Θεωρια για το ΑΙ και για classification models etc.

**Δομή εκπαίδευσης**:

Για την εκπαίδευση του ΑΙ μοντέλου θα γίνει διαχωρισμός των δεδομένων σε 2 μέρη:

- Σετ εκπαίδευσης (80%): Το μοντέλο θα μαθαίνει μοτίβα από αυτά τα δεδομένα.

- Σετ δοκιμών (20%): Το μοντέλο θα κάνει προβλέψεις σε αυτά τα δεδομένα για να αξιολογήσει την ακρίβεια.

**Python Libraries**:

Για χειρισμό και προεπεξεργασία δεδομένων:

* **Pandas**: Φόρτωση, καθαρισμός και χειρισμός δεδομένων.
* **Numpy**: Χειρισμό αριθμητικών πράξεων.

Για Οπτικοποίηση Δεδομένων:

* **Matplotlib**: Για δημιουργία γραφικών παραστάσεων (π.χ. διαγράμματα διασποράς, ιστογράμματα).
* **Seaborn**: Για οπτικοποίηση.

Για μοντέλα μηχανικής μάθησης και AI:

* **scikit-learn**: Εκπαίδευση στο μοντέλο μηχανικής μάθησης (Random Forest, Logistic Regression, κ.λπ.).
* **XGBoost**: Μοντέλο ενίσχυσης κλίσης υψηλής απόδοσης.
* **Tensorflow**/**Keras** (Ίσως): Για δοκιμή στη Deep Learning.

Για την αξιολόγηση του μοντέλου:

* **sklearn.metrics**: Ελέγξτε την ακρίβεια, την ακρίβεια, την ανάκληση, τη μήτρα σύγχυσης.

Basic Features:

**1. Features Αποτελεσμάτων Οδηγού**

* **Πόντοι καριέρας ανά αγώνα**= Συνολικοί πόντοι / Συνολικοί αγώνες
* **Win Rate** = Συνολικές νίκες / Συνολικοί αγώνες
* **Εμφανίσεις βάθρου ανά αγώνα** = Συνολικά βάθρα / Συνολικοί αγώνες
* **Top-10 Finish Rate** = Αριθμός αποτελεσμάτων στις 10 πρώτες θέσεις / Συνολικοί αγώνες
* **Μέση τελική θέση** = Μέση τελική θέση για όλες τις σεζόν
* **Συνολικοί αγώνες** = Συνολικοί αγώνες που συμμετείχε ο οδηγός
* **Το καλύτερο αποτέλεσμα για κάθε σεζόν** = Το καλύτερο αποτέλεσμα για μια σεζόν

**2. Features Αποτελεσμάτων Ομάδας**

* **Δυναμικότητα ομάδας** = (Συνολικές νίκες ανα σεζόν) / (Συνολικοί αγώνες ανα σεζόν)
* **Team Consistency** = Τυπική απόκλιση των πόντων της ομάδας μεταξύ των εποχών (οι χαμηλότερες τιμές υποδηλώνουν συνέπεια)
* **Πόντοι ανά αγώνα (Ομάδα)** = Συνολικοί πόντοι ομάδας / Αγώνες της σεζόν
* **Συνολικά αποτελέσματα 1η και 2ης θέσης** = Πόσες φορές ένα από τους οδηγούς της ομάδας τελείωσε στην 1η και 2η θέση

**3. Features Ανα Πίστα**

* **Ποσοστό νικών στην πίστα** = Νίκες στην πίστα / Συνολικοί αγώνες στην πίστα
* **Επίδοση οδηγού στην πίστα** = Επίδοση του οδηγού στην πίστα vs. Την μέση επίδοση του
* **Επιδόσεις για καιρικές συνθήκες** Επιδόσεις για αγώνες με βροχή vs. Αγώνες χωρίς βροχή
* **Qualifying vs. Race Performance** = Διαφορά μεταξύ της μέσης θέσης εκκίνησης και της μέσης θέσης τερματισμού σε μια πίστα

**4. Features Αποτελεσμάτων στους προκριματικούς αγώνες**

* **Μέση προκριματική θέση** = Συνολικό αποτέλεσμα των προκριματικών / των συνολικών προκριματικών αγώνων
* **Pole Position Rate** = Αριθμός πρώτων θέσεων στα προκριματικά / Συνολικοί αγώνες
* **Συνέπεια στα αποτελέσματα προκριματικών αγώνων** = Τυπική απόκλιση θέσεων εκκίνησης

