Velocity Insight

(A Fusion of Speed and Data-Driven F1 Race Predictions)

CMPE 257 | San Jose State University

Team Introduction

Hanlin Chen

Shrivaikunth Krishnakumar

Sravan

Sai Chetan

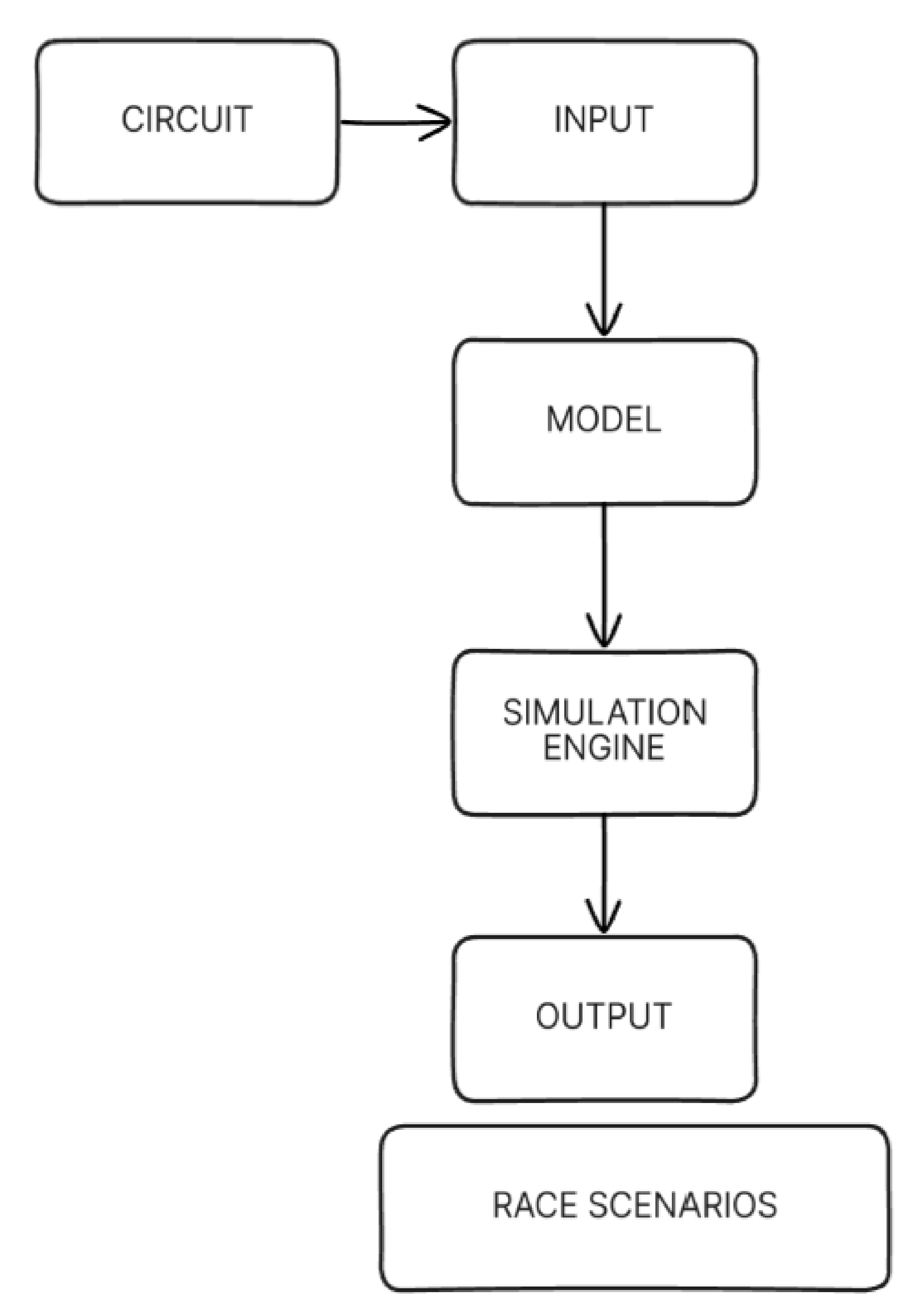
Brian Lam

Problem Statement

- Formula 1 is a high-speed motorsport with strategic decisions like Pit stops, tire choices and driver performance

