A background image of an F1 race track with several cars. The cars are semi-transparent, showing the underlying track and the text. The track is grey with red and white curbs. The cars are in various colors: orange, yellow, black, red, and blue.

# **Predicting Average Lap Times for the 2024 F1 Abu Dhabi GP**

**Kimaya & Ainsley  
Machine Learning**

# Introduction

**Problem Type:** Regression

**Goal:** Predict each driver's *average race lap time* (s) for the 2024 Abu Dhabi GP

**Why It Matters:**

- Lap time depends on driver skill, tyre, circuit, weather, and car performance
- Helps evaluate predictive power of ML methods
- Shows how features like tyre age, weather, etc. influence strategic decisions on track

**Approach:**

- FastF1 for data extraction
- Feature engineering (tyre age, circuit type, weather, avg speed, team, etc.)
- One-hot encoding for categorical variables
- Random Forest model
- Train on races 1–23 → Test on Abu Dhabi





## Dataset Overview

**Source:** FastF1 open-access & weather data

Includes:

- Lap times (converted to seconds)
- Driver, team, tyre compound
- Lap number, stint info
- Rainfall flag
- Circuit metadata (street vs permanent)

## Data Preprocessing

Removed laps <30 sec (invalid in/out laps or errors)

- Dropped missing lap times
- Rainfall = 1, No Rain = 0
- **TyreAge:** cumulative laps since tyre change
- **Normalization:** Avg Speed =  $\text{Track Length (km)} / (\text{LapTime}/3600)$

## Train–Test Split

**Train:** Races 1–23

(all 2024 races before Abu Dhabi)

**Test:** Abu Dhabi only

(future race prediction)

- Ensures realistic evaluation
- No hyperparameter tuning performed

## Model Selection

**Baseline:** Linear Regression → rejected; poor performance

**Main Model:**

- **Random Forest Regressor**
  - 500 trees
  - max\_depth = 15
  - random\_state = 42
- **Reasons for choosing RF:**
  - Captures nonlinear relationships
  - Works well with tabular data
  - Robust to noise
  - Provides feature importance

