



ΠΑΝΕΠΙΣΤΗΜΙΟ
ΔΥΤΙΚΗΣ ΑΤΤΙΚΗΣ

Τμήμα Μηχανικών
Βιομηχανικής Σχεδίασης & Παραγωγής

Διπλωματική Εργασία

**Τίτλος: Εφαρμογές Μηχανικής Μάθησης στην
Αυτοκινητοβιομηχανία**

του φοιτητή

Παπαδόπουλο Κωνσταντίνο

Αριθμός Μητρώου: 18389263

Επιβλέπων Καθηγητής:

Νικολάου Γρηγόρης

Αθήνα-Αιγάλεω, Ιούνιος 2023

Εξεταστική Επιτροπή:

Α/α	ΟΝΟΜΑ ΕΠΩΝΥΜΟ	ΒΑΘΜΙΑΔΑ/ΙΔΙΟΤΗΤΑ	ΨΗΦΙΑΚΗ ΥΠΟΓΡΑΦΗ
	Νικολάου Γρηγόρης	Λέκτορας	
	Βασιλειάδου Σουλτάνα	Επίκουρη Καθηγήτρια	
	Δρόσος Χρήστος	ΕΔΙΠ	

ΔΗΛΩΣΗ ΣΥΓΓΡΑΦΕΑ ΠΤΥΧΙΑΚΗΣ/ΔΙΠΛΩΜΑΤΙΚΗΣ ΕΡΓΑΣΙΑΣ

Ο κάτωθι υπογεγραμμένος Παπαδόπουλος Κωνσταντίνος του Αχιλλέως, με αριθμό μητρώου 18389263 φοιτητής του Πανεπιστημίου Δυτικής Αττικής της Σχολής Μηχανικών του Τμήματος Βιομηχανικής Σχεδίασης και Παραγωγής, δηλώνω υπεύθυνα ότι:

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

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

Ο Δηλών

ΠΑΠΑΔΟΠΟΥΛΟΣ

ΚΩΝΣΤΑΝΤΙΝΟΣ



Ευχαριστίες

Θα ήθελα να ευχαριστήσω πρώτα από όλους και ξεχωριστά, τον επιβλέποντα καθηγητή μου, κύριο Νικολάου Γρηγόριο, για την υποστήριξη, την καθοδήγηση και την εμπιστοσύνη που μου έχει δείξει, για την αμέριστη αρωγή και στήριξή του σε όλη την ακαδημαϊκή μου πορεία και την καθοδήγησή του σε αυτή την εργασία. Του είμαι βαθιά ευγνώμων για όσα μου έχει προσφέρει.

Επίσης, θα ήθελα να ευχαριστήσω την οικογένεια μου, γιατί ήταν πάντα κοντά μου κι με στήριξε από την αρχή μέχρι και την περάτωση των σπουδών μου.

Τέλος θα ήθελα να ευχαριστήσω τον μέντορα της πρακτικής άσκησης στην αυτοκινητοβιομηχανία «TOYOTA» κύριο Thomas Huylebroeck, του οποίου οι γνώσεις και η καθοδήγηση που μου προσέφερε, αποτέλεσαν σημαντικά εφόδια για τη σύνταξη αυτής της εργασίας.

Πίνακας Περιεχομένων

Ευχαριστίες.....	2
Περίληψη.....	4
Abstract.....	5
Κεφάλαιο 1 : Εισαγωγή στην αυτοκινητοβιομηχανία.....	6
1.1 .1 Ιστορική Αναδρομή.....	6
1.1.2 Οι πρώτες καινοτομίες στο χώρο της αυτοκινητοβιομηχανίας.....	6
1.1.3 Οι καινοτομίες της αυτοκινητοβιομηχανίας τον 21ο αιώνα.....	9
1.1.4 Συμπεράσματα.....	12
Κεφάλαιο 2: Μηχανική Μάθηση και η χρήση της στην σύγχρονη αυτοκινητοβιομηχανία.....	13
2.1 Τι είναι Μηχανική Μάθηση.....	13
2.2 Η χρήση της μηχανικής μάθησης για την επίλυση σύγχρονων προβλημάτων στην αυτοκινητοβιομηχανία.....	14
2.2.1 Μηχανική μάθηση στην ανάπτυξη ηλεκτρικών οχημάτων.....	14
2.3 Χρήση μηχανικής μάθησης για ασφάλεια και προστασία δεδομένων.....	16
2.3.1 Η μηχανική μάθηση στην υπηρεσία της ασφάλειας των δεδομένων των αυτοκινήτων.....	16
2.4 Καλύτερη κατανόηση των καταναλωτικών αναγκών.....	17
2.5 Βελτίωση ενεργειακής απόδοσης και μείωση εκπομπών.....	17
Κεφάλαιο 3 Μελέτη κατανάλωσης καυσίμου στην σύγχρονη αυτοκινητοβιομηχανία.....	20

3.1 Τι είναι ο γύρος μετρήσεων WLTP;	20
3.2 Μελέτη σύγκρισης πραγματικής κατανάλωσης καυσίμου και κύκλου WLTP	21
3.3 Συλλογή Δεδομένων:	21
3.3.1 Απαιτήσεις Δεδομένων	21
3.3.2 Αναζήτηση πηγών	22
3.3.3 Εύρεση και αξιολόγηση πηγών	22
3.3.4 Πηγές εμπλουτισμού της βάσης δεδομένων	23
3.4.1 Επιλογή πηγών	24
3.4.2 Πίνακας αξιολόγησης πηγών	26
3.4.3 Συμπεράσματα	27
3.5 Εξαγωγή δεδομένων	27
3.5.1 Διάγραμμα Ροής των δεδομένων	27
3.5.2 Μεθοδολογία και προγραμματισμός εξαγωγής URL	29
3.5.3 Μεθοδολογία και προγραμματισμός εξαγωγή πληροφοριών οχημάτων	31
3.5.4 A.D.A.C. web scraping	34
3.5.5 Καθαρισμός βάσης δεδομένων ανεφοδιασμού (Table)	38
3.5.6 Εμπλουτισμός βάσης δεδομένων (Spritmonitor)	39
3.5.7 Οπτικοποίηση Δεδομένων	40
3.5.8 Συμπεράσματα	46
Κεφάλαιο 4 Μοντελοποίηση οδηγικής συμπεριφοράς	48
4.1 Συλλογή και μελέτη δεδομένων	48
4.2 Κατασκευή Μοντέλου	56
4.3 Εφαρμογή των δεδομένων στο Νευρωνικό δίκτυο	59
4.4 Αξιολόγηση Μοντέλου	60
4.5 Αξιολόγηση ακριβείας μοντέλου	64

Συμπεράσματα.....	65
Βιβλιογραφία.....	66

Περίληψη

Στην εποχή του 21^{ου} αιώνα, η τεχνολογία που αναπτύσσεται από την αυτοκινητοβιομηχανία φαίνεται να εξελίσσεται με ραγδαίους ρυθμούς. Παρόλα αυτά, σύγχρονες προκλήσεις όπως η ρύπανση του περιβάλλοντος και οι συνεχώς αυξανόμενες απαιτήσεις των καταναλωτών δεν παύουν να δημιουργούν την ανάγκη για μία ευρύτερη τεχνολογική εξέλιξη όσο αφορά την κατασκευή μηχανοκίνητων οχημάτων.

Στο πλαίσιο αυτό, ακολουθώντας την τεχνολογική εξέλιξη και ανταποκρινόμενη στις σύγχρονες ανάγκες στον χώρο της αυτοκινητοβιομηχανίας, λαμβάνει χώρα η παρούσα εργασία. Στην παρούσα εργασία παρουσιάζεται και αναλύεται ο τρόπος με τον οποίο ο κλάδος της μηχανικής μάθησης μπορεί να συνεισφέρει στην επίλυση σύγχρονων προκλήσεων που αντιμετωπίζει η σύγχρονη αυτοκινητοβιομηχανία.

Η ανάλυση αυτή συνοδεύεται με μια μελέτη της ποσοστιαίας απόκλισης της πραγματικής κατανάλωσης καυσίμου σε σύγκριση με τις εργαστηριακές δοκιμές που πραγματοποιούνται από τους κατασκευαστές αυτοκινήτων. Μετά την εξαγωγή συμπερασμάτων της μελέτης αυτής, ακολουθεί η δημιουργία ενός Νευρωνικού Δικτύου που μπορεί να αναγνωρίσει και να ταξινομήσει κινήσεις απότομες και επικίνδυνες των οδηγών, που προκύπτουν από επιθετική οδηγική συμπεριφορά. Ευελπιστούμε ότι η παρούσα μελέτη θα βοηθήσει την αυτοκινητοβιομηχανία να κατανοήσει βαθύτερα την σημασία παραμέτρων που επηρεάζουν την κατανάλωση καυσίμου.

Λέξεις κλειδιά: Αυτοκινητοβιομηχανία, Νευρωνικά Δίκτυα, Ταξινόμηση, Κατανάλωση καυσίμου, Ποσοστιαία απόκλιση, επιστήμη των δεδομένων.

Abstract

In the 21st century, the technology developed by the automotive industry seems to be evolving rapidly. However, modern challenges, such as the environmental crisis and the continuous consumers' needs, do not stop creating the need for further technological development, when it comes to motor vehicle manufacturing.

This paper, includes a in depth analysis on how Machine Learning can contribute to the solving of modern challenges faced by the automotive industry.

This analysis is accompanied with an in depth research of the percentage deviation gap between real-time fuel consumption and the lab prediction made by car manufacturers. After extracting results from the research, a Neural Network that will be able to recognize and predict driving maneuvers that come from aggressive driving behavior.

Key words: Automotive industry, Neural networks, Classification, Fuel consumption, Percentage deviation, Data science

Κεφάλαιο 1: Εισαγωγή στην αυτοκινητοβιομηχανία

1.1 Ιστορική αναδρομή

Η αυτοκινητοβιομηχανία αποτελεί έναν από τους σημαντικότερους οικονομικούς τομείς στην παγκόσμια αγορά. Ο συγκεκριμένος τομέας ασχολείται με τον σχεδιασμό, την κατασκευή, την ανάπτυξη και την εμπορία μηχανοκίνητων οχημάτων και ανταλλακτικών. Ως κλάδος αποτελεί έναν από τους πιο κερδοφόρους παγκοσμίως, με ετήσιο εισόδημα 2,86 τρισεκατομμυρίων δολαρίων το 2021, σημειώνοντας αριθμό πωλήσεων μεγαλύτερο από 66,2 εκατομμύρια τεμάχια (*Economic and Market Report State of the EU auto industry Full-year 2022*).

Η εν λόγω βιομηχανία έκανε τα πρώτα της βήματα το 1885 με την εφεύρεση του πρώτου κινητήρα εσωτερικής καύσης από τον Karl Benz. Τούτη η καινοτομία παρείχε έναν αξιόπιστο και αποτελεσματικό μέσο μεταφοράς, που σηματοδότησε την αρχή μιας νέας εποχής που θα άλλαζε τον ρου της ιστορίας του ανθρώπου : της εποχής του αυτοκινήτου. Η τεχνολογική αυτή πρόοδος προκάλεσε μια σειρά καινοτομιών, οδηγώντας σε οχήματα ταχύτερα, πιο άνετα, πιο ασφαλή και ευκολότερα στη χρήση, συγκριτικά με τους έως τότε συνηθέστερους τρόπους μετακίνησης (*The Complete Encyclopedia of Motorcars 1885–1968. New York: E. P. Dutton*).

Το έτος 1913 ο Henry Ford καινοτόμησε στην αγορά, δημιουργώντας την πρώτη γραμμή μαζικής παραγωγής αυτοκινήτων, μειώνοντας δραστικά τόσο το κόστος όσο και το χρόνο παραγωγής του τότε μοντέλου Ford Model T από 12 ώρες το τεμάχιο στις 2 ώρες. Τούτη η καινοτομία οδήγησε σε ραγδαία αύξηση των πωλήσεων –με πάνω από 15.000 παραγγελίες στην πρώτη σειρά. Η μείωση του κόστους παραγωγής οδήγησε στη μείωση του κόστους πώλησης, με αποτέλεσμα όλο και μεγαλύτερο πλήθος καταναλωτών να έχει την οικονομική δυνατότητα να αποκτήσει ένα προσωπικό αυτοκίνητο, μετατρέποντας το αυτοκίνητο από μέσο πολυτελείας, σε εργαλείο πρώτης ανάγκης και εντάσσοντάς το στην καθημερινότητα των ανθρώπων. (Ford Model T: The Car That Put the World on Wheels Lindsay Brooke).

1.1.2 Οι πρώτες καινοτομίες στο χώρο της αυτοκινητοβιομηχανίας Ύστερα από την εκβιομηχάνιση της αγοράς αυτοκινήτων, ακολούθησε μια συνεχής περίοδος ακμής στην βιομηχανία. Όλο και περισσότεροι καινούριοι κατασκευαστικές εταιρείες ιδρύονταν, δημιουργώντας έτσι ένα σημαντικό ανταγωνιστικό κλίμα στην συγκεκριμένη αγορά. Το φαινόμενο του αυξημένου ανταγωνισμού οδήγησε στην ολοένα αυξανόμενη ανάγκη για τεχνολογική εξέλιξη. Χάρη στο γεγονός αυτό, καθ' όλη τη διάρκεια του 20^{ου} αιώνα παρατηρούνταν αμέτρητες καινοτομίες όπου κάθε μια από αυτές συνεισέφερε στη διαμόρφωση του σύγχρονου αυτοκινήτου και της σύγχρονης αυτοκινητοβιομηχανίας. Μερικές από τις σημαντικότερες καινοτομίες που συνέβαλαν στην εξέλιξη των αυτοκινήτων είναι :

Ο εξοπλισμός των αυτοκινήτων με υδραυλικά φρένα όπως στο μοντέλο «Duesenberg Model A το 1921.» Η εν λόγω καινοτομία χρησιμοποιούσε σωλήνες που περιείχαν μια μίξη από είναι ένα μείγμα γλυκολών και άλλων προσθέτων που παρέχουν τις απαραίτητες ιδιότητες - για ένα αποτελεσματικό φρενάρισμα-, όπως υψηλό σημείο βρασμού, χαμηλή συμπίεστικότητα και αντοχή στη διάβρωση. Το σύστημα φρένων χρησιμοποιούσε αυτά τα υγρά ώστε να μεταφέρει την ανάλογη δύναμη από το πάτημα του φρένου στις 4 ρόδες του αυτοκινήτου. Η τεχνολογία αυτή οδήγησε στην ασφαλέστερη και ευκολότερη χρήση των φρένων, όπου ακόμα και στη σημερινή εποχή συνεχίζει να χρησιμοποιείται στην κατασκευή των σύγχρονων αυτοκινήτων.

Η ένταξη της ανεξάρτητης ανάρτησης στις μπροστινές ρόδες. Η εταιρεία “General Motors” αποτέλεσε τον πρώτο κατασκευαστή ο οποίος ενέταξε την τεχνολογία τούτη στα αυτοκίνητά του δίνοντας την δυνατότητα στις δύο μπροστινές ρόδες να περιστρέφονται η μία ανεξάρτητα από την άλλη, μειώνοντας έτσι το αντίκτυπο της μίας ρόδας στην άλλη όταν το αυτοκίνητο διέσχισε κακοτράχαλους δρόμους, αυξάνοντας έτσι την ασφάλεια του αυτοκινήτου.

Η δημιουργία του πρώτου πλήρως αυτόματου κιβωτίου ταχυτήτων. Κατά τη διάρκεια του Β' Παγκοσμίου Πολέμου (1940) μηχανικοί της “General Motors” όπως ο Earl Thompson και William H.Grauer κατασκεύασαν εξόπλισαν τα μοντέλα «Oldsmobile Series 60» και «Oldsmobile Series 70» με τα πρώτα αυτόματα κιβώτια ταχυτήτων. Η τεχνολογία αυτή έφερε μια ακόμα επανάσταση στην οδηγική εμπειρία εξαλείφοντας την ανάγκη για χειροκίνητη αλλαγή, καθιστώντας την οδήγηση πιο βολική και προσβάσιμη σε ένα ευρύτερο φάσμα οδηγών. Επρόκειτο για ένα κιβώτιο, που άλλαζε ταχύτητες αυτόματα με βάση την ταχύτητα και τις συνθήκες φορτίου του οχήματος. Χρησιμοποιούσε μια σύνδεση υγρού, γνωστή και ως μετατροπέας ροπής, για τη μεταφορά ισχύος από τον κινητήρα στο κιβώτιο ταχυτήτων, επιτρέποντας ομαλές και απρόσκοπτες αλλαγές ταχυτήτων χωρίς την ανάγκη για πεντάλ συμπλέκτη. Η εισαγωγή μιας τέτοιας τεχνολογίας σηματοδότησε ένα σημαντικό ορόσημο στην ιστορία της αυτοκινητοβιομηχανίας και άνοιξε το δρόμο για την ευρεία υιοθέτηση των αυτόματων κιβωτίων ταχυτήτων τα επόμενα χρόνια. Σήμερα, τα αυτόματα κιβώτια αποτελούν κανόνα στα περισσότερα επιβατικά οχήματα, προσφέροντας ευκολία στη χρήση και βελτιωμένη οδηγική άνεση.

Η καθιέρωση της ζώνης ασφαλείας ως αναγκαίο εξάρτημα κάθε αυτοκινήτου. Παρόλου που οι πρώτες ζώνες ασφαλείας είχαν ήδη εμφανιστεί από τις αρχές 20^{ου} αιώνα, η επίσημη αναγνώρισή τους ως κρίσιμο χαρακτηριστικό ασφαλείας δεν είχε καθιερωθεί μέχρι και τη δεκαετία του 1950. Η σουηδική αυτοκινητοβιομηχανία Volvo ήταν αυτή που έφερε σημαντικές εξελίξεις στην ασφάλεια των ζωνών ασφαλείας. Ο μηχανικός της Volvo, Nils Bohlin, εφηύρε τη ζώνη ασφαλείας τριών σημείων το 1958. Αυτός ο σχεδιασμός ήταν πολύ πιο αποτελεσματικός από τον συμβατικό ιμάντα δύο σημείων, καθώς προστατεύει τόσο το άνω όσο και το κάτω μέρος του σώματος, ενώ κατανέμει ομοιόμορφα την πρόσκρουση σε όλο το σώμα κατά τη διάρκεια μιας σύγκρουσης, ελαχιστοποιώντας έτσι τον κίνδυνο τραυματισμού. Το 1959, η Volvo ενσωμάτωσε τη ζώνη ασφαλείας τριών σημείων στο μοντέλο Volvo PV 544. Αναγνωρίζοντας τον κρίσιμο ρόλο της εφεύρεσης του Bohlin στην ενίσχυση της δημόσιας ασφάλειας, η Volvo μοιράστηκε ελεύθερα το δίπλωμα ευρεσιτεχνίας για τη ζώνη ασφαλείας τριών σημείων με άλλους κατασκευαστές αυτοκινήτων. Κατά συνέπεια, η ζώνη ασφαλείας τριών σημείων έγινε σταδιακά ένα παγκοσμίως αποδεκτό χαρακτηριστικό ασφαλείας στα αυτοκίνητα. Παρόλα αυτά, η εφαρμογή των νόμων για τις ζώνες ασφαλείας στις Ηνωμένες Πολιτείες ξεκίνησε τη δεκαετία του 1960 με το Ουισκόνσιν να είναι η πρώτη πολιτεία που επέβαλε τη χρήση ζώνης ασφαλείας το 1961, δεν ήταν μέχρι το 1984 που η Νέα Υόρκη έγινε η πρώτη πολιτεία που απαιτούσε από όλους τους επιβάτες αυτοκινήτων να δεθούν. Αν και πολλοί οδηγοί αρχικά αντιστάθηκαν στη χρήση ζωνών ασφαλείας, οι δυνατότητες διάσωσης και ο ρόλος τους στην πρόληψη τραυματισμών έχουν γίνει ευρέως αποδεκτοί με την πάροδο του χρόνου. Η χρήση ζώνης ασφαλείας αποτελεί πλέον νομική απαίτηση σε πολλές χώρες σε όλο τον κόσμο.

- Η πρώτη εμφάνιση καινοτομιών είχε σκοπό την μείωση ρύπων και την βελτιστοποίηση της απόδοσης του κινητήρα. Πιο συγκεκριμένα, κατά τη διάρκεια της δεκαετίας του 1960, η αυτοκινητοβιομηχανία γνώρισε σημαντικές εξελίξεις στα συστήματα ελέγχου εκπομπών. Η εποχή αυτή σηματοδότησε την αρχή σοβαρών προσπαθειών για την αντιμετώπιση των περιβαλλοντικών επιπτώσεων των εκπομπών των αυτοκινήτων. Οι κυβερνήσεις σε όλο τον κόσμο άρχισαν να εφαρμόζουν κανονισμούς για τον έλεγχο και τη μείωση των εκπομπών, οδηγώντας στην εισαγωγή διαφόρων τεχνολογιών ελέγχου των εκπομπών. Για τη μείωση των εκπομπών υδρογονανθράκων εισήχθησαν συστήματα εξαερισμού και συστήματα ελέγχου εκπομπών που προέρχονται από τις εξατμίσεις. Επιπλέον, οι εξελίξεις στην τεχνολογία του καρμπυρατέρ και η εισαγωγή πρώιμων καταλυτικών μετατροπέων βοήθησαν στη βελτιστοποίηση των αναλογιών μείγματος καυσίμου-αέρα και στη μείωση των εκπομπών ρύπων. Τα γεγονότα αυτά έθεσαν τα θεμέλια για μελλοντικές τεχνολογίες ελέγχου των εκπομπών CO₂ και για αυστηρότερα πρότυπα εκπομπών.
- Η εισαγωγή του Anti-lock Braking System (ABS). Το 1981, η Mercedes-Benz αποτέλεσε την πρώτη αυτοκινητοβιομηχανία που προσέφερε ένα σύστημα ζώνης ασφαλείας αερόσακου ως βασικό εξοπλισμό στο μοντέλο Mercedes-Benz S-Class. Ταυτόχρονα έθεσε σε εφαρμογή την παραγωγή αυτοκινήτων με λειτουργία ABS. Η ABS είναι μια λειτουργία ασφαλείας που εμποδίζει τους τροχούς να κλειδώνουν ή να γλιστρούν κατά τη διάρκεια βαριάς πέδησης, ειδικά σε ολισθηρές επιφάνειες.

- Η ανάπτυξη συστημάτων On Board Diagnostics (OBD). Τα συστήματα OBD έχουν σχεδιαστεί για να ανιχνεύουν και να αναφέρουν δυσλειτουργίες του κινητήρα, του κιβωτίου ταχυτήτων και των συστημάτων εκπομπών. Χρησιμοποιούν αισθητήρες σε όλο το όχημα για την παρακολούθηση παραμέτρων όπως η ταχύτητα του κινητήρα, η εισαγωγή αέρα, το μείγμα καυσίμου και οι εκπομπές καυσαερίων. Όταν εντοπίζεται σφάλμα, το σύστημα OBD δημιουργεί έναν κωδικό διαγνωστικού προβλήματος (DTC) που βοηθά τους τεχνικούς να εντοπίσουν και να διορθώσουν το πρόβλημα. Τα προηγούμενα συστήματα OBD-I ήταν ειδικά για τον κατασκευαστή και χρησιμοποιούσαν ιδιόκτητα διαγνωστικά πρωτόκολλα. Ωστόσο, η εισαγωγή του OBD-II στα μέσα της δεκαετίας του 1990 τυποποίησε τα διαγνωστικά συστήματα σε διαφορετικές μάρκες οχημάτων. Το OBD-II ανέθεσε έναν τυπικό διαγνωστικό σύνδεσμο και ένα σύνολο γενικών διαγνωστικών κωδικών βλάβης, διευκολύνοντας τους τεχνικούς να διαγνώσουν και να επισκευάσουν οχήματα από διάφορους κατασκευαστές.
- Ο 20ος αιώνας έκλεισε με την εδραίωση χρήσης συστημάτων GPS. Το Global Positioning System (GPS) είναι ένα δορυφορικό σύστημα πλοήγησης που παρέχει ακριβείς πληροφορίες τοποθεσίας και χρόνου οπουδήποτε στη Γη. Ενώ η τεχνολογία GPS αναπτύχθηκε από το Υπουργείο Άμυνας των ΗΠΑ τη δεκαετία του 1970, έγινε πιο προσιτή και χρησιμοποιήθηκε ευρέως τη δεκαετία του 1990. Στη δεκαετία του 1990, η τεχνολογία GPS άρχισε να εμφανίζεται σε εμπορικές εφαρμογές, συμπεριλαμβανομένων των συστημάτων πλοήγησης αυτοκινήτων. Τα συστήματα πλοήγησης GPS στα αυτοκίνητα βασίζονταν σε σήματα από πολλαπλούς δορυφόρους για να καθορίσουν την ακριβή θέση του οχήματος και να παρέχουν κατευθύνσεις στον οδηγό. Στην αρχή, αυτά τα συστήματα ήταν ακριβά και ήταν κυρίως διαθέσιμα σε οχήματα υψηλής τεχνολογίας, αλλά η δημοτικότητά τους αυξήθηκε καθώς η τεχνολογία προχωρούσε και το κόστος μειώθηκε. Η ενσωμάτωση της τεχνολογίας GPS στα αυτοκίνητα έφερε επανάσταση στην πλοήγηση, επιτρέποντας στους οδηγούς να βρίσκουν τους προορισμούς τους πιο εύκολα και με ακρίβεια. Επίσης, άνοιξε το δρόμο για άλλες υπηρεσίες που βασίζονται στην τοποθεσία, όπως η διαχείριση του στόλου, η παρακολούθηση κλεμμένων οχημάτων και η οδική βοήθεια έκτακτης ανάγκης.

1.1.3 Οι καινοτομίες της αυτοκινητοβιομηχανίας τον 21^ο αιώνα.

Με την είσοδο στον 21^ο αιώνα οι ανάγκες της αγοράς – ιδίως της ευρωπαϊκής – φαίνονται να πιο απαιτητικές από ποτέ. Η επίδραση της κλιματικής αλλαγής, ώθησε την παγκόσμια αυτοκινητοβιομηχανία να παράγει τα λεγόμενα «πράσινα αυτοκίνητα», δηλαδή αυτοκίνητα φιλικά προς το περιβάλλον. Σκοπός της κίνησης αυτής ήταν να μπορέσει να τηρήσει τις συνεχώς απαιτητικότερες νομοθεσίες, οι οποίες αποσκοπούν στην μείωση αποτυπώματος διοξειδίου του άνθρακα και την προστασία του περιβάλλοντος από τους ρύπους των αυτοκινήτων. Πρωτοπόρος στην κατασκευή τέτοιων κινητήρων ήταν η εταιρεία Toyota.

Ο αιώνας ξεκίνησε με την εισαγωγή του πρώτου μαζικής παραγωγής υβριδικού αυτοκινήτου από την Toyota το «Toyota Prius». Πρόκειται για ένα πρωτοποριακό υβριδικό όχημα μαζικής παραγωγής που έπαιξε καθοριστικό ρόλο στη διάδοση της υβριδικής τεχνολογίας. Συνδύασε έναν βενζινοκινητήρα εσωτερικής καύσης με ηλεκτρικό κινητήρα και μπαταρία, επιτρέποντας βελτιωμένη απόδοση καυσίμου και μειωμένες εκπομπές σε σύγκριση με τα παραδοσιακά βενζινοκίνητα αυτοκίνητα. Η επιτυχία του Prius, με πάνω από 500.000 πωλήσεις μέχρι το 2005, άνοιξε το δρόμο για την ευρύτερη αποδοχή και υιοθέτηση υβριδικών οχημάτων, ενθαρρύνοντας άλλες αυτοκινητοβιομηχανίες να αναπτύξουν τα δικά τους υβριδικά μοντέλα και συμβάλλοντας στη συνολική στροφή προς πιο πράσινες επιλογές μεταφοράς.

- Μέσα στην ίδια δεκαετία εδραιώθηκε στην αγορά η τεχνολογία της άμεσης έγχυσης (Direct Injection). Η συγκεκριμένη τεχνολογία, έφερε επανάσταση, ενισχύοντας σημαντικά την αποδοτικότητα καυσίμου, την ισχύ και τον έλεγχο των εκπομπών σε βενζινοκινητήρες. Η άμεση έγχυση είναι ένα σύστημα παροχής καυσίμου που ψεκάζει καύσιμο απευθείας στον θάλαμο καύσης σε υψηλή πίεση, επιτρέποντας πιο ακριβή έλεγχο του μείγματος αέρα-καυσίμου. Αυτή η τεχνολογία αντικατέστησε το παραδοσιακό σύστημα έγχυσης καυσίμου θύρας, το οποίο εισήγαγε καύσιμο στην πολλαπλή εισαγωγής πριν φτάσει στον θάλαμο καύσης. Ένα από τα βασικά πλεονεκτήματα του άμεσου ψεκασμού είναι η ικανότητά του να βελτιώνει την απόδοση καυσίμου. Με τον ακριβή έλεγχο της ποσότητας και του χρονισμού της έγχυσης καυσίμου, τα συστήματα άμεσου ψεκασμού μπορούν να επιτύχουν καλύτερη απόδοση καύσης, με αποτέλεσμα τη μειωμένη κατανάλωση καυσίμου. Αυτή η τεχνολογία επιτρέπει επίσης υψηλότερους λόγους συμπίεσης, γεγονός που ενισχύει περαιτέρω την απόδοση και την ισχύ εξόδου. Τα συστήματα άμεσης έγχυσης συμβάλλουν επίσης στη βελτίωση του ελέγχου των εκπομπών. Με την έγχυση καυσίμου απευθείας στον θάλαμο καύσης, το μείγμα ψεκάζεται και κατανέμεται καλύτερα, οδηγώντας σε πληρέστερη καύση. Αυτό έχει ως αποτέλεσμα χαμηλότερες εκπομπές ρύπων όπως οξείδια του αζώτου (NOx) και σωματίδια. Επιπλέον, η άμεση έγχυση επιτρέπει τη μείωση του μεγέθους του κινητήρα χωρίς να θυσιάζεται η απόδοση. Η υψηλότερη απόδοση και η ισχύς των κινητήρων άμεσης έγχυσης επιτρέπουν στους κατασκευαστές να παράγουν μικρότερους, ελαφρύτερους κινητήρες που παρέχουν τα ίδια ή και υψηλότερα επίπεδα απόδοσης σε σύγκριση με τους μεγαλύτερους κινητήρες με συμβατική έγχυση καυσίμου.