- We developed a Machine Learning model to predict lap times for each driver and build a simulation engine to provide optimal pit stops and race scenarios

- The goal is to provide teams with race scenarios and optimal pit stops to make data driven decisions



Solution Overview

Data Collection

- Historical race data (2021–2024) sourced via FastF1 API
- Verified race metadata with external datasets Kaggle for Cross Verification and consistency

Approach

- Built Supervised regression models for lap time prediction.
- Focused on single circuit to remove cross-track variability
- Engineered race specific features like: tire degradation, stint management

Prediction

- Predicated lap-by-lap performance for each driver
- Forecasted pit-stop timings, Tire compound selection

Strategy

- Enabled tactical decisions making by predicting race evolution

Solution Approach

- **Data Preprocessing**

  Cleaned & structured **Abu Dhabi** raw race data for consistent analysis.

- **Feature Engineering**

  Create features like tire degradation, Stint progress and Pit stops

  Designed features to capture race dynamics

- **Data Split**

  Split seasons 2021-2023 for training, and 2024 reserved for testing.

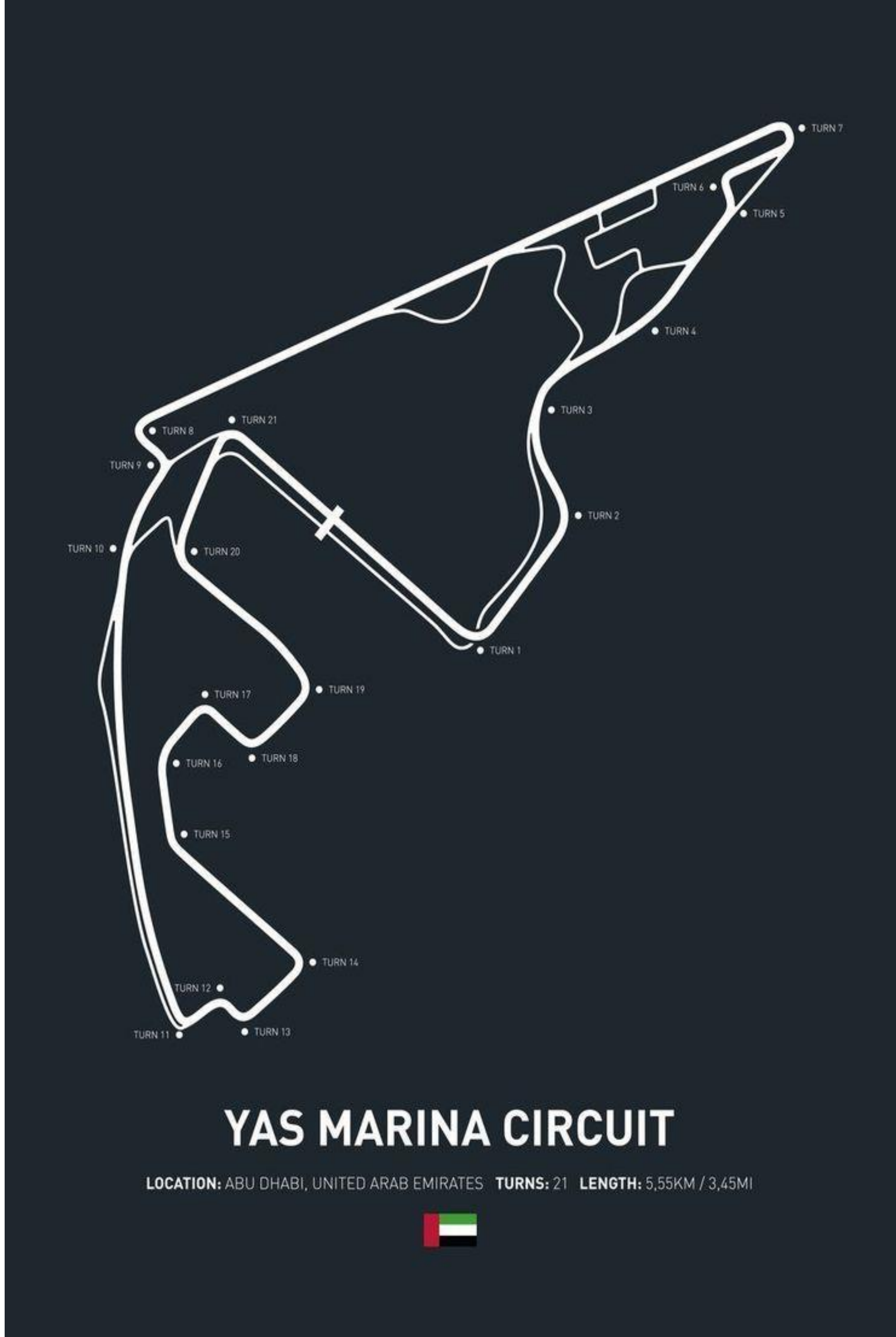
- **Model Training**

  Trained Random forest and Gradient Boosting model for lap time predictions.

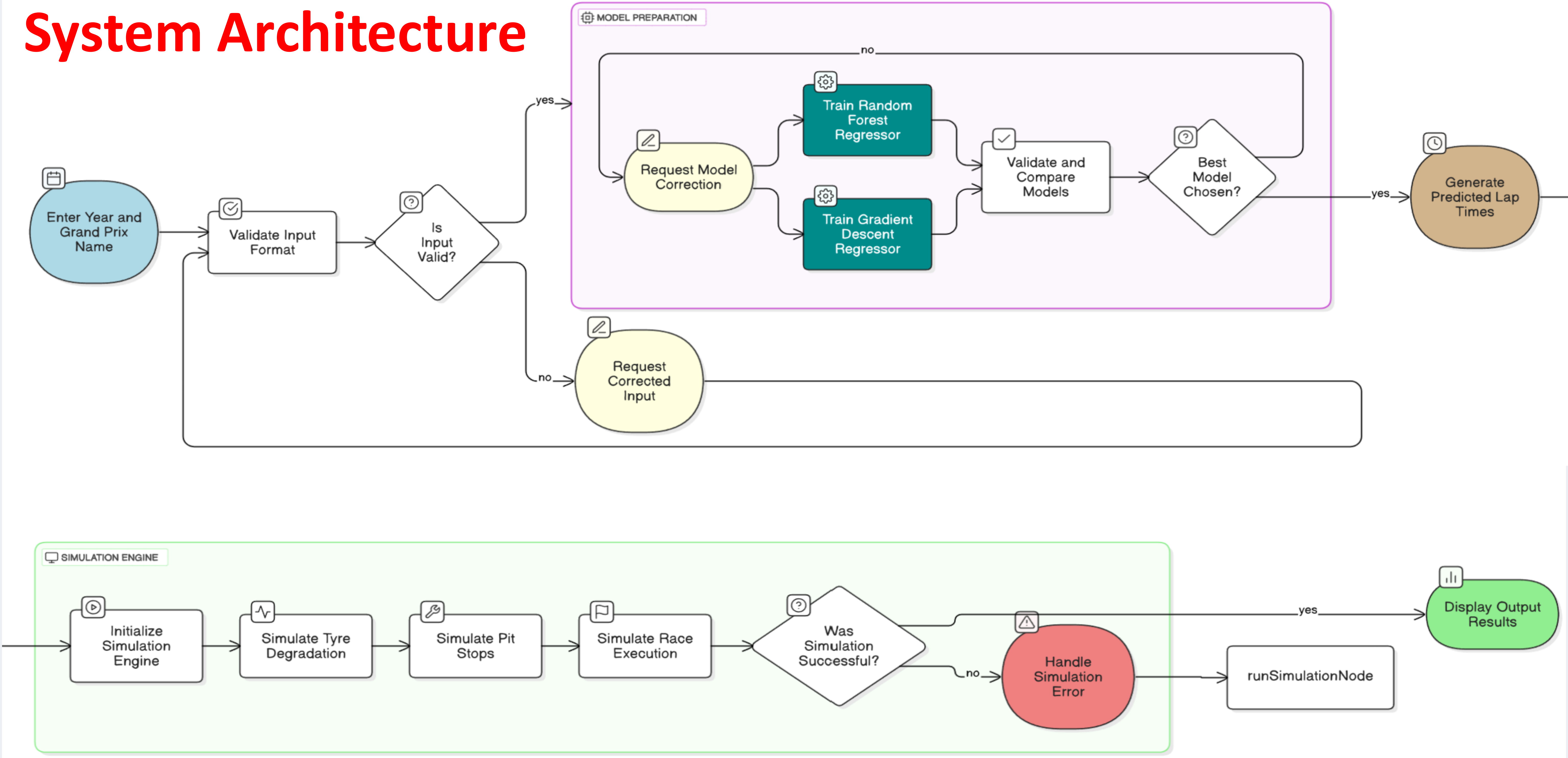
- **Simulation and Iterative Refinement**

