

ΙΟΝΙΟ ΠΑΝΕΠΙΣΤΗΜΙΟ
ΤΜΗΜΑ ΠΛΗΡΟΦΟΡΙΚΗΣ



Πτυχιακή Εργασία Με Θέμα :
«Χρήση Μηχανικής Μάθησης Για την Ανάλυση
Δεδομένων Αεροπορικών Πτήσεων»

Σωτήριος Μήτσος

A.M: Π2017087

Επιβλέπων: Ανδρέας Καναβός

Επιβλέπων

Ανδρέας Καναβός

Επιτροπή

Στέργιος Παλαμάς

Άννα Σωτηροπούλου

Πίνακας Περιεχομένων

Εισαγωγή.....	7
ΚΕΦΑΛΑΙΟ 1 : ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ.....	10
1.1. Ορισμός Μηχανικής Μάθησης.....	10
1.2. Σημασία Μηχανικής Μάθησης	11
1.3. Τύποι Μηχανικής Μάθησης.....	12
1.3.1. Εποπτευόμενη Μάθηση	12
1.3.2. Μη Εποπτευόμενη Μάθηση	14
1.3.3. Ημι-Εποπτευόμενη Μάθηση.....	16
1.3.4. Ενισχυτική μάθηση.....	17
1.4. Εφαρμογές Μηχανικής Μάθησης	18
1.5. Προκλήσεις Μηχανικής Μάθησης.....	20
ΚΕΦΑΛΑΙΟ 2 : ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ ΚΑΙ ΑΕΡΟΠΟΡΙΚΗ ΒΙΟΜΗΧΑΝΙΑ	23
2.1. Εισαγωγή	23
2.2. Συμβολή της μηχανικής μάθησης στην αεροδιαστημική μηχανική	24
2.3. Συμβολή της μηχανικής μάθησης στην βελτίωση της εμπειρίας των πελατών	32
2.4. Προκλήσεις.....	43
ΚΕΦΑΛΑΙΟ 3 : ΠΕΙΡΑΜΑΤΙΚΟ ΤΜΗΜΑ ΕΡΓΑΣΙΑΣ – ΜΕΡΟΣ 1.....	48
3.1. Εισαγωγή – Βασική Φιλοσοφία	48
3.2. Παρουσίαση Συνόλου Δεδομένων.....	49
3.3. Προεπεξεργασία Δεδομένων	63
3.4. Χρησιμοποιούμενα Μοντέλα	64
3.4.1. Λογιστική Παλινδρόμηση.....	64
3.4.2. Decision Tree	67
3.4.3. Random Forest	69
3.4.4. Naïve Bayes.....	71
3.4.5. Αλγόριθμος k-NN.....	72
3.5. Χρησιμοποιούμενες Μετρικές Αξιολόγησης	74
3.5.1. Accuracy.....	74

3.5.2. Μετρική Precision.....	75
3.5.3. Μετρική Recall.....	75
ΚΕΦΑΛΑΙΟ 4 : ΠΕΙΡΑΜΑΤΙΚΟ ΤΜΗΜΑ ΕΡΓΑΣΙΑΣ – ΜΕΡΟΣ 2.....	77
4.1. Αποτελέσματα εφαρμογής μοντέλων	77
4.2. Συγκριτικά Αποτελέσματα	80
4.3. Σχολιασμός Αποτελεσμάτων.....	82
ΣΥΜΠΕΡΑΣΜΑΤΑ	83
ΠΑΡΑΡΤΗΜΑ ΚΩΔΙΚΑ.....	86
ΒΙΒΛΙΟΓΡΑΦΙΑ – ΠΗΓΕΣ	112

Λίστα Εικόνων

Εικόνα 1 : Μηχανική Μάθηση και Αεροδιαστημική Μηχανική [31]	25
Εικόνα 2 : Ιστόγραμμα συχνοτήτων της Flight Distance	54
Εικόνα 3 : Ιστόγραμμα συχνοτήτων της Departure Delay	54
Εικόνα 4 : Ιστόγραμμα συχνοτήτων της Arrival Delay.....	55
Εικόνα 5 : Ιστόγραμμα συχνοτήτων της Age	55
Εικόνα 6 : Box Plot Διάγραμμα της Flight Distance.....	56
Εικόνα 7 : Box Plot Διάγραμμα της Departure Delay	56
Εικόνα 8 : Box Plot Διάγραμμα της Age.....	57
Εικόνα 9 : Κατανομή εγγραφών για τη στήλη satisfaction	57
Εικόνα 10 : Κατανομή εγγραφών για τη στήλη Gender.....	58
Εικόνα 11 : Κατανομή εγγραφών για τη στήλη Gender Type	58
Εικόνα 12 : Κατανομή εγγραφών για τη στήλη Type of Travel	59
Εικόνα 13 : Κατανομή εγγραφών για τη στήλη Class.....	59
Εικόνα 14 : Συνδυαστικό γράφημα κατανομής εγγραφών για τις στήλες satisfaction και Gender	60
Εικόνα 15 : Συνδυαστικό γράφημα κατανομής για τις στήλες satisfaction και Customer Type..	60
Εικόνα 16 : Συνδυαστικό γράφημα κατανομής εγγραφών για τις στήλες satisfaction και Type of Travel.....	61
Εικόνα 17 : Συνδυαστικό γράφημα κατανομής εγγραφών για τις στήλες satisfaction και Class	61
Εικόνα 18 : γράφημα συσχέτισης μεταξύ των στηλών Departure Delay και satisfaction	62
Εικόνα 19 : Γράφημα συσχέτισης μεταξύ των στηλών Arrival Delay και satisfaction	62
Εικόνα 20 : Μητρώο Συσχέτισης μεταξύ των στηλών του συνόλου δεδομένων	63
Εικόνα 21 : Μοντέλο Λογιστικής Παλινδρόμησης.....	65
Εικόνα 22 : Decision Tree	67
Εικόνα 23 : Random Forest.....	70
Εικόνα 24 : Naive Bayes	71
Εικόνα 25 : Αλγόριθμος k-NN	73
Εικόνα 26 : Συγκριτικό γράφημα για τη μετρική Accuracy.....	81
Εικόνα 27 : Συγκριτικό γράφημα για μετρική Precision	81
Εικόνα 28 : Συγκριτικό γράφημα για τη μετρική Recall	82

Λίστα Πινάκων

Πίνακας 1 : Στήλες συνόλου δεδομένων	51
Πίνακας 2 : Διαστάσεις συνόλου δεδομένων.....	51
Πίνακας 3 : Πλήθος ελλιπών τιμών ανά στήλη του συνόλου δεδομένων	53
Πίνακας 4 : Στατιστικές ιδιότητες στηλών του συνόλου δεδομένων – Μέρος 1	53
Πίνακας 5 : Στατιστικές ιδιότητες στηλών του συνόλου δεδομένων - Μέρος 2.....	53
Πίνακας 6 : Αποτελέσματα μετρικής Accuracy	76
Πίνακας 7 : Αποτελέσματα μετρικής Precision	76
Πίνακας 8 : Αποτελέσματα μετρικής Recall.....	78
Πίνακας 9 : Μοντέλο με την καλύτερη επίδοση για κάθε μετρική.....	82
Πίνακας 10 : Μοντέλο με τη χειρότερη επίδοση σε κάθε μοντέλο	83

Εισαγωγή

Η μηχανική μάθηση έχει αναδειχθεί ως μετασχηματιστική δύναμη στον κλάδο των αερομεταφορών, προσφέροντας σημαντικές ευκαιρίες για καινοτομία, αποτελεσματικότητα και ασφάλεια. Με την αξιοποίηση τεράστιων ποσοτήτων δεδομένων που παράγονται από σύγχρονα αεροσκάφη, οι αλγόριθμοι μηχανικής μάθησης μπορούν να βελτιστοποιήσουν διάφορες πτυχές των πτητικών λειτουργιών, από την προγνωστική συντήρηση έως τη βελτιστοποίηση της διαδρομής. Η τεχνολογία αυτή επιτρέπει την ακριβέστερη πρόβλεψη πιθανών ζητημάτων, μειώνοντας έτσι τον κίνδυνο μη προγραμματισμένης συντήρησης και ενισχύοντας τη συνολική αξιοπιστία των αεροσκαφών. Η ενσωμάτωση της μηχανικής μάθησης στην αεροπορία όχι μόνο βελτιώνει την επιχειρησιακή αποτελεσματικότητα αλλά συμβάλλει επίσης σημαντικά στη μείωση του κόστους που συνδέεται με τη συντήρηση και το χρόνο διακοπής λειτουργίας του αεροσκάφους.

Μία από τις κρίσιμες εφαρμογές της μηχανικής μάθησης στην αεροπορία είναι η ενίσχυση της ασφάλειας μέσω προηγμένων αναλύσεων δεδομένων. Τα μοντέλα μηχανικής μάθησης μπορούν να αναλύσουν τα πρότυπα και τις ανωμαλίες στα δεδομένα πτήσης για να προβλέψουν και να αποτρέψουν πιθανούς κινδύνους για την ασφάλεια πριν εμφανιστούν. Για παράδειγμα, αυτά τα μοντέλα μπορούν να εντοπίσουν σημάδια μηχανικής βλάβης, κόπωσης του πιλότου ή ασυνήθιστων καιρικών συνθηκών που θα μπορούσαν να θέσουν σε κίνδυνο την ασφάλεια της πτήσης. Παρέχοντας έγκαιρες προειδοποιήσεις και χρήσιμες πληροφορίες, η μηχανική μάθηση βοηθά με τον τρόπο αυτό τις αεροπορικές εταιρείες και τους ρυθμιστικούς φορείς να εφαρμόσουν προληπτικά μέτρα, διασφαλίζοντας έτσι ασφαλέστερους ουρανούς τόσο για τους επιβάτες όσο και για το πλήρωμα.

Επιπλέον, η μηχανική μάθηση φέρνει επανάσταση στην εμπειρία των πελατών όσο αφορά κλάδο των αερομεταφορών. Οι εξατομικευμένες υπηρεσίες και οι συστάσεις που βασίζονται στις προτιμήσεις και τις συμπεριφορές των επιβατών γίνονται όλο και πιο συχνές. Οι αεροπορικές εταιρείες χρησιμοποιούν αλγόριθμους μηχανικής μάθησης για την ανάλυση δεδομένων των πελατών, επιτρέποντάς τους να προσαρμόζουν υπηρεσίες όπως ψυχαγωγία κατά την πτήση, προτιμήσεις γεύματος, ακόμη και προσφορές μάρκετινγκ. Αυτό το επίπεδο εξατομίκευσης όχι μόνο ενισχύει την ικανοποίηση των πελατών αλλά και ενισχύει την πίστη και ενισχύει τα έσοδα. Καθώς η αεροπορική βιομηχανία συνεχίζει να αγκαλιάζει τη μηχανική μάθηση, οι δυνατότητες για περαιτέρω καινοτομία και βελτίωση τόσο της λειτουργικής αποτελεσματικότητας όσο και της εμπειρίας των πελατών παραμένουν τεράστιες και συναρπαστικές.

Στόχος της συγκεκριμένης εργασίας είναι να θίξει ζητήματα που σχετίζονται με την εφαρμογή μηχανικής μάθησης στην αεροπορική βιομηχανία και πιο συγκεκριμένα με το σκέλος της ανάλυσης δεδομένων αεροπορικών πτήσεων. Η διάρθρωση της εργασίας έχει ως εξής:

Στο **πρώτο κεφάλαιο** παρουσιάζονται βασικές έννοιες που σχετίζονται με τη μηχανική μάθηση, δίνοντας έμφαση στους τύπους της, το πεδίο εφαρμογής της, καθώς και τις προκλήσεις που σχετίζονται με τη χρήση της.

Στο **δεύτερο κεφάλαιο** αναλύεται το πεδίο της εφαρμογής της μηχανικής μάθησης στην αεροπορική βιομηχανία, δίνοντας έμφαση στην ανάλυση των δεδομένων λειτουργίας των αεροσκαφών με στόχο την αποτελεσματικότερη και ασφαλή λειτουργία τους, αλλά και στην ανάλυση των δεδομένων των πελατών με στόχο την αποτελεσματικότερη και εξατομικευμένη εξυπηρέτησή τους κατά τη διάρκεια των πτήσεων.

Στο **τρίτο κεφάλαιο** γίνεται η παρουσίαση του πρακτικού μέρους της εργασίας, εστιάζοντας στην παρουσίαση του συνόλου δεδομένων που θα χρησιμοποιηθεί, στα μοντέλα μηχανικής μάθησης, αλλά και στις μετρικές αξιολόγησης.

Στο **τέταρτο κεφάλαιο** παρουσιάζονται τα αποτελέσματα της πειραματικής διαδικασίας, όπου καταγράφονται σημαντικά και ελπιδοφόρα αποτελέσματα. Η παρουσίαση γίνεται μέσω συγκριτικών πινάκων, αλλά και διαγραμμάτων.

Τέλος, εξάγονται τα συμπεράσματα και προτείνονται επεκτάσεις της παρούσας εργασίας.

ΚΕΦΑΛΑΙΟ 1 : ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ

1.1. Ορισμός Μηχανικής Μάθησης

Η μηχανική μάθηση (ML) είναι ένα επιστημονικό πεδίο της τεχνητής νοημοσύνης (AI) που επικεντρώνεται στην κατασκευή συστημάτων υπολογιστών που μαθαίνουν από δεδομένα [1 – 6]. Το ευρύ φάσμα τεχνικών που περιλαμβάνει η μηχανική μάθηση επιτρέπει στις εφαρμογές λογισμικού να βελτιώνουν την απόδοσή τους με την πάροδο του χρόνου.

Οι αλγόριθμοι μηχανικής μάθησης εκπαιδεύονται για να βρουν σχέσεις και μοτίβα στα δεδομένα. Χρησιμοποιούν ιστορικά δεδομένα ως είσοδο για να κάνουν προβλέψεις, να ταξινομήσουν πληροφορίες, να συγκεντρώσουν σημεία δεδομένων, να μειώσουν τη διαστατικότητα και ακόμη και να βοηθήσουν στη δημιουργία νέου περιεχομένου, όπως αποδεικνύεται από νέες εφαρμογές που χρησιμοποιούν μηχανική μάθηση, όπως το ChatGPT, το dall-E 2 και το GitHub Copilot.

Η μηχανική μάθηση είναι ευρέως εφαρμόσιμη σε πολλές βιομηχανίες. Οι μηχανές συστάσεων, για παράδειγμα, χρησιμοποιούνται από το ηλεκτρονικό εμπόριο, τα μέσα κοινωνικής δικτύωσης και τους οργανισμούς ειδήσεων ώστε να προτείνουν περιεχόμενο με βάση την προηγούμενη συμπεριφορά ενός πελάτη. Οι αλγόριθμοι μηχανικής μάθησης και η μηχανική όραση είναι ένα κρίσιμο συστατικό των αυτοκινούμενων αυτοκινήτων, βοηθώντας τα να πλοηγηθούν με ασφάλεια στους δρόμους. Στην υγειονομική περίθαλψη, η μηχανική μάθηση χρησιμοποιείται για τη διάγνωση και την πρόταση σχεδίων θεραπείας. Άλλες κοινές περιπτώσεις χρήσης μηχανικής μάθησης περιλαμβάνουν ανίχνευση απάτης, φιλτράρισμα ανεπιθύμητων μηνυμάτων, ανίχνευση απειλών κακόβουλου λογισμικού, προβλεπτική συντήρηση και αυτοματοποίηση επιχειρηματικών διαδικασιών.

Ενώ η μηχανική μάθηση είναι ένα ισχυρό εργαλείο για την επίλυση προβλημάτων, τη βελτίωση των επιχειρηματικών λειτουργιών και την αυτοματοποίηση εργασιών, είναι επίσης μια πολύπλοκη και προκλητική τεχνολογία, που απαιτεί βαθιά τεχνογνωσία και σημαντικούς πόρους. Η επιλογή του σωστού αλγορίθμου για μια εργασία απαιτεί μια ισχυρή κατανόηση των μαθηματικών και της στατιστικής. Οι αλγόριθμοι μηχανικής μάθησης συχνά απαιτούν μεγάλες ποσότητες δεδομένων καλής ποιότητας για την παραγωγή αποτελεσμάτων που θα χαρακτηρίζονται από ακρίβεια. Τα ίδια τα αποτελέσματα μπορεί να είναι δύσκολο να κατανοηθούν-ιδιαίτερα τα αποτελέσματα που παράγονται από πολύπλοκους αλγόριθμους, όπως τα νευρωνικά δίκτυα βαθιάς μάθησης που εμπνέουν τη λειτουργία τους από τον ανθρώπινο εγκέφαλο.

1.2. Σημασία Μηχανικής Μάθησης

Η μηχανική μάθηση έχει διαδραματίσει προοδευτικά κεντρικό ρόλο στην ανθρώπινη κοινωνία [7–10]. Η εκπαίδευση των μηχανών ώστε να μαθαίνουν από τα δεδομένα και να βελτιώνονται με την πάροδο του χρόνου επέτρεψε στους οργανισμούς να αυτοματοποιήσουν τις συνήθεις εργασίες, απελευθερώνοντάς τους για πιο δημιουργική και στρατηγική εργασία.

Η μηχανική μάθηση εκτελεί επίσης χειρωνακτικές εργασίες που είναι πέρα από την ικανότητά μας να εκτελούμε σε μεγάλη κλίμακα-για παράδειγμα, επεξεργασία των τεράστιων ποσοτήτων δεδομένων που παράγονται σήμερα από ψηφιακές συσκευές. Η ικανότητα της μηχανικής μάθησης να εξάγει μοτίβα και γνώσεις από τεράστια σύνολα δεδομένων έχει γίνει ανταγωνιστικός παίκτης σε τομείς που κυμαίνονται από τη χρηματοοικονομική και το λιανικό εμπόριο έως την υγειονομική περίθαλψη και την επιστημονική ανακάλυψη. Πολλές από τις σημερινές κορυφαίες εταιρείες, καθιστούν τη μηχανική μάθηση κεντρικό μέρος των δραστηριοτήτων τους.

Καθώς ο όγκος των δεδομένων που παράγονται από τις σύγχρονες κοινωνίες συνεχίζει να πολλαπλασιάζεται, η μηχανική μάθηση πιθανότατα θα γίνει ακόμη πιο ζωτική για τον άνθρωπο και απαραίτητη για την ίδια τη μηχανική νοημοσύνη. Η τεχνολογία όχι μόνο μας βοηθά να κατανοήσουμε τα δεδομένα που δημιουργούμε, αλλά επιπλέον η αφθονία των δεδομένων που δημιουργούμε ενισχύει περαιτέρω τις δυνατότητες μάθησης που βασίζονται στα δεδομένα της μηχανικής μάθησης.

Συμπερασματικά, η μηχανική μάθηση είναι ένας δρόμος προς την τεχνητή νοημοσύνη, η οποία με τη σειρά της τροφοδοτεί τις εξελίξεις στη μηχανική μάθηση, που βελτιώνουν επίσης την τεχνητή νοημοσύνη και σταδιακά θολώνουν τα όρια μεταξύ της μηχανικής νοημοσύνης και της ανθρώπινης νοημοσύνης.

1.3. Τύποι Μηχανικής Μάθησης

1.3.1. Εποπτευόμενη Μάθηση

Η εποπτευόμενη μάθηση περιλαμβάνει την εκπαίδευση μιας μηχανής και του αλγόριθμου της χρησιμοποιώντας επισημασμένα δεδομένα εκπαίδευσης και απαιτεί σημαντική ποσότητα ανθρώπινης καθοδήγησης [11–14]. Είναι μια από τις πιο δημοφιλείς μορφές μηχανικής μάθησης και είναι σε θέση να εκπαιδεύσει μοντέλα για την εκτέλεση εργασιών στην ταξινόμηση, την παλινδρόμηση ή την πρόβλεψη. Η εποπτευόμενη μάθηση χρησιμοποιείται συνήθως για τη δημιουργία συστημάτων συστάσεων, την ανίχνευση ανεπιθύμητων μηνυμάτων εισερχομένων και την πρόβλεψη των αξιών της αγοράς μετοχών και κατοικιών.

Για να λειτουργήσει, η εποπτευόμενη μάθηση απαιτεί σημαντική ανθρώπινη παρέμβαση λόγω της χρήσης επισημασμένων συνόλων δεδομένων. Τα δεδομένα

πρέπει να χωρίζονται σε χαρακτηριστικά (τα δεδομένα εισόδου) και ετικέτες (τα δεδομένα εξόδου) .

Τα χαρακτηριστικά περιγράφουν μεμονωμένες, μετρήσιμες μονάδες δεδομένων. Οι ετικέτες χρησιμοποιούνται για την ομαδοποίηση δεδομένων με συγκεκριμένα χαρακτηριστικά και συχνά εκχωρούνται χειροκίνητα από τους ανθρώπους για να βοηθήσουν στην εξήγηση του πλαισίου ορισμένων δεδομένων.

Με την εποπτευόμενη μάθηση, τα επισημασμένα δεδομένα εισόδου και εξόδου τροφοδοτούνται συνεχώς και τροφοδοτούνται εκ νέου σε συστήματα εκπαιδευμένα από τον άνθρωπο, προσφέροντας καθοδήγηση σε πραγματικό χρόνο. Αυτό βοηθά τις προβλέψεις να αυξάνουν την ακρίβεια μετά την τροφοδοσία κάθε νέου συνόλου δεδομένων στο σύστημα. Οι άνθρωποι παρέχουν επίσης ανατροφοδότηση σχετικά με την ακρίβεια του αλγορίθμου μηχανικής μάθησης κατά τη διάρκεια αυτής της διαδικασίας, η οποία τον βοηθά να μάθει με την πάροδο του χρόνου.

Η εποπτευόμενη μάθηση, όπως και κάθε ένας από αυτούς τους τύπους μηχανικής μάθησης, χρησιμεύει ως ομπρέλα για συγκεκριμένους αλγόριθμους και στατιστικές μεθόδους. Στη συνέχεια αναφέρονται παραδείγματα, που εμπίπτουν στην εποπτευόμενη μάθηση.

1.3.1.1. Ταξινόμηση

Με σκοπό την περαιτέρω κατηγοριοποίηση δεδομένων, οι αλγόριθμοι ταξινόμησης αποτελούν ένα εξαιρετικό εργαλείο που κινείται προς τη συγκεκριμένη κατεύθυνση. Μερικοί δημοφιλείς αλγόριθμοι ταξινόμησης που χρησιμοποιούνται για την ταξινόμηση δεδομένων περιλαμβάνουν τους : KNN, SVM, Decision Tree, Naive Bayes, και Random Forest.

1.3.1.2. Παλινδρόμηση

Οι αλγόριθμοι παλινδρόμησης χρησιμοποιούνται συχνά ως εργαλεία για την πρόβλεψη τάσεων. Αυτοί οι αλγόριθμοι προσδιορίζουν τις σχέσεις μεταξύ των αποτελεσμάτων και άλλων ανεξάρτητων μεταβλητών προκειμένου να πραγματοποιήσουν ακριβείς προβλέψεις. Οι αλγόριθμοι γραμμικής παλινδρόμησης είναι οι πιο ευρέως χρησιμοποιούμενοι, και ακολουθούν οι αλγόριθμοι λογιστικής παλινδρόμησης.

1.3.2. Μη Εποπτευόμενη Μάθηση

Με τη μη εποπτευόμενη μάθηση, τα ακατέργαστα δεδομένα, που δεν φέρουν ετικέτα, επεξεργάζονται από το σύστημα, πράγμα που σημαίνει λιγότερη εργασία για τον άνθρωπο [15– 18]. Οι αλγόριθμοι μάθησης χωρίς επίβλεψη ανακαλύπτουν μοτίβα ή ανωμαλίες σε μεγάλα, μη δομημένα σύνολα δεδομένων που διαφορετικά μπορεί να μην εντοπιστούν από τον άνθρωπο. Αυτή η δυνατότητα επιτρέπει την εκτέλεση εργασιών που σχετίζονται με την ομαδοποίηση ή τη μείωση των διαστάσεων.

Οι αλγόριθμοι μάθησης χωρίς επίβλεψη λειτουργούν αναλύοντας τα διαθέσιμα δεδομένα και ομαδοποιώντας πληροφορίες με βάση ομοιότητες και διαφορές, δημιουργώντας έτσι σχέσεις μεταξύ των σημείων δεδομένων. Η τμηματοποίηση των πελατών, η υπολογιστική όραση και η ανίχνευση παραβιάσεων μπορούν να εφαρμόσουν μη εποπτευόμενη μάθηση.

Οι δύο πιο συνηθισμένοι τύποι μάθησης χωρίς επίβλεψη παρατίθενται στη συνέχεια.

1.3.2.1. Ομαδοποίηση

Οι αλγόριθμοι ομαδοποίησης είναι το πιο ευρέως χρησιμοποιούμενο παράδειγμα μη εποπτευόμενης μηχανικής μάθησης. Αυτοί οι αλγόριθμοι επικεντρώνονται σε ομοιότητες εντός των ακατέργαστων δεδομένων και στη συνέχεια ομαδοποιούν αυτές τις πληροφορίες ανάλογα. Πιο απλά, αυτοί οι αλγόριθμοι παρέχουν δομή σε ακατέργαστα δεδομένα. Οι αλγόριθμοι ομαδοποίησης χρησιμοποιούνται συχνά σε δεδομένα μάρκετινγκ για να συγκεντρώσουν πληροφορίες πελατών (ή δυνητικών πελατών), καθώς και για τον εντοπισμό απάτης.

Ορισμένοι αλγόριθμοι ομαδοποίησης περιλαμβάνουν ιεραρχική ομαδοποίηση και ομαδοποίηση k-means.

1.3.2.2. Μείωση διαστάσεων

Η μείωση της διαστατικότητας είναι η διαδικασία μείωσης της ποσότητας των χαρακτηριστικών μέσα σε ένα σύνολο δεδομένων, διατηρώντας παράλληλα σημαντικές ιδιότητες των δεδομένων. Αυτό γίνεται για να μειωθεί ο χρόνος επεξεργασίας, ο χώρος αποθήκευσης, η πολυπλοκότητα και η υπερβολική εκπαίδευση σε ένα μοντέλο μηχανικής μάθησης.

Οι δύο κύριες μέθοδοι για την εφαρμογή της μείωσης των διαστάσεων περιλαμβάνουν την επιλογή χαρακτηριστικών και την εξαγωγή χαρακτηριστικών. Η επιλογή χαρακτηριστικών περιλαμβάνει την επιλογή ενός υποσύνολου σχετικών χαρακτηριστικών από το αρχικό σύνολο χαρακτηριστικών για χρήση ως είσοδος σε ένα μοντέλο, το οποίο βοηθά στην απλοποίηση του μοντέλου και στη βελτίωση της ακρίβειας των εξόδων. Η εξαγωγή χαρακτηριστικών περιλαμβάνει την εξαγωγή νέων, σημαντικών χαρακτηριστικών από τα αρχικά ακατέργαστα δεδομένα για εισαγωγή, η οποία επικεντρώνεται στην περικοπή περιττών δεδομένων και στην επιλογή των χαρακτηριστικών που θα βελτιώσουν περισσότερο την παραγωγή.

Οι δημοφιλείς αλγόριθμοι μείωσης διαστάσεων περιλαμβάνουν την ανάλυση κύριων συστατικών (PCA) , τη μη αρνητική παραγοντοποίηση πινάκων (NMF) , τη γραμμική ανάλυση διακρίσεων (LDA) και τη γενικευμένη ανάλυση διακρίσεων (GDAs) .

1.3.3. Ημι-Εποπτευόμενη Μάθηση

Η ημι-εποπτευόμενη μάθηση προσφέρει ένα ισορροπημένο μείγμα τόσο εποπτευόμενης όσο και μη εποπτευόμενης μάθησης [19–22]. Με την ημι-εποπτευόμενη μάθηση, λαμβάνεται μια υβριδική προσέγγιση καθώς μικρές ποσότητες επισημασμένων δεδομένων επεξεργάζονται παράλληλα με μεγαλύτερα κομμάτια ακατέργαστων δεδομένων. Αυτή η στρατηγική δίνει ουσιαστικά στους αλγόριθμους ένα προβάδισμα όσον αφορά τον εντοπισμό σχετικών προτύπων και την πραγματοποίηση ακριβών προβλέψεων σε σύγκριση με αλγόριθμους μάθησης χωρίς επίβλεψη, χωρίς το χρόνο, την προσπάθεια και το κόστος που σχετίζονται με αλγόριθμους εποπτευόμενης μάθησης.

Η ημι-εποπτευόμενη μάθηση χρησιμοποιείται συνήθως σε εφαρμογές που κυμαίνονται από την ανίχνευση απάτης έως την αναγνώριση ομιλίας καθώς και την ταξινόμηση εγγράφων κειμένου. Επειδή η ημι-εποπτευόμενη μάθηση χρησιμοποιεί επισημασμένα δεδομένα και μη επισημασμένα δεδομένα, συχνά βασίζεται σε τροποποιημένους εποπτευόμενους και μη εποπτευόμενους αλγόριθμους εκπαιδευμένους και για τους δύο τύπους δεδομένων.

Στη συνέχεια παρατίθενται δύο χαρακτηριστικές περιπτώσεις προσεγγίσεων

1.3.3.1. Αυτό-Εκπαίδευση

Οι αλγόριθμοι αυτοεκπαίδευσης χρησιμοποιούν ένα προϋπάρχον, εποπτευόμενο μοντέλο ταξινόμησης, που εκπαιδεύεται σε ένα μικρό μέρος των επισημασμένων δεδομένων. Ο ψευδο-ετικετοποιητής χρησιμοποιείται στη συνέχεια για να πραγματοποιήσει προβλέψεις για το υπόλοιπο του συνόλου δεδομένων, το οποίο δεν διαθέτει ετικέτα. Οι ετικέτες που παράγονται από αυτή τη διαδικασία ονομάζονται ψευδό-ετικέτες και προστίθενται πίσω στο σύνολο δεδομένων με ετικέτα. Αυτές οι ενέργειες πραγματοποιούνται επανειλημμένα από το μοντέλο μέχρι να επισημανθούν όλα τα δείγματα δεδομένων, βελτιώνοντας την ακρίβειά του με την πάροδο του χρόνου.

1.3.3.2. Διάδοση ετικετών

Οι αλγόριθμοι διάδοσης ετικετών εκχωρούν ετικέτες σε μη επισημασμένες παρατηρήσεις με την διάδοση ή την κατανομή ετικετών μέσω ενός συνόλου δεδομένων με την πάροδο του χρόνου, συνήθως σε ένα νευρωνικό δίκτυο γραφημάτων. Η διάδοση ετικετών μπορεί να χρησιμοποιηθεί για τον γρήγορο εντοπισμό κοινοτήτων, την ανίχνευση ανώμαλης συμπεριφοράς ή την επιτάχυνση εκστρατειών μάρκετινγκ. Για παράδειγμα, εάν σε έναν πελάτη σε ένα γράφημα αρέσει ένα συγκεκριμένο προϊόν, είναι πολύ πιθανόν να αρέσει και σε έναν άλλον πελάτη, ο οποίο διασυνδέεται μαζί του στο γράφημα.

1.3.4. Ενισχυτική μάθηση

Με την ενισχυτική μάθηση, τα προγράμματα είναι εξοπλισμένα με αισθητήρες, που συνήθως αναφέρονται ως ευφυείς πράκτορες, που ανταποκρίνονται στο περιβάλλον τους για να λαμβάνουν αποφάσεις ανεξάρτητα με το εάν επιτυγχάνουν ένα επιθυμητό αποτέλεσμα [23–25].

