνιών. Ειδικότερα, οι στήλες 'Date_birth', 'Date_driving_license', 'Date_start_contract', 'Date_last_renewal', 'Date_next_renewal', 'Date_lapse' δίνονται στην μορφή HH/MM/YYYY. Αρχικά, για κάθε μία από αυτές τις μεταβλητές κρατήσαμε το έτος (YYYY) και στην συνέχεια πραγματοποιώντας τις κατάλληλες αφαιρέσεις δημιουργήσαμε νέες στήλες στο dataset που πήραν την θέση αυτών που αναφέρθηκαν νωρίτερα. Έτσι, δημιουργήσαμε τις στήλες: 'Age' που προσδιορίζει την ηλικία του πελάτη, 'Years_driving' που προσδιορίζει πόσα χρόνια οδηγεί ο πελάτης, 'Year_on_road' που προσδιορίζει πόσα χρόνια κυκλοφορεί το κάθε όχημα υπό την κατοχή συγκεκριμένου πελάτη, 'Policy Duration' που υποδεικνύει την διάρκεια του εκάστοτε συμβολαίου σε χρόνια και 'Years_on_policy' που προσδιορίζει πόσα χρόνια ο πελάτης βρίσκεται στον ίδιο τύπο συμβολαίου. Πρέπει να σημειωθεί πως το dataset περιέχει δεδομένα μέχρι και το 2019. Για να έχουμε μια σωστή εικόνα των χρονολογιών σε όλες αυτές τις μεταβλητές που δημιουργήσαμε, χρησιμοποιήσαμε ως σημείο αναφοράς την χρονολογία τελευταίας ανανέωσης του συμβολαίου. Για παράδειγμα η μεταβλητή 'Age' προκύπτει από την αφαίρεση: 'Age' = 'Date_last_renewal' - 'Date_birth'.

Επιπλέον, δημιουργήσαμε μία ακόμη νέα στήλη με όνομα 'accidents' για να υπάρχει μια συσχέτιση μεταξύ των αριθμών των ατυχημάτων με τα χρόνια που ένας πελάτης είναι ασφαλισμένος στην εταιρεία.

Διαχειριστήκαμε την απουσία τιμών με δύο τρόπους. Στην στήλη 'Length' αντικαταστήσαμε τα κενά πεδία με τον μέσο όρο των τιμών της στήλης. Στην στήλη 'Type_fuel' αντικαταστήσαμε τα κενά πεδία με την τιμή 'Unknown'.

Μετά από δοκιμές διαπιστώσαμε πως κάποιες μεταβλητές του dataset δεν συνεισέφεραν καθόλου στην βελτίωση της απόδοσης και παραλείφθηκαν. Οι στήλες που χρησιμοποιήθηκαν τελικά είναι οι: 'Seniority', 'Premium', 'Type_risk', 'Area', 'Power', 'Second_driver', 'Years_on_road', 'R_claims_history', 'Years_on_policy', 'accidents', 'Value_vehicle', 'Age', 'Years_driving', 'Distribution_channel', 'N_claims_history', 'Cylinder_capacity', 'Weight', 'Length', 'Type_fuel', 'Payment', 'Contract_year', 'Policies_in_force', 'Lapse'.

4 Εκπαίδευση μοντέλου

XGBoost Parameters

(Στις παρενθέσεις βρίσκεται το συνηθισμένο εύρος τιμών, δηλαδή το εύρος που έγιναν οι δοκιμές για να πετύχουμε το μέγιστο επιθυμητό αποτέλεσμα.)

Objective (linear):

Το 'reg:squarederror' είναι το προεπιλεγμένο και ευρέως χρησιμοποιούμενο loss function για regression tasks, παρέχοντας ένα σαφές πλαίσιο για τη βελτιστοποίηση του μοντέλου XGBoost. Συγκεκριμένα, είναι κατάλληλο για την πρόβλεψη ασφάλιστρων, καθώς στοχεύει στην επίτευξη υψηλής ακρίβειας στην πρόβλεψη της συνεχούς τιμής των ασφάλιστρων.