  Built a optimization algorithm based simulation engine.

  Improved models by adjusting features and validating techniques



System Architecture



Model Architecture

Random Forest Regressor

- Best performing model in our case
- Builds independent decision trees and combine their predictions.
- Fast, robust and captured lap time variations effectively

Gradient Boosting Regressor

- Trained for comparison
- Builds model Sequentially. Tries to correct errors of previous one
- Great for capturing small patterns and reduces bias over time

Why both Gradient Boosting Regressor and Random Forest Regressor ?

- GBR is better for fine-tuned, high accurate predictions and performs better on large, multi-circuits data
- RF offer quick and stable baseline predictions. Handled circuit-level data better
- Together, they allow comparison between precision vs generalization for race simulation

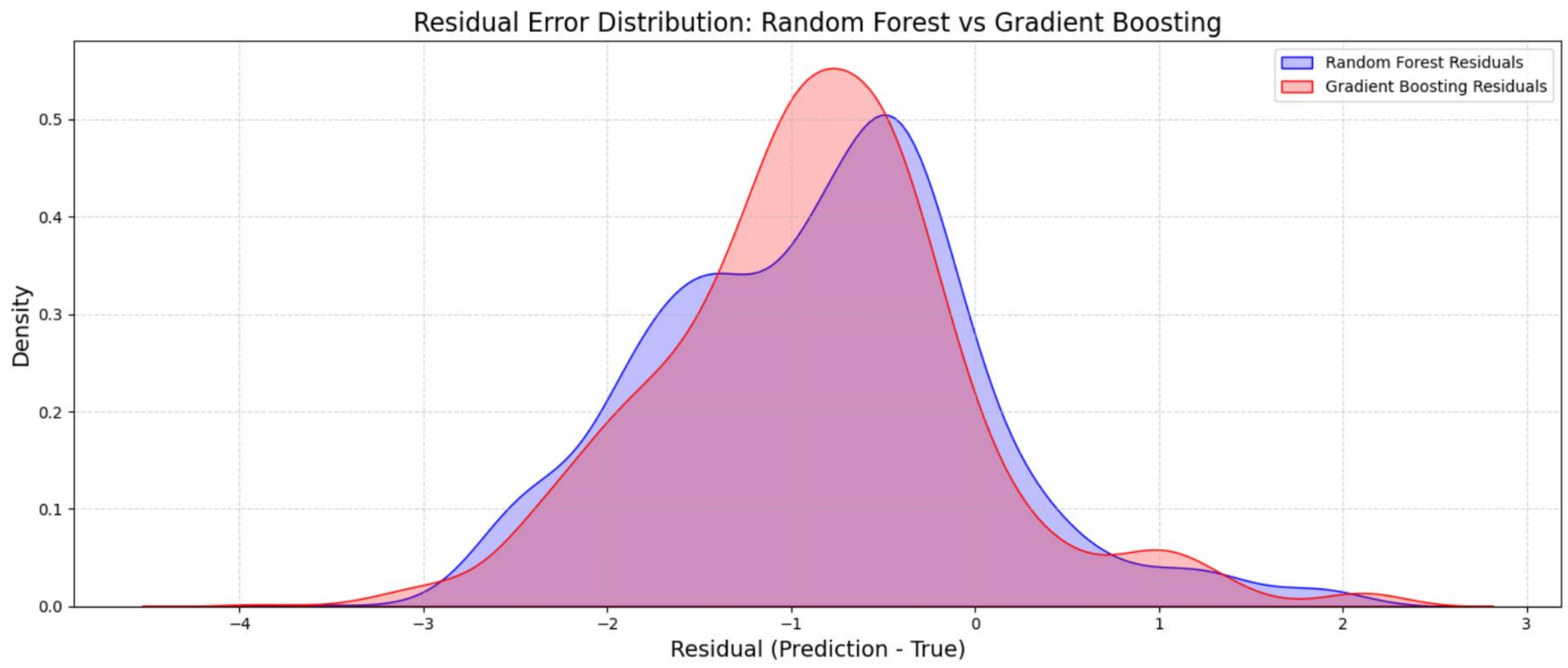
Results

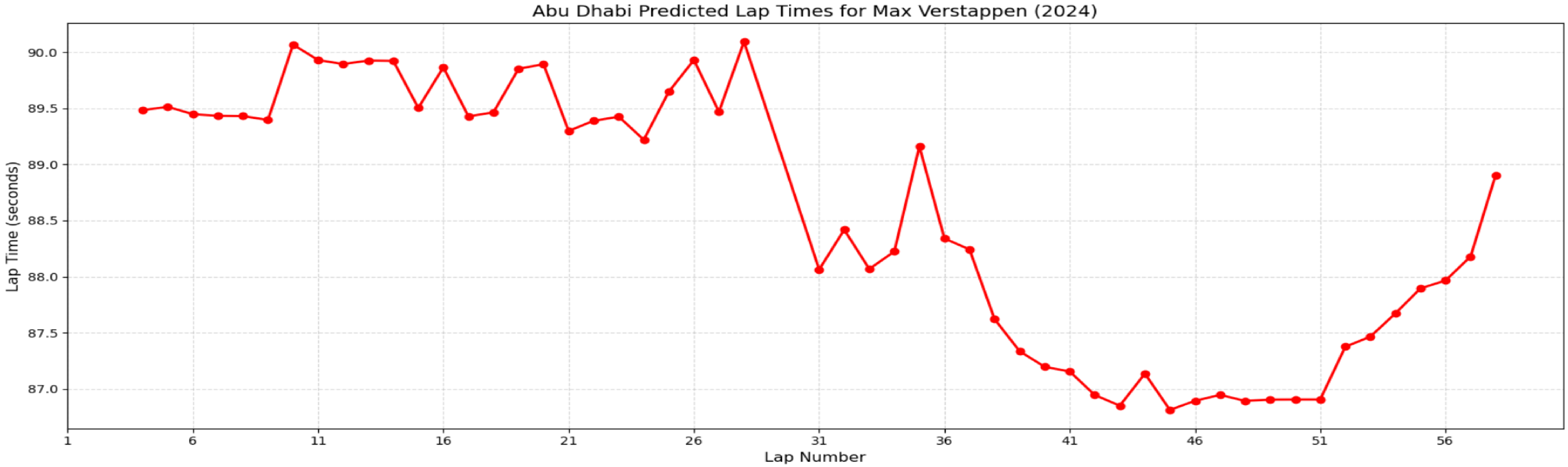
Performance Metrics – Random Forest

- Mean Absolute Error (MAE): 1.012 seconds
- Median Absolute Error (MAE): 0.867 seconds
- Root Mean Squared Error (RMSE): 1.22 seconds
- R^2 Score = 0.7168

