

PREDIÇÃO DO RITMO DE PILOTOS NA *FORMULA 1* POR ANÁLISE DE PADRÕES COM REDES NEURAIS

Maria Eduarda de Oliveira Furini¹, Lucas Marim da Silva², Nykolas Fornaziero dos Santos³

¹Centro Universitário das Faculdades Associadas de Ensino - UNIFAE, São João da Boa Vista - SP, Brasil (maria.furini@sou.fae.br)

^{2,3}Centro Universitário das Faculdades Associadas de Ensino - UNIFAE, São João da Boa Vista - SP, Brasil

Resumo: Este estudo apresenta o desenvolvimento e a avaliação de uma arquitetura para o processamento e a análise de séries temporais da telemetria de *Formula 1*. Para isso, contou-se com a *API OpenF1* como fonte de dados da temporada de 2025. A pesquisa focou na construção de uma Rede Convolucional 1D, implementada em framework *PyTorch*, resultando em 18 modelos. Para a pista de Suzuka, o *Mean Absolut Error* ficou abaixo de 0,25s e 96,5% das previsões dentro de 1 segundo do real.

Palavras-chave: Análise de Padrões; *Formula 1*; Previsão de Tempo de Volta; Redes Neurais Convolucionais; Telemetria

INTRODUÇÃO

A *Formula 1* é reconhecida como um dos ambientes mais avançados tecnologicamente, sendo marcada por inovações, tanto na engenharia aplicada nos carros, quanto na ciência dos dados. Embora a visibilidade esteja nas ultrapassagens, a competitividade não se limita ao desempenho apresentado nas pistas, sendo estendida à capacidade das equipes em coletar, processar e analisar um grande volume de informações ao longo de todas as sessões que compreendem um final de semana de corrida. Nesse cenário, a análise de dados torna-se fator determinante para o sucesso de pilotos e equipes ao longo da temporada (Preosti, 2021).

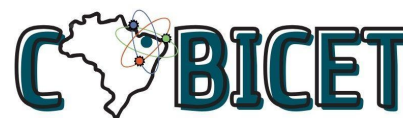
O ambiente da *Formula 1* pode ser considerado altamente dinâmico e complexo, sendo influenciado por múltiplas variáveis, como desgaste de pneus, condições climáticas, estratégia de *pit stop* e interações entre pilotos, como ultrapassagens, incidentes, entre outras. Além disso, eventos imprevisíveis, como *safety car*, bandeiras vermelhas, mudanças das condições de pista, exigem que decisões estratégicas tenham de ser realizadas ao longo da corrida (Mistry e Smith, 2025). Tais características tornam as tomadas de decisão processos desafiadores, dependendo tanto da análise de dados quanto da capacidade de adaptação em um contexto em que cada décimo de segundo conta.

É possível notar, no entanto, um crescimento exponencial na utilização de técnicas de inteligência artificial e aprendizado de máquina no contexto da análise de dados, especialmente em cenários

complexos e com grande volume de variáveis e informações. Tais abordagens permitem identificar padrões não lineares e realizar previsões mais robustas, superando limitações e métodos tradicionais (Bansal et al., 2025). Entretanto, sua aplicação na *Formula 1* apresenta desafios, especialmente no que diz respeito ao processamento e integração de dados dinâmicos e à adaptação em tempo real.

Assim sendo, o desenvolvimento de ferramentas que auxiliem na tomada de decisão pode significar uma vantagem relevante em um ambiente tão competitivo. Vale ressaltar que a aplicação desses modelos não se limita ao uso por equipes, podendo também contribuir para analistas, comentaristas e fãs, ampliando a compreensão sobre as decisões estratégicas ao longo das corridas e o desempenho dos pilotos (Jafri, 2024). Nesse sentido, deve-se evidenciar que o uso de tais modelos não tem como objetivo substituir a decisão humana, mas atuar como suporte analítico, fornecendo informações que auxiliem engenheiros e estrategistas na avaliação de diferentes cenários. Ademais, fatores como intuição, experiência e a capacidade de assumir riscos são atributos humanos relevantes, tornando a integração entre análise preditiva e decisão humana, elemento central no contexto estratégico.

Segundo Provost e Fawcett (2013), a ciência de dados permite compreender determinado panorama sob uma perspectiva baseada em estatística, auxiliando na tomada de decisões. A efetividade da análise e processamento de dados está diretamente



relacionada à engenharia de dados, responsável pela coleta, armazenamento, integração e preparação de informações, garantindo qualidade e disponibilidade para uso em modelos preditivos (Belcic e Stryker, 2024). No contexto da *Formula 1*, a coleta de dados envolve a obtenção massiva de informações como tempo de volta, velocidade, aceleração, freio, rotações do motor por minuto e eventos de corrida, constituindo a base para o desenvolvimento de soluções preditivas aplicadas ao esporte.

As Redes Neurais Artificiais (*ANNs*, do inglês *Artificial Neural Networks*) buscam replicar o funcionamento biológico dos neurônios de forma computacional e simplificada (Haykin, 2007). Diferentemente de sistemas tradicionais, que dependem de regras previamente definidas, as redes neurais aprendem padrões diretamente a partir dos dados, sendo capazes de adaptarem seu comportamento conforme as informações são apresentadas em uma etapa de treinamento (Haykin, 2007).

As Redes Neurais Convolucionais (*CNNs*, do inglês *Convolutional Neural Networks*) são uma classe especializada de redes neurais eficazes no reconhecimento de padrões locais em dados estruturados. Embora originalmente desenvolvidas para o processamento de imagens bidimensionais (*Conv2D*), em que o filtro percorre as duas dimensões espaciais da imagem, as *CNNs* adaptadas para dados unidimensionais (*Conv1D*) demonstraram alto desempenho em tarefas de análise de séries temporais, como classificação de sensores e detecção de anomalias (Kiranyaz *et al.*, 2021).

O princípio fundamental de uma *CNN* é a operação de convolução: um filtro (ou *kernel*) de tamanho fixo desliza sobre a sequência de entrada, calculando o produto interno entre seus pesos e o trecho correspondente de cada posição, permitindo ao modelo detectar padrões locais como um pico de frenagem ou uma zona de aceleração, independentemente de onde ocorrem na sequência (Li *et al.*, 2021). Combinadas às camadas convolucionais, aplicam-se operações de *pooling*, que reduzem a dimensionalidade da representação, tornando o modelo mais robusto a pequenas variações na posição dos padrões detectados e diminuindo o risco de *overfitting* (Carvalho *et al.*, 2025).

