

Modelo preditivo para estratégia de pneus na Fórmula 1

Rafael Russo¹

¹Faculdade de Computação e Informática – Universidade Presbiteriana Mackenzie (MACK)
São Paulo – SP – Brazil

10401195@mackenzista.com.br

Abstract. *In the motorsport industry, a race engineer's responsibilities include deciding on tire strategy. In a Formula 1 Grand Prix, drivers receive 13 sets of tires and must manage their use for the session and the race itself [Formula 1 2025a]. Free practice and qualifying generate telemetry data for the best car setup and choice of tire strategy. However, with the large amount of data coming from telemetry, in-depth analysis becomes very difficult. So, using a predictive model, it is possible to train with historical data and generate a strategy for each grand prix. This way, race engineers can make a more informed choice from the millions of pieces of data generated by telemetry.*

Resumo. *No ambiente de automobilismo, um engenheiro de corrida tem, entre suas responsabilidades, a responsabilidade da decisão de estratégia de pneus. Em um Grande Prêmio da Fórmula 1, pilotos recebem 13 sets de pneus e devem administrar o uso para sessão e a corrida em si [Formula 1 2025a]. Os treinos livres e classificatórios geram dados de telemetria para melhor setup do carro e escolha de estratégias de pneus. Porém, com o alto número de dados provenientes da telemetria, uma análise mais aprofundada se torna muito difícil de ser feita. Assim, utilizando um modelo preditivo, é possível que ele faça um treinamento com dados históricos e gere uma estratégia para cada grande prêmio. Assim, engenheiros de corrida conseguem fazer uma escolha mais fundada nos milhões de dados gerados pela telemetria.*

1. Introdução

Um fim de semana normal para competições de Fórmula 1 é composto por três treinos livres, um treino classificatório e a corrida. Alguns grandes prêmios têm essa organização diferente para acomodar as corridas Sprints, sendo então um treino livre, dois treinos classificatórios, uma corrida Sprint e a corrida principal [Formula 1 2025b]. Todos os treinos e corridas geram dados de telemetria para que times possam melhorar o setup do carro, entender falhas e de modo geral ter um melhor desempenho. Engenheiros de corrida utilizam esses dados para por exemplo ajudar o piloto durante corridas a entender pontos de melhoria em uma volta. Assim, o ideal é que os pilotos testem os três compostos disponíveis para ter um melhor entendimento de como cada um performa na pista e estratégia de pneus ser feita adequadamente.

Porém, pela grande quantidade de dados de telemetria que são gerados, para que uma análise tome em conta tudo registrado, demandaria muito tempo. Além do mais, levar em conta dados históricos de forma manual deixaria essa análise exaustiva. Assim, a utilização de modelos preditivos se torna muito interessante, já que consegue levar em

conta tanto os dados históricos quanto os atuais e é rápida. Com uma ferramenta assim disponível para o engenheiro de corrida, suas decisões agora levam em conta todos esses dados de telemetria.

Assim, o objetivo do projeto é produzir um modelo preditivo (Opção Framework) capaz de observar os dados de telemetria das sessões de treino livre e classificatório para prever a melhor estratégia de pneus.

2. Descrição do Problema

Carros de Formula 1, em média, geram de 60 a 90 Mb de dados de telemetria por volta [Mercedes-AMG Petronas Formula One Team 2022]. Essa enorme quantidade de informações é composta por variáveis que vão desde os dados de desempenho do motor e do sistema de freios até as nuances do comportamento do pneu em diferentes condições de pista e clima. Processar e analisar esse volume de dados manualmente é inviável, o que torna a tomada de decisão baseada em análises tradicionais lenta e sujeita a erros.

O desafio reside em desenvolver um modelo preditivo capaz de integrar tanto os dados históricos quanto os dados gerados nas sessões de treino livre e classificatório. Esse modelo deve identificar padrões e correlações complexas que influenciam a performance dos pneus, considerando aspectos como clima e tempos de volta. A implementação dessa solução requer a utilização de técnicas avançadas de redes neurais recorrentes, especificamente LSTMs ou GRUs, que possibilitem a análise de grandes volumes de informações como visto por [Rondelli 2023].

3. Aspecto Ético

Os aspectos éticos no desenvolvimento deste modelo preditivo devem ser considerados com ênfase na função de suporte que ele oferece, e não como um substituto do engenheiro de corrida. O modelo será concebido para fornecer subsídios na definição das estratégias de pneus, agregando informações valiosas a partir dos dados históricos, mas a decisão final sempre deverá ser do engenheiro ou piloto.

4. Dataset

O dataset contém dados de todas as voltas de treinos livres, classificatórios e corridas do ano de 2019, provenientes da API FastF1. O formato do DataFrame é (58947, 35). A análise exploratória foi realizada no notebook `Entregas_IA/1_Bimestre/dataset.ipynb`.

O dataset contém dados de todas as voltas de treinos livres, classificatórios e corridas da temporada de 2019 de Fórmula 1. Os dados foram obtidos via API FastF1 [theOehrly 2025]. O DataFrame possui 58947 linhas e 35 colunas. A análise exploratória foi realizada no notebook `Entregas_IA/1_Bimestre/dataset.ipynb`.

