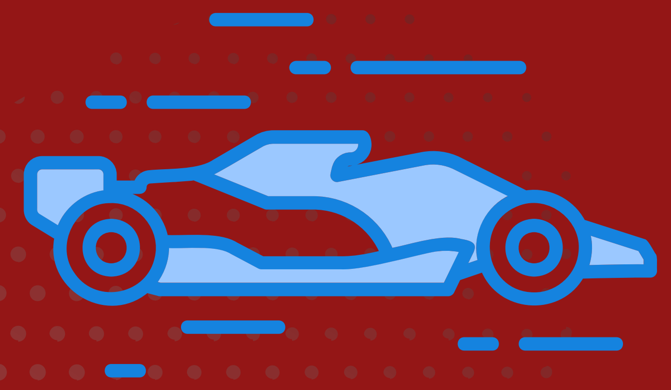
F1 OPTIMAL PIT STOP PREDICTOR

By : Kaloyan Dragiychev



CONTEXT

F1 teams make 1,200+ tactical decisions per race in split seconds.

What if AI could predict the optimal strategy before they even know it?

THE BIG IDEA: AI-powered tactical predictor beyond Formula 1

- Real-time race strategy recommendations
- Transforms reactive decisions into predictive intelligence

VISION: From manual calculations → AI-driven race strategy

PROBLEM

THE PROBLEM:

- F1 strategy = extreme complexity (tire degradation, weather, competitors)
- Some teams rely on manual calculations + historical precedents
- Critical decisions made in seconds with incomplete information
- Current approach: reactive, not predictive

MY SOLUTION:

- Historical data training (FastF1 telemetry)
- Predictive analytics + interactive visualization

DATA PREP AND FEATURES

Data Collection

- Collected F1 data from 2022, 2023, and 2024 seasons using FastF1 library
- Total: 74,605 lap records from 44 complete races
- Loaded lap times, weather data, tire information, and pit stop data
- Set up caching system to store data locally for faster access

Data Cleaning

- Fixed data types
- Handled missing values: 96.5% missing pit times (normal since most laps don't have pit stops)
- Checked data quality and consistency across all seasons

Feature Engineering

Created new features from the original data:

Basic Features:

- NumberOfPitStopsMade
- RaceFractionCompleted
- IsSafetyCar and IsVSC flags from track status data

Time-Based Features:

- PreviousLapTimeSeconds1 and PreviousLapTimeSeconds2 (lap times from 1-2 laps ago)
- Rolling averages and statistics over 3, 5, and 8 lap windows
- Trend analysis (lap time changes over time)

Target Variable:

- PittedInNextNRows = 1 if driver pits in next 3 laps, 0 if not
- This creates a binary classification problem

ROADMAP

1

Initial Prototype

- Developed a rule-based pit stop suggestion system using the FastF1 library
- Implementing tire wear thresholds and lap time degradation detection.

15	00:01:58	SOFT	2	1	N/A	–
16	00:01:37	SOFT	2	2	N/A	–
17	00:01:37	SOFT	2	3	-0.08	–
18	00:01:37	SOFT	2	4	-0.11	–
19	00:01:37	SOFT	2	5	-0.01	–
20	00:01:37	SOFT	2	6	+0.06	–
21	00:01:37	SOFT	2	7	-0.14	–
22	00:01:37	SOFT	2	8	-0.07	–
23	00:01:37	SOFT	2	9	-0.17	–
24	00:01:37	SOFT	2	10	-0.34	–
25	00:01:37	SOFT	2	11	-0.35	–
26	00:01:37	SOFT	2	12	-0.04	–
27	00:01:37	SOFT	2	13	-0.35	–
28	00:01:37	SOFT	2	14	-0.07	–
29	00:01:37	SOFT	2	15	-0.15	Pit Window Open (Lap Count)
30	00:01:37	SOFT	2	16	-0.18	Pit Window Open (Lap Count)
31	00:01:37	SOFT	2	17	-0.29	Pit Window Open (Lap Count)
32	00:01:37	SOFT	2	18	-0.03	Consider Pit (Lap Count)
33	00:01:37	SOFT	2	19	-0.04	Consider Pit (Lap Count)
34	00:01:37	SOFT	2	20	+0.11	Consider Pit (Lap Count)
35	00:01:37	SOFT	2	21	+0.13	Consider Pit (Lap Count)
36	00:01:40	SOFT	2	22	+3.24	High Degradation / Consider Pit (Lap Count)

ROADMAP

2

Random Forest

- **Why Random Forest?:** Handles mixed data types, robust to outliers, interpretable, ease of use.
- **The Good:** 81% accuracy, stable performance
- **The Problem:** Only 36% F1-score - missing too many pit opportunities
- Key Insight: "Even balanced class weights couldn't solve the fundamental imbalance problem"

```
--- Final Model Evaluation (Test Set) ---
Test Accuracy: 0.8171
Test Precision (PittedInNextNRows=1): 0.2550
Test Recall (PittedInNextNRows=1): 0.6221
Test F1-Score (PittedInNextNRows=1): 0.3617
Test ROC AUC: 0.8088

Test Confusion Matrix:
[[8992 1779]
 [ 370  609]]
```

