
PROPOSTA DE ORIENTAÇÃO DE TCC I

Aluno: Marcos Paulo de Oliveira Pereira

Matrícula: 21.1.1090

Orientador/Departamento: Prof. Alexandre Magno de Sousa – DECSI

Coorientador/Departamento:

Disciplina: CSI495 – Trabalho de Conclusão de Curso 1

Título Provisório: Análise Preditiva e Otimização de Estratégias na Fórmula 1 Utilizando Simulação de Monte Carlo e Aprendizado de Máquina

Área de pesquisa:

Código	Nome
1.02.02.08-0	Análise de Dados
1.03.03.04-9	Sistemas de Informação

1. Introdução

A estratégia na Fórmula 1 moderna é um problema de otimização multivariável executado sob alta incerteza. Fatores como o desgaste dos pneus, condições climáticas e eventos estocásticos, como a entrada do *Safety Car*, criam um cenário de alta complexidade onde decisões ótimas precisam ser tomadas em tempo real [3].

Este trabalho teve como ponto de partida o desenvolvimento de uma aplicação semelhante na disciplina de Projeto e Análise de Experimentos (PAE), que utilizou análise exploratória de dados e modelos de regressão linear para criar um “Strategy Score”. Esse Strategy Score seria uma métrica quantitativa para determinar a qualidade de uma estratégia, se baseando em posições ganhas e posição de chegada. Conforme detalhado na apresentação final do projeto, essa abordagem inicial, embora útil para identificar variáveis-chave, possui uma natureza determinística e estática, sendo incapaz de simular o desenrolar dinâmico e probabilístico de uma corrida.

A literatura recente explora soluções mais sofisticadas para este desafio, como Redes Neurais Profundas (EDNN) [3], Aprendizado por Reforço (RL) [8] e algoritmos de Monte Carlo Tree Search (MCTS) [6]. Este TCC, portanto, surge como a evolução natural da análise inicial de PAE, posicionando-se de forma distinta em relação à literatura. O objetivo é desenvolver um modelo híbrido e computacionalmente eficiente, que integra a simulação de Monte Carlo com modelos de aprendizado de máquina. O diferencial deste trabalho reside em focar na simulação e otimização de estratégias, utilizando exclusivamente ferramentas *open-source* (Python, FastF1, Scikit-learn), para demonstrar que *insights* estratégicos valiosos podem ser gerados sem a necessidade de infraestruturas computacionais massivas ou conjuntos de dados privados.

2. Motivação e Justificativa

A investigação deste tema fundamenta-se nos seguintes pilares:

1. **Complexidade e Relevância do Domínio:** A estratégia de corrida na Fórmula 1 é um fator crítico que pode determinar o resultado das competições, sendo modelada na literatura até mesmo como um problema de teoria dos jogos para otimizar duelos diretos na pista [1]. As equipes investem recursos consideráveis em simulação e análise de dados para obter vantagens competitivas, o que valida a relevância acadêmica e prática de se pesquisar e desenvolver ferramentas de otimização para este fim , a literatura atual justifica a utilização desse tipo de simulação para tornar os resultados mais realistas, modelando as incertezas e avaliando a robustez do problema [4]. A fórmula 1 é um esporte de impacto também na vida externa, tendo papel importante no desenvolvimento de motores para carros de rua e em combustíveis bio-degradáveis;
2. **Limitação de Modelos Estáticos:** A investigação inicial, realizada na disciplina de Projeto e Análise de Experimentos, demonstrou que modelos determinísticos como a regressão linear são insuficientes para capturar a natureza dinâmica e as incertezas de uma corrida. Modelos como redes neurais apresentam desempenho superior a esse tipo de abordagem [9] até mesmo em uma simples previsão de tempo de volta. Essa limitação é corroborada pela literatura, que aponta que corridas são frequentemente decididas por eventos estocásticos, como acidentes e safety cars, exigindo abordagens probabilísticas para uma análise robusta [4]. Além disso, trabalhos recentes mostram que a simples previsão de variáveis isoladas, como o tempo de volta, não se traduz diretamente em uma previsão acurada do resultado final [5], o que evidencia a necessidade de simular a interação complexa entre os eventos de corrida. Essa limitação cria uma clara necessidade de evoluir para abordagens mais robustas, que, através de simulações, possam gerar um conjunto de dados mais robusto e que será explorado por um modelo adequado;
3. **Oportunidade de Inovação Acessível:** A literatura recente confirma a viabilidade de técnicas avançadas para o problema, como a aplicação de Aprendizado por Reforço (RL) para otimização em tempo real [8]. No entanto, tais abordagens demandam alto custo computacional e possuem elevada complexidade de implementação, criando uma lacuna para soluções que sejam não apenas eficazes, mas também acessíveis. Este trabalho se justifica ao propor um modelo híbrido que busca democratizar a análise de dados esportivos avançada, utilizando ferramentas de código aberto e oferecendo uma alternativa viável à complexidade dos sistemas de RL ou proprietários.