Os dados de telemetria da *Formula 1* constituem séries temporais multivariadas: para cada piloto e sessão, são registrados dados a uma taxa de 3,7 Hz como valores de velocidade (km/h), posição do acelerador (0–100%), estado do freio (booleano) e rotação do motor (rpm) (OpenF1, 2026). Para que esses dados possam ser utilizados como entrada em uma *CNN*, aplica-se a técnica de janela deslizante,

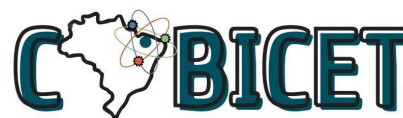
que consiste em percorrer a série com uma janela de tamanho fixo, extraíndo subconjuntos sobrepostos da sequência onde cada janela extraída se torna um exemplo de treinamento independente (Carvalho *et al.*, 2025). Esse procedimento é amplamente utilizado em tarefas de aprendizado de máquina sobre dados temporais de sensores biomédicos (Kiranyaz *et al.*, 2021), tendo sido adaptado neste trabalho para o contexto de telemetria automotiva.

A aplicação de técnicas de inteligência artificial na *Formula 1* tem crescido de forma significativa nos últimos anos, impulsionada tanto pela disponibilidade de grandes volumes de dados, quanto pelo avanço das ferramentas de aprendizado de máquina. Preosti (2021) destaca que as equipes utilizam modelos preditivos para otimizar estratégias durante as corridas, monitorando o desempenho dos carros e antecipando decisões.

Jafri (2024) propôs um modelo de aprendizado de máquina para prever resultados de corridas de *Formula 1* com base em dados históricos de qualificação, estratégia de pneus e posição de largada. O trabalho demonstra que algoritmos como *Random Forest* conseguem capturar padrões relevantes em dados históricos, apresentando limitações na adaptação a cenários dinâmicos durante a corrida. Bansal *et al.* (2025) realizaram uma análise de abordagens de aprendizado de máquina para previsão de desempenho na *Formula 1*, concluindo que modelos mais complexos, como redes neurais profundas, tendem a superar abordagens tradicionais quando há volume suficiente de dados de treinamento, mas ressaltam a importância da engenharia de atributos adequada.

No contexto de redes neurais recorrentes aplicadas à *Formula 1*, Mistry e Smith (2025) desenvolveram o *AutoF1*, um sistema baseado em *LSTM* (do inglês, *Long Short-Term Memory*) capaz de simular decisões estratégicas de corrida, especialmente o momento mais oportuno para *pit stops*. Embora o sistema demonstre resultados promissores na simulação de estratégias, seu foco está restrito à tomada de decisão sobre parada nos *boxes*, não abordando a previsão do ritmo individual dos pilotos com base em telemetria de alta frequência.

Zhao (2024) propôs um modelo de previsão de tempo de volta baseado em redes neurais profundas, utilizando dados históricos de voltas e características da pista, obtendo erros médios da ordem de poucos segundos. No entanto, o trabalho utiliza exclusivamente dados agregados por volta, sem incorporar dados de telemetria de alta frequência, limitando a granularidade das previsões e a capacidade do modelo de capturar o estilo individual de cada piloto.



O presente trabalho diferencia-se ao utilizar telemetria amostrada a 3,7 Hz, captando o comportamento detalhado do piloto ao longo de cada volta, e não apenas o tempo agregado. Diante do contexto apresentado, é proposto o desenvolvimento de um modelo preditivo baseado em redes neurais convolucionais, capaz de analisar e identificar padrões de desempenho de pilotos a partir de dados de telemetria em séries temporais. Tal abordagem tem como objetivo prever tempos de volta e auxiliar na identificação de padrões de pilotagem individuais de cada piloto em cada pista, servindo de apoio à decisão estratégica na *Formula 1*. A proposta envolve a coleta, processamento e análise de dados considerando o contexto regulatório da temporada de 2025, bem como a aplicação de técnicas de aprendizado de máquina, engenharia de atributos e a avaliação do modelo construído.

MATERIAL E MÉTODOS

Este trabalho adota uma abordagem experimental quantitativa, estruturada pelas etapas de coleta de dados, pré-processamento, construção e treinamento do modelo, avaliação de desempenho e análise comparativa com os acontecimentos reais.

1. Coleta de Dados

Os dados utilizados para a construção do modelo foram coletados a partir de uma *API (Application Programming Interface)* pública especializada em dados da *Formula 1*. A *API OpenF1* (OpenF1, 2026) disponibiliza dados de telemetria com frequência de aproximadamente 3,7 Hz, contendo os campos de *timestamp* (ISO 8601), velocidade (km/h), acelerador (0-100%), freio (booleano) e rotação do motor (rpm). Além disso, são coletadas informações sobre de cada volta (número de volta, *timestamp* início e duração em segundos) e acerca do calendário da temporada e a identificação das sessões de corrida. A coleta é realizada por meio de requisições *HTTP* à *API*, que retorna os dados no formato *JSON*, sendo estes posteriormente convertidos e persistidos localmente para uso nas etapas subsequentes do pipeline.

Os dados foram coletados para os seis pilotos de maior destaque na temporada de 2025: Verstappen (#1), Norris (#4), Leclerc (#16), Hamilton (#44), Russell (#63) e Piastri (#81), contemplando três pistas com características físicas e de telemetria distintas: o *GP* do Japão (Suzuka), o *GP* da Bélgica (Spa-Francorchamps), e o *GP* de Monte Carlo (Mônaco). Os dados brutos foram persistidos localmente no formato *Apache Parquet*, que oferece armazenamento binário comprimido com preservação de tipos de dados.

2. Pré-Processamento

O pré-processamento dos dados ocorreu em quatro etapas sequenciais, seguindo as diretrizes de

preparação de dados para modelos de aprendizado de máquina descritas por Carvalho *et al.* (2025). Primeiramente, foram removidas as voltas com campo *lap_duration* ausente, situação observada em casos de abandono ou dados incompletos da *API*, sendo esse um procedimento padrão de limpeza de dados recomendado pelos autores para garantir a integridade do conjunto de treinamento.