Οι ευφυείς πράκτορες αυτό-εκπαιδεύονται λαμβάνοντας ανταμοιβή για επιθυμητές συμπεριφορές ή τιμωρούνται για ανεπιθύμητες συμπεριφορές. Με την αντίληψη και την αλληλεπίδραση με το περιβάλλον τους, αυτοί οι πράκτορες μαθαίνουν μέσω δοκιμής και του λάθους, επιτυγχάνοντας τελικά τη βέλτιστη επάρκεια μέσω θετικής ενίσχυσης κατά τη διάρκεια της μαθησιακής διαδικασίας. Η ενισχυτική μάθηση χρησιμοποιείται συχνά στη ρομποτική και στα αυτό-οδηγούμενα αυτοκίνητα, βοηθώντας τις μηχανές να αποκτήσουν συγκεκριμένες δεξιότητες και συμπεριφορές.

1.4. Εφαρμογές Μηχανικής Μάθησης

Η μηχανική μάθηση (ML) έχει επιφέρει επανάσταση σε διάφορες βιομηχανίες αυτοματοποιώντας εργασίες, ενισχύοντας τις διαδικασίες λήψης αποφάσεων και παρέχοντας πληροφορίες μέσω ανάλυσης Δεδομένων [26 – 27]. Στην υγειονομική περίθαλψη, η μηχανική μάθηση χρησιμοποιείται για τη βελτίωση της διαγνωστικής ακρίβειας και της εξατομίκευσης της θεραπείας. Οι αλγόριθμοι μπορούν να αναλύσουν ιατρικές εικόνες για να ανιχνεύσουν ανωμαλίες όπως όγκους με μεγάλη ακρίβεια, συχνά ξεπερνώντας τους ανθρώπινους εμπειρογνώμονες. Επιπλέον, τα μοντέλα μηχανικής μάθησης μπορούν να προβλέψουν τα αποτελέσματα των ασθενών με βάση ιστορικά δεδομένα, βοηθώντας τους γιατρούς στη δημιουργία εξατομικευμένων σχεδίων θεραπείας. Αυτή η εφαρμογή όχι μόνο βελτιώνει τη φροντίδα των ασθενών αλλά και βελτιστοποιεί τους νοσοκομειακούς πόρους και μειώνει το κόστος.

Στον χρηματοπιστωτικό τομέα, οι αλγόριθμοι μηχανικής μάθησης μετασχηματίζουν τον τρόπο με τον οποίο οι εταιρείες διαχειρίζονται τους κινδύνους, εντοπίζουν την απάτη και ύποπτες συναλλαγές. Τα χρηματοπιστωτικά ιδρύματα χρησιμοποιούν μοντέλα μηχανικής μάθησης για να αναλύσουν τεράστιες ποσότητες δεδομένων

συναλλαγών ώστε να εντοπίσουν ύποπτες δραστηριότητες που θα μπορούσαν να υποδηλώνουν απάτη. Με αυτόν τον τρόπο, μπορούν να αποτρέψουν οικονομικά εγκλήματα σε πραγματικό χρόνο. Επιπλέον, η μηχανική μάθηση χρησιμοποιείται στην αλγοριθμική διαπραγμάτευση, που βοηθά στη λήψη αποφάσεων σε κλάσματα δευτερολέπτου με βάση τα δεδομένα της αγοράς για τη μεγιστοποίηση των κερδών. Η διαχείριση κινδύνων επωφελείται επίσης από τη μηχανική μάθηση παρέχοντας προγνωστικά αναλυτικά στοιχεία που αξιολογούν τους πιθανούς κινδύνους στις επενδύσεις και τις αλλαγές της αγοράς, βοηθώντας τις επιχειρήσεις να λαμβάνουν τεκμηριωμένες αποφάσεις.

Οι βιομηχανίες λιανικής και ηλεκτρονικού εμπορίου αξιοποιούν επιπλέον τη μηχανική μάθηση για να βελτιώσουν τις εμπειρίες των πελατών και να βελτιστοποιήσουν τις λειτουργίες τους. Τα συστήματα συστάσεων που υποστηρίζονται από τη μηχανική μάθηση αναλύουν τη συμπεριφορά και τις προτιμήσεις των πελατών ώστε να προτείνουν προϊόντα που είναι πιθανό να αγοράσει ένας χρήστης, αυξάνοντας έτσι τις πωλήσεις και την ικανοποίηση των πελατών. Η διαχείριση αποθεμάτων βλέπει επίσης σημαντικές βελτιώσεις μέσω αλγορίθμων μηχανικής μάθησης που προβλέπουν τη ζήτηση για προϊόντα, διασφαλίζοντας ότι τα επίπεδα αποθεμάτων διατηρούνται χωρίς υπερβολική αποθήκευση. Επιπλέον, η μηχανική μάθηση βοηθά στις στρατηγικές τιμολόγησης αναλύοντας την τιμολόγηση των ανταγωνιστών, τη ζήτηση των πελατών και τις τάσεις της αγοράς για τον καθορισμό βέλτιστων τιμών.

Τέλος, η μηχανική μάθηση είναι καθοριστικής σημασίας στον τομέα των αυτόνομων οχημάτων. Αυτά τα οχήματα βασίζονται σε αλγόριθμους μηχανικής μάθησης για την επεξεργασία δεδομένων από αισθητήρες και κάμερες, επιτρέποντάς τους να πλοηγούνται στους δρόμους, να αναγνωρίζουν σήματα κυκλοφορίας και να αποφεύγουν τα εμπόδια με ασφάλεια. Η ικανότητα συνεχούς μάθησης της

μηχανικής μάθησης επιτρέπει σε αυτά τα συστήματα να βελτιώνονται με την πάροδο του χρόνου, προσαρμόζοντας τις νέες συνθήκες οδήγησης και τα σενάρια. Πέρα από τα αυτοκίνητα, η μηχανική μάθηση εφαρμόζεται επίσης στον τομέα της εφοδιαστικής για τη βελτιστοποίηση διαδρομών και τη διαχείριση στόλου, εξασφαλίζοντας έγκαιρες παραδόσεις και μειώνοντας το λειτουργικό κόστος. Αυτή η ευρεία εφαρμογή της μηχανικής μάθησης στις μεταφορές όχι μόνο ενισχύει την ασφάλεια και την αποτελεσματικότητα αλλά και ανοίγει το δρόμο για καινοτόμες λύσεις κινητικότητας.

1.5. Προκλήσεις Μηχανικής Μάθησης

Ανεπαρκή Δεδομένα Εκπαίδευσης: Το κύριο ζήτημα που προκύπτει κατά τη χρήση αλγορίθμων μηχανικής μάθησης είναι η έλλειψη ποιότητας καθώς και ποσότητας δεδομένων. Αν και τα δεδομένα διαδραματίζουν ζωτικό ρόλο στην επεξεργασία αλγορίθμων μηχανικής μάθησης, πολλοί επιστήμονες δεδομένων ισχυρίζονται ότι τα ανεπαρκή δεδομένα, τα θορυβώδη δεδομένα και τα ακάθαρτα δεδομένα εξαντλούν εξαιρετικά τους αλγόριθμους μηχανικής μάθησης. Για παράδειγμα, μια απλή εργασία απαιτεί χιλιάδες δείγματα δεδομένων και μια προηγμένη εργασία όπως η αναγνώριση ομιλίας ή εικόνας χρειάζεται εκατομμύρια παραδείγματα δειγμάτων δεδομένων. Επιπλέον, η ποιότητα των δεδομένων είναι σημαντική για την ιδανική λειτουργία των αλγορίθμων, αλλά η απουσία ποιότητας δεδομένων βρίσκεται επίσης σε εφαρμογές μηχανικής μάθησης. Η ποιότητα των δεδομένων μπορεί να επηρεαστεί από ορισμένους παράγοντες ως εξής :

- **Θορυβώδη δεδομένα:** Είναι υπεύθυνα για μια ανακριβή πρόβλεψη που επηρεάζει την απόφαση καθώς και την ακρίβεια στις εργασίες ταξινόμησης.
- **Εσφαλμένα δεδομένα:** Είναι υπεύθυνα για ελαττωματικό προγραμματισμό και αποτελέσματα που λαμβάνονται σε μοντέλα μηχανικής μάθησης. Ως εκ

τούτου, λανθασμένα δεδομένα μπορεί να επηρεάσουν την ακρίβεια των αποτελεσμάτων επίσης.

- **Γενίκευση των δεδομένων εξόδου:**Μερικές φορές, διαπιστώνεται επίσης ότι η γενίκευση των δεδομένων εξόδου γίνεται πολύπλοκη, γεγονός που έχει ως αποτέλεσμα συγκριτικά κακές μελλοντικές ενέργειες.

Μη αντιπροσωπευτικά δεδομένα εκπαίδευσης:Για να βεβαιωθούμε ότι το μοντέλο εκπαίδευσης γενικεύεται καλά ή όχι, πρέπει να διασφαλίσουμε ότι τα δείγματα δεδομένων εκπαίδευσης πρέπει να είναι αντιπροσωπευτικά των νέων περιπτώσεων που πρέπει να γενικεύσουμε. Τα δεδομένα εκπαίδευσης πρέπει να καλύπτουν όλες τις περιπτώσεις.

Επιπλέον, εάν χρησιμοποιούνται μη αντιπροσωπευτικά δεδομένα εκπαίδευσης στο μοντέλο, αυτό οδηγεί σε λιγότερο ακριβείς προβλέψεις. Ένα μοντέλο μηχανικής μάθησης λέγεται ότι είναι ιδανικό αν προβλέπει καλά για γενικευμένες περιπτώσεις και παρέχει ακριβείς αποφάσεις. Εάν υπάρχουν λιγότερα δεδομένα εκπαίδευσης, τότε θα υπάρχει θόρυβος δειγματοληψίας στο μοντέλο, που ονομάζεται μη αντιπροσωπευτικό σύνολο εκπαίδευσης. Δεν θα είναι ακριβές στις προβλέψεις.

Υπερεκπαίδευση:Η υπερεκπαίδευση είναι ένα από τα πιο κοινά ζητήματα που αντιμετωπίζουν οι μηχανικοί και οι επιστήμονες δεδομένων. Κάθε φορά που ένα μοντέλο μηχανικής μάθησης εκπαιδεύεται με τεράστιο όγκο δεδομένων, αρχίζει να καταγράφει θόρυβο και ανακριβή δεδομένα στο σύνολο δεδομένων εκπαίδευσης. Επηρεάζει αρνητικά την απόδοση του μοντέλου. Ο κύριος λόγος πίσω από την υπερβολική εκπαίδευση είναι η χρήση μη γραμμικών μεθόδων που χρησιμοποιούνται σε αλγόριθμους μηχανικής μάθησης καθώς δημιουργούν μη ρεαλιστικά μοντέλα δεδομένων.

Υποεκπαίδευση: Η υποεκπαίδευση είναι ακριβώς το αντίθετο της υπερβολικής εκπαίδευσης. Κάθε φορά που ένα μοντέλο μηχανικής μάθησης εκπαιδεύεται με λιγότερες ποσότητες δεδομένων, παρέχει ελλιπή και ανακριβή δεδομένα και καταστρέφει την ακρίβεια του μοντέλου μηχανικής μάθησης.

Η υποεκπαίδευση συμβαίνει όταν το μοντέλο είναι πολύ απλό για να κατανοήσει τη βασική δομή των δεδομένων. Αυτό συμβαίνει γενικά όταν έχουμε περιορισμένα δεδομένα στο σύνολο δεδομένων και προσπαθούμε να δημιουργήσουμε ένα γραμμικό μοντέλο με μη γραμμικά δεδομένα. Σε τέτοια σενάρια, οι κανόνες του μοντέλου μηχανικής μάθησης γίνονται πολύ εύκολοι για να εφαρμοστούν σε αυτό το σύνολο δεδομένων και το μοντέλο αρχίζει να πραγματοποιεί λανθασμένες προβλέψεις.

Πολυπλοκότητα διαδικασίας της μηχανικής μάθησης : Η διαδικασία μηχανικής μάθησης είναι πολύ περίπλοκη, γεγονός το οποίο επίσης ένα άλλο σημαντικό ζήτημα που αντιμετωπίζουν οι μηχανικοί και οι επιστήμονες δεδομένων. Ωστόσο, η μηχανική μάθηση και η τεχνητή νοημοσύνη είναι πολύ νέες τεχνολογίες, αλλά βρίσκονται ακόμη σε πειραματική φάση και αλλάζουν συνεχώς με την πάροδο του χρόνου. Επιπλέον, περιλαμβάνει την ανάλυση των δεδομένων, την αφαίρεση της μεροληψίας δεδομένων, τα δεδομένα εκπαίδευσης, την εφαρμογή σύνθετων μαθηματικών υπολογισμών κ.λπ., καθιστώντας τη διαδικασία πιο περίπλοκη και αρκετά κουραστική [28–30].

ΚΕΦΑΛΑΙΟ 2 : ΜΗΧΑΝΙΚΗ ΜΑΘΗΣΗ ΚΑΙ ΑΕΡΟΠΟΡΙΚΗ ΒΙΟΜΗΧΑΝΙΑ

2.1. Εισαγωγή

Η αεροπορική βιομηχανία έχει υποστεί τεράστιο μετασχηματισμό καθώς η τεχνολογία έχει προχωρήσει και έχουν αναπτυχθεί νέες ψηφιακές δυνατότητες. Οι έξυπνες λύσεις μπορούν να ενισχύσουν την αποτελεσματικότητα, να μειώσουν το κόστος και να ενισχύσουν την παραγωγικότητα στον βιομηχανικό τομέα. Τα προηγμένα συστήματα ενσωματώνουν μια ποικιλία τεχνολογιών αιχμής, συμπεριλαμβανομένης της αυτοματοποίησης, της ρομποτικής, της τεχνητής νοημοσύνης (AI) , της μηχανικής μάθησης, της μικτής πραγματικότητας και του Διαδικτύου των πραγμάτων (IoT) . Η ψηφιοποίηση έχει αλλάξει το παράδειγμα της βιομηχανίας για την έξυπνη αεροπορία. Οι πρόσφατα υιοθετημένες καινοτόμες ψηφιακές προσεγγίσεις προωθούν την αποτελεσματικότητα, την ασφάλεια στη λειτουργική διαδικασία και αυξάνουν την ικανοποίηση των επιβατών κατανοώντας καλύτερα τις ανάγκες, τις προτιμήσεις και τις συνήθειές τους. Η ψηφιοποίηση επιπλέον έχει ενισχύσει τη συνεργασία και την επικοινωνία μεταξύ των αεροπορικών εταιρειών, των αεροδρομίων και άλλων φορέων της αεροπορίας. Με τη μηχανική μάθηση, η αεροπορική βιομηχανία έχει την ευκαιρία να μεταμορφώσει την αεροδιαστημική μηχανική και να βελτιώσει την εμπειρία των επιβατών [31- 34].

Η μηχανική μάθηση είναι ζωτικής σημασίας για την ψηφιοποίηση, την ερμηνεία και τον εντοπισμό χαρακτηριστικών, προτύπων και τάσεων στα ψηφιακά δεδομένα για την απόκτηση πολύτιμων γνώσεων και τη λήψη τεκμηριωμένων αποφάσεων. Παρέχει ισχυρά εργαλεία για τη δημιουργία αποτελεσματικών, αξιόπιστων και ασφαλών σχεδίων αεροσκαφών, κατασκευής και εκπαίδευσης. Οι εφαρμογές μηχανικής μάθησης στον σχεδιασμό, στην παραγωγή, στην επαλήθευση και

επικύρωση και στις υπηρεσίες έχουν αυξήσει τον αυτοματισμό και τις εξορθολογισμένες διαδικασίες στην αεροπορική βιομηχανία. Η μετασχηματιστική μηχανική μάθηση έχει επίδραση στην κατασκευή, την αυτοματοποίηση και την ανάλυση δεδομένων στην αεροπορική βιομηχανία με τη χρήση ψηφιακής μοντελοποίησης και προσομοίωσης. Η πλούσια σε δεδομένα αεροπορική βιομηχανία είναι έτοιμη να αξιοποιήσει την επανάσταση της μηχανικής μάθησης. Η μηχανική μάθηση βελτιστοποιεί τα δίκτυα μεταφορών, προβλέπει τη συμπεριφορά των πελατών και παρέχει προσαρμοσμένες υπηρεσίες για τη βελτίωση της εμπειρίας των επιβατών. Οι αεροπορικές εταιρείες μπορούν να βελτιστοποιήσουν τις λειτουργίες, να βελτιώσουν την αφοσίωση και να αυξήσουν τα έσοδα αναλύοντας τα δεδομένα των επιβατών. Επιπλέον, η μηχανική μάθηση παρακολουθεί και αναλύει πολλές φάσεις διέλευσης επιβατών, συμπεριλαμβανομένων των περιόδων άφιξης, αναχώρησης και αναμονής για τη μείωση των καθυστερήσεων και τη βελτίωση της εμπειρίας των πελατών. Επιπροσθέτως, οι αεροπορικές εταιρείες χρησιμοποιούν κινητές εφαρμογές για να προσφέρουν ενημερώσεις σε πραγματικό χρόνο και εξατομικευμένες υπηρεσίες. Έξυπνες υπηρεσίες όπως προσαρμοσμένα εισιτήρια, παρακολούθηση αποσκευών και παρακολούθηση πτήσεων παρέχονται από αεροπορικές εταιρείες [35– 39].

2.2. Συμβολή της μηχανικής μάθησης στην αεροδιαστημική μηχανική

Η μηχανική μάθηση παρέχει ισχυρά εργαλεία ανάλυσης και βελτιστοποίησης για πολύπλοκα προβλήματα στην αεροδιαστημική μηχανική. Η προηγμένη μηχανική μάθηση στον αεροδιαστημικό σχεδιασμό, την κατασκευή, τις δοκιμές και τις υπηρεσίες καλύπτονται σε αυτήν την ενότητα.



Εικόνα 1 : Μηχανική Μάθηση και Αεροδιαστημική Μηχανική [31]

Η μηχανική μάθηση έχει γίνει αναπόσπαστο μέρος του σύγχρονου σχεδιασμού αεροσκαφών. Οι σχεδιαστές αεροσκαφών μπορούν να δημιουργήσουν πιο αποτελεσματικά και ασφαλέστερα αεροσκάφη, καθώς και να βελτιστοποιήσουν την απόδοση των αεροσκαφών, χρησιμοποιώντας ισχυρούς αλγόριθμους. Ο πολλαπλασιασμός της επιστήμης των δεδομένων και της μηχανικής που βασίζεται σε μοντέλα έχει καταστήσει δυνατό τον σύγχρονο σχεδιασμό. Ο πολυεπιστημονικός σχεδιασμός χρησιμοποιεί ανάλυση δεδομένων και μείωση διαστάσεων για την αντιμετώπιση παραμέτρων σχεδιασμού υψηλής διάστασης. Η μείωση των δεδομένων βελτιώνει την ακρίβεια υπολογισμού και τις διεπιστημονικές αλληλεπιδράσεις στην πρόωση αεριωθούμενων αεροσκαφών, τα φτερά και την σύνθετη. Ο σχεδιασμός ενός αεροσκάφους είναι ένα πρόβλημα βελτιστοποίησης πολλαπλών στόχων. Οι αντικειμενικές συναρτήσεις, οι αλγόριθμοι πρόβλεψης και

τα αναλυτικά εργαλεία έχουν μεταμορφώσει την αεροδυναμική, τις δομές και τα συστήματα ελέγχου στο σχεδιασμό αεροσκαφών [40–42].

Η εφαρμογή της μηχανικής μάθησης στην βελτιστοποίηση της αεροδυναμικής έχει ως αποτέλεσμα πιο αποδοτική, σταθερή και εύκολη στον έλεγχο πτήση. Τα δεδομένα που συλλέγονται από δοκιμές αεροδυναμικής σήραγγας, δοκιμές πτήσης και προσομοιώσεις εκπαιδεύονται σε ένα μοντέλο μηχανικής μάθησης, το οποίο μπορεί στη συνέχεια να χρησιμοποιηθεί για τον προσδιορισμό των βέλτιστων παραμέτρων σχεδιασμού. Αυτές οι παράμετροι χρησιμοποιούνται για τη δημιουργία τρισδιάστατων μοντέλων του αεροσκάφους, τα οποία στη συνέχεια δοκιμάζονται σε εικονικό περιβάλλον για να εξασφαλιστεί η επιθυμητή αεροδυναμική απόδοση. Η μηχανική μάθηση μπορεί να δημιουργήσει νέες δομές αεροσκαφών και να βελτιστοποιήσει τις επιδόσεις τους για πολύ ελαφρύτερα και ισχυρότερα αεροσκάφη. Ένα μεγάλο σύνολο δεδομένων των υφιστάμενων δομών αεροσκαφών, συμπεριλαμβανομένων των υλικών, της αεροδυναμικής, της δομής και του κόστους, συλλέγεται και αναλύεται με τη χρήση μηχανικής μάθησης. Με βάση τα πρότυπα που προσδιορίζονται στο σύνολο δεδομένων, το εκπαιδευμένο μοντέλο δημιουργεί νέες δομές αεροσκαφών. Οι παραγόμενες δομές αεροσκαφών αξιολογούνται σε περιβάλλον προσομοίωσης υπολογιστή. Οι σχεδιασμένες δομές αεροσκαφών επιτυγχάνουν την επιθυμητή απόδοση και ασφάλεια μέσω μιας επαναληπτικής διαδικασίας. Η μηχανική μάθηση μπορεί να δημιουργήσει και να βελτιστοποιήσει συστήματα ελέγχου που κάνουν τα αεροσκάφη πιο ευέλικτα. Συλλέγονται και αναλύονται δεδομένα από το σύστημα ελέγχου αεροσκαφών, όπως αρχεία επιδόσεων και ασφάλειας αεροσκαφών. Ένα μαθηματικό μοντέλο σχεδιάζεται για συστήματα ελέγχου αεροσκαφών ώστε να αποκαλύψει τις σχέσεις μεταξύ μεταβλητών εισόδου και εξόδου. Το εκπαιδευμένο μοντέλο δοκιμάζεται σε προσομοιωτή ή πραγματικό αεροσκάφος για περαιτέρω προσαρμογές [43–47].

Η μηχανική που βασίζεται σε μοντέλα υποστηρίζει ακριβή πρόβλεψη και βελτιστοποίηση σύνθετων σχεδίων αεροσκαφών. Το ψηφιακό δίδυμο συλλαμβάνει τις φυσικές πληροφορίες κόσμου ακριβώς και εξασφαλίζει τυποποιημένο σχήμα στοιχείων. Διαφορετικά σήματα από εκατοντάδες υποσυστήματα μοντελοποιούνται και τα φυσικά λειτουργικά συστήματα αντιπροσωπεύονται αποτελεσματικά με εικονικές απεικονίσεις. Η Airbus και η Boeing επενδύουν σε μεγάλο βαθμό σε πλατφόρμες μηχανικής μάθησης. Η Airbus δημιούργησε λύσεις σχεδιασμού αεροσκαφών που συνδυάζουν 3D μοντελοποίηση και μηχανική μάθηση. Οι λύσεις μπορούν να αναλύσουν εξαρτήματα αεροσκαφών όπως φτερά και ατράκτους, να βελτιστοποιήσουν το βάρος, τις επιδόσεις και άλλες μετρήσεις για να καθορίσουν τον καλύτερο σχεδιασμό αεροσκαφών και να διαγνώσουν πιθανά προβλήματα των αεροσκαφών. Η Boeing χρησιμοποιεί μηχανική μάθηση για να αναλύσει το σχεδιασμό διαφόρων εξαρτημάτων αεροσκαφών, επιτρέποντας στους μηχανικούς να σχεδιάζουν εξαρτήματα αεροσκαφών γρήγορα και με ακρίβεια. Η χρήση της μηχανικής μάθησης προκαλεί μια αλλαγή παραδείγματος στο σχεδιασμό αεροσκαφών, με αποτέλεσμα τη βελτίωση των επιδόσεων των αεροσκαφών και την ενίσχυση της ασφάλειας για το μέλλον της αεροπορίας[48, 49].

Η κατασκευή αεροσκαφών είναι μια πολύπλοκη διαδικασία που περιλαμβάνει την κατασκευή εξαρτημάτων, τη συναρμολόγηση, τη δοκιμή και την επιθεώρηση. Η μηχανική μάθηση μειώνει το κόστος, βελτιώνει την ασφάλεια και ενισχύει την αποτελεσματικότητα στην κατασκευή αεροσκαφών. Στον έλεγχο της διαδικασίας, την αυτόνομη ανίχνευση και την επιλογή υλικού, η μηχανική μάθηση παίζει σημαντικό ρόλο. Η μηχανική μάθηση εντοπίζει τάσεις, ανιχνεύει μη φυσιολογική συμπεριφορά, προτείνει διορθωτικές ενέργειες, συνιστά βέλτιστες ακολουθίες συναρμολόγησης, αναλαμβάνει την αναμενόμενη απόδοση και προτείνει πιθανές

τροποποιήσεις, επιτρέποντας στους κατασκευαστές να λαμβάνουν καλύτερες αποφάσεις [50, 51].

Επιπλέον, η μηχανική μάθηση επαναπροσδιορίζει τις τυπικές μορφές δεδομένων και βοηθά στην ισχυρή μεταφορά δεδομένων μεταξύ πλατφορμών. Η ανάλυση στοιχείων των παραμέτρων μερών των αεροσκαφών συμπυκνώνει τον κύκλο κατασκευής. Οι παράμετροι των εξαρτημάτων χρησιμοποιούνται στην τυποποίηση των αεροσκαφών, η οποία έχει ενημερώσει τις μελλοντικές αποφάσεις και τις εξορθολογισμένες διαδικασίες. Η μηχανική μάθηση χρησιμοποιεί αισθητήρες και αρχεία καταγραφής δεδομένων για την ανίχνευση ανωμαλιών στη διαδικασία, τη μείωση του χρόνου διακοπής λειτουργίας και τη βελτίωση της ασφάλειας και της αξιοπιστίας του αεροσκάφους. Οι αυτοματοποιημένες μηχανές παράγουν τα λεπτομερή μέρη αεροσκαφών με την υψηλότερη ακρίβεια και το χαμηλότερο κόστος.

Στην αυτόνομη ανίχνευση, η μηχανική μάθηση παρακολουθεί και ελέγχει τις διαδικασίες παραγωγής. Η μηχανική μάθηση αναγνωρίζει τα πρότυπα στα δεδομένα, αναλύει τις αποκλίσεις από την αναμενόμενη παραγωγή και εντοπίζει αόρατα κατασκευαστικά ελαττώματα όπως ρωγμές και στρέβλωση. Για να βελτιωθεί η ακρίβεια και η αξιοπιστία της κατασκευής αεροσκαφών, η εποπτευόμενη μάθηση, η μη εποπτευόμενη μάθηση και η βαθιά μάθηση χρησιμοποιούνται στην αυτόνομη ανίχνευση. Η εκμάθηση των μηχανών μετρά το λάθος για να ελαχιστοποιήσει τους χρόνους λειτουργίας και να αυξήσει την αποδοτικότητα των χειριστών. Οι αισθητήρες ανιχνεύουν σφάλματα σε πραγματικό χρόνο, αναγνωρίζουν επαναλαμβανόμενα μοτίβα και προβλέπουν ανωμαλίες. Οι ακραίες τιμές ανιχνεύονται χρησιμοποιώντας ανάλυση κύριων συστατικών και εξαγωγή χαρακτηριστικών για τη βελτίωση της ποιότητας των αυτοματοποιημένων επιθεωρήσεων [52– 54].

Η μηχανική μάθηση βελτιώνει την επιλογή υλικού ανιχνεύοντας αλλαγές παραμέτρων, φιλτράροντας κατάλληλα υλικά, συντομεύοντας τους χρόνους δοκιμών και καθορίζοντας τις καλύτερες επιλογές. Τα ψηφιακά μοντέλα χρησιμοποιούνται για την προσομοίωση νέων υλικών ή δομών με βάση διάφορα κριτήρια όπως το κόστος, η αντοχή, το βάρος και η ανθεκτικότητα. Η παλινδρόμηση χρησιμοποιείται για να διερευνήσει τη σχέση μεταξύ διαφόρων παραμέτρων όπως ο χαρακτηρισμός του υλικού, ο δομικός χαρακτηρισμός, η τριβή, η αντίσταση στην πίεση και η ευαισθησία στη θερμοκρασία. Η χρήση ψηφιακών προσομοιώσεων μειώνει σημαντικά το κόστος ανάπτυξης νέων υλικών. Τέλος, η υπολογιστική όραση και η θερμική ανίχνευση χρησιμοποιούνται στη διάγνωση νέων υλικών και στην ανίχνευση ελαττωμάτων [55– 57].

Είναι αδιαμφισβήτητο το γεγονός ότι τα συστήματα που βασίζονται στη μηχανική μάθηση παίρνουν σταδιακά τη θέση της χειρωνακτικής εργασίας. Για παράδειγμα, η GE Aviation χρησιμοποιεί μηχανική μάθηση στην κατασκευή αεροσκαφών. Τα εργαλεία ανάλυσης που λειτουργούν με τεχνητή νοημοσύνη χρησιμοποιούνται για την προγνωστική συντήρηση και τη βελτιστοποίηση της απόδοσης των κινητήρων αεριωθούμενων αεροσκαφών. Η GE Aviation χρησιμοποιεί επίσης μηχανική μάθηση για τη βελτίωση της αποτελεσματικότητας των γραμμών παραγωγής κινητήρων αεριωθούμενων αεροσκαφών [58, 59].

Η μηχανική μάθηση παρέχει αποτελεσματική και ακριβή ανάλυση δεδομένων για δοκιμές αεροσκαφών, επιτρέποντας την ανάπτυξη καλύτερων μοντέλων αεροσκαφών και τη βελτίωση της απόδοσης των αεροσκαφών. Οι μεθοδολογίες που βασίζονται σε δεδομένα εξορθολογίζουν τη διαδικασία δοκιμών, διασφαλίζοντας παράλληλα την ποιότητα των δοκιμών. Η μηχανική μάθηση χρησιμοποιείται σε επίπεδο υποσυστήματος και ολοκληρωμένου συστήματος για τη μείωση της διάστασης των δεδομένων που συλλέγονται από αισθητήρες, την εξαγωγή κύριων

χαρακτηριστικών, την ανάλυση σχέσεων παραμέτρων και την απεικόνιση βελτιστοποιημένων προτύπων δεδομένων. Υπάρχουν πολλοί διαφορετικοί τύποι αεροσκαφών και δεδομένων σήματος, όπως επιταχυνσιόμετρα, αισθητήρες θερμοκρασίας και αισθητήρες πίεσης. Κάθε διαδικασία του σύνθετου και δυναμικού συστήματος ολοκλήρωσης έχει βελτιωθεί σημαντικά μέσω της μοντελοποίησης δεδομένων και της βελτιστοποίησης περιορισμών. Η αυτόματη επαλήθευση και επικύρωση έχουν μειώσει την ανάγκη για χρονοβόρα χειρωνακτική εργασία και έχουν μετριάσει το πρόβλημα της έλλειψης προσωπικού. Μια μεγάλη ποσότητα διακριτών δεδομένων σε πραγματικό χρόνο επεξεργάζεται η αεροπορική βιομηχανία για να εξασφαλίσει την ασφάλεια των πτήσεων. Οι αυτοματοποιημένες δοκιμές πτήσης επιτρέπουν στο προσωπικό να επικεντρώνεται εκ νέου σε κρίσιμες πληροφορίες, να ανιχνεύει γρήγορα μη φυσιολογικές δραστηριότητες, να αυξάνει την αποτελεσματικότητα των δοκιμών και να συντομεύει τους κύκλους δοκιμών. Η απλοποίηση δεδομένων μετατρέπει τα ακατάστατα δεδομένα αισθητήρων σε κανονικά μοτίβα.