Η εισαγωγή της άμεσης έγχυσης επέφερε σημαντικές εξελίξεις στην αποδοτικότητα καυσίμου, την ισχύ και τον έλεγχο των εκπομπών σε βενζινοκινητήρες. Η τεχνολογία αυτή έπαιξε καθοριστικό ρόλο στις προσπάθειες της αυτοκινητοβιομηχανίας να ανταποκριθεί σε όλο και πιο αυστηρούς περιβαλλοντικούς κανονισμούς, παρέχοντας παράλληλα βελτιωμένες επιδόσεις στους καταναλωτές.

Το 2003 η Honda ενσωμάτωσε το πρώτο σύστημα προειδοποίησης αναχώρησης λωρίδας (Lane Departure Warning) στο «Honda Inspire». Η στα οχήματα βελτίωσε σημαντικά την οδική ασφάλεια μειώνοντας τον κίνδυνο παρασυρόμενων λωρίδων και πιθανών συγκρούσεων. Η LDW χρησιμοποιεί τεχνολογία κάμερας ή αισθητήρα για την παρακολούθηση της θέσης του οχήματος εντός της λωρίδας κυκλοφορίας και ειδοποιεί τον οδηγό εάν παρεκκλίνει ακούσια από τη λωρίδα κυκλοφορίας του χωρίς σηματοδότηση. Αυτό το προηγμένο σύστημα βοήθειας οδηγού έχει αποδειχθεί αποτελεσματικό στην πρόληψη ατυχημάτων που προκαλούνται από απροσεξία, υπνηλία ή απόσπαση της προσοχής του οδηγού. Μελέτες έχουν δείξει ότι τα συστήματα LDW συμβάλλουν στη μείωση των ατυχημάτων που σχετίζονται με την αναχώρηση από τη λωρίδα κυκλοφορίας, καθιστώντας τα ένα ανεκτίμητο εργαλείο για την προώθηση ασφαλέστερων οδικών συνηθειών και την ελαχιστοποίηση των τροχαίων ατυχημάτων.

Παρά το γεγονός ότι η ιδέα ενός πλήρως ηλεκτρικού αυτοκίνητου υπάρχει ήδη από τα μέσα του 20^{ου} αιώνα και ταυτόχρονα πολλές εταιρίες από το 2000 και μετά είχαν ξεκινήσει να αναπτύσσουν και να εξάγουν ηλεκτρικά αυτοκίνητα στην αγορά, η μεγαλύτερη επανάσταση στην ηλεκτροκίνηση ήρθε από την Tesla Motors στις αρχές της δεκαετίας του 2010.

Μέσω ενός συνδυασμού πρωτοποριακής τεχνολογίας, αμείλικτης καινοτομίας και στρατηγικού μάρκετινγκ, η Tesla αναδιαμόρφωσε την αντίληψη και την υιοθέτηση των ηλεκτρικών οχημάτων, καθιστώντας τα μια βιώσιμη και επιθυμητή εναλλακτική λύση στα παραδοσιακά βενζινοκίνητα αυτοκίνητα. Ένας από τους βασικούς τρόπους με τους οποίους η Tesla άλλαξε τη βιομηχανία είναι πιέζοντας τα όρια της σειράς EV. Τα πρώτα μοντέλα της Tesla, όπως το Roadster και το Model S, καυχήθηκαν για εντυπωσιακές δυνατότητες εμβέλειας, ανακουφίζοντας το άγχος της εμβέλειας που ήταν από καιρό ανησυχία για τους πιθανούς αγοραστές EV. Με την προβολή των δυνατοτήτων για ταξίδια μεγάλων αποστάσεων, η Tesla απέδειξε ότι τα ηλεκτρικά οχήματα θα μπορούσαν να είναι πρακτικά για καθημερινή χρήση και θα μπορούσαν να απευθύνονται σε μια ευρύτερη αγορά.

Επιπλέον, η δέσμευση της Tesla για την οικοδόμηση ενός ισχυρού δικτύου υπερσυμπιεστών έχει αντιμετωπίσει μια άλλη σημαντική ανησυχία για τους ιδιοκτήτες EV: την υποδομή φόρτισης. Η ευρεία διαθεσιμότητα των υπερσυμπιεστών επιτρέπει στους οδηγούς της Tesla να ξεκινήσουν μακρινά ταξίδια με ηρεμία, αντιμετωπίζοντας αποτελεσματικά τους περιορισμούς της εμβέλειας των ηλεκτρικών οχημάτων και συμβάλλοντας περαιτέρω στην αποδοχή τους.

Η καινοτόμος προσέγγιση της Tesla στον σχεδιασμό και την τεχνολογία των οχημάτων έχει επίσης θέσει νέα βιομηχανικά πρότυπα. Η εστίαση της εταιρείας στην απόδοση, τα χαρακτηριστικά αιχμής και τις δυνατότητες αυτόνομης οδήγησης ώθησε άλλες αυτοκινητοβιομηχανίες να επενδύσουν σε μεγάλο βαθμό στην ανάπτυξη ηλεκτρικών οχημάτων και να ανταγωνιστούν όσον αφορά την εμβέλεια, την επιτάχυνση και τη συνολική εμπειρία των χρηστών. Ο αντίκτυπος της Tesla στην αυτοκινητοβιομηχανία δεν μπορεί να υπερεκτιμηθεί. Η ακλόνητη δέσμευση της εταιρείας για την προώθηση της τεχνολογίας EV, την επέκταση της εμβέλειας, τη δημιουργία υποδομών φόρτισης και την πρόκληση παραδοσιακών μοντέλων πωλήσεων επιτάχυνε την υιοθέτηση ηλεκτρικών οχημάτων και αναμόρφωσε το μέλλον των μεταφορών.

Δεν θα έπρεπε βέβαια να παραληφθεί μια από τις σημαντικότερες σύγχρονες τεχνολογικές επαναστάσεις στην σύγχρονη αυτοκινητοβιομηχανία αφορά την δημιουργία των πρώτων αυτοκινούμενων οχημάτων.

Η έλευση της τεχνολογίας self-driving έχει σημειώσει σημαντικές ανακαλύψεις, λόγω των ταχέων εξελίξεων στους αλγόριθμους μηχανικής μάθησης και στις δυνατότητες επεξεργασίας δεδομένων. Τα αυτόνομα οχήματα αξιοποιούν έναν συνδυασμό τεχνολογιών αισθητήρων, συμπεριλαμβανομένων ραντάρ, L.I.D.A.R. και φωτογραφικών μηχανών, για τη συλλογή δεδομένων σε πραγματικό χρόνο για το περιβάλλον.

Τα αυτοκινούμενα οχήματα έχουν αναδειχθεί ως μετασχηματιστική δύναμη στην αυτοκινητοβιομηχανία. Στην καρδιά αυτής της Επανάστασης βρίσκεται η μηχανική μάθηση, ένας κλάδος της τεχνητής νοημοσύνης που επιτρέπει στα αυτόνομα αυτοκίνητα να αντιλαμβάνονται το περιβάλλον τους, να λαμβάνουν τεκμηριωμένες αποφάσεις και να περιηγούνται σε πολύπλοκα περιβάλλοντα. Ένα από τα βασικά πλεονεκτήματα της μηχανικής μάθησης στην τεχνολογία αυτόνομης οδήγησης είναι η ικανότητά της να προσαρμόζεται και να βελτιώνεται με την πάροδο του χρόνου. Η συνεχής μάθηση και η βελτιστοποίηση βάσει δεδομένων επιτρέπουν στα αυτοκινούμενα οχήματα να βελτιώσουν την απόδοσή τους και την ανταπόκρισή τους σε διάφορα σενάρια. Αυτή η επαναληπτική διαδικασία μάθησης εξασφαλίζει ότι οι αλγόριθμοι εξελίσσονται για να χειριστούν ένα ευρύ φάσμα συνθηκών οδήγησης, συμπεριλαμβανομένων των δύσκολων καιρικών συνθηκών, των ζωνών κατασκευής και των απροσδόκητων εμποδίων, με αποτέλεσμα η ιδέα ενός κόσμου που μελλοντικά ο άνθρωπος θα πραγματοποιεί τις καθημερινές του μετακινήσεις, χωρίς την ανάγκη για οδήγηση, να καθίσταται όλο και πιο ρεαλιστική.

1.1.4 Συμπέρασμα

Από την ιστορική ανασκόπηση της ιστορίας της αυτοκινητοβιομηχανίας τον 21^ο αιώνα προκύπτει ότι κάθε δεκαετία φέρνει και μια σειρά τεχνολογικών επαναστάσεων. Στην καρδιά των επαναστάσεων που συμβαίνουν τα τελευταία δέκα χρόνια, βρίσκεται η μηχανική μάθηση, ένας κλάδος της τεχνητής νοημοσύνης που επιτρέπει την δημιουργία μοντέλων τα οποία έχουν τη δυνατότητα να πραγματοποιούν σύνθετες πράξεις και συσχετίσεις δεδομένων, δίνοντας αξιόπιστες και τεκμηριωμένες λύσεις σε σύγχρονα προβλήματα που αντιμετωπίζει η αυτοκινητοβιομηχανία. Στο παρακάτω κεφάλαιο θα παρουσιαστούν τα πλεονεκτήματα της μηχανικής μάθησης και η εφαρμογή της στον κλάδο τούτο.

Κεφάλαιο 2: Μηχανική Μάθηση και η χρήση της στην σύγχρονη αυτοκινητοβιομηχανία

Στο τέλος του προηγούμενου κεφαλαίου αναφέρθηκε πως η μηχανική μάθηση έχει συνεισφέρει σημαντικά στην ανάπτυξη τεχνολογίας αυτοκινούμενων οχημάτων. Αξίζει όμως να σημειωθεί πως η μηχανική μάθηση δεν είναι αποκλειστικά χρήσιμη μόνο στο συγκεκριμένο κομμάτι της αυτοκινητοβιομηχανίας. Είναι γεγονός ότι η χρήση της μοντέλων μηχανικής μάθησης έχει ήδη προσφέρει πολλές λύσεις σε σύγχρονες προκλήσεις που αντιμετωπίζουν τα τμήματα έρευνας και ανάπτυξης των κατασκευαστών αυτοκινήτων

2.1 Τι είναι Μηχανική Μάθηση

Ορισμός:

Η μηχανική μάθηση αποτελεί μία υποκατηγορία της τεχνητής νοημοσύνης (AI) που επικεντρώνεται στην ανάπτυξη αλγορίθμων και μοντέλων ικανών να μαθαίνουν και να λαμβάνουν προβλέψεις ή αποφάσεις χωρίς ρητό προγραμματισμό. Χρησιμοποιεί στατιστικές τεχνικές για να επιτρέψει στα συστήματα υπολογιστών να αναλύουν και να ερμηνεύουν αυτόματα πολύπλοκα πρότυπα δεδομένων και στη συνέχεια να μαθαίνουν από αυτά για να βελτιώσουν την απόδοσή τους με την πάροδο του χρόνου, με βασικό σκοπό να μιμηθεί τον τρόπο με τον οποίο ο άνθρωπος μαθαίνει, εξάγει συμπεράσματα και πραγματοποιεί προβλέψεις.

Η θεμελιώδης αρχή της μηχανικής μάθησης περιλαμβάνει αλγόριθμους εκπαίδευσης σε ένα όγκο δεδομένων, γνωστό ως δεδομένα εκπαίδευσης, για να τους επιτρέψει να αναγνωρίσουν και να γενικεύσουν τα πρότυπα. Αυτά τα πρότυπα στη συνέχεια αξιοποιούνται για να κάνουν προβλέψεις ή αποφάσεις όταν αντιμετωπίζουν νέα, αόρατα δεδομένα.

Η μηχανική μάθηση περιλαμβάνει διάφορες τεχνικές και αλγόριθμους, όπως η επιβλεπόμενη μάθηση (supervised learning), η μη επιβλεπόμενη μάθηση (unsupervised learning), η μάθηση με ενίσχυση (reinforcement learning) και η ημι-επιβλεπόμενη μάθηση (semi-supervised learning)

Η επιβλεπόμενη μάθηση περιλαμβάνει αλγόριθμους κατάρτισης που χρησιμοποιούν επισημασμένα δεδομένα (labeled data) για να μάθουν πρότυπα και να κάνουν προβλέψεις.

Η μη επιβλεπόμενη μάθηση, από την άλλη πλευρά, ασχολείται με μη επισημασμένα δεδομένα (unlabeled data) για να ανακαλύψει εγγενή πρότυπα ή ομαδοποιήσεις.

Η ημι-επιβλεπόμενη μάθηση βρίσκεται μεταξύ της επιβλεπόμενης μάθησης και της μη επιβλεπόμενης μάθησης. Αντιμετωπίζει την πρόκληση της μάθησης από μερικούς επισημασμένα δεδομένα, όπου μόνο ένα μικρό μέρος του συνόλου δεδομένων είναι επισημασμένο, ενώ η πλειοψηφία παραμένει χωρίς ετικέτα.

Η μάθηση με ενίσχυση επικεντρώνεται σε αλγόριθμους εκπαίδευσης μέσω ενός συστήματος ανταμοιβών και τιμωριών για την ενθάρρυνση της βέλτιστης συμπεριφοράς

Η μηχανική μάθηση έχει φέρει επανάσταση σε πολλούς τομείς, συμπεριλαμβανομένης του computer vision, της επεξεργασίας φυσικής γλώσσας, της αναγνώρισης ομιλίας, των συστημάτων συστάσεων και της ανάλυσης χρηματοπιστωτικών αγορών.

Ένα από τα βασικά πλεονεκτήματα της μηχανικής μάθησης στην τεχνολογία αυτο-οδήγησης είναι η ικανότητά της να προσαρμόζεται και να βελτιώνεται με την πάροδο του χρόνου. Η συνεχής μάθηση και η βελτιστοποίηση βάσει δεδομένων επιτρέπουν στα αυτο-οδηγούμενα οχήματα να βελτιώσουν την απόδοσή τους και την ανταπόκρισή τους σε διάφορα σενάρια. Αυτή η επαναληπτική διαδικασία μάθησης εξασφαλίζει ότι οι αλγόριθμοι εξελίσσονται για να χειριστούν ένα ευρύ φάσμα συνθηκών οδήγησης, συμπεριλαμβανομένων των δύσκολων καιρικών συνθηκών, των ζωνών κατασκευής και των απροσδόκητων εμποδίων.

Η μηχανική μάθηση έχει αναδειχθεί ως βασικός μοχλός στις τεχνικές καινοτομίες των αυτοκινούμενων οχημάτων. Αξιοποιώντας τη δύναμη των αλγορίθμων μηχανικής μάθησης, τα αυτόνομα αυτοκίνητα μπορούν να αντιληφθούν το περιβάλλον τους, να λάβουν τεκμηριωμένες αποφάσεις και να πλοηγηθούν με αυξανόμενα επίπεδα αυτονομίας. Αν και εξακολουθούν να υπάρχουν προκλήσεις, η ενσωμάτωση της μηχανικής μάθησης στην τεχνολογία αυτο-οδήγησης υπόσχεται ασφαλέστερους δρόμους, βελτιωμένη αποτελεσματικότητα των μεταφορών και μια νέα εποχή κινητικότητας. Παρακάτω αναλύεται πως η μηχανική μάθηση μπορεί να συντελέσει στην επίλυση των προβλημάτων που αντιμετωπίζει η σύγχρονη βιομηχανία αυτοκινήτων.

2.2 Η χρήση της μηχανικής μάθησης για την επίλυση σύγχρονων προβλημάτων στην αυτοκινητοβιομηχανία

Όσο αφορά τη βιομηχανία αυτοκινήτων η μηχανική μάθηση έχει τη δυνατότητα να δώσει λύσεις, σε πολλούς και διαφορετικούς κλάδους της από το κομμάτι της έρευνας και ανάπτυξης μέχρι την διοίκηση ανθρώπινου δυναμικού και την επιχειρηματικότητα.

2.2.1 Μηχανική μάθηση στην ανάπτυξη ηλεκτρικών οχημάτων

Η μηχανική μάθηση μπορεί να διαδραματίσει καθοριστικό ρόλο στην βελτιστοποίηση της τεχνολογίας ηλεκτρικών οχημάτων (EVs). Με την αξιοποίηση αλγορίθμων και τεχνικών μηχανικής μάθησης, οι ερευνητές είναι σε θέση να αντιμετωπίσουν διάφορες προκλήσεις που σχετίζονται με τα ηλεκτρικά οχήματα, συμπεριλαμβανομένης της τεχνολογίας μπαταριών, της βελτιστοποίησης εμβέλειας και της υποδομής φόρτισης. Ακολουθεί μια ανάλυση του τρόπου με τον οποίο η μηχανική μάθηση μπορεί να συμβάλει σε καθέναν από αυτούς τους τομείς:

- **Τεχνολογία μπαταριών:** η μηχανική μάθηση μπορεί να χρησιμοποιηθεί για τη βελτίωση της απόδοσης και της αποτελεσματικότητας των μπαταριών, με τη δημιουργία. Οι αλγόριθμοι μηχανικής μάθησης μπορούν να αναλύσουν δεδομένα

απόδοσης της μπαταρίας, συμπεριλαμβανομένης της θερμοκρασίας, των ρυθμών φόρτισης/εκφόρτισης και των προτύπων υποβάθμισης, για να εντοπίσουν τις βέλτιστες συνθήκες λειτουργίας και να παρατείνουν τη διάρκεια ζωής της μπαταρίας.

- **Βελτιστοποίηση εύρους:** Η μηχανική μάθηση μπορεί να βοηθήσει στη βελτιστοποίηση της εμβέλειας των ηλεκτρικών οχημάτων αναλύοντας τα πρότυπα οδήγησης, τα δεδομένα κίνησης, τις καιρικές συνθήκες και άλλες μεταβλητές. Συλλέγοντας και αναλύοντας δεδομένα από διάφορες πηγές, οι αλγόριθμοι μηχανικής μάθησης μπορούν να δημιουργήσουν ακριβείς προβλέψεις εμβέλειας, επιτρέποντας στους οδηγούς να σχεδιάζουν τα ταξίδια τους πιο αποτελεσματικά.
- Η ικανότητα της μηχανικής μάθησης να πραγματοποιεί προβλέψεις με πολύ υψηλή ακρίβεια θα μπορούσε να φανεί χρήσιμη στην διασφάλιση των προσωπικών δεδομένων τόσο του οχήματος όσο και του οδηγού. Προσωπικά δεδομένα μπορεί να είναι η θέση του οχήματος, ο αριθμός της άδειας του οδηγού, ή ο αριθμός του διπλώματος οδήγησης, ή οποιοδήποτε άλλο στοιχείο, ευάλωτο σε οποιονδήποτε κακόβουλο επιθυμεί να το αντιγράψει, ή να το υποκλέψει. Παρόλα αυτά, ήδη πολλές εταιρίες προσανατολίζονται σε μοντέλα μηχανικής μάθησης προκειμένου, τα όποια είναι σε θέση να αναγνωρίσουν μοτίβα κυβερνοεπιθέσεων και να τις αποκλείσουν.
- Παρακάτω παρατίθενται στοιχεία όπου η χρήση της μηχανικής μάθησης μπορεί να βοηθήσει στην ασφάλεια και προστασία των προσωπικών δεδομένων

2.3 Χρήση μηχανικής μάθησης για ασφάλεια και προστασία δεδομένων

Τα περισσότερα σύγχρονα οχήματα είναι συνδεδεμένα με το διαδίκτυο μέσα από το οποίο εξυπηρετούν ορισμένες ανάγκες του καταναλωτή και στέλνουν δεδομένα στους servers του κατασκευαστή σχετικά με την απόδοση του αυτοκινήτου. Το γεγονός αυτό όμως καθιστά εύάλωτο το αυτοκίνητο σε «κυβερνοεπιθέσεις», θέτοντας σε κίνδυνο τόσο τον καταναλωτή όσο και τον κατασκευαστή. Η μηχανική μάθηση ωστόσο έχει τη δυνατότητα να συμβάλει σημαντικά στην αντιμετώπιση των συγκεκριμένων κυβερνοεπιθέσεων.

2.3.1 Η μηχανική μάθηση στην υπηρεσία της ασφάλειας των δεδομένων των αυτοκινήτων

Υπάρχουν διάφοροι τρόποι με τους οποίους η χρήση της μηχανικής μάθησης μπορεί να βοηθήσει στην ασφάλεια των δεδομένων των αυτοκινήτων και κατ'επέκταση των οδηγών. Οι τομείς που η μηχανική μάθησης μπορεί να συμβάλει σε αυτό είναι οι εξής:

1. **Ανίχνευση Ανωμαλιών:** Με τη χρήση αλγορίθμων μηχανικής μάθησης, είναι δυνατόν να ανιχνευθούν ανωμαλίες στη συμπεριφορά του αυτοκινήτου. Οι αλγόριθμοι μπορούν να εκπαιδευτούν να αναγνωρίζουν τα κανονικά μοτίβα λειτουργίας του αυτοκινήτου και να εντοπίζουν ανεπιθύμητες ενέργειες ή συμπεριφορές που υποδηλώνουν κυβερνοεπίθεση.
2. **Πρόβλεψη Κινδύνων:** Οι αλγόριθμοι μηχανικής μάθησης μπορούν να αναλύσουν δεδομένα από διάφορες πηγές, όπως αισθητήρες και συστήματα πλοήγησης, για να προβλέψουν πιθανούς κινδύνους ή επιθέσεις. Αυτό μπορεί να βοηθήσει το σύστημα να αντιδράσει άμεσα και να λάβει μέτρα προστασίας.
3. **Ανίχνευση Κακόβουλου Λογισμικού:** Οι τεχνικές μηχανικής μάθησης μπορούν να ανιχνεύουν κακόβουλο λογισμικό που προσπαθεί να εισέλθει ή να παραμείνει κρυφό στο σύστημα του αυτοκινήτου. Με την ανίχνευση αυτού του είδους λογισμικού, μπορούν να ληφθούν μέτρα ασφαλείας για την αποτροπή πιθανών επιθέσεων.

Είναι αντιληπτό από τα παραπάνω ότι η μηχανική μάθηση μπορεί να αποβεί χρήσιμη στην προστασία του οδηγού καθώς μπορεί να προβλέψει και να κατανοήσει τις ανάγκες του και ως καταναλωτή.

2.4 Καλύτερη κατανόηση των καταναλωτικών αναγκών

Η μηχανική μάθηση μπορεί να αξιοποιηθεί για την ανάλυση μεγάλου όγκου δεδομένων από διάφορες πηγές, όπως κοινωνικά δίκτυα, ιστοσελίδες, και φόρουμ, προκειμένου να κατανοηθούν οι προτιμήσεις και οι απόψεις των καταναλωτών. Με τη χρήση αλγορίθμων μηχανικής μάθησης, όπως οι αλγόριθμοι συστάσεων και οι αναλυτικοί αλγόριθμοι, μπορούν να ανιχνευθούν τάσεις, προτεραιότητες και προτιμήσεις των καταναλωτών. Αυτή η κατανόηση μπορεί να βοηθήσει τις αυτοκινητοβιομηχανίες να προσαρμόσουν τις προσφερόμενες υπηρεσίες, τον σχεδιασμό των οχημάτων, και τις μάρκετινγκ εκστρατείες σύμφωνα με τις προτιμήσεις του κοινού.

Επιπλέον, η μηχανική μάθηση μπορεί να βελτιώσει την εμπειρία χρήστη με την ανάπτυξη εξατομικευμένων λύσεων. Με βάση τα προηγούμενα δεδομένα και τα στοιχεία των καταναλωτών, οι αλγόριθμοι μηχανικής μάθησης μπορούν να παράγουν προσαρμοσμένες συστάσεις για υπηρεσίες, προϊόντα και χαρακτηριστικά των οχημάτων, προσφέροντας έτσι μια εξατομικευμένη εμπειρία για τους χρήστες. Αυτό μπορεί να περιλαμβάνει εξατομίκευση της ρύθμισης του καθίσματος, των προτιμήσεων για τον ήχο ή τον κλιματισμό, και την παροχή πληροφοριών και ψυχαγωγίας μέσω των συστημάτων ψυχαγωγίας και ενημέρωσης.

2.5 Βελτίωση ενεργειακής απόδοσης και μείωση εκπομπών

Ο συνδυασμός της ενεργειακής κρίσης μαζί με την χρόνια περιβαλλοντική κρίση, καθιστούν πλέον αναγκαία τη στροφή σε πιο φιλικά προς το περιβάλλον αυτοκίνητα. Αυτό σημαίνει πως, οι κινητήρες των αυτοκινήτων να αποδίδουν με χρήση μικρότερης ποσότητας καυσίμου να έχουν αυξημένη απόδοση. Η μηχανική μάθηση μπορεί να συμβάλει στη βελτίωση της ενεργειακής απόδοσης και τη μείωση των εκπομπών των αυτοκινήτων με αρκετούς τρόπους.

1. **Βελτιστοποίηση της κίνησης:** Οι αλγόριθμοι μηχανικής μάθησης μπορούν να αναλύουν μεγάλο όγκο δεδομένων για να εξάγουν πληροφορίες σχετικά με τα πρότυπα οδήγησης και τις συνθήκες κυκλοφορίας. Με βάση αυτές τις πληροφορίες, μπορούν να αναπτυχθούν

προγράμματα κατεύθυνσης που βελτιστοποιούν την κίνηση, μειώνοντας την κατανάλωση καυσίμου και τις εκπομπές. (Sivaraman, V., & Trivedi, 2014)

2. **Πρόβλεψη Κίνησης:** Οι αλγόριθμοι μηχανικής μάθησης επίσης έχουν τη δυνατότητα να προβλέψουν τις συνθήκες κυκλοφορίας. Αυτή η πρόβλεψη μπορεί να χρησιμοποιηθεί για τη βελτιστοποίηση της λειτουργίας του συστήματος υβριδικής κίνησης, για παράδειγμα, με την προσαρμογή στην κατανομή της ισχύος μεταξύ του κινητήρα καύσης και του ηλεκτροκινητήρα με στόχο την επίτευξη της μέγιστης απόδοσης. (Lin, Y., 2019)
3. **Βελτιστοποίηση του Συστήματος Κίνησης:** Η βελτιστοποίηση των συστημάτων κίνησης, όπως η στρατηγική επιλογής ταχυτήτων και η προσαρμογή τους στις συνθήκες οδήγησης, αποτελεί μια από τις συνηθέστερες εφαρμογές της μηχανικής μάθησης, καθώς αρκετά αυτοκίνητα στην αγορά λειτουργούν με αυτόματο κιβώτιο ταχυτήτων.
4. **Ανίχνευση δυσλειτουργιών:** Οι αλγόριθμοι μηχανικής μάθησης έχουν την ικανότητα να ανιχνεύουν προβλήματα και αποτυχίες στα συστήματα οχημάτων, εγκαίρως. Αυτό επιτρέπει την πρόληψη σοβαρών βλαβών και επιτρέπει την εκτέλεση προληπτικής συντήρησης, προσφέροντας βελτιωμένη αξιοπιστία και μειώνοντας τα κόστη συντήρησης.

Η μηχανική μάθηση όμως εκτός από την ασφάλεια που παρέχει στον οδηγό, μπορεί να συμβάλει σημαντικά στην αντιμετώπιση της ενεργειακής κρίσης καθώς μπορεί να ελέγχει την ποσότητα του καυσίμου, μέσω της οδικής συμπεριφοράς. Είναι μια πολύ ενδιαφέρουσα καινοτομία που μπορεί να φέρει σημαντικά οικονομικά οφέλη. Στο επόμενο κεφάλαιο αναλύεται η εφαρμογή της μηχανικής μάθησης στον έλεγχο των καυσίμων όπως εφαρμόστηκε με επιτυχία στην αυτοκινητοβιομηχανία της Toyota . Είναι μία πειραματική έρευνα που φιλοδοξεί να αλλάξει τα δεδομένα τόσο στην οδική συμπεριφορά όσο και στην μείωση στην κατανάλωση των καυσίμων, επιφέροντας σημαντικά οικονομικά οφέλη τόσο στις αυτοκινητοβιομηχανίες όσο και στους καταναλωτές.

Κεφάλαιο 3 Μελέτη κατανάλωσης καυσίμου στην σύγχρονη αυτοκινητοβιομηχανία.

Σύμφωνα με έρευνες που πραγματοποιήθηκαν από την Ευρωπαϊκή Ένωση το 2017 και το 2019, παρατηρήθηκε πως σε όλη τη αυτοκινητοβιομηχανία υπάρχει ένα κενό μεταξύ της πραγματικής κατανάλωσης καυσίμου και αυτής που προκύπτει από τον κύκλο μετρήσεων - WLTP- που πραγματοποιούν οι κατασκευαστές αυτοκινήτων για τα αυτοκίνητα που εισάγονται στην αγορά. (Preparation for collection and monitoring of real-world fuel consumption data for light and heavy duty vehicles, Report for European Commission – DG Climate Action)

3.1 Τι είναι ο γύρος μετρήσεων WLTP;

Η WLTP, ή αλλιώς Παγκόσμια εναρμονισμένη διαδικασία δοκιμής ελαφρών οχημάτων (Worldwide Harmonized Light Vehicles Test Procedure), αποτελεί ένα Παγκοσμίως Εναρμονισμένο Πρότυπο για τον προσδιορισμό των επιπέδων ρύπων, εκπομπών διοξειδίου του άνθρακα και κατανάλωσης καυσίμου παραδοσιακών και υβριδικών αυτοκινήτων, καθώς και της γκάμας των πλήρως ηλεκτρικών οχημάτων. Αναπτύχθηκε από την Οικονομική Επιτροπή των Ηνωμένων Εθνών για την Ευρώπη (OEE / HE) και έγινε υποχρεωτική στην Ευρωπαϊκή Ένωση το 2017. Ο κύκλος WLTP έχει σχεδιαστεί για να παρέχει πιο ρεαλιστικές συνθήκες δοκιμών σε σύγκριση με την προηγούμενη δοκιμή του νέου Ευρωπαϊκού κύκλου οδήγησης (NEDC). Περιλαμβάνει ένα ευρύτερο φάσμα καταστάσεων οδήγησης, όπως αστικό, προαστιακό, κεντρικό δρόμο και αυτοκινητόδρομο.