Eval_metric ('rmse', 'mae'):

Για το eval_metric, επιλέξαμε το 'rmse' (Root Mean Squared Error) επειδή είναι ένα ευρέως αποδεκτό μέτρο για regression tasks. Αναλυτικότερα, δίνει μεγαλύτερη βαρύτητα σε μεγαλύτερες αποκλίσεις (λάθη) μεταξύ πραγματικών και προβλεπόμενων τιμών, γεγονός που μπορεί να είναι χρήσιμο στην ασφαλιστική πρόβλεψη όπου τα μεγάλα λάθη μπορεί να είναι πιο σοβαρά.

Learning_rate (0.001-0.3):

Ο ρυθμός μάθησης (learning rate) επηρεάζει πόσο γρήγορα προσαρμόζεται το μοντέλο κατά την εκπαίδευση. Επιλέχθηκε η τιμή 0.01 έπειτα από δοκιμές, διότι όντας χαμηλή σχετικά τιμή επιβραδύνει τη διαδικασία μάθησης, μειώνοντας την πιθανότητα υπερβολικής προσαρμογής (overfitting).

Max_depth (3-10):

Το μέγιστο βάθος των δέντρων στο μοντέλο. Ένας μεγαλύτερος αριθμός επιτρέπει πιο περίπλοκα δέντρα, αλλά μπορεί να οδηγήσει σε overfitting, επομένως επιλέχθηκε η τιμή 6.

Min_child_weight (1-10):

Ελάχιστο βάρος που πρέπει να έχει ένας κόμβος για να διαχωριστεί. Επηρεάζει την ευαισθησία του μοντέλου στις διακυμάνσεις στα δεδομένα. Ύστερα από δοκιμές προτιμήθηκε ο αριθμός βάρους 5.

Subsample (0.5-1):

Αναφέρεται στο ποσοστό των δειγμάτων που θα χρησιμοποιηθούν σε κάθε επανάληψη της εκπαίδευσης. Ένα ποσοστό 0.8 σημαίνει ότι το 80

Colsample_bytree (0.5-1):

Το ποσοστό των χαρακτηριστικών (columns) που θα επιλεγθούν τυχαία για να κατασκευάσουν κάθε δέντρο στο μοντέλο. Ένα ποσοστό 0.8 σημαίνει ότι το 80

N_estimators (100-2000):

Ο αριθμός των δέντρων που θα κατασκευαστούν στο μοντέλο. Δοκιμάστηκαν διάφοροι αριθμοί, συμπεριλαμβανομένων μεγαλύτερων από 1000, αλλά δεν υπήρχε μεγάλη βελτίωση στην απόδοση του μοντέλου, ενώ ο χρόνος εκπαίδευσης αυξανόταν σημαντικά. Έτσι, ως optimal θεωρήθηκε η τιμή 1000.

Seed:

Για να καταφέρουμε να εξασφαλίσουμε επαναληψιμότητα στα δεδομένα μας και να αφαιρεθεί η τυχαιότητα, επιλέξαμε έναν σταθερό αριθμό seed 42.

Random forest parameters:

N_estimators (100-1000):

Επιλέχθηκε η τιμή 700, διότι παρατηρήθηκε ότι, παρόλο που η αύξηση της τιμής βελτίωνε την απόδοση του RF, ο χρόνος εκτέλεσης αυξανόταν σημαντικά λόγω της δημιουργίας πολλών δέντρων.

Max_depth (10-20):

Επιλέχθηκε η τιμή 15, καθώς η αύξηση της τιμής αυξάνει τον κίνδυνο overfitting, αν και προσφέρει σημαντική βελτίωση στην απόδοση του RF.

Min_samples_split (2-10):

Επιλέχθηκε η τιμή 4, που βρίσκεται στη μέση του εύρους τιμών, για να ισορροπήσει μεταξύ της πρόληψης του overfitting και της διατήρησης της απόδοσης του μοντέλου.