Επιπλέον, η μηχανική μάθηση ενσωματώνει την προηγούμενη εμπειρία δοκιμών στην ανάλυση δεδομένων και τις φυσικές πληροφορίες για την εκτέλεση λήψης αποφάσεων σε πραγματικό χρόνο κατά τη διάρκεια της φάσης δοκιμών. Η Lockheed Martin αναπτύσσει νέες εφαρμογές μηχανικής μάθησης για δοκιμές και αξιολόγηση σχεδίων αεροσκαφών. Οι μηχανικοί και οι πιλότοι δοκιμών είναι ικανοί να διεξάγουν δομικές, προωθητικές, αεροηλεκτρονικές και περιβαλλοντικές δοκιμές. Οι δοκιμές των αεροσκαφών διασφαλίζουν ότι τα δοκιμασμένα αεροσκάφη πληρούν τα υψηλότερα δυνατά πρότυπα και παρέχουν στους επιβάτες ασφαλέστερη εμπειρία [60, 61].

Τα στάδια της αεροδιαστημικής μπορούν να βελτιωθούν αντίστροφα με τη χρήση δεδομένων, συμπεριλαμβανομένης της βελτιωμένης προβλεπτικής συντήρησης, της

βελτιστοποιημένης απόδοσης και της βελτιωμένης ασφάλειας, αξιοπιστίας και αποτελεσματικότητας. Τα πιθανά προβλήματα εντοπίζονται και αντιμετωπίζονται πριν εμφανιστούν, με αποτέλεσμα χαμηλότερο κόστος συντήρησης και λειτουργίας. Οι πληροφορίες που συλλέγονται βελτιώνουν τα ψηφιακά μοντέλα του αεροσκάφους, επιτρέποντας πιο ακριβείς προβλέψεις και προσομοιώσεις επιδόσεων, συντήρησης και άλλων επιχειρησιακών πτυχών. Τα δεδομένα βελτιστοποιούν περαιτέρω την επεξεργασία των αισθητήρων, επιτρέποντας πιο ακριβείς σχεδιασμούς αεροσκαφών. Επιπλέον, τα δεδομένα μπορούν να χρησιμοποιηθούν για την ταχεία αξιολόγηση και τροποποίηση των υφιστάμενων σχεδίων αεροσκαφών ώστε να καταστούν πιο αποτελεσματικά και ασφαλέστερα. Πολλοί επιβάτες χάνουν την εμπιστοσύνη τους στην αεροπορική εταιρεία ως αποτέλεσμα των καθυστερήσεων πτήσης που προκαλούνται από βλάβες αεροσκαφών. Η μηχανική μάθηση αναγνωρίζει πρότυπα δεδομένων σε προηγούμενες πτήσεις για να προβλέψει την πιθανότητα καθυστέρησης πτήσης. Η μηχανική μάθηση προσδιορίζει την καθυστέρηση που προκαλείται από μηχανικά προβλήματα, επιτρέποντάς τους να επιλυθούν γρήγορα και να αποφευχθούν στο μέλλον [62– 63].

Επιπροσθέτως, η μηχανική μάθηση επεξεργάζεται τεράστιες ποσότητες δεδομένων του συστήματος αεροσκαφών για να υποστηρίξει τη γρήγορη αυτοδιάγνωση, να εντοπίσει σφάλματα σε πραγματικό χρόνο και να λάβει ακριβείς αποφάσεις. Οι υπηρεσίες του αεροδρομίου παράγουν μεγάλο όγκο δεδομένων, συμπεριλαμβανομένων πληροφοριών ελέγχου επιβατών, λεπτομερειών φορτίου, χρονοδιαγραμμάτων πτήσεων, αρχείων καταγραφής ασφαλείας και ενημερώσεων καιρού, τα οποία μπορούν να χρησιμοποιηθούν για τη βελτιστοποίηση των λειτουργιών του αεροδρομίου και τη διασφάλιση της ασφάλειας των επιβατών. Τα αυτόνομα συστήματα που βασίζονται στη μηχανική μάθηση λαμβάνουν τεκμηριωμένες αποφάσεις για το προσωπικό του αεροδρομίου και παρέχουν

καλύτερες υπηρεσίες στους επιβάτες. Τα συστήματα αυτά βελτιστοποιούν τις λειτουργίες για τη μείωση των καθυστερήσεων, τη βελτίωση των προγραμμάτων πτήσεων και την πρόβλεψη των αναγκών των επιβατών [64– 65].

2.3. Συμβολή της μηχανικής μάθησης στην βελτίωση της εμπειρίας των πελατών

Στην αεροπορική βιομηχανία, η μηχανική μάθηση γίνεται όλο και πιο διαδεδομένη. Η χρήση της ενισχύει την ασφάλεια, βελτιώνει την εμπειρία των επιβατών και εξορθολογίζει τη διαδικασία του ταξιδιού, παρέχοντας τελικά ένα πιο αποτελεσματικό, ασφαλές, απρόσκοπτο και ευχάριστο ταξίδι από την είσοδο στην πύλη επιβίβασης [66].

Πριν ταξιδέψουν, οι επιβάτες μπορούν εύκολα να προγραμματίσουν ταξίδια, να κλείσουν εισιτήρια και να αγοράσουν βοηθητικά προϊόντα στο διαδίκτυο με μηχανική μάθηση. Η μηχανική μάθηση παρέχει στους επιβάτες εξατομικευμένες συστάσεις με βάση τις προτιμήσεις και την τοποθεσία τους. Τα πρότυπα και οι προτιμήσεις των επιβατών στο παρελθόν αναλύονται με αλγόριθμους. Οι αλγόριθμοι μπορούν να προτείνουν διαδρομές, προορισμούς και αξιοθέατα προσαρμοσμένα στις συγκεκριμένες ανάγκες και επιθυμίες τους. Μέσω της μηχανικής μάθησης, οι επιβάτες αποκτούν μια σειρά επιλογών, συμπεριλαμβανομένων των καταλληλότερων τρόπων μεταφοράς, των χαμηλότερων ναύλων και των πιο άνετων ταξιδιών. Αυτό δίνει τη δυνατότητα στους επιβάτες να λαμβάνουν τεκμηριωμένες αποφάσεις και να επιλέγουν την καταλληλότερη επιλογή για το ταξίδι τους, με αποτέλεσμα εξοικονόμηση χρόνου και κόστους [67- 69].

Η μηχανική μάθηση ενημερώνει την κυκλοφορία σε πραγματικό χρόνο, προτείνει εναλλακτικές διαδρομές και διασφαλίζει ότι οι επιβάτες φτάνουν εγκαίρως. Επιπλέον, βελτιστοποιεί τις διαδρομές λαμβάνοντας υπόψη την κυκλοφορία σε πραγματικό χρόνο, τον καιρό και άλλες συνθήκες για να εξασφαλίσει την ταχύτερη και πιο αποτελεσματική διαδρομή, καθιστώντας τον προγραμματισμό του ταξιδιού πιο ευχάριστο και χωρίς άγχος. Για παράδειγμα, η Uber έχει αναπτύξει κλιμακωτά, αξιόπιστα, εύχρηστα και αυτοματοποιημένα εργαλεία μηχανικής μάθησης για την πρόβλεψη κατά προσέγγιση χρόνων άφιξης. Αυτή η διαδικασία περιλαμβάνει διαχείριση δεδομένων, εκπαίδευση δεδομένων, αξιολόγηση μοντέλων, ανάπτυξη μοντέλων, πρόβλεψη και παρακολούθηση προβλέψεων. Επιπλέον, η μηχανική μάθηση χρησιμοποιείται από τους Χάρτες Google για τη βελτιστοποίηση των δρομολογίων για τους επιβάτες. Οι επιβάτες ενημερώνονται για τυχόν καθυστερήσεις ή αλλαγές στη διαδρομή τους σε πραγματικό χρόνο και συνιστώνται εναλλακτικές διαδρομές με βάση τις τρέχουσες συνθήκες. Η Delta AirLines χρησιμοποιεί μηχανική μάθηση για την τιμολόγηση εισιτηρίων, την ανάλυση της συμπεριφοράς των επιβατών και την παροχή εμπειριών αυτοεξυπηρέτησης [70– 72].

Κατά τη διάρκεια του check-in, η μηχανική μάθηση μεταμορφώνει γρήγορα την αεροπορική βιομηχανία. Τα μηχανήματα αυτοεξυπηρέτησης check-in κερδίζουν δημοτικότητα ως μέσο παροχής ακριβέστερων και αποτελεσματικών διαδικασιών. Η μηχανική μάθηση μπορεί να αναγνωρίσει διάφορους τύπους επιβατών και να προσφέρει μια πιο εξατομικευμένη εμπειρία check-in. Με την αυτοματοποίηση της διαδικασίας, οι αεροπορικές εταιρείες μπορούν να εξοικονομήσουν χρήματα, ενώ παράλληλα αναλύουν δεδομένα για να αποκτήσουν γνώσεις σχετικά με τις προτιμήσεις των επιβατών για τη βελτίωση των υπηρεσιών και την αύξηση της αφοσίωσης. Οι επιβάτες εισάγουν τα στοιχεία τους, όπως το όνομα και τον αριθμό πτήσης τους, στο περίπτερο αυτοεξυπηρέτησης. Στη συνέχεια, το σύστημα αποκτά

πρόσβαση σε μια βάση δεδομένων με αποθηκευμένες πληροφορίες επιβατών και χρησιμοποιεί μηχανική μάθηση για να εντοπίσει γρήγορα τον επιβάτη και να τον ελέγξει για την πτήση του χωρίς την ανάγκη χειροκίνητης επαλήθευσης, μειώνοντας την πιθανότητα ανθρώπινου λάθους, συντομεύοντας τους χρόνους αναμονής και παρέχοντας πιο οικονομικά αποδοτικές λύσεις. Για παράδειγμα, το Διεθνές Αεροδρόμιο του Χονγκ Κονγκ χρησιμοποιεί μηχανική μάθηση για να αυτοματοποιήσει το σύστημα check-in του αεροδρομίου. Το αεροδρόμιο χρησιμοποιεί συνδυασμό αναγνώρισης προσώπου, σαρώσεων διαβατηρίων και άλλων βιομετρικών δεδομένων για την επαλήθευση της ταυτότητας των επιβατών και τη διευκόλυνση μιας ταχύτερης και ασφαλέστερης διαδικασίας check-in. Αυτές οι τεχνολογίες μπορούν να μειώσουν τις μεγάλες ουρές στα σημεία εισόδου και να βελτιστοποιήσουν την αποδοτικότητα των υπηρεσιών

Επιπροσθέτως, η μηχανική μάθηση με δυνατότητα αυτόματης απόδοσης ετικετών αποσκευών βελτιώνει τις διαδικασίες του αεροδρομίου μειώνοντας παράλληλα τη χειρωνακτική εργασία και τους χρόνους αναμονής των επιβατών. Το σύστημα αναγνωρίζει γρήγορα το σχήμα και το μέγεθος των αποσκευών, αποδίδει έναν μοναδικό κωδικό και τις επισημαίνει με ακρίβεια. Στη συνέχεια, ο κωδικός χρησιμοποιείται για την παρακολούθηση των αποσκευών και τη διασφάλιση της ασφαλούς άφιξής τους. Το αεροδρόμιο Changi της Σιγκαπούρης χρησιμοποιεί μηχανική μάθηση για αυτο-σήμανση, η οποία αναγνωρίζει τους επιβάτες, τους εκχωρεί μια μοναδική κάρτα επιβίβασης και εκχωρεί μια ετικέτα αναγνώρισης στις αποσκευές τους. Η Cathay Pacific απαιτεί από τους επιβάτες να αποκτήσουν την κάρτα επιβίβασης και την ετικέτα αποσκευών τους από το περίπτερο πριν προχωρήσουν στον μετρητή πτώσης αποσκευών. Η εφαρμογή της αυτοματοποιήσεως πτώσης αποσκευών έχει τη δυνατότητα να μειώσει σημαντικά τους χρόνους αναμονής των επιβατών στα αεροδρόμια, εξαλείφοντας παράλληλα την ανάγκη για

το προσωπικό του αεροδρομίου να ελέγχει χειροκίνητα τους επιβάτες. Αυτό επιτρέπει στους επιβάτες να διαθέσουν το χρόνο τους σε πιο κρίσιμα καθήκοντα. Τα συστήματα υπολογιστικής όρασης διαβάζουν τις πληροφορίες εισιτηρίων και τις συγκρίνουν με τις πληροφορίες στο διαβατήριό. Μόλις επικυρωθούν οι πληροφορίες, το σύστημα αναγνώρισης προσώπου που βασίζεται στη μηχανική μάθηση αναγνωρίζει τον επιβάτη. Το σύστημα ανακτά γρήγορα τις πληροφορίες πτήσης και δημιουργεί τις κατάλληλες ετικέτες. Ο επιβάτης τοποθετεί τις τσάντες του στη μεταγωγική ζώνη και το μηχάνημα σαρώνει τις ετικέτες για να κατευθύνει τις τσάντες στον σωστό προορισμό. Η Cathay Pacific και το Διεθνές Αεροδρόμιο του Χονγκ Κονγκ έχουν εφαρμόσει εγκαταστάσεις self-service bag drop για να βοηθήσουν τους επιβάτες να εκτυπώσουν ετικέτες αποσκευών και να ελέγξουν τις αποσκευές τους μόνοι τους, μειώνοντας τη διαδικασία πτώσης τσαντών σε περίπου 1 λεπτό. Τα διεθνή αεροδρόμια Heathrow και Los Angeles υιοθετούν προηγμένο σύστημα για να διευκολύνουν τη διαδικασία check-in για τους επιβάτες. Η μηχανική μάθηση μπορεί να εξορθολογίσει τη διαδικασία check-in και να μειώσει τους χρόνους αναμονής αναλύοντας την ώρα της ημέρας, τον αριθμό των επιβατών και άλλους παράγοντες. Τα ημερήσια δεδομένα χρησιμοποιούνται για τη βελτιστοποίηση της διάταξης του προσωπικού και του check-in, βελτιώνοντας την εμπειρία των επιβατών [73– 75].

Η μηχανική μάθηση αναλύει τεράστιες ποσότητες δεδομένων για τον εντοπισμό προτύπων, την πρόβλεψη πιθανών απειλών για την ασφάλεια, τον εντοπισμό δόλιων εγγράφων, ασυνεπειών και ανωμαλιών. Τα αεροδρόμια μπορούν να χρησιμοποιούν μηχανική μάθηση για τον εντοπισμό πιθανών κινδύνων και τη λήψη προληπτικών μέτρων για τη διασφάλιση της ασφάλειας των ταξιδιωτών. Η μηχανική μάθηση μαθαίνει από σύνολα δεδομένων και αναγνωρίζει πρότυπα συμπεριφοράς για τον εντοπισμό και την πρόληψη ζητημάτων ασφάλειας στα

αεροδρόμια. Τα αεροδρόμια αντιμετωπίζουν κάθε επιβάτη ως σημείο δεδομένων, δημιουργώντας ένα προσωπικό προφίλ για τον καθένα. Οι πληροφορίες κράτησης, το ιστορικό ταξιδιού και οι μεμονωμένες λεπτομέρειες χρησιμοποιούνται ως παράμετροι προγνωστικής ανάλυσης. Αυτά τα δομημένα και μη δομημένα δεδομένα χρησιμοποιούνται σε μη εποπτευόμενη μάθηση και το αεροδρόμιο αποδίδει βαθμολογία κινδύνου σε κάθε άτομο. Η κατανομή του Gauss χρησιμοποιείται στη μαθηματική μοντελοποίηση ώστε να μάθει πολύπλοκες μη γραμμικές σχέσεις μεταξύ των χαρακτηριστικών των επιβατών. Σε ένα σύστημα ταξινόμησης, κάθε επιβάτης έχει διαφορετική κατηγορία και ο υψηλότερος κίνδυνος εντοπίζεται εύκολα. Οι αλγόριθμοι υπολογιστικής όρασης αναλύουν δεδομένα από κάμερες ασφαλείας για να ανιχνεύσουν ύποπτη συμπεριφορά, παράνομες δραστηριότητες και πιθανές απειλές. Τα αεροδρόμια χρησιμοποιούν τεχνολογία αναγνώρισης προσώπου για τον εντοπισμό ατόμων που έχουν επισημανθεί ως πιθανές απειλές για την ασφάλεια. Το σύστημα επιτρέπει στα αεροδρόμια να εντοπίζουν γρήγορα και με ακρίβεια άτομα που ενδέχεται να αποτελούν απειλή, αποτρέποντας την τρομοκρατία, το λαθρεμπόριο και άλλες εγκληματικές δραστηριότητες. Η μηχανική μάθηση προσδιορίζει τους ταξιδιώτες που ενδέχεται να ενέχουν κίνδυνο με βάση τα ταξιδιωτικά τους πρότυπα, όπως εάν έχουν επισκεφθεί συγκεκριμένες χώρες ή ιστότοπους [76–78].

Τα αυτοματοποιημένα συστήματα ανίχνευσης και αυτοελέγχου αποσκευών, συμπεριλαμβανομένων των συστημάτων ανίχνευσης εκρηκτικών, των συστημάτων ανίχνευσης απαγορευμένων αντικειμένων και των τεχνολογιών 3D CT, χρησιμοποιούνται εκτενώς στα αεροδρόμια. Τα συστήματα αυτά βοηθούν τους φορείς εκμετάλλευσης να εντοπίζουν ύποπτα στοιχεία για περαιτέρω εξέταση. Η Υπηρεσία Ασφάλειας Μεταφορών (TSA) ανιχνεύει ύποπτα αντικείμενα, πιθανά εκρηκτικά και άλλα απαγορευμένα αντικείμενα σε τσάντες μεταφοράς και

αεροδρόμια χρησιμοποιώντας τεχνολογία 3D απεικόνισης και μοντέλα μηχανικής μάθησης. Η μηχανική μάθηση χρησιμοποιεί συνθετικά σύνολα δεδομένων εκπαίδευσης και προσομοιωτές για να μιμηθεί πραγματικές και ποικίλες απειλές. Τα σενάρια μελετώνται γρήγορα και διεξοδικά, αναδημιουργούνται, μοντελοποιούνται και κατανοούνται για τη δημιουργία συνθετικών βιβλιοθηκών για την πλήρη εκπαίδευση και δοκιμή αλγορίθμων για πολύπλοκες ιδιότητες απειλής. Η μηχανική μάθηση χρησιμοποιείται όλο και περισσότερο για την αυτοματοποίηση διαδικασιών ελέγχου των συνόρων που είναι πιο αποτελεσματικές και ασφαλείς από τις χειροκίνητες μεθόδους. Βιομετρικές πληροφορίες όπως δακτυλικά αποτυπώματα, φωτογραφίες, σαρώσεις ίριδας, αριθμοί διαβατηρίων και θεωρήσεων συλλέγονται από τον ταξιδιώτη για αυτοματοποιημένο έλεγχο των συνόρων. Με βάση τα δεδομένα που συλλέγονται, τα προφίλ των ταξιδιωτών δημιουργούνται για να προχωρούν αυτόματα. Όταν η μηχανική μάθηση ανιχνεύει ασυνήθιστα πρότυπα σε δεδομένα, όπως ύποπτη συμπεριφορά ή πιθανές απειλές για την ασφάλεια, ειδοποιεί τις αρχές και ξεκινά έρευνα [79– 80].

Επιπλέον, η μηχανική μάθηση βελτιώνει την εμπειρία επιβίβασης και διαχείρισης αποσκευών μειώνοντας τους χρόνους αναμονής, βελτιστοποιώντας τη ροή της κυκλοφορίας και παρέχοντας στους επιβάτες έξυπνες υπηρεσίες. Πιο συγκεκριμένα, η μηχανική μάθηση αναλύει τις πληροφορίες κράτησης, τον αριθμό επιβατών, το μέγεθος του αεροσκάφους, τον αριθμό αποσκευών για να καθορίσει μια αποτελεσματική σειρά επιβίβασης που ελαχιστοποιεί τον χρόνο επιβίβασης. Η μηχανική μάθηση παρακολουθεί τους επιβάτες αναλύοντας δεδομένα όπως οι χρόνοι άφιξης και ελέγχου ασφαλείας για να προσαρμόσει ανάλογα τη διαδικασία επιβίβασης. Ανιχνεύει πρότυπα στη διαδικασία επιβίβασης, όπως η προτεραιότητα επιβίβασης, η ταχύτητα επιβίβασης και ο χρόνος επιβίβασης, για τη μεγιστοποίηση της αποτελεσματικότητας. Επιπροσθέτως, αναλύει την τρέχουσα διαδικασία

επιβίβασης για τον εντοπισμό σημείων συμφόρησης, τη μείωση της συμφόρησης και τη βελτίωση της ροής των επιβατών. Τέλος, η μηχανική μάθηση παρέχει εξατομικευμένες εμπειρίες επιβίβασης αναλύοντας δεδομένα επιβατών, όπως ιστορικό ταξιδιού, προτιμήσεις και προηγούμενες εμπειρίες πτήσης [81– 82].

Οι αυτοματοποιημένες λύσεις αποσκευών που χρησιμοποιούν μηχανική μάθηση παρέχουν βελτιωμένα και αποτελεσματικά αεροπορικά ταξίδια. Τα αεροδρόμια επωφελούνται από τη μηχανική μάθηση χρησιμοποιώντας λιγότερο χρόνο και πόρους για τη μεταφορά αποσκευών, γεγονός που μειώνει το κόστος και βελτιώνει την εμπειρία του ταξιδιώτη. Τα αεροδρόμια μπορούν να καθορίσουν τις καλύτερες διαδρομές αποσκευών και τις πιο αποτελεσματικές τεχνικές φόρτωσης και εκφόρτωσης αξιολογώντας δεδομένα από σαρωτές αποσκευών και πληροφορίες επιβατών. Εντοπίζει πιθανά ζητήματα που ενδέχεται να προκύψουν κατά τη διάρκεια της διαδικασίας χειρισμού αποσκευών για την πρόβλεψη παραπόνων των καταναλωτών και την αντιμετώπισή τους εκ των προτέρων. Ο αυτοματοποιημένος χειρισμός αποσκευών εξαλείφει την ανάγκη για χειροκίνητη διαλογή και μειώνει την πιθανότητα ανθρώπινου λάθους. Χρησιμοποιώντας μηχανική μάθηση, το προσωπικό του αεροδρομίου μπορεί να παρακολουθεί με μεγαλύτερη ακρίβεια και αποτελεσματικότητα τις πληροφορίες επιβατών και αποσκευών. Μετά τη σάρωση των ετικετών αποσκευών, οι αλγόριθμοι μηχανικής μάθησης αξιολογούν τα σαρωμένα δεδομένα και αναθέτουν τον σωστό επιβάτη στις πληροφορίες. Οι ανωμαλίες, οι κατεστραμμένες και οι χαμένες αποσκευές μπορούν να αναγνωριστούν χρησιμοποιώντας αλγόριθμους αναγνώρισης εικόνας. Το Διεθνές Αεροδρόμιο του Κρασνοντάρ χρησιμοποιεί ρομπότ για να αυτοματοποιήσει τη διαδικασία παραλαβής και φόρτωσης αποσκευών, μειώνοντας το ανθρώπινο λάθος, την απώλεια και το ποσό ζημιών και τον χρόνο που απαιτείται για τη φόρτωση και εκφόρτωση των αεροσκαφών. Το ρομπότ είναι ένας χειριστής εξοπλισμένος με

κάμερες, σαρωτές γραμμικού κώδικα, ανιχνευτές και κάποιο προηγμένο αλγόριθμο. Ένα αυτοματοποιημένο σύστημα ανίχνευσης αντικειμένων βασισμένο σε όραση υπολογιστή ανιχνεύει απαγορευμένα αντικείμενα σε εικόνες ακτίνων Χ. Το σύστημα μπορεί να ανιχνεύσει απαγορευμένα αντικείμενα σε ασπρόμαυρες εικόνες ακτίνων Χ 2D, έγχρωμες εικόνες ακτίνων Χ και εικόνες αξονικής τομογραφίας 3D με έως και 30 εικόνες αποσκευών ανά δευτερόλεπτο. Το σύστημα ανιχνεύει τις απειλές με μεγαλύτερη ακρίβεια και αποτελεσματικότητα, μειώνοντας τους ψευδείς συναγερμούς, εξοικονομώντας πολύτιμο χρόνο, αυξάνοντας την ασφάλεια των επιβατών και αφαιρώντας το ανθρώπινο λάθος από τη διαδικασία ελέγχου αποσκευών. Η μηχανική μάθηση βελτιώνει την ακρίβεια, τη βιωσιμότητα και την οικονομική αποδοτικότητα της διαδικασίας διαχείρισης αποσκευών. Το αεροδρόμιο του Αϊντχόβεν χρησιμοποιεί ένα σύστημα χειρισμού αποσκευών με μηχανική μάθηση που επιτρέπει στους επιβάτες να φωτογραφίζουν απλά τις αποσκευές τους, να τις καταθέτουν και να τις παραλαμβάνουν όταν φτάνουν στον προορισμό χωρίς να χρειάζονται ετικέτες. Ο αλγόριθμος αναγνώρισης εικόνας συνδέεται με το σύστημα αποσκευών. Οι αλγόριθμοι κατηγοριοποιούν τις αποσκευές και στη συνέχεια τις συγκρίνουν με το καταχωρημένο σύνολο δεδομένων εικόνας, ανακτώντας ακριβή δεδομένα όπως η προέλευση, ο τύπος, το χρώμα, η ταξινόμηση IATA, ο κατασκευαστής και οι διαστάσεις των αποσκευών. Το σύστημα αποσκευών που βασίζεται στη μηχανική μάθηση εξαλείφει την ανάγκη για ετικέτες αποσκευών και μηχανές εκτύπωσης ετικετών, καθιστώντας το αεροδρόμιο του Αϊντχόβεν πιο φιλικό προς το περιβάλλον [83 – 84].

Επιπροσθέτως, η μηχανική μάθηση βοηθά το προσωπικό του αεροδρομίου να λαμβάνει πιο ενημερωμένες επιχειρησιακές αποφάσεις. Χρησιμοποιώντας εξελιγμένους αλγόριθμους, το προσωπικό του αεροδρομίου μπορεί να προβλέψει και να παρακολουθήσει με ακρίβεια τις ώρες άφιξης και αναχώρησης των πτήσεων,

καθώς και άλλες σχετικές πληροφορίες. Τα μεγάλα αεροδρόμια περιέχουν τεράστιο όγκο δεδομένων από συστήματα και εξοπλισμό όπως ασφάλεια, αποσκευές, φορτίο, κυκλοφορία και επιβάτες. Η μηχανική μάθηση συλλέγει όλα τα σχετικά δεδομένα και τα μετατρέπει σε πληροφορίες με δυνατότητα δράσης, επιτρέποντας στους χειριστές να μεγιστοποιήσουν την αποτελεσματικότητά τους και να βελτιστοποιήσουν την αποτελεσματικότητα του αεροδρομίου. Η μηχανική μάθηση βελτιώνει τις λειτουργίες των πελατών, παρέχει εξατομικευμένες υπηρεσίες και επιτρέπει μεμονωμένες στρατηγικές. Οι δραστηριότητες επιβατών, όπως το check-in, τα έγγραφα αναχώρησης, τα σημεία ελέγχου ασφαλείας, η επιβίβαση και η αποβίβαση στο αεροπλάνο απλοποιούνται. Τα αεροδρόμια δημιουργούν προφίλ πελατών αναλύοντας τα δεδομένα των επιβατών για να προσδιορίσουν τις ανάγκες τους και να παρέχουν προσαρμοσμένες λύσεις. Οι πράκτορες εξυπηρέτησης πελατών χρησιμοποιούν τα προφίλ για να εντοπίζουν πελάτες με παρόμοιες ανάγκες και να τους παρέχουν τις ίδιες λύσεις, μειώνοντας τους χρόνους αναμονής και αυξάνοντας την ικανοποίηση. Το αεροδρόμιο προσαρμόζει τα προγράμματα των πτήσεων, την χωρητικότητα των αεροσκαφών, τους συντελεστές φόρτωσης και τα ποσοστά μεταφοράς με βάση τα μεταβαλλόμενα πρότυπα των επιβατών. Κατά τις ώρες αιχμής, το αεροδρόμιο συντονίζει δυναμικά το μήκος της ουράς και ανοίγει αμέσως εφεδρικό εξοπλισμό. Η μηχανική μάθηση χρησιμοποιείται στο αεροδρόμιο Schiphol του Άμστερνταμ για την παροχή εξατομικευμένων υπηρεσιών καθώς και για τη διορθωτική και προγνωστική συντήρηση των περιουσιακών στοιχείων. Το συγκεκριμένο αεροδρόμιο αναλύει τη συμπεριφορά και τις προτιμήσεις των πελατών χρησιμοποιώντας προγνωστικά αναλυτικά στοιχεία για να παρέχει προσαρμοσμένες συστάσεις για την πύλη επιβίβασης, πληροφορίες πτήσης και ταξιδιωτικές συμβουλές. Το αεροδρόμιο Heathrow αναλύει δεδομένα που δημιουργούνται από τα αεροσκάφη, βελτιώνει τις διαδικασίες ανακύκλωσης

αεροσκαφών και μειώνει τις καθυστερήσεις των επιβατών χρησιμοποιώντας μηχανική μάθηση [85– 87].