3. Definição do Problema

O problema central consiste em desenvolver um modelo capaz de simular o resultado de uma corrida de Fórmula 1 a partir de diferentes estratégias de pit stop, considerando as incertezas inerentes ao esporte. Para tal, o modelo deverá ser alimentado por um conjunto de dados públicos e ser capaz de avaliar quantitativamente qual estratégia oferece a maior probabilidade de sucesso.

Os dados para este trabalho serão extraídos por meio da API *FastF1*, uma biblioteca Python de código aberto que fornece acesso a dados detalhados de sessões de Fórmula 1. As principais variáveis disponíveis e que serão consideradas incluem:

- **Dados de Telemetria por Volta:** Tempo de volta, tempos de setor, velocidade máxima;
- **Dados dos Pneus:** Composto do pneu (macio, médio, duro), idade do pneu (número de voltas), composto da pirelli (C1, C2, C3, C4, C5, C6);
- **Dados da Corrida:** Posição no grid, posição a cada volta, informações sobre pit stops (volta da parada, duração), informações do controle de prova;
- **Dados de Clima e Temperatura:** Temperatura de pista, temperatura dor ar, umidade, pressão do ar, probabilidade de chuva;
- **Eventos Estocásticos:** Períodos de *Safety Car* (real ou virtual).

Existem restrições importantes a serem consideradas. A solução deve ser computacionalmente eficiente, projetada para ser executada em hardware de consumidor padrão, sem depender de clusters de alta performance. Além disso, o modelo se restringirá aos dados publicamente disponíveis, o que exclui telemetrias internas e dados em tempo real que as equipes possuem para a tomada dessas decisões normalmente.

Diante deste contexto, o problema de pesquisa que este trabalho busca responder é:

Como a integração da ideia da simulação de Monte Carlo com modelos de aprendizado de máquina pode aprimorar a análise e a tomada de decisão estratégica na Fórmula 1, oferecendo uma alternativa computacionalmente acessível às soluções de grande escala utilizadas pelas equipes?

4. Objetivos Geral e Específicos

O objetivo geral desse trabalho é o desenvolvimento e a validação de um modelo computacional híbrido que integra simulação de Monte Carlo e aprendizado de máquina, visando a análise e otimização de estratégias na Fórmula 1 de forma computacionalmente eficiente e com o uso de dados públicos, através do *FastF1*. Para que o objetivo geral seja alcançado, os seguintes objetivos específicos deverão ser cumpridos:

- Estudar e aplicar técnicas para a coleta, tratamento e análise de dados de corridas de Fórmula 1, utilizando a biblioteca *FastF1*;
- Desenvolver modelos de aprendizado de máquina para estimar variáveis dinâmicas da corrida, como a degradação dos pneus e os tempos de volta esperados, que servirão como parâmetros para o simulador;
- Construir um simulador de corridas baseado no método de Monte Carlo, capaz de projetar múltiplos resultados possíveis a partir de diferentes sequências de decisões estratégicas;
- Implementar uma métrica quantitativa para avaliar e comparar a eficácia das estratégias simuladas, aprimorando o conceito de “*Strategy Score*”;
- Realizar a validação do modelo, comparando as estratégias sugeridas pela ferramenta com os resultados reais de corridas passadas para aferir sua acurácia e potencial.

5. Trabalhos Relacionados

Diversos trabalhos têm buscado abordar o problema de otimização de estratégias na Fórmula 1. A Tabela 1 apresenta um resumo de contribuições relevantes, evidenciando suas principais características e limitações.

Observa-se na literatura que, enquanto os trabalhos com modelos preditivos e de aprendizado por reforço avançam no campo da otimização, ainda há uma lacuna em soluções que sejam simultaneamente eficazes e acessíveis. Este TCC parte da limitação identificada no trabalho anterior do autor (PAE), cuja abordagem de regressão linear se mostrou insuficiente para modelar fatores não-lineares e a alta estocasticidade de eventos de corrida. A proposta deste trabalho busca, portanto, preencher essas lacunas por meio de uma abordagem híbrida que combina a simulação de Monte Carlo com modelos de aprendizado de máquina mais robustos, como o *Gradient Boosting* ou *Random Forest*.

Tabela 1: Resumo dos trabalhos relacionados à análise e otimização de estratégias na Fórmula 1.