Η διαδικασία δοκιμής WLTP αποτελείται από:

- Συλλογή δεδομένων πραγματικής οδήγησης: τα πραγματικά δεδομένα οδήγησης συλλέγονται παγκοσμίως για τη δημιουργία ρεαλιστικών προφίλ χρήσης. Τα δεδομένα αντιπροσωπεύουν την καθημερινή χρήση από τους οδηγούς σε διαφορετικές περιοχές.
- Ανάπτυξη κύκλου δοκιμής: με βάση τα δεδομένα που συλλέγονται, αναπτύσσεται ο κύκλος δοκιμής. Περιλαμβάνει τέσσερα μέρη με ποικίλες μέσες ταχύτητες: χαμηλή, μεσαία, υψηλή και εξαιρετικά υψηλή, προσομοιώνοντας την αστική και την υπεραστική οδήγηση.
- Εργαστηριακές δοκιμές: τα οχήματα ελέγχονται σε ελεγχόμενες εργαστηριακές συνθήκες. Οι δοκιμές γίνονται σε δυναμόμετρο πλαισίου. Η διαδικασία περιλαμβάνει πιο ρεαλιστικές συνθήκες δοκιμής, όπως υψηλότερη μέση ταχύτητα και μέγιστη ταχύτητα, μεγαλύτερες αποστάσεις δοκιμής, πιο δυναμικές και αντιπροσωπευτικές επιταχύνσεις και επιβραδύνσεις και μικρότερες στάσεις.
- Υπολογισμός εκπομπών και κατανάλωσης καυσίμου: μετά τις δοκιμές, οι τιμές εκπομπών CO₂ και κατανάλωσης καυσίμου υπολογίζονται με βάση τα αποτελέσματα.
- Το WLTP έχει σχεδιαστεί για να παρέχει στους καταναλωτές πιο ακριβείς πληροφορίες σχετικά με την κατανάλωση καυσίμου και τις εκπομπές ενός οχήματος, βοηθώντας τους να επιλέξουν πιο φιλικά προς το περιβάλλον οχήματα. Ωστόσο, ενώ το WLTP παρέχει πιο ρεαλιστικές συνθήκες δοκιμών από το προηγούμενο NEDC, εξακολουθεί να είναι εργαστηριακή δοκιμή και οι τιμές που λαμβάνονται ενδέχεται να διαφέρουν από την πραγματική απόδοση στο δρόμο.

3.2 Μελέτη σύγκρισης πραγματικής κατανάλωσης καυσίμου και κύκλου WLTP

Στόχοι της έρευνας:

Με αφορμή, λοιπόν, την παραπάνω έρευνα, διεξήχθη μια μελέτη η οποία έχει ως βασικούς στόχους:

- 1) Να διερευνηθεί αν οι κατασκευαστές αυτοκινήτων στην Ευρώπη ανακοινώνουν ακριβή στοιχεία όσο αφορά την κατανάλωση καυσίμων.
- 2) Ποιοι παράγοντες μπορεί να οδηγούν στο κενό μεταξύ WLTP και πραγματικής κατανάλωσης καυσίμου.

Για την πραγματοποίηση της εν λόγω μελέτης ακολουθήθηκε η παρακάτω διαδικασία:

- 1) Συλλογή Δεδομένων
- 2) Οργάνωση και «καθαρισμός» δεδομένων
- 3) Εμπλουτισμός βάσης δεδομένων
- 4) Οπτικοποίηση δεδομένων και εξαγωγή συμπερασμάτων

3.3 Συλλογή Δεδομένων:

3.3.1 Απαιτήσεις Δεδομένων

Πρώτου επιλεχθούν οι πηγες από τις οποίες θα αντληθούν δεδομένα, κρίνεται απαραίτητο να υπάρχει μία αρχική εικόνα για το ποια θα είναι τα επιθυμητά χαρακτηριστικά της βάσης δεδομένων.

Τα χαρακτηριστικά που είναι απαραίτητα για την βάση δεδομένων αποτελούν:

- Όνομα κατασκευαστή (Brand Name)
- Μοντέλο (Model)
- Είδος κινητήρα (Engine Type)
- Έτος πώλησης (Year)
- (Μέση) Κατανάλωση καυσίμου σε λίτρα/100χλμ (Average_FC)
- Κατανάλωση κύκλου WLTP (WLTP_FC)
- Μάζα (Weight)

Για να διαχωριστούν αυτοκίνητα ίδιου μοντέλου αλλά διαφορετικής γενιάς. Σε κάθε γενιά παραγωγής μπορεί ο κατασκευαστής να αλλάζει κάποιο υλικό ή να προσθαφαιρεί εξαρτήματα, με αποτέλεσμα να αλλάζει και η μάζα του αυτοκινήτου.

Αριθμός πωλήσεων μοντέλου (Cars Sold)

Χρησιμοποιείται για να αναγνωριστούν τα αυτοκίνητα με το μεγαλύτερο μερίδιο στην αγορά – άρα και αυτά που συναντώνται συχνότερα στους ευρωπαϊκούς δρόμους. Έτσι δημιουργείται μια πιο ξεκάθαρη εικόνα για το ποια αυτοκίνητα πρέπει να εξεταστούν στη μελέτη.

3.3.2 Αναζήτηση πηγών

Γνωρίζοντας ποιες είναι οι απαιτήσεις δεδομένων για την μελέτη, ξεκινάει η αναζήτηση πηγών που περιέχουν τα δεδομένα που είναι αναγκαία. Προτού, όμως ξεκινήσει η αναζήτηση αυτή, κρίνεται απαραίτητο να οριστούν μερικά συγκεκριμένα κριτήρια τα οποία θα εξασφαλίσουν μια αποτελεσματική και αξιόπιστη συλλογή δεδομένων. Μία πηγή λοιπόν οφείλει να πληροί τις παρακάτω προϋποθέσεις:

Να περιέχει δεδομένα που αφορούν την Ευρωπαϊκή αυτοκινητοβιομηχανία αποκλειστικά. Αυτό γιατί τα αυτοκίνητα που κυκλοφορούν στους Ευρωπαϊκούς δρόμους πληρούν συγκεκριμένες προϋποθέσεις και περνάνε από ειδικές δοκιμές προκειμένου η κατανάλωσή τους να είναι χαμηλότερη από τα όρια της νομοθεσίας που ορίζει η Ευρωπαϊκή Ένωση. (<https://eur-lex.europa.eu/>)

Η πηγή να περιέχει πρόσφατα δεδομένα, από το 2020 μέχρι σήμερα. Η συγκεκριμένη μελέτη αφορά τα σύγχρονα αυτοκίνητα της αγοράς της Ευρώπης. Άρα αν χρησιμοποιηθούν παλαιότερα ενδέχεται να μην εξαχθούν αντιπροσωπευτικά αποτελέσματα. Ωστόσο, αυτοκίνητα που παρήχθησαν το 2018 και 2019 θα χρησιμοποιούνταν αποκλειστικά σε περίπτωση που τα πιο πρόσφατα αποτελούσαν μικρό δείγμα.

Τέλος η πηγή οφείλει να παρέχει αξιόπιστα και ρεαλιστικά δεδομένα. Αυτό σημαίνει ότι .θα πρέπει να αποτελεί πιστοποιημένο πάροχο δεδομένων είτε στην ακαδημαϊκή κοινότητα είτε στην αυτοκινητοβιομηχανία.

3.3.3 Εύρεση και αξιολόγηση πηγών

Αφού ορίστηκαν τα κριτήρια με τα οποία θα διεξαχθεί η αναζήτηση πηγών, βρέθηκαν 6 πιθανές πηγές από τις οποίες μπορούν να αντληθούν δεδομένα.

Αναλυτικότερα, οι πηγές που βρέθηκαν, αναλύονται παρακάτω:

Πηγές πραγματικής κατανάλωσης καυσίμου

spritmonitor.de: Γερμανική ιστοσελίδα που επιτρέπει στους χρήστες της να παρακολουθούν και να αναλύουν την κατανάλωση καυσίμου. Βασίζεται στην κοινότητα, με τα μέλη να εισάγουν δεδομένα σχετικά με την κατανάλωση των οχημάτων τους σε πραγματικές συνθήκες. Αυτό την καθιστά πολύτιμο πόρο για να αντληθούν δεδομένα πραγματικής κατανάλωσης καυσίμου.

mile21.eu: Διαδικτυακή πλατφόρμα, χρηματοδοτούμενη από την Ευρωπαϊκή Ένωση, που επιτρέπει στους χρήστες να παρακολουθούν τη δική τους χρήση καυσίμων.

Οι δύο αυτές πηγές παρέχουν αρκετές πληροφορίες για την πραγματική κατανάλωση καυσίμου. Ωστόσο, δεν δίνουν αρκετές πληροφορίες για να ομαδοποιηθούν και να οργανωθούν σε μια βάση δεδομένων. Επιπλέον, στο «spritmonitor.de» δεν εμπεριέχονται παντού οι τιμές του κύκλου WLTP. Για αυτό βρέθηκαν άλλες 4 πηγές που θα χρησιμοποιηθούν για τον εμπλουτισμό των δεδομένων, δηλαδή την συμπλήρωση των χαρακτηριστικών που λείπουν από τις βασικές πηγές.

3.3.4 Πηγές εμπλουτισμού της βάσης δεδομένων:

- adac.de: Ο επίσημος ιστότοπος της Allgemeiner Deutscher Automobil-Club. Πρόκειται για τον μεγαλύτερο σύνδεσμο της Γερμανικής και της Ευρωπαϊκής αυτοκινητοβιομηχανίας. Ο ιστότοπος προσφέρει ένα ευρύ φάσμα πληροφοριών και υπηρεσιών που σχετίζονται με τα αυτοκίνητα και τις μεταφορές. Πιο συγκεκριμένα περιλαμβάνονται:
- FleetNews.co.uk: Αγγλική ιστοσελίδα που παρέχει πολλές και διάφορες πληροφορίες σχετικά με την αυτοκινητοβιομηχανία. Ο ιστότοπος παρέχει πληθώρα πληροφοριών, συμπεριλαμβανομένων:
 - Ειδήσεις
 - Αναλύσεις από εξειδικευμένους επαγγελματίες
 - Κριτικές
 - Πληροφορίες για λογισμικό

ACEA (European Automobile Manufacturers Association): Εμπορική ομάδα με έδρα τις Βρυξέλλες που εκπροσωπεί τους 15 μεγάλους κατασκευαστές αυτοκινήτων, φορτηγών, φορτηγών και λεωφορείων με έδρα την Ευρώπη. Επιπλέον, υποστηρίζει πολιτικές και νομοθεσίες σε επίπεδο Ε.Ε. που είναι ευνοϊκές για την αυτοκινητοβιομηχανία και τέλος παρέχει τεχνική εμπειρογνομosύνη και δεδομένα σχετικά με την αυτοκινητοβιομηχανία.

Όλες οι πληροφορίες που συλλέχθηκαν, αξιολογήθηκαν ως προς την ορθότητά τους, στον βαθμό που καθίσταται αυτό εφικτό, μετά την διασταύρωση τους με τις επίσημες πληροφορίες του κατασκευαστή.

Μετά την μελέτη των προαναφερθέντων πηγών, έγινε καταγραφή για τον τρόπο που καλύπτει η κάθε πηγή τις ανάγκες της βάσης δεδομένων που θα δημιουργηθεί. Η εν λόγω καταγραφή εκφράζεται από τον ακόλουθο πίνακα.

Πίνακας 1

	Spritmonitor	Mile21	Adac	Κατασκευαστής	Fleet News	ACEA
Κατασκευαστής	○	○	○	○	○	○
Μοντέλο	○	○	○	○	○	-
Είδος κινητήρα	○	○	○	○	○	-
Μέση κατανάλωση	○	○	-	-	-	-
Κατανάλωση WLTP κύκλου	-	○	○	○	○	-
Βάρος	-	-	○	○	-	-
Αριθμός πωλήσεων	-	-	-	-	-	○

Κατανάλωση (πραγματική)
καυσίμου

Εμπλουτισμός
των δεδομένων

Προτού επιλεγθούν η πηγές που θα χρησιμοποιηθούν, αξίζει να αναφερθεί η κύρια μέθοδος συλλογής των δεδομένων. Οι πηγές περιέχουν τεράστιο όγκο δεδομένων, γεγονός που καθιστά την χειροκίνητη συλλογή των δεδομένων αδύνατη. Για το λόγο αυτό θα χρησιμοποιηθεί μια μέθοδος συλλογής δεδομένων με το όνομα «web scraping».

Τι είναι η μέθοδος web scraping;

Πρόκειται για μια μέθοδο εξαγωγής δεδομένων από το διαδίκτυο. Ο τρόπος με τον οποίο αυτό επιτυγχάνεται, περιλαμβάνει μια σειρά από αυτοματοποιημένες ενέργειες που πραγματοποιούνται συνήθως μέσω ενός προγραμματισμένου λογισμικού (web crawler ή bot). Οι ενέργειες εξαγωγής δεδομένων ακολουθούν την παρακάτω μεθοδολογία:

1. Αίτημα πρόσβασης στον ιστότοπο (URL request).
2. «Διάβασμα» του κώδικα του γραφικού περιβάλλοντος της ιστοσελίδα –συνήθως είναι γραμμένος σε κώδικα της γλώσσας HTML.
3. Εντοπισμός των δεδομένων που πρέπει να εξαχθούν,
4. Αποθήκευση των δεδομένων σε οργανωμένες στήλες ενός αρχείου –συνήθως csv ή excel.

Νομικά πλαίσια web scraping

Σύμφωνα με την πιο πρόσφατη νομοθεσία της Ε.Ε. (2000/31/EC Digital Services Act 15/12/2020), η "αναπαραγωγή δημοσίως διαθέσιμου περιεχομένου" δεν είναι παράνομη. Αυτό σημαίνει πως αν δεν εξαχθούν πληροφορίες που περιέχουν προσωπικά δεδομένα, δεν τίθεται κάποιο νομικό ζήτημα εντός Ευρωπαϊκής Ένωσης. Παρόλα αυτά ορισμένες ιστοσελίδες μπορεί να απαγορεύουν τη χρήση web scraping στους όρους χρήσης τους, γεγονός που εμποδίζει την εξαγωγή δεδομένων από αυτές.

3.4.1 Επιλογή πηγών:

Είναι πλέον γνωστές τόσο οι πιθανές πηγές που θα χρησιμοποιηθούν για την δημιουργία της βάσης δεδομένων, όσο και η μεθοδολογία που θα χρησιμοποιηθεί για την εξαγωγή τους, καθίσταται δυνατό να αξιολογηθεί το ποιες από αυτές θα είναι οι πιο χρήσιμες. Φυσικά, για τη διαλογή τους ορίστηκαν ξανά μερικά κριτήρια, σύμφωνα με τα οποία εξετάστηκε η κάθε μία από τις πηγές.

Μια πηγή οφείλει να πληροί τις παρακάτω προϋποθέσεις:

- Αξιοπιστία : Είναι τα δεδομένα ρεαλιστικά;
- Πληρότητα: Να λείπει χαμηλός αριθμός τιμών
- Προγραμματιστική δυσκολία: Να μπορεί να φτιαχτεί το πρόγραμμα εξαγωγής δεδομένων σε από χρόνο, χωρίς να χρειαστεί να γίνει πιο πολύπλοκη η διαδικασία.
- Ποικιλία δεδομένων: Πόσες διαφορετικές απαιτήσεις ικανοποιούνται από κάθε πηγή.

Αξιολόγηση πηγών:

Εφόσον ορίστηκαν τα κριτήρια με τα οποία θα επιλεχθούν οι πηγές, κάθε μία από αυτές εξετάζεται με βάσει αυτών.

➤ Spritmonitor:

- Αξιοπιστία : Επαρκής αξιοπιστία. Είναι γεγονός πως πηγές δεδομένων που έχουν καταχωρηθεί από ανθρώπους που ενδέχεται να μην είναι σχετικοί με την αυτοκινητοβιομηχανία ή να μην χειρίζονται σωστά την πλατφόρμα της ιστοσελίδας. Παρόλα αυτά η ιστοσελίδα παρέχει δεδομένα από κάθε ανεφοδιασμό του κάθε χρήστη. Αυτό καθιστά πολύ εύκολη την εκκαθάριση ακραίων τιμών που δύναται να αλλοιώσουν το αποτέλεσμα της έρευνας.
- Πληρότητα: Μερική έλλειψη τιμών. Παρόλα αυτά η πηγή παρέχει αρκετές πληροφορίες, ώστε να είναι εφικτή η εξαγωγή συμπερασμάτων
- Προγραμματιστική δυσκολία: Εφικτή. Στα πλαίσια του προγραμματισμού τόσο για τη συλλογή όσο και για το ξεκαθάρισμα των δεδομένων, δεν φαίνεται να υπάρχει κάποιο σοβαρό εμπόδιο. Η ιστοσελίδα είναι αρκετά δομημένη αρά είναι εφικτό να εξαχθούν ορθώς ομαδοποιημένα τα δεδομένα.
- Ποικιλία δεδομένων: Καλύπτει σχεδόν όλες τις βασικές κατηγορίες δεδομένων που είναι αναγκαίες, με εξαίρεση την κατανάλωση WLTP κύκλου.

➤ Mile21:

- Αξιοπιστία : Σίγουρα πρόκειται για αξιόπιστη πηγή, αφού επρόκειτο για οργανισμό χρηματοδοτούμενο και αξιολογημένο από την Ε.Ε.. Ωστόσο, υπάρχει μια αβεβαιότητα σχετικά με τον εντοπισμό στην ακραίων τιμών, εφόσον η ιστοσελίδα παρέχει αποκλειστικά τον μέσο όρο της κατανάλωσης.
- Πληρότητα: Πολύ μικρή πληρότητα. Παρόλο που έχει σχεδόν μηδενική έλλειψη τιμών, ο πληθυσμός των δεδομένων είναι αρκετά χαμηλός (λιγότερα από 1700 δείγματα).
- Προγραμματιστική δυσκολία: Το γραφικό περιβάλλον του ιστότοπου είναι κρυπτογραφημένο, γεγονός που απαιτεί υπερβολικό χρόνο και προσπάθεια, ειδικότερα για μια βάση δεδομένων που αποτελείται μόνο από 1700 καταγραφές.
- Ποικιλία δεδομένων: Παρέχει όλες τις βασικές πληροφορίες για την διεξαγωγή της έρευνας, ακόμη και τις τιμές του WLTP κύκλου.

➤ ADAC:

- Αξιοπιστία : Πολύ υψηλή αξιοπιστία. Απαρτίζεται αποκλειστικά από δεδομένα που έχουν καταχωρήσει οι ίδιοι κατασκευαστές.
- Πληρότητα: Πολύ υψηλή. Περιέχει όλες τις πληροφορίες για τα πιο πρόσφατα αυτοκίνητα που κυκλοφορούν στην ευρωπαϊκή αγορά.
- Προγραμματιστική δυσκολία: Μέτρια. Παρόλη την ύπαρξη μιας πολυστρωματικής σελιδοποίησης, τα δεδομένα είναι ξεκάθαρα και αρμονικά οργανωμένα στο γραφικό περιβάλλον.

- Ποικιλία δεδομένων: Καλύπτει όλες τις κατηγορίες που λείπουν από τις πηγές πραγματικής κατανάλωσης καυσίμου, με εξαίρεση τον αριθμό πωλήσεων.
- Fleet News:
 - Αξιοπιστία : Χαμηλή. Πρόκειται για δημοσιογραφική ιστοσελίδα, γεγονός που κρίνει απαραίτητη τη διασταύρωση των δεδομένων με άλλες πηγές.
 - Πληρότητα: Παρέχει πληροφορίες για όλα τα πρόσφατα μοντέλα αυτοκινήτων, χωρίς να λείπει κάποια τιμή.
 - Προγραμματιστική δυσκολία: Χαμηλή. Πολύ απλή δομή του γραφικού περιβάλλοντος.
 - Ποικιλία δεδομένων: Το μόνο ουσιαστικό περιεχόμενο που προσφέρει ως πηγή δεδομένων αποτελεί η κατανάλωση κύκλου WLTP.
- ACEA:
 - Αξιοπιστία : Υψηλή. Όλα τα δεδομένα είναι καταγραμμένα απευθείας από την Ε.Ε.
 - Πληρότητα: Δεν υπάρχει έλλειψη τιμών. Ωστόσο η τιμές της κατανάλωσης WLTP περιέχουν ένα ευρύτερο μέσο όρο για το κάθε μοντέλο και δεν υπάρχει κάποια περαιτέρω εμβάθυνση.
 - Προγραμματιστική δυσκολία: Πολύ υψηλή. Όλα τα δεδομένα είναι σε αρχεία μορφής pdf, με επιπλέον μεγάλο αριθμό σελιδοποίησης. Για να μπορέσει λοιπόν να γίνει η εξαγωγή των δεδομένων θα πρέπει το πρόγραμμα να κατεβάσει στον τοπικό δίσκο όλα τα αρχεία και να ανοίξει το καθένα ξεχωριστά, κάτι που απαιτεί αρκετό χρόνο και υπολογιστικούς πόρους.
 - Ποικιλία δεδομένων: Παρέχονται πληροφορίες για τον αριθμό πωλήσεων και κατανάλωση κύκλου WLTP.

3.4.2 Πίνακας αξιολόγησης πηγών: Ο παρακάτω πίνακας εμφανίζει μια οπτικοποίηση της αξιολόγησης όλων το πηγών, μέσα από τον οποίο μπορούν να αναγνωριστούν οι πιο αποδοτικές πηγές εξαγωγής δεδομένων

	Κατανάλωση καυσίμων		Εμπλουτισμός δεδομένων			
	Spritmonitor	Mile21	ADAC	Κατασκευαστής	Fleet News	ACEA
Αξιοπιστία	○	△	○	○	△	○
Πληρότητα	○	○	○	○	○	△
Προγραμματιστική προσπάθεια	△	✗	○	✗	○	△
Ποικιλία δεδομένων	○	△	○	○	△	✗

Πίνακας 2

○: Ιδανικό

△: Επαρκεί μερικώς. Απαιτείται προσοχή.

✗: Δεν πληροί το κριτήριο

3.4.3 Συμπεράσματα:

Από την αξιολόγηση που προηγήθηκε, αποφασίστηκε πως θα χρησιμοποιηθεί το spritmonitor.de ως πηγή πραγματικών μετρήσεων κατανάλωσης καυσίμου και επίσης θα χρησιμοποιηθεί το ADAC για προστεθούν τα δεδομένα που λείπουν. Τέλος θα χρησιμοποιηθεί και η ACEA, για την εύρεση των αυτοκινήτων με το μεγαλύτερο μερίδιο αγοράς, καθώς αποτελεί τη μοναδική αξιόπιστη πηγή που μπορεί να δώσει πληροφορίες για τον ακριβή αριθμό πωλήσεων.

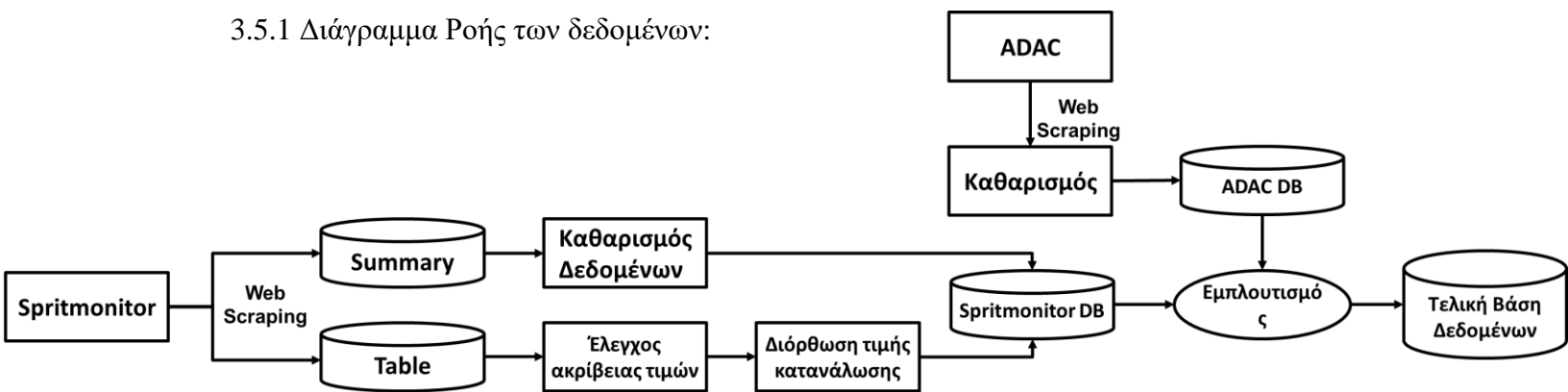
3.5 Εξαγωγή δεδομένων

Για την εξαγωγή των δεδομένων, πρόκειται να δημιουργηθούν 2 προγράμματα τα οποία θα αποκτούν πρόσβαση (μετά από συγκατάθεση του παρόχου) και θα αποθηκεύουν τα δεδομένα σε αρχεία csv.

Η γλώσσα προγραμματισμού που θα χρησιμοποιηθεί, είναι η Python, καθώς η πρόσβαση σε βιβλιοθήκες, προσφέρει μεγαλύτερη ευελιξία αλλά και ευκολία.

Προτού ξεκινήσει η εξαγωγή δεδομένων, κρίνεται απαραίτητη η κατανόηση της ευρύτερης εικόνας της διαδικασίας αυτής. Είναι αναγκαίο να οριστεί μια ροή που θα ακολουθήσουν έως ότου καταλήξουν σε μία ομογενή και ολοκληρωμένη βάση δεδομένων.

3.5.1 Διάγραμμα Ροής των δεδομένων:



Πίνακας 3

Με τη χρήση Web scraping, τα δεδομένα θα εξάγονται από το spritmonitor και θα αποθηκεύονται σε 2 ξεχωριστές βάσεις δεδομένων. Η βάση με το όνομα Summary θα περιέχει όλες τις γενικές πληροφορίες του κάθε αυτοκινήτου όπως όνομα κατασκευαστή, μοντέλο, έτος παραγωγής κ.ο.κ.. Η δεύτερη βάση, με το όνομα «Table»

θα περιέχει όλες τις μετρήσεις κατανάλωσης καυσίμου του κάθε αυτοκινήτου ξεχωριστά. Σκοπός της είναι η επαλήθευση της ορθότητας των δεδομένων που εξήχθησαν από την ιστοσελίδα. Μετά την επαλήθευση, η τιμή της κατανάλωσης καυσίμου αποθηκεύεται στο αντίστοιχο αυτοκίνητο, ολοκληρώνοντας έτσι την πρώτη βασική βάση δεδομένων για την μελέτη.

Στη συνέχεια ακολουθεί η εξαγωγή των δεδομένων για τον εμπλουτισμό της βάσης δεδομένων του spritmonitor. Τα δεδομένα, αφού εξαχθούν από την ιστοσελίδα του ADAC, ομαδοποιούνται, ξεκαθαρίζονται και ύστερα αποθηκεύονται στην 2^η βασική βάση δεδομένων («ADAC DB»).

Τέλος, χρησιμοποιείται η βάση εμπλουτισμού για προστεθεί η τιμή του κύκλου WLTP σε κάθε αυτοκίνητο από την βάση δεδομένων του spritmonitor.

Spritmonitor web scraping

Η εξαγωγή δεδομένων από τον ιστότοπο spritmonitor ακολουθεί τα παρακάτω βήματα:

1. Προ-φιλτράρισμα της βάσης δεδομένων για την απομόνωση των αναγκαίων δεδομένων.
2. Εξαγωγή και αποθήκευση της διεύθυνσης URL κάθε μεμονωμένου οχήματος.
3. Εξαγωγή πληροφοριών οχήματος από το κάθε ξεχωριστό URL.
4. Αποθήκευση αποκομμένων πληροφοριών στις βάσεις Summary και Table. Προ-φιλτράρισμα:

Το Spritmonitor αποτελεί μια διαδικτυακή βάση δεδομένων με πραγματική κατανάλωση καυσίμου συνολικά 622.500 διαφορετικών μηχανοκίνητων οχημάτων.

Για να επιταχυνθεί η διαδικασία εξαγωγής δεδομένων, η βάση δεδομένων προφιλτράρεται με τα ακόλουθα κριτήρια

- a. Vehicle type: Automobile
- b. Fuel: Diesel, Gasoline
- c. Minimum mileage: 1500km
- d. Year of manufacturing: 2020 – 2023

Χρησιμοποιώντας αυτές τις ρυθμίσεις φίλτρου, το συνολικό ποσό των οχημάτων έχει μειωθεί σε περίπου 28.000.