Min_samples_leaf (2-10):

Επιλέχθηκε η τιμή 2, διότι παρατηρήθηκε ότι η αύξηση της τιμής αυτής παραμέτρου βελτίωνε μόνο την απόδοση στο type_risk4, ενώ έπεφτε η συνολική απόδοση σε κάθε άλλο τύπο οχήματος.

Max_features ('sqrt', 'log2', ή float 0.1-1):

Συνήθως, σε αλγορίθμους RF προτιμάται η τιμή 'sqrt', και αυτή επιλέχθηκε και από εμάς για καλύτερη απόδοση.

Bootstrap (True, False):

Επιλέξαμε να έχουμε bootstrap, καθώς βελτιώνονται σημαντικά τα αποτελέσματα σε σύγκριση με την επιλογή της μη ύπαρξης bootstrap.

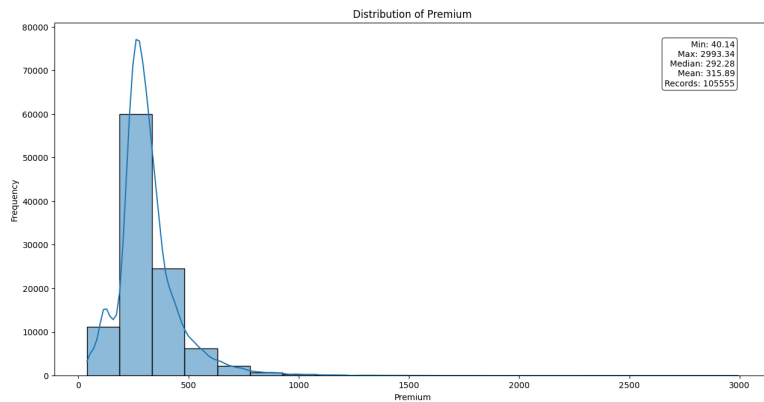
N_jobs (-1, 1, αριθμός πυρήνων):

Με την επιλογή της τιμής -1, χρησιμοποιούμε όλους τους διαθέσιμους πυρήνες.

Random_state:

Πρόκειται για τιμή τυχαιότητας όπως το seed στο XGBoost. Έχει επιλεγεί και εδώ η τιμή 42.

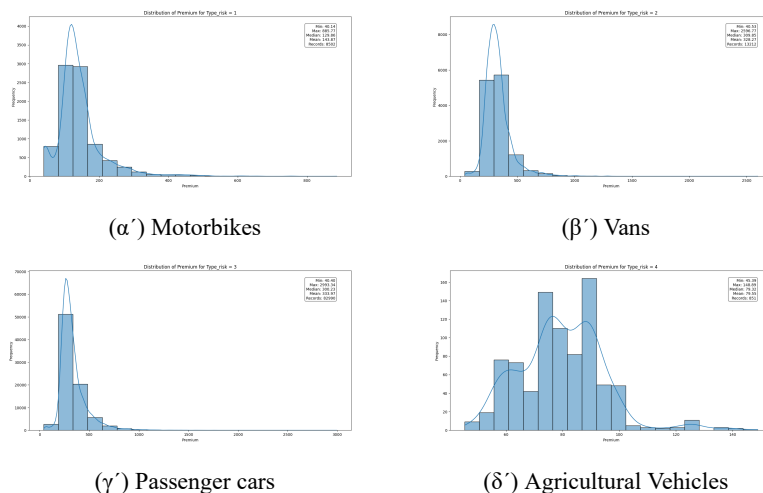
5 Διαγράμματα



Τα διαγράμματα

Παρατήρηση 1 Στο πρώτο διάγραμμα παρουσιάζεται η κατανομή των ασφαλίσεων για όλες τις καταχωρίσεις. Τα επόμενα διαγράμματα δείχνουν την κατανομή των ασφαλίσεων για κάθε κατηγορία οχημάτων ξεχωριστά.