## Evaluation Metrics

**Final Results:**

- **RMSE:** 1.240 sec
- **MAE:** 1.038 sec
- **R<sup>2</sup>:** 0.281

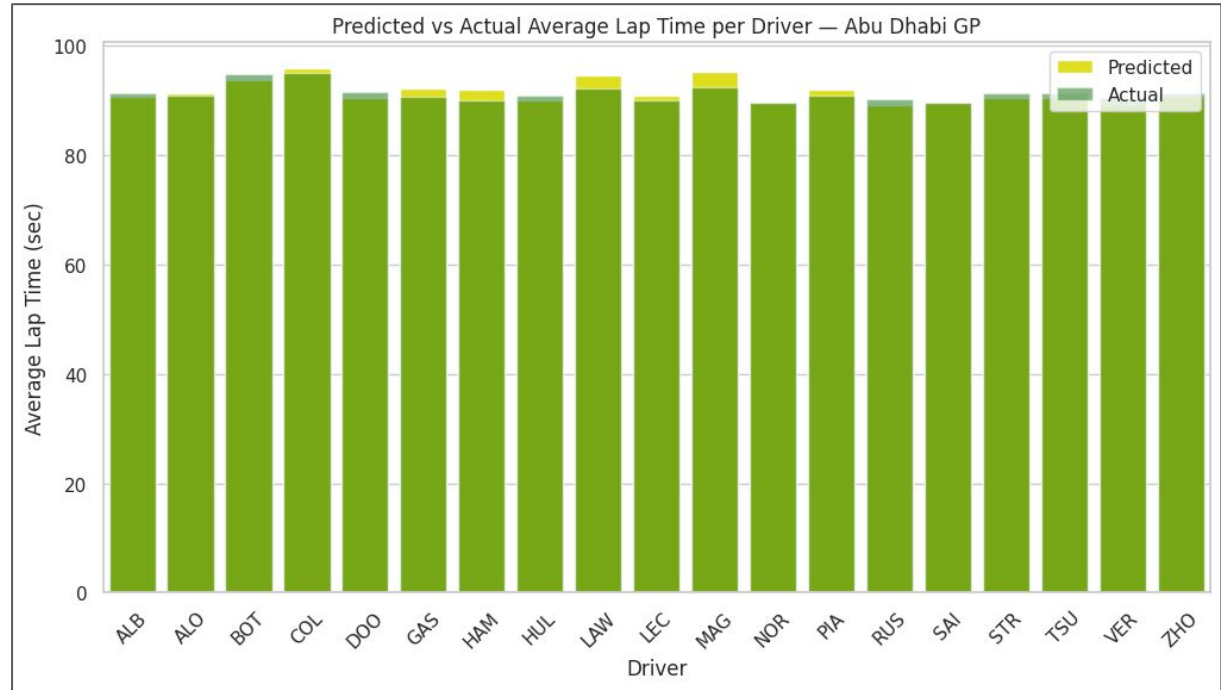
**Interpretation:**

- Predictions generally within ~1 sec
- Moderate accuracy given unpredictable race conditions
- Model captures ~28% of lap-time variance
- Reasonable for complex F1 pace prediction

# Predicted vs Actual Lap Times

- Most predictions within 1–2 seconds, RF model captures general race-pace trends well
- Over Predictions:
  - MAG (+2.90),
  - LAW (+2.27),
  - HAM (+1.94)
- Under Predictions:
  - VER (−1.28),
  - BOT, DOO (∼1s)

The model tends to struggle when drivers perform **significantly better or worse** in Abu Dhabi than their season averages.



This indicates that:

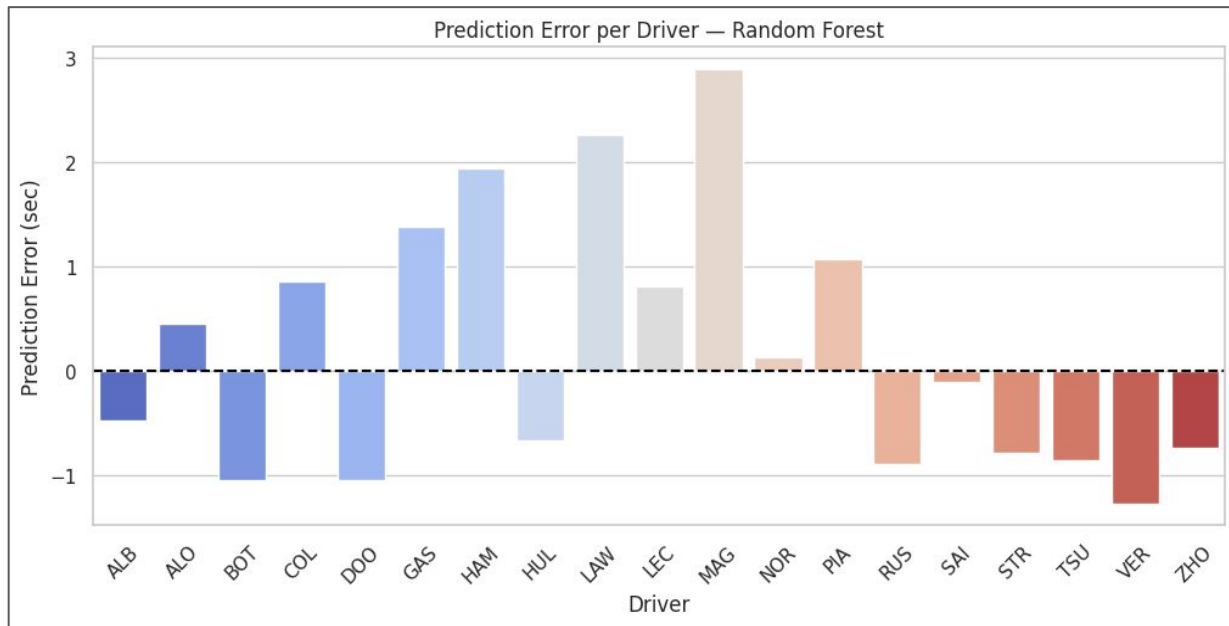
- Driver-specific Abu Dhabi performance differed from typical patterns
- The model lacks features that capture **race-specific form**, fuel load, ERS usage, and team setup choices

## Prediction Error per Driver

- Positive = predicted too slow
- Negative = predicted too fast

Clear pattern: some drivers

- overpredicted
  - (MAG, LAW)
- underpredicted
  - (VER, DOO)
- Largest errors come from drivers whose Abu Dhabi pace differed from their season trend.
- Shows model is good at typical performance, weaker with unexpected results.