**5. Features Μακροπρόθεσμων Τάσεων & Σεζόν**

* **Performance Delta** = Διαφορά στους πόντους/κατάταξης/τερματικής θέσης σε σύγκριση με την προηγούμενη σεζόν
* **Δείκτης Momentum** = Κινούμενος μέσος όρος των τελευταίων 5 θέσεων τερματισμού
* **Win Streaks** = Αριθμός συνεχόμενων νικών σε μια σεζόν
* **Συντελεστής ηλικίας οδηγού** = Κανονικοποίηση επίδοσης οδηγού βασισμένα στις ηλικιακές κατηγορίες (rookie, in his prime, veteran)
* **Championship Lead Stability** = Πόσες φορές ο οδηγός ήταν πρώτος σε πόντους στο πρωτάθλημα την σεζόν

**6. Features Για Προβλέψεις Μελλοντικών Αποτελέσματων**

* **Σκορ συνεργείας μεταξύ οδηγού και ομάδος** = Αλλαγές στην επίδοση κατά την αλλαγή ομάδας
* **Αποτελεσματικότητα pit stop** = Μέσος χρόνος pit stop ανα αγώνα
* **DNFs (Did Not Finish/Δεν τελείωσε τον αγώνα)** = Ρυθμός βλαβών του αυτοκινήτου που οδηγεί σε απόσυρση από τον αγώνα
* **Σύγκριση συμπαικτών** = Διαφορά πόντων ανάμεσα στους συμπαίκτες
* **Αποτελεσματικότητα εκκίνησης** = Κερδισμένες θέσεις στον πρώτο γύρω κατά μέσο όρο

**7. Feature Για Θέματα Μηχανικής Μάθησης**

* Δημιουργία «**lag features»** (π.χ., μέσο αποτέλεσμα για τους τελευταίους 3 τελευταίους αγώνες)
* Χρήση κυλιόμενων μέσων όρων για την εξομάλυνση των διακυμάνσεων
* Μετατροπή κατηγορικών μεταβλητών (π.χ., οδηγοί, ομάδες, πίστες) σε one-hot encoding ή embeddings
* Χρήση time-series decomposition για να καταγράψουμε μακροπρόθεσμες τάσεις

**Trend Analysis Tasks & Models**

Οι τάσεις επικεντρώνονται στην ιστορική εξέλιξη των επιδόσεων για να βρούν μοτίβα στις δεξιότητες και στον τρόπο οδήγησης του οδηγού, την κυριαρχία της ομάδας, τα χαρακτηριστικά της πίστας και τις στρατηγικές αλλαγές.

1. **Τάσεις Απόδοσης Οδηγού**

**Trend Tasks:**

* **Driver Consistency Over Time** → Ποιοι οδηγοί είναι πιο συνεπής όσον αφορά τα αποτελέσματα τους;
* **Driver Peak Performance Years** → Πότε ένας οδηγός φτάνει στο αποκορύφομα της καριέρας του;
* **Declining or Improving Performance** → Πως η απόδοση του οδηγού αλλάζει ανα σεζόν;

**Features:**  
- **Rolling Average of Finishing Positions** (μέσω αποτέλεσμα για του τελευταίους 5, 10, 20 αγώνες)  
- **Τυπική απόκληση αποτελεσμάτων** (συνέπεια στους πόντους του οδηγού)  
- **Ηλικία vs Performance Curve (Καμπύλη απόδοσης)**  
- **Ρυθμός συσσώρευσης πόντων** (Συνολικοί πόντοι / δυνατοί πόντοι)  
- **Τάσεις στο Win Rate στην καριέρα του οδηγού**

**Μοντέλα για Trend Analysis:**  
**- Time-Series Analysis** (ARIMA, Exponential Smoothing)  
**- Rolling Mean & Moving Averages** (for visualizing consistency)  
**- Clustering (K-Means, DBSCAN)** (ομαδοποίηση οδηγών βασισμένη σε τάσεις επιδόσεων)

1. **Τάσεις Ομάδας**

**Trend Tasks:**

* **Team Dominance Evolution** → Ποιες ομάδες βελτιώνονται/χειρωτερεύουν;
* **Reliability Trends** → Ποιες ομάδες έχουν τις περισσότερες μηχανικές αποτυχίες;
* **Race Strategy Trends** → Οι στρατηγικές στα pit stop βελτιώνονται ή χειρωτερεύουν.