Vehicle type
Automobile

Make
All

Model
All

Fuel type
All

Fuel sort
All

Gearing type
All

Exact modelname
[Empty]

Year of manufacture
2020 to 2023

Engine power
[Empty] to [Empty] kW

Minimum kilometres
1500 km

Activity within
Don't care

Sorting
Consumption increas

Search

Consumption: All manufactures - All models

Picture	Vehicle	Mileage ↑	Quantity	User
	Polestar Polestar 2 SM LR FWD Electricity, 170 kW	0,12		ChrisJJ
	Hyundai IONIQ Premium Plug-in hybrid gasoline, 104 kW	0,15		Flyer
	Mercedes-Benz A250e Plug-in hybrid gasoline, 119 kW	0,21		Diereh
	Cupra Formentor Gasoline, 228 kW	0,21		torinco
	Renault intens Plug-in hybrid gasoline, 116 kW	0,24		daniabba
	Mercedes-Benz 250e Plug-in hybrid gasoline, 160 kW	0,26		kpmg
	Opel Grandland X Hybrid Plug-in hybrid gasoline, 133 kW	0,32		Tyran
	Mitsubishi Basis Plug-in hybrid gasoline, 138 kW	0,34		E_Landgraf

Εξαγωγή και αποθήκευση URLs

Εικόνα 1

Αφού λοιπόν έχουν απομονωθεί όλα τα αναγκαία για την παρούσα έρευνα αυτοκίνητα, ακολουθεί η εξαγωγή και αποθήκευση όλων των διευθύνσεων URL και η αποθήκευσή τους σε ένα αρχείο txt.

3.5.2 Μεθοδολογία και προγραμματισμός εξαγωγής URL

1 Εντοπισμός URL

2 Αποθήκευση όλων των URL στο txt αρχείο

3 Αίτημα πρόσβασης στην επόμενη σελίδα

4 Επανάληψη

Next >

```
sprit_link_final_text - Notepad
File Edit Format View Help
https://www.spritmonitor.de/en/detail/1301322.html
https://www.spritmonitor.de/en/detail/1274533.html
https://www.spritmonitor.de/en/detail/1370884.html
https://www.spritmonitor.de/en/detail/1203734.html
https://www.spritmonitor.de/en/detail/1364589.html
https://www.spritmonitor.de/en/detail/1273038.html
https://www.spritmonitor.de/en/detail/1266085.html
https://www.spritmonitor.de/en/detail/1196723.html
https://www.spritmonitor.de/en/detail/1109363.html
https://www.spritmonitor.de/en/detail/1328015.html
https://www.spritmonitor.de/en/detail/1385498.html
https://www.spritmonitor.de/en/detail/1220700.html
https://www.spritmonitor.de/en/detail/1210578.html
https://www.spritmonitor.de/en/detail/1182626.html
https://www.spritmonitor.de/en/detail/1268963.html
https://www.spritmonitor.de/en/detail/1264266.html
https://www.spritmonitor.de/en/detail/1345922.html
https://www.spritmonitor.de/en/detail/1334529.html
https://www.spritmonitor.de/en/detail/1378366.html
https://www.spritmonitor.de/en/detail/1390545.html
https://www.spritmonitor.de/en/detail/1357068.html
```

Εικόνα 2

Η παραπάνω μεθοδολογία εφαρμόζεται με τη χρήση του παρακάτω κώδικα:

```
import scrapy

class LinksSpider(scrapy.Spider):
    name = 'links'
    allowed_domains = ['www.spritmonitor.de'] #allowed domain definition
    #definition of the URL used to start the scraping process
    start_urls = ['https://www.spritmonitor.de/en/overview/0-All_manufactures/0-All_models.html?vehicletype=1&constyear_s=2020&constyear_e=2022&page=1']

    custom_settings = {'COOKIES_ENABLED': False,
                       'DOWNLOAD_DELAY': 3,
                       'AUTOTHROTTLER_ENABLED': True,
                       'AUTOTHROTTLER_DEBUG': True,
                       'RANDOMIZE_DOWNLOAD_DELAY': True
                      }

    def parse(self, response): #parsing function definition
        base_url = 'https://www.spritmonitor.de'

        links = response.css('td.description a::attr(href)') #location of URL unique identifier in HTML code
        for link in links:
            yield{
                'link': base_url + str(link.get()) #storing all URLs of cars inside the txt file
            }
        next_page = response.xpath('//*[@id="searchresultbox"]/div/a[last()]/@href').extract() #next page's URL location in HTML
        if next_page:
            next_url = ''.join(map(str, next_page))
            yield response.follow(next_url, callback=self.parse)
```

Εικόνα 3

Το πρόγραμμα αποτελείται από μια κλάση που χρησιμοποιεί την βιβλιοθήκη «scrapy». Η συγκεκριμένη κλάση αποτελεί μια από τις βασικές κλάσεις που προσφέρει η scrapy. Πρόκειται για μια κλάση γνωστή ως web crawler. Οι web crawlers αποτελούν λογισμικά που πλοηγούνται στο διαδίκτυο, διαβάζουν δεδομένα και τα αποθηκεύουν. Για την ορθή λειτουργία της κλάσης αυτής αρχικά ορίζονται ορισμένες παράμετροι και χαρακτηριστικά σύμφωνα με τα οποία θα λειτουργεί ο web crawler. Συγκεκριμένα, δίνεται όνομα για το λογισμικό το οποίο θα χρησιμοποιηθεί για να δοθεί εντολή λειτουργίας του λογισμικού. Έπειτα ορίζεται το επιτρεπόμενο εύρος πλοήγησης του λογισμικού καθώς και το πρώτο URL από το οποίο θα ξεκινήσει να αντλεί πληροφορίες, έτσι ώστε το πρόγραμμα να πλοηγείται αποκλειστικά στην ιστοσελίδα που του έχει οριστεί και όχι σε όλο τον υπόλοιπο ιστό. Προτού ξεκινήσει η διαδικασία εξαγωγής δεδομένων, ορίζονται στην ενσωματωμένη λειτουργία του scrapy συγκεκριμένοι κανόνες που θα πρέπει το λογισμικό (custom_settings). Αυτοί οι

κανόνες κρίνονται απαραίτητοι τόσο για τη βελτίωση της λειτουργίας του λογισμικού αλλά και τη μείωση φόρτου του server στον οποίο πλοηγείται το λογισμικό.

Εφόσον λοιπόν, έχει δημιουργηθεί ένα περιβάλλον που επιτρέπει την ομαλή λειτουργία για το λογισμικό, γίνεται χρήση μίας συνάρτησης ανάλυσης (parsing function). Η συγκεκριμένη συνάρτηση εντοπίζει όλα τα URL των αυτοκινήτων που περιέχει η κάθε καρτέλα και με τη χρήση της εντολής «yield», τα αποθηκεύει στο αρχείο txt.

Τέλος υπάρχει μια συνθήκη «if» η οποία ελέγχει αν ύπαρξη URL της αμέσως επόμενης σελίδας είναι αληθής ή όχι. Αν είναι αληθής, το πρόγραμμα κάνει αίτημα να αποκτήσει πρόσβαση και στην επόμενη σελίδα μέχρι αυτή η συνθήκη να είναι λανθασμένη (False).

3.5.3 Μεθοδολογία και προγραμματισμός εξαγωγή πληροφοριών οχημάτων

Έχοντας συλλέξει όλα τα URL των αυτοκινήτων που θα χρησιμοποιηθούν για την βάση Spritmonitor, ακολουθεί η εξαγωγή των πληροφοριών του κάθε αυτοκινήτου και η αποθήκευσή τους στις 2 ξεχωριστές βάσεις δεδομένων. Αυτό που παρατηρήθηκε στην ιστοσελίδα είναι πως το κάθε αυτοκίνητο στο URL του, περιέχει και τον αποκλειστικό αναγνωριστικό κωδικό του στη βάση δεδομένων του Spritmonitor. Οπότε μαζί με τις πληροφορίες που θα εξαχθούν και θα αποθηκευτούν, θα αποθηκευτούν επίσης και όλοι οι αποκλειστικοί κωδικοί, καθώς θα διευκολύνουν την διαδικασία της αντιστοίχισης επιβεβαιωμένης

Πληροφορίες που εξάγονται

1 Οχημα
 Gasoline, year 2021, 100 kW (136 PS), direct shift gearbox

2 Περιγραφή οχήματος

3 Μέση κατανάλωση καυσίμου
 Consumption: 9.32 l/100km
 CO₂ emission: 217 g/km
 Fuel costs: 17,03 EUR/100km

4 Πίνακας ανεφοδιασμών

Consumption histogram
 Monthly Consumption

1 Εντοπισμός πληροφοριών στην HTML

```

<div id="navi">...</div>
<div id="contentcontainer">
  <script language="Javascript" text="text/javascript">...</script>
  <script language="Javascript" text="text/javascript">...</script>
  <script language="Javascript" text="text/javascript">...</script>
  <div id="content-left">
    <div>
      <div id="vehiclegpic">...</div>
      <div id="vehicledetails">
        <h1>BMW - 2er - 218i GC</h1>
        " Gasoline, year 2021, 100 kW (136 PS), direct shift gearbox"
      </div>
    </div>
  </div>

```

2 Εξαγωγή πληροφοριών με χρήση κώδικα

```

response.css('#vehicledetails h1::text').get()
'BMW - 2er - 218i GC'

```

Εικόνα 4

κατανάλωσης καυσίμου με το αυτοκίνητο που της αντιστοιχεί.

Η σελίδα του κάθε αυτοκινήτου όπως και φαίνεται από την παραπάνω αναπαράσταση, είναι ήδη χωρισμένη σε 2 μέρη (γενικές πληροφορίες και Πίνακα ανεφοδιασμών). Προκειμένου λοιπόν να αποθηκευτούν τα δεδομένα σε 2 ξεχωριστές βάσεις δεδομένων, θα δημιουργηθούν στο πρόγραμμα εξαγωγής δεδομένων 2 ξεχωριστοί αγωγοί δεδομένων (data pipelines), οι οποίοι θα αποθηκεύουν αποκλειστικά τα δεδομένα που τους έχουν οριστεί σε αντίστοιχες βάσεις δεδομένων.

```
import scrapy
from scrapy import Request
import re
from ..items import SpritLinksItem
import csv

class SpritCharSpider(scrapy.Spider):
    rate = 0.5

    custom_settings = {
        'COOKIES_ENABLED': False,
        'CONCURRENT_REQUESTS': 1,
        'DOWNLOAD_DELAY': 1,
        'FEEDS': {'data/%(name)s/%(name)s_%(time)s.csv':{'format': 'csv'}}
    }

    name = 'sprit_second_csv'
    allowed_domains = ['www.spritmonitor.de']
    with open("C:\\Users\\KPA5367\\sprit_link_final_text.txt") as f: #iterate through the URL text file
        start_urls = [url.strip() for url in f.readlines()] #assing all links to a list in a form of string

    def requests(self, response): #The function responsible for making the requests in the website

    def parse(self, response): #definition of a function which parses and saves the data acquired from the approved (by website's host) requests
        #Define Items which will follow the default pipeline
        s_item = SpritLinksItem()
        #####STRING FILTER#####
        def scrape_filter(string,prev_str,next_str): #saves the value between 2 strings
            y = re.search(prev_str + '(.*?)' + next_str,string)
            new_val = y.group(1)
            return new_val
        def string_filter(text): #removes ' ', '\n', '\t' from data
            new_text = text.replace('\t','').replace('\n','').replace(' ','')
            return new_text
        #####
        link = str(response.url)
        if 'page=' not in link: #if cased used in case of the existence of pagination in the table section
            #assign each of the data's location to variables
            ID =str(response.url.replace('https://www.spritmonitor.de/en/detail/', '').replace('.html','')),
            Title = str(response.css('#vehicledetails h1::text').get()),
            Description = str(response.xpath('//*[@id="vehicledetails"]/text()[2]').get())
            Description = string_filter(Description)

            Avg_FC =str(response.css('td strong::text').get())

            Avg_FC = Avg_FC.replace('.', '').replace(', ', '')
            #Matching variables with their equivalent position in the database which they will be saved
            s_item['ID'] = ID
            s_item['Title'] = Title
            s_item['Description'] =Description
            s_item['Avg_FC'] = Avg_FC
            s_item['Title'] = Title
            yield s_item #write all scraped variables to their columns in "Summary" dataset
```

Όπως και στο προηγούμενο κομμάτι ακολουθείται παρόμοια διαδικασία. Πιο συγκεκριμένα ορίζεται μία κλάση με συγκεκριμένα, η οποία ακολουθεί ορισμένους κανόνες. Σε αυτό ωστόσο το πρόγραμμα, τα αιτήματα πρόσβασης και το διάβασμα των τιμών θα γίνουν σε διαφορετικές συναρτήσεις, καθώς στην περίπτωση αυτή το διάβασμα και η αποθήκευση των δεδομένων είναι πιο σύνθετα. Ακόμη το πρόγραμμα θα χρειαστεί να πραγματοποιήσει επιπλέον αιτήματα (εσωτερικά της συνάρτησης), όταν ο πίνακας ανεφοδιασμών (refueling table) έχει παραπάνω από μια σελίδες.

Αρχικά εντοπίζονται στον κώδικα HTML οι τοποθεσίες των δεδομένων που θα αποθηκευτούν στη βάση δεδομένων με τα γενικές πληροφορίες των αυτοκινήτων (Summary Database), τα οποία και αποθηκεύονται στη προεπιλεγμένη τοποθεσία που ορίστηκε στην αρχή. Στα δεδομένα της βάσης δεδομένων «Summary» αποθηκεύονται επιπλέον και οι ξεχωριστοί κωδικοί (IDs) του κάθε αυτοκινήτου, ώστε στη συνέχεια να αντιστοιχηθούν με τα δεδομένα της «Table».

```

for i in range(len(date)): #dates only variable with no nan values so the loop never stops when its not supposed to
#get a list of all the values found in location of each column from the table
try:
    Odometer = response.css('td.fuelkmpos').extract()[i].replace('.', '') or 'na'
except IndexError: #filter nan values to avoid errors
    Odometer = 'na'

Distance = response.css('td.trip').extract()[i] or 'na'
Refuel_Quantity = response.css('td.quantity').extract()[i] or 'na'
Refuel_Type = (response.css('td.fuelsort ::attr(onmouseover)').extract()[i]).replace('(', '').replace(')', '').replace(' ', '') or 'na'
FC = response.css('#content-left > div.itemtable_wrapper > table > tbody > tr > td.consumption').extract()[i] or 'na'
Date = response.css('#content-left > div.itemtable_wrapper > table > tbody > tr > td.fueldate ::text').extract()[i] or 'na'

Odometer = scrape_filter(Odometer, '<td class="fuelkmpos">', '</td>') or 'na'
try:
    Distance = scrape_filter(Distance, '\t\t \n \t\t \t\t \t\t \t\t \t\t \t\t</td>' or 'na'
    Distance = str(Distance).replace('.', '')
except AttributeError:
    Distance = 'na'
Refuel_Quantity = scrape_filter(Refuel_Quantity, '\n \t\t \t\t \t\t \t\t \t\t \t\t</td>' or 'na'
Refuel_Quantity = str(Refuel_Quantity).replace('.', '')
Refuel_Type = scrape_filter(Refuel_Type, 'showTooltip(', ')',
try:
    FC = scrape_filter(FC, '<td class="consumption" onmouseover="showTooltip", ' 1/100km')
    FC = str(FC).replace('(', '').replace(')', '').replace(' ', '')
except AttributeError:
    FC = 'na'

#create a dictionary with all the lists created from the scraped page
table = { 'ID' : id_,
          'Date' : Date,
          'Odometer' : Odometer,
          'Distance' : Distance,
          'Refuel_Quantity' : Refuel_Quantity,
          'Refuel_Type' : Refuel_Type,
          'Fuel_Consumption' : FC
        }
writer.writerow(table) #write each row of the table in the id specific

#check for pagination inside the table
next_page = response.xpath('//*[@id="content-left"]/div/a[last()]/@href').extract() #location of "next page" button in refuelling table

if next_page is not None: #if a string is found yield a request before proceeding to next car
    next_url = ''.join(map(str, next_page))
    next_url = 'https://www.spritmonitor.de' + next_url

```

Εικόνα 6

Αφού ολοκληρωθεί αυτή η διεργασία, δημιουργούνται σε ξεχωριστή τοποθεσία αρχεία csv, όπου το καθένα έχει για όνομα τον ξεχωριστό κωδικό του αυτοκινήτου που του αντιστοιχεί.

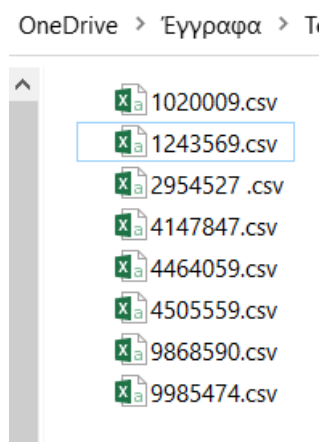
Υστερα το πρόγραμμα

- εντοπίζει την τοποθεσία όλων των ημερομηνιών που βρίσκονται στον πίνακα ανεφοδιασμών και τις αποθηκεύει σε μια λίστα.

- Με τη χρήση μιας επανάληψης «for» , με εύρος ίσο με τον αριθμό των ημερομηνιών στον πίνακα, καταγράφονται και αποθηκεύονται τα δεδομένα της κάθε σειράς του πίνακα (ένδειξη κοντέρ, χιλιόμετρα ταξιδιών, είδος καυσίμου, κατανάλωση καυσίμου...).
- Όταν ολοκληρωθεί η επανάληψη καταγραφής δεδομένων του πίνακα, ελέγχεται αν υπάρχει επόμενη σελίδα με δεδομένα στον πίνακα, όπου σε περίπτωση που υπάρχει πραγματοποιείται καινούριο αίτημα για την εξαγωγή των δεδομένων της επόμενης σελίδας.
- Με το που εκτελεστούν οι προηγούμενες εντολές, το πρόγραμμα συνεχίζει στην εξαγωγή δεδομένων του επόμενου αυτοκινήτου.

Αφού ολοκληρωθεί και η εκτέλεση του 2^{ου} προγράμματος, έχουν διαμορφωθεί 2 βάσεις δεδομένων «Summary» και «Table», οι οποίες έχουν τις εξής μορφές:

Table Database



Summary Database

Description	ID	Title
Diesel, year 2020, 110 kW (150 PS), direct shi...	1148393	Volkswagen - Passat - Elegance
Hybrid Gasoline, year 2020, 192 kW (261 PS), a...	1183059	Volvo - XC40 - recharge
Hybrid Gasoline, year 2020, 90 kW (122 PS), au...	1159167	Toyota - Corolla Hybrid - 125H
Gasoline, year 2020, 74 kW (101 PS), direct sh...	1256442	Hyundai - i20 - 1.0 T-GDI DCT
Hybrid Gasoline, year 2022, 90 kW (122 PS), co...	1370338	Toyota - Corolla Hybrid - Trek

Εικόνα 7

3.5.4 A.D.A.C. web scraping

Η βάση δεδομένων του «adac.de» αποτελείται από μια πολυστρωματική μορφή. Προκειμένου να δημιουργηθεί μια δομημένη και αποτελεσματική μεθοδολογία που μπορεί να ακολουθηθεί από το πρόγραμμα, η διαδικασία εξαγωγής δεδομένων χωρίζεται σε 6 βήματα:

- Ορισμός παραμέτρων για να προσδιορισθεί εάν μια διεύθυνση URL αξίζει να εξαχθεί ή όχι.
- Εξαγωγή και αποθήκευση URL όλων των κατασκευαστών αυτοκινήτων που απαιτούνται για τη μελέτη των ανταγωνιστών.
- Εξαγωγή URL από όλα τα σχετικά μοντέλα μέσα στις διευθύνσεις URL κατασκευαστών που αποθηκεύτηκαν προηγουμένως.

- Εξαγωγή διευθύνσεων URL από τις σχετικές γενιές μοντέλων.
- Εξαγωγή διευθύνσεων URL των σχετικών εκδόσεων από όλες τις σχετικές γενιές μοντέλων
- Αποθήκευση των απαραίτητων δεδομένων από κάθε έκδοση.

Ορισμός Παραμέτρων:

Η βάση δεδομένων του ADAC αποτελείται από μια τεράστια πληθώρα αυτοκινήτων, πολλά από τα οποία όμως συναντώνται σπάνια ή και ποτέ στους ευρωπαϊκούς δρόμους, ή –εξαιτίας των χαρακτηριστικών τους - δεν συνεισφέρουν ουσιαστικά στην έρευνα. Προκειμένου λοιπόν το πρόγραμμα να λειτουργεί πιο αποτελεσματικά, καθορίστηκαν ορισμένες παράμετροι για να αποφασιστεί εάν το υποκείμενο στρώμα αξίζει να αποθηκευτεί ή όχι:

- Κατασκευαστής οχημάτων: εξακολουθεί να δραστηριοποιείται στην ΕΕ; Υπάρχει μεγάλος πληθυσμός αυτοκινήτων στη βάση δεδομένων spritmonitor;
- Μοντέλο οχήματος: πωλείται το μοντέλο στην ΕΕ (Υπάρχουν καταγραφές πωλήσεων στο ACEA) ;
- Γενιά μοντέλου οχήματος: γενιά που πωλείται το 2020 έως το 2023;
- Έκδοση μοντέλου οχήματος: Καθώς η συγκεκριμένη έρευνα αφορά αυτοκίνητα που κινούνται με βενζίνη ή πετρέλαιο, τα μοντέλα που κινούνται με: ρεύμα, υγραέριο, φυσικό αέριο και υδρογόνο εξαιρούνται.

Μεθοδολογία εξαγωγής δεδομένων από το ADAC



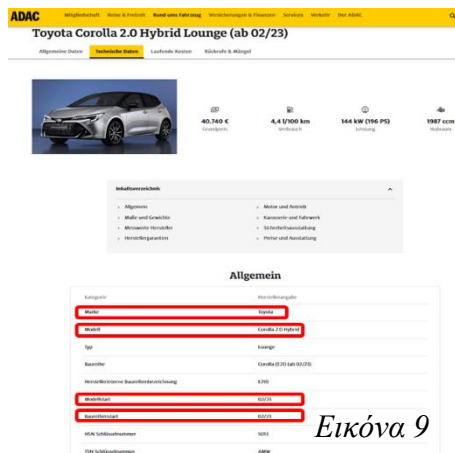
Εικόνα 8

Μόλις ολοκληρωθεί η παραπάνω διαδικασία εξαγωγής URL, όλα τα URL των μοντέλων που είναι αναγκαία για τον εμπλουτισμό της βάσης του Spritmonitor, θα είναι αποθηκευμένα σε ένα αρχείο txt μέσα από το οποίο θα εξαχθούν τα παρακάτω χαρακτηριστικά:

Οργάνωση και «καθαρισμός» δεδομένων

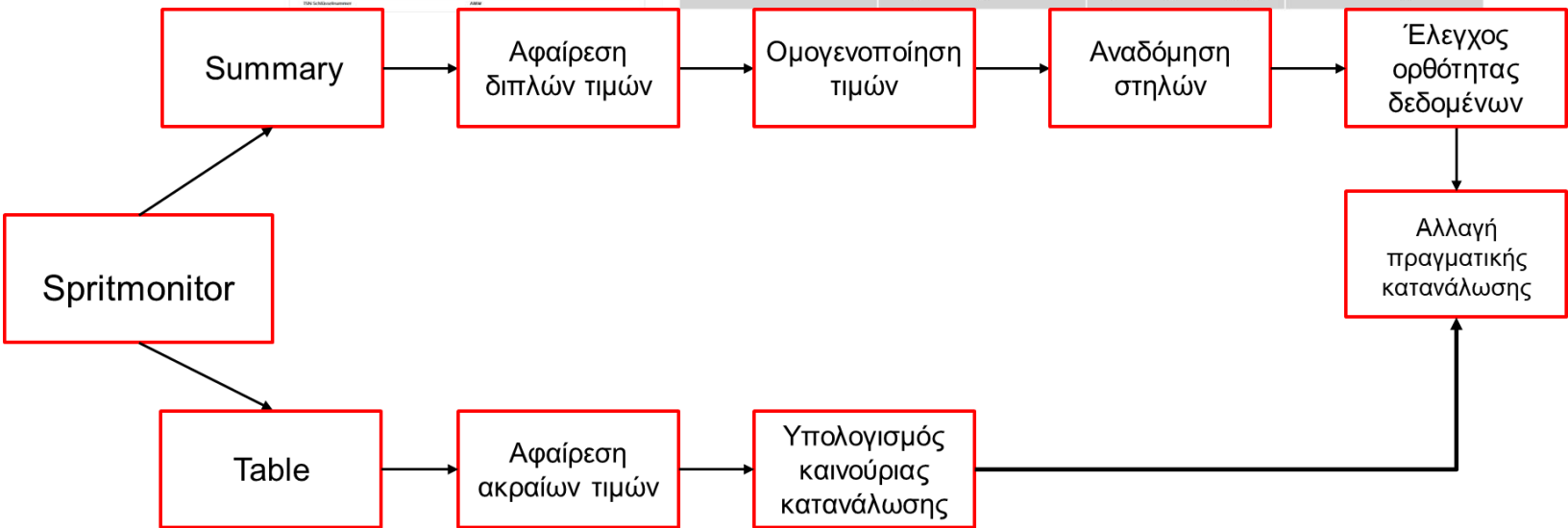
Το πρώτο βήμα για την κατασκευή της τελικής βάσης δεδομένων, αφορά την μετατροπή της βάσης δεδομένων του Spritmonitor σε μια οργανωμένη και ομογενή βάση δεδομένων, με μορφή τέτοια που θα επιτρέπει τον εύκολο εμπλουτισμό με τη βάση του ADAC.

Η οργάνωση και ο καθαρισμός της βάσης Spritmonitor, πραγματοποιείται με δυο παράλληλες διαδικασίες – 1 για την οργάνωση του Summary και 1 για του Table – οι οποίες εμφανίζονται στο παρακάτω διάγραμμα



Εικόνα 9

Brand	Period of production	Rated engine power	WLTP_FC
Model	Engine type	System power [kW]	Weight
Variant name	Engine displacement	System power [hp]	Emission regulation



Πίνακας 4

- Αφαίρεση των διπλών καταχωρήσεων οχημάτων:. Δεδομένου ότι κάθε όχημα στο Spritmonitor έχει ένα μοναδικό αναγνωριστικό κωδικό (ID), είναι απαραίτητο να αφαιρεθούν όσα αυτοκίνητα υπάρχουν πάνω από μια φορές.

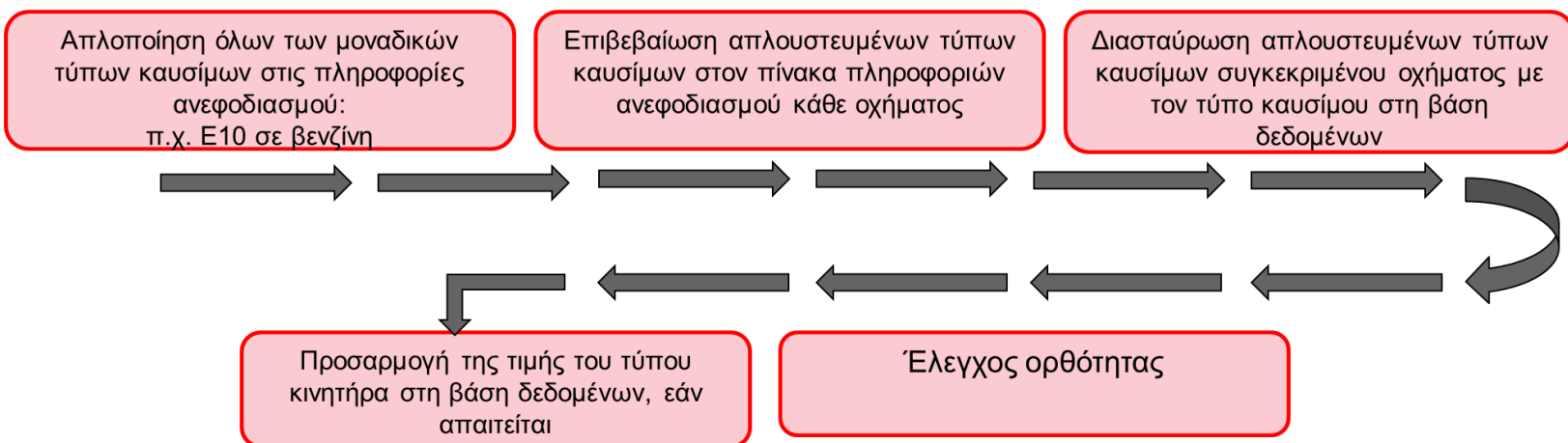
- Ομογενοποίηση δεδομένων: Κατά την εξαγωγή των δεδομένων παρατηρείται πως η μορφή των δεδομένων δεν είναι ομογενής. Υπάρχουν αρκετοί χαρακτήρες, οι οποίοι έχουν παραμείνει από την διαδικασία εξαγωγής δεδομένων και είναι αναγκαίο να αφαιρεθούν. Μετά από διεξοδική μελέτη παρατηρήθηκε ότι όλοι οι ανεπιθύμητοι ακολουθούν πολλά και διαφορετικά μοτίβα. Αφού ανιχνεύθηκαν όλα τα μοτίβα αφαιρέθηκαν όλοι οι περιττοί χαρακτήρες.
- Αναδόμηση στηλών: Η «ωμή» μορφή των πληροφοριών που υπάρχουν στα δεδομένα, έχουν μια συγκεκριμένη δομή. Πιο συγκεκριμένα, η στήλη που περιέχει τον τίτλο του αυτοκινήτου αποτελείται από το όνομα του κατασκευαστή (brand), το όνομα της σειράς (model) και τέλος η παραλλαγή της σειράς (variant). Ταυτόχρονα, το δεύτερο μέρος των πληροφοριών του αυτοκινήτου περιέχει με αντίστοιχη δομή:
 1. Το είδος του καυσίμου (fuel type)
 2. Έτος παραγωγής (year)
 3. Ισχύς σε kW
 4. Είδος μετάδοσης ταχυτήτων (transmission)

Έτσι λοιπόν, προκειμένου να υπάρξει μια πιο σωστά δομημένη βάση δεδομένων, κρίνεται απαραίτητο, οι δύο ομάδες πληροφοριών να διαχωριστούν.