Τα αυτόνομα ρομπότ και οχήματα προσφέρουν έναν αποτελεσματικό, ασφαλή και οικονομικά αποδοτικό τρόπο μετακίνησης επιβατών και εμπορευμάτων σε όλο τον τερματικό σταθμό, παρέχοντας υπηρεσίες όπως έλεγχος ροής επιβατών, παράδοση αποσκευών, ασφάλεια επιβατών και μεταφορά επιβατών. Τα chatbots μηχανικής μάθησης είναι κρίσιμα εργαλεία για τα αεροδρόμια για τη βελτίωση της εμπειρίας των επιβατών. Ανταποκρίνονται με ανθρώπινο τρόπο χρησιμοποιώντας επεξεργασία φυσικής γλώσσας, παρέχοντας αποτελεσματικά πληροφορίες, μειώνοντας παράλληλα το φόρτο εργασίας του προσωπικού του αεροδρομίου. Αυτά τα chatbots μπορούν να παρακολουθούν και να διαχειρίζονται τις λειτουργίες του αεροδρομίου, να παρακολουθούν τα σχόλια των πελατών, να εντοπίζουν βελτιώσεις και να προειδοποιούν για διάφορα θέματα. Η AirAsia δημιούργησε chatbots χρησιμοποιώντας μηχανική μάθηση για να παρέχει στους επιβάτες μια πιο απρόσκοπτη και φιλική προς το χρήστη εμπειρία. Τα ρομπότ που βασίζονται στη μηχανική μάθηση, βοηθούν τους ταξιδιώτες να περιηγούνται στα αεροδρόμια βοηθώντας στην πλοήγηση και το χειρισμό αποσκευών. Χρησιμοποιώντας υπολογιστική όραση, τα ρομπότ-οδηγοί προγραμματίζονται ώστε να αναγνωρίζουν διάφορους τύπους αποσκευών και να εντοπίζουν τους κατάλληλους μεταφορικούς ιμάντες. Για τη μεταφορά αποσκευών, το αεροδρόμιο της Φρανκφούρτης χρησιμοποιεί ρομπότ - οδηγούς γνωστά ως FRANbots. Τα ρομπότ έχουν σχεδιαστεί για να μεταφέρουν αυτόνομα τις αποσκευές στον τελικό προορισμό τους, εξαλείφοντας την ανάγκη για χειροκίνητο χειρισμό και βελτιώνοντας την αποτελεσματικότητα της μεταφοράς αποσκευών. Τα FRANbots ήταν εξοπλισμένα με αισθητήρες, οθόνες αφής και τεχνολογία υπολογιστικής όρασης για να επιτρέψουν την προηγμένη πλοήγηση και την αποφυγή εμποδίων. Οι επιβάτες

εισάγουν τις πληροφορίες πτήσης και τον προορισμό τους στο FRANbot, το οποίο θα τους πλοηγηθεί στον προορισμό τους και θα αφήσει τις αποσκευές τους. Με την αυξημένη δημοτικότητα των αεροπορικών ταξιδιών και την αυξανόμενη ζήτηση για έγκαιρες αναχωρήσεις και αφίξεις, τα αυτόνομα ρομπότ θεωρούνται όλο και περισσότερο ως βιώσιμη λύση στην ενίσχυση της ακρίβειας των αεροδρομίων. Το αεροδρόμιο Incheon στη Νότια Κορέα και το αεροδρόμιο του Μονάχου στη Γερμανία αναπτύσσουν αυτόνομα ρομπότ για να βοηθήσουν τους επιβάτες να κινούνται γύρω από το αεροδρόμιο γρήγορα και με ευκολία. Τα ρομπότ είναι εξοπλισμένα με προηγμένους αισθητήρες και μηχανική μάθηση για την παρακολούθηση της κίνησης των επιβατών, τη διασφάλιση της ακρίβειας των κινήσεων των επιβατών, τη διατήρηση των επιβατών στην ώρα τους και την αποφυγή καθυστερήσεων. Παρέχοντας στους επιβάτες ταχύτερη και αποτελεσματικότερη εξυπηρέτηση, το αεροδρόμιο είναι σε θέση να μειώσει το χρόνο και το ανθρώπινο δυναμικό που απαιτείται για την ομαλή λειτουργία. Τα αυτόνομα οχήματα προορίζονται να αποτελέσουν μια αποτελεσματική λύση για το χειρισμό αποσκευών. Η Delta Airlines χρησιμοποιεί αυτοκινούμενα οχήματα για να βοηθήσει στην καθυστερημένη παράδοση αποσκευών. Οι αυτοκινούμενες ηλεκτρικές αναπηρικές καρέκλες βελτιώνουν την προσβασιμότητα για τους επιβάτες με ειδικές ανάγκες και αποτελούν ένα βήμα προς τα εμπρός στον ταξιδιωτικό κλάδο. Οι αυτόνομες ηλεκτρικές αναπηρικές καρέκλες είναι εξοπλισμένες με διάφορους αισθητήρες και κάμερες για να εξασφαλίσουν ίση πρόσβαση στα αεροπορικά ταξίδια και συνεπή ταξιδιωτική εμπειρία για τους επιβάτες με ειδικές ανάγκες [88– 90].

Τέλος, η ψυχαγωγία παίζει σημαντικό ρόλο στο ταξίδι των επιβατών. Σήμερα, υπάρχουν πολλές επιλογές ψυχαγωγίας διαθέσιμες για να κάνουν το ταξίδι πιο ευχάριστο. Οι αεροπορικές εταιρείες χρησιμοποιούν μηχανική μάθηση για να παρέχουν πιο εξατομικευμένη ψυχαγωγία κατά την πτήση. Ορισμένες αεροπορικές

εταιρείες παρέχουν εξατομικευμένη μουσική ως καταπραΰντικό και ηρεμιστικό εργαλείο για τους νευρικούς επιβάτες. Με τη συλλογή δεδομένων επιβατών, οι αεροπορικές εταιρείες προσαρμόζουν τις προσφορές πτήσεων ώστε να ανταποκρίνονται καλύτερα στις ατομικές τους προτιμήσεις, παρέχουν στους επιβάτες τις αγαπημένες τους ταινίες και επιλογές φαγητού στις πτήσεις. Οι προσαρμοσμένες υπηρεσίες βελτιώνουν την ικανοποίηση των επιβατών, βοηθούν τις αεροπορικές εταιρείες να κατανοήσουν καλύτερα τις ανάγκες των πελατών, να οικοδομήσουν την αφοσίωση των πελατών και να αυξήσουν τη συνολική κερδοφορία. Τα δεδομένα επιβατών συλλέγονται, καθαρίζονται και οργανώνονται για να εκπαιδεύσουν ένα μοντέλο μηχανικής μάθησης με σκοπό την παροχή εξατομικευμένων υπηρεσιών. Με την πρόοδο της τεχνολογίας, οι αεροπορικές εταιρείες αναζητούν ευκαιρίες για να επαναπροσδιορίσουν την ταξιδιωτική εμπειρία και να ενισχύσουν τη συνολική αποδοτικότητα των επιχειρήσεων [91].

2.4. Προκλήσεις

Ασφάλεια και ιδιωτικότητα: Οι αλγόριθμοι μηχανικής μάθησης εκπαιδεύονται σε δεδομένα που περιέχουν προσωπικές πληροφορίες επιβατών, από τις οποίες μπορούν να συναχθούν ευαίσθητες πληροφορίες. Οι κακόβουλοι παράγοντες μπορούν να σχεδιάσουν προκατειλημμένους αλγόριθμους για να επηρεάσουν τη λήψη αποφάσεων και τα αποτελέσματα. Η μηχανική μάθηση είναι ευάλωτη σε επιθέσεις στον κυβερνοχώρο και άλλες κακόβουλες δραστηριότητες [31]. Η υπηρεσία εξατομίκευσης που βασίζεται στη μηχανική μάθηση εγείρει ανησυχίες για την προστασία της ιδιωτικής ζωής. Η μηχανική μάθηση αναλύει τεράστιες ποσότητες δεδομένων από διάφορες πηγές, συμπεριλαμβανομένου του συστήματος πτήσης, των κοινωνικών μέσων, των αναζητήσεων στο διαδίκτυο και ακόμη και των φυσικών τοποθεσιών. Αυτό επιτρέπει τη δημιουργία λεπτομερών προφίλ του επιβάτη, τα

οποία στη συνέχεια μπορούν να χρησιμοποιηθούν για την παροχή εξατομικευμένων υπηρεσιών. Ωστόσο, οι πληροφορίες που συλλέγονται μπορούν να καταχραστούν. Η εξατομικευμένη υπηρεσία πτήσης που βασίζεται στη μηχανική μάθηση μπορεί να οδηγήσει σε αθέμιτες διακρίσεις, συμπεριλαμβανομένης της δημιουργίας "μαύρων λιστών" επιβατών που θεωρούνται ανεπιθύμητοι, της άρνησής τους να έχουν πρόσβαση σε ορισμένες υπηρεσίες ή της διαφορετικής μεταχείρισής τους. Οι αλγόριθμοι μπορούν να οδηγήσουν σε μεροληπτική λήψη αποφάσεων ή αθέμιτη τιμολόγηση. Οι αεροπορικές εταιρείες πρέπει να ακολουθούν αυστηρούς κανονισμούς περί απορρήτου για να διασφαλίσουν ότι οι εξατομικευμένες υπηρεσίες που βασίζονται στη μηχανική μάθηση παραμένουν ηθικές. Πριν από τη συλλογή και την αποθήκευση προσωπικών δεδομένων για τους επιβάτες, οι αεροπορικές εταιρείες θα πρέπει να λάβουν τη ρητή συγκατάθεσή τους. Οι αεροπορικές εταιρείες πρέπει να είναι διαφανείς σχετικά με τον τρόπο χρήσης των δεδομένων και να διασφαλίζουν ότι οι αποφάσεις που βασίζονται σε δεδομένα είναι δίκαιες και χωρίς διακρίσεις. Αυτές οι υπηρεσίες θα πρέπει να επιτρέπουν στους επιβάτες να επιλέξουν να εξαιρεθούν και να διαγράψουν τα δεδομένα τους.

Τεχνολογία: Καθώς η τεχνολογία εξελίσσεται και γίνεται πιο εξελιγμένη, η αεροπορική βιομηχανία πρέπει να συμβαδίζει με τις μεταβαλλόμενες τάσεις. Τα δεδομένα στην αεροπορική βιομηχανία είναι μεγάλου όγκου και ποικιλίας, καθιστώντας τα παραδοσιακά εργαλεία διαχείρισης δεδομένων αναποτελεσματικά. Τα αεροπορικά δεδομένα αποτελούνται από μεγάλο αριθμό δομημένων ή μη δομημένων δεδομένων που παράγονται από διάφορες πηγές. Τα δεδομένα παράγονται με ταχύ ρυθμό από τα μέσα κοινωνικής δικτύωσης, τα αρχεία καταγραφής ιστού και τους αισθητήρες. Η μηχανική μάθηση έχει περιορισμένη ισχύ επεξεργασίας και ταχύτητα. Υπάρχει συχνά πολύς θόρυβος στα αεροπορικά δεδομένα. Οι αλγόριθμοι μηχανικής μάθησης συχνά δυσκολεύονται να

αναγνωρίσουν και να μάθουν, με αποτέλεσμα τη δημιουργία ανακριβών μοντέλων. Η μηχανική μάθηση στερείται ερμηνείας, καθιστώντας δύσκολο για τους ανθρώπους να κατανοήσουν τις υποκείμενες διαδικασίες και αποφάσεις των μοντέλων. Η έλλειψη κατανόησης μπορεί να οδηγήσει σε δυσπιστία. Επιπλέον, η μηχανική μάθηση είναι επιρρεπής σε προκατάληψη, η οποία μπορεί να οδηγήσει σε ανακριβή και άδικα αποτελέσματα. Τα μοντέλα μηχανικής μάθησης εκπαιδεύονται στα δεδομένα του πραγματικού κόσμου, τα οποία μπορούν να αντικατοπτρίζουν κοινωνικές προκαταλήψεις. Αυτός ο τύπος προκατάληψης μπορεί να έχει σοβαρές συνέπειες για τους αλγόριθμους, καθώς μπορεί να οδηγήσει σε αθέμιτες αποφάσεις. Για να μειωθεί ο κίνδυνος μεροληψίας στη μηχανική μάθηση, τα δεδομένα που χρησιμοποιούνται για την εκπαίδευση του μοντέλου πρέπει να είναι ακριβή και αντιπροσωπευτικά του πληθυσμού που προορίζεται να εξυπηρετήσει.

Κόστος: Η ανάπτυξη εφαρμογών μηχανικής απαιτεί σημαντική ποσότητα υπολογιστικής ισχύος, λογισμικού και ενέργειας. Το κόστος υλοποίησης των αλγορίθμων μηχανικής μάθησης μπορεί να εξαρτάται από διάφορους παράγοντες, όπως η πολυπλοκότητα και το μέγεθος του συνόλου δεδομένων, ο αριθμός των μοντέλων και των αλγορίθμων που χρησιμοποιούνται, το απαιτούμενο υλικό και λογισμικό και η διάρκεια και οι πόροι που απαιτούνται για την εκπαίδευση και την ανάπτυξη. Το κόστος υλικού και λογισμικού μπορεί να είναι αρκετά υψηλό, ειδικά για αλγόριθμους βαθιάς μάθησης και μεγάλα σύνολα δεδομένων. Ο χρόνος και οι πόροι που απαιτούνται για την εκπαίδευση και την ανάπτυξη μοντέλων είναι δαπανηροί και απαιτούν εξειδικευμένες γνώσεις. Τα δεδομένα που χρησιμοποιούνται στους αλγόριθμους μηχανικής μάθησης πρέπει να είναι ακριβή και ενημερωμένα, γεγονός που μπορεί να είναι δύσκολο λόγω αλλαγών στο περιβάλλον και στη διαδικασία συλλογής δεδομένων. Η διατήρηση και ενημέρωση αλγορίθμων μηχανικής μάθησης είναι μια χρονοβόρα και δαπανηρή διαδικασία. Οι

αεροπορικές εταιρείες πρέπει να εξετάσουν προσεκτικά τα πλεονεκτήματα και τα μειονεκτήματα της χρήσης τεχνολογίας μηχανικής μάθησης για να διασφαλίσουν ότι οι πόροι τους χρησιμοποιούνται αποτελεσματικά.

Αποδοχή: Οι αεροπορικές εταιρείες αντιμετωπίζουν προκλήσεις στην αποδοχή της μηχανικής μάθησης λόγω ανησυχιών σχετικά με τον πιθανό αντίκτυπό τους σε ευρύτερες εφαρμογές. Για να αντιμετωπιστεί αυτό, οι αεροπορικές εταιρείες πρέπει να διασφαλίσουν ότι οι αιτήσεις τους μπορούν να εξηγηθούν, παρέχοντας πληροφορίες σχετικά με την υποκείμενη διαδικασία λήψης αποφάσεων. Η ποιότητα των αλγορίθμων μηχανικής μάθησης εξαρτάται σε μεγάλο βαθμό από την ποιότητα των δεδομένων στα οποία εκπαιδεύονται και οποιαδήποτε προκατάληψη στα δεδομένα μπορεί να επηρεάσει αρνητικά ακόμη και τους πιο προηγμένους εξελιγμένους αλγόριθμους. Ως εκ τούτου, οι αεροπορικές εταιρείες πρέπει να λάβουν μέτρα για τον εντοπισμό και την εξάλειψη πιθανών πηγών μεροληψίας στα δεδομένα τους. Για να διασφαλιστεί ότι οι χρήστες εμπιστεύονται τους αλγόριθμους μηχανικής μάθησης, τα δεδομένα πρέπει να είναι ακριβή και αμερόληπτα, οι αλγόριθμοι πρέπει να είναι διαφανείς και εξηγήσιμοι και το σύστημα πρέπει να είναι ασφαλές έναντι πιθανών επιθέσεων στον κυβερνοχώρο. Ωστόσο, οι αλγόριθμοι μηχανικής μάθησης είναι συχνά πολύπλοκοι και δύσκολο να κατανοηθούν και τα σύνολα δεδομένων που χρησιμοποιούνται για την εκπαίδευσή τους μπορεί να περιέχουν σφάλματα ή να μην είναι αντιπροσωπευτικά ολόκληρου του πληθυσμού, με αποτέλεσμα ανακριβή ή προκατειλημμένα αποτελέσματα. Οι αλγόριθμοι εξελίσσονται επίσης συνεχώς, καθιστώντας δύσκολη την εξασφάλιση συνεπών αποτελεσμάτων.

Κανονισμοί: Οι αυστηροί κυβερνητικοί κανονισμοί και η συμμόρφωση είναι απαραίτητοι για να παραμείνει η βιομηχανία ασφαλής. Η μηχανική μάθηση έχει πολλές και ποικίλες πιθανές εφαρμογές και ο αντίκτυπός της στη ζωή μας

αυξάνεται ραγδαία. Ωστόσο, υπάρχουν αρκετά κενά στο ισχύον ρυθμιστικό σύστημα που πρέπει να καλυφθούν. Πολλοί ρυθμιστικοί φορείς αγωνίζονται να καθορίσουν τη μηχανική μάθηση. Για να συμβαδίζουν με τις συνεχώς μεταβαλλόμενες δυνατότητες των τεχνολογιών, αυτά τα πρότυπα πρέπει να αναθεωρηθούν και να ενημερωθούν. Οι εσωτερικές λειτουργίες της μηχανικής μάθησης είναι σε μεγάλο βαθμό άγνωστες, καθιστώντας δύσκολο για τους ρυθμιστικούς φορείς να διασφαλίσουν ότι οι αλγόριθμοι δεν είναι προκατειλημμένοι ή διακριτικοί στις αποφάσεις τους. Μεγάλο μέρος του κανονισμού αφήνεται επί του παρόντος σε μεμονωμένες χώρες και οργανισμούς, με αποτέλεσμα ένα συνονθύλευμα νόμων και κανονισμών. Ένα παγκόσμιο κανονιστικό πλαίσιο θα διασφαλίσει ότι όλες οι χώρες τηρούν τα ίδια πρότυπα και θα παρέχουν ίσους όρους ανταγωνισμού για την ανάπτυξη και τη χρήση της τεχνολογίας. Απαιτούνται νόμοι και κανονισμοί για την αντιμετώπιση των ηθικών επιπτώσεων της μηχανικής μάθησης. Πρέπει να υπάρχει ένα πλαίσιο για να διασφαλιστεί ότι οι αλγόριθμοι δεν λαμβάνουν προκατειλημμένες αποφάσεις με βάση παράγοντες όπως η φυλή ή το φύλο. Υπάρχει ανάγκη για την εφαρμογή παγκόσμιων κανονισμών για την εποπτεία της προόδου και της χρήσης της τεχνολογίας αιχμής, παράλληλα με εξειδικευμένους νόμους και πολιτικές που στοχεύουν στην αντιμετώπιση των ηθικών ανησυχιών, της εμπιστευτικότητας των δεδομένων και των πιθανών αρνητικών επιπτώσεων. Η ρύθμιση της χρήσης των δεδομένων προωθεί τη διαφάνεια και τη συνεργασία των παραγωγών δεδομένων και των καταναλωτών δεδομένων. Η αυξημένη προσβασιμότητα των δεδομένων μπορεί να οδηγήσει σε πιο δίκαιες διαδικασίες λήψης αποφάσεων.

ΚΕΦΑΛΑΙΟ 3:ΥΛΟΠΟΙΗΣΗ

3.1. Εισαγωγή – Βασική Φιλοσοφία

Βασικός στόχος του πρακτικού μέρους της διπλωματικής εργασίας είναι η εφαρμογή μοντέλων μηχανικής μάθησης έτσι ώστε με βάση σύνολο δεδομένων , το οποίο θα παρουσιαστεί στην συνέχεια, να πραγματοποιηθεί πρόβλεψη της ικανοποίησης ή μη πελατών αεροπορικών εταιρειών.

Τα μοντέλα μηχανικής μάθησης, τα οποία εφαρμόστηκαν είναι τα παρακάτω :

- Logistic Regression
- Decision Trees
- Random Forest
- Naïve Bayes
- KNN

Οι μετρικές οι οποίες χρησιμοποιήθηκαν για την αξιολόγηση των μοντέλων είναι :

- Accuracy
- Precision
- Recall

Επιπλέον, η υλοποίηση πραγματοποιήθηκε σε γλώσσα Python, μια αρκετά ισχυρή γλώσσα όσο αφορά την εφαρμογή μοντέλων μηχανικής μάθησης και την ανάλυση μεγάλου όγκου δεδομένων [92 –98].

Στις επόμενες ενότητες που θα ακολουθήσουν :

- Θα πραγματοποιηθεί ανάλυση του συνόλου δεδομένων.
- Θα παρατεθούν οι μέθοδοι προ-επεξεργασίας του.

- Θα παρατεθούν τα αποτελέσματα εφαρμογής των μοντέλων μηχανικής μάθησης

3.2. Παρουσίαση Συνόλου Δεδομένων

Αρχικά, παρουσιάζεται πίνακας με την περιγραφή των στηλών του συνόλου δεδομένων :

Όνομα Στήλης	Περιγραφή
Gender	Φύλο των επιβατών
Customer Type	Τύπος των πελατών
Age	Ηλικία των επιβατών
Type of Travel	Σκοπός ταξιδιού
Class	Ταξιδιωτική κλάση των επιβατών
Flight Distance	Απόσταση του ταξιδιού
Inflight wifi service	Επίπεδο ικανοποίησης από τις παροχές Wi-Fi
Departure/Arrival time convenient	Επίπεδο ικανοποίησης από την ώρα αναχώρησης / άφιξης
Ease on Online booking	Επίπεδο ικανοποίησης από την διαδικασία online κράτησης

Gate location	Επίπεδο ικανοποίησης από την τοποθεσία της πύλης
Food and drink	Επίπεδο ικανοποίησης από τα φαγητά και ποτά
Online boarding	Επίπεδο ικανοποίησης από την online έκδοση boarding pass
Seat comfort	Επίπεδο ικανοποίησης από την άνεση των θέσεων
Inflight entertainment	Επίπεδο ικανοποίησης από την ψυχαγωγία κατά την διάρκεια της πτήσης
On-board service	Επίπεδο ικανοποίησης από τις υπηρεσίες επιβίβασης
Leg room service	Επίπεδο ικανοποίησης από τον χώρο των ποδιών
Baggage handling	Επίπεδο ικανοποίησης από την διαχείριση αποσκευών
Check -in service	Επίπεδο ικανοποίησης από την υπηρεσία check-in

Inflight service	Επίπεδο ικανοποίησης από την υπηρεσία πτήσης
Cleanliness	Επίπεδο ικανοποίησης από την καθαριότητα
Departure Delay	Πλήθος λεπτών καθυστέρησης αναχώρησης
Arrival Delay	Πλήθος λεπτών καθυστέρησης άφιξης
Satisfaction	Επίπεδο ικανοποίησης από την αεροπορική εταιρεία (δυνατές τιμές : satisfaction και neutral or dissatisfaction)

Πίνακας 1 : Στήλες συνόλου δεδομένων

Επιπλέον, στον επόμενο πίνακα, παρουσιάζονται οι διαστάσεις του συνόλου δεδομένων :

Πλήθος γραμμών	129880
Πλήθος στηλών	23

Πίνακας 2 : Διαστάσεις συνόλου δεδομένων

Ακολουθεί επίσης πίνακας με το πλήθος των ελλιπών τιμών ανά στήλη του συνόλου δεδομένων :

Όνομα Στήλης	Πλήθος ελλιπών τιμών
Gender	0
Customer Type	0

Age	0
Type of Travel	0
Class	0
Flight Distance	0
Inflight wifi service	0
Departure/Arrival time convenient	0
Ease on Online booking	0
Gate location	0
Food and drink	0
Online boarding	0
Seat comfort	0
Inflight entertainment	0
On-board service	0
Leg room service	0
Baggage handling	0
Check -in service	0
Inflight service	0
Cleanliness	0
Departure Delay	0

Arrival Delay	393
Satisfaction	0

Πίνακας 3 : Πλήθος ελλειπών τιμών ανά στήλη του συνόλου δεδομένων

Επιπλέον, παρατίθενται οι στατιστικές ιδιότητες ενός μέρους των στηλών του συνόλου δεδομένων :

	Age	Flight Distance	Inflight wifi service	Departure/Arrival time convenient	Ease of Online booking	Gate location	Food and drink	Online boarding
count	129880.000000	129880.000000	129880.000000	129880.000000	129880.000000	129880.000000	129880.000000	129880.000000
mean	39.427957	1190.316392	2.728696	3.057599	2.756876	2.976925	3.204774	3.252633
std	15.119360	997.452477	1.329340	1.526741	1.401740	1.278520	1.329933	1.350719
min	7.000000	31.000000	0.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	27.000000	414.000000	2.000000	2.000000	2.000000	2.000000	2.000000	2.000000
50%	40.000000	844.000000	3.000000	3.000000	3.000000	3.000000	3.000000	3.000000
75%	51.000000	1744.000000	4.000000	4.000000	4.000000	4.000000	4.000000	4.000000
max	85.000000	4983.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000

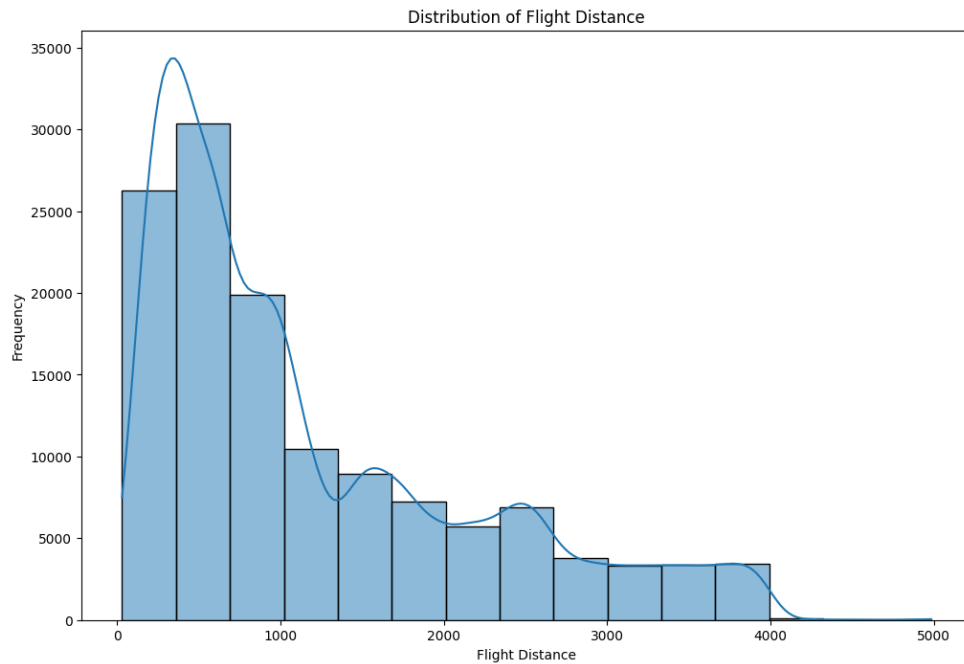
Πίνακας 4 : Στατιστικές ιδιότητες στηλών του συνόλου δεδομένων – Μέρος 1

Ακολουθούν οι στατιστικές ιδιότητες ενός δεύτερου μέρους των στηλών του συνόλου δεδομένων :

	Seat comfort	Inflight entertainment	On-board service	Leg room service	Baggage handling	Checkin service	Inflight service	Cleanliness	Departure Delay in Minutes	Arrival Delay in Minutes
count	129880.000000	129880.000000	129880.000000	129880.000000	129880.000000	129880.000000	129880.000000	129880.000000	129880.000000	129487.000000
mean	3.441361	3.358077	3.383023	3.350878	3.632114	3.306267	3.642193	3.286326	14.713713	15.091129
std	1.319289	1.334049	1.287099	1.316252	1.180025	1.266185	1.176669	1.313682	38.071126	38.465650
min	0.000000	0.000000	0.000000	0.000000	1.000000	0.000000	0.000000	0.000000	0.000000	0.000000
25%	2.000000	2.000000	2.000000	2.000000	3.000000	3.000000	3.000000	2.000000	0.000000	0.000000
50%	4.000000	4.000000	4.000000	4.000000	4.000000	3.000000	4.000000	3.000000	0.000000	0.000000
75%	5.000000	4.000000	4.000000	4.000000	5.000000	4.000000	5.000000	4.000000	12.000000	13.000000
max	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	5.000000	1592.000000	1584.000000

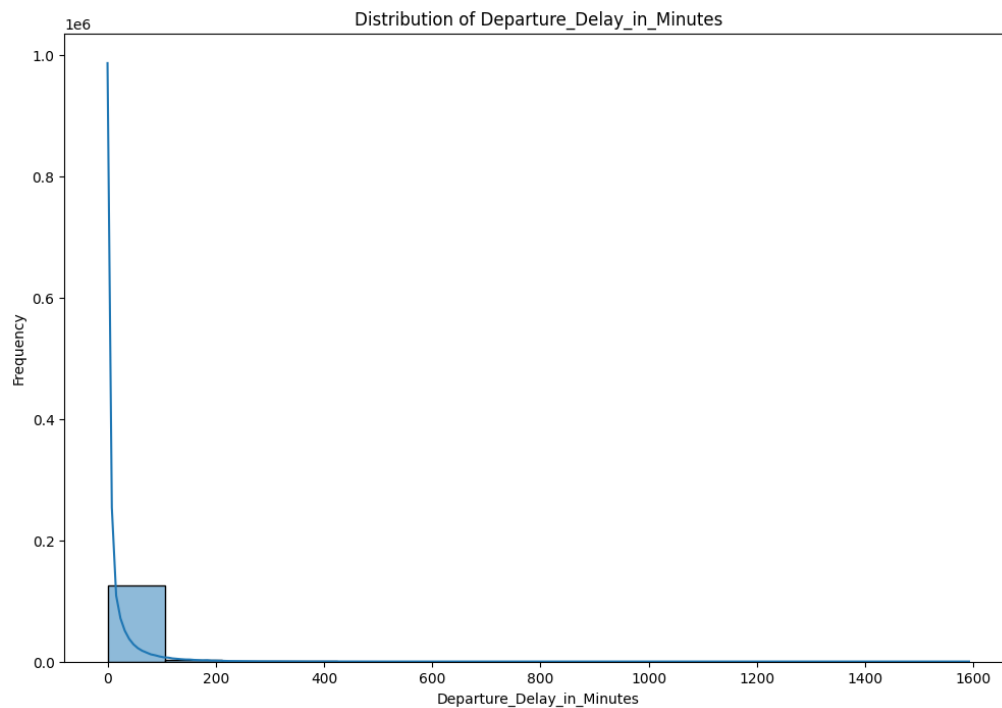
Πίνακας 5 : Στατιστικές ιδιότητες στηλών του συνόλου δεδομένων - Μέρος 2

Το ιστόγραμμα συχνοτήτων της στήλης Flight Distance :



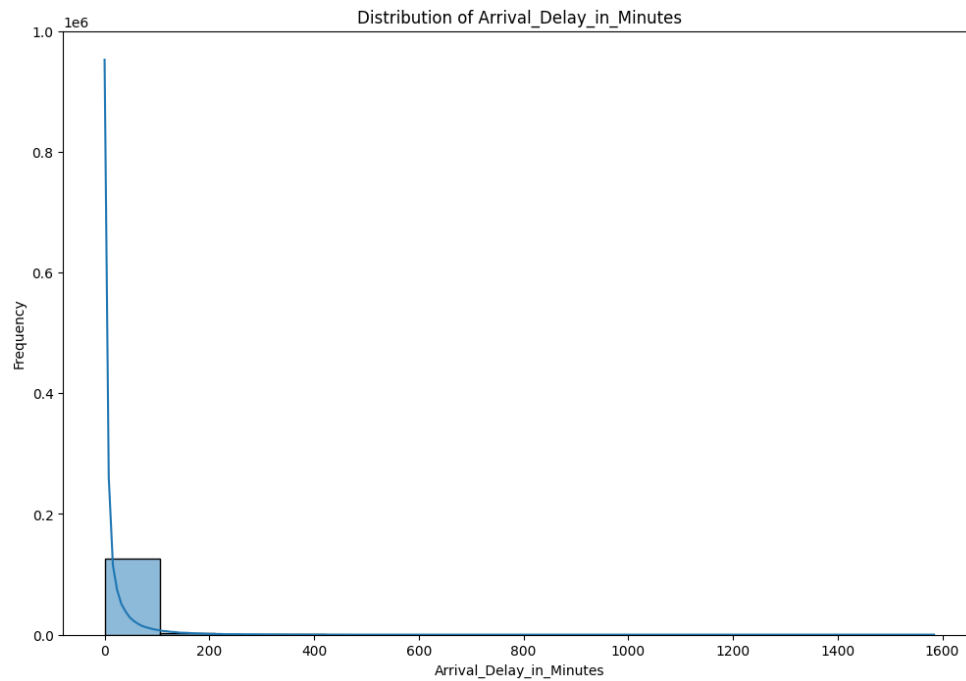
Εικόνα 2 : Ιστόγραμμα συχνοτήτων της Flight Distance