Εύκολα διαπιστώνεται από το πρώτο ολικό διάγραμμα ότι οι ασφαλιστικές τιμές πάνω από 500 είναι ελάχιστες και δεν επηρεάζουν σημαντικά το τελικό αποτέλεσμα. Ωστόσο, όταν κατηγοριοποιήσαμε τα δεδομένα, παρατηρήσαμε ότι η διασπορά των τιμών στα



Εικ. 1: Comparison of Different Vehicle Types

αγροτικά οχήματα ήταν μεγαλύτερη, με αποτέλεσμα η διαφοροποίηση του μοντέλου ανά κατηγορία οχήματος να είναι απαραίτητη.

6 User Interface

Το user interface της εφαρμογής αναπτύχθηκε με χρήση της Python και ειδικότερα του framework Kivy, καθώς και της συλλογής από γραφικά στοιχεία KivyMD. Στην εφαρμογή μας υπάρχουν πέντε διαφορετικές "οθόνες":

1. Η οθόνη του login
2. Η οθόνη συμπλήρωσης στοιχείων του πελάτη που πρόκειται να ασφαλίσει το όχημά του.
3. Η οθόνη συμπλήρωσης στοιχείων του οχήματος που πρόκειται να ασφαλιστεί.
4. Η οθόνη συμπλήρωσης στοιχείων παλαιότερων συμβολαίων που είχε ο πελάτης στην εταιρεία.
5. Η οθόνη παρουσίασης της προτεινόμενης ετήσιας τιμής χρέωσης του πελάτη με βάση τα στοιχεία που συμπληρώθηκαν.

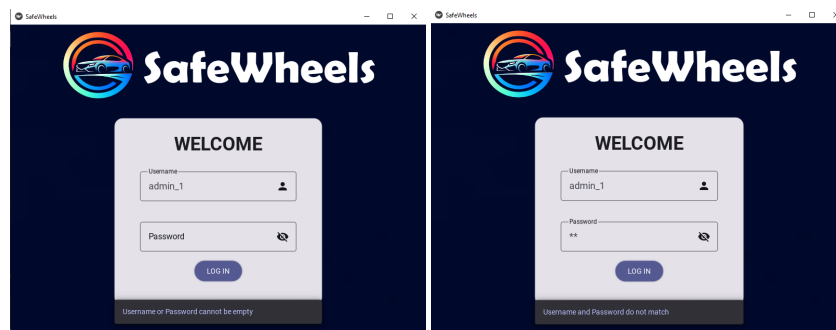
6.1 Login

Η οθόνη του login είναι η αρχική οθόνη της εφαρμογής [Εικ. 2] στην οποία ο χρήστης - υπάλληλος της εταιρείας θα πρέπει να συμπληρώσει τα σωστά στοιχεία συνδεσής του (username και password) και στην συνέχεια να πατήσει το κουμπί "LOG IN". Σε περίπτωση που ένα από τα δύο πεδία μένει κενό, ο χρήστης θα λάβει το αντίστοιχο μήνυμα λάθους [Εικ. 3α']. Σε περίπτωση που τα στοιχεία σύνδεσης δεν είναι σωστά, ο χρήστης θα λάβει διαφορετικό μήνυμα λάθους [Εικ. 3β'].

Έχει προβλεφθεί και δημιουργηθεί μόνο ένας λογαριασμός υπαλλήλου. Έτσι, για να πάμε στην επόμενη οθόνη θα πρέπει να συμπληρώσουμε στο πεδίο username: **admin_1** και στο πεδίο password: **12345**. Αφού συμπληρώσουμε αυτά τα στοιχεία σωστά και πατήσουμε το κουμπί "LOG IN" μεταφερόμαστε στην δεύτερη οθόνη [Εικ. 4].



