

Modelos de árvore de decisão para embasamento na escolha de estratégia de pneus na Fórmula 1

Rafael Russo

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

10401195@mackenzista.com.br

Abstract. *In Formula 1, tire strategy plays a critical role in a driver's performance and can be affected by variables such as the weather, the track and driving style. Deciding on the right compounds and the ideal time to change them is essential for minimizing the time lost in pit stops and maintaining a good race pace [Motorsport engineer]. This paper proposes the use of historical race data to train an ML model that helps predict tire strategies. Using data such as weather conditions, circuit characteristics and information from drivers and teams, the prediction of which compounds to use and for how many laps was tested. The results showed that simpler models without rigorous implementation have difficulty generalizing these strategies.*

Resumo. *Na Fórmula 1, a estratégia de pneus desempenha um papel crítico para o desempenho de um piloto, podendo ser afetada por variáveis como o clima, a pista e o estilo de pilotagem. A decisão correta dos compostos e o momento ideal para trocá-los é essencial para minimizar o tempo perdido nos pit stops e manter um bom ritmo de prova [Motorsport engineer]. Neste trabalho, propõe-se o uso de dados históricos de corridas para o treinamento de um modelo de ML que auxilia na predição de estratégias de pneus. A partir de dados como condições climáticas, características do circuito e informações dos pilotos e equipes, foi testado a predição de compostos a serem utilizados e por quantas voltas. Os resultados mostraram que modelos mais simples sem um rigor em sua implementação tem dificuldade em generalizar essas estratégias.*

1. Vídeo

Link do vídeo de apresentação: <https://youtu.be/tBIzlUFODeg>

2. Introdução

Em um Grande Prêmio (GP) da Fórmula 1, realizado ao longo do fim de semana e compreendendo três dias de trabalhos, acontecem 3 treinos livres, 1 treino classificatório e 1 corrida. A corrida conclui o GP, definindo um vencedor e a quantidade de pontos obtida pelos pilotos. Cada equipe recebe 13 jogos de pneus para pista seca, de compostos diferentes (duro, médio e macio); 4 jogos de pneus intermediários e 3 jogos de pneus de chuva [Pirelli 2025]. Apesar da maioria dos pneus serem utilizados durante as sessões de treino e de classificação, o planejamento da alocação é um fator determinante no resultado da prova. Durante a corrida cada piloto deve usar pelo menos dois compostos

diferentes e maximizar a vida útil de cada jogo para otimizar seu desempenho. Assim, existe a necessidade de escolha dos compostos com antecedência.

A definição dos compostos a serem usados é uma decisão estratégica, pois cada troca envolve um *pit stop* que implica no piloto perder posições devido ao tempo gasto, mas entrega condições ao piloto de ter voltas com tempos muito menores. Além disso, cada composto se comporta de maneira diferente: pneus macios têm alta aderência, mas têm uma vida útil pequena; pneus duros oferecem vida útil muito longa, mas pouca aderência e tempos de volta maiores; pneus médios se encontram no meio desse espectro e tentam atingir um compromisso entre desempenho e durabilidade [Pirelli 2025]. Essa estratégia deve levar em conta diversos fatores como a pista, posição de largada, condições climáticas, aderência dos diferentes compostos ao asfalto, estilo de pilotagem dentre outros quesitos, fatores que podem ser observados ou previstos.

Atualmente, a escolha de quais compostos devem ser usados pode ser embasada pelos dados de telemetria gerados pelos carros, mostrando, por exemplo, como o carro está reagindo à pista. Como visto em [Todd et al. 2025], o uso da telemetria de corridas passadas pode ajudar a estimar o desgaste dos pneus, algo essencial na decisão de estratégia. Entretanto, apesar do número massivo de dados gerados por cada carro, a maioria deles não está disponível ao público, já que isso revelaria possíveis segredos de cada equipe e piloto. No entanto, mesmo com informações limitadas, é possível extrair *insights* valiosos, como visto em [Rondelli 2023, Krishnan et al. 2024] que fazem uso de dados da *API Fast F1*, tanto de telemetria quanto de meteorologia, para prever estratégias de pneus.

Assim, o objetivo do projeto é produzir um modelo preditivo (Opção Framework) capaz de observar os dados gerais de clima e a corrida para prever a melhor estratégia de pneus.