**Features:**  
- **Τάσεις στους πόντους της ομάδας ανα σεζόν**  
- **Win Rate διαχρονικά** (Νίκες ομάδας ανά σεζόν)  
- **Average Pit Stop Times Over Seasons** (Θετικά αποτελέματα της στρατιγικής που ακολουθείται)  
- **Technical Failure Rate (DNFs)** (Μηχανικές βλάβες ανα σεζόν)

**Models for Trend Analysis:**  
- **Time-Series Forecasting (ARIMA, LSTMs)**  
- **Moving Averages for Trend Smoothing**  
- **Regression Analysis to Find Long-Term Patterns**

1. **Τάσεις Πίστας**

**Trend Tasks:**

* **Overtaking Evolution** → Υπάρχουν πίστες που γίνεται όλο και δυσκολότερο να γίνουν προσπεράσεις;
* **Weather Impact Trends** → Ποιες πίστες επηρεάζονται πιο πολύ από την βροχή;
* **Qualifying vs. Race Performance Trends** → Τα τρία καλύτερα αποτελέσματα στα προκριματικά μεταφράζονται σε νίκες;

**Features:**  
- **Μέσος όρος θέσεων που αποκτήθηκαν σε αυτή την πίστα με την πάροδο του χρόνου**  
- **Ιστορική συχνότητα βροχόπτωσης σε αγώνες για κάθε πίστα**  
- **Κέρδος ή απώλεια θέσεις στα προκριματικά στην πίστα**  
- **Παρακολούθηση στις τάσεις συχνότητας του αυτοκινήτου ασφαλείας (Safety / car)**

**Models for Trend Analysis:**  
**- Time-Series Heatmaps** (Πως οι πίστες αλλάζουν στην πάροδο του χρόνου)  
**- Rolling Averages for Track-Based Performance**  
**- PCA for Dimensionality Reduction of Track-Specific Data**

**Prediction Tasks & Models**

Οι προβλέψεις περιλαμβάνουν: Την πρόβλεψη νικητών αγώνων, τερματισμούς στο βάθρο (1ος, 2ος, 3ος), DNF (Did not Finish) οδηγούς που δεν τελείωσαν τον αγώνα και τα γενικά αποτελέσματα του πρωταθλήματος χρησιμοποιώντας μοντέλα μηχανικής εκμάθησης.

* 1. **Predicting Race Winner (Classification)**

**Task:** Πρόβλεψε ποιος οδηγός θα νικήση έναν συγκεκριμένο αγώνα.

**Features:**   
**- Μέσος όρος θέσεις στους τελευταίους 5 αγώνες** (Momentum)   
**- Win Rate ανά πίστα** (Ιστορικό οδηγού στην πίστα)   
**- Πόντοι ομάδας ανά αγώνα (Τελευταίοι 5 αγώνες)**   
**- Θέση στα προκριματικά** (Σημαντικό σε αγώνες που δεν γίνονται εύκολα προσπεράσεις)

**Best Models:**  
- **Gradient Boosting (XGBoost, LightGBM, CatBoost)**  
- **Random Forest** (Καλό για δομημένα (Structured) δεδομένα)  
- **Neural Networks (Deep Learning for non-linear patterns)**

**Τύπος ταξινόμισης:** **Ομαδοποίηση πολλών κατηγοριών** (Κάθε οδηγός θα είναι ξεχωριστή ομάδα)

* 1. **Predicting Podium Finish (Binary Classification)**

**Task:** Πρόβλεψε εάν ένας οδηγός θα τερματήσει στους 3 πρώτους.

**Features:**  
- **Ρυθμός τερματισμού στο βάθρο τις τελευταίες 5 σεζόν**  
- **Μέσος όρος πόντων τους τελευταίους 5 αγώνες**  
- **Consistency του σκόρ στα προκριματικά του οδηγού**  
- **Προσαρμοστικότητα σε χειροτέρευση ελαστικών σε αυτή την πίστα**

**Best Models:**  
- **Logistic Regression (baseline model)**  
- **Random Forest (captures feature importance well)**  
- **XGBoost (handles complex interactions)**

**Τύπος ταξινόμησης:** **Διαδική ταξινόμηση** (1 = Podium, 0 = No Podium)

* 1. **Predicting Championship Winner (Season-Long Prediction)**

**Task:** Πρόβλεψε ποιος οδηγός θα κερδίσει το πρωτάθλημα πριν ή κατά την διάρκεια του πρωταθλήματος.

**Features:**  
- **Προβάδισμα σε πόντους στο τρέχον πρωτάθλημα**  
- **Πιθανότητα νίκης πρωταθλήματος βάση ιστορικού (Βασισμένο στη διαφορά πόντων των προηγουμένων πρωταθλημάτων)**  
- **Team Car Development Rate (Βελτίωση στους πόντους κατά την σεζόν)**  
- **Κούραση οδηγού / Παράγοντας ηλικίας**