Εικ. 2: Οθόνη login



(α') Κενό πεδίο

(β') Λάθος στοιχεία

Εικ. 3: Μηνύματα λάθους στο login

6.2 Στοιχεία πελάτη

Στην οθόνη συμπλήρωσης στοιχείων του πελάτη [Εικ. 4], ο χρήστης-υπάλληλος της εταιρείας θα πρέπει να συμπληρώσει τα στοιχεία του ανθρώπου που ενδιαφέρεται να

ασφαλίσει το όχημά του. Ειδικότερα, υπάρχουν δύο πεδία επιλογής ημερομηνίας ("Date of Birth", "Licence Issue Date"), τα οποία όταν επιλεγθούν εμφανίζεται ένα ημερολόγιο [Εικ. 5α'] προκειμένου να επιλεγθεί η κατάλληλη ημερομηνία. Πατώντας το πεδίο "Area" ο χρήστης βλέπει ένα μενού δύο επιλογών [Εικ. 5β'] από τις οποίες θα πρέπει να επιλέξει μία. Στο πεδίο "Seniority" ο χρήστης θα πρέπει να πληκτρολογήσει έναν αριθμό που προσδιορίζει τα χρόνια που ο συγκεκριμένος πελάτης είναι ασφαλισμένος στην εταιρεία.

Αφού συμπληρωθούν όλα τα πεδία ο χρήστης θα πρέπει να πατήσει το κουμπί "NEXT" για να μεταφερθεί στην επόμενη οθόνη [Εικ. 6]. Σε περίπτωση που επιθυμεί να πάει στην οθόνη του login [Εικ. 2] θα πρέπει να πατήσει το κουμπί "GO BACK".

Εικ. 4: Οθόνη συμπλήρωσης στοιχείων πελάτη

(α') Ημερολόγιο

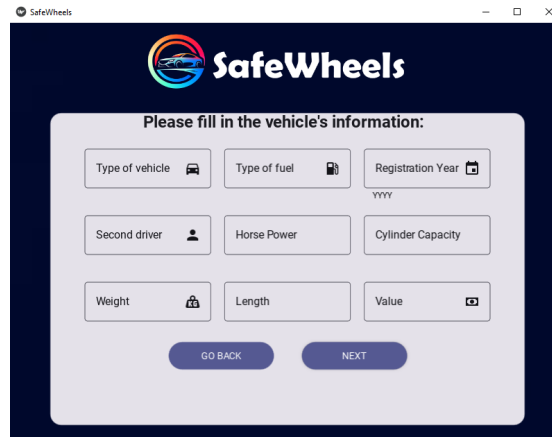
(β') Μενού επιλογών

Εικ. 5: Παραδείγματα εισόδου

6.3 Στοιχεία οχήματος

Σε αυτήν την οθόνη [Εικ. 6], ο χρήστης-υπάλληλος της εταιρείας θα πρέπει να συμπληρώσει τα στοιχεία του οχήματος το οποίο επιθυμεί να ασφαλίσει ο πελάτης. Ειδικότερα, για τα πεδία "Type of vehicle", "Type of fuel", "Second driver" ο χρήστης θα πρέπει να πατήσει καθένα από αυτά τα πεδία και να επιλέξει μία από τις επιλογές που εμφανίζονται στο κάθε μενού επιλογών (ενδεικτικά [Εικ. 7α']). Στα υπόλοιπα πεδία ο χρήστης θα πρέπει να πληκτρολογήσει έναν αριθμό σύμφωνα με τις υποδείξεις του κάθε πεδίου (ενδεικτικά [Εικ. 7β']).

Για να μεταβεί στην επόμενη οθόνη [Εικ. 8] θα πρέπει να πατήσει το κουμπί "NEXT" ενώ για να πάει στην προηγούμενη οθόνη [Εικ. 4] συμπλήρωσης των στοιχείων του πελάτη θα πρέπει να πατήσει το κουμπί "GO BACK".