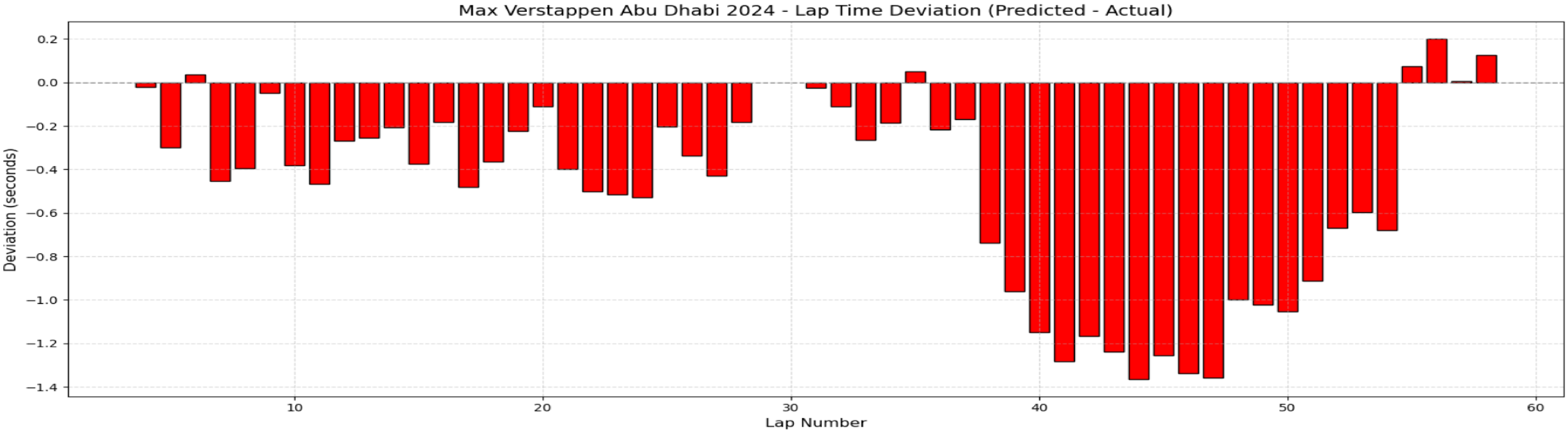
Performance Metrics – Gradient Boosting

- Mean Absolute Error (MAE): 1.017 seconds
- Median Absolute Error (MAE): 0.925 seconds
- Root Mean Squared Error (RMSE): 1.21 seconds
- R^2 Score = 0.6656