Em seguida, as séries de telemetria foram segmentadas por volta, utilizando os *timestamps* de início e a duração registrada para delimitar o intervalo de amostras correspondente a cada volta. Na terceira etapa, voltas afetadas por condições de corrida neutralizada foram removidas por meio de filtragem baseada em contexto de corrida. Os eventos de controle de cada sessão foram consultados na *API OpenF1* para identificar períodos de *Safety Car*, *Virtual Safety Car* e Bandeira Vermelha, descartando as voltas que se sobrepõem a esses intervalos. Para sessões sem dados de controle disponíveis, aplica-se um filtro estatístico alternativo, removendo voltas cujo tempo se afaste mais de 2,0 desvios padrões da média do piloto. Essa abordagem é recomendada por Carvalho *et al.* (2025) como estratégia para garantir que o modelo aprenda a partir de exemplos representativos do comportamento esperado.

Em decorrência dessa filtragem, o número de voltas válidas não é uniforme entre pilotos e pistas, uma vez que cada piloto é afetado de forma distinta por neutralizações, incidentes e abandonos ao longo de cada corrida.

Por fim, cada canal de telemetria foi normalizado individualmente pelo seu valor máximo global na sessão, mapeando os valores para o intervalo [0, 1]. Essa normalização por máximo é amplamente adotada em pré-processamento de redes neurais para garantir que canais com escalas muito diferentes contribuam de forma equilibrada para o aprendizado (Carvalho *et al.*, 2025). O tempo de volta foi normalizado para o intervalo [0, 1] com base no mínimo e máximo observados para aquele piloto naquela pista, sendo os parâmetros de escala preservados para reconversão durante a avaliação.

3. Janelamento Deslizante

Após o pré-processamento, as séries temporais por volta foram transformadas em janelas deslizantes para alimentar a rede neural. Cada janela possui tamanho fixo de 150 amostras, correspondendo a aproximadamente 40 segundos de telemetria a 3,7 Hz, com passo de 10 amostras entre janelas consecutivas. Para uma volta típica de aproximadamente 333 amostras em Suzuka, esse processo gera em torno de 18 janelas por volta.

Cada janela é representada por um tensor de dimensões (4, 150), em que 4 corresponde aos canais

de telemetria (velocidade, acelerador, freio e rotações por minuto) e 150 corresponde à dimensão temporal. O rótulo associado a cada janela é o tempo de volta normalizado da volta à qual aquela janela pertence. Esse design permite que o modelo possa aprender a associar padrões de telemetria em escala de dezenas de segundos com o ritmo geral da volta, capturando informações de setores completos da pista.

4. Arquitetura do Modelo

A arquitetura proposta consiste em uma rede convolucional unidimensional, implementada com a biblioteca *PyTorch* (PyTorch, [2026?]), sem utilização de arquiteturas pré-treinadas, conforme ilustrado na Figura 1. A escolha da *CNN*, em comparação a outras arquiteturas como redes recorrentes ou perceptrons multicamada, justifica-se pela capacidade da operação de convolução em capturar padrões locais na telemetria de forma eficiente. Além disso, a combinação de camadas *Conv1D* com *Global Average Pooling* favorece a extração de atributos em séries temporais com um número reduzido de parâmetros, diminuindo o risco de explosão de parâmetros e de overfitting em comparação a arquiteturas totalmente conectadas (Kiranyaz *et al.*, 2021).

O modelo é composto por um bloco convolucional e uma camada de regressão na saída. O bloco convolucional é formado por duas camadas *Conv1D* em sequência, sendo a primeira composta por 64 filtros com *kernel* de tamanho 7, responsável por detectar padrões locais de curta duração, como zonas de frenagem. A segunda camada possui 128 filtros e *kernel* de tamanho 5, responsável por capturar padrões de maior escala temporal, como o comportamento ao longo de um setor. Cada camada convolucional é seguida pela função de ativação *ReLU* (Almada, 2019, p. 8-9). Uma camada de *Global Average Pooling* ao final do bloco reduz a dimensão temporal para um único valor por canal, produzindo um vetor de 128 elementos.

A camada de regressão (saída) recebe o vetor de 128 características e o processa por meio de uma função Linear (128, 64), *ReLU*, *Dropout* com taxa de 0,2 para reduzir o risco de *overfitting* e Linear (64, 1), produzindo o tempo de volta normalizado como saída. O modelo é treinado com otimizador Adam com taxa de aprendizado de 0,001 e função de perda *MSE Loss* (erro quadrático médio). O treinamento foi realizado por 150 épocas, com divisão dos dados em 70% para treino e 30% para teste (Carvalho *et al.*, 2025). Os hiperparâmetros como número de camadas convolucionais, quantidade de filtros (64 e 128), tamanhos de *kernel* (7 e 5), taxa de aprendizado de 0,001, taxa de *dropout* de 0,2 e número de 150 épocas, foram definidos empiricamente a partir de experimentos preliminares, nos quais se observou a

estabilização das curvas de erro de treino e teste, sem ganhos relevantes adicionais com configurações maiores.

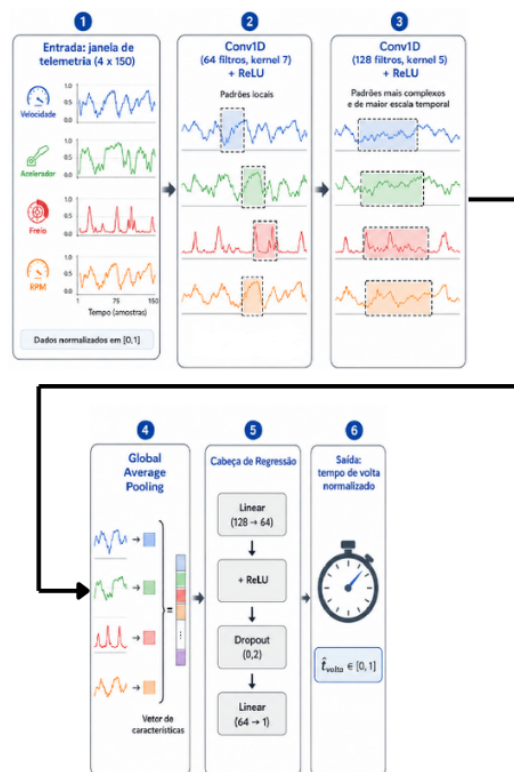


Figura 1: Arquitetura da Rede Convolucional 1D para Previsão do Tempo de Volta

Um modelo independente, foi treinado para cada combinação de piloto e pista, totalizando 18 combinações (6 pilotos x 3 pistas). Esta estratégia permite que cada modelo aprenda os padrões específicos de telemetria de um único piloto em uma única pista, evitando a confusão introduzida por diferenças de desempenho entre carros de equipes distintas, sendo esse um problema identificado em experimentos preliminares nos quais a mistura de múltiplos pilotos em um único modelo resultou em erros superiores a 10 segundos.