Description	ID	Title	Brand	Model	Variant	Extra	Ftype	year	kw(PS)	transmission
Diesel, year 2020, 110 kW (150 PS), direct shi...	1148393	Volkswagen - Passat - Elegance	Volkswagen	Passat	Elegance	None	Diesel	year 2020	110 kW (150 PS)	direct shift gearbox
Hybrid Gasoline, year 2020, 192 kW (261 PS), a...	1183059	Volvo - XC40 - recharge	Volvo	XC40	recharge	None	Hybrid Gasoline	year 2020	192 kW (261 PS)	automatic
Hybrid Gasoline, year 2020, 90 kW (122 PS), au...	1159167	Toyota - Corolla Hybrid - 125H	Toyota	Corolla Hybrid	125H	None	Hybrid Gasoline	year 2020	90 kW (122 PS)	automatic
Gasoline, year 2020, 74 kW (101 PS), direct sh...	1256442	Hyundai - i20 - 1.0 T-GDI DCT	Hyundai	i20	1.0 T	GDI DCT	Gasoline	year 2020	74 kW (101 PS)	direct shift gearbox
Hybrid Gasoline, year 2022, 90 kW (122 PS), co...	1370338	Toyota - Corolla Hybrid - Trek	Toyota	Corolla Hybrid	Trek	None	Hybrid Gasoline	year 2022	90 kW (122 PS)	continuously variable

Εικόνα 10

- Έλεγχος ορθότητας δεδομένων: Κατά την μελέτη των δεδομένων στην ιστοσελίδα παρατηρήθηκε ότι ορισμένα αυτοκίνητα δεν είχαν σωστές πληροφορίες. Συγκεκριμένα παρατηρήθηκε πως αρκετοί χρήστες μπορεί να είχαν εισάγει είδος κινητήρα το οποίο δεν υπάρχει για το συγκεκριμένο μοντέλο (π.χ. Brand: Volkswagen | Model: ID.3| Engine type: Gasoline - αδύνατο καθώς το συγκεκριμένο μοντέλο έχει μόνο ηλεκτρικούς κινητήρες -). Για αυτό το λόγο, δημιουργήθηκε μια μεθοδολογία η οποία θα χρησιμοποιηθεί η βάση δεδομένων με τους ανεφοδιασμούς (Table database) για να ελεγχθεί αν ταιριάζουν τα είδη του καυσίμου.



3.5.5 Καθαρισμός βάσης δεδομένων ανεφοδιασμού (Table)

- Αφαίρεση ακραίων τιμών:

1. Σε ορισμένες περιπτώσεις, ο χρήστης έκανε μια προσωρινή εισαγωγή που στη συνέχεια διορθώθηκε, με την προσθήκη μιας καινούριας. Καθώς αυτά μπορούν να παραμορφώσουν τον υπολογισμό της μέσης κατανάλωσης καυσίμου, πρέπει να αφαιρεθούν.

01.05.2022	26.379	5.479,0	216,00	95	☔	🚗	🚗	🚗	333,07	3,94	AC
10.12.2021	20.900	1.368,0	37,00	95	☔	🚗	🚗	🚗	44,65	2,70	AC
10.12.2021		1.400,0	37,00	95	☔	🚗	🚗	🚗	44,65	2,64	AC
01.12.2021	19.532	1.633,0	146,11	95	☔	🚗	🚗	🚗	187,95	8,95	AC

Εικόνα 11

Ίδια ημερομηνία και δεν εισήχθη απόσταση στο κοντέρ → Ανάγκη για αφαίρεση

2. Σε ορισμένες περιπτώσεις, οι χρήστες ξεχνούν να συμπληρώσουν 1 ανεφοδιασμό, με αποτέλεσμα μια ασυνήθιστα χαμηλή κατανάλωση καυσίμου σε σύγκριση με τη μέση κατανάλωση καυσίμου. Τέτοιες περιπτώσεις οφείλουν να εντοπισθούν και να αφαιρεθούν.

29.05.2022	50.369	491,0	36,40	95	☀️	🚗	🚗	🚗	78,95	7,41
24.05.2022	49.878	446,0	35,06	95	☀️	🚗	🚗	🚗	72,89	7,86
10.05.2022	49.432	374,0	28,18	95	☀️	🚗	🚗	🚗	57,18	7,53
03.05.2022	49.058	755,0	34,88	95	☀️	🚗	🚗	🚗	69,73	4,62
26.04.2022	48.303	444,0	36,70	95	☀️	🚗	🚗	🚗	70,42	8,27

Εικόνα 12

Αδύνατο με 30 λίτρα να έχει διανύσει ο οδηγός 700 χιλιόμετρα.

Αφού έχουν φιλτραριστεί (στον βαθμό που είναι εφικτό) όλες οι καταγραφές κατανάλωσης καυσίμου, επανυπολογίζεται ο μέσος όρος της κατανάλωσης καυσίμου και αντικαθιστά την ήδη υπάρχουσα τιμή.

3.5.6 Εμπλουτισμός βάσης δεδομένων (Spritmonitor)

Μετά την οργάνωση και τον καθαρισμό της βάσης δεδομένων του Spritmonitor, ακολουθεί η προσθήκη της κατανάλωσης κύκλου WLTP.

Προ επεξεργασία δεδομένων:

Προτού εφαρμοστεί η μεθοδολογία εμπλουτισμού δεδομένων, κρίνεται απαραίτητο να πραγματοποιηθεί μια σύμπτυξη των δεδομένων της βάσης δεδομένων του adac. Πιο συγκεκριμένα η βάση δεδομένων του adac περιέχει αρκετά πιο αναλυτικά τις διαφορετικές εκδόσεις του κάθε αυτοκινήτου σε σχέση με τις πληροφορίες που παρέχει η βάση δεδομένων του Spritmonitor. Τούτο σημαίνει ότι υπάρχουν κάποιες πολύ μικρές αλλαγές μεταξύ ίδιων μοντέλων, οι οποίες όμως δεν επηρεάζουν την κατανάλωση κύκλου WLTP, ή τον επηρεάζουν σε τόσο μικρό βαθμό που κρίνεται αμελητέος. Για το λόγο αυτό το λόγο, τα αυτοκίνητα με κοινά χαρακτηριστικά και ίδιες ή σχεδόν ίδιες τιμές κατανάλωσης ενώνονται σύμφωνα με το παρακάτω σχεδιάγραμμα

Brand	Series	Model	Enginetype	Displacement	Systempowerkw	RatedpwrKW	Systempowerps	WLTP_FC	Weight
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.0	Otto	998 ccm	53 53 kW	72	4,6 l/100 km	1055 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.0	Otto	998 ccm	53 53 kW	72	4,7 l/100 km	1055 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.0	Otto	998 ccm	53 53 kW	72	4,6 l/100 km	1055 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5	Otto	1497 ccm	82 82 kW	112	5,1 l/100 km	1115 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5	Otto	1497 ccm	82 82 kW	112	4,9 l/100 km	1140 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5	Otto	1497 ccm	82 82 kW	112	5,1 l/100 km	1115 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5	Otto	1497 ccm	82 82 kW	112	5,1 l/100 km	1115 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5	Otto	1497 ccm	82 82 kW	112	4,9 l/100 km	1140 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5	Otto	1497 ccm	82 82 kW	112	5,1 l/100 km	1115 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5	Otto	1497 ccm	82 82 kW	112	4,9 l/100 km	1140 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5	Otto	1497 ccm	82 82 kW	112	5,1 l/100 km	1115 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5	Otto	1497 ccm	82 82 kW	112	5,1 l/100 km	1115 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5 Hybrid	Voll-Hybrid	1497 ccm	74 54 kW	100	3,7 l/100 km	1055 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5 Hybrid	Voll-Hybrid	1497 ccm	74 54 kW	100	3,7 l/100 km	1055 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5 Hybrid	Voll-Hybrid	1497 ccm	74 54 kW	100	3,7 l/100 km	1055 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5 Hybrid	Voll-Hybrid	1497 ccm	74 54 kW	100	3,7 l/100 km	1055 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5 Hybrid	Voll-Hybrid	1497 ccm	74 54 kW	100	3,7 l/100 km	1055 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5 Hybrid	Voll-Hybrid	1497 ccm	74 54 kW	100	3,7 l/100 km	1055 kg	
Toyota	Yaris (XP13) (04/17 - 08/20)	Yaris 1.5 Hybrid	Voll-Hybrid	1497 ccm	74 54 kW	100	3,7 l/100 km	1055 kg	

Model	Brand	Variant	Enginetype	start_date	end_date	Systempowerkw	Systempowerps	WLTP_FC	Weight
Yaris 1.0	Toyota	Yaris	Gasoline	2017	2020	53	72	4.633333333	1055
Yaris 1.5	Toyota	Yaris	Gasoline	2017	2020	82	112	5.033333333	1123.333333
Yaris 1.0	Toyota	Yaris	Gasoline	2020	present	53	71	5.775	960
Yaris 1.5	Toyota	Yaris	Gasoline	2020	present	92	125	4.905263158	1076.842105
Yaris 1.5 Hybrid	Toyota	Yaris	Hybrid Gasoline	2017	2020	74	100	3.7	1055
Yaris 1.5 Hybrid	Toyota	Yaris	Hybrid Gasoline	2020	present	85	116	4.425	1090

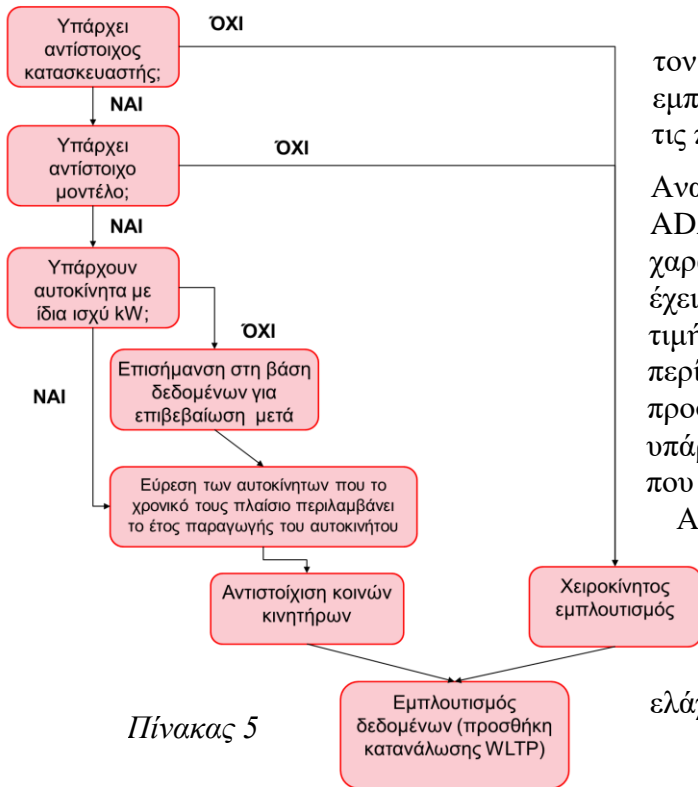
Εικόνα 13

Μεθοδολογία εμπλουτισμού:

Βασικός στόχος του εμπλουτισμού της βάσης δεδομένων του Spritmonitor, αποτελεί την αντιστοίχιση των χαρακτηριστικών των αυτοκινήτων της και της βάσης εμπλουτισμού (ADAC). Με τη χρήση της γλώσσας Python και των βιβλιοθηκών Pandas και Numpry, προγραμματίστηκε ένα λογισμικό που πραγματοποιεί την εν λόγω αντιστοίχιση. Αναλυτικότερα, το παρόν λογισμικό διασταυρώνει τις πληροφορίες μεταξύ των δύο βάσεων δεδομένων, και αντιστοιχίζει τα αυτοκίνητα σύμφωνα με τα παρακάτω χαρακτηριστικά:

1. Κατασκευαστής
2. Μοντέλο

3. Είδος κινητήρα (Gasoline, Diesel, Hybrid Gasoline, Plug-in Hybrid Gasoline)
4. Έτος παραγωγής
5. Ισχύς κινητήρα σε kW



Πίνακας 5

Το διπλανό διάγραμμα περιγράφει τον αλγόριθμο που χρησιμοποιείται για τον εμπλουτισμό της βάσης δεδομένων του Spritmonitor με τις πληροφορίες του ADAC.

Αναλυτικότερα, το πρόγραμμα αναζητά στη βάση του ADAC κατασκευαστή και μοντέλο, καθώς αυτά τα χαρακτηριστικά είναι σχεδόν σίγουρο ότι ο χρήστης θα έχει καταγράψει σωστά. Στη συνέχεια ελέγχεται αν η τιμή της ισχύος σε kW εισάχθηκε σωστά ή όχι. Σε περίπτωση που δεν εισάχθηκε σωστά το πρόγραμμα προσπαθεί να προσεγγίσει την κοντινότερη τιμή που υπάρχει στο ADAC σε διάστημα $\pm 3kW$ (η απόκλιση που παρατηρήθηκε από την πραγματική τιμή ισχύος).

Αν και πάλι δεν υπάρχει κάποια αντιστοίχιση, ελέγχεται αργότερα η τιμή της ισχύος χειροκίνητα. Μετά από αυτό το βήμα, είναι σίγουρο ότι θα υπάρχει αντιστοίχιση τόσο στην χρονιά παραγωγής όσο και στο είδος του κινητήρα (με ελάχιστες εξαιρέσεις που

ούτως ή άλλως δεν παρείχαν ακριβείς πληροφορίες εξ' αρχής).

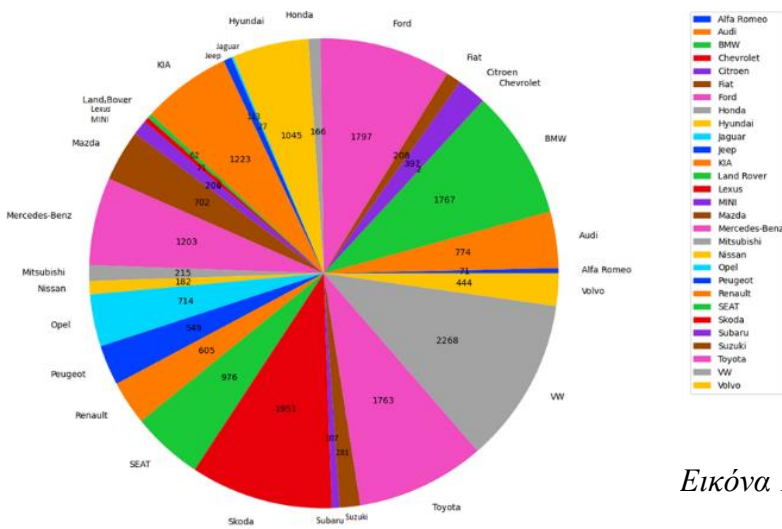
Τέλος αφού το πρόγραμμα έχει καταλήξει σε ένα μοναδικό αυτοκίνητο, προστίθεται η τιμή του WLTP κύκλου στη βάση δεδομένων. Όσα αυτοκίνητα δεν υπήρχαν στην βάση δεδομένων του ADAC (λιγότερα από 100), μελετήθηκαν και εμπλουτίστηκαν χειροκίνητα.

Συνολικά, περίπου 28000 μοναδικά βενζινοκίνητα και πετρελαιοκίνητα οχήματα μπόρεσαν να εξαχθούν από το Spritmonitor με ημερομηνία παραγωγής μεταξύ 2020 και 2023.

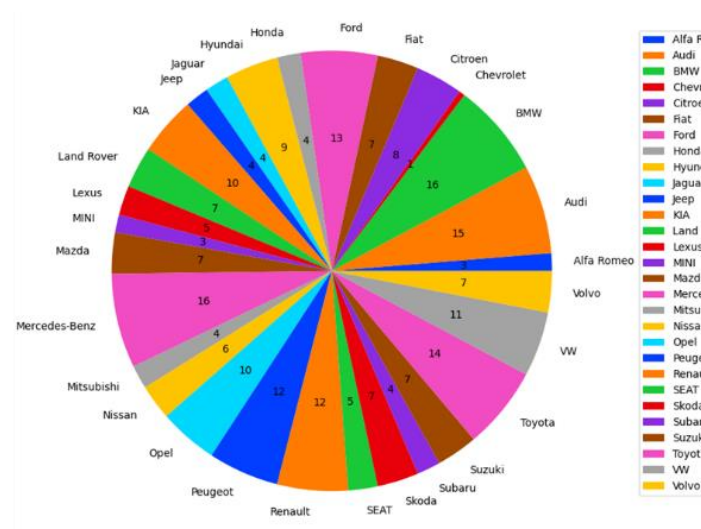
Μετά τον εμπλουτισμό, η τελική βάση δεδομένων περιέχει περίπου 20000 οχήματα, αποτελούμενα από 29 μάρκες και 231 μοντέλα:

3.5.7 Οπτικοποίηση Δεδομένων

Πληθυσμός αυτοκινήτων ανά κατασκευαστή



Πληθυσμός μοντέλων ανά κατασκευαστή



Εικόνα 14

Μελέτη ποσοστιαίας απόκλισης

Κατηγοριοποίηση αυτοκινήτων:

Εφόσον, έχει δημιουργηθεί μια οργανωμένη βάση δεδομένων, ακολουθεί η ανάλυση και η μελέτη της ποσοστιαίας απόκλισης μεταξύ κατανάλωσης WLTP κύκλου και πραγματικής. Το πρώτο βήμα της μελέτης αυτής, σχετίζεται με τον διαχωρισμό των αυτοκινήτων. Τούτο συμβαίνει καθώς, δεν δύναται να μελετηθούν και να συγκριθούν όλα τα αυτοκίνητα μεταξύ τους. Αυτό οφείλεται στην ύπαρξη μεγάλης διαφοροποίησης σε χαρακτηριστικά όπως βάρος, ιπποδύναμη, μέγεθος κινητήρα στην αγορά αυτοκινήτων. Έτσι καθίσταται αναγκαίο να χωριστούν τα αυτοκίνητα της βάσης δεδομένων σε διαφορετικές κατηγορίες προκειμένου να μελετηθούν και να συγκριθούν όμοια δεδομένα.

Ο διαχωρισμός αυτός πραγματοποιείται σύμφωνα με τα ευρωπαϊκά πρότυπα κατηγοριοποίησης (REGULATION (EEC) No.4064/89 MERGER PROCEDURE). Οι κατηγορίες (segments) στις οποίες θα χωριστούν τα αυτοκίνητα αποτελούν:

- Segment A (αυτοκίνητα πόλης)
- Segment B (μικρά αυτοκίνητα)
- Segment B-SUV (μικρού μεγέθους αυτοκίνητα τύπου SUV)
- Segment C (μεσαίου μεγέθους αυτοκίνητα)
- Segment C-SUV (μεσαίου μεγέθους αυτοκίνητα τύπου SUV)
- Segment D-SUV (μεγάλα αυτοκίνητα τύπου SUV)

Οι παραπάνω κατηγορίες καλύπτουν τόσο το μεγαλύτερο μέρος της αγοράς όσο και το μεγαλύτερο μέρος της βάσης δεδομένων (πάνω από 95% των αυτοκινήτων).

Αφού, διαχωριστούν σε αντίστοιχες ομάδες, θα μελετηθούν τα αυτοκίνητα που κατέχουν το μεγαλύτερο μερίδιο της σύγχρονης ευρωπαϊκής αγοράς. Μέσω αυτής της προσέγγισης, επιλέχθηκε να εστιάσουμε στα αυτοκίνητα με το μεγαλύτερο μερίδιο αγοράς στην Ευρώπη, προκειμένου να μελετηθεί η πραγματική κατανάλωση καυσίμου. Μέσω αυτής της επιλογής, περιλαμβάνεται ένας μεγαλύτερος αριθμός οδηγών που κυκλοφορούν στους ευρωπαϊκούς δρόμους. Έτσι, η πραγματική κατανάλωση καυσίμου μπορεί να μελετηθεί με μεγαλύτερη ακρίβεια και αντιπροσωπευτικότητα.

Η επιλογή αυτή επιτρέπει την απόκτηση μιας πληρέστερης εικόνας για την κατανάλωση καυσίμου στον πραγματικό κόσμο, υπερβαίνοντας τους εργαστηριακούς κύκλους δοκιμών. Με την ενσωμάτωση δεδομένων από το μεγαλύτερο μερίδιο αγοράς, μπορεί να αναδειχθούν οι πραγματικές τάσεις στην κατανάλωση καυσίμου.

Αφού ολοκληρωθεί η επιλογή και ο διαχωρισμός, τα δεδομένα ομαδοποιούνται κατά αυτοκίνητο και μοντέλο (brand, model).

Σύμφωνα λοιπόν με τις πληροφορίες που αντλήθηκαν από το ACEA τα αυτοκίνητα που κατείχαν, τόσο ένα σημαντικό μερίδιο αγοράς στην κατηγορία τους όσο και ικανοποιητικό πληθυσμό έτσι ώστε να μπορεί η μελέτη τους να οδηγηθεί στην εξαγωγή συμπερασμάτων.

Πιο συγκεκριμένα τα αυτοκίνητα που μελετήθηκαν αποτελούν:

- Segment A: Toyota Aygo, Hyundai I10, Fiat 500, Fiat Panda
- Segment B: Peugeot 208, Toyota Yaris, Opel Corsa, Citroen C3
- Segment B-SUV : Peugeot 2008, Ford Puma, Toyota Yaris Cross, Renault Captur, Volkswagen T-Cross
- Segment C: Volkswagen Golf, Toyota Corolla, Peugeot 308, Ford Focus, KIA Ceed
- Segment C-SUV: Volkswagen T-Roc, Nissan Qashqai, Peugeot 3008, Toyota, C-HR, Kia Niro, Renault Arkana,
- Segment D-SUV: Hyundai Tucson, Volkswagen Tiguan, Kia Sportage, Ford Kuga, Toyota RAV4, Skoda Kodiaq.

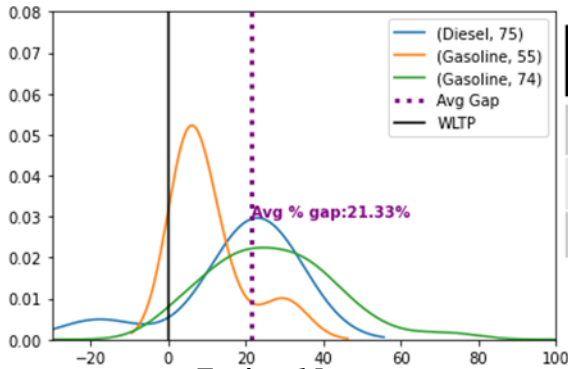
Αφού λοιπόν, επιλέχθηκαν τα αυτοκίνητα προς μελέτη, υπολογίζεται η τιμή της μέσης ποσοστιαίας απόκλισης τους σύμφωνα με τον παρακάτω τύπο:

$$\delta = \frac{|Πραγματική\ κατανάλωση - WLTP\ κατανάλωση|}{WLTP\ κατανάλωση} * 100\%$$

Μελέτη ποσοστιαίας απόκλισης:

Μετά το διαχωρισμό των αυτοκινήτων και τον υπολογισμό της ποσοστιαίας απόκλισης μεταξύ «πραγματικής» κατανάλωσης καυσίμου και κύκλου WLTP, δημιουργήθηκαν με τη χρήση της βιβλιοθήκης matplotlib, διαγράμματα κατανομής τα οποία περιγράφουν τις τάσεις ποσοστιαίας απόκλισης για κάθε μοντέλο ξεχωριστά. Επιπλέον, στα διαγράμματα κατανομής των μοντέλων δημιουργούνται ξεχωριστές γραμμές κατανομής για κάθε διαφορετικό είδος κινητήρα. Καθίσταται προφανές πως ο κάθε διαφορετικός κινητήρας ενδέχεται να

Peugeot 208



Εικόνα 15

Κινητήρας	Μέση Απόκλιση	5ο εκατοστημόριο	95ο εκατοστημόριο	Διάμεσος της Απόκλισης
Gasoline 55kW	10,42	4,50	29,19	5,30
Gasoline 74kW	26,78	3,38	50,36	26,36
Diesel 75 kW	17,87	-6,28	30,01	21,42

Πίνακας 6

έχει και διαφορετικές τάσεις ποσοστιαίας απόκλισης. Με στόχο λοιπόν την εξαγωγή, στατιστικά ορθών αποτελεσμάτων όλοι οι διαφορετικοί κινητήρες μελετήθηκαν ανάλογα με το είδος του κινητήρα τους (βενζινοκινητήρας, πετρελαίου, υβριδικός, plug-in υβριδικός) και την ισχύ του σε kw (kilowatt).

Το παρακάτω παράδειγμα απεικονίζει την μέθοδο της μελέτης της ποσοστιαίας απόκλισης για το μοντέλο Peugeot 208. Όπως και φαίνεται, έχουν δημιουργηθεί 3 διαφορετικές κατανομές για κάθε διαφορετικό κινητήρα και ύστερα καταγράφηκαν όλες οι μετρήσεις που είναι απαραίτητες για την εξαγωγή συμπερασμάτων όσο αφορά την τάση της ποσοστιαίας απόκλισης της κατανάλωσης καυσίμου.

Κατηγορία	Min	Max
A	4.9%	17.1%
B	12.4%	21.3%
B-SUV	4.7%	30.5%
C	12.3%	27.7%
C-SUV	4.5%	31.7%
D-SUV	4.6%	10.5%

Πίνακας 7

Αποτελέσματα Έρευνας

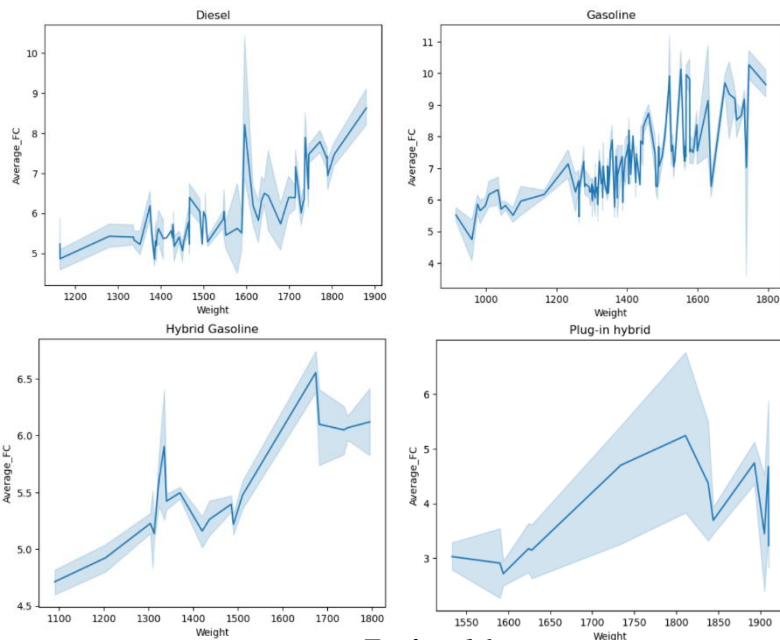
Μετά την διερεύνηση όλων των αυτοκινήτων υπολογίστηκε το εύρος ποσοστιαίας απόκλισης για κάθε κατηγορία αυτοκινήτων, που απεικονίζεται από τον παρακάτω πίνακα:

Ο διπλάνος πίνακας περιλαμβάνει τη μέση τιμή της ελάχιστης και της μέγιστης ποσοστιαίας απόκλισης όλων των αυτοκινήτων με εξαίρεση τα plug-in υβριδικά (PHEVs).

Παρατηρήσεις:

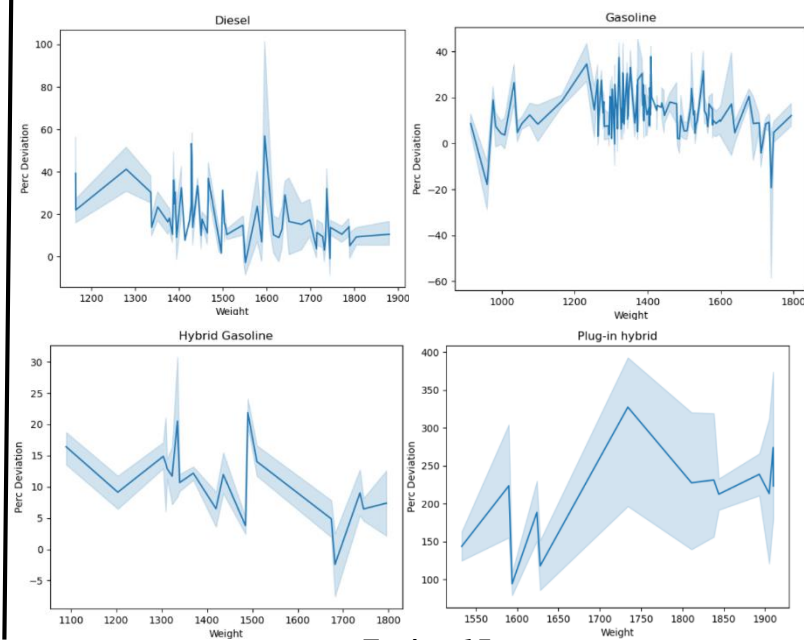
- Οι κατηγορίες αυτοκινήτου τύπου SUV φαίνεται να έχουν μεγαλύτερο εύρος τιμών ποσοστιαίας απόκλισης σε σύγκριση με τις υπόλοιπες.
- Όταν η ισχύς του οχήματος αυξάνεται, το μέσο και το μέγιστο χάσμα μεταξύ της πραγματικής κατανάλωσης καυσίμου και της αναγνωρισμένης τιμής μετατοπίζεται προς μια υψηλότερη τιμή. Αυτό δίνει μια ένδειξη για ένα πιο σπορ / επιθετικό στυλ οδήγησης από αυτό του κύκλου WLTP. Αναλυτικότερα, κατά διαδικασία δοκιμών στον κύκλο WLTP, η κατανάλωση των αυτοκινήτων μετριέται σε περιβάλλον, κάτω από πολύ συγκεκριμένες συνθήκες. Ωστόσο, η υψηλότερη ισχύ κινητήρα δίνει την δυνατότητα στον οδηγό να αναπτύξει υψηλότερες ταχύτητες, από αυτές του κύκλου WLTP, γεγονός όπου σύμφωνα με πρόσφατη έρευνα (Al-Shaebi, 2017), οδηγούν σε αυξημένη κατανάλωση, άρα και σε υψηλότερη ποσοστιαία απόκλιση.
- Ενώ η μέση πραγματική κατανάλωση καυσίμου τείνει να αυξάνεται με το βάρος του αυτοκινήτου, η ποσοστιαία απόκλιση δεν παρουσιάζει την ίδια τάση.