The model captures the overall pace hierarchy but misses unexpected performance spikes, likely due to unmodeled factors such as car setup changes, track evolution, in-race incidents, or strategic differences.

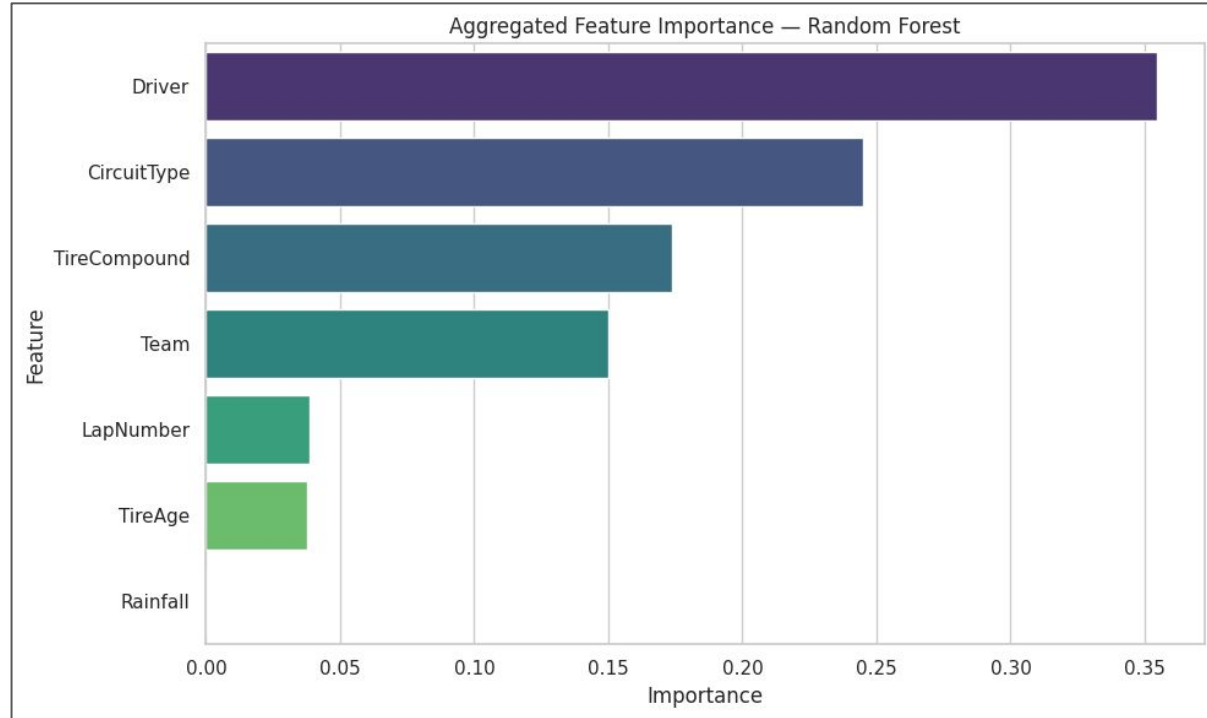
# Feature Importance

Top Influential Features:

1. **Driver** (~0.36): Driver identity is the strongest predictor of lap time
2. **CircuitType** (~0.24)
3. **Tyre Compound** (~0.17)
4. **Team** (~0.15)

Lower Influence:

- Lap Number & TireAge (<0.05)
- Rainfall (minimal impact since Abu Dhabi dry)



Performance is mainly driven by driver talent, circuit type, and tyre choice, low importance of TireAge suggests the model would benefit from more detailed tyre-degradation features.

Rainfall shows minimal importance, which is expected given the absence of wet sessions in the dataset.



## Limitations

No telemetry features (speed, throttle, braking, ERS)

Simplified weather variable

No fuel-load modeling

Driver variability between weekends not modeled

Timing anomalies or unusual slow laps may remain

## Conclusion

Accurate predictions (within ~1 second for most drivers)

Identified which features most affect lap times

Modeled realistic future-race performance

## Future Work

Add telemetry features for richer modeling & higher accuracy

Test more ML models + hyperparameter tuning, also test on more features

Improve UI/UX and dataset flexibility

Add weather variables (track temp, wind, humidity)



Thank you!