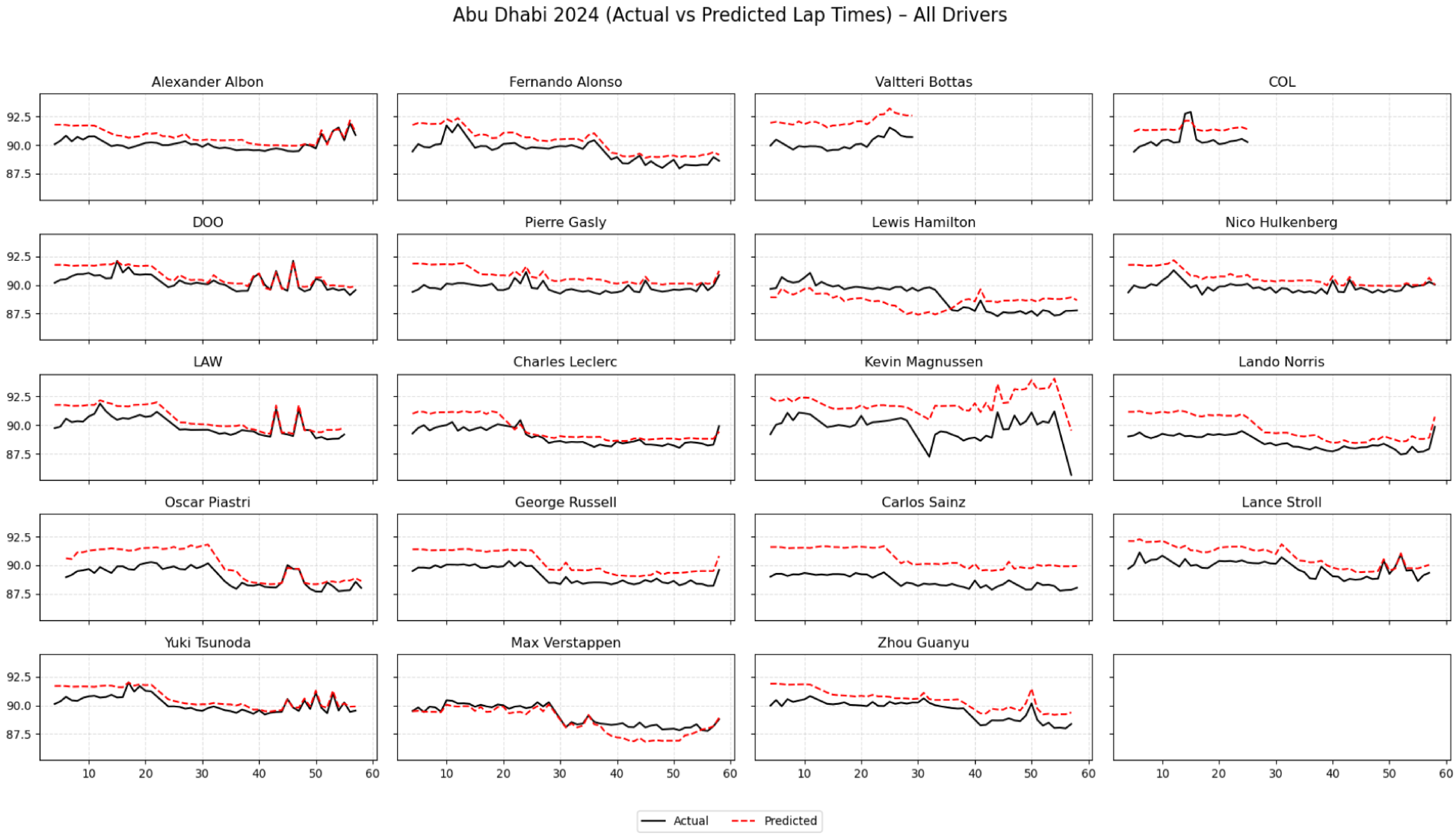




Abu Dhabi Predicted Lap Times for Max Verstappen (2024)
Lap Time (seconds)
Lap Number



Max Verstappen Abu Dhabi 2024 - Lap Time Deviation (Predicted - Actual)
Deviation (seconds)
Lap Number



Abu Dhabi 2024 (Actual vs Predicted Lap Times) – All Drivers
Alexander Albon
Fernando Alonso
Valtteri Bottas
COL
DOO
Pierre Gasly
Lewis Hamilton
Nico Hulkenberg
LAW
Charles Leclerc
Kevin Magnussen
Lando Norris
Oscar Piastri
George Russell
Carlos Sainz
Lance Stroll
Yuki Tsunoda
Max Verstappen
Zhou Guanyu
92.5
90.0
87.5
10
20
30
40
50
60
Actual
Predicted

Simulation Engine

- The engine simulated lap-by-lap performance for top drivers (Lewis Hamilton, Max Verstappen, etc.)
- Optimal pit strategies suggested a **one-stop** and **two-stint** approach:
  - Start on **HARD compound** for long initial stints (35–50 laps)
  - Switch to **SOFT** (or **MEDIUM** for Leclerc) for final sprint to the end.
  - Predicted total race times closely matched actual leader timing:
- E.g. Max Verstappen gap to Hamilton predicted as **+0.217s**, very minimal!

Race Simulation for Abu Dhabi grand prix - 2024:

Driver	Stint 1	Compound 1	Stint 2	Compound 2	Stint 3	Compound 3	Gap	Total Race Time
Max Verstappen	1-21	HARD	22-42	SOFT	43-58	MEDIUM	Leader	1:21:52.484