Ακολουθεί το ιστόγραμμα συχνοτήτων της στήλης Departure Delay :



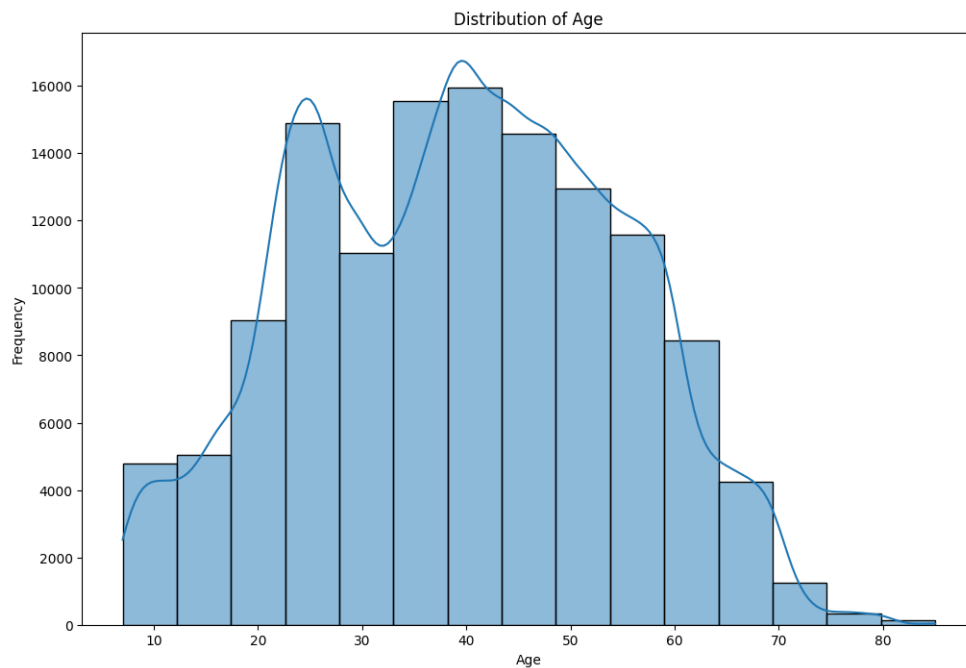
Εικόνα 3 : Ιστόγραμμα συχνοτήτων της Departure Delay

Ακολουθεί το ιστόγραμμα συχνοτήτων της στήλης Arrival Delay :



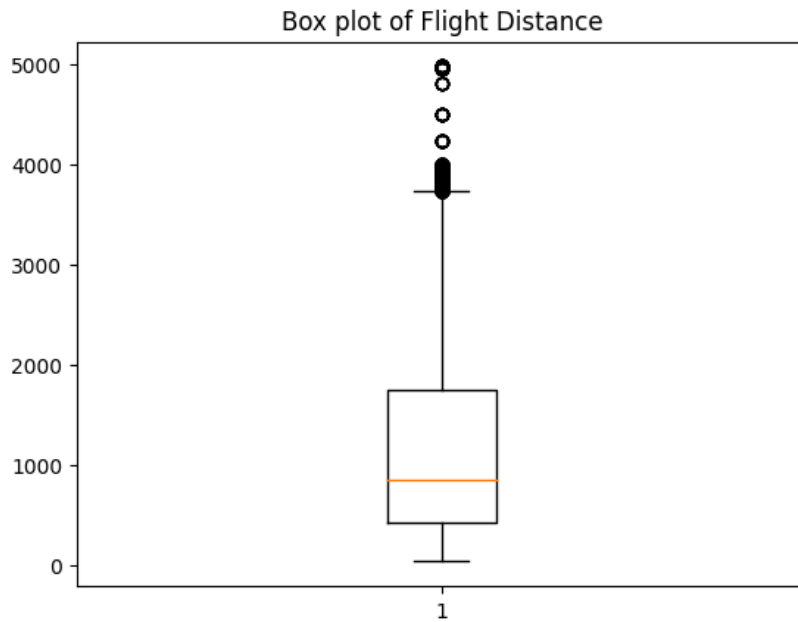
Εικόνα 4 : Ιστόγραμμα συχνοτήτων της Arrival Delay

Το αντίστοιχο ιστόγραμμα για τη στήλη Age :

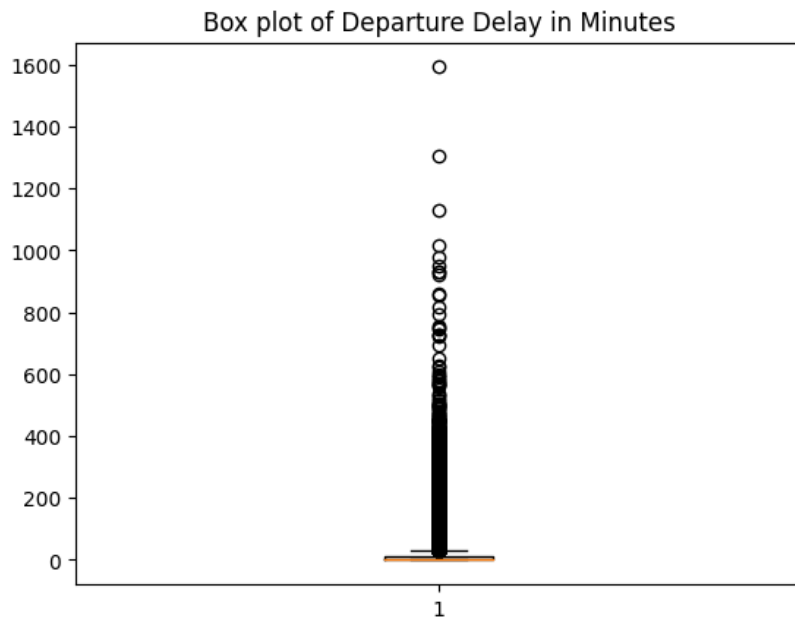


Εικόνα 5 : Ιστόγραμμα συχνοτήτων της Age

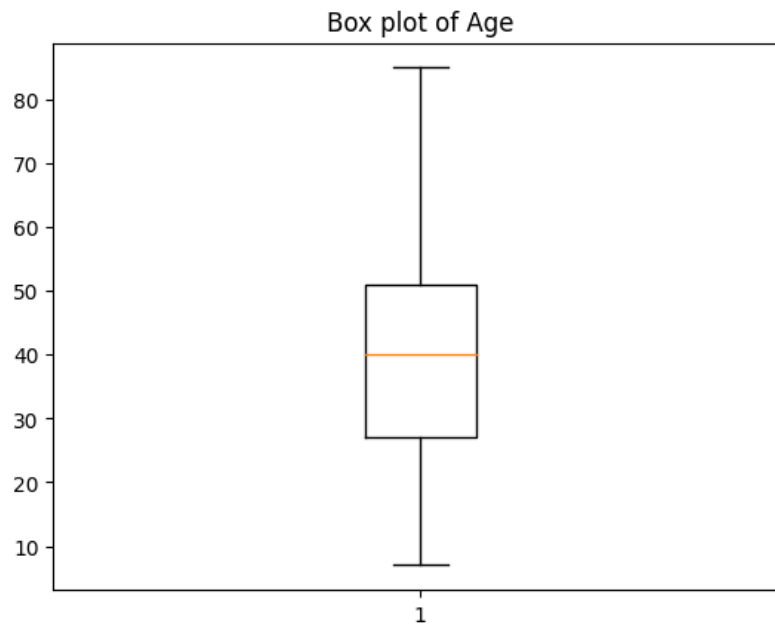
Στην συνέχεια, ακολουθούν τα BoxPlot διαγράμματα για τις στήλες : Flight Distance, Departure Delay, και Age :



Εικόνα 6 : Box Plot Διάγραμμα της Flight Distance

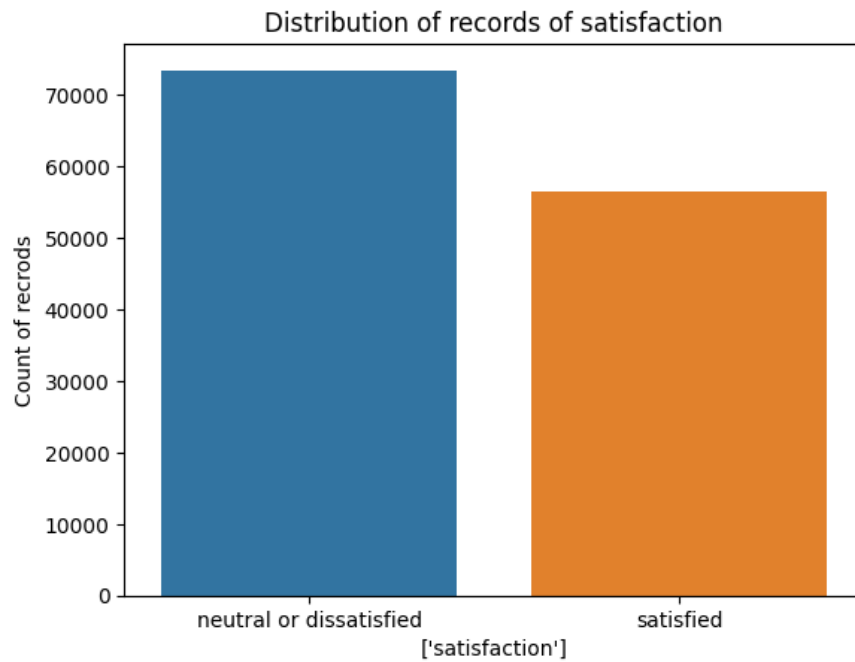


Εικόνα 7 : Box Plot Διάγραμμα της Departure Delay



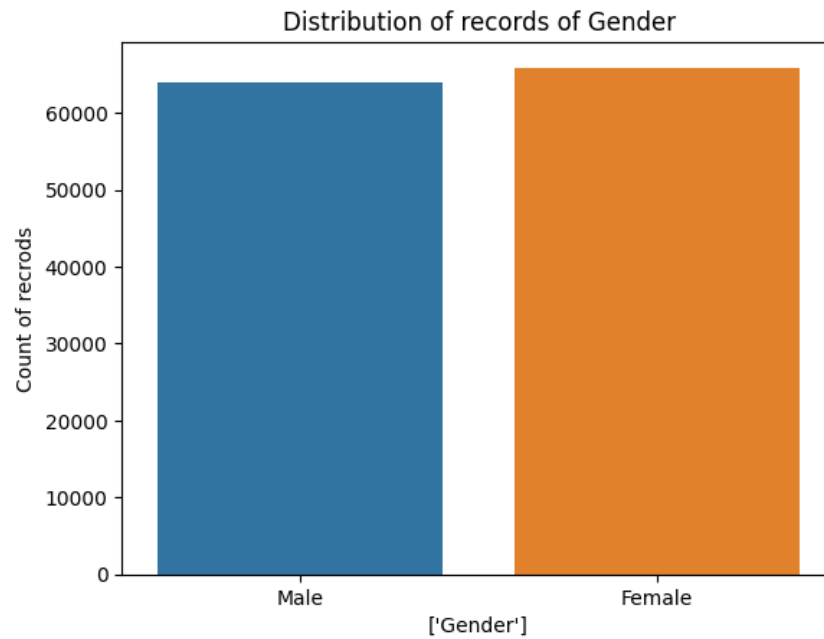
Εικόνα 8 : Box Plot Διάγραμμα της Age

Επιπλέον, παρατίθεται η κατανομή των εγγραφών όσο αφορά τη στήλη satisfaction :



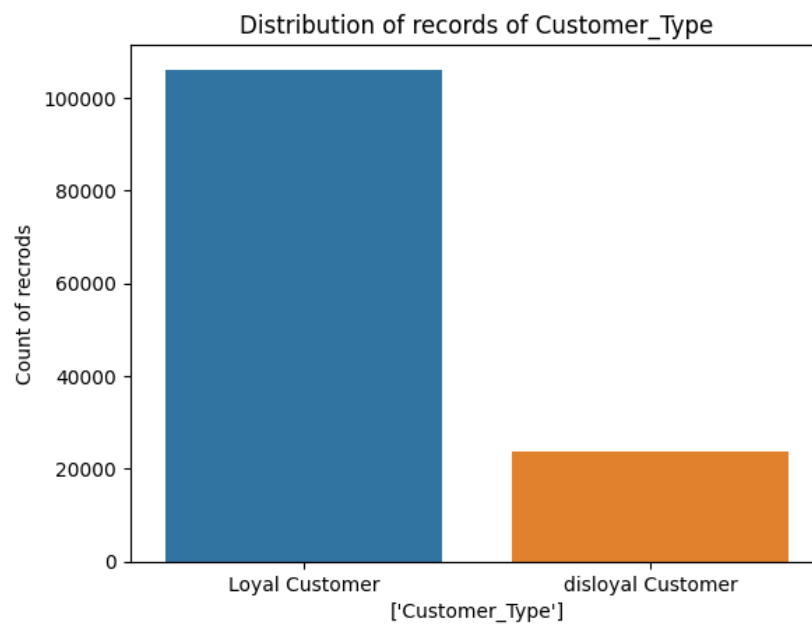
Εικόνα 9 : Κατανομή εγγραφών για τη στήλη satisfaction

Αντίστοιχο γράφημα όσο αφορά τη στήλη Gender :



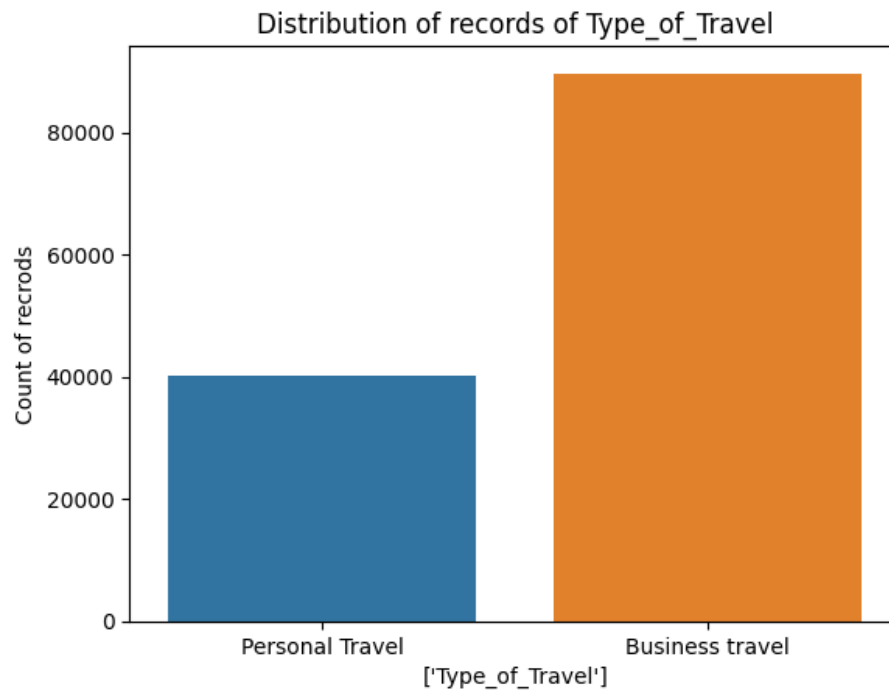
Εικόνα 10 : Κατανομή εγγραφών για τη στήλη Gender

Αντίστοιχο γράφημα όσο αφορά τη στήλη Customer Type :



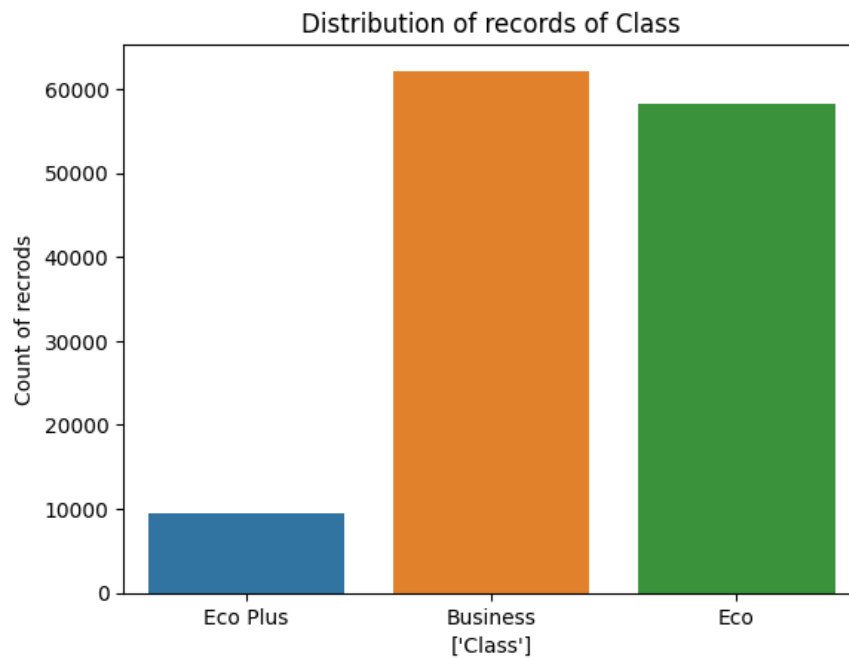
Εικόνα 11 : Κατανομή εγγραφών για τη στήλη Gender Type

Αντίστοιχο γράφημα όσο αφορά τη στήλη Type of Travel :



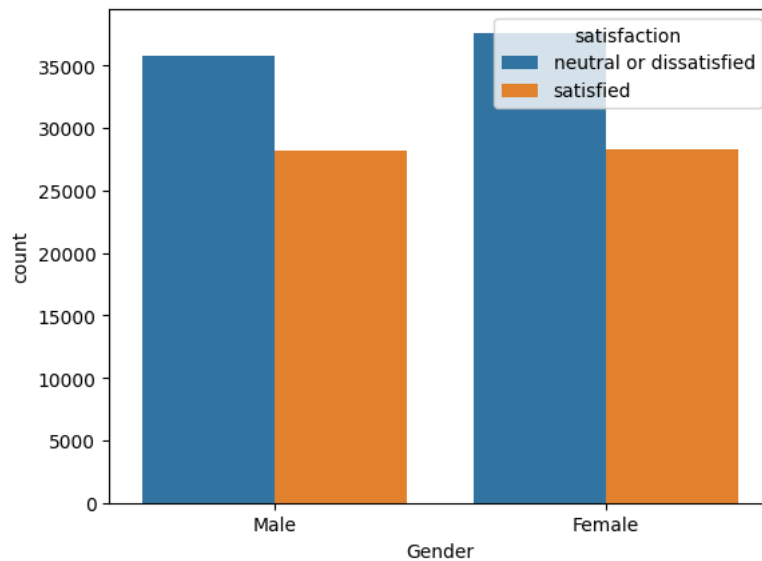
Εικόνα 12 : Κατανομή εγγραφών για τη στήλη Type of Travel

Ένα τελευταίο τέτοιου τύπου γράφημα όσο αφορά τη στήλη Class :



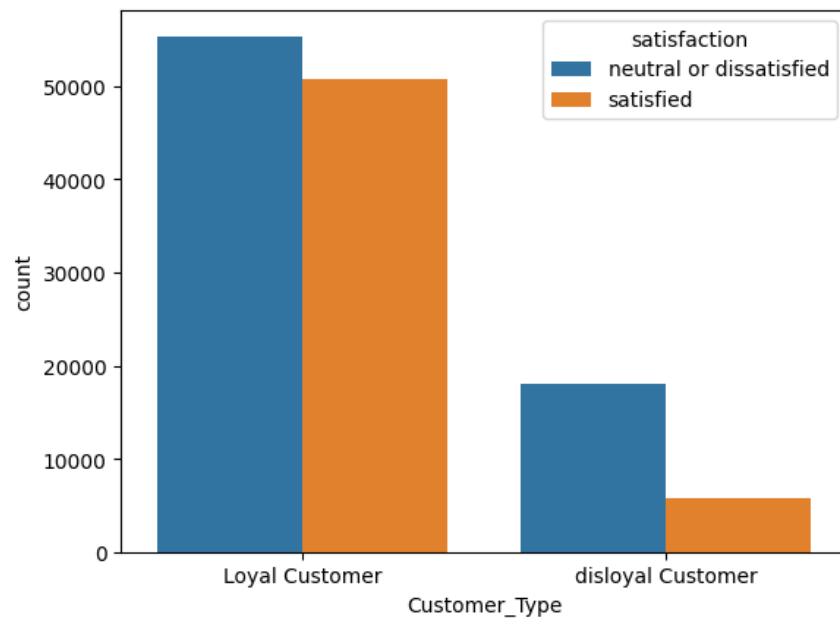
Εικόνα 13 : Κατανομή εγγραφών για τη στήλη Class

Ένα συνδυαστικό γράφημα κατανομής των εγγραφών όσο αφορά τις στήλες satisfaction και Gender :



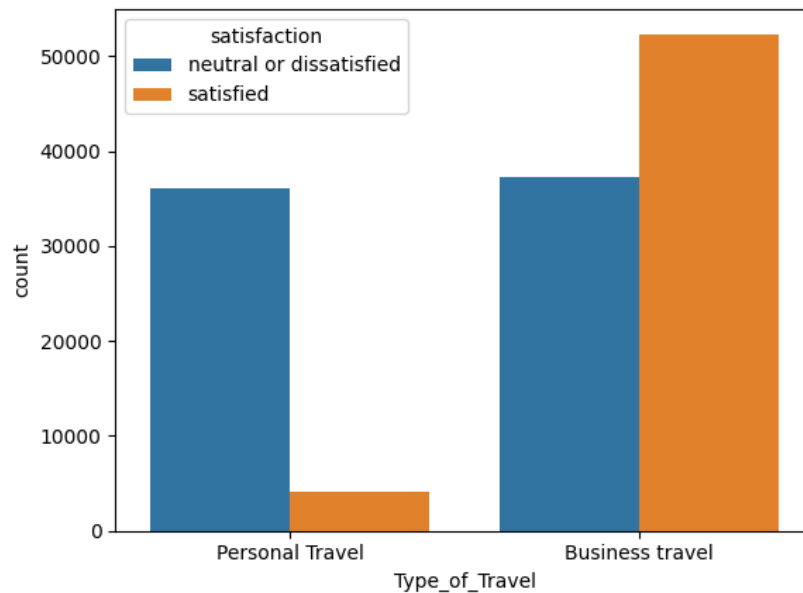
Εικόνα 14 : Συνδυαστικό γράφημα κατανομής εγγραφών για τις στήλες satisfaction και Gender

Το αντίστοιχο συνδυαστικό γράφημα κατανομής των εγγραφών όσο αφορά τις στήλες satisfaction και Customer Type :



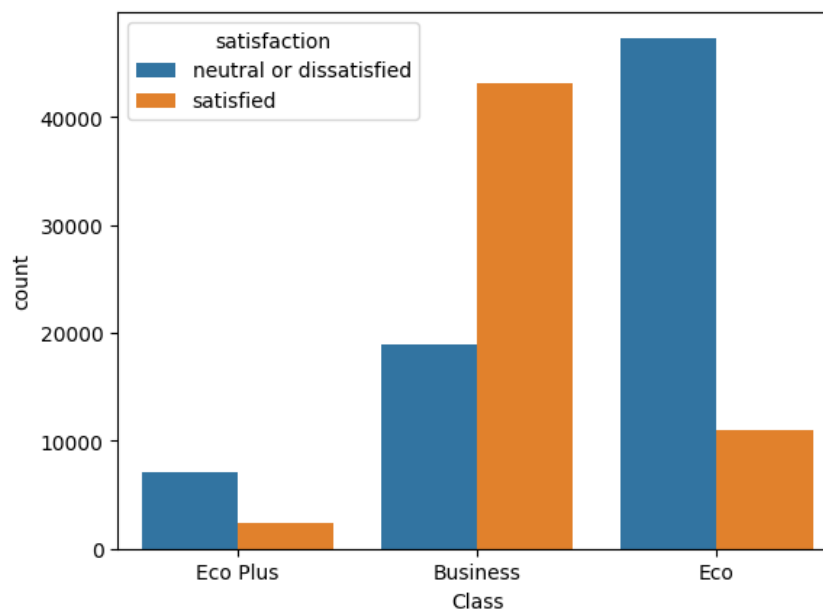
Εικόνα 15 : Συνδυαστικό γράφημα κατανομής για τις στήλες satisfaction και Customer Type

Το αντίστοιχο συνδυαστικό γράφημα κατανομής των εγγραφών όσο αφορά τις στήλες satisfaction και Type of Travel :



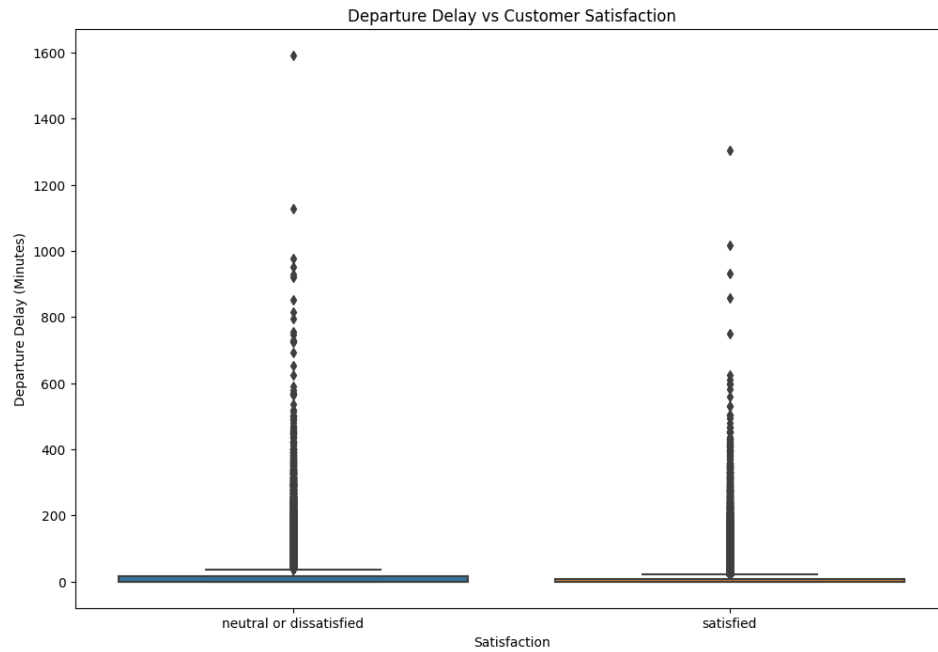
Εικόνα 16 : Συνδυαστικό γράφημα κατανομής εγγραφών για τις στήλες satisfaction και Type of Travel

Το τελευταίο συνδυαστικό γράφημα κατανομής των εγγραφών όσο αφορά τις στήλες satisfaction και Class :



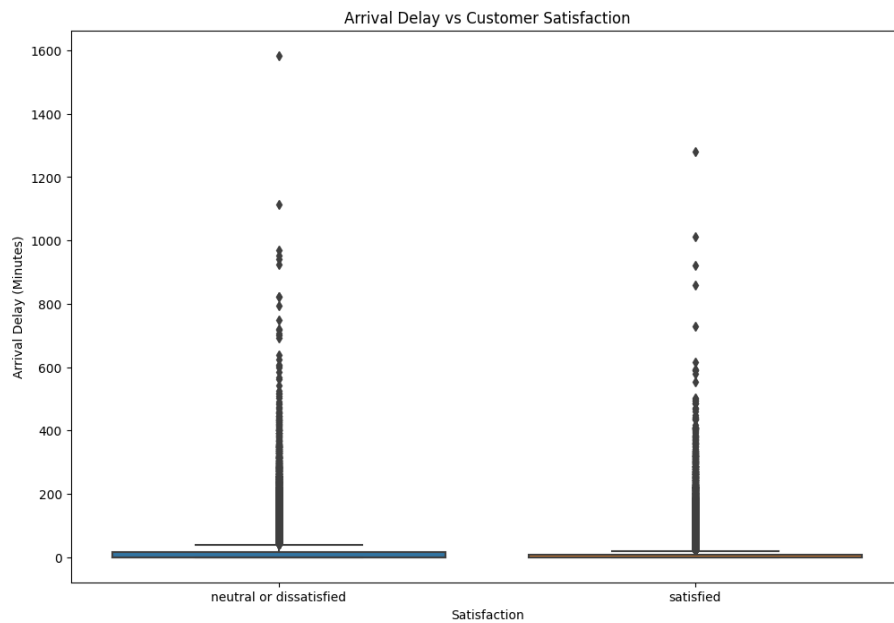
Εικόνα 17 : Συνδυαστικό γράφημα κατανομής εγγραφών για τις στήλες satisfaction και Class

Ακολουθεί το γράφημα συσχέτισης μεταξύ των στηλών Departure Delay και satisfaction :



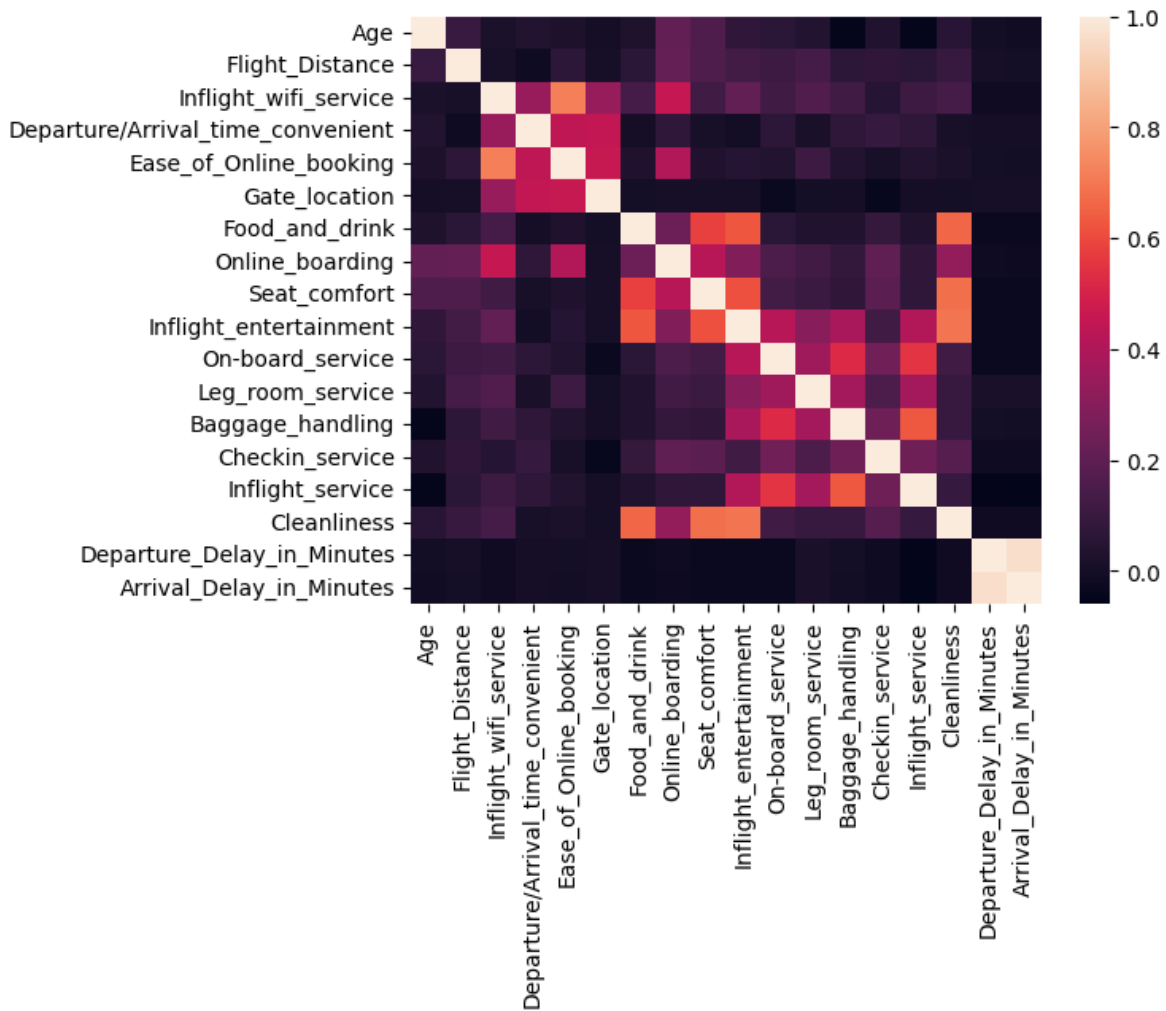
Εικόνα 18 : γράφημα συσχέτισης μεταξύ των στηλών *Departure Delay* και *satisfaction*

Και το γράφημα συσχέτισης μεταξύ των στηλών Arrival Delay και satisfaction :



Εικόνα 19 : Γράφημα συσχέτισης μεταξύ των στηλών *Arrival Delay* και *satisfaction*

Τέλος, παρατίθεται το μητρώο συσχέτισης μεταξύ των στηλών του συνόλου δεδομένων :



Εικόνα 20 : Μητρώο Συσχέτισης μεταξύ των στηλών του συνόλου δεδομένων

3.3. Προ-επεξεργασία Δεδομένων

Όσο αφορά την προ-επεξεργασία των δεδομένων, αρχικά δόθηκε λύση στο πρόβλημα ύπαρξης ελλιπών τιμών της στήλης Arrival Delay. Η τεχνική η οποία χρησιμοποιήθηκε ήταν η απόδοση της μέσης τιμής της στήλης στις ελλιπείς τιμές.