3. Descrição do Problema

A definição dos compostos de pneus a serem usados em uma corrida de Fórmula 1 é uma decisão estratégica crucial, pois cada troca envolve um *pit stop* que implica no piloto perder posições devido ao tempo gasto, mas entrega condições ao piloto de ter voltas com tempos muito menores. Durante a corrida, cada piloto deve utilizar pelo menos dois compostos distintos [Parkes 2015] e maximizar a vida útil de cada jogo para otimizar seu desempenho. Cada composto se comporta de maneira diferente: pneus macios têm alta aderência, mas uma vida útil pequena; pneus duros oferecem vida útil muito longa, mas pouca aderência e tempos de volta maiores; pneus médios se encontram no meio desse espectro, tentando atingir um compromisso entre desempenho e durabilidade [Pirelli 2025]. Um *pit stop* típico demora entre 20 e 30 segundos, enquanto pneus novos podem reduzir os tempos de volta em 0.5 a 1.5 segundos, tornando a decisão do *pit stop* crítica para que o ganho de performance supere o tempo perdido [Motorsport engineer].

Essa estratégia deve levar em conta diversos fatores como a pista, posição de largada, condições climáticas, aderência dos diferentes compostos ao asfalto, estilo de pilotagem, dentre outros quesitos que podem ser observados ou previstos. O uso da telemetria de corridas passadas, por exemplo, pode ajudar a estimar o desgaste dos pneus, algo essencial na decisão de estratégia [Todd et al. 2025]. Mesmo com informações limitadas, é possível extrair *insights* valiosos, como demonstrado em trabalhos que utilizam dados da

API pública *Fast F1* para fins de análise e previsão em Fórmula 1 [Krishnan et al. 2024, Rondelli 2023, Cheteles 2024].

Trabalhos como os de [Krishnan et al. 2024, Todd et al. 2025] tentam generalizar a estratégia de pneus por meio de aprendizagem de máquina. Modelos de árvore de decisão, como o *Random Forest*, são usados em [Krishnan et al. 2024] para prever tempos de volta e esquemas de *pit stop*. Modelos de árvore de decisão também se mostraram equiparáveis a outras abordagens, como visto em [Todd et al. 2025], onde o *XGBoost* foi o melhor modelo testado para estimar o desgaste de pneus com boa precisão. Assim, o objetivo de tais modelos é auxiliar os pilotos e engenheiros de corrida na elaboração de combinações de compostos, aproveitando os dados históricos de outras corridas, algo difícil de ser feito sem o uso de aprendizado de máquina.

4. 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.

5. Dataset

O conjunto de dados utilizado para este projeto é composto por todos os *stints* de corrida das temporadas de Fórmula 1 compreendidas entre os anos de 2019 e 2024 obtidos na *API Fast F1*. Esse intervalo de anos foi escolhido, pois corridas anteriores a 2019 usam compostos diferentes dos permitidos no regulamento atual. Assim, foi obtido 7.095 linhas, onde cada uma representa um *stint* individual de um piloto. Para cada *stint*, são registradas informações como o número sequencial do *stint* daquele piloto na corrida, o composto de pneu utilizado (duro, médio, macio, intermediário e chuva) e o número de voltas que o piloto permaneceu com ele. Além desses dados específicos do *stint*, cada linha contém informações contextuais mais amplas, incluindo o nome do piloto, o circuito onde a corrida ocorreu, o ano, a posição de largada do piloto, sua equipe e diversas variáveis climáticas pertinentes ao evento.

Como o objetivo do projeto é prever toda a estratégia de pneus de um piloto para uma corrida, foi necessário realizar uma transformação no *dataset* inicial. A estrutura dos dados foi modificada para que cada linha correspondesse a uma corrida completa de um piloto específico em um determinado circuito e ano. Essa reorganização permite treinar modelos capazes de aprender e prever as sequências de compostos e a duração de seus usos, ou seja, as estratégias de pneus. Como parte do tratamento de dados, as entradas referentes a *stints* com o composto identificado como “UNKNOWN” foram removidas. Essa exclusão não comprometeu a robustez da amostra, pois representavam um volume muito pequeno de casos. Adicionalmente, os nomes das pistas foram padronizados para conterem apenas o nome do circuito, eliminando assim pequenas diferenças de nomenclatura que poderiam surgir devido ao nome do evento específico de cada ano e garantindo maior consistência nos dados. Além disso, corridas com uma estratégia de apenas um composto foram removidas por não cumprirem a regra de pelo menos dois compostos diferentes.