**Best Models:**  
- **XGBoost & CatBoost** (best for structured data)  
- **Recurrent Neural Networks (LSTMs for sequence modeling)**  
- **Random Forest for ensemble learning**

**Τύπος ταξινόμησης:** **Ομαδοποίηση πολλών κατηγοριών** (Προβλέποντας ποιος οδηγός θα κερδίσει)

* 1. **Predicting a Driver's Finishing Position (Regression)**

**Task:** Προέβλεψε την ακριβής θέση που θα τερματίσει ο οδηγός.

**Features:**  
- **Θέση προριματικών**  
- **Μέσος όρος θέσης στου τελευταίους 5 αγώνες**  
- **Καιρικές συνθήκες (Wet vs. Dry)**  
- **Δυναμικότητα ομάδας (Πόντοι ανα αγώνα για τους τελευταίους 5 αγώνες)**

**Best Models:**  
- **Gradient Boosting Regression (XGBoost, LightGBM, CatBoost)**  
- **Random Forest Regression**  
- **Neural Networks for Deep Learning Regression**

**Τύπος ταξινόμησης:** **Regression** (Προβλέποντας συνεχές θέσεις)

* 1. **Predicting If a Driver Will DNF (Did Not Finish)**

**Task:** Προέβλεψε εάν ένας οδηγός δεν θα τελειώσει τον αγώνα.

**Features:**  
- **Ιστορικό συχνότητας DNF** (Οδηγού & Ομάδας)  
- **Ρυθμός στις μηχανικές βλάβες (Τελευταία 3 χρόνια)**  
- **Wet Weather Risk Score**  
- **Πιθανότητα σύγκρουσης στο πρώτο γύρω (Βασισμένη στην θέση εκκίνησης & στην πίστα)**

**Best Models:**   
- **Logistic Regression (baseline for binary classification)**  
- **Random Forest (captures non-linearities in crash data)**  
- **XGBoost (handles imbalanced DNF data better than others)**

**Τύπος ταξινόμησης:** **Διαδική ταξινόμηση** (1 = DNF, 0 = Τερμάτισε)

* 1. **Predicting Constructor Championship Winner**

**Task:** Προέβλεψε ποια ομάδα θα κερδίσει το πρωτάθλημα.

**Features:**  
- **Ρυθμός συλλογής πόντων της ομάδας**  
- **Αξιοπιστία αυτοκινήτου (Μηχανικές βλαβες ανα σεζόν)**  
- **Budget Ομάδας & Έξοδα ανάπτυξης (Εάν υπάρχουν)**  
- **Σύγκριση ικανοτήτων οδηγού και συμπαίκτη**

**Best Models:**

* **CatBoost (handles categorical variables well)**
* **XGBoost (most powerful gradient boosting model for structured data)**
* R**andom Forest (for ensemble learning & feature importance ranking)**

**Τύπος ταξινόμησης:** **Ομαδοποίηση πολλαπλών κατηγοριών** (Προέβλεψε ποια ομάδα θα κερδίσει)

**Τελική Περίληψη**

**Ανάλυση τάσεων (Κατανοώντας μοτίβα)**

* **Μοντέλα:** Ανάλυση Time-Series, Κινούμενα μέσα, Clustering
* **Tasks:** Εξέλιξη απόδοσης οδηγού, Τάσεις στην κυριαρχεία ομάδων, Προσαρμοστικότητα στις πίστες

**Προβλέψεις (Προβλέποντας αποτελέσματα)**

* **Μοιντέλα:** XGBoost, Random Forest, Neural Networks
* **Tasks:** Νικητές αγώνων, Podium Finish, Νικητής πρωταθλήματος, DNFs

**Πλάνο**

1. Καθορισμός στόχου (Τάσεις & Προβλέψεις και λύση προβλήματος)
2. Συλλογή & Καθαρισμός δεδομένων
3. Διερευνητική Ανάλυση Δεδομένων (E.D.A.) δηλαδή οπτικοποίηση και κατανόηση
4. Επιλογή των Feature που θα χρησιμοποιήσουμε
5. Κατασκευή & Αξιολόγηση Μοντέλων
6. Ανάλυση τάσεων (προαιρετικό)
7. Οπτικοποίηση Τάσεων και Προβλέψεων
8. Συμπέρασμα & Αναφορά