Επιπλέον, προκειμένου να καταστεί εφικτή η εφαρμογή των μοντέλων μηχανικής μάθησης, πραγματοποιήθηκε αντιστοίχιση αλφαριθμητικών με ακέραιες τιμές για τις παρακάτω στήλες :

- Customer Type
- Satisfaction
- Gender
- Type Of Travel
- Class

Επιπροσθέτως, αξίζει να αναφερθεί ότι εφαρμόστηκε η τεχνική MinMax για την κανονικοποίηση των τιμών των παρακάτω στηλών :

- Flight Distance
- Arrival Delay
- Departure Delay

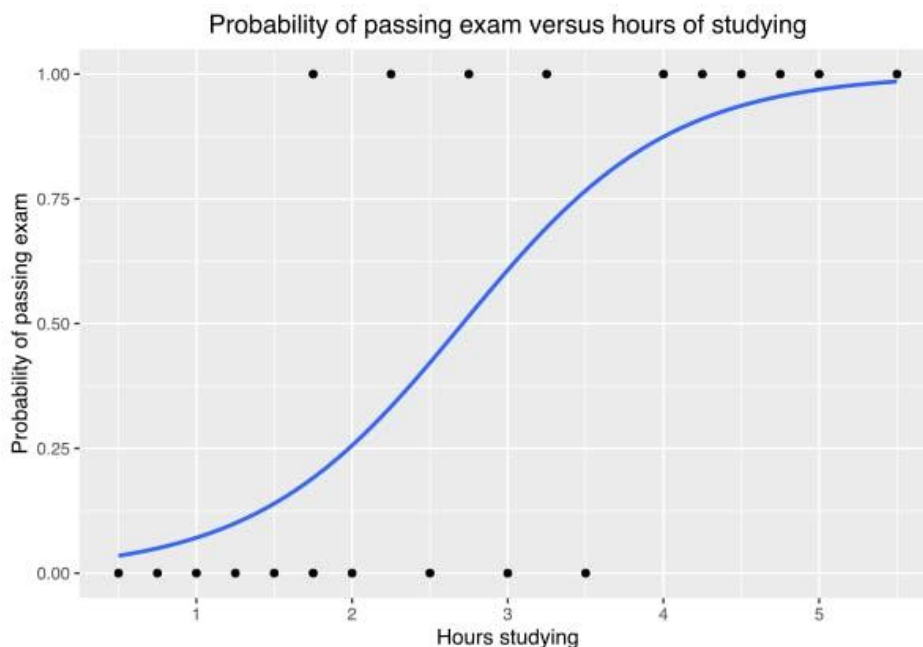
Τέλος, η αναλογία των δεδομένων εκπαίδευσης και δεδομένων ελέγχου σε όλες τις περιπτώσεις μοντέλων ήταν 80% - 20%.

3.4. Χρησιμοποιούμενα Μοντέλα

3.4.1. Λογιστική Παλινδρόμηση

Η λογιστική παλινδρόμηση είναι ένα στατιστικό μοντέλο που χρησιμοποιείται για την πρόβλεψη της πιθανότητας ενός δυαδικού αποτελέσματος με βάση μία ή περισσότερες ανεξάρτητες μεταβλητές [99–101]. Ο πρωταρχικός σκοπός της στη μηχανική μάθηση είναι να ταξινομήσει τα δεδομένα σε διαφορετικές κατηγορίες και να κατανοήσει τη σχέση μεταξύ των ανεξάρτητων μεταβλητών και των μεταβλητών αποτελεσμάτων.

Η θεμελιώδης διαφορά μεταξύ γραμμικής και λογιστικής παλινδρόμησης έγκειται στη μεταβλητή έκβασης. Η γραμμική παλινδρόμηση χρησιμοποιείται όταν η μεταβλητή έκβασης είναι συνεχής, ενώ η λογιστική παλινδρόμηση χρησιμοποιείται όταν η μεταβλητή έκβασης είναι δυαδική ή κατηγορηματική.



Εικόνα 21 : Μοντέλο Λογιστικής Παλινδρόμησης

Στην ανάλυση της λογιστικής παλινδρόμησης, οι παραδοχές της γραμμικότητας και της ανεξαρτησίας είναι σημαντικές επειδή διασφαλίζουν ότι οι σχέσεις μεταξύ των ανεξάρτητων και των εξαρτημένων μεταβλητών είναι συνεπείς. Αυτό επιτρέπει την πραγματοποίηση προβλέψεων που χαρακτηρίζονται από ακρίβεια. Η παραβίαση αυτών των παραδοχών μπορεί να θέσει σε κίνδυνο την εγκυρότητα της ανάλυσης και τη χρησιμότητά της στη λήψη τεκμηριωμένων αποφάσεων, υπογραμμίζοντας έτσι τη σημασία αυτών των παραδοχών.

Όσο αφορά τους τύπους της λογιστικής παλινδρόμησης είναι οι παρακάτω :

- **Δυναδική Παλινδρόμηση:** Η δυναδική λογιστική παλινδρόμηση χρησιμοποιείται όταν η μεταβλητή έκβασης έχει μόνο δύο κατηγορίες και ο στόχος είναι να προβλεφθεί η πιθανότητα μιας παρατήρησης που ανήκει σε μία από τις δύο κατηγορίες με βάση τις ανεξάρτητες μεταβλητές.
- **Πολυωνυμική Παλινδρόμηση:** Η πολυωνυμική λογιστική παλινδρόμηση χρησιμοποιείται όταν η μεταβλητή έκβασης έχει περισσότερες από δύο κατηγορίες. Σε αυτή την περίπτωση, το μοντέλο λογιστικής παλινδρόμησης θα εκτιμήσει τις πιθανότητες μιας παρατήρησης που ανήκει σε κάθε κατηγορία σε σχέση με μια κατηγορία αναφοράς με βάση τις ανεξάρτητες μεταβλητές.

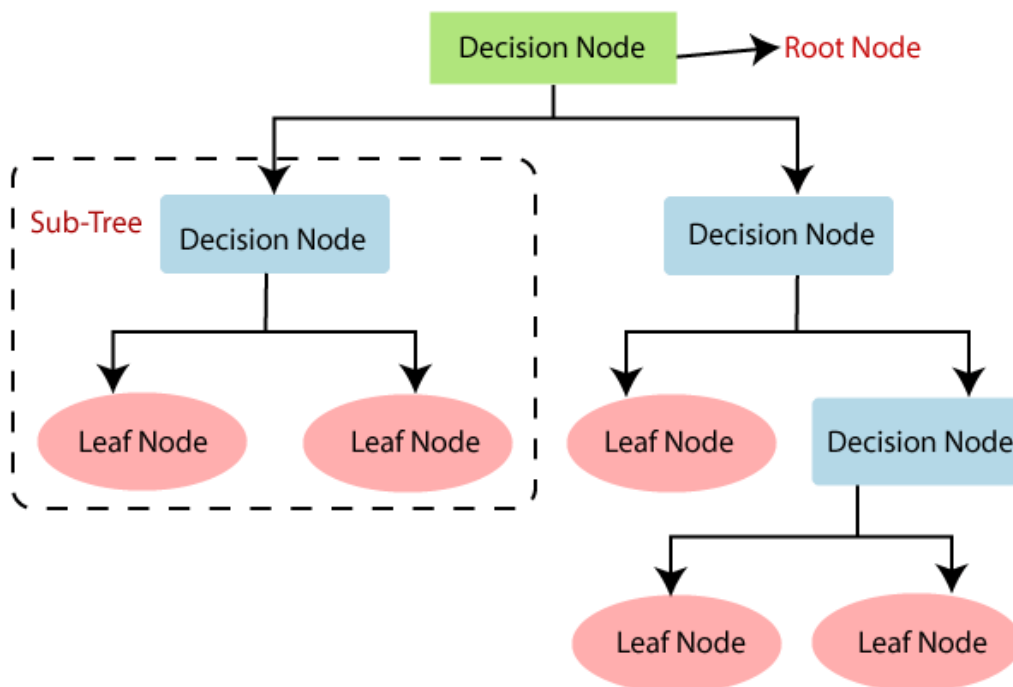
Τέλος, όσο αφορά τις προκλήσεις του μοντέλου λογιστικής παλινδρόμησης, συνοψίζονται στη συνέχεια :

- **Μη ισορροπημένα σύνολα δεδομένων:** Τα μη ισορροπημένα σύνολα δεδομένων οδηγούν σε προκατειλημμένες προβλέψεις προς την τάξη της πλειοψηφίας και οδηγούν σε ανακριβείς αξιολογήσεις για την τάξη της μειοψηφίας. Αυτή η ανισότητα στην αναπαράσταση της τάξης εμποδίζει την ικανότητα του μοντέλου να λαμβάνει σωστά υπόψη την λιγότερο αντιπροσωπευόμενη ομάδα, επηρεάζοντας τη συνολική προγνωστική του απόδοση.
- **Πολυγραμμικότητα:** Η πολυγραμμικότητα προκύπτει από πολύ συσχετισμένες μεταβλητές πρόβλεψης, καθιστώντας δύσκολο τον προσδιορισμό των μεμονωμένων επιπτώσεων κάθε μεταβλητής στο αποτέλεσμα. Η ισχυρή αλληλεξάρτηση μεταξύ των προγνωστικών περιπλέκει περαιτέρω τη διαδικασία μοντελοποίησης, επηρεάζοντας την αξιοπιστία της ανάλυσης λογιστικής παλινδρόμησης

3.4.2. Decision Tree

Το δέντρο αποφάσεων είναι μια εποπτευόμενη τεχνική μάθησης που μπορεί να χρησιμοποιηθεί τόσο για προβλήματα ταξινόμησης όσο και για προβλήματα παλινδρόμησης, αλλά κυρίως προτιμάται για την επίλυση προβλημάτων ταξινόμησης [102–104]. Είναι ένας ταξινομητής με δομή δέντρου, όπου οι εσωτερικοί κόμβοι αντιπροσωπεύουν τα χαρακτηριστικά ενός συνόλου δεδομένων, οι κλάδοι αντιπροσωπεύουν τους κανόνες απόφασης και κάθε κόμβος φύλλου αντιπροσωπεύει το αποτέλεσμα.

Σε ένα δέντρο αποφάσεων, υπάρχουν δύο κόμβοι, οι οποίοι είναι ο κόμβος απόφασης και ο κόμβος φύλλου. Οι κόμβοι αποφάσεων χρησιμοποιούνται για τη λήψη οποιασδήποτε απόφασης και έχουν πολλαπλούς κλάδους, ενώ οι κόμβοι φύλλων είναι η έξοδος αυτών των αποφάσεων και δεν περιέχουν περαιτέρω κλάδους.



Εικόνα 22 : Decision Tree

Οι αποφάσεις εκτελούνται με βάση τα χαρακτηριστικά του δεδομένου συνόλου δεδομένων. Είναι μια γραφική αναπαράσταση που περιλαμβάνει όλες τις πιθανές

λύσεις σε ένα πρόβλημα/απόφαση με βάση δεδομένες συνθήκες. Ονομάζεται δέντρο απόφασης επειδή, παρόμοιο με ένα δέντρο, ξεκινά με τον ριζικό κόμβο, ο οποίος επεκτείνεται σε περαιτέρω κλαδιά και κατασκευάζει μια δομή που μοιάζει με δέντρο.

Σε ένα δέντρο αποφάσεων, για την πρόβλεψη της τάξης του δεδομένου συνόλου δεδομένων, ο αλγόριθμος ξεκινά από τον ριζικό κόμβο του δέντρου. Στη συνέχεια ο αλγόριθμος συγκρίνει τις τιμές του χαρακτηριστικού ρίζας με το χαρακτηριστικό εγγραφής (πραγματικό σύνολο δεδομένων) και, με βάση τη σύγκριση, ακολουθεί τον κλάδο και μεταπηδά στον επόμενο κόμβο.

Για τον επόμενο κόμβο, ο αλγόριθμος συγκρίνει και πάλι την τιμή του χαρακτηριστικού με τους άλλους υπο-κόμβους και προχωρά περαιτέρω. Συνεχίζει τη διαδικασία μέχρι να φτάσει στον κόμβο - φύλλο του δέντρου. Η πλήρης διαδικασία μπορεί να γίνει καλύτερα κατανοητή χρησιμοποιώντας τον παρακάτω αλγόριθμο :

Βήμα-1: Ξεκινά με τον ριζικό κόμβο, έστω S , ο οποίος περιέχει το πλήρες σύνολο δεδομένων.

Βήμα-2: Βρίσκει το καλύτερο χαρακτηριστικό στο σύνολο δεδομένων χρησιμοποιώντας το μέτρο επιλογής χαρακτηριστικών (ASM).

Βήμα-3: Χωρίζει το S σε υποσύνολα που περιέχουν πιθανές τιμές για τα καλύτερα χαρακτηριστικά.

Βήμα-4: Δημιουργεί τον κόμβο του δέντρου, ο οποίος περιέχει το καλύτερο χαρακτηριστικό.

Βήμα-5: Δημιουργεί αναδρομικά νέα δέντρα αποφάσεων χρησιμοποιώντας τα υποσύνολα του συνόλου δεδομένων που δημιουργήθηκαν στο βήμα 3. Συνεχίζει αυτή τη διαδικασία μέχρι να επιτευχθεί ένα στάδιο όπου δεν μπορεί να ταξινομήσει περαιτέρω τους κόμβους.

Συνοπτικά, τα πλεονεκτήματα χρήσης του δέντρου αποφάσεων είναι τα παρακάτω :

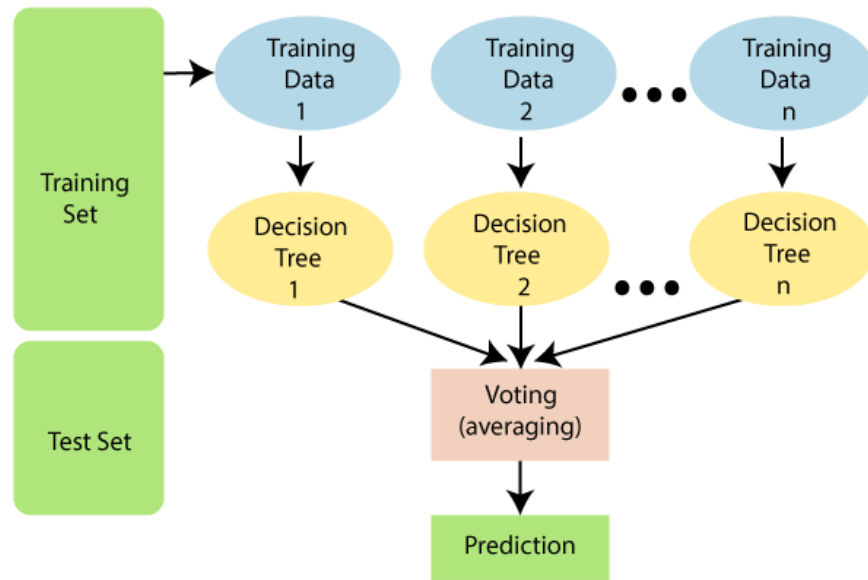
- Είναι απλό να κατανοηθεί καθώς ακολουθεί την ίδια διαδικασία που ακολουθεί ένας άνθρωπος ενώ λαμβάνει οποιαδήποτε απόφαση στην πραγματική ζωή.
- Μπορεί να είναι πολύ χρήσιμο για την επίλυση προβλημάτων που σχετίζονται με αποφάσεις.
- Υπάρχει λιγότερη απαίτηση καθαρισμού δεδομένων σε σύγκριση με άλλους αλγόριθμους.

3.4.3. Random Forest

Ο αλγόριθμος Random Forest είναι ένας δημοφιλής αλγόριθμος μηχανικής μάθησης που ανήκει στην τεχνική εποπτευόμενης μάθησης [105–107]. Μπορεί να χρησιμοποιηθεί τόσο για προβλήματα ταξινόμησης όσο και για προβλήματα παλινδρόμησης σε μηχανική μάθηση. Βασίζεται στην έννοια της μάθησης του συνόλου, η οποία είναι μια διαδικασία συνδυασμού πολλαπλών ταξινομητών για την επίλυση ενός σύνθετου προβλήματος και τη βελτίωση της απόδοσης του μοντέλου.

Όπως υποδηλώνει το όνομά του, είναι ένας ταξινομητής που περιέχει έναν αριθμό δέντρων αποφάσεων σε διάφορα υποσύνολα του δεδομένου συνόλου δεδομένων και παίρνει τον μέσο όρο για να βελτιώσει την προγνωστική ακρίβεια αυτού του συνόλου δεδομένων. Αντί να βασίζεται σε ένα δέντρο αποφάσεων, ο Random Forest λαμβάνει την πρόβλεψη από κάθε δέντρο και με βάση τη μέθοδο της πλειοψηφίας, προβλέπει την τελική έξοδο.

Ο μεγαλύτερος αριθμός δέντρων στο δάσος οδηγεί σε μεγαλύτερη ακρίβεια και αποτρέπει το πρόβλημα της υπερβολικής εκπαίδευσης.



Εικόνα 23 : Random Forest

Όσο αφορά τα βήματα λειτουργίας του αλγορίθμου, παρατίθενται στη συνέχεια :

Η διαδικασία εργασίας μπορεί να εξηγηθεί στα παρακάτω βήματα και διάγραμμα :

Βήμα-1:Επιλέγει K τυχαία σημεία δεδομένων από το σετ εκπαίδευσης.

Βήμα-2:Δημιουργεί τα δέντρα αποφάσεων που σχετίζονται με τα επιλεγμένα σημεία δεδομένων (υποσύνολα).

Βήμα-3:Επιλέγει τον αριθμό N για δέντρα αποφάσεων που θέλει να δημιουργήσει.

Βήμα-4:Επαναλαμβάνει Το Βήμα 1 & 2.

Βήμα-5:Για νέα σημεία δεδομένων, λαμβάνει τις προβλέψεις κάθε δέντρου αποφάσεων και αντιστοιχίζει τα νέα σημεία δεδομένων στην κατηγορία που πλειοψηφεί.

Συνοπτικά, τα πλεονεκτήματα του μοντέλου είναι τα εξής :

- Είναι ικανό να εκτελεί τόσο εργασίες ταξινόμησης όσο και παλινδρόμησης.
- Είναι ικανό να χειρίζεται μεγάλα σύνολα δεδομένων με υψηλή διαστατικότητα.

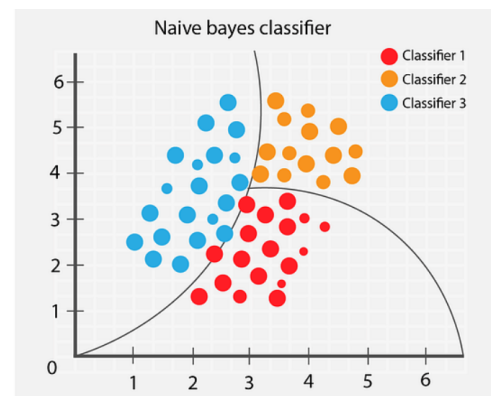
3.4.4. Naive Bayes

Ο αφελής ταξινομητής Bayes είναι ένας απλός αλλά εξαιρετικά αποτελεσματικός πιθανολογικός αλγόριθμος μηχανικής μάθησης που χρησιμοποιείται για εργασίες ταξινόμησης [108–110]. Βασίζεται στο θεώρημα του Bayes, το οποίο περιγράφει την πιθανότητα ενός γεγονότος με βάση την προηγούμενη γνώση των συνθηκών που μπορεί να σχετίζονται με το γεγονός. Η "naive" πτυχή του ταξινομητή προέρχεται από την υπόθεση ότι όλα τα χαρακτηριστικά του συνόλου δεδομένων είναι ανεξάρτητα το ένα από το άλλο, μια απλοποίηση που σπάνια ισχύει στα δεδομένα του πραγματικού κόσμου, αλλά συχνά αποδίδει εκπληκτικά καλά αποτελέσματα στην πράξη. Αυτή η υπόθεση επιτρέπει στον αλγόριθμο να υπολογίζει την πιθανότητα διαφορετικών αποτελεσμάτων εξετάζοντας κάθε χαρακτηριστικό ανεξάρτητα και οδηγώντας με τον τρόπο αυτό σε αποτελεσματικό υπολογισμό ακόμη και για μεγάλα σύνολα δεδομένων.

$$P(A|B) = \frac{P(B|A) P(A)}{P(B)}$$

using Bayesian probability terminology, the above equation can be written as

$$\text{Posterior} = \frac{\text{prior} \times \text{likelihood}}{\text{evidence}}$$



Εικόνα 24 : Naive Bayes

Ένα από τα βασικά πλεονεκτήματα του Naive Bayes είναι η απόδοσή του σε δεδομένα υψηλής διάστασης, καθιστώντας το ιδιαίτερα κατάλληλο για προβλήματα ταξινόμησης κειμένου όπως ανίχνευση ανεπιθύμητων μηνυμάτων, ανάλυση συναισθημάτων και κατηγοριοποίηση εγγράφων. Παρά την απλότητά του, μπορεί να χειριστεί αυτές τις πολύπλοκες εργασίες υπολογίζοντας την πιθανότητα ενός κειμένου που ανήκει σε μια συγκεκριμένη κατηγορία με βάση την παρουσία και τη συχνότητα των λέξεων μέσα σε αυτό.

Εφαρμόζοντας την αφελή υπόθεση, αντιμετωπίζει κάθε λέξη ανεξάρτητα και υπολογίζει την εκ των υστέρων πιθανότητα για κάθε κατηγορία, αναθέτοντας το κείμενο στην κατηγορία με την υψηλότερη πιθανότητα. Αυτό το χαρακτηριστικό επιτρέπει στον ταξινομητή να είναι γρήγορος και αποτελεσματικός, ακόμη και με μεγάλο αριθμό χαρακτηριστικών.

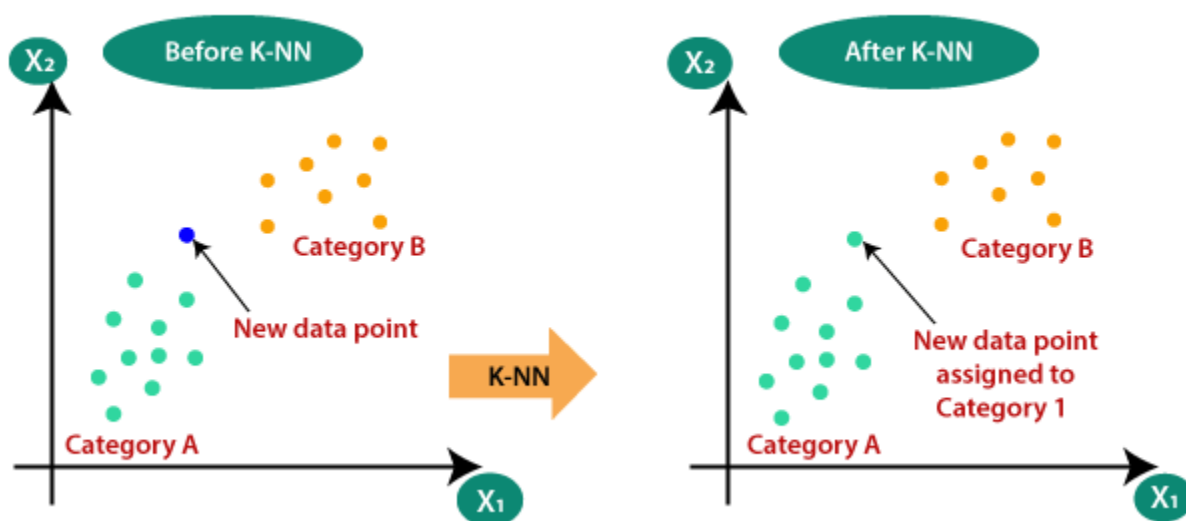
Ωστόσο, ο ταξινομητής Naive Bayes έχει τους περιορισμούς του. Η υπόθεση της ανεξαρτησίας των χαρακτηριστικών μπορεί να οδηγήσει σε μη βέλτιστη απόδοση όταν τα χαρακτηριστικά συσχετίζονται έντονα, καθώς δεν καταγράφει τις σχέσεις μεταξύ τους. Επιπλέον, μπορεί να δυσκολευτεί με σύνολα δεδομένων όπου οι πιθανότητες κλάσης είναι πολύ διαφορετικές ή όπου τα χαρακτηριστικά έχουν πολύ διαφορετικά εύρη τιμών.

3.4.5. Αλγόριθμος k-NN

Ο αλγόριθμος k-NN είναι ένας από τους απλούστερους αλγόριθμους μηχανικής μάθησης που βασίζονται στην τεχνική εποπτευόμενης μάθησης [111]. Ο αλγόριθμος k-NN υποθέτει την ομοιότητα μεταξύ της νέας περίπτωσης / δεδομένων και των διαθέσιμων περιπτώσεων και τοποθετεί τη νέα περίπτωση στην κατηγορία που είναι πιο παρόμοια με τις διαθέσιμες κατηγορίες. Αξίζει να αναφερθεί ότι αποθηκεύει όλα

τα διαθέσιμα δεδομένα και ταξινομεί ένα νέο σημείο δεδομένων με βάση την ομοιότητα. Αυτό σημαίνει ότι όταν εμφανίζονται νέα δεδομένα, τότε μπορούν εύκολα να ταξινομηθούν σε μια κατηγορία.

Ο αλγόριθμος k-NN μπορεί να χρησιμοποιηθεί για παλινδρόμηση καθώς και για ταξινόμηση, αλλά κυρίως χρησιμοποιείται για τα προβλήματα ταξινόμησης. Είναι ένας μη παραμετρικός αλγόριθμος, που σημαίνει ότι δεν κάνει καμία υπόθεση για τα υποκείμενα δεδομένα. Ο αλγόριθμος k - NN στη φάση εκπαίδευσης αποθηκεύει μόνο το σύνολο δεδομένων και όταν παίρνει νέα δεδομένα, τότε ταξινομεί αυτά τα δεδομένα σε μια κατηγορία που είναι πολύ παρόμοια με τα νέα δεδομένα.



Εικόνα 25 : Αλγόριθμος k-NN

Τα βήματα λειτουργίας του αλγορίθμου συνοψίζονται στην συνέχεια :

Βήμα-1:Επιλέγει τον αριθμό k των γειτόνων.

Βήμα-2:Υπολογίζει την ευκλείδεια απόσταση σε σχέση με όλα τα σημεία του συνόλου δεδομένων.

Βήμα-3 :Λαμβάνει τους k πλησιέστερους γείτονες σύμφωνα με την υπολογισμένη ευκλείδεια απόσταση.

Βήμα-4:Μεταξύ αυτών των k γειτόνων, υπολογίζει τον αριθμό των σημείων δεδομένων σε κάθε κατηγορία.

Βήμα-5:Αντιστοιχίζει τα νέα σημεία δεδομένων σε εκείνη την κατηγορία η οποία πλειοψηφεί.

3.5. Χρησιμοποιούμενες Μετρικές Αξιολόγησης

3.5.1. Accuracy

Η μέτρηση Accuracy είναι ένα θεμελιώδες μέτρο που χρησιμοποιείται για την αξιολόγηση της απόδοσης των αλγορίθμων ταξινόμησης [112–114]. Υποδεικνύει το ποσοστό των σωστά ταξινομημένων περιπτώσεων από τον συνολικό αριθμό περιπτώσεων στο σύνολο δεδομένων. Συγκεκριμένα, η μετρική accuracy υπολογίζεται ως ο λόγος του αριθμού των αληθινών θετικών και αληθινών αρνητικών προβλέψεων προς τον συνολικό αριθμό των προβλέψεων. Μαθηματικά, εκφράζεται ως εξής :

$$Accuracy = \frac{\text{Πλήθος σωστών προβλέψεων}}{\text{Συνολικό πλήθος προβλέψεων}}$$

Ενώ η μετρική accuracy είναι μια απλή και ευρέως χρησιμοποιούμενη μετρική, είναι πιο χρήσιμη σε σενάρια όπου οι τάξεις είναι ισορροπημένες, πράγμα που σημαίνει ότι υπάρχουν περίπου ίσοι αριθμοί περιπτώσεων σε κάθε τάξη. Ωστόσο, σε καταστάσεις με ανισορροπημένες τάξεις—όπου μια τάξη ξεπερνά σημαντικά τις άλλες—η μετρική accuracy μπορεί να είναι παραπλανητική.

3.5.2. Μετρική Precision

Η μέτρηση Precision είναι ένα κρίσιμο μέτρο για την αξιολόγηση της απόδοσης των αλγορίθμων ταξινόμησης. Η μετρική Precision ποσοτικοποιεί την ακρίβεια των θετικών προβλέψεων που γίνονται από τον ταξινομητή. Ορίζεται ως ο λόγος των αληθινών θετικών προβλέψεων προς τον συνολικό αριθμό θετικών προβλέψεων (τόσο αληθινών θετικών όσο και ψευδών θετικών). Μαθηματικά, η ακρίβεια εκφράζεται ως :

$$Precision = \frac{\text{Πλήθος αληθινών θετικών προβλέψεων}}{\text{Συνολικό πλήθος θετικών προβλέψεων}}$$

Ωστόσο, η μετρική αυτή από μόνη της δεν παρέχει μια πλήρη εικόνα της απόδοσης ενός ταξινομητή, καθώς δεν λαμβάνει υπόψη ψευδώς αρνητικές περιπτώσεις (περιπτώσεις όπου οι θετικές περιπτώσεις ταξινομούνται λανθασμένα ως αρνητικές).