- **Time** (`pandas.Timedelta`): Tempo da sessão em que o tempo da volta foi registrado (final da volta).
- **Driver** (`str`): Identificador do piloto (três letras).
- **DriverNumber** (`str`): Número do piloto.
- **LapTime** (`pandas.Timedelta`): Tempo registrado da volta. Para verificar se foi deletado, consulte a coluna `Deleted`.

- **LapNumber** (float): Número da volta registrada.
- **Stint** (float): Número do stint (período contínuo com o mesmo conjunto de pneus).
- **PitOutTime** (pandas.Timedelta): Tempo da sessão em que o carro saiu do pitlane.
- **PitInTime** (pandas.Timedelta): Tempo da sessão em que o carro entrou no pitlane.
- **Sector1Time, Sector2Time, Sector3Time** (pandas.Timedelta): Tempos registrados nos respectivos setores da volta.
- **Sector1SessionTime, Sector2SessionTime, Sector3SessionTime** (pandas.Timedelta): Tempo da sessão em que cada setor foi registrado.
- **SpeedI1, SpeedI2** (float): Velocidades nos speedtraps dos setores 1 e 2 (km/h).
- **SpeedFL** (float): Velocidade na linha de chegada (km/h).
- **SpeedST** (float): Velocidade na reta mais longa (pode ser incerta).
- **IsPersonalBest** (bool): Indica se é a melhor volta pessoal oficial do piloto.
- **Compound** (str): Composto de pneus (SOFT, MEDIUM, HARD, INTERMEDIATE, WET, etc).
- **TyreLife** (float): Número de voltas dirigidas com esse pneu (inclui voltas de outras sessões).
- **FreshTyre** (bool): Indica se o pneu era novo no início do stint (TyreLife = 0).
- **Team** (str): Nome da equipe.
- **LapStartTime** (pandas.Timedelta): Tempo da sessão no início da volta.
- **LapStartDate** (pandas.Timestamp): Timestamp (data e hora) no início da volta.
- **TrackStatus** (str): Status da pista.
- **Position** (float): Posição do piloto ao final da volta (pode ser NaN).
- **Deleted** (Optional[bool]): Indica se a volta foi deletada (ex: violação de limites da pista).
- **DeletedReason** (str): Motivo da exclusão da volta (quando disponível).
- **FastF1Generated** (bool): Indica se os dados foram gerados/interpolados pela FastF1.
- **IsAccurate** (bool): Indica se os tempos estão sincronizados corretamente (não garante precisão absoluta).
- **session_type** (str): Tipo de sessão (treino, classificação, corrida).
- **Year** (int): Ano do evento.
- **EventName** (str): Nome do evento (ex: Grande Prêmio do Brasil).
- **Round** (int): Número da rodada na temporada.

5. Metodologia e Resultados Esperados

O projeto utilizará dados históricos de telemetria de corridas (como tempos de volta, estratégia de pneus, stint, velocidades, status da pista e condições de sessão) para treinar um modelo preditivo baseado em redes neurais recorrentes (LSTM/GRU). O uso de telemetria possibilita a captura de padrões temporais e sequenciais, como visto em [Krishnan et al. 2024].

Primeiro, deverá ser feita a extração de múltiplos dados de eventos e anos e tratar valores ausentes, normalizar variáveis e tratar features categorias como o composto

sendo usado. Somado a isso, deve-se criar novas features considerando médias de valores numéricos, número de pontos de frenagem e etc. para melhorar a performance do modelo, visto em [Cheteles 2024].

Com os dados tratados, eles serão divididos em treino, teste e validação para que seja feito uma análise imparcial das métricas do modelo. Assim, usando o dataset de treinamento, treinamos a rede neural para prever a melhor estratégia de pneus em dias de corrida considerando as outras sessões.

É esperado que o modelo consiga prever a estratégia ideal para as corridas considerando pontos como número de paradas e performance de cada pneu segundo os treinos anteriores.

Referências

Cheteles, B. O.-A. (2024). Feature importance versus feature selection in predictive modeling for formula 1 race standings.

Formula 1 (2025a). The beginner's guide to f1 tyres. <https://www.formula1.com/en/latest/article/the-beginners-guide-to-formula-1-tyres.61SvF0Kfg29UR2SPhakDqd>. Acessado: 2025-04-07.

Formula 1 (2025b). The beginner's guide to the f1 weekend. <https://www.formula1.com/en/latest/article/the-beginners-guide-to-the-formula-1-weekend.5RFZzGXNhEi9AEuMXwo987>. Acessado: 2025-04-07.

Krishnan, A., Noe, B., and Patel, A. (2024). Utilizing telemetry data for machine learning-driven lap time predictions and race strategies. University of Rochester, Project work.

Mercedes-AMG Petronas Formula One Team (2022). How much data does an f1 car generate? <https://www.youtube.com/watch?v=JDkePD2-scY>. Acessado: 2025-04-07.

Rondelli, M. (2023). The future of formula 1 racing: Neural networks to predict tyre strategy. Advisor: Elena Loli Piccolomini; Co-advisor: Davide Evangelista. Degree Program in Computer Science for Management (L-DM270).

theOehrly (2025). *FastF1 Documentation*. Acessado: 2025-04-07.