Trabalho	Contribuição Principal	Limitação / Diferença
Fatima e Johrendt [3]	Propõe uma Rede Neural (EDNN) para prever o ranking final e a volta ótima do pit stop.	Foco em previsão, não em otimização de múltiplos cenários. Alto custo computacional.
Thomas et al. [8]	Aplica Aprendizado por Reforço (RL) para otimização de estratégia em tempo real, com foco em explicabilidade.	Altíssima complexidade e custo computacional, sendo pouco acessível para pesquisa independente.
Aguad e Thraves [1]	Usa Teoria dos Jogos para modelar a estratégia como um duelo 1x1, otimizando decisões relativas a um oponente.	Modelo complexo para escalar para um grid completo; não aborda a incerteza geral da corrida (acidentes, clima).
Piccinotti (2021) [6]	Introduz um planejador com Monte Carlo Tree Search (MCTS) para otimização de estratégia.	Exige simuladores dedicados e possui limitações de escalabilidade com a profundidade da busca.
Heilmeier et al. [4]	Valida o uso de Monte Carlo para modelar incertezas (Safety Car) e avaliar a robustez das estratégias.	Foca na simulação, mas não usa ML para prever o desempenho dinâmico (ex: degradação de pneus).
Zhao (2024) [9]	Usa uma Rede Neural Profunda (DNN) para a tarefa específica de prever o tempo de volta na qualificação.	Foco muito restrito na previsão de uma única métrica; não simula a corrida ou a interação estratégica.
Bansal et al. [2]	Usa Gradient Boosting para prever o resultado da corrida, mostrando a posição de largada como fator dominante.	Abordagem estática e puramente preditiva; não simula nem otimiza a estratégia para alterar o resultado.

Quiroga García [7]	Desenvolve um modelo de otimização determinístico (MIQP) para achar a estratégia ótima em um cenário perfeito.	Ignora eventos aleatórios, resultando em uma solução que pode ser muito arriscada em um cenário real.
Jafri (2024) [5]	Usa Redes LSTM para prever resultados, concluindo que prever o tempo de volta não é suficiente para prever o final.	Foco em previsão. A própria conclusão reforça a necessidade de simular a interação estratégica.
Projeto de PAE (Autor)	Propôs um <i>Strategy Score</i> inicial com regressão linear para identificar variáveis-chave de desempenho.	Modelo determinístico e estatístico, incapaz de simular cenários probabilísticos e incertezas da corrida.

6. Cronograma de Atividades

O desenvolvimento do Trabalho de Conclusão de Curso I seguirá o cronograma de atividades apresentado na Tabela 2.

Tabela 2: Cronograma de Atividades - TCC I.

Etapa da Metodologia	10/25	11/25	12/25	01/26	02/26	03/26
1. Revisão Sistemática da Literatura	✓	✓				
2. Coleta e Pré-processamento dos Dados	✓	✓				
3. Modelagem e Engenharia de Atributos		✓	✓			
4. Treinamento e Validação dos Modelos			✓	✓		
5. Desenvolvimento do Simulador			✓	✓	✓	
6. Execução das Simulações e Análise					✓	✓
7. Escrita da Monografia			✓	✓	✓	✓
8. Apresentação do TCC I nos seminários						✓

Referências

- [1] Felipe Aguad and Charles Thraves. Optimizing pit stop strategies in formula 1 with dynamic programming and game theory. *European Journal of Operational Research*, 319(3):908–919, 2024.
- [2] Aayam Bansal, Aadit Arora, Lakshay Bhati, Kushagra Sethia, Ishani Verma, and Naisha Kapoor. Advanced machine learning approaches for formula 1 race performance prediction: A comprehensive analysis of championship point forecasting. *ResearchGate*, July 2025. Preprint.
- [3] Syeda Sitara Wishal Fatima and Jennifer Johrendt. Deep-Racing: An Embedded Deep Neural Network (EDNN) Model to Predict the Winning Strategy in Formula One Racing. *International Journal of Machine Learning*, 13(3):97–103, July 2023.
- [4] Alexander Heilmeier, Michael Graf, Johannes Betz, and Markus Lienkamp. Application of monte carlo methods to consider probabilistic effects in a race simulation for circuit motorsport. *Applied Sciences*, 10(12):4229, 2020.

- [5] Ali Jafri. Predicting formula 1 race outcomes: A machine learning approach. Technical report, New York University Abu Dhabi, 2024. Capstone Project.
- [6] Diego Piccinotti. Open loop planning for formula 1 race strategy identification. Master's thesis, Politecnico di Milano, Milan, Italy, April 2021. Master of Science in Computer Science and Engineering.
- [7] Mariana Paz Quiroga García. Optimization of pit stop strategies in formula 1 racing: a data-driven approach. Master's thesis, Universidad de Chile, Facultad de Ciencias Físicas y Matemáticas, 2024.
- [8] Devin Thomas, Junqi Jiang, Avinash Kori, Aaron Russo, Steffen Winkler, Stuart Sale, Joseph McMillan, Francesco Belardinelli, and Antonio Rago. Explainable reinforcement learning for formula one race strategy. *arXiv preprint arXiv:2501.04068*, 2025. To be published in SAC 2025.
- [9] Zhixuan Zhao. Deep neural network-based lap time forecasting of formula 1 racing. In *Proceedings of the 4th International Conference on Signal Processing and Machine Learning*, 2024.