ROADMAP

2

XGBoost

- **Problem:** Random Forest couldn't capture complex relationships between features
- **Solution**: XGBoost builds many small decision trees that learn from each other's mistakes
- **Optimization Strategy**: RandomizedSearchCV with 80 iterations, GroupKFold CV
- **SMOTE Enhancement**:
  - The Experiment: "What if we balance the training data?"
  - Result: Precision jumped but recall inconsistent
  - Lesson: "Synthetic samples help, but we need smarter learning"

```
Test
accuracy : 0.868
precision: 0.428
recall   : 0.581
F1       : 0.493
ROC-AUC  : 0.859
confusion:
 [[6818  727]
 [ 392  545]]
```

ROADMAP

3

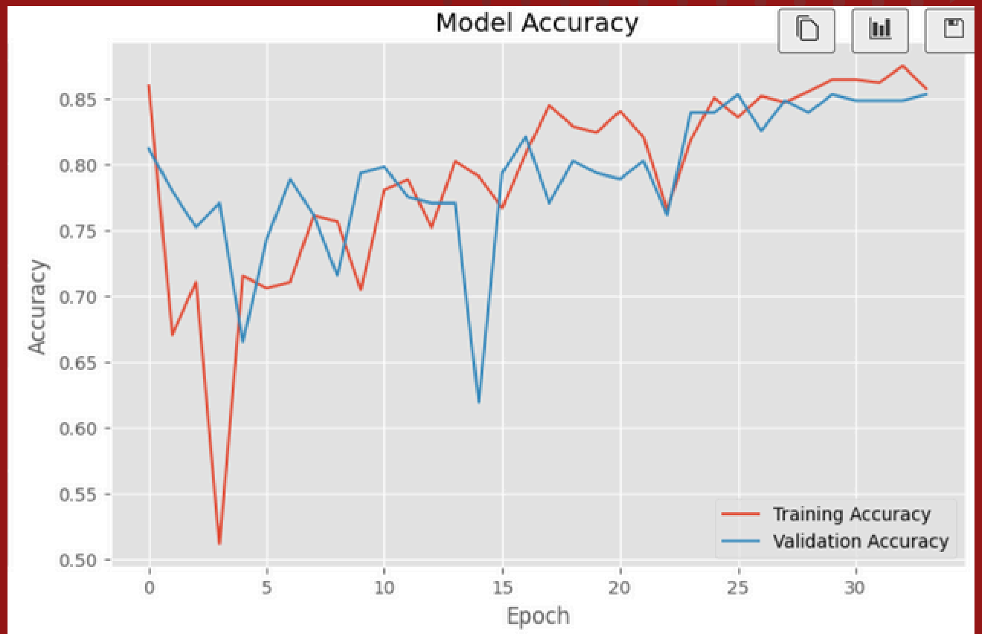
LSTM

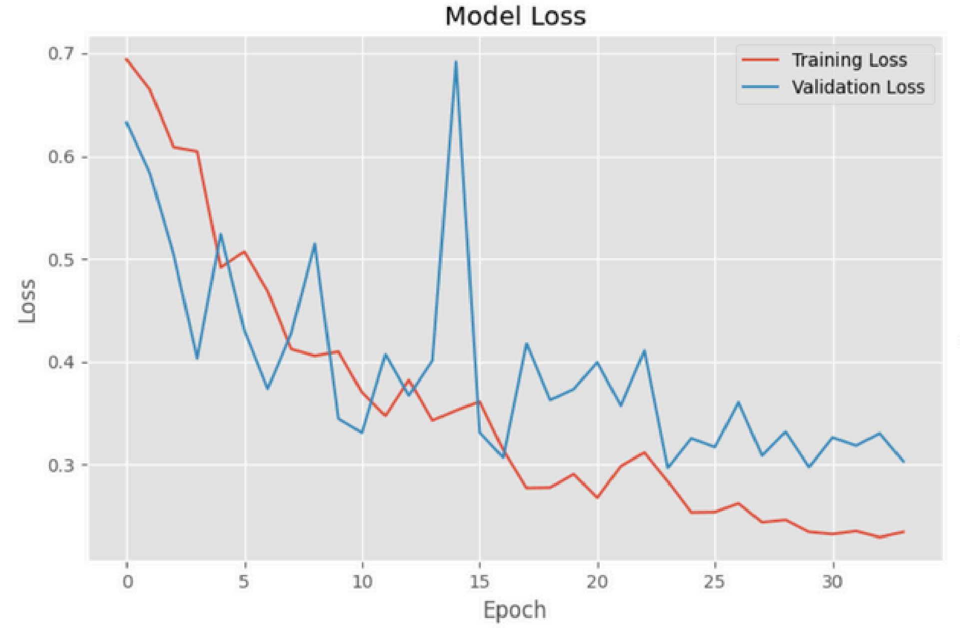
LSTM Advantages:

- **Temporal Memory:** Remembers tire degradation patterns from 5 laps ago
- **Sequential Learning:** Understands that lap N depends on laps N-1, N-2, etc.

Interesting Technical Stats:

- **LSTM Sequential Learning**: 5-lap temporal windows capturing tire degradation patterns
- **New Class Imbalance Handling**: 11.17x weight adjustment (pit stops are rare events!)
- **Interactive Analysis**: Live driver/race selection with prediction visualization





DEMO