Διαγράμματα Μάζας- Κατανάλωσης



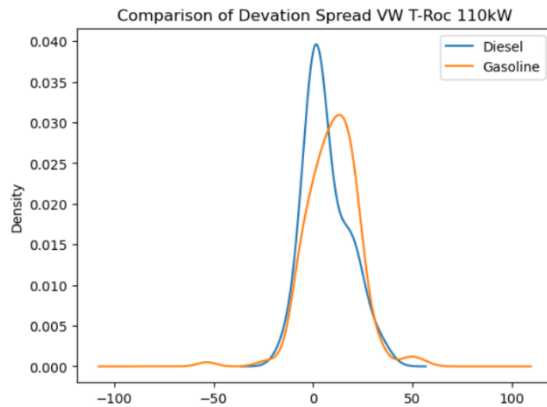
Εικόνα 16

Διαγράμματα Μάζας- Ποσοστιαίας Απόκλισης



Εικόνα 17

- Οι κινητήρες ντίζελ τις περισσότερες φορές έχουν στενότερη εξάπλωση ποσοστιαίας απόκλισης σε σύγκριση με αυτούς της βενζίνης. Αυτό μπορεί να εξηγηθεί από την υψηλότερη ενεργειακή πυκνότητα του ντίζελ σε σύγκριση με το καύσιμο βενζίνης. Ο ισχυρισμός αυτός φαίνεται και από το παρακάτω παράδειγμα, στο οποίο μελετάται η ποσοστιαία απόκλιση για το VW T-Roc. Παρόλο που και οι δυο κινητήρες έχουν ισχύ 110 kW και σχεδόν ίση μέση ποσοστιαία απόκλιση, ο κινητήρας ντίζελ φαίνεται να έχει στενότερη εξάπλωση.



Εικόνα 18

- Οι υβριδικοί κινητήρες έχουν αρκετά παρόμοια συμπεριφορά κατά μέσο όρο με αυτή των κινητήρων βενζίνης. Ωστόσο, δείχνουν μεγαλύτερη εξάπλωση καθώς η κατανάλωση καυσίμου είναι πιο ευαίσθητη στο στυλ οδήγησης, εξαιτίας της μικρότερης μέσης κατανάλωσης. Με άλλα λόγια, όταν ένα αυτοκίνητο έχει χαμηλότερη τιμή κατανάλωσης από ένα άλλο, η ποσοστιαία μεταβολή της κατανάλωσης θα είναι μεγαλύτερη. Αυτό συμβαίνει επειδή η ποσοστιαία μεταβολή υπολογίζεται σε σχέση με την αρχική τιμή. Όταν η αρχική τιμή είναι χαμηλή, οποιαδήποτε μικρή αλλαγή στην κατανάλωση θα έχει μεγαλύτερη επίδραση στην ποσοστιαία μεταβολή. Για παράδειγμα, αν ένα αυτοκίνητο έχει αρχική κατανάλωση 6 λίτρα ανά 100 χιλιόμετρα και μετά από βελτιώσεις η κατανάλωση μειωθεί σε 5 λίτρα ανά 100 χιλιόμετρα, η ποσοστιαία μεταβολή θα είναι -16.7%. Αντίθετα, αν ένα άλλο αυτοκίνητο έχει αρχική κατανάλωση 10 λίτρα ανά 100 χιλιόμετρα και μετά από βελτιώσεις η κατανάλωση μειωθεί σε 9 λίτρα ανά 100 χιλιόμετρα, η ποσοστιαία μεταβολή θα είναι -10%.
- Όσο αφορά τα plug-in υβριδικά αυτοκίνητα, οι τιμές της ποσοστιαίας απόκλισης τους μπορεί να κυμανθεί από 5% έως και 400%. Αυτή η διακύμανση ενδέχεται να οφείλεται πιθανότατα στον τρόπο χρήσης των plug-in υβριδικών οχημάτων. Πιο συγκεκριμένα, επειδή τα αυτοκίνητα αυτά αποτελούνται και από μηχανή καύσης αλλά και από ηλεκτρικό κινητήρα, η τιμή της κατανάλωσης WLTP προκύπτει από τον συνδυασμό και των δύο κινητήρων. Αυτό σημαίνει πως κατά τον κύκλο δοκιμών, το αυτοκίνητο λειτουργεί σε ρυθμό οικονομικής οδήγησης, που σημαίνει πως η χρήση του κινητήρα καύσης είναι αρκετά πιο σπάνια από αυτή του ηλεκτρικού, οδηγώντας σε πολύ μικρή κατανάλωση. Επιπλέον

Ωστόσο, το φαινόμενο που παρατηρήθηκε κατά την περιήγηση στη βάση δεδομένων έδειξε πως ένα μεγάλο ποσοστό οδηγών έκανε χρήση του κινητήρα καύσης για πολλά περισσότερα χιλιόμετρα σε σχέση με τον κύκλο WLTP, γεγονός που πιθανότατα οδήγησε στην κατανάλωση καυσίμου να είναι αυξημένη κατά πολύ μεγάλο ποσοστό, καθώς στην διαδικασία εργαστηριακών δοκιμών οι κατασκευαστές κάνουν χρήση του κινητήρα καύσης μόνο σε υψηλές ταχύτητες, γεγονός που σημαίνει πως η χρήση του κινητήρα χρησιμοποιείται μονάχα στην περίπτωση αυτή, δημιουργώντας έτσι μια εξαιρετικά χαμηλή απόλυτη κατανάλωση.

Μοντέλο	Κινητήρας	Μέση Απόκλιση	5ο εκατοστημόριο	95ο εκατοστημόριο	Διάμεσος της Απόκλισης
Hyundai Tucson PHEV	PHEV 195 kW	238,65%	93,25	405,18	251,07
Μοντέλο	Κινητήρας	Μέση Απόκλιση	5ο εκατοστημόριο	95ο εκατοστημόριο	Διάμεσος της Απόκλισης
Toyota RAV4 PHEV	PHEV 225 kW	223,31%	7,5	463,0	192,0
Μοντέλο	Κινητήρας	Μέση Απόκλιση	5ο εκατοστημόριο	95ο εκατοστημόριο	Διάμεσος της Απόκλισης
Ford Kuga PHEV	PHEV 195 kW	212,42%	41,48	435,11	172,46

Πίνακας 8

3.5.8 Συμπεράσματα:

Η έρευνα που πραγματοποιήθηκε, αποτέλεσε ένα σημαντικό βήμα για την ευρύτερη κατανόηση της αξιοπιστίας που προσφέρει ο κύκλος δοκιμών WLTP που πραγματοποιείται από τους κατασκευαστές και ταυτόχρονα αποτέλεσε εφαλτήριο για την μελέτη των αιτιών που οι συνθήκες τα κριτήρια και οι κανόνες σύμφωνα με τις οποίες πραγματοποιούνται οι εργαστηριακές δοκιμές οι οποίες εκτιμούν την κατανάλωση των αυτοκινήτων της αγοράς.

Τα αποτελέσματα που εξάχθηκαν, οδηγούν στο συμπέρασμα πως υπάρχουν παράμετροι όπως η οδηγική συμπεριφορά και η τεχνογνωσία του μέσου οδηγού οι οποίες θα μπορούσαν να ληφθούν υπ' όψη σε πιο έντονο βαθμό. Οι κύκλοι δοκιμών πραγματοποιούνται από εξειδικευμένο προσωπικό με βαθιά γνώση μηχανικής των αυτοκινήτων. Το εν λόγω προσωπικό γνωρίζει ακριβώς τον τρόπο με τον οποίο λειτουργεί ένα αυτοκίνητο ώστε να έχει την μέγιστη απόδοση καυσίμου. Στην περίπτωση όμως του μέσου ευρωπαϊκού οδηγού, είναι αρκετά σπάνιο ο γνωρίζει πως ένα αυτοκίνητο λειτουργεί στην μέγιστη απόδοση, δημιουργώντας έτσι μια εμφανή απόκλιση σε σύγκριση με τους κύκλους δοκιμών.

Αξίζει επίσης να σημειωθεί, πως θα πρέπει να διαφοροποιηθούν περαιτέρω οι δοκιμές του κύκλου WLTP για κάθε διαφορετική κατηγορία οχημάτων. Τα τεχνικά χαρακτηριστικά της κάθε κατηγορίας οχημάτων φαίνεται να είναι διαφορετικά, σε βαθμό που οι διαδικασία κύκλου WLTP – για την κάθε κατηγορία – να πραγματοποιείται κάτω από διαφορετικές συνθήκες.

Κεφάλαιο 4: Μοντελοποίηση οδηγικής συμπεριφοράς.

Από την μελέτη που προηγήθηκε φαίνεται πως ένα από τα πιο συχνά φαινόμενα που οδηγούν στην ύπαρξη απόκλισης μεταξύ πραγματικής κατανάλωσης και εργαστηριακών δοκιμών, αφορά τον ανθρώπινου παράγοντα. Σύμφωνα με την έρευνα που διεξήχθη από τους Emilia M.Szumaska και Rafal Jurecki, η επιθετική οδήγηση μπορεί να αυξήσει τόσο την κατανάλωση καυσίμου όσο και τις εκπομπές διοξειδίου του άνθρακα έως και 40%. (Szumska, E., & Jurecki, R., 2020)

Ένας από τους τρόπους που θα μπορούσε ο εργαστηριακός έλεγχος να πετύχει ακριβέστερες εκτιμήσεις κατανάλωσης καυσίμου, είναι να συμπεριλάβει μανούβρες που προκύπτουν από επικίνδυνη οδηγική συμπεριφορά. Με αφορμή λοιπόν την κατάσταση αυτή, προκύπτει η ανάγκη δημιουργίας ενός μοντέλου το οποίο θα μπορεί να διαβάζει δεδομένα κατεύθυνσης του αυτοκινήτου και να βρίσκεται σε θέση να αναγνωρίζει επιθετικές μανούβρες.

Για τον λόγο αυτό, στο παρόν κεφάλαιο αφορά την δημιουργία ενός μοντέλου μηχανικής μάθησης το οποίο θα έχει ως είσοδο μετρήσεις ταχύμετρου και γυροσκοπίου και θα αναγνωρίζει αν υπάρχει επιθετική μανούβρα και εφόσον υπάρχει, να την εντάξει στην ανάλογη κατηγορία.

4.1 Συλλογή και μελέτη δεδομένων

Κριτήρια επιλογής πηγών:

Για την δημιουργία του μοντέλου, κρίνεται αναγκαία η εύρεση δεδομένων τα οποία θα χρησιμοποιηθούν για την εκπαίδευση του μοντέλου. Όπως και στην έρευνα που προηγήθηκε πρέπει να επιλεγθούν δεδομένα τα οποία ικανοποιούν τις απαιτήσεις μοντέλου από αξιόπιστες πηγές.

Έτσι λοιπόν διεξήχθη μια αναζήτηση κατά την οποία βρέθηκε μια πληθώρα από datasets τα οποία περιείχαν μετρήσεις από κίνηση αυτοκινήτων. Για την επιλογή ενός dataset, ήταν απαραίτητο να περιέχει μετρήσεις από ταχύμετρο και γυροσκόπιο και για τους 3 άξονες (x,y,z) . Μετά την αναζήτηση των πηγών βρέθηκε ένα μοναδικό dataset το οποίο εμπεριέχει πληροφορίες για διάφορες επικίνδυνες μανούβρες οδήγησης. Το συγκεκριμένο dataset, συγκροτεί μετρήσεις από διαδρομές 3 οχημάτων κατηγορίας B: Ford Fiesta 1.4, Ford Fiesta 1.25, Hyundai i20. Μέσα σε αυτό συμπεριλαμβάνονται οι 4 μανούβρες που προκύπτουν σε καταστάσεις επιθετικής οδήγησης:

- Απότομη επιτάχυνση (label: 1).
- Απότομη δεξιά στροφή ή ελιγμός (label: 2).
- Απότομη αριστερή στροφή ή ελιγμός (label: 3).
- Απότομο φρενάρισμα (label: 4).

Μερικές σημαντικές επιπρόσθετες πληροφορίες σχετικά με το dataset αποτελούν:

- Ποσοστό δειγματοληψίας: Μέσος όρος 2 δειγμάτων (σειρές) ανά δευτερόλεπτο
- Μονάδες μέτρησης:
 - Επιταχυνσιόμετρο: m/s^2
 - Γυροσκόπιο: μοίρες ανά δευτερόλεπτο ($^{\circ}/s$)

Οι μετρήσεις πραγματοποιήθηκαν υπό την εποπτεία των Asim Sinan Yuksel και Şerafettin Atmaca, ως μέρος της έρευνάς τους στο Suleyman Demirel University.

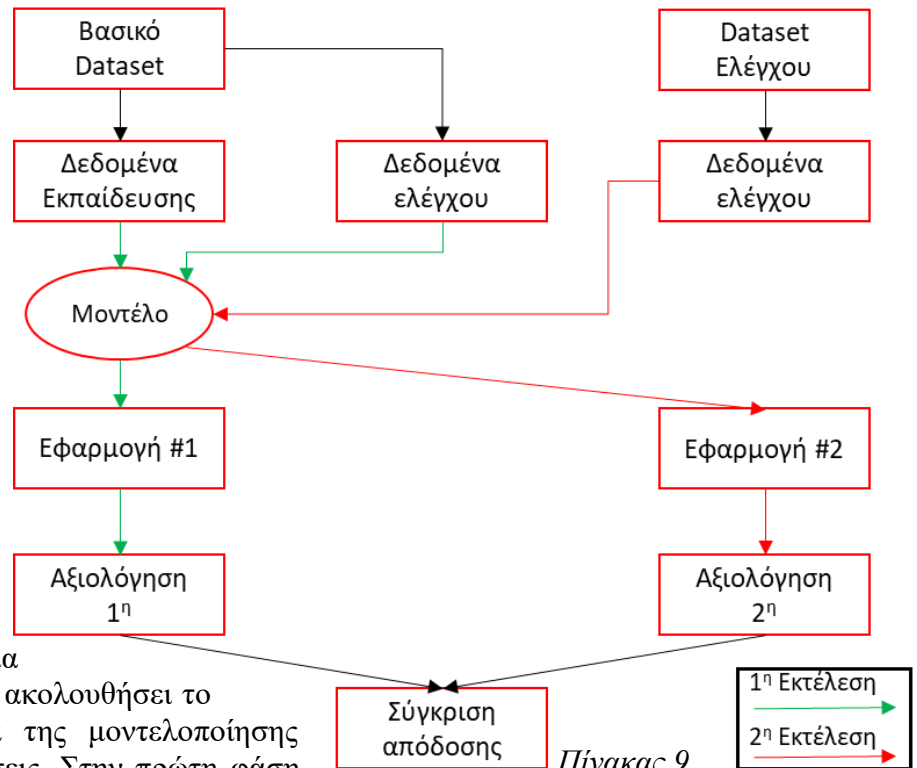
Οι απαιτήσεις του μοντέλου, προκύπτουν από τον επιθυμητό στόχο του μοντέλου.

Πέρα από το εν λόγω dataset το μοντέλο θα εκπαιδευτεί σε 2^η φάση από ένα dataset με το ίδιο ποσοστό δειγματοληψίας, ίδιας κατηγορίας αυτοκίνητα και ίδιων μονάδων μέτρησης των χαρακτηριστικών με το βασικό. Πρόκειται για dataset που δημιουργήθηκε υπό την εποπτεία των Ion Cojocaru, Paul-Stefan Popescu και Cristian Mihaescu του πανεπιστημίου της Κραϊόβας. Το συγκεκριμένο dataset αντί για επικίνδυνες μανούβρες οδήγησης παρέχει πληροφορίες για τη γενικότερη συμπεριφορά του οδηγού – Επιθετική, κανονική ή αργή . Η χρήση αυτού του dataset εξυπηρετεί το ρόλο της αξιολόγησης της αποτελεσματικότητας του πρώτου μοντέλου, ώστε να ελεγχθεί το πόσο ακριβές μπορεί να είναι το μοντέλο αν εκπαιδευτεί με καινούρια δεδομένα.

Η χρήση του 2^{ου} dataset κρίνεται επίσης αναγκαία, καθώς υπάρχουν πολλές διαφορετικές αιτίες που ενδέχεται να επηρεάσουν την κατανάλωση καυσίμου, πέρα από την οδηγική συμπεριφορά. Συνθήκες οδήγησης με πρόσφυση, ανεπαρκής ποσότητα αέρα στα ελαστικά και καιρικές συνθήκες δύναται να αυξήσουν σημαντικά την απόλυτη μέση κατανάλωση καυσίμου. Τούτο σημαίνει πως ακόμα και όταν η συμπεριφορά του οδηγού είναι μη επιθετική, η κατανάλωση του αυτοκινήτου του να αυξηθεί. Αφού λοιπόν βασικός σκοπός του μοντέλου αποτελεί η ρύθμιση και η βελτιστοποίηση της κατανάλωσης, κρίνεται απαραίτητο το μοντέλο να διαθέτει την ικανότητα να επεξεργάζεται και δεδομένα μη επιθετικής οδηγικής συμπεριφοράς, έτσι ώστε να μπορεί να εφαρμοστεί σε πραγματικές (καθημερινές) συνθήκες.

Αναγνώριση προβλήματος και καθορισμός μεθοδολογίας

Το πρόβλημα που καλείται να λύσει το μοντέλο αποτελεί ένα χαρακτηριστικό πρόβλημα ταξινόμησης πολλαπλών τάξεων (Multiclass Classification). Ένα πρόβλημα ταξινόμησης πολλαπλών τάξεων αφορά ένα είδος προβλήματος μηχανικής μάθησης όπου στόχος είναι να η ταξινόμηση περιπτώσεων σε περισσότερες από δύο αμοιβαία αποκλειστικές τάξεις. Με άλλα λόγια, περιλαμβάνει την ανάθεση μιας μόνο παρουσίας σε μία από τις διάφορες πιθανές κατηγορίες ή τάξεις. Στην προκειμένη περίπτωση, πρέπει να γίνει ταξινόμηση ανάμεσα στις 4 κατηγορίες επικίνδυνων μανουβρών οδήγησης.



Το παραπάνω διάγραμμα περιγράφει την ροή που θα ακολουθήσει το πρόγραμμα. Η διαδικασία της μοντελοποίησης πραγματοποιείται σε 2 φάσεις. Στην πρώτη φάση εφαρμόζεται το μοντέλο που θα δημιουργηθεί με δεδομένα εκπαίδευσης (training) και ελέγχου (testing) από το βασικό dataset. Έπειτα αξιολογείται η ακρίβεια ταξινόμησης της πρώτης εκτέλεσης. Έπειτα χρησιμοποιείται το 2^ο dataset (ελέγχου) για την αξιολόγησης της ικανότητας του μοντέλου να ξεχωρίζει μια επιθετική οδηγική συμπεριφορά από μία μη επιθετική. Επιπλέον, αυτό πραγματοποιείται προκειμένου να αποδειχθεί πως το μοντέλο θα μπορέσει να εφαρμοστεί και να είναι αποτελεσματικό ακόμα και αν εκπαιδευτεί με άγνωστα σε αυτό δεδομένα, κάτι που πραγματοποιείται στο τελευταίο βήμα, στο οποίο συγκρίνονται οι μετρήσεις αποτελεσματικότητας και των δυο εφαρμογών του μοντέλου.

Πίνακας 9

Επιλογή Μεθόδου εκμάθησης του μοντέλου

Για την δημιουργία του μοντέλου ταξινόμησης, εξετάστηκαν 2 μέθοδοι υλοποίησης του μοντέλου.

Η πρώτη αφορούσε την χρήση ενός αλγορίθμου μηχανικής μάθησης και συγκεκριμένα τη χρήση Random Forest ή και Δέντρων Αποφάσεων (Decision Trees).

Η δεύτερη προσέγγιση, συνιστούσε την δημιουργία ενός νευρωνικού δικτύου.

Τα κριτήρια επιλογής της κατάλληλης μεθόδου μηχανικής μάθησης κρίνονται από την μορφή και τις σχέσεις που έχουν τα δεδομένα μέσα στο dataset.

Μελέτη δεδομένων

Το βασικό dataset αποτελείται από 18734 καταγραφές δεδομένων και έχει συνολικά 61 στήλες με χαρακτηριστικά. Τα χαρακτηριστικά ακολουθούν μια συγκεκριμένη δομή. Κατά τη διάρκεια της δειγματοληψίας είναι καταγεγραμμένες οι ακόλουθες μετρήσεις:

- AccX (Accelerometer X) : Δείγμα μέτρησης ταχύμετρου στον X άξονα

- AccY (Accelerometer Y) : Δείγμα μέτρησης ταχύμετρου στον Y άξονα
- AccZ (Accelerometer Z) : Δείγμα μέτρησης ταχύμετρου στον Z άξονα
- GyroX (Gyroscope X) : Δείγμα μέτρησης γυροσκοπίου στον X άξονα
- GyroY (Gyroscope Y) : Δείγμα μέτρησης γυροσκοπίου στον Y άξονα
- GyroZ (Gyroscope Z) : Δείγμα μέτρησης γυροσκοπίου στον Z άξονα

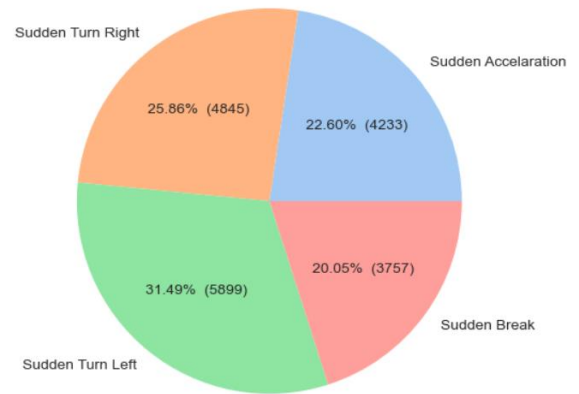
Το κάθε δείγμα, διαχωρίζεται σε επιμέρους μετρήσεις και δημιουργούνται οι στήλες του dataset. Συγκεκριμένα, η βάση δεδομένων περιλαμβάνει του ακόλουθους υπολογισμούς των μετρήσεων που πραγματοποιήθηκαν κατά τη δειγματοληψία:

- Mean: Αντιπροσωπεύει τη μέση (μέση) τιμή της επιτάχυνσης στην κατεύθυνση του αντίστοιχου άξονα. Δίνει μια ένδειξη της μέσης τιμής που βιώνει το αυτοκίνητο στο δείγμα που λήφθηκε. (π.χ. AccMeanX: Accelerometer, Mean, X-axis)
- Covariance (Cov): Αφορά τη συνδιακύμανση των τιμών επιτάχυνσης στην κατεύθυνση του άξονα X. Η συνδιακύμανση μετρά τη σχέση μεταξύ δύο μεταβλητών, σε αυτήν την περίπτωση, μπορεί να υποδεικνύει πώς η μέτρηση σε σχέση με άλλους παράγοντες ή μεταβλητές (π.χ. AccCovX: Accelerometer, Covariance, X-axis)
- Skewness (Skew): Αντιπροσωπεύει την ασυμμετρία των τιμών των μετρήσεων. Η ασυμμετρία είναι ένα στατιστικό μέτρο που περιγράφει την ασυμμετρία μιας κατανομής. Παρέχει πληροφορίες σχετικά με το σχήμα και τη συμμετρία της κατανομής των τιμών (π.χ. AccSkewX: Accelerometer, Skewness, X-axis).
- Kurtosis (Kurt): Αναφέρεται στην κύρτωση των τιμών της μέτρησης στην κατεύθυνση του άξονα X. Η κύρτωση είναι ένα στατιστικό μέτρο που περιγράφει το σχήμα μιας κατανομής. Παρέχει πληροφορίες σχετικά με την παρουσία και την έκταση των ακραίων τιμών ή των ακραίων τιμών (π.χ. AccKurtX: Accelerometer, Kurtosis, X-axis).
- Sum: Παρέχει ένα συνολικό μέτρο της συνολικής επιτάχυνσης ή γωνίας κατεύθυνσης του αυτοκινήτου που παρατηρείται κατά τη μέτρηση (π.χ. AccSumX: Accelerometer, Sum, X-axis).
- Minimum (Min): Ελάχιστη τιμή από το σύνολο των μετρήσεων του δείγματος (π.χ. AccMinX: Accelerometer, Minimum, X-axis).
- Maximum (Max): Μέγιστη τιμή από το σύνολο των μετρήσεων του δείγματος (π.χ. AccMaxX: Accelerometer, Maximum, X-axis).
- Variance (Var): Διακύμανση της κάθε μέτρησης σε σχέση με τη μέση τιμή του δείγματος (π.χ. AccVarX: Accelerometer, Variance, X-axis).
- Standard Deviation (Std): Διακύμανση της κάθε μέτρησης σε σχέση με τη μέση τιμή του δείγματος (π.χ. AccStdX: Accelerometer, Standard Deviation, X-axis).

- Median: Διάμεσος των τιμών από τις μετρήσεις του δείγματος (π.χ. AccMedianX: Accelerometer, Median, X-axis).

Τέλος το dataset περιέχει μια στήλη η οποία δηλώνει το είδος της επικίνδυνης μανούβρας με τη μορφή αριθμητικής τιμής:

1. Απότομη Επιτάχυνση (Ετικέτα Κατηγορίας: **1**).
2. Απότομη Δεξιά Στροφή (Ετικέτα Κατηγορίας: **2**).
3. Απότομη Αριστερή Στροφή (Ετικέτα Κατηγορίας: **3**).
4. Απότομο Φρενάρισμα (Ετικέτα Κατηγορίας: **4**).



Εικόνα 19

Γενικότερα, ο πληθυσμός των κατηγοριών φαίνεται να είναι μοιρασμένος ισάξια καθώς και οι 4 κατηγορίες αποτελούν ένα σημαντικό ποσοστό του πληθυσμού

Από τις 18.734 σειρές:

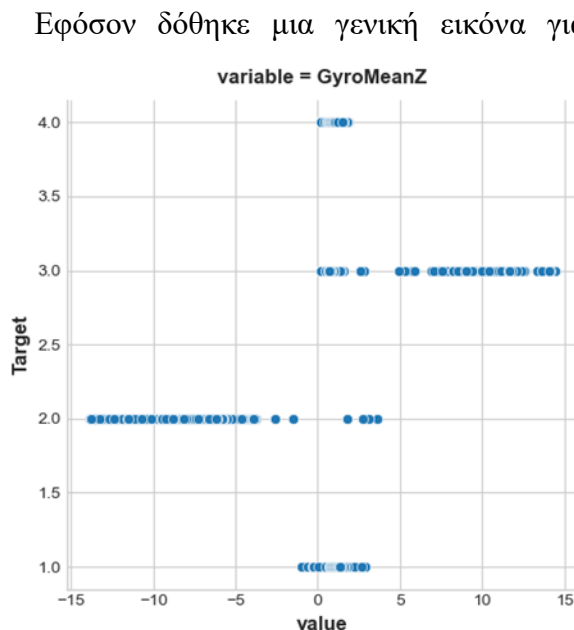
Οι 4.233 αφορούν απότομη επιτάχυνση (22%),

οι 3.757 απότομο φρενάρισμα (20%),

οι 4.45 απότομη δεξιά στροφή (25%)

και οι υπόλοιπες 5.899

απότομη αριστερή (31%)



Εικόνα 20

την μορφή των δεδομένων και το τι αντιπροσωπεύει το καθένα ξεκινάει η διαδικασία ανάλυσης των δεδομένων, με στόχο την εύρεση των χαρακτηριστικών που επηρεάζουν περισσότερο την τιμή της κάθε επικίνδυνης μανούβρας, έτσι ώστε να επιλεγθούν τα κατάλληλα χαρακτηριστικά για την δημιουργία του μοντέλου. Για τη διευκόλυνση της κατανόησης των δεδομένων χρησιμοποιήθηκαν οι βιβλιοθήκες matplotlib και seaborn, με σκοπό την

καλύτερη οπτικοποίηση των δεδομένων. Σε αρχικό στάδιο μελετήθηκε η άμεση σχέση που έχει το κάθε χαρακτηριστικό με τον στόχο. Έτσι λοιπόν έγινε χρήση Implots μεταξύ κάθε χαρακτηριστικού ξεχωριστά και της κατηγορίας

Παρατηρήσεις:

Οι μετρήσεις του επιταχυνσιόμετρου στους άξονες X και Y φαίνεται να επηρεάζουν περισσότερο την απότομη επιτάχυνση και το απότομο φρενάρισμα.

Οι μετρήσεις του γυροσκοπίου στον άξονα Z φαίνεται να επηρεάζουν περισσότερο την απότομη αριστερή ή δεξιά στροφή.

Στα περισσότερα χαρακτηριστικά δεν υπάρχει ξεκάθαρη εικόνα και για τους 4 ελιγμούς. Στις περισσότερες περιπτώσεις, το κάθε χαρακτηριστικό βοηθάει στην κατηγοριοποίηση ενός ζευγαριού τιμών, με **τις τιμές των υπολοίπων 2 χαρακτηριστικών να έχουν μεγαλύτερο εύρος σε σημείο η μία να καλύπτει την άλλη.**

Επιλογή Μεθόδου υλοποίησης του μοντέλου

Μετά την αρχική μελέτη των δεδομένων αποφασίστηκε να χρησιμοποιηθεί ένα τεχνητό νευρωνικό δίκτυο για την δημιουργία του μοντέλου.