5. Avaliação de Desempenho

O desempenho de cada modelo foi avaliado no conjunto de teste (30% dos dados não vistos durante o treinamento) por meio de duas métricas padrões de regressão.

O *MAE* (erro médio absoluto) representa a média dos valores absolutos das diferenças entre o tempo previsto e o tempo real, fornecendo uma medida



direta e interpretável do erro em segundos. É definido conforme a Equação 1:

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (1)$$

O MSE (erro quadrático médio) eleva as diferenças ao quadrado antes de calcular a média, penalizando proporcionalmente erros maiores, sendo utilizado como função de perda durante o treinamento. É definido conforme a Equação 2:

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (2)$$

em que y_i representa o tempo de volta real, $\hat{y}_i$ o tempo de volta previsto pelo modelo e n o número total de amostras avaliadas.

Como métrica complementar, foi também calculado o RMSE (raiz do erro quadrático médio), definido como a raiz quadrada do MSE, conforme a Equação 3:

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (3)$$

Por penalizar mais fortemente os erros de maior magnitude, o RMSE é sensível à presença de previsões muito distantes do valor real, complementando o MAE na identificação de voltas atípicas.

6. Ferramentas

A implementação foi realizada na linguagem Python (Python Software Foundation, 2001-2026), muito utilizada em projetos de aprendizado de máquina e ciência de dados contendo bibliotecas especializadas, sendo elas:

- *PyTorch* (framework de aprendizado profundo utilizado para a construção, treinamento e avaliação dos modelos de redes neurais convolucionais. Oferece suporte nativo a computação com tensores e diferenciação automática (PyTorch, [2026?]). A escolha do PyTorch, em detrimento de alternativas como TensorFlow ou Keras, deveu-se à sua flexibilidade na definição explícita de arquiteturas e ao controle direto sobre o laço de treinamento, características adequadas ao desenvolvimento de modelos personalizados por piloto e pista.;

- *Pandas* e *NumPy*: utilizadas para manipulação, transformação e análise dos dados de telemetria coletados, incluindo operações de filtragem, segmentação por volta e normalização dos canais (Pandas, NumPy Team 2026);
- *Matplotlib*: empregada para a geração de gráficos dos resultados, incluindo curvas de aprendizado, dispersão entre tempo previsto e real, e análises comparativas de ritmo entre pilotos (The Matplotlib Development Team, 2012-2026).

O armazenamento dos dados brutos foi realizado no formato *Apache Parquet*, um formato de armazenamento binário comprimido amplamente utilizado em pipelines de ciência de dados por sua eficiência de leitura e preservação de tipos de dados (Apache Parquet, 2026). Houve ainda, colaboração e compartilhamento de código por meio do *Google Colab* (Google Research, 2026), plataforma baseada em nuvem que oferece acesso gratuito a unidades de processamento gráfico.

Adicionalmente, foram utilizadas ferramentas de inteligência artificial generativa como suporte ao desenvolvimento. O modelo *Claude* (Anthropic, versão Sonnet 4.6, 2026) foi empregado como auxílio na revisão textual, enquanto o modelo *ChatGPT* (OpenAI, versão GPT-5, 2026) contribuiu na geração da Figura 1.. Vale ressaltar que o uso dessas ferramentas ocorreu de forma complementar ao trabalho dos autores, sem substituir a análise, a interpretação dos resultados ou a tomada de decisões metodológicas.

RESULTADOS E DISCUSSÃO

1. Eficiência do Aprendizado por Épocas

As curvas de aprendizado registram a evolução do MSE para os blocos de dados de treino e teste ao longo das 150 épocas de treinamento, para cada um dos seis pilotos em cada pista.

1.1 Mônaco (Monte Carlo)

A Figura 2 e a Figura 3 apresentam as curvas de aprendizado dos seis pilotos na pista de Mônaco. O comportamento geral é de convergência gradual e suave, com os modelos atingindo os melhores valores de validação entre as épocas 143 e 150. Todos os pilotos apresentaram curvas de treino e validação sobrepostas ou muito próximas durante todo o treinamento, evidenciando ausência de um *overfitting* significativo.

Entre os pilotos, Leclerc apresenta o padrão mais ruidoso, com picos pontuais na curva de validação até a época 80, após a qual se estabiliza. Russell, por sua vez, apresenta a convergência mais homogênea dos seis modelos. Os valores finais de *MSE* giram em

torno de 0,008 a 0,012, indicando boa capacidade preditiva para essa pista.