The image shows a web application window titled "SafeWheels". The main heading is "Please fill in the vehicle's information:". Below this, there are nine input fields arranged in a grid. The first row contains "Type of vehicle" (with a car icon), "Type of fuel" (with a fuel pump icon), and "Registration Year" (with a calendar icon and a "YYYY" placeholder). The second row contains "Second driver" (with a person icon), "Horse Power", and "Cylinder Capacity". The third row contains "Weight" (with a scale icon), "Length", and "Value" (with a price tag icon). At the bottom of the form are two buttons: "GO BACK" and "NEXT".

Εικ. 6: Οθόνη συμπλήρωσης στοιχείων οχήματος προς ασφάλιση

(α') Παράδειγμα μενού επιλογών

(β') Παράδειγμα συμπλήρωσης πεδίου

Εικ. 7: Παραδείγματα εισόδου

6.4 Στοιχεία συμβολαίου

Στην οθόνη συμπλήρωσης στοιχείων συμβολαίου [Εικ. 8] ο χρήστης-υπάλληλος της εταιρείας θα πρέπει να συμπληρώσει στοιχεία που αφορούν παλαιότερα συμβόλαια που είχε ο πελάτης στην εταιρεία. Υπάρχουν τρία πεδία επιλογής ημερομηνίας ("Start Contract", "Last renewal", "Next renewal"), τα οποία όταν επιλεγθούν εμφανίζεται ημερολόγιο (ενδεικτικά [Εικ. 9α']). Τα πεδία "Distribution Channel" και "Payment" είναι πεδία που κατά την επιλογή τους εμφανίζεται μενού επιλογών (ενδεικτικά [Εικ. 9β']), ενώ τα υπόλοιπα πρέπει να συμπληρωθούν με πληκτρολόγηση σύμφωνα με τις υποδείξεις των πεδίων.

Για την μετάβαση στην προηγούμενη οθόνη [Εικ. 6] ο χρήστης θα πρέπει να πατήσει το κουμπί "GO BACK". Διαφορετικά, βρισκόμαστε στο σημείο που έχουν συμπληρωθεί όλα τα στοιχεία και ο χρήστης θα πρέπει να πατήσει το κουμπί "CALCULATE" προκειμένου να μεταβεί στην τελευταία οθόνη [Εικ. 10].

The screenshot shows the 'SafeWheels' logo at the top. Below it, a light blue box contains the text 'Please fill in the contract information:'. Inside this box are nine input fields arranged in a 3x3 grid. The first three fields are 'Start Contract', 'Last renewal', and 'Next renewal', each with a calendar icon and a placeholder 'MM/DD/YYYY'. The next three are 'Distribution Channel', 'Payment', and 'Claims', each with a car icon. The last three are 'Policies', 'Lapse', and 'Ratio claims', each with a car icon. At the bottom of the box are two buttons: 'GO BACK' and 'CALCULATE'.

Εικ. 8: Οθόνη συμπλήρωσης στοιχείων οχήματος προς ασφάλιση

The left screenshot shows the 'SafeWheels' form with a calendar pop-up for the 'Start Contract' field. The calendar is for May 2024, and the date '10' is selected. The right screenshot shows the same form with a dropdown menu open for the 'Payment' field. The dropdown shows two options: 'Half-yearly' and 'Annual'.