Ο βασικότερος λόγος που επιλέχθηκε το Τεχνητό Νευρωνικό δίκτυο αποτελεί η τόσο έλλειψη γραμμικότητας στη σχέση μεταξύ των δεδομένων, όσο και ύπαρξη παρόμοιων τιμών μεταξύ των διαφορετικών κατηγοριών. Τα ANN είναι ικανά να ανιχνεύουν πολύπλοκες, μη γραμμικές σχέσεις στα δεδομένα, το οποίο είναι ζωτικής σημασίας όταν αντιμετωπίζουμε την περίπλοκη και δυναμική φύση των επιθετικών κινήσεων οδήγησης.

Ορισμένα χαρακτηριστικά από τη μια να βοηθούν πάρα πολύ στην κατηγοριοποίηση ενός ζεύγους τιμών, αλλά από την άλλη να δημιουργούν αρκετό θόρυβο και να δυσκολεύουν την ταξινόμηση ενός άλλου ζεύγους. Με τη χρήση όμως ενός τεχνητού νευρωνικού δικτύου, επιτυγχάνεται αποτελεσματικότερη ρύθμιση των βαρών, σε σύγκριση με τις συμβατικές μεθόδους μηχανικής μάθησης, μειώνοντας έτσι το ενδεχόμενο σύγχυσης του μοντέλου.

Τεχνητά Νευρωνικά δίκτυα (Artificial Neural Networks)

Πρόκειται για ένα υπολογιστικό μοντέλο που μιμείται τη δομή και τη λειτουργικότητα των νευρωνικών δικτύων του ανθρώπινου εγκεφάλου. Είναι ένας αλγόριθμος μηχανικής μάθησης που αποτελείται από διασυνδεδεμένους κόμβους, γνωστούς ως τεχνητούς νευρώνες ή μονάδες, οργανωμένους σε στρώματα.

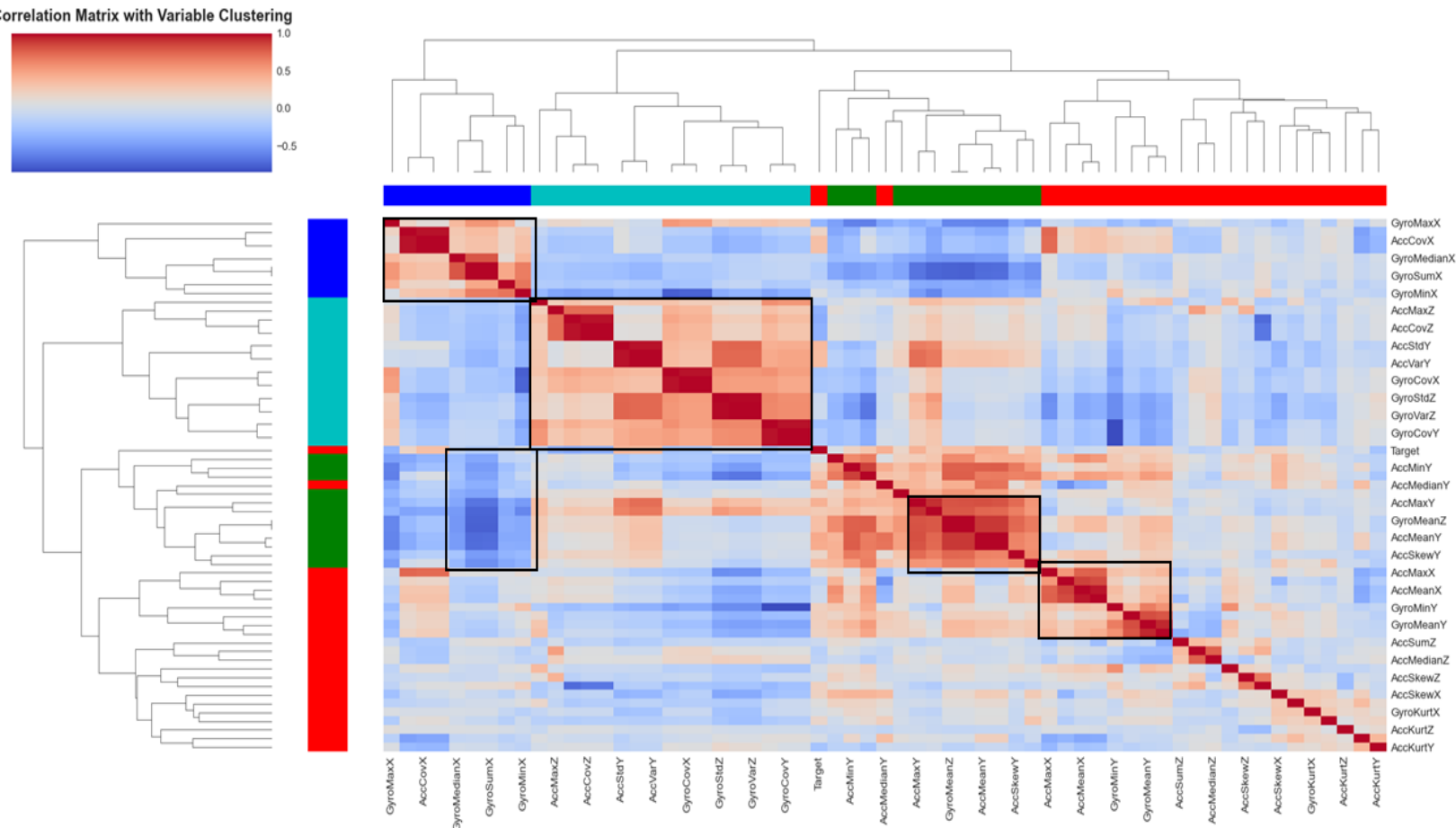
Τα τεχνητά νευρωνικά δίκτυα χρησιμοποιούν τεχνητούς νευρώνες για να λαμβάνουν σήματα εισόδου, να εφαρμόζουν βάρη σε αυτά και να περνούν το σταθμισμένο άθροισμα μέσω μιας λειτουργίας ενεργοποίησης για να δημιουργήσουν μια έξοδο. Τα βάρη καθορίζουν τη δύναμη των συνδέσεων μεταξύ των νευρώνων και προσαρμόζονται κατά τη διάρκεια της μαθησιακής διαδικασίας για τη βελτιστοποίηση της απόδοσης του δικτύου.

Τα ANN χαρακτηρίζονται από την πολύ-επίπεδη αρχιτεκτονική τους, η οποία συνήθως αποτελείται από ένα στρώμα εισόδου, ένα ή περισσότερα κρυμμένα στρώματα και ένα στρώμα εξόδου. Το στρώμα εισόδου λαμβάνει τα αρχικά δεδομένα, τα κρυμμένα στρώματα εκτελούν ενδιάμεσους υπολογισμούς και το στρώμα εξόδου παράγει το τελικό αποτέλεσμα ή πρόβλεψη.

Κατά τη διάρκεια της εκπαίδευσης, ένα τεχνητό νευρωνικό δίκτυο μαθαίνει να προσεγγίζει περίπλοκα μοτίβα και σχέσεις στα δεδομένα προσαρμόζοντας επαναλαμβανόμενα τα βάρη με βάση μια λειτουργία στόχου ή σφάλματος. Τεχνικές όπως η αντίστροφη διάδοση χρησιμοποιούνται συνήθως για τη διάδοση σφαλμάτων προς τα πίσω μέσω του δικτύου και την ενημέρωση των βαρών.

Αφού αποκτήθηκε μια γενική εικόνα για τα δεδομένα και αφού ορίστηκε η μέθοδος που θα χρησιμοποιηθεί για το μοντέλο, ξεκινάει η διαδικασία εύρεσης του σωστού μείγματος στοιχείων εισόδου του μοντέλου.

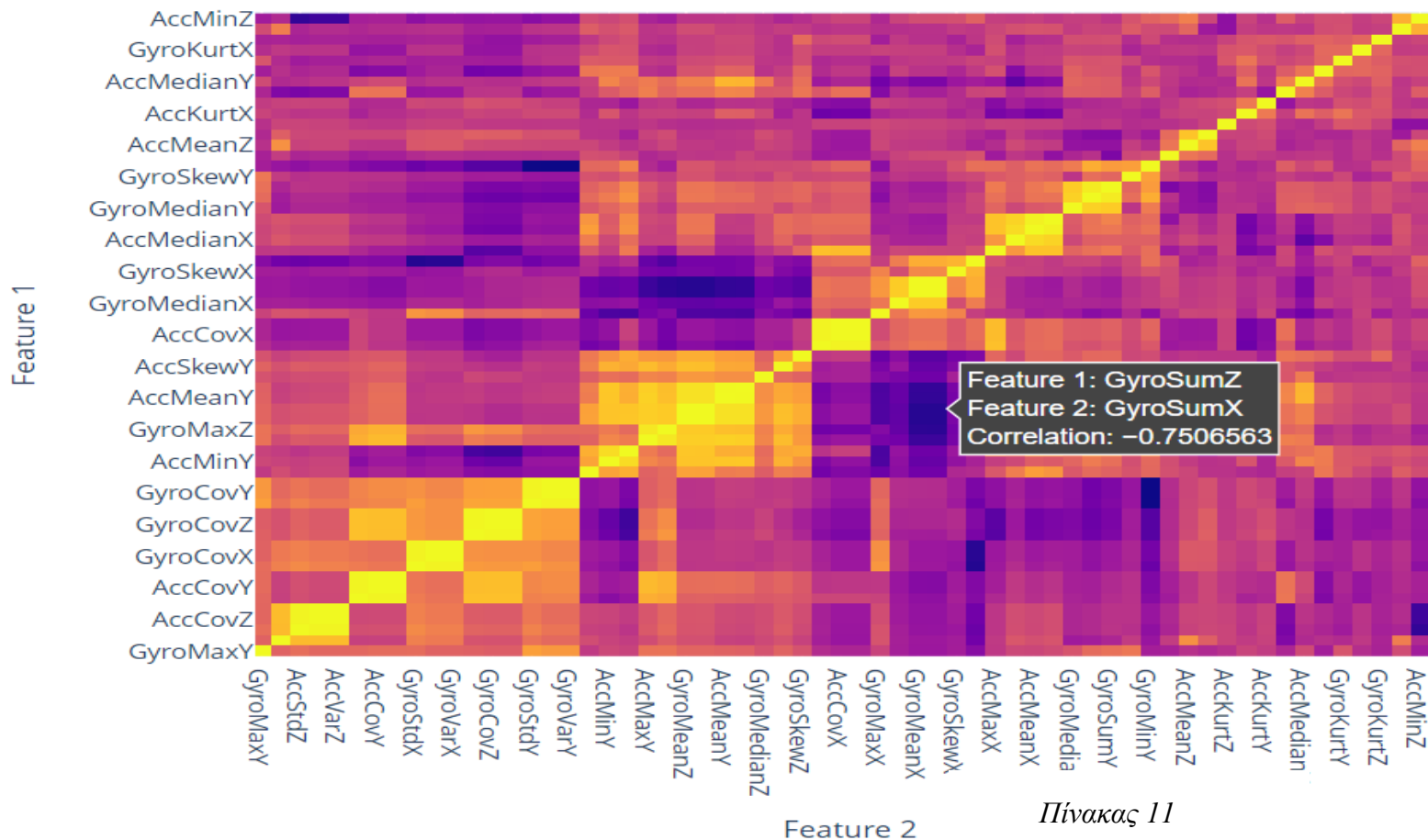
Σε πρώτο στάδιο μελετήθηκε σε βάθος η σχέση μεταξύ όλων των μεταβλητών με τη βοήθεια ενός πίνακα συσχέτισης με μεταβλητή ομαδοποίηση. Καθώς το dataset περιέχει έναν πολύ μεγάλο όγκο χαρακτηριστικών ένας τυπικός πίνακας συσχετίσεων δεν αρκεί για την μελέτη γραμμικών συσχετίσεων μεταξύ των δεδομένων. Με την εφαρμογή ιεραρχικής ομαδοποίησης στον πίνακα συσχέτισης δημιουργήθηκαν 4 ομάδες οι οποίες έχουν κοινά χαρακτηριστικά, δημιουργώντας έτσι ένα πιο εύκολο περιβάλλον για τη μελέτη των σχέσεων των δεδομένων.



Πίνακας 10

Όπως φαίνεται και στο γράφημα μπορούν διακριθούν ήδη 5 σημεία που αποτελούνται από ομάδες δεδομένων που έχουν υψηλή συσχέτιση μεταξύ τους. Αφού πλέον έχουν ξεκαθαριστεί οι συνδυασμοί των δεδομένων προς μελέτη, δημιουργείται ένας τυπικός πίνακας συσχετίσεων με τη λειτουργία εμφάνισης ένδειξης της τιμής των συσχετίσεων όταν ο κέρσορας του ποντικιού ακουμπάει στο κάθε σημείο του πίνακα για ευκολότερη ανάλυση.

Hover Cluster Map



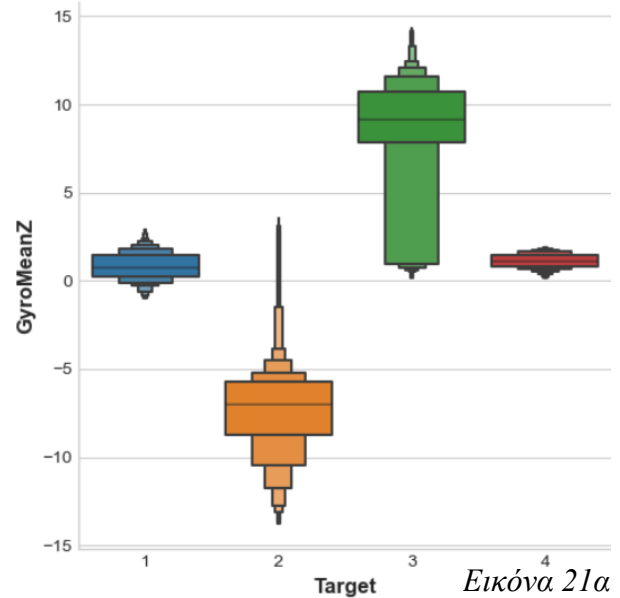
Αφού μελετήθηκαν όλοι οι – από τις ομάδες - επιλεγμένοι συνδυασμοί δεδομένων καταχώθηκαν σε ένα πίνακα που περιέχει τις μεταξύ τους συσχετίσεις σε φθίνουσα σειρά.

Το επόμενο στάδιο, ανάλυσης δεδομένων αφορά την ιεράρχηση δεδομένων βάσει του “mutual information value”. Πρόκειται για ένα μέτρο της εξάρτησης μεταξύ δύο μεταβλητών. Στην περίπτωση αυτή -με χρήση της «mutual_info_classifier» από την βιβλιοθήκη sk learn- η ποσότητα πληροφοριών που παρέχει ένα χαρακτηριστικό ποσοτικοποιείται ως προς τη σχέση του με την τιμή της εξόδου. Μια υψηλότερη τιμή αμοιβαίας πληροφορίας υποδεικνύει ότι το χαρακτηριστικό είναι πιο ενημερωτικό για την πρόβλεψη της μεταβλητής στόχου.

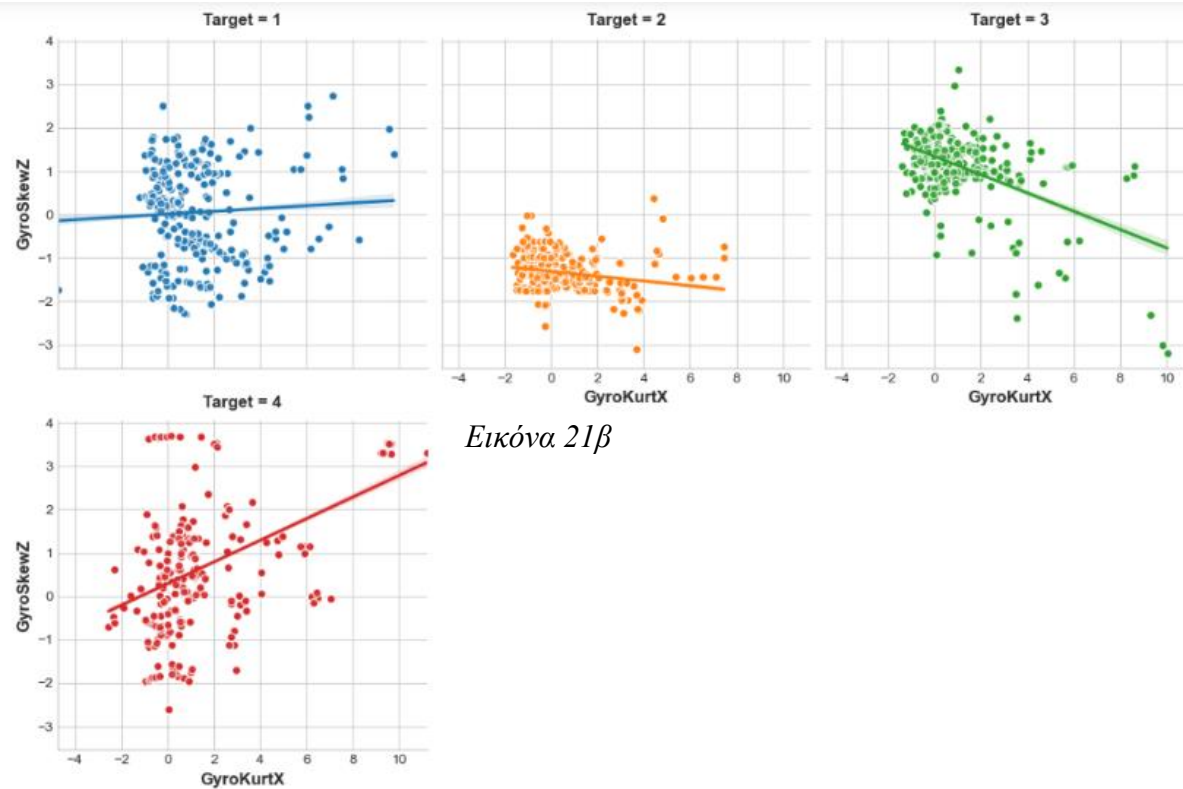
Μετά την ιεράρχηση των τιμών, επιλέγεται ένα αρχικό δείγμα από χαρακτηριστικά που είχαν τόσο υψηλό ποσό mutual information όσο και υψηλή συσχέτιση. Στο σύνολο αυτό πραγματοποιήθηκε έλεγχος με τη χρήση boxplots (Εικόνα 21α) για τον εντοπισμό και την αφαίρεση χαρακτηριστικών με υψηλό ποσοστό ακραίων τιμών, οι οποίες ενδεχομένως να δυσκολεύανε τη ταξινόμηση.

Μετά από την διερεύνηση όλων των τιμών παρατηρήθηκε πως υπάρχει μόνο 2 χαρακτηριστικά που δίνουν σαφείς πληροφορίες για τον διαχωρισμό αριστερής από της δεξιά στροφή (GyroMeanZ: Gyroscope, Mean value, Z-axis και GyroSumZ: Gyroscope, Sum, Z-axis). Έτσι ξεκίνησε η διαδικασία δημιουργία καινούριων χαρακτηριστικών (feature engineering) για τη δημιουργία χαρακτηριστικών που διευκολύνουν τον διαχωρισμό μεταξύ των στροφών. Με τη χρήση Implots ανακαλύφθηκαν δυο σχέσεις οι οποίες μπορούσε ξεκάθαρα να διαχωριστεί η αριστερή από την δεξιά στροφή.

Σύμφωνα με τον διπλανό πίνακα φαίνεται πως αν συνδυαστούν τα δεδομένα GyroKurtX (Gyroscope, Kurtosis, X-axis) και GyroSkewZ (Gyroscope,



Skewness, Z-axis) φαίνεται πως οι 2 κλάσεις των στροφών έχουν αρκετά διαφορετικές τιμές. Έτσι λοιπόν δημιουργείται ένα καινούριο χαρακτηριστικό που αποτελεί το πηλίκο των τιμών. Οι τιμές του συνδυασμού των 2 τιμών καταχωρούνται ως καινούρια στήλη στο dataset. Όπως φαίνεται και στα παραπάνω



διαγράμματα (21β) ο συνδυασμός των δυο προαναφερθέντων χαρακτηριστικών βοηθάει πολύ στον διαχωρισμό μεταξύ των δυο στροφών (Target =2 –κίτρινο- & Target =3 –πράσινο-), καθώς οι τιμές που παίρνουν φαίνεται να διαφέρουν σημαντικά. Η απότομη αριστερή στροφή (πράσινο διάγραμμα) φαίνεται να έχει παίρνει αρκετά μεγαλύτερες τιμές συγκριτικά με την απότομη δεξιά στροφή (κίτρινο διάγραμμα). Για το λόγο αυτό δημιουργήθηκε ένα καινούριο χαρακτηριστικό GyroSkewZ/GyroKurtX_ratio το οποίο μετά από πειραματική μελέτη αποδείχθηκε ότι βοηθάει αρκετά το μοντέλο να ξεχωρίζει τις αριστερές από τις δεξιές στροφές.

4.2 Κατασκευή Μοντέλου

Μετά από την διερεύνηση όλων των χαρακτηριστικών, επιλέχθηκαν 13 χαρακτηριστικά εισόδου για την εκμάθηση του μοντέλου.

Επιλεγμένα χαρακτηριστικά εισόδου:

- AccMeanX: Accelerometer Mean X-axis
- AccMeanY: Accelerometer Mean Y-axis
- AccSkewY: Accelerometer Skewness Y-axis
- AccSumX: Accelerometer Sum X-axis
- AccSumY: Accelerometer Sum Y-axis
- AccMinY: Accelerometer Minimum Y-axis
- AccMinX: Accelerometer Minimum X-axis
- AccMaxX: Accelerometer Maximum X-axis
- GyroMeanZ: Gyroscope Mean Z-axis
- GyroMinZ: Gyroscope Minimum Z-axis
- GyroMaxZ: Gyroscope Maximum Z-axis
- GyroSumZ: Gyroscope Sum Z-axis
- GyroSkewZ/GyroKurtX_ratio

Κατασκευή Νευρωνικού Δικτύου

Μετά την επιλογή των χαρακτηριστικών εισόδου, κατασκευάστηκε με τη βοήθεια της βιβλιοθήκης “tensorflow” το τεχνητό νευρωνικό δίκτυο που θα αναγνωρίζει τις 4 διαφορετικές μανούβρες επιθετικής οδήγησης.

Tensorflow

Η TensorFlow είναι μια βιβλιοθήκη Python ανοιχτού κώδικα που χρησιμοποιείται ευρέως για την κατασκευή και την εκπαίδευση μοντέλων μηχανικής μάθησης και βαθιάς μάθησης. Παρέχει ένα ολοκληρωμένο σύνολο εργαλείων και λειτουργιών για αριθμητικούς υπολογισμούς, νευρωνικά δίκτυα και μεγάλης κλίμακας κατανεμημένη εκπαίδευση. Η χρήση της επιτρέπει τη δημιουργία ενός ολοκληρωμένου οικοσυστήματος για την ανάπτυξη και την ανάπτυξη μοντέλων μηχανικής μάθησης, προσφέροντας ευελιξία, επεκτασιμότητα και υποστήριξη τόσο για περιπτώσεις έρευνας όσο και για περιπτώσεις χρήσης παραγωγής. Χρησιμοποιείται ευρέως από ερευνητές, επιστήμονες δεδομένων και προγραμματιστές για ένα ευρύ φάσμα εργασιών μηχανικής μάθησης και βαθιάς μάθησης.

Μεθοδολογία: Το πρώτο βήμα για την δημιουργία του Νευρωνικού δικτύου αφορά τον χωρισμό των δεδομένων σε δεδομένα εκπαίδευσης (training) και δεδομένα επαλήθευσης (testing). Ο αριθμός των δειγμάτων που χρησιμοποιήθηκε για έλεγχο προέκυψε από τον υπολογισμό της διαμέσου του πληθυσμού των διαφορετικών επικίνδυνων ελιγμών.

Με άλλα λόγια, όπως είχε προαναφερθεί, ο συνολικός πληθυσμός του dataset 18,734 τιμές, από τις οποίες, οι 4,233 αφορούν απότομη επιτάχυνση (22%), οι 3,757 απότομο φρενάρισμα (20%), οι 4,845 απότομη δεξιά στροφή (25%) οι υπόλοιπες 5,899 απότομη αριστερή στροφή (31%). Για την ισάξια δειγματοληψία δεδομένων ελέγχου υπολογίστηκε:

1) Διάμεσος πληθυσμών = 4,839

2) Ποσοστό δεδομένων ελέγχου = $\frac{\text{Διάμεσος πληθυσμών}}{\text{Συνολικός πληθυσμος}} * 100\% = \frac{4,839}{18,734} * 100\% = 24\%$

Έτσι λοιπόν χρησιμοποιήθηκε το 24% των δεδομένων για έλεγχο και το υπόλοιπο 76% για εκπαίδευση.

Μετά τον διαχωρισμό σε δεδομένα εκπαίδευσης και δεδομένα επαλήθευσης, τα δεδομένα εισόδου κανονικοποιούνται. Στην προκειμένη περίπτωση, η κανονικοποίηση των δεδομένων κρίνεται απαραίτητη, καθώς τα δεδομένα δεν μοιράζονται την ίδια μονάδα μέτρησης. Η κανονικοποίησή τους όμως για να έρθουν όλα τα δεδομένα – ανεξαρτήτως μονάδας μέτρησης – σε παρόμοια κλίμακα, γεγονός που θα βοηθήσει περαιτέρω το Νευρωνικό δίκτυο να:

- Χρησιμοποιεί πιο αποτελεσματικά αλγορίθμους βελτιστοποίησης.
- Αποφύγει το Bias. Εάν τα χαρακτηριστικά εισόδου έχουν διαφορετικές κλίμακες, το νευρωνικό δίκτυο μπορεί να αποδώσει δυσανάλογη σημασία σε χαρακτηριστικά με Μεγαλύτερες τιμές. Η κανονικοποίηση των δεδομένων διασφαλίζει ότι κάθε χαρακτηριστικό συμβάλλει δίκαια στη διαδικασία μάθησης και βοηθά στην πρόληψη της προκατάληψης προς συγκεκριμένα χαρακτηριστικά με βάση την κλίμακα τους.
- Είναι πιο αποτελεσματικό στη γενίκευση. Όταν τα δεδομένα εισόδου κανονικοποιούνται, το δίκτυο γίνεται πιο ισχυρό σε παραλλαγές στις τιμές χαρακτηριστικών εισόδου. Αυτή η βελτιωμένη ικανότητα γενίκευσης μπορεί να οδηγήσει σε καλύτερες επιδόσεις όταν το μοντέλο κάνει προβλέψεις για νέα και άγνωστα σε αυτό δεδομένα.

Προτού κατασκευαστεί το Νευρωνικό δίκτυο, είναι απαραίτητο να μετατραπούν οι τιμές της εξόδου να μετατραπούν σε μονοδιάστατους πίνακες. Η επιθυμητή έξοδος του μοντέλου είναι ένας μονοδιάστατος πίνακας (1x4), όπου η τιμή του κάθε στοιχείου του θα συμβολίζει το ποσοστό ομοιότητας με την επικίνδυνη οδηγική μανούβρα. Αξίζει να σημειωθεί πως αν το ποσοστό ομοιότητας και των 4 στοιχείων του πίνακα δεν ξεπερνάει το 50%-55% τότε η μανούβρα μπορεί να χαρακτηριστεί ως μη επικίνδυνη. Για το λόγο αυτό οι 4 τιμές της εξόδου μετατρέπονται ως εξής:

- Απότομη επιτάχυνση (label:1) → [1 , 0, 0, 0]
- Απότομη δεξιά στροφή (label:2) → [0 , 1, 0, 0]

- Απότομη αριστερή στροφή (label:3) $\rightarrow [0, 0, 1, 0]$
- Απότομη πέδηση (label:4) $\rightarrow [0, 0, 0, 1]$

Δομή Νευρωνικού δικτύου

Αφού ολοκληρώθηκε η προ επεξεργασία όλων των δεδομένων, κατασκευάζεται ένα Νευρωνικό δίκτυο που αποτελείται από συνολικά 5 layers (1 input layer, 2 hidden layers και 1 output layer).

- **Input layer:** Το 1^ο layer του νευρωνικού δικτύου αποτελείται από ένα «dense» στρώμα με 20 μονάδες/νευρώνες. Με τη χρήση της παραμέτρου «input_shape» καθορίζεται το σχήμα των δεδομένων εισόδου, το οποίο στην περίπτωση αυτή είναι [13], υποδεικνύοντας ότι τα δεδομένα εισόδου έχουν 13 χαρακτηριστικά. Η συνάρτηση ενεργοποίησης που χρησιμοποιείται σε αυτό το layer είναι η ενεργοποίηση της Rectified Linear Unit (ReLU*).

Hidden Layers: Τα δυο layers που ακολουθούν, έχουν και αυτά «Dense» μορφή με τον αριθμό των νευρώνων να τριπλασιάζεται κάθε φορά σε σχέση με αυτόν του προηγούμενου layer. Ως συνάρτηση ενεργοποίησης ξαναχρησιμοποιείται η ReLU*.

*Η συνάρτηση ενεργοποίησης ReLU που χρησιμοποιείται στα layers για να εισάγει μη γραμμικότητα, επιτρέποντας στο μοντέλο να μάθει σύνθετες σχέσεις στα δεδομένα.

- **Output Layer:** Το παρών layer εξόδου του νευρωνικού δικτύου αποτελείται από ένα «Dense» layer με 4 νευρώνες, το οποίο αντιστοιχεί στον αριθμό των τάξεων ή κατηγοριών στην έξοδο. Η συνάρτηση ενεργοποίησης που χρησιμοποιείται σε αυτό το στρώμα είναι το «softmax», το οποίο χρησιμοποιείται συχνά για προβλήματα ταξινόμησης πολλαπλών κλάσεων. Υπολογίζει τις πιθανότητες των διαφόρων τάξεων και εξασφαλίζει ότι το άθροισμα των πιθανοτήτων είναι 1.