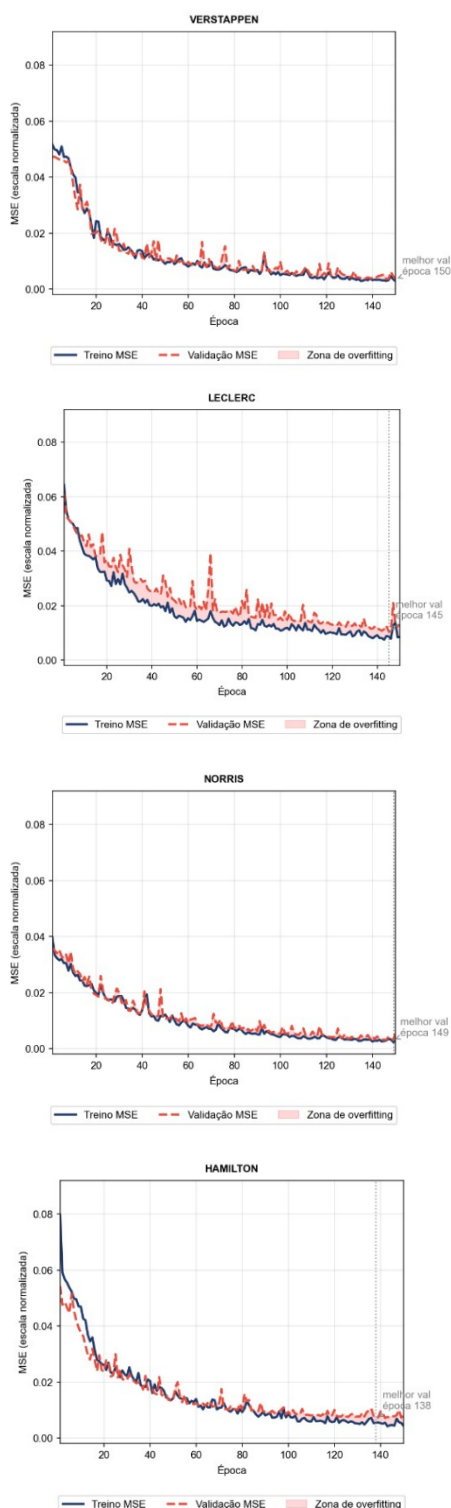


Figura 2: Curva de aprendizado por piloto (Verstappen, Leclerc, Norris e Hamilton) – Mônaco

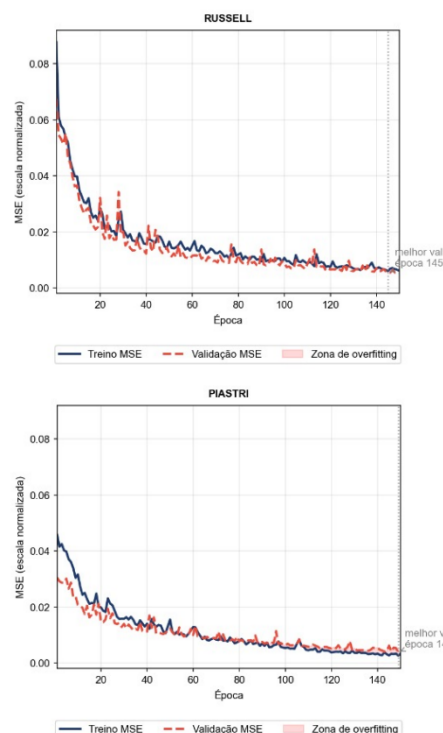


Figura 3: Curva de aprendizado por piloto (Russell e Piastri) – Mônaco

1.2 Bélgica (Spa-Francorchamps)

As curvas de aprendizado em Spa-Francorchamps, apresentadas na Figura 4 e na Figura 5, revelam uma dinâmica de treinamento mais desafiadora em comparação a Mônaco. Hamilton se destaca por apresentar o melhor valor de validação na época 128, convergindo mais cedo que os demais pilotos. Norris, em contraste, apresenta o maior ruído nas curvas de validação entre as épocas 40 e 70, com a zona de *overfitting* mais expressiva do conjunto.

Esse comportamento é consistente com a natureza da pista: uma vez que possui o maior comprimento entre as pistas avaliadas e maior variabilidade de tempos de volta em razão de mudanças meteorológicas e variações de combustível ao longo da corrida. Essas condições tendem a introduzir maior ruído no sinal de telemetria, dificultando a extração de padrões estáveis nas primeiras épocas.

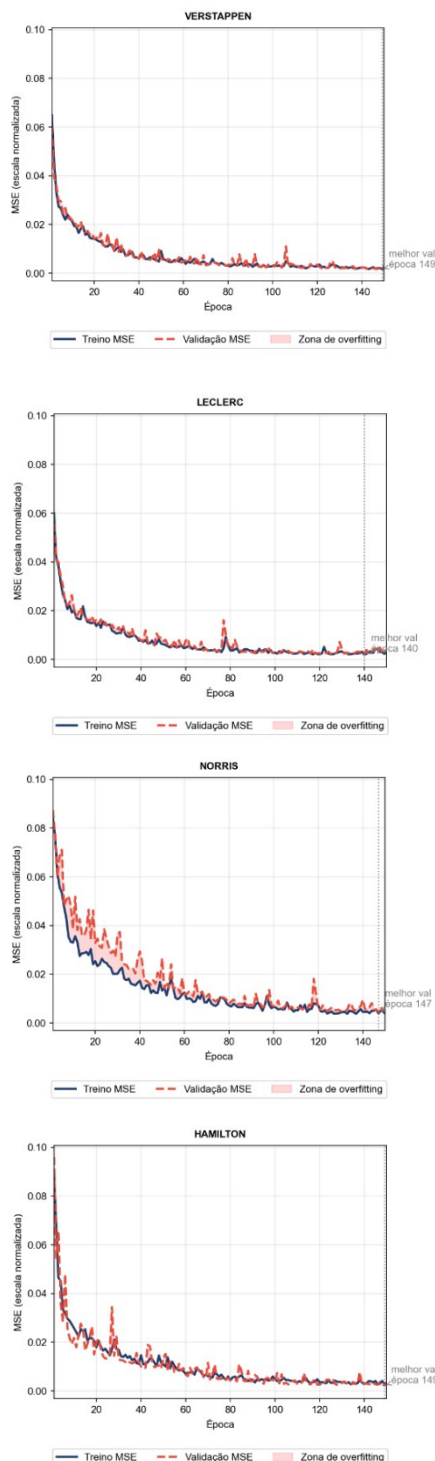


Figura 4: Curvas de aprendizado por piloto (Verstappen, Leclerc, Norris e Hamilton)- Spa-Francorchamps

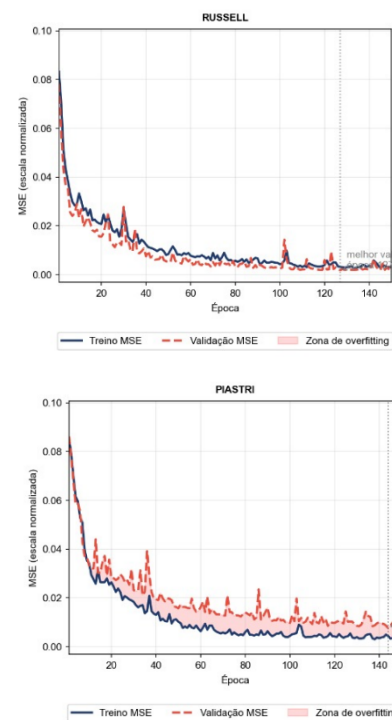


Figura 5: Curvas de aprendizado por piloto (Russell e Piastri) - Spa-Francorchamps

1.3 Japão (Suzuka)

A pista de Suzuka apresenta as curvas de aprendizado mais heterogêneas, conforme ilustrado na Figura 6 e na Figura 7. Verstappen e Norris convergem de forma estável e suave, com curvas de treino e validação bem alinhadas desde as primeiras épocas. Já Leclerc e Russell apresentam picos pontuais expressivos, mas que não impedem a convergência ao final.

