Θεωρια για το ΑΙ και για classification models etc.

Basic Features:

**1. Features Αποτελεσμάτων Οδηγού**

* **Career Points per Race** = Total points / Total races
* **Win Rate** = Total wins / Total races
* **Podium Rate** = Total podiums / Total races
* **Top-10 Finish Rate** = Number of top-10 finishes / Total races
* **Average Finish Position** = Mean finishing position across seasons
* **Career Race Count** = Total races participated in
* **Season Best Finish** = Best position achieved in a season

**2. Features Αποτελεσμάτων Ομάδας**

* **Team Dominance Index** = (Total wins per season) / (Total races per season)
* **Constructor Consistency** = Standard deviation of team points across seasons (lower values indicate consistency)
* **Points Per Race (Team)** = Total team points / Races in a season
* **Number of 1-2 Finishes** = Number of times the team secured both 1st and 2nd place

**3. Features Ανα Πίστα**

* **Track Win Percentage** = Wins at a circuit / Total races at that circuit
* **Circuit Specialist Score** = (Driver’s performance at a circuit vs. their average performance)
* **Weather Performance** = Performance in wet races vs. dry races
* **Qualifying vs. Race Performance** = Difference between average grid position and average finish position at a track

**4. Features Αποτελεσμάτων στους προκριματικούς αγώνες**

* **Average Qualifying Position** = Mean of grid positions
* **Pole Position Rate** = Number of pole positions / Total races
* **Qualifying Consistency** = Standard deviation of grid positions

**5. Features Μακροπρόθεσμων Τάσεων & Σεζόν**

* **Performance Delta** = Change in points/rank/finishing position compared to the previous season
* **Momentum Index** = Weighted moving average of past 5 races' finishing positions
* **Win Streaks** = Number of consecutive wins in a season
* **Driver Age Factor** = Normalize driver performance based on age buckets (young, prime, veteran)
* **Championship Lead Stability** = Number of times the driver led the championship throughout a season

**6. Features Για Προβλέψεις Μελλοντικών Αποτελέσματων**

* **Driver-Team Synergy Score** = Performance changes when switching teams
* **Pit Stop Efficiency** = Average pit stop time per race
* **Technical DNFs (Did Not Finish)** = Rate of car failures leading to DNFs
* **Teammate Comparison** = Difference in points between teammates
* **Race Start Efficiency** = Positions gained in first lap on average

**7. Feature Για Θέματα Μηχανικής Μάθησης**

* Create **lag features** (e.g., past 3 races’ average finish)
* Use **rolling averages** to smooth fluctuations
* Convert categorical variables (drivers, teams, circuits) into **one-hot encoding or embeddings**
* Use time-series decomposition to capture long-term trends

--------------------------------------------------------------------------------------------------------------------------------------

**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** (total points earned / possible points)  
- **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)** (group drivers based on performance patterns)

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)** (Μηχανιές 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