Lewis Hamilton	1-50	HARD	51-51	MEDIUM	52-58	SOFT	+0:00:04.529	1:21:57.013
Lando Norris	1-38	HARD	39-40	MEDIUM	41-58	SOFT	+0:01:26.402	1:23:18.886
Charles Leclerc	1-37	HARD	38-49	MEDIUM	50-58	SOFT	+0:01:33.070	1:23:25.554
Oscar Piastri	1-30	HARD	31-38	MEDIUM	39-58	SOFT	+0:01:39.256	1:23:31.740

Main causes for deviations (small differences observed):

- **Tire degradation variance:** Real-world tire wear can be slightly unpredictable due to temperature changes and track conditions.
- **Traffic and overtakes:** Being stuck behind slower cars may increase lap times unexpectedly.
- **Driver behavior:** Aggressive vs conservative driving styles were not fully captured.
- **Unmodeled race incidents:** Minor contact, yellow flags, or slight mechanical issues during race were not part of simulation assumptions.

Scope for Future Enhancement

- Generalize Model for Multiple Circuits
- Handle Dynamic race Scenarios like Safety cars, yellow flags more accurately
- Develop Real-Time Strategy Engine that can dynamically re-optimize race strategies

- **In conclusion**, teams that can effectively leverage machine learning can gain a significant strategic edge in Formula 1.

Thank You

We appreciate your time and attention.

Looking forward to discussing next steps and answering questions.