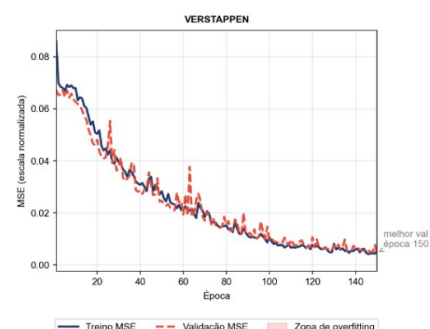
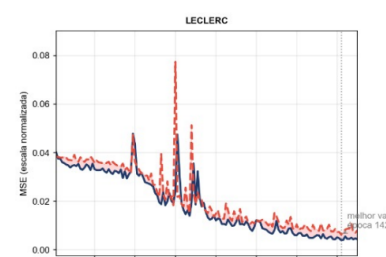


Figura 6: Curvas de aprendizado por piloto (Verstappen) - Suzuka



A precisão das previsões de tempo de volta é apresentada por meio dos gráficos de tempo real x tempo previsto (em segundos), em que a linha tracejada representa o Previsto igual ao Real. Pontos próximos à diagonal indicam alta acurácia, pontos acima da diagonal indicam superestimação do tempo de volta e os pontos abaixo indicam subestimação.

2.1 Mônaco (Monte Carlo)

Os resultados para a pista de Mônaco são apresentados na Figura 8 e na Figura 9. O modelo atingiu desempenho consistente para a maioria dos pilotos, com *MAE* entre 0,31s para Norris e 0,36s Piastri, sendo esses os valores mais baixos do grupo principal. Russell representa o único caso destoante da pista, com *MAE* de 0,68s, sendo possível observar que os pontos estão distribuídos de forma mais dispersa e com viés de superestimação para voltas no intervalo de 76 a 82 segundos.

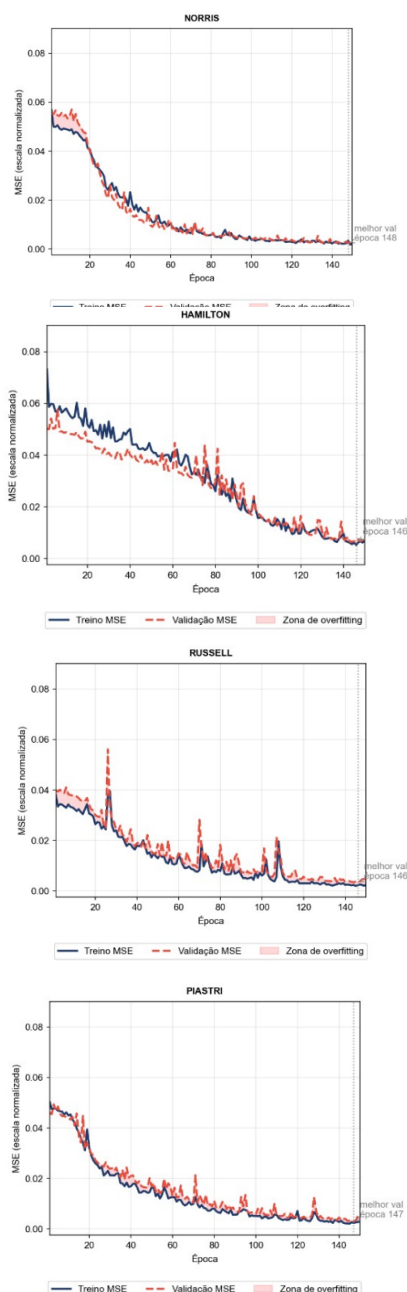


Figura 7: Curvas de aprendizado por piloto (Leclerc, Norris, Hamilton, Russell e Piastri) - Suzuka

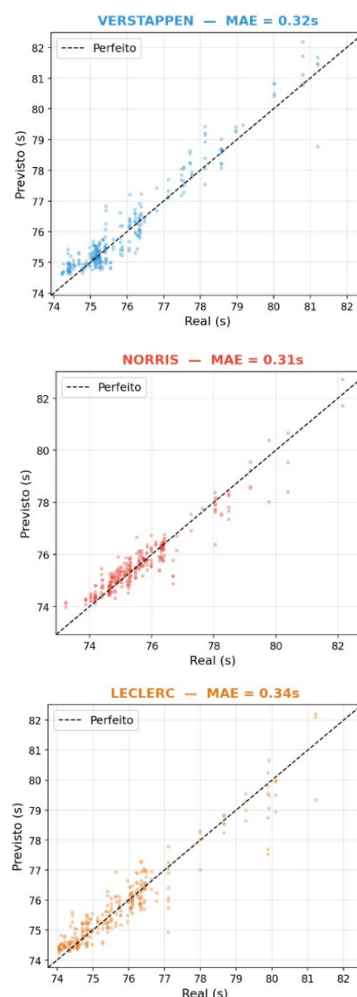
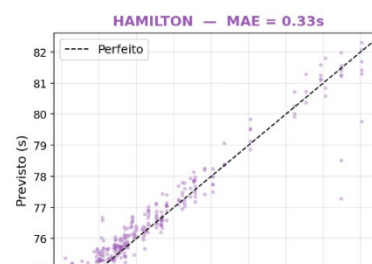


Figura 8: Comparação entre Tempo Real e Tempo Previsto - Por Piloto (Verstappen, Norris e Leclerc) - Mônaco

2. Desempenho Preditivo por Pista



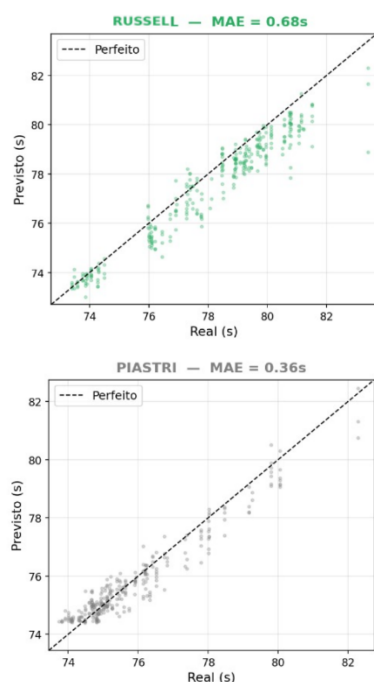


Figura 9: Comparação entre Tempo Real e Tempo Previsto - Por Piloto (Hamilton, Russell e Piastri)- Mônaco

2.2 Bélgica (Spa-Francorchamps)

Conforme ilustrado na Figura 10 e na Figura 11, a pista de Spa-Francorchamps apresenta os maiores erros médios absolutos entre as três pistas avaliadas. Os MAEs variam de 0,77s para Leclerc, a 1,22s de Piastri, com Hamilton registrando 1,09s e Piastri o pior desempenho individual do conjunto completo de experimentos.