(α') Παράδειγμα ημερολογίου

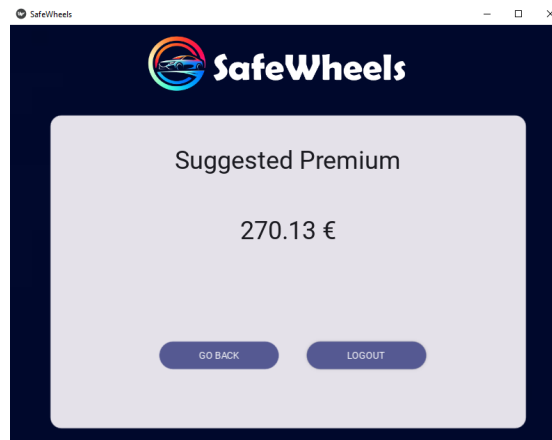
(β') Παράδειγμα μενού επιλογών

Εικ. 9: Παραδείγματα εισόδου

6.5 Οθόνη προτεινόμενης χρέωσης

Σε αυτήν την οθόνη [Εικ. 10], ο χρήστης-υπάλληλος της εταιρείας βλέπει την προτεινόμενη τιμή ασφάλισης του οχήματος που πρόκειται να ασφαλίσει ο πελάτης. Πατώντας το κουμπί "GO BACK" μπορεί να μεταβεί στην προηγούμενη οθόνη [Εικ. 8], ενώ πατώντας το κουμπί "LOGOUT" μεταβαίνει στην αρχική οθόνη της εφαρμογής [Εικ. 2].

Αξίζει να σημειωθεί πως ο χρήστης μπορεί να πηγαίνει προς τα πίσω και να αλλάζει στοιχεία ανά πάσα στιγμή. Κάθε φορά που πατά το κουμπί "CALCULATE" της προτελευταίας οθόνης [Εικ. 8], όλα τα ανανεωμένα στοιχεία δίνονται στο μοντέλο προκειμένου να βγει η προτεινόμενη τιμή χρέωσης.



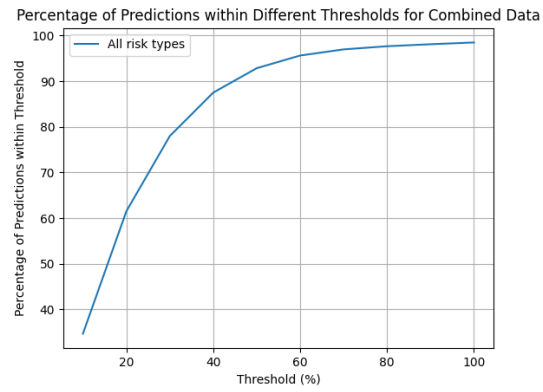
Εικ. 10: Οθόνη παρουσίασης προτεινόμενης τιμής χρέωσης

6.6 Γενικές παρατηρήσεις για το ui

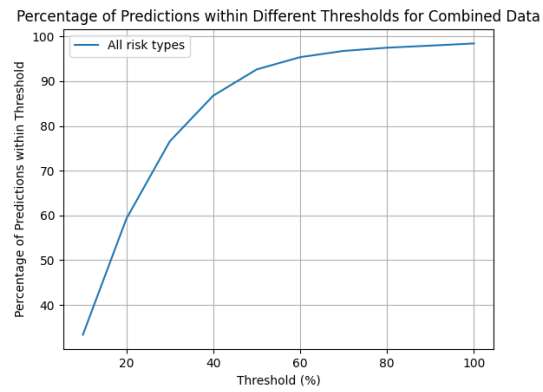
//κενά πεδία

7 Οπτικοποίηση και αξιολόγηση αποτελεσμάτων

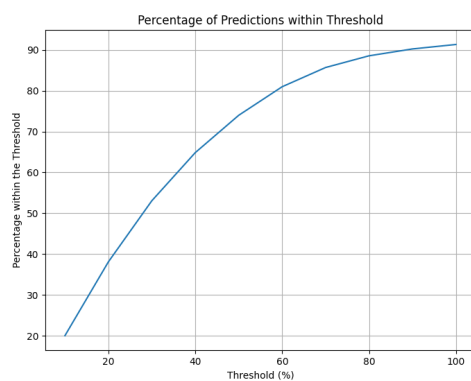
Όπως φαίνεται απο τα παρακάτω διαγράμματα, από τα μοντέλα που χρησιμοποιήσαμε τα καλύτερα αποτελέσματα μας έδωσε το XGBoost.



Εικ. 11: Μοντέλο XGBoost που εκπαιδεύθηκε στο σύνολο των δεδομένων



Εικ. 12: Μοντέλο Random Forest που εκπαιδεύθηκε στο σύνολο των δεδομένων



Εικ. 13: Μοντέλο Νευρωνικού δικτύου που εκπαιδεύθηκε στο σύνολο των δεδομένων