3.5.3. Μετρική Recall

Η μέτρηση Recall είναι ένα βασικό μέτρο για την αξιολόγηση της απόδοσης των αλγορίθμων ταξινόμησης, ιδιαίτερα σε πλαίσια όπου ο προσδιορισμός όλων των σχετικών περιπτώσεων είναι κρίσιμος. Η μετρική αυτή ποσοτικοποιεί την ικανότητα του ταξινομητή να αναγνωρίζει σωστά όλες τις θετικές περιπτώσεις στο σύνολο δεδομένων. Ορίζεται ως ο λόγος των αληθινών θετικών προβλέψεων προς τον συνολικό αριθμό των πραγματικών θετικών περιπτώσεων (τόσο αληθινών θετικών όσο και ψευδών αρνητικών). Μαθηματικά, η μετρική εκφράζεται ως :

$$Recall = \frac{\text{Πλήθος αληθινών θετικών προβλέψεων}}{\text{Πλήθος αληθινών θετικών} + \text{Πλήθος ψευδών αρνητικών προβλέψεων}}$$

Η μετρική είναι ιδιαίτερα σημαντική σε σενάρια όπου η απουσία θετικής περίπτωσης (ψευδώς αρνητική) μπορεί να έχει σοβαρές συνέπειες.

ΚΕΦΑΛΑΙΟ 4: ΑΞΙΟΛΟΓΗΣΗ

4.1. Αποτελέσματα εφαρμογής μοντέλων

4.1.1. Ακολουθούν τα αποτελέσματα για κάθε μοντέλο μετά τον διαχωρισμό των δεδομένων σε 80% για εκπαίδευση και 20% για δοκιμή:

Όσο αφορά την μετρική Accuracy, παρατίθενται ο παρακάτω συγκριτικός πίνακας :

Μοντέλο	Τιμή Accuracy
Logistic Regression	0.873653
Decision Tree	0.943756
Random Forest	0.961233
Naïve Bayes	0.866608
KNN	0.912920

Πίνακας 6 : Αποτελέσματα μετρικής Accuracy

Όσο αφορά την μετρική Precision, παρατίθενται ο παρακάτω συγκριτικός πίνακας :

Μοντέλο	Τιμή Precision
Logistic Regression	0.834332
Decision Tree	0.934120
Random Forest	0.936586
Naïve Bayes	0.821032
KNN	0.867624

Πίνακας 7 : Αποτελέσματα μετρικής Precision

Τέλος, όσο αφορά την μετρική Recall, παρατίθενται ο παρακάτω συγκριτικός πίνακας :

Μοντέλο	Τιμή Recall
Logistic Regression	0.871161
Decision Tree	0.937009
Random Forest	0.973720
Naïve Bayes	0.866760
KNN	0.928464

Πίνακας 8 : Αποτελέσματα μετρικής Recall

4.1.2. Ακολουθούν τα αποτελέσματα για κάθε μοντέλο μετά τον διαχωρισμό των δεδομένων σε 70% για εκπαίδευση και 30% για δοκιμή:

Όσο αφορά την μετρική Accuracy, παρατίθενται ο παρακάτω συγκριτικός πίνακας :

Μοντέλο	Τιμή Accuracy
Logistic Regression	0.872036
Decision Tree	0.943794
Random Forest	0.961708
Naïve Bayes	0.866492
KNN	0.910789

Πίνακας 9 : Αποτελέσματα μετρικής Accuracy

Όσο αφορά την μετρική Precision, παρατίθενται ο παρακάτω συγκριτικός πίνακας :

Μοντέλο	Τιμή Precision
Logistic Regression	0.832205
Decision Tree	0.938004
Random Forest	0.938475
Naïve Bayes	0.821136
KNN	0.865587

Πίνακας 100 : Αποτελέσματα μετρικής Precision

Τέλος, όσο αφορά την μετρική Recall, παρατίθενται ο παρακάτω συγκριτικός πίνακας :

Μοντέλο	Τιμή Recall
Logistic Regression	0.868723
Decision Tree	0.933388
Random Forest	0.972722
Naïve Bayes	0.865682
KNN	0.865682

Πίνακας 111: Αποτελέσματα μετρικής Recall

4.1.2. Ακολουθούν τα αποτελέσματα για κάθε μοντέλο μετά τον διαχωρισμό των δεδομένων σε 60% για εκπαίδευση και 40% για δοκιμή: Όσο αφορά την μετρική Accuracy, παρατίθενται ο παρακάτω συγκριτικός πίνακας :

Μοντέλο	Τιμή Accuracy
Logistic Regression	0.871882
Decision Tree	0.944295
Random Forest	0.961176
Naïve Bayes	0.866107
KNN	0.909128

Πίνακας 12: Αποτελέσματα μετρικής Accuracy

Όσο αφορά την μετρική Precision, παρατίθενται ο παρακάτω συγκριτικός πίνακας :

Μοντέλο	Τιμή Precision
Logistic Regression	0.831485
Decision Tree	0.935652
Random Forest	0.938260
Naïve Bayes	0.818756
KNN	0.862775

Πίνακας 13 : Αποτελέσματα μετρικής Precision

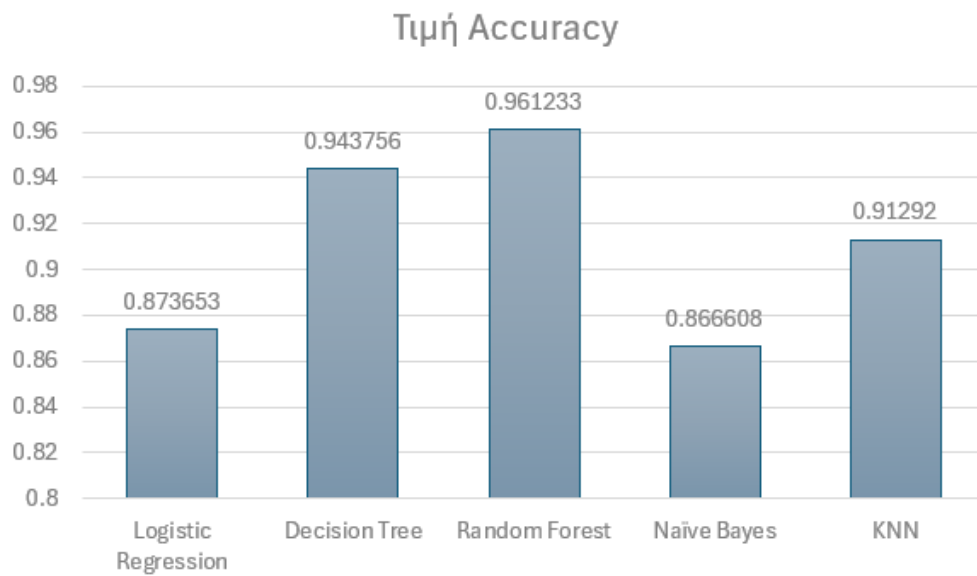
Τέλος, όσο αφορά την μετρική Recall, παρατίθενται ο παρακάτω συγκριτικός πίνακας :

Μοντέλο	Τιμή Recall
Logistic Regression	0.868726
Decision Tree	0.936397
Random Forest	0.971625
Naïve Bayes	0.866470
KNN	0.923550

Πίνακας 144: Αποτελέσματα μετρικής Recall

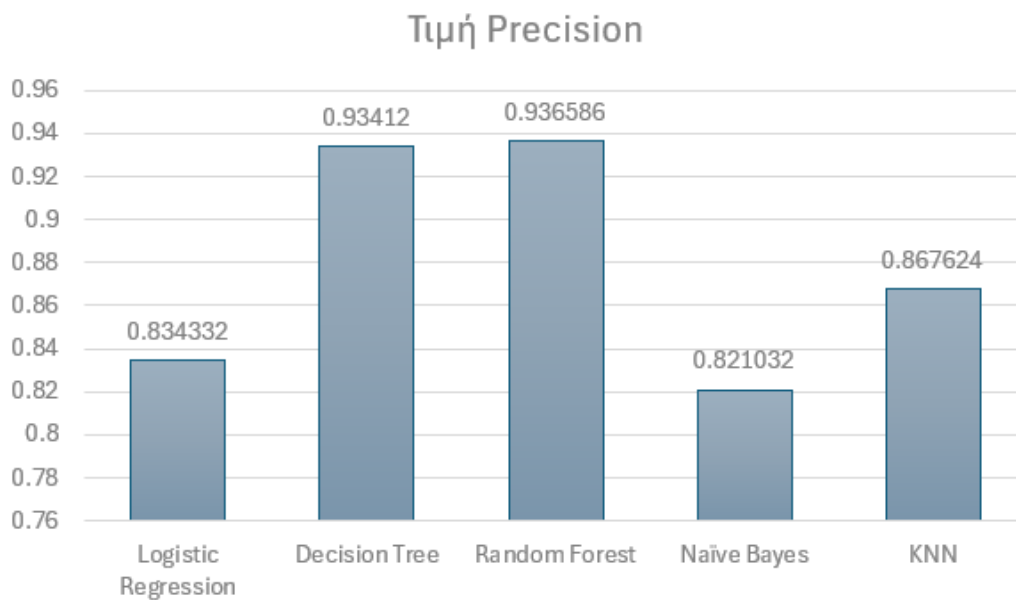
4.2. Συγκριτικά Αποτελέσματα

Για την καλύτερη οπτικοποίηση των αποτελεσμάτων και την εξαγωγή πληρέστερων συμπερασμάτων παρατίθεται το παρακάτω συγκριτικό γράφημα όσο αφορά τη μετρική Accuracy :



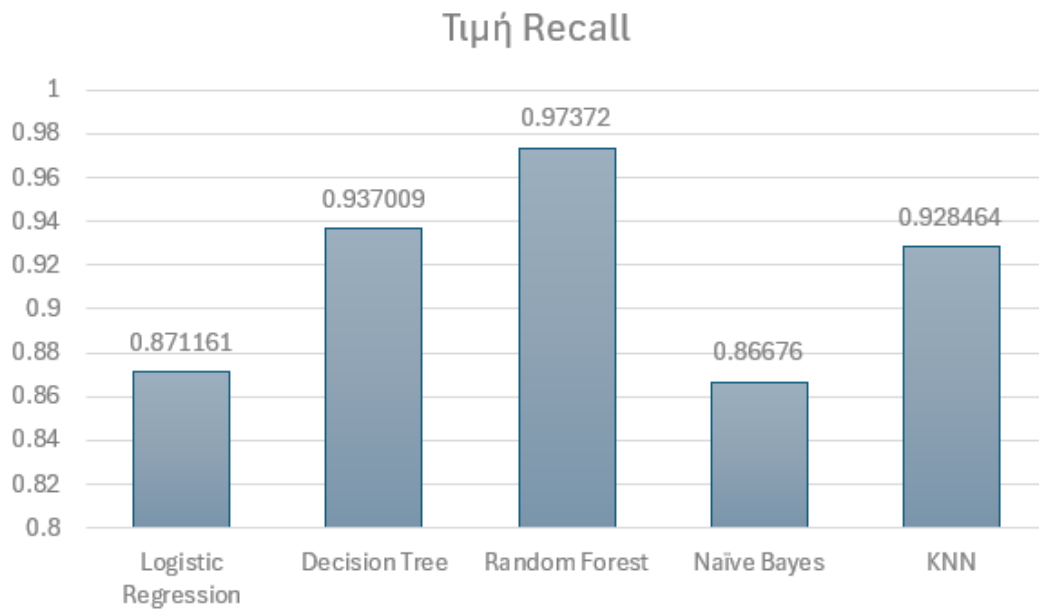
Εικόνα 26 : Συγκριτικό γράφημα για τη μετρική Accuracy

Το αντίστοιχο συγκριτικό γράφημα όσο αφορά τη μετρική Precision :



Εικόνα 27 : Συγκριτικό γράφημα για μετρική Precision

Και το γράφημα όσο αφορά τη μετρική Recall :



Εικόνα 28 : Συγκριτικό γράφημα για τη μετρική Recall

4.3. Σχολιασμός Αποτελεσμάτων

Στον παρακάτω πίνακα, για κάθε μετρική καταγράφεται το μοντέλο, στο οποίο επετεύχθησαν τα καλύτερα αποτελέσματα :

Μετρική	Μοντέλο με την καλύτερη επίδοση
Accuracy	Random Forest
Precision	Random Forest
Recall	Random Forest

Πίνακας 15 : Μοντέλο με την καλύτερη επίδοση για κάθε μετρική

Επιπλέον, παρατίθεται ο πίνακας με το μοντέλο με τη χειρότερη επίδοση για κάθε μετρική :

Μετρική	Μοντέλο με τη χειρότερη επίδοση
Accuracy	Naïve Bayes
Precision	Naïve Bayes
Recall	Naïve Bayes

Πίνακας 16 : Μοντέλο με τη χειρότερη επίδοση σε κάθε μοντέλο

Επομένως, από τους παραπάνω συγκριτικούς πίνακες διαπιστώνεται ότι το μοντέλο Random Forest παρουσίασε τη καλύτερα αποτελέσματα σε όλες τις μετρικές, σε αντίθεση με το μοντέλο Naïve Bayes. Βέβαια στο σημείο αυτό, αξίζει να αναφερθεί ότι στη περίπτωση του Random Forest, οι επιδόσεις άγγιξαν τις βέλτιστες δυνατές τιμές. Η χειρότερη συμπεριφορά του Naïve Bayes ωστόσο δεν σημαίνει άσχημες τιμές στις μετρικές (πάνω από 0.8 σε όλες τις μετρικές).

ΣΥΜΠΕΡΑΣΜΑΤΑ

Συμπερασματικά, η ενσωμάτωση της μηχανικής μάθησης στην αεροπορική βιομηχανία σηματοδοτεί ένα σημαντικό άλμα προς τα εμπρός στην ενίσχυση της λειτουργικής αποτελεσματικότητας, της ασφάλειας και της ικανοποίησης των πελατών. Η ικανότητα επεξεργασίας και ανάλυσης τεράστιων ποσοτήτων δεδομένων σε πραγματικό χρόνο επέτρεψε στις αεροπορικές εταιρείες να βελτιστοποιήσουν τις λειτουργίες τους, να μειώσουν το κόστος και να ελαχιστοποιήσουν το χρόνο διακοπής λειτουργίας. Η προγνωστική συντήρηση, που υποστηρίζεται από τη μηχανική μάθηση, διασφαλίζει ότι τα πιθανά ζητήματα εντοπίζονται πριν γιγαντωθούν, βελτιώνοντας έτσι την αξιοπιστία και τη μακροζωία των αεροσκαφών. Αυτή η τεχνολογική πρόοδος όχι μόνο ωφελεί τις αεροπορικές εταιρείες μειώνοντας τις λειτουργικές διαταραχές, αλλά και ενισχύει τη συνολική εμπειρία των επιβατών εξασφαλίζοντας πιο συνεπή και αξιόπιστη εξυπηρέτηση.

Επιπλέον, η εφαρμογή της μηχανικής μάθησης στην ασφάλεια της αεροπορίας είναι μια κρίσιμη εξέλιξη που δεν μπορεί να παραληφθεί. Με την αξιοποίηση προηγμένων αναλύσεων δεδομένων, τα μοντέλα μηχανικής μάθησης μπορούν να προβλέψουν και να μετριάσουν τους πιθανούς κινδύνους όσο αφορά την ασφάλεια, μειώνοντας έτσι σημαντικά τον κίνδυνο ατυχημάτων. Τα προληπτικά μέτρα που διευκολύνονται από αυτές τις γνώσεις είναι ζωτικής σημασίας για τη διατήρηση υψηλών προτύπων ασφάλειας στον κλάδο. Η ικανότητα πρόβλεψης και αντιμετώπισης ζητημάτων όπως μηχανικές βλάβες, κόπωση του πιλότου και δυσμενείς καιρικές συνθήκες ενισχύει την ασφάλεια τόσο των επιβατών όσο και του πληρώματος, ενισχύοντας την εμπιστοσύνη στα αεροπορικά ταξίδια.

Τέλος, ο αντίκτυπος της μηχανικής μάθησης στην εμπειρία των πελατών στην αεροπορία είναι μεγάλος. Η δυνατότητα εξατομίκευσης των υπηρεσιών και των

προσφορών με βάση τις ατομικές προτιμήσεις και συμπεριφορές έχει επαναπροσδιορίσει το ταξίδι των επιβατών. Από την προσαρμοσμένη ψυχαγωγία κατά την πτήση έως τις προσαρμοσμένες επιλογές γευμάτων και το στοχευμένο μάρκετινγκ, η μηχανική μάθηση ενισχύει την ικανοποίηση των πελατών. Καθώς η αεροπορική βιομηχανία συνεχίζει να εξελίσσεται, η υιοθέτηση της μηχανικής μάθησης υπόσχεται ακόμη μεγαλύτερες καινοτομίες και βελτιώσεις. Οι συνεχιζόμενες εξελίξεις στον τομέα αυτό πρόκειται να φέρουν επανάσταση στα αεροπορικά ταξίδια, καθιστώντας τα πιο αποτελεσματικά, ασφαλή και ευχάριστα για τους επιβάτες παγκοσμίως.

Όσο αφορά το πρακτικό τμήμα της εργασίας, τα επιλεγμένα μοντέλα κατέγραψαν πολύ ικανοποιητικά αποτελέσματα όσο αφορά τη πρόβλεψη της ικανοποίησης πελατών αεροπορικών εταιριών, με το μοντέλο Random Forest να πλησιάζει το βέλτιστα δυνατά αποτελέσματα. Επομένως, πρόκειται για μια αρκετά ελπιδοφόρα πειραματική διαδικασία, η οποία στο μέλλον μπορεί να επεκταθεί με τους παρακάτω προτεινόμενους τρόπους:

- Χρήση περισσότερων μοντέλων με σκοπό τη δημιουργία περισσότερων συγκριτικών αποτελεσμάτων
- Εύρεση ογκωδέστερου συνόλου δεδομένων, προκειμένου να πραγματοποιηθεί πληρέστερη εκπαίδευση των μοντέλων
- Δοκιμή εφαρμογής μοντέλων βαθιάς μάθησης (νευρωνικά δίκτυα)

ΠΑΡΑΡΤΗΜΑ ΚΩΔΙΚΑ

```
import pandas as pd

import numpy as np

import matplotlib.pyplot as plt

import seaborn as sns

from sklearn.linear_model import LinearRegression

from sklearn.linear_model import LogisticRegression

from sklearn.svm import LinearSVC

from sklearn.tree import DecisionTreeClassifier

from sklearn.ensemble import RandomForestClassifier

from sklearn.naive_bayes import GaussianNB

from sklearn.neighbors import KNeighborsClassifier

from sklearn.metrics import accuracy_score, precision_score, recall_score, f1_score

from sklearn.model_selection import train_test_split

from sklearn.preprocessing import MinMaxScaler

from sklearn.preprocessing import StandardScaler

import tensorflow as tf

from tensorflow.keras.models import Sequential

from tensorflow.keras.layers import Dense


data = pd.read_csv ( ' data.csv ' )


data = data.dropna ( subset= [ "satisfaction" ] )
```

data

	aa	id	Gender	Customer Type	Age	Type of Travel	¥
0	0	70172	Male	Loyal Customer	13	Personal Travel	
1	1	5047	Male	disloyal Customer	25	Business travel	
2	2	110028	Female	Loyal Customer	26	Business travel	
3	3	24026	Female	Loyal Customer	25	Business travel	
4	4	119299	Male	Loyal Customer	61	Business travel	
...	...	...	...	...	...	...	...
129875	25971	78463	Male	disloyal Customer	34	Business travel	
129876	25972	71167	Male	Loyal Customer	23	Business travel	
129877	25973	37675	Female	Loyal Customer	17	Personal Travel	
129878	25974	90086	Male	Loyal Customer	14	Business travel	
129879	25975	34799	Female	Loyal Customer	42	Personal Travel	

	Class	Flight Distance	Inflight wifi service	¥
0	Eco Plus	460	3	
1	Business	235	3	
2	Business	1142	2	
3	Business	562	2	
4	Business	214	3	
...	...	...	...	...
129875	Business	526	3	
129876	Business	646	4	
129877	Eco	828	2	

129878	Business	1127	3
129879	Eco	264	2

	Departure/Arrival time convenient ...	Inflight entertainment ¥
0	4 ...	5
1	2 ...	1
2	2 ...	5
3	5 ...	2
4	3 ...	3
...		...
129875	3 ...	4
129876	4 ...	4
129877	5 ...	2
129878	3 ...	4
129879	5 ...	1

	On-board service	Leg room service	Baggage handling	Checkin service ¥
0	4	3	4	4
1	1	5	3	1
2	4	3	4	4
3	2	5	3	1
4	3	4	4	3
...	...	...	...	...
129875	3	2	4	4
129876	4	5	5	5

129877	4	3	4	5
129878	3	2	5	4
129879	1	2	1	1

	Inflight service	Cleanliness	Departure Delay in Minutes	¥
0	5	5		25
1	4	1		1
2	4	5		0
3	4	2		11
4	3	3		0
...	...	...		...
129875	5	4		0
129876	5	4		0
129877	4	2		0
129878	5	4		0
129879	1	1		0

	Arrival Delay in Minutes	satisfaction
0	18.0	neutral or dissatisfied
1	6.0	neutral or dissatisfied
2	0.0	satisfied
3	9.0	neutral or dissatisfied
4	0.0	satisfied
...	...	...
129875	0.0	neutral or dissatisfied

```

129876          0.0          satisfied
129877          0.0  neutral or dissatisfied
129878          0.0          satisfied
129879          0.0  neutral or dissatisfied

```

```
[ 129880 rows x 25 columns ]
```

```
data.shape
```

```
( 129880, 25 )
```

```
data = data.drop ( columns= [ ' aa ', ' id ' ] )
```

```
data.shape
```

```
( 129880, 23 )
```

```
data.info ( )
```

```
<class ' pandas.core.frame.DataFrame ' >
```

```
Int64Index : 129880 entries, 0 to 129879
```

```
Data columns ( total 23 columns ) :
```

#	Column	Non-Null Count	Dtype
0	Gender	129880 non-null	object

1	Customer Type	129880	non-null	object
2	Age	129880	non-null	int64
3	Type of Travel	129880	non-null	object
4	Class	129880	non-null	object
5	Flight Distance	129880	non-null	int64
6	Inflight wifi service	129880	non-null	int64
7	Departure/Arrival time convenient	129880	non-null	int64
8	Ease of Online booking	129880	non-null	int64
9	Gate location	129880	non-null	int64
10	Food and drink	129880	non-null	int64
11	Online boarding	129880	non-null	int64
12	Seat comfort	129880	non-null	int64
13	Inflight entertainment	129880	non-null	int64
14	On-board service	129880	non-null	int64
15	Leg room service	129880	non-null	int64
16	Baggage handling	129880	non-null	int64
17	Checkin service	129880	non-null	int64
18	Inflight service	129880	non-null	int64
19	Cleanliness	129880	non-null	int64
20	Departure Delay in Minutes	129880	non-null	int64
21	Arrival Delay in Minutes	129487	non-null	float64
22	satisfaction	129880	non-null	object

dtypes : float64 (1) , int64 (17) , object (5)

memory usage : 23.8+ MB

data.describe ()

	Age	Flight Distance	Inflight wifi service	¥
count	129880.000000	129880.000000	129880.000000	
mean	39.427957	1190.316392	2.728696	
std	15.119360	997.452477	1.329340	
min	7.000000	31.000000	0.000000	
25%	27.000000	414.000000	2.000000	
50%	40.000000	844.000000	3.000000	
75%	51.000000	1744.000000	4.000000	
max	85.000000	4983.000000	5.000000	

	Departure/Arrival time convenient	Ease of Online booking	¥
count	129880.000000	129880.000000	
mean	3.057599	2.756876	
std	1.526741	1.401740	
min	0.000000	0.000000	
25%	2.000000	2.000000	
50%	3.000000	3.000000	
75%	4.000000	4.000000	
max	5.000000	5.000000	

	Gate location	Food and drink	Online boarding	Seat comfort	¥
count	129880.000000	129880.000000	129880.000000	129880.000000	
mean	2.976925	3.204774	3.252633	3.441361	

std	1.278520	1.329933	1.350719	1.319289
min	0.000000	0.000000	0.000000	0.000000
25%	2.000000	2.000000	2.000000	2.000000
50%	3.000000	3.000000	3.000000	4.000000
75%	4.000000	4.000000	4.000000	5.000000
max	5.000000	5.000000	5.000000	5.000000

	Inflight entertainment	On-board service	Leg room service	¥
count	129880.000000	129880.000000	129880.000000	
mean	3.358077	3.383023	3.350878	
std	1.334049	1.287099	1.316252	
min	0.000000	0.000000	0.000000	
25%	2.000000	2.000000	2.000000	
50%	4.000000	4.000000	4.000000	
75%	4.000000	4.000000	4.000000	
max	5.000000	5.000000	5.000000	

	Baggage handling	Checkin service	Inflight service	Cleanliness	¥
count	129880.000000	129880.000000	129880.000000	129880.000000	
mean	3.632114	3.306267	3.642193	3.286326	
std	1.180025	1.266185	1.176669	1.313682	
min	1.000000	0.000000	0.000000	0.000000	
25%	3.000000	3.000000	3.000000	2.000000	
50%	4.000000	3.000000	4.000000	3.000000	
75%	5.000000	4.000000	5.000000	4.000000	

max	5.000000	5.000000	5.000000	5.000000
-----	----------	----------	----------	----------

	Departure Delay in Minutes	Arrival Delay in Minutes
count	129880.000000	129487.000000
mean	14.713713	15.091129
std	38.071126	38.465650
min	0.000000	0.000000
25%	0.000000	0.000000
50%	0.000000	0.000000
75%	12.000000	13.000000
max	1592.000000	1584.000000

data.columns

```
Index ( [ ' Gender ', ' Customer Type ', ' Age ', ' Type of Travel ', ' Class ',
,
' Flight Distance ', ' Inflight wifi service ',
' Departure/Arrival time convenient ', ' Ease of Online booking ',
' Gate location ', ' Food and drink ', ' Online boarding ', ' Seat comfort ',
' Inflight entertainment ', ' On-board service ', ' Leg room service ',
' Baggage handling ', ' Checkin service ', ' Inflight service ',
' Cleanliness ', ' Departure Delay in Minutes ', ' Arrival Delay in Minutes ',
' satisfaction ' ],
dtype= ' object ' )
```

```
data.rename ( columns=lambda x : x.replace ( ' ', ' _ ' ), inplace=True )
```

data

	Gender	Customer_Type	Age	Type_of_Travel	Class	¥
0	Male	Loyal Customer	13	Personal Travel	Eco Plus	
1	Male	disloyal Customer	25	Business travel	Business	
2	Female	Loyal Customer	26	Business travel	Business	
3	Female	Loyal Customer	25	Business travel	Business	
4	Male	Loyal Customer	61	Business travel	Business	
...	...	...	...	...	...	...
129875	Male	disloyal Customer	34	Business travel	Business	
129876	Male	Loyal Customer	23	Business travel	Business	
129877	Female	Loyal Customer	17	Personal Travel	Eco	
129878	Male	Loyal Customer	14	Business travel	Business	
129879	Female	Loyal Customer	42	Personal Travel	Eco	

	Flight_Distance	Inflight_wifi_service	¥
0	460	3	
1	235	3	
2	1142	2	
3	562	2	
4	214	3	
...	...	...	...
129875	526	3	
129876	646	4	

129877	828	2
129878	1127	3
129879	264	2

	Departure/Arrival_time_convenient	Ease_of_Online_booking	¥
0	4	3	
1	2	3	
2	2	2	
3	5	5	
4	3	3	
...	...	...	
129875	3	3	
129876	4	4	
129877	5	1	
129878	3	3	
129879	5	2	

	Gate_location	...	Inflight_entertainment	On-board_service	¥
0	1	...	5	4	
1	3	...	1	1	
2	2	...	5	4	
3	5	...	2	2	
4	3	...	3	3	
...	...	...	...	...	
129875	1	...	4	3	

129876	4 ...	4	4
129877	5 ...	2	4
129878	3 ...	4	3
129879	5 ...	1	1

	Leg_room_service	Baggage_handling	Checkin_service	Inflight_service	¥
0	3	4	4	5	
1	5	3	1	4	
2	3	4	4	4	
3	5	3	1	4	
4	4	4	3	3	
...	...	...	...	...	
129875	2	4	4	5	
129876	5	5	5	5	
129877	3	4	5	4	
129878	2	5	4	5	
129879	2	1	1	1	

	Cleanliness	Departure_Delay_in_Minutes	Arrival_Delay_in_Minutes	¥
0	5	25	18.0	
1	1	1	6.0	
2	5	0	0.0	
3	2	11	9.0	
4	3	0	0.0	
...	...	...	...	