O desempenho inferior possivelmente em Spa se deve a maior variabilidade de condições climáticas, a extensão da pista (sendo essa a mais longa da temporada) e a sensibilidade do tempo de volta a mudanças de vento e temperatura aumentam o ruído dos dados de telemetria.

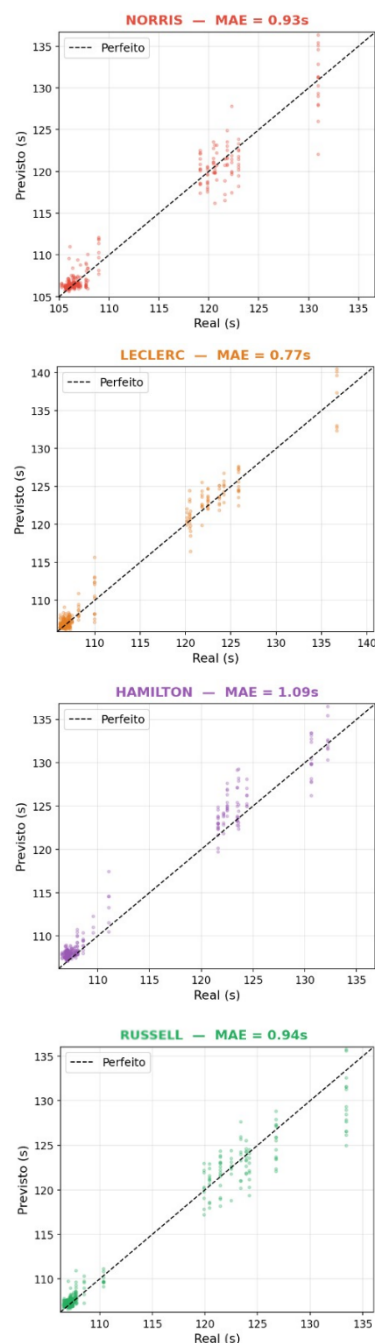


Figura 10: Comparação entre Tempo Real e Tempo Previsto - Por Piloto (Verstappen, Norris, Leclerc, Hamilton e Russell) - Spa-Francorchamps



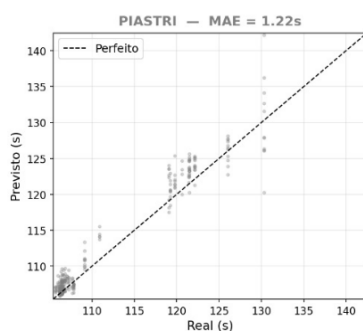


Figura 11: Comparação entre Tempo Real e Tempo Previsto - Por Piloto (Piastrì) - Spa-Francorchamps

2.3 Japão (Suzuka)

Suzuka representa o melhor desempenho do modelo em termos absolutos, conforme apresentado na Figura 12 e na Figura 13. Os MAEs variam de 0,12s de Norris a 0,25s Russell, com todos os pilotos abaixo de 0,25 segundos. Os gráficos de dispersão mostram pontos extremamente concentrados ao redor da diagonal de referência, com dispersão reduzida mesmo para os outliers mais distantes.

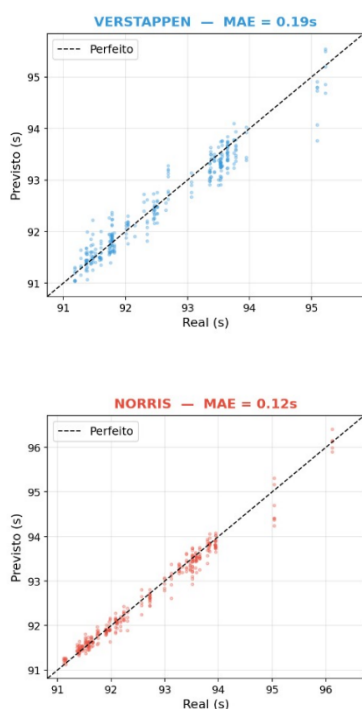


Figura 12: Comparação entre Tempo Real e Tempo Previsto - Por Piloto (Verstappen e Norris) - Suzuka

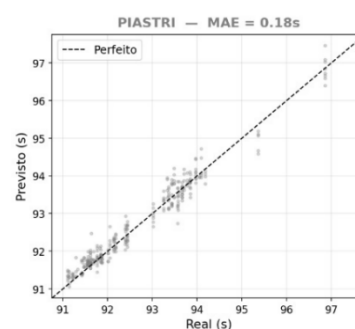
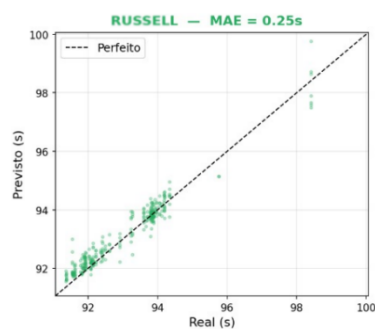
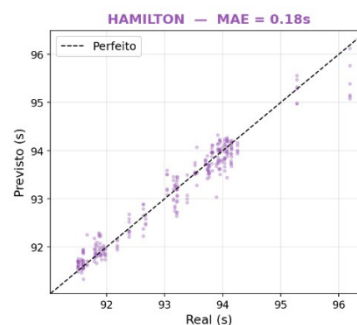
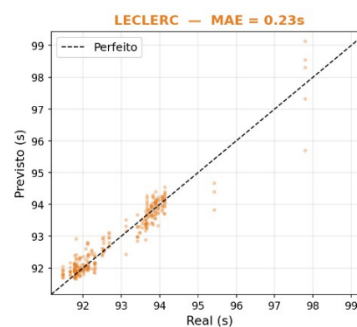
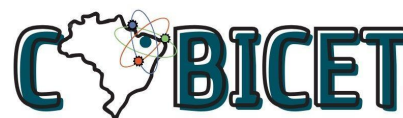