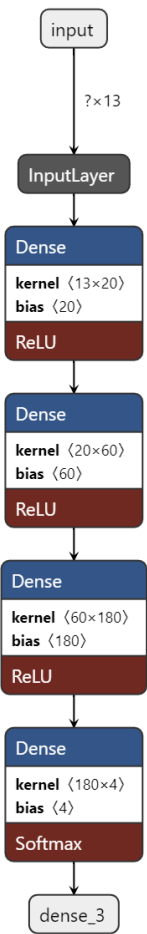
4.3 Εφαρμογή των δεδομένων στο Νευρωνικό δίκτυο

Τα δεδομένα εκπαίδευσης και ελέγχου από το βασικό dataset χρησιμοποιούνται για την εκμάθηση του νευρωνικού δικτύου.

Για την εκμάθηση του μοντέλου χρησιμοποιήθηκε ο αλγόριθμος βελτιστοποίησης adam. Ο Adam optimizer είναι ένας δημοφιλής αλγόριθμος βελτιστοποίησης που χρησιμοποιείται για την εκπαίδευση νευρωνικών δικτύων. Μερικά βασικά χαρακτηριστικά του Adam περιλαμβάνουν προσαρμογή ρυθμών μάθησης, διόρθωση bias.

Ο Adam προσαρμόζει το ρυθμό μάθησης για κάθε παράμετρο. Αυτός ο προσαρμοστικός ρυθμός μάθησης επιτρέπει ταχύτερη σύγκλιση κάνοντας μεγαλύτερες ενημερώσεις για σπάνια ενημερωμένες παραμέτρους και μικρότερες ενημερώσεις για συχνά ενημερωμένες παραμέτρους. Η ευελιξία του συγκεκριμένου αλγορίθμου, προστατεύει τις τιμές των βαρών που αποκτούν οι τα χαρακτηριστικά από ακραίες τιμές, γεγονός που ελαχιστοποιεί σημαντικά τις πιθανότητες του μοντέλου να κάνει overfit.

Για την εκμάθηση του μοντέλου χρησιμοποιήθηκε η «categorical_crossentropy» ως παράμετρος απώλειας (loss function). Στο πλαίσιο των νευρωνικών δικτύων, η παράμετρος απώλειας αναφέρεται στη συνάρτηση που χρησιμοποιείται για την ποσοτικοποίηση της ανομοιότητας μεταξύ της προβλεπόμενης εξόδου του μοντέλου και της πραγματικής εξόδου κατά τη διάρκεια της φάσης εκπαίδευσης. Η λειτουργία απώλειας παίζει καθοριστικό ρόλο στην καθοδήγηση του μοντέλου για την ελαχιστοποίηση των σφαλμάτων και την ενίσχυση των προβλεπτικών δυνατοτήτων του.



Εικόνα 22

Στην προκειμένη περίπτωση η «categorical_crossentropy loss function» υπολογίζει την απόκλιση του ποσοστού ομοιότητας που προβλέπει το μοντέλο από τα δεδομένα ελέγχου από την πραγματική κατηγορία.

Κατά την εκπαίδευση χρησιμοποιήθηκε «batch size» ίσο με 20, δηλαδή το μοντέλο – κατά στην εκπαίδευση του – χρησιμοποιούσε 20 δείγματα από τα δεδομένα εκπαίδευσης, με το μοντέλο να εξετάζει τα δεδομένα εκπαίδευσης συνολικά 4 φορές (epochs).

4.4 Αξιολόγηση Μοντέλου

Η αξιολόγηση της αποτελεσματικότητας του νευρωνικού δικτύου διεξάχθηκε σε 2 φάσεις. Στην πρώτη φάση αξιολόγησης εξετάστηκε το ποσοστό ακριβείας αναγνώρισης συγκεκριμένου επικίνδυνου οδηγικού ελιγμού σε σχέση με τις υπόλοιπες επικίνδυνες μανούβρες. Η 2^η φάση της αξιολόγησης αφορά την ικανότητα του μοντέλου να ξεχωρίζει μια επικίνδυνη μανούβρα από μια αμυντική.

Αξιολόγηση ταξινόμησης μεταξύ επικίνδυνων οδηγικών ελιγμών

Για να αξιολογηθεί η αποτελεσματικότητα με την οποία το μοντέλο μπορεί να ξεχωρίσει έναν οδηγικό ελιγμό, από τις υπόλοιπες οδηγικές κινήσεις, χρησιμοποιήθηκε το δείγμα των δεδομένων ελέγχου που έμεινε από το βασικό dataset – που συγκροτεί το 24% των δεδομένων. Αρχικά το μοντέλο αξιολογείται ως προς την ακρίβεια προβλέψεων κατά την εκπαίδευση του μοντέλου και ως προς την απώλεια (loss) .



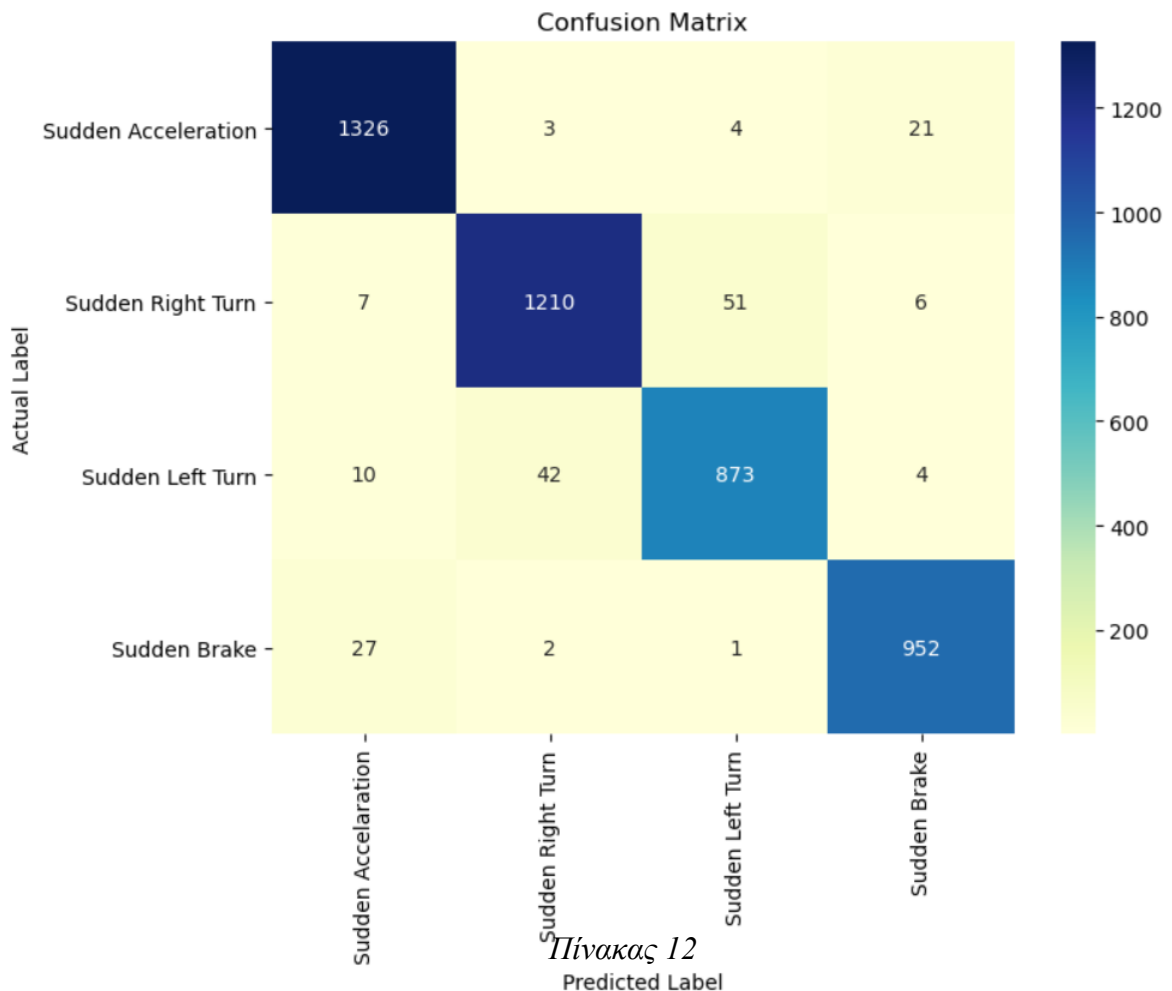
Εικόνα 23

Όπως φαίνεται στα 2 παραπάνω διαγράμματα, μετά τις 2 πρώτες εκπαιδεύσεις παρατηρήθηκε πως η τιμή της ακρίβειας αγγίζει το 99% με την τιμή της απώλειας να πέφτει κάτω από το 5%.

Αυτό υποδηλώνει πως τα δεδομένα εισόδου ασκούν μεγάλη επιρροή στην τελική τιμή της εξόδου, καθώς επίσης και την υψηλή ακρίβεια εκμάθησης του μοντέλου.

Μετά την μελέτη ακριβείας και απώλειας ακολουθεί η μελέτη του πίνακα σύγχυσης. Ο πίνακας σύγχυσης βοηθάει στην κατανόηση των ακριβών προβλέψεων που πραγματοποίησε το μοντέλο. Στην προκειμένη περίπτωση από τα 4.539 δείγματα που χρησιμοποιήθηκαν για έλεγχο:

- 1.354 δείγματα εκφράζανε απότομη επιτάχυνση
- 1.274 δείγματα εκφράζανε απότομη δεξιά στροφή
- 929 δείγματα εκφράζανε απότομη αριστερή στροφή
- 982 δείγματα εκφράζανε απότομο φρενάρισμα



Σε γενικότερες γραμμές το μοντέλο φαίνεται να πραγματοποιεί αρκετά ακριβείς προβλέψεις. Αξίζει να σημειωθεί ωστόσο, ότι το μοντέλο μπερδεύει τις αριστερές από τις δεξιές στροφές σε ένα ποσοστό 5%. Πέρα από αυτή την λεπτομέρεια το μοντέλο φαίνεται να μπορεί να αναγνωρίζει αρκετά αποτελεσματικά τους διάφορους ελιγμούς, που προκύπτουν από επιθετική οδήγηση.

Παρατήρηση: Όλες οι έξοδοι του μοντέλου από τα δεδομένα ελέγχου προβλέπανε με παραπάνω από 55% ποσοστό ομοιότητας με έναν συγκεκριμένο ελιγμό.

Αξιολόγηση αναγνώρισης επιθετικής συμπεριφοράς

Το τελευταίο στάδιο της αξιολόγησης αφορά την αξιολόγηση της ικανότητας του μοντέλου να αναγνωρίζει αν μια μανούβρα θεωρείται επιθετική ή όχι. Για την αξιολόγηση αυτή χρησιμοποιήθηκε το dataset από τους Ion Cojocar and Paul-Stefan Popescu του Πανεπιστημίου της Κραϊόβα.

Το εν λόγω dataset δεν περιέχει συγκεκριμένες κατηγορίες επιθετικών ελιγμών οδήγησης αλλά κατηγορίες στυλ οδήγησης (Επιθετική, κανονική, αργή). Σκοπός της εφαρμογής του μοντέλου στα δεδομένα του 2ου dataset είναι η διερεύνηση για το πώς το μοντέλο αντιμετωπίζει δεδομένα μη επιθετικής οδήγησης σε σύγκριση με τα δεδομένα επιθετικής οδήγησης.

Προ επεξεργασία δεδομένων

Το δεύτερο dataset δεν περιέχει τις παραπάνω παραμέτρους που είχαν υπολογιστεί στο βασικό dataset. Για να δημιουργηθούν τα αντίστοιχα υπολογίζονται οι τιμές των απαραίτητων

AccX	AccY	AccZ	GyroX	GyroY	GyroZ	Class	Timestamp	AccMeanX	AccMeanY	...	GyroMeanZ	GyroSumZ	GyroMaxZ
0.758194	-0.217791	0.457263	0.000000	0.000000	0.000000	AGGRESSIVE	818922	1.620288	-3.865569	...	0.033788	0.135154	0.225257
0.667560	-0.038610	0.231416	-0.054367	-0.007712	0.225257	AGGRESSIVE	818923	1.620288	-3.865569	...	0.033788	0.135154	0.225257
2.724449	-7.584121	2.390926	0.023824	0.013668	-0.038026	AGGRESSIVE	818923	1.620288	-3.865569	...	0.033788	0.135154	0.225257
2.330950	-7.621754	2.529024	0.056810	-0.180587	-0.052076	AGGRESSIVE	818924	1.620288	-3.865569	...	0.033788	0.135154	0.225257
2.847215	-6.755621	2.224640	-0.031765	-0.035201	0.035277	AGGRESSIVE	818924	0.892821	-2.036033	...	0.034208	0.136834	0.077427
...
0.915688	-2.017489	1.687505	0.450360	0.384845	-1.236468	SLOW	3583789	-1.270033	-1.282950	...	-0.298331	-1.193325	1.190500
-1.934203	0.914925	-0.096013	0.321468	0.649350	-0.477162	SLOW	3583790	-0.727225	0.533927	...	-0.106672	-0.426689	0.054291
-0.222845	0.747304	-0.887430	0.361174	-0.406836	0.054291	SLOW	3583790	-0.727225	0.533927	...	-0.106672	-0.426689	0.054291
-0.349423	0.067261	0.394368	-0.132405	0.020159	-0.004963	SLOW	3583791	-0.727225	0.533927	...	-0.106672	-0.426689	0.054291
-0.402428	0.406218	-0.423009	-0.053603	-0.006720	0.001145	SLOW	3583791	-0.727225	0.533927	...	-0.106672	-0.426689	0.054291

Εικόνα 24

παραμέτρων (π.χ. AccMeanX, GyroSumZ...) βάσει του χρονικού σημείου που πραγματοποιήθηκε η μέτρηση (timestamp). Ουσιαστικά τα δεδομένα είναι καταγεγραμμένα σε χρονολογική σειρά. Προκειμένου λοιπόν να έρθουν στην ίδια μορφή με τα δεδομένα του πρώτου dataset, ομαδοποιήθηκαν οι τιμές των ώστε να δημιουργηθούν τα αντίστοιχα δείγματα.

Αφού ολοκληρωθεί η προ επεξεργασία, του dataset, το μοντέλο προσπαθεί να αναγνωρίσει το ποσοστό ομοιότητας κάθε τιμής.

Το συγκεκριμένο dataset αποτελείται από 6.728 τιμές από τις οποίες:

- Οι 1.927 αφορούν επιθετική οδήγηση.
- Οι 2.197 αφορούν φυσιολογική οδήγηση.
- Οι 2.604 αφορούν αργή οδήγηση.

Για να επιτευχθεί ο στόχος που προαναφέρθηκε, τα δεδομένα που αποτελούσαν φυσιολογική και αργή οδήγηση ομαδοποιήθηκαν σε μια κατηγορία με συνολικό πληθυσμό 4.801. Μετά τον διαχωρισμό τούτο, το dataset χωρίστηκε σε 2 Dataframes (1 με επιθετική και 1 για μη επιθετική οδήγηση).

Εφαρμογή μοντέλου στο 2^ο dataset:

Μετά την εφαρμογή του μοντέλου στο 2^ο dataset διαπιστώθηκε ότι το μοντέλο φαίνεται να προβλέπει έναν από τους 4 ελιγμούς με μεγάλη ποσοστιαία διαφορά από τους άλλους 3

Δείγμα εξόδου επιθετικής οδήγησης → [0.68194395, 0.07812294, 0.227552 , 0.0123812]

Όταν όμως το ίδιο μοντέλο εφαρμόστηκε στα δεδομένα μη επιθετικής οδήγησης το μοντέλο παρουσίαζε μια σύγχυση στην εκτίμηση ποσοστού ομοιότητας με έναν συγκεκριμένο απότομο ελιγμό.

Δείγμα εξόδου μη επιθετικής οδήγησης → [0.38582784,0.15390920,0.25263148,0.20763148]

Πιο συγκεκριμένα από τις 4.801 προβλέψεις μη επιθετικής οδήγησης το μοντέλο εμφάνισε σύγχυση στις 3.996 ενώ στις υπόλοιπες 805 πραγματοποίησε προβλέψεις στις οποίες μια από τις 4 κατηγορίες ελιγμών είχε ένα από 4 πιθανά ποσοστά ομοιότητας 45% ή μεγαλύτερο.

Επιπλέον όταν εφαρμόστηκε το ίδιο μοντέλο στα δεδομένα επιθετικής οδήγησης, από τα 1.927 δεδομένα, το μοντέλο προέβλεψε έναν από τους 4 πιθανούς ελιγμούς με βεβαιότητα μεγαλύτερη του 45% τις 1.715 ενώ παρουσίασαν σύγχυση (όλα το ποσοστά ομοιότητας <45%) οι 212 από αυτές.

	Πλήθος	Προβλέψεις (>45%)	Προβλέψεις (<45%)
Aggressive	1927	1715	212
Not Aggressive	4801	805	3996

Πίνακας 13

4.5 Αξιολόγηση ακριβείας μοντέλου:

Για την ορθότερη αξιολόγηση του μοντέλου τόσο ως προς τη δυνατότητα αναγνώρισης μανούβρας όσο και στην γενικότερη ανίχνευση ύπαρξης επιθετικής οδηγικής συμπεριφοράς, χρησιμοποιήθηκαν:

- Το Precision score, το οποίο εκφράζει την ικανότητα του μοντέλου να μην ταξινομεί ένα δείγμα σε κατηγορία που δεν του ανήκει.

$$\text{Precision} = \frac{\text{True Positives}}{\text{True Positives} + \text{False Positives}}$$

- Το Recall score, που περιγράφει την ικανότητα του μοντέλου να προβλέψει θετικά τα δείγματα.

$$\text{Recall} = \frac{\text{True Positive}}{\text{True Positive} + \text{False Negative}}$$

- Το F1-score, το οποίο μετράει τον μέσο όρο του Recall και του Precision score.

$$\text{F1-score} = 2 * \left(\frac{\text{Precision} * \text{Recall}}{\text{Precision} + \text{Recall}} \right)$$

Αξιολόγηση ταξινόμησης μεταξύ επικίνδυνων ελιγμών :

	Precision	Recall	F1-score
Sudden Acceleration	0,9679	0,9793	0,9736
Sudden Right Turn	0,9626	0,9498	0,9561
Sudden Left Turn	0,9397	0,9397	0,9397
Sudden Braking	0,9685	0,9695	0,9690

Πίνακας 14

Συμπεράσματα:

Το νευρωνικό δίκτυο που δημιουργήθηκε φαίνεται να μπορεί να ταξινομεί με μεγάλη ακρίβεια ελιγμούς επιθετικής οδήγησης.

Επιπλέον παρατηρήθηκε πως σε δεδομένα μη επιθετικής οδήγησης το μοντέλο δεν μπορεί να ξεχωρίσει με σιγουριά έναν από τους 4 πιθανούς ελιγμούς. Το φαινόμενο αυτό ενδεχομένως οφείλεται στο ότι οι μετρήσεις του επιταχυνσιόμετρου και του γυροσκοπίου να μην βρίσκονται στο ίδιο εύρος τιμών με αυτές της επιθετικής οδήγησης. Το φαινόμενο αυτό θα μπορούσε να αποτελέσει εφελκτήριο για να αποκτήσει το μοντέλο την ικανότητα να ξεχωρίζει την επιθετική από τη μη επιθετική οδήγηση. Για τη πραγματοποίηση του εν λόγω εγχειρήματος, είναι απαραίτητο να συλλεχθούν περισσότερα και πιο πλήρη δεδομένα.

Αξίζει ακόμη να σημειωθεί το συγκεκριμένο μοντέλο λειτουργεί αποτελεσματικά σε βενζινοκίνητα αυτοκίνητα Β κατηγορίας (μικρά). Ενδεχομένως αυτοκίνητα με ισχυρότερους κινητήρες και διαφορετικό καύσιμο (diesel, LPG, CNG) να χρειάζονται ένα διαφορετικά δομημένο μοντέλο.

Ο τελικός στόχος του μοντέλου είναι να μπορέσει να αναγνωρίζει επιθετικούς ελιγμούς του οδηγού από τα δεδομένα που καταγράφουν τα επιταχυνσιόμετρα και τα γυροσκόπια των σύγχρονων αυτοκινήτων, έτσι ώστε ο οδηγός να ενημερώνεται όταν πραγματοποιεί κάποιον επιθετικό ελιγμό.

Βιβλιογραφία

- [1] Berry, I. M. (2007). The Effects of Driving Style and Vehicle Performance on the Real-World Fuel Consumption of U.S. Light-Duty Vehicles.
- [2] Castignani, G., Derrmann, T., Frank, R., & Engel, T. (2017). Smartphone-Based Adaptive Driving Maneuver Detection: A Large-Scale Evaluation Study. *IEEE Transactions on Intelligent Transportation Systems*, 18(9), 2330–2339. <https://doi.org/10.1109/TITS.2016.2646760>
- [3] Cho, W., & Choi, E. (2017). DTG big data analysis for fuel consumption estimation. *Journal of Information Processing Systems*, 13(2), 285–304. <https://doi.org/10.3745/JIPS.04.0031>
- [4] Cojocar, I., Popescu, P. S., & Mihăescu, M. C. (2022). Driver Behaviour Analysis based on Deep Learning Algorithms. *RoCHI - International Conference on Human-Computer Interaction*, 108–114. <https://doi.org/10.37789/rochi.2022.1.1.18>
- [5] Khademi, Z., Ebrahimi, F., & Kordy, H. M. (2023). A review of critical challenges in MI-BCI: From conventional to deep learning methods. In *Journal of Neuroscience Methods* (Vol. 383). Elsevier B.V. <https://doi.org/10.1016/j.jneumeth.2022.109736>
- [6] Krotov, V. (2018). *Legality and Ethics of Web Scraping*. <https://www.researchgate.net/publication/324907302>
- [7] Meseguer, J. E., Calafate, C. T., Cano, J. C., & Manzoni, P. (n.d.). *Assessing the Impact of Driving Behavior on Instantaneous Fuel Consumption*. <http://www.drivingstyles.info>.
- [8] Mullick, S., Pabitra, S., & Khilar, M. (n.d.). *Driver Behavior Profiling using Machine Learning*.
- [9] Loman, M., Šarkan, B., & Skrúčaný, T. (2021). Comparison of fuel consumption of a passenger car depending on the driving style of the driver. *Transportation Research Procedia*, 55, 458–465. <https://doi.org/10.1016/J.TRPRO.2021.07.009>
- [10] Nokes, T., Stephenson, S., Kaar, A.-L., Norris, J., Tweed, J., Brannigan, C., Sindano, H., & Scarbrough, T. (n.d.). Preparation for collection and monitoring of real-world fuel consumption data for light and heavy-duty vehicles | i Executive summary.
- [11] Paschalidis, E., Hajiseyedjavadi, F., Wei, C., Solernou, A., Hamish Jamson, A., Merat, N., Romano, R., & Boer, E. R. (2020). Deriving metrics of driving comfort for autonomous vehicles: A dynamic latent variable model of speed choice. *Analytic Methods in Accident Research*, 28. <https://doi.org/10.1016/j.amar.2020.100133>
- [12] Tsiakmakis, S., Fontaras, G., Cubito, C., Pavlovic, J., Anagnostopoulos, K., & Ciuffo, B. (2017). *From NEDC to WLTP: effect on the type-approval CO 2 emissions of light-duty vehicles*. doi.org/10.2760/93419
- [13] Zhao, L., Yao, W., Wang, Y., & Hu, J. (2020). Machine Learning-Based Method for Remaining Range Prediction of Electric Vehicles. *IEEE Access*, 8, 212423–212441. doi.org/10.1109/ACCESS.2020.3039815

- [14] de Vlieger, I., de Keukeleere, D., & Kretzschmar, J. G. (2000). Environmental effects of driving behaviour and congestion related to passenger cars. *Atmospheric Environment*, 34(27), 4649–4655. doi.org/10.1016/S1352-2310(00)00217-X
- [15] Allen, John Paul. "Gasoline direct injection: an enabling technology for future SI engines." SAE Technical Paper Series, 2000, doi:10.4271/2000-01-0253.
- [16] Gao, Bingzhao, et al. "A review of gasoline direct injection and the influence of injector deposits on spray and combustion." *Energy Conversion and Management*, vol. 157, 2018, pp. 389-410, doi:10.1016/j.enconman.2017.12.023
- [17] Kittelson, David B., and Tim J. Johnson. "Gasoline engine particulate matter emissions: effects of direct injection and fuel composition." *SAE International Journal of Engines*, vol. 1, no. 1, 2009, pp. 1317-1329, doi:10.4271/2008-01-1068.
- [18] Teece, D. J. (2018) "Tesla and the Reshaping of the Auto Industry," *Management and Organization Review*. Cambridge University Press, 14(3), pp. 501–512. doi: 10.1017/mor.2018.33.
- [19] Deur, Joško, et al. "Analysis of Electric Power Steering Systems and Their Potential for Energy Saving." *SAE International Journal of Passenger Cars - Mechanical Systems*, vol. 5, no. 2, 2012, pp. 957-967, doi:10.4271/2012-01-0589.
- [20] Kusano, Kristofer D., and Jon T. Hardy. "Evaluation of Lane Departure Warning Systems." *Traffic Injury Prevention*, vol. 6, no. 3, 2005, pp. 248-255.
- [21] Carlson, J. (2019). How Tesla created a new paradigm for the car industry. *Journal of Strategic Innovation and Sustainability*, 14(2), 100-112.
- [22] Lambert, F. (2019). Tesla disrupts the auto industry, but dealerships fight back: a legal battle ensues. *Journal of Innovation and Entrepreneurship*.
- [23] Luebkehan, C. (2017). Tesla Motors: A case study in disruptive innovation. *Journal of Business and Economics*,
- [24] Sheth, J. N., & Sisodia, R. S. (2015). The Tesla phenomenon: An investigation of the disruptive nature of Tesla Motors, Inc. *Journal of Business Research*
- [25] Anderson, James M., and Nidhi Kalra. "Characterizing the Safety and Security of Tesla's Autopilot System." (2021)
- [26] Sivaraman, Sushma, and Chris Urmson. "The prospects of autonomous vehicles: Myth versus reality." *Annual Review of Control, Robotics, and Autonomous Systems* 1 (2018)
- [27] Bojarski, Mariusz, et al. "End to end learning for self-driving cars." (2016)
- [28] Li, Yan, et al. "Deep reinforcement learning for autonomous driving." *IEEE Intelligent Transportation Systems Magazine* 9.4 (2017)
- [29] Koopman, Philip, et al. "Challenges in Autonomous Vehicle Testing and Validation." (2020).
- [30] Murphy, K. P. (2012). *Machine learning: a probabilistic perspective*. MIT Press.
- [31] Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning: data mining, inference, and prediction*. Springer.

- [32] Shibl M, Ismail L and Massoud A 2020 Machine Learning-Based Management of Electric Vehicles Charging: Towards Highly-Dispersed Fast Chargers *Energies* 13 5429 Online: <http://dx.doi.org/10.3390/en13205429>
- [33] McLaughlin, S., et al. "Intrusion Detection and Prevention in Cyber-Physical Systems: A Survey." *ACM Computing Surveys* (2016)
- [34] Chen, Y., et al. "A Survey of Machine Learning Techniques for Cybersecurity." *Proceedings of the 2018 International Conference on Computing and Artificial Intelligence*, 2018
- [35] Krotov, A., et al. "Anomaly Detection in Automotive Networks with Recurrent Neural Networks." *Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security*, 2016
- [36] Kazanas, P., et al. "Predicting Customer Satisfaction in the Automotive Industry using Machine Learning Techniques." *2019 IEEE International Conference on Big Data (Big Data)*, 2019
- [37] Xu, H., et al. "Predicting Users' Experience on Autonomous Vehicles Using Machine Learning Methods." *2018 IEEE 21st International Conference on Intelligent Transportation Systems (ITSC)*, 2018
- [38] Zhang, C., et al. "Predicting User Preferences for Vehicle Customization using Machine Learning Techniques." *2019 IEEE Intelligent Transportation Systems Conference (ITSC)*, 2019
- [39] Abualigah, L. M., et al. "User Experience Analysis for Intelligent Vehicles Using Machine Learning Techniques." *Information*, vol. 10, no. 11, 2019, article 343
- [40] Wang, D., et al. "Intelligent fault diagnosis and prognosis for electric vehicle driving motor based on machine learning techniques."
- [41] Lin, Y., et al. "Energy management strategy of plug-in hybrid electric vehicle based on driving pattern recognition and prediction." *Energies*, vol. 12, no. 4, 2019
- [42] Sivaraman, V., & Trivedi, M. M. "Guest editorial: Special issue on machine learning for intelligent transportation systems." *IEEE Transactions on Intelligent Transportation Systems*
- [43] Jia, S., Hui, F., Li, S., Zhao, X. and Khattak, A.J. (2020), Long short-term memory and convolutional neural network for abnormal driving behaviour recognition. *IET Intell. Transp. Syst.*, 14: 306-312. <https://doi.org/10.1049/iet-its.2019.0200>
- [44] R. Malekian et al. "Design and Implementation of a Wireless OBD II Fleet Management System." *IEEE Sensors Journal*, 17 (2017): 1154-1164. <https://doi.org/10.1109/JSEN.2016.2631542>.
- [45] B. Akin et al. "DSP Applications in Electric and Hybrid Electric Vehicles [In the Spotlight]." *IEEE Signal Processing Magazine*, 29 (2012): 136-133. <https://doi.org/10.1109/MSP.2012.2185863>
- [46] Javier E. Meseguer et al. "Assessing the impact of driving behavior on instantaneous fuel consumption." *2015 12th Annual IEEE Consumer Communications and Networking Conference (CCNC)* (2015): 443-448. <https://doi.org/10.1109/CCNC.2015.7158016>

Σύνδεσμοι:

[47] <https://www.spritmonitor.de/>

[48] <https://www.adac.de/>

[49] <https://global.toyota/en/>

[50] <https://digital-strategy.ec.europa.eu/en/>

[51] <https://eur-lex.europa.eu/legal-content/>

[52] <https://alternative-fuels-observatory.ec.europa.eu/>

[53] <https://www.thehenryford.org/collections-and-research/>

Μουσεία:

[54] Autoworld: National car museum, Brussels