129875	4	0	0.0
129876	4	0	0.0
129877	2	0	0.0
129878	4	0	0.0
129879	1	0	0.0

satisfaction

0	neutral or dissatisfied
1	neutral or dissatisfied
2	satisfied
3	neutral or dissatisfied
4	satisfied
...	...
129875	neutral or dissatisfied
129876	satisfied
129877	neutral or dissatisfied
129878	satisfied
129879	neutral or dissatisfied

[129880 rows x 23 columns]

data.isna (). sum ()

Gender	0
Customer_Type	0

Age	0
Type_of_Travel	0
Class	0
Flight_Distance	0
Inflight_wifi_service	0
Departure/Arrival_time_convenient	0
Ease_of_Online_booking	0
Gate_location	0
Food_and_drink	0
Online_boarding	0
Seat_comfort	0
Inflight_entertainment	0
On-board_service	0
Leg_room_service	0
Baggage_handling	0
Checkin_service	0
Inflight_service	0
Cleanliness	0
Departure_Delay_in_Minutes	0
Arrival_Delay_in_Minutes	393
satisfaction	0
dtype :	int64

```
plt.figure ( figsize= ( 12, 8 ) )
```

```
sns.histplot ( data=data, x= ' Flight_Distance ', bins=15, kde=True )
```

```
plt.title ( ' Distribution of Flight Distance ' )
```

```
plt.xlabel ( ' Flight Distance ' )
```

```
plt.ylabel ( ' Frequency ' )
```

```
plt.show ( )
```

```
plt.figure ( figsize= ( 12, 8 ) )
```

```
sns.histplot ( data=data, x= ' Departure_Delay_in_Minutes ', bins=15, kde=True )
```

```
plt.title ( ' Distribution of Departure_Delay_in_Minutes ' )
```

```
plt.xlabel ( ' Departure_Delay_in_Minutes ' )
```

```
plt.ylabel ( ' Frequency ' )
```

```
plt.show ( )
```

```
plt.figure ( figsize= ( 12, 8 ) )
```

```
sns.histplot ( data=data, x= ' Arrival_Delay_in_Minutes ', bins=15, kde=True )
```

```
plt.title ( ' Distribution of Arrival_Delay_in_Minutes ' )
```

```
plt.xlabel ( ' Arrival_Delay_in_Minutes ' )
```

```
plt.ylabel ( ' Frequency ' )
```

```
plt.show ( )
```

```
plt.figure ( figsize= ( 12, 8 ) )
```

```
sns.histplot ( data=data, x= ' Age ' , bins=15, kde=True )
```

```
plt.title ( ' Distribution of Age ' )
```

```
plt.xlabel ( ' Age ' )
```

```
plt.ylabel ( ' Frequency ' )
```

```
plt.show ( )
```

```
plt.boxplot ( data [ ' Flight_Distance ' ] )
```

```
plt.title ( ' Box plot of Flight Distance ' )
```

```
plt.show ( )
```

```
plt.boxplot ( data [ ' Departure_Delay_in_Minutes ' ] )
```

```
plt.title ( ' Box plot of Departure Delay in Minutes ' )
```

```
plt.show ( )
```

```
plt.boxplot ( data [ ' Age ' ] )
```

```
plt.title ( ' Box plot of Age ' )
```

```
plt.show ( )
```

```
categorical_columns = [ ' satisfaction ', ' Gender ', ' Customer_Type ',  
                        ' Type_of_Travel ',  
                        ' Class ', ' Seat_comfort ',  
                        ' Departure/Arrival_time_convenient ', ' Food_and_drink ', ' Gate_location ',  
                        ' Inflight_wifi_service ', ' Inflight_entertainment ',  
                        ' Ease_of_Online_booking ', ' On-board_service ', ' Leg_room_service ',  
                        ' Baggage_handling ', ' Checkin_service ', ' Cleanliness ', ' Online_boarding  
                        ' ]
```

```
for column in categorical_columns :
```

```
sns.countplot ( data=data, x=column )
```

```
plt.title ( f ' Distribution of records of {column} ' )
```

```
plt.xlabel ( [ column ] )
```

```
plt.ylabel ( ' Count of records )
```

```
plt.show ( )
```

```
sns.countplot ( x= ' Gender ' , hue= ' satisfaction ' , data=data )
```

```
plt.show ( )
```

```
sns.countplot ( x= ' Customer_Type ' , hue= ' satisfaction ' , data=data )
```

```
plt.show ( )
```

```
sns.countplot ( x= ' Type_of_Travel ' , hue= ' satisfaction ' , data=data )
```

```
plt.show ( )
```

```
sns.countplot ( x= ' Class ' , hue= ' satisfaction ' , data=data )
```

```
plt.show ( )
```

```
sns.countplot ( x= ' Seat_comfort ', hue= ' satisfaction ', data=data )
```

```
plt.show ( )
```

```
sns.countplot ( x= ' Gate_location ', hue= ' satisfaction ', data=data )
```

```
plt.show ( )
```

```
sns.countplot ( x= ' On-board_service ', hue= ' satisfaction ', data=data )
```

```
plt.show ( )
```

```
plt.figure ( figsize= ( 12, 8 ) )
```

```
sns.boxplot ( data=data, x= ' satisfaction ', y= ' Departure_Delay_in_Minutes ' )
```

```
plt.title ( ' Departure Delay vs Customer Satisfaction ' )
```

```
plt.xlabel ( ' Satisfaction ' )
```

```
plt.ylabel ( ' Departure Delay ( Minutes ) ' )
```

```
plt.show ( )
```

```

plt.figure ( figsize= ( 12, 8 ) )

sns.boxplot ( data=data, x= ' satisfaction ', y= ' Arrival_Delay_in_Minutes ' )

plt.title ( ' Arrival Delay vs Customer Satisfaction ' )

plt.xlabel ( ' Satisfaction ' )

plt.ylabel ( ' Arrival Delay ( Minutes ) ' )

plt.show ( )

sns.heatmap ( data.corr ( ) )

#replace missing values in arrival delay in minutes

missing_indices = data [ data [ ' Arrival_Delay_in_Minutes ' ] .isnull ( ) ] .index

data.loc [ missing_indices, ' Arrival_Delay_in_Minutes ' ] = data [ '
Arrival_Delay_in_Minutes ' ] .mean ( )

data.isna ( ) .sum ( )

```

```
Gender                                0
```

```
Customer_Type                        0
```

```
Age                                  0
```

Type_of_Travel	0
Class	0
Flight_Distance	0
Inflight_wifi_service	0
Departure/Arrival_time_convenient	0
Ease_of_Online_booking	0
Gate_location	0
Food_and_drink	0
Online_boarding	0
Seat_comfort	0
Inflight_entertainment	0
On-board_service	0
Leg_room_service	0
Baggage_handling	0
Checkin_service	0
Inflight_service	0
Cleanliness	0
Departure_Delay_in_Minutes	0
Arrival_Delay_in_Minutes	0
satisfaction	0
dtype : int64	

data.dtypes

Gender	object
--------	--------

Customer_Type	object
Age	int64
Type_of_Travel	object
Class	object
Flight_Distance	int64
Inflight_wifi_service	int64
Departure/Arrival_time_convenient	int64
Ease_of_Online_booking	int64
Gate_location	int64
Food_and_drink	int64
Online_boarding	int64
Seat_comfort	int64
Inflight_entertainment	int64
On-board_service	int64
Leg_room_service	int64
Baggage_handling	int64
Checkin_service	int64
Inflight_service	int64
Cleanliness	int64
Departure_Delay_in_Minutes	int64
Arrival_Delay_in_Minutes	float64
satisfaction	object
dtype : object	

#convert object types to integers

```
data [ ' Customer_Type ' ] = data [ ' Customer_Type ' ] .map ( { ' Loyal Customer ' :  
0, ' disloyal Customer ' : 1} )
```

```
data [ ' satisfaction ' ] = data [ ' satisfaction ' ] .map ( { ' neutral or dissatisfied  
' : 0, ' satisfied ' : 1} )
```

```
data [ ' Gender ' ] = data [ ' Gender ' ] .map ( { ' Male ' : 1, ' Female ' : 2} )
```

```
data [ ' Type_of_Travel ' ] = data [ ' Type_of_Travel ' ] .map ( { ' Personal Travel '  
: 1, ' Business travel ' : 2} )
```

```
data [ ' Class ' ] = data [ ' Class ' ] .map ( { ' Eco ' : 1, ' Business ' : 2 ,  
' Eco Plus ' : 3} )
```

```
data.dtypes
```

Gender	int64
Customer_Type	int64
Age	int64
Type_of_Travel	int64
Class	int64
Flight_Distance	int64
Inflight_wifi_service	int64
Departure/Arrival_time_convenient	int64
Ease_of_Online_booking	int64
Gate_location	int64

```

Food_and_drink                int64
Online_boarding                int64
Seat_comfort                  int64
Inflight_entertainment         int64
On-board_service               int64
Leg_room_service               int64
Baggage_handling               int64
Checkin_service                int64
Inflight_service               int64
Cleanliness                    int64
Departure_Delay_in_Minutes     int64
Arrival_Delay_in_Minutes       float64
satisfaction                    int64
dtype : object

```

```

y = data [ 'satisfaction' ]

```

```

x = data.drop ( 'satisfaction' , axis=1 )

```

```

#normalize data

```

```

columns_to_normalize = [ 'Flight_Distance' , 'Arrival_Delay_in_Minutes' , '
Departure_Delay_in_Minutes' ]

```

```

scaler = MinMaxScaler ( )

```

```

x [ columns_to_normalize ] = scaler.fit_transform ( x [ columns_to_normalize ] )

```

```
X_train, X_test, y_train, y_test = train_test_split ( x, y, test_size=0.2, random_state=42 )
```

```
models = {}
```

```
models [ ' Logistic Regression ' ] = LogisticRegression ( max_iter=1000 )
```

```
models [ ' Decision Trees ' ] = DecisionTreeClassifier ( )
```

```
models [ ' Random Forest ' ] = RandomForestClassifier ( )
```

```
models [ ' Naive Bayes ' ] = GaussianNB ( )
```

```
models [ ' K-Nearest Neighbor ' ] = KNeighborsClassifier ( )
```

```
accuracy, precision, recall = {}, {}, {}
```

```
for key in models.keys ( ) :
```

```
    # Fit the classifier
```

```
    models [ key ] .fit ( X_train, y_train )
```

```
    # Make predictions
```

```
    predictions = models [ key ] .predict ( X_test )
```

```
    # Calculate metrics
```

```
    accuracy [ key ] = accuracy_score ( predictions, y_test )
```

```
    precision [ key ] = precision_score ( predictions, y_test )
```

```
    recall [ key ] = recall_score ( predictions, y_test )
```

```
results = pd.DataFrame ( index=models.keys ( ) , columns= [ ' Accuracy ' , ' Precision ' ,
' Recall ' ] )
```

```
results [ ' Accuracy ' ] = accuracy.values ( )
```

```
results [ ' Precision ' ] = precision.values ( )
```

```
results [ ' Recall ' ] = recall.values ( )
```

```
results
```

	Accuracy	Precision	Recall
Logistic Regression	0.873653	0.834332	0.871161
Decision Trees	0.943756	0.934120	0.937009
Random Forest	0.961233	0.936586	0.973720
Naive Bayes	0.866608	0.821032	0.866760
K-Nearest Neighbor	0.912920	0.867624	0.928464

ΒΙΒΛΙΟΓΡΑΦΙΑ – ΠΗΓΕΣ

- [1]. Choi, R. Y., Coyner, A. S., Kalpathy - Cramer, J., Chiang, M. F., & Campbell, J. P. (2020) . Introduction to machine learning, neural networks, and deep learning. Translational vision science & technology, 9 (2) , 14 - 14.
- [2]. El Bouchefry, K., & de Souza, R. S. (2020) . Learning in big data : Introduction to machine learning. In Knowledge discovery in big data from astronomy and earth observation (pp. 225 - 249) . Elsevier.
- [3]. Kubat, M. (2017) . An introduction to machine learning. Springer.
- [4]. Badillo, S., Banfai, B., Birzele, F., Davydov, I. I., Hutchinson, L., Kam-Thong, T., ... & Zhang, J. D. (2020) . An introduction to machine learning. Clinical pharmacology & therapeutics, 107 (4) , 871 - 885.
- [5]. Mahesh, B. (2020) . Machine learning algorithms - a review. International Journal of Science and Research (IJSR) .[Internet], 9 (1) , 381 - 386.
- [6]. El Naqa, I., & Murphy, M. J. (2015) . What is machine learning? (pp. 3 - 11) . Springer International Publishing.
- [7]. Ge, Z., Song, Z., Ding, S. X., & Huang, B. (2017) . Data mining and analytics in the process industry : The role of machine learning. Ieee Access, 5, 20590 - 20616.

- [8]. Mitchell, T. M. (2006) . The discipline of machine learning (Vol. 9, p. 3) .
Pittsburgh : Carnegie Mellon University, School of Computer Science, Machine
Learning Department.
- [9]. Sarker, M. (2024) . Revolutionizing healthcare : the role of machine learning in
the health sector. Journal of Artificial Intelligence General science (JAIGS) ISSN :
3006 - 4023, 2 (1) , 36 - 61.
- [10]. Ranft, B., & Stiller, C. (2016) . The role of machine vision for intelligent vehicles.
IEEE Transactions on Intelligent vehicles, 1 (1) , 8 - 19.
- [11]. Nasteski, V. (2017) . An overview of the supervised machine learning methods.
Horizons. b, 4 (51 - 62) , 56.
- [12]. Jiang, T., Gradus, J. L., & Rosellini, A. J. (2020) . Supervised machine learning : a
brief primer. Behavior therapy, 51 (5) , 675 - 687.
- [13]. Singh, A., Thakur, N., & Sharma, A. (2016, March) . A review of supervised
machine learning algorithms. In 2016 3rd international conference on computing for
sustainable global development (INDIACom) (pp. 1310 - 1315) . Ieee.
- [14]. Osisanwo, F. Y., Akinsola, J. E. T., Awodele, O., Hinmikaiye, J. O., Olakanmi, O., &
Akinjobi, J. (2017) . Supervised machine learning algorithms : classification and
comparison. International Journal of Computer Trends and Technology (IJCTT) , 48
(3) , 128 - 138.

- [15]. Usama, M., Qadir, J., Raza, A., Arif, H., Yau, K. L. A., Elkhatib, Y., ... & Al - Fuqaha, A. (2019) . Unsupervised machine learning for networking : Techniques, applications and research challenges. IEEE access, 7, 65579 - 65615.
- [16]. Hahne, F., Huber, W., Gentleman, R., Falcon, S., Gentleman, R., & Carey, V. J. (2008) . Unsupervised machine learning. Bioconductor case studies, 137 - 157.
- [17]. Naeem, S., Ali, A., Anam, S., & Ahmed, M. M. (2023) . An unsupervised machine learning algorithms : Comprehensive review. International Journal of Computing and Digital Systems.
- [18]. Hu, W., Singh, R. R., & Scalettar, R. T. (2017) . Discovering phases, phase transitions, and crossovers through unsupervised machine learning : A critical examination. Physical Review E, 95 (6) , 062122.
- [19]. Zhu, X., & Goldberg, A. B. (2022) . Introduction to semi - supervised learning. Springer Nature.
- [20]. Van Engelen, J. E., & Hoos, H. H. (2020) . A survey on semi - supervised learning. Machine learning, 109 (2) , 373 - 440.
- [21]. Reddy, Y. C. A. P., Viswanath, P., & Reddy, B. E. (2018) . Semi - supervised learning : A brief review. Int. J. Eng. Technol, 7 (1.8) , 81.
- [22]. Hady, M. F. A., & Schwenker, F. (2013) . Semi - supervised learning. Handbook on Neural Information Processing, 215 - 239.

- [23]. François - Lavet, V., Henderson, P., Islam, R., Bellemare, M. G., & Pineau, J. (2018) . An introduction to deep reinforcement learning. Foundations and Trends® in Machine Learning, 11 (3 - 4) , 219 - 354.
- [24]. Kulkarni, P. (2012) . Reinforcement and systemic machine learning for decision making (Vol. 1) . John Wiley & Sons.
- [25]. Li, Y. (2017) . Deep reinforcement learning : An overview. arXiv preprint arXiv : 1701.07274.
- [26]. Shinde, P. P., & Shah, S. (2018, August) . A review of machine learning and deep learning applications. In 2018 Fourth international conference on computing communication control and automation (ICCUBEA) (pp. 1 - 6) . IEEE.
- [27]. Carbonell, J. G., Michalski, R. S., & Mitchell, T. M. (1983) . An overview of machine learning. Machine learning, 3 - 23.
- [28]. L'heureux, A., Grolinger, K., Elyamany, H. F., & Capretz, M. A. (2017) . Machine learning with big data : Challenges and approaches. Ieee Access, 5, 7776 - 7797.
- [29]. Paleyes, A., Urma, R. G., & Lawrence, N. D. (2022) . Challenges in deploying machine learning : a survey of case studies. ACM computing surveys, 55 (6) , 1 - 29.
- [30]. Zhou, L., Pan, S., Wang, J., & Vasilakos, A. V. (2017) . Machine learning on big data : Opportunities and challenges. Neurocomputing, 237, 350 - 361.

- [31]. Jiang, Y., Tran, T. H., & Williams, L. (2023). Machine learning and mixed reality for smart aviation: Applications and challenges. *Journal of Air Transport Management*, 111, 102437.
- [32]. Abeyratne, R. (2020). *Aviation in the Digital age*. Springer International Publishing.
- [33]. Molchanova, K. (2020). A review of digital technologies in aviation industry. *Logistics and Transport*, 47 (3-4), 69-77.
- [34]. Kuisma, N. (2018). Digitalization and its impact on commercial aviation.
- [35]. Duraisamy, K., Iaccarino, G., & Xiao, H. (2019). Turbulence modeling in the age of data. *Annual review of fluid mechanics*, 51, 357-377.
- [36]. Pereira, B. A., Lohmann, G., & Houghton, L. (2022). Technology trajectory in aviation: Innovations leading to value creation (2000–2019). *International Journal of Innovation Studies*, 6 (3), 128-141.
- [37]. Chinchani, S., & Shaikh, A. A. (2022). A review on machine learning, big data analytics, and design for additive manufacturing for aerospace applications. *Journal of Materials Engineering and Performance*, 31 (8), 6112-6130.
- [38]. Brunton, S. L., & Kutz, J. N. (2022). *Data-driven science and engineering: Machine learning, dynamical systems, and control*. Cambridge University Press.
- [39]. Avram, B. (2017). Ancillaries in the aviation industry: importance, trends, going digital. *Expert Journal of Marketing*, 5 (2), 53-65.

- [40]. Bowcutt, K., Kuruvila, G., Grandine, T., & Cramer, E. (2008, April). Advancements in multidisciplinary design optimization applied to hypersonic vehicles to achieve performance closure. In 15th AIAA international space planes and hypersonic systems and technologies conference (p. 2591).
- [41]. Bons, N., & Martins, J. R. (2020). Aerostructural wing design exploration with multidisciplinary design optimization. In AIAA Scitech 2020 Forum (p. 0544).
- [42]. Baran, I., Cinar, K., Ersoy, N., Akkerman, R., & Hattel, J. H. (2017). A review on the mechanical modeling of composite manufacturing processes. Archives of computational methods in engineering, 24, 365-395.
- [43]. Dong, Y., Tao, J., Zhang, Y., Lin, W., & Ai, J. (2021). Deep learning in aircraft design, dynamics, and control: Review and prospects. IEEE Transactions on Aerospace and Electronic Systems, 57 (4), 2346-2368.
- [44]. Le Clainche, S., Ferrer, E., Gibson, S., Cross, E., Parente, A., & Vinuesa, R. (2023). Improving aircraft performance using machine learning: A review. Aerospace Science and Technology, 108354.
- [45]. Sabater, C., Stürmer, P., & Bekemeyer, P. (2022). Fast predictions of aircraft aerodynamics using deep-learning techniques. AIAA Journal, 60 (9), 5249-5261.
- [46]. Eroglu, B., Sahin, M. C., & Ure, N. K. (2020). Autolanding control system design with deep learning based fault estimation. Aerospace Science and Technology, 102, 105855.

- [47]. Ravaioli, U. J., Cunningham, J., McCarroll, J., Gangal, V., Dunlap, K., & Hobbs, K. L. (2022, March). Safe reinforcement learning benchmark environments for aerospace control systems. In 2022 IEEE Aerospace Conference (AERO) (pp. 1-20). IEEE.
- [48]. Airbus. Design a Starting Point for the Innovations of Tomorrow. Available online: <https://www.airbus.com/en/products-services/commercial-aircraft/the-life-cycle-of-an-aircraft/design>
- [49]. Let's Connect: Digital thread advances manufacturing. Available online: <https://www.boeing.com/features/innovation-quarterly/2022/10/digital-thread-advances-manufacturing>
- [50]. Caesarendra, W., Wijaya, T., Pappachan, B. K., & Tjahjowidodo, T. (2019, July). Adaptation to industry 4.0 using machine learning and cloud computing to improve the conventional method of deburring in aerospace manufacturing industry. In 2019 12th International Conference on Information & Communication Technology and System (ICTS) (pp. 120-124). IEEE.
- [51]. Shafi, I., Mazhar, M. F., Fatima, A., Alvarez, R. M., Miró, Y., Espinosa, J. C. M., & Ashraf, I. (2023). Deep learning-based real time defect detection for optimization of aircraft manufacturing and control performance. Drones, 7 (1), 31.

- [52]. Malik, S., Rouf, R., Mazur, K., & Kontsos, A. (2020). The industry Internet of Things (IIoT) as a methodology for autonomous diagnostics in aerospace structural health monitoring. *Aerospace*, 7 (5), 64.
- [53]. Lu, L., Hou, J., Yuan, S., Yao, X., Li, Y., & Zhu, J. (2023). Deep learning-assisted real-time defect detection and closed-loop adjustment for additive manufacturing of continuous fiber-reinforced polymer composites. *Robotics and Computer-Integrated Manufacturing*, 79, 102431.
- [54]. Manohar, K., Hogan, T., Buttrick, J., Banerjee, A. G., Kutz, J. N., & Brunton, S. L. (2018). Predicting shim gaps in aircraft assembly with machine learning and sparse sensing. *Journal of manufacturing systems*, 48, 87-95.
- [55]. Verpoort, P. C., MacDonald, P., & Conduit, G. J. (2018). Materials data validation and imputation with an artificial neural network. *Computational Materials Science*, 147, 176-185.
- [56]. Conduit, B. D., Jones, N. G., Stone, H. J., & Conduit, G. J. (2017). Design of a nickel-base superalloy using a neural network. *Materials & Design*, 131, 358-365.
- [57]. Green, A. G., Conduit, G., & Krüger, F. (2018). Quantum order-by-disorder in strongly correlated metals. *Annual Review of Condensed Matter Physics*, 9, 59-77.
- [58]. Sacco, C., Radwan, A. B., Beatty, T., & Harik, R. (2019). Machine learning based AFP inspection: A tool for characterization and integration. *SAMPE 2019-Charlotte*, NC, May 2019.

- [59]. GE aviation. Available online: <https://www.ge.com/research/clients/ge-businesses/ge-aviation>
- [60]. Lockheed Martin aircraft. Available online: <https://www.lockheedmartin.com/en-us/capabilities/aircraft.html>
- [61]. Boeing. Boeing Testing Services. Available online: <https://www.boeing.com/company/key-orgs/boeing-testing-services/index.page>
- [62]. Huang, H., & Zhu, J. (2021). A short review of the application of machine learning methods in smart airports. In Journal of Physics: Conference Series (Vol. 1769, No. 1, p. 012010). IOP Publishing.
- [63]. See, K. F., Ülkü, T., Forsyth, P., & Niemeier, H. M. (2023). Twenty years of airport efficiency and productivity studies: A machine learning bibliometric analysis. Research in Transportation Business & Management, 46, 100771.
- [64]. Mamdouh, M., Ezzat, M., & Hefny, H. A. (2020). Airport resource allocation using machine learning techniques. Inteligencia Artificial, 23 (65), 19-32.
- [65]. Muelaner, J. E., & Maropoulos, P. (2011, September). Integrated dimensional variation management in the digital factory. In 7th International Conference on Digital Enterprise Technology.
- [66]. Huang, H., & Zhu, J. (2021). A short review of the application of machine learning methods in smart airports. In Journal of Physics: Conference Series (Vol. 1769, No. 1, p. 012010). IOP Publishing.

- [67]. Jain, P. K., & Pamula, R. (2021). Content-based airline recommendation prediction using machine learning techniques. *Machine Learning Algorithms for Industrial Applications*, 185-194.
- [68]. Noviantoro, T., & Huang, J. P. (2022). Investigating airline passenger satisfaction: Data mining method. *Research in Transportation Business & Management*, 43, 100726.
- [69]. Hasib, K. M., Habib, M. A., Towhid, N. A., & Showrov, M. I. H. (2021, February). A novel deep learning based sentiment analysis of twitter data for us airline service. In *2021 International Conference on Information and Communication Technology for Sustainable Development (ICICT4SD)* (pp. 450-455). IEEE.
- [70]. Antwi, C. O., Ren, J., Owusu-Ansah, W., Mensah, H. K., & Aboagye, M. O. (2021). Airport self-service technologies, passenger self-concept, and behavior: An attributional view. *Sustainability*, 13(6), 3134.
- [71]. Hong Kong International Airport. Self-service check-in. <https://www.hkairport-technovation.com/en/self-service-check-in.html>
- [72]. Thamaraiselvan, N., Arul, S. T., & Kasilingam, D. (2019). Understanding the intention to use self service technologies in the airline industry. *International Journal of Services, Economics and Management*, 10(2), 89-109.

- [73]. Cathay Pacific. Self-service Facilities and Network. Available online: https://www.cathaypacific.com/cx/en_HK/check-in/airport-self-service-network.html
- [74]. Changi Airport Group. Fast and Seamless Travel (FAST). Available online: <https://www.changiairport.com/en/airport-guide/departing/checking-in/fast-check-in.html> (accessed on 9 January 2023)
- [75]. Checking in at Heathrow. <https://www.heathrow.com/departures/checking-in>
- [76]. Rodríguez-Sanz, Á., de Marcos, A. F., Pérez-Castán, J. A., Comendador, F. G., Valdés, R. A., & Loreiro, Á. P. (2021). Queue behavioural patterns for passengers at airport terminals: A machine learning approach. *Journal of Air Transport Management*, 90, 101940.
- [77]. Zhang, Z. (2022). Smart solutions to airport security in post-COVID-19 era. *Aca. J. Bus. Manag*, 4, 100-106.
- [78]. Gota, D. I., Puscasiu, A., Fanca, A., Valean, H., & Miclea, L. (2020, October). Threat objects detection in airport using machine learning. In 2020 21th International Carpathian Control Conference (ICCC) (pp. 1-6). IEEE.
- [79]. Liang, K. J., Heilmann, G., Gregory, C., Diallo, S. O., Carlson, D., Spell, G. P., ... & Carin, L. (2018, April). Automatic threat recognition of prohibited items at aviation checkpoint with x-ray imaging: a deep learning approach. In *Anomaly Detection and Imaging with X-Rays (ADIX) III* (Vol. 10632, p. 1063203). SPIE.

- [80]. Gui, G., Liu, F., Sun, J., Yang, J., Zhou, Z., & Zhao, D. (2019). Flight delay prediction based on aviation big data and machine learning. *IEEE Transactions on Vehicular Technology*, 69(1), 140-150.
- [81]. Abdelaziz, S. G., Hegazy, A. A., & Elabbassy, A. (2010). Study of airport self-service technology within experimental research of check-in techniques case study and concept. *International Journal of Computer Science Issues (IJCSI)*, 7(3), 30.
- [82]. Altexsoft. How Airline Industry Streamlines Check-In and Boarding with Digital Self-Services. Available online: <https://www.altexsoft.com/blog/travel/how-airline-industry-streamlines-check-in-and-boarding-with-digital-self-services/>(accessed on 10 January 2023).
- [83]. Sørensen, R. A., Nielsen, M., & Karstoft, H. (2020). Routing in congested baggage handling systems using deep reinforcement learning. *Integrated Computer-Aided Engineering*, 27(2), 139-152.
- [84]. Intelligent tracking: a baggage management revolution. Available online: <https://www.sita.aero/resources/White-papers/intelligent-tracking-a-baggage-management-revolution>
- [85]. Lambelho, M., Mitici, M., Pickup, S., & Marsden, A. (2020). Assessing strategic flight schedules at an airport using machine learning-based flight delay and cancellation predictions. *Journal of air transport management*, 82, 101737.

- [86]. Sridhar, B., Chatterji, G. B., & Evans, A. D. (2020). Lessons Learned in the Application of Machine Learning Techniques to Air Traffic Management. In AIAA AVIATION 2020 FORUM (p. 2882).
- [87]. Orsini, F., Gastaldi, M., Mantecchini, L., & Rossi, R. (2019, June). Neural networks trained with WiFi traces to predict airport passenger behavior. In 2019 6th international conference on models and Technologies for Intelligent Transportation Systems (MT-ITS) (pp. 1-7). IEEE.
- [88]. Airport Benchmarking. Intelligent Robots at Incheon Airport. Available online: <https://www.airportbenchmarking.com/intelligent-robots-at-incheon-airport/>(accessed on 16 January 2023).
- [89]. ANA Group. ANA and Panasonic Partner to Test Self-Driving Electric Wheelchairs at Narita Airport. Available online: <https://www.anahd.co.jp/group/en/pr/201905/20190516.html> (accessed on 16 January 2023).
- [90]. Etihad Airways and Abu Dhabi airports complete guest trials of autonomous wheelchairs. Available online: <https://www.etihad.com/en-us/news/etih-airways-and-abu-dhabi-airports-complete-guest-trials-of-autonomous-wheelchairs>
- [91]. Steiner, M., Eggert, A., Ulaga, W., & Backhaus, K. (2016). Do customized service packages impede value capture in industrial markets?. *Journal of the Academy of marketing Science*, 44, 151-165.

- [92]. Raschka, S., Patterson, J., & Nolet, C. (2020). Machine learning in python: Main developments and technology trends in data science, machine learning, and artificial intelligence. *Information*, 11(4), 193.
- [93]. Pedregosa, F., Varoquaux, G., Gramfort, A., Michel, V., Thirion, B., Grisel, O., ... & Duchesnay, É. (2011). Scikit-learn: Machine learning in Python. *the Journal of machine Learning research*, 12, 2825-2830.
- [94]. Sarkar, D., Bali, R., & Sharma, T. (2018). Practical machine learning with Python. Book" Practical Machine Learning with Python, 25-30.
- [95]. Subasi, A. (2020). Practical machine learning for data analysis using python. Academic Press.
- [96]. Randles, B. M., Pasquetto, I. V., Golshan, M. S., & Borgman, C. L. (2017, June). Using the Jupyter notebook as a tool for open science: An empirical study. In 2017 ACM/IEEE Joint Conference on Digital Libraries (JCDL) (pp. 1-2). IEEE.
- [97]. Grotov, K., Titov, S., Sotnikov, V., Golubev, Y., & Bryksin, T. (2022, May). A large-scale comparison of Python code in Jupyter notebooks and scripts. In *Proceedings of the 19th International Conference on Mining Software Repositories* (pp. 353-364).
- [98]. Wang, J., Li, L., & Zeller, A. (2020, June). Better code, better sharing: on the need of analyzing jupyter notebooks. In *Proceedings of the ACM/IEEE 42nd international conference on software engineering: new ideas and emerging results* (pp. 53-56).

- [99]. Hosmer Jr, D. W., Lemeshow, S., & Sturdivant, R. X. (2013). Applied logistic regression. John Wiley & Sons.
- [100]. Sperandei, S. (2014). Understanding logistic regression analysis. *Biochemia medica*, 24(1), 12-18.
- [101]. Stoltzfus, J. C. (2011). Logistic regression: a brief primer. *Academic emergency medicine*, 18(10), 1099-1104.
- [102]. Kotsiantis, S. B. (2013). Decision trees: a recent overview. *Artificial Intelligence Review*, 39, 261-283.
- [103]. Song, Y. Y., & Ying, L. U. (2015). Decision tree methods: applications for classification and prediction. *Shanghai archives of psychiatry*, 27(2), 130.
- [104]. Potharst, R., & Bioch, J. C. (2000). Decision trees for ordinal classification. *Intelligent Data Analysis*, 4(2), 97-111.
- [105]. Parmar, A., Katariya, R., & Patel, V. (2019). A review on random forest: An ensemble classifier. In *International conference on intelligent data communication technologies and internet of things (ICICI) 2018* (pp. 758-763). Springer International Publishing.
- [106]. Kulkarni, V. Y., & Sinha, P. K. (2013). Random forest classifiers: a survey and future research directions. *Int. J. Adv. Comput*, 36(1), 1144-1153.
- [107]. Shaik, A. B., & Srinivasan, S. (2019). A brief survey on random forest ensembles in classification model. In *International Conference on Innovative Computing and*

Communications: Proceedings of ICICC 2018, Volume 2 (pp. 253-260). Springer Singapore.

- [108]. Webb, G. I., Keogh, E., & Miikkulainen, R. (2010). Naïve Bayes. Encyclopedia of machine learning, 15(1), 713-714.
- [109]. Murphy, K. P. (2006). Naive bayes classifiers. University of British Columbia, 18(60), 1-8.
- [110]. Berrar, D. (2019). Bayes' Theorem and Naive Bayes Classifier.
- [111]. Zhang, Z. (2016). Introduction to machine learning: k-nearest neighbors. Annals of translational medicine, 4(11).
- [112]. Juba, B., & Le, H. S. (2019, July). Precision-recall versus accuracy and the role of large data sets. In Proceedings of the AAAI conference on artificial intelligence (Vol. 33, No. 01, pp. 4039-4048).
- [113]. Powers, D. M. (2020). Evaluation: from precision, recall and F-measure to ROC, informedness, markedness and correlation. arXiv preprint arXiv:2010.16061.
- [114]. Hossin, M., & Sulaiman, M. N. (2015). A review on evaluation metrics for data classification evaluations. International journal of data mining & knowledge management process, 5(2), 1.
- [115]. <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>