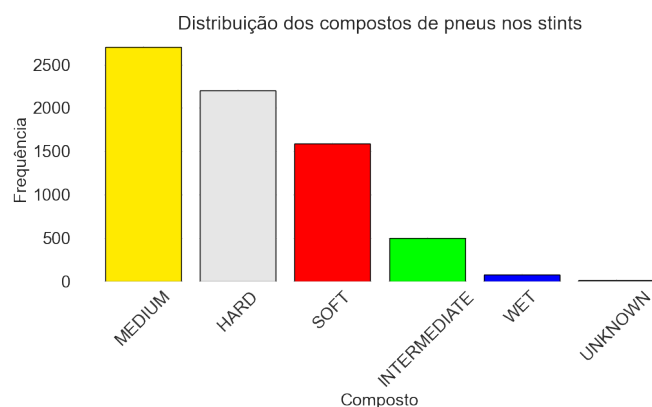


Figura 1. Histograma da frequência de uso de cada composto.

Observando a figura 1, o dataset final obteve poucas amostras de corridas com chuva, algo esperado. Assim, esses registros foram removidos e os únicos compostos agora são duro, médio e macio.

6. Metodologia e Resultados Esperados

Este estudo caracteriza-se como uma pesquisa quantitativa, de caráter empírico e reprodutível, baseada em dados históricos de corridas de Fórmula 1. O caminho metodológico adotado para o trabalho contempla as seguintes etapas:

1. **Revisão da literatura** para identificação das técnicas de ML adequadas ao tipo de problema e seu contexto. A pesquisa deve contemplar pontos não explorados envolvendo o problema e diferentes tipos de abordagem;
2. **Definição da base de dados** a ser utilizada para o treinamento do modelo;
3. **Análise e seleção de características** a partir da base de dados extraída. Deve ser feita uma análise exploratória para compreender possíveis desbalanços no *dataset* e outros tipos de inconsistências. Elas então devem ser corrigidas para tornar a amostra de dados imparcial e balanceada;
4. **Treinamento do modelo** levantado na revisão da literatura como promissor para o problema em questão;
5. **Avaliação de desempenho** do modelo calculando métricas pertinentes;

Com fins de simplificar o modelo, as estratégias com mais de quatro pit stops foram removidas, critério adotado para estabelecer um escopo de análise consistente.

Para este projeto, utilizando os dados tratados da FastF1, foi desenvolvido uma arquitetura de ensemble com *Random Forest*. Foram implementados dois modelos independentes de *RF* com o auxílio da classe *MultiOutputClassifier*, da biblioteca *scikit-learn*. Essa classe permite treinar múltiplos classificadores de forma paralela, tratando cada saída do modelo como uma tarefa separada. Ela foi utilizada para prever simultaneamente cinco compostos de pneus e cinco durações de stint, sendo cada composto e stint uma saída independente. O primeiro modelo prevê um vetor de quatro compostos, podendo conter valores nulos caso o número de pit stops seja menor. O segundo estima o tamanho dos stints também em um vetor de quatro valores.

Devido à independência entre os dois modelos, esse arranjo pode gerar inconsistências, como uma divergência entre o número de compostos previstos e o número de

stints estimados, o que compromete a coerência da estratégia simulada. Para as predições de compostos foi avaliada a acurácia, precisão, recall e F1 Score, enquanto para as predições de tamanho de stints foi calculado o MAE, MSE, RMSE e R2 Score.

7. Resultados

Com o modelo treinado, foi utilizado a partição dos dados de validação para obter as métricas definidas. Os resultados foram:

Tabela 1. Métricas de Desempenho do Modelo

Acurácia	Precisão	Recall	F1-Score	MAE	MSE	RMSE	R ² Score
0.763	0.559	0.555	0.554	4.699	64.312	8.019	0.371

A princípio as métricas calculadas não aparentam estar tão ruins, porém essas métricas não conseguem fornecer detalhes o suficiente de cada composto. Assim, foi calculado as métricas de classificação de forma individual para cada composto.

