Θεωρια για το ΑΙ και για classification models etc.

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**

**Circuit-Specific Trends**

**Trend Tasks:**

* **Overtaking Evolution** → Are some circuits becoming harder to overtake?
* **Weather Impact Trends** → Which tracks are most affected by rain?
* **Qualifying vs. Race Performance Trends** → Are pole positions translating into wins?

**Features:**  
✔ **Average Positions Gained at This Track Over Time**  
✔ **Historical Rain Race Frequency at Each Circuit**  
✔ **Qualifying-to-Finish Position Gain/Loss at This Track**  
✔ **Track Safety Car Frequency Trends**

**Models for Trend Analysis:**  
**Time-Series Heatmaps** (how circuits change over time)  
**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.

**Predicting Race Winner (Classification)**

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

**Features:**   
**- Last 5 Races Finishing Position Average** (momentum)   
**- Circuit-Specific Win Rate** (driver history at this track)   
**- Team Points Per Race (Last 5 Races)**   
**- Qualifying Position** (important at tracks with low overtaking)

**Best Models:**  
- **Gradient Boosting (XGBoost, LightGBM, CatBoost)**  
- **Random Forest** (good for structured data)  
- **Neural Networks (Deep Learning for non-linear patterns)**

**Classification Type:** **Multi-Class Classification** (each driver is a class)

**Predicting Podium Finish (Binary Classification)**

**Task:** Predict whether a driver will finish in the top 3.

**Features:**  
- **Podium Rate Last 5 Seasons**  
- **Last 5 Races Average Points**  
- **Driver Qualifying Consistency Score**  
- **Tire Degradation Adaptability at This Circuit**

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

**Classification Type:** **Binary Classification** (1 = Podium, 0 = No Podium)

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

**Task:** Forecast which driver will win the championship before or during the season.

**Features:**  
✔ **Current Championship Points Lead**  
✔ **Historical Championship Win Probability (Based on Points Gaps in Previous Years)**  
✔ **Team Car Development Rate (Improvement in Points Over a Season)**  
✔ **Driver Fatigue / Age Factor**

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

**Classification Type:** **Multi-Class Classification** (predicting which driver wins)

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

**Task:** Predict the exact finishing position of a driver.

**Features:**  
✔ **Qualifying Position**  
✔ **Last 5 Races Finishing Position Average**  
✔ **Weather Conditions (Wet vs. Dry)**  
✔ **Team Strength (Points Per Race Last 5 Races)**

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

**Classification Type:** **Regression** (predicting continuous finishing positions)

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

**Task:** Predict if a driver will fail to finish a race.

**Features:**  
✔ **Historical DNF Rate** (driver & team)  
✔ **Mechanical Failure Rate (Last 3 Years)**  
✔ **Wet Weather Risk Score**  
✔ **First Lap Collision Probability (Based on Grid Position & Track)**

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

**Classification Type:** **Binary Classification** (1 = DNF, 0 = Finished)

**Predicting Constructor Championship Winner**

**Task:** Predict which team will win the Constructor's Championship.

**Features:**  
✔ **Team Points Accumulation Rate**  
✔ **Car Reliability Factor (Technical Failures per Season)**  
✔ **Team Budget & Development Spending (if available)**  
✔ **Driver-Teammate Strength Comparison**

**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)**

**Classification Type:** **Multi-Class Classification** (predict which team wins)

**Final Summary**

**Trend Analysis (Understanding Patterns)**

* **Models:** Time-Series Analysis, Moving Averages, Clustering
* **Tasks:** Driver performance evolution, team dominance trends, track adaptation

**Predictions (Forecasting Outcomes)**

* **Models:** XGBoost, Random Forest, Neural Networks
* **Tasks:** Race winner, podium finish, championship winner, DNFs