Com uma visualização mais detalhada da performance do modelo, tornou-se clara a explicação das métricas gerais estarem boas: os primeiros compostos são onde residem a maior parte dos dados, enquanto os últimos compostos possuem a menor quantidade de dados. Assim, o desempenho para um modelo prever o primeiro composto é extremamente bom, mas para um quinto composto ele é quase inexistente.

Tabela 2. Relatório de Classificação para Composto_1

Classe	Precisão	Revocação	Pontuação F1	Suporte
HARD	0.25	0.36	0.29	28
MEDIUM	0.83	0.74	0.78	266
SOFT	0.70	0.78	0.74	124
Accuracy			0.73	418

Tabela 3. Relatório de Classificação para Composto_2

Classe	Precisão	Revocação	Pontuação F1	Suporte
HARD	0.82	0.76	0.79	276
MEDIUM	0.46	0.59	0.52	109
SOFT	0.48	0.36	0.41	33
Accuracy			0.68	418

Tabela 4. Relatório de Classificação para Composto.3

Classe	Precisão	Revocação	Pontuação F1	Suporte
HARD	0.60	0.62	0.61	107
MEDIUM	0.31	0.37	0.34	62
NONE	0.72	0.69	0.70	192
SOFT	0.40	0.35	0.37	57
Accuracy			0.58	418

Tabela 5. Relatório de Classificação para Composto.4

Classe	Precisão	Revocação	Pontuação F1	Suporte
HARD	0.53	0.53	0.53	15
MEDIUM	0.50	0.35	0.41	20
NONE	0.92	0.95	0.93	351
SOFT	0.44	0.38	0.41	32
Accuracy			0.86	418

Tabela 6. Relatório de Classificação para Composto.5

Classe	Precisão	Revocação	Pontuação F1	Suporte
MEDIUM	0.00	0.00	0.00	2
NONE	0.98	0.98	0.98	404
SOFT	0.54	0.58	0.56	12
Accuracy			0.96	418

8. Conclusão

O problema da escolha de estratégia na Fórmula 1 é de crucial importância, e a busca por otimizar decisões e obter vantagens sobre os adversários impulsiona uma pesquisa contínua. É evidente que este não é um problema simples, tanto que as abordagens descritas na literatura frequentemente envolvem soluções complexas, como o uso de redes neurais recorrentes.

A tentativa de utilizar modelos de *Random Forest* de forma independente para prever o desempenho ou as características associadas a estratégias de pneus em cenários de Fórmula 1 apresenta limitações consideráveis. A principal barreira identificada reside na escassez de dados históricos que correlacionem cada composto a cada *stints* durante uma corrida. Muitos *stints*, a depender do composto utilizado, são necessariamente mais curtos, pois cada *pit stop* acarreta uma perda de tempo significativa, o que pode, por sua vez, dificultar a previsão precisa de estratégias que envolvem *stints* mais longos ou um número reduzido de paradas.

Assim, arquiteturas de modelos que permitam uma influência mútua e uma análise conjunta entre os diferentes compostos de pneus são essenciais. Essa abordagem integrada é vital, pois permite que o conhecimento adquirido sobre o comportamento de um tipo de pneu, mesmo que com base em dados mais abundantes, possa informar e refinar as previsões para outros compostos que possuam dados históricos mais limitados, refletindo de forma mais fiel as complexas interdependências das estratégias de corrida.

Referências

- Cheteles, O.-A. (2024). Feature Importance versus Feature Selection in Predictive Modeling for Formula 1 Race Standings. Publisher: University of Twente.
- Krishnan, A., Noe, B., and Patel, A. (2024). Utilizing Telemetry Data for Machine Learning-Driven Lap Time Predictions and Race Strategies.
- Motorsport engineer. How Pit Stop Strategy Leads To Victories In Formula 1.
- Parkes, I. (2015). How the 2016 Formula 1 tyre rules will work.
- Pirelli (2025). Beginner's guide to F1 tyres.
- Rondelli, M. (2023). *The Future of Formula 1 Racing: Neural Networks to Predict Tyre Strategy*. Tesi di laurea, Università di Bologna.
- Todd, J., Jiang, J., Russo, A., Winkler, S., Sale, S., McMillan, J., and Rago, A. (2025). Explainable Time Series Prediction of Tyre Energy in Formula One Race Strategy. arXiv:2501.04067.