Figura 13: Comparação entre Tempo Real e Tempo Previsto - Por Piloto (Leclerc, Hamilton, Russell e Piastrì) - Suzuka

CONCLUSÃO

O presente trabalho propôs e avaliou um modelo preditivo baseado em redes neurais convolucionais unidimensionais para previsão de tempos de volta na *Formula 1* a partir de dados de telemetria pública. Os resultados demonstram que o objetivo principal foi alcançado: o modelo é capaz de aprender os padrões individuais de pilotagem de cada piloto em cada pista



e prever tempos de volta com erro absoluto inferior a 1 segundo em duas das três pistas avaliadas, atingindo erro médio absoluto inferior a 0,25 segundos em Suzuka. A proximidade entre as curvas de treino e validação ao longo das 150 épocas confirma que a camada *Dropout* e a estratégia de treinamento individualizado por piloto e pista foram eficazes no controle de generalização, resultando em convergência estável e sem indícios relevantes de *overfitting*.

Em relação às limitações identificadas, destaca-se a sensibilidade do modelo a eventos atípicos de corrida. Os picos observados nas curvas de aprendizado evidenciam que a filtragem de outliers não eliminou completamente os efeitos de bandeiras vermelhas, *safety cars* e incidentes que alteram momentaneamente o ritmo dos pilotos. O desempenho inferior em Spa-Francorchamps reforça que condições climáticas variáveis e circuitos de maior extensão representam desafios adicionais para a estabilidade do sinal de telemetria e, consequentemente, para a qualidade das previsões.

Como oportunidades futuras, observa-se a incorporação de variáveis contextuais de estado de corrida, como tipo de pneu, número de voltas desde o último pit stop e informações sobre *safety car*, que permitiriam ao modelo distinguir entre voltas de ritmo normal e voltas perturbadas por eventos externos. Espera-se também aprimorar a arquitetura do modelo, garantindo maior acurácia no objetivo proposto. Por fim, deseja-se desenvolver uma interface de visualização em tempo real que consuma as previsões do modelo e as exiba de forma interativa para analistas, comentaristas e fãs, ampliando o impacto prático do trabalho além do contexto acadêmico.

REFERÊNCIAS

ALMADA, M. Modelos de neurônios em redes neurais artificiais. ResearchGate, 2019. Disponível em: <https://doi.org/10.13140/RG.2.2.32042.16323>. Acesso em: 22 Abr. 2026.

ANTHROPIC. Claude (versão Sonnet 4.6). San Francisco: Anthropic, 2026. Disponível em: <https://claude.ai>. Acesso em: 2026.

APACHE SOFTWARE FOUNDATION. Apache Parquet. 2026. Disponível em: <https://parquet.apache.org>. Acesso em: 2026.

BANSAL, Aayam et al. Advanced Machine Learning Approaches for Formula 1 Race Performance Prediction: A Comprehensive Analysis of Championship Point Forecasting. Research Gate, jul. 2025. Disponível em:

<https://doi.org/10.13140/RG.2.2.20910.01607>. Acesso em: 22 Mar. 2026.

BELCIC, Ivan; STRYKER, Cole. O que é engenharia de dados? IBM.com, [2024?]. Disponível em: <https://www.ibm.com/br-pt/think/topics/data-engineering>. Acesso em: 19 Abr. 2026.

CARVALHO, André C. P. L. F. de; MENEZES, Angelo G.; BONIDIA, Robson P. Ciência de Dados - Fundamentos e Aplicações. LTC, Rio de Janeiro, 2025.

GOOGLE RESEARCH. Google Colaboratory. Disponível em: <https://colab.research.google.com>. Acesso em: 2026.

HAYKIN, Simon. Redes Neurais: Princípios e Prática. 2. ed. Bookman, Porto Alegre, 2007.

JAFRI, Ali. Predicting Formula 1 Race Outcomes: A Machine Learning Approach. NYUAD Capstone Seminar Reports, Abu Dhabi, 2024. Disponível em: https://aliabdullahjafri.com/static/media/Ali_Jafri_CapstoneProject1_Fall2024.c7244022875d46bec5d9.pdf. Acesso em: 22 Mar. 2026.

KIRANYAZ, Serkan et al. 1D Convolutional Neural Networks and Applications: A Survey. Mechanical Systems and Signal Processing, v. 151, art. 107398, abr. 2021. Disponível em: <https://doi.org/10.1016/j.ymssp.2020.107398>. Acesso em: 28 Abr. 2026.

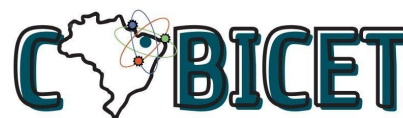
LI, Zewen et al. A Survey of Convolutional Neural Networks: Analysis, Applications, and Prospects. IEEE Transactions on Neural Networks and Learning Systems, v. 33, n. 12, p. 6999–7019, 2021. Disponível em: https://www.researchgate.net/publication/352306122_A_Survey_of_Convolutional_Neural_Networks_Analysis_Applications_and_Prospects. Acesso em: 10 Mar. 2026.

MISTRY, Vivek; SMITH, William Alfred Peter. AutoF1: An RNN-Based Approach to Simulating Strategic Decision-Making in Formula 1. IEEE Intelligent Transportation Systems Conference (ITSC), 21 jul. 2025. Disponível em: <https://eprints.whiterose.ac.uk/id/eprint/229478/>. Acesso em: 24 Mar. 2026.

NUMPY TEAM. NumPy. 2026. Disponível em: <https://numpy.org>. Acesso em: 2026.

OPENAI. ChatGPT (versão GPT-5). San Francisco: OpenAI, 2026. Disponível em: <https://chatgpt.com>. Acesso em: 2026.

OPENF1. OpenF1 API - Introduction. [entre 2023 e 2026]. Disponível em: <https://openf1.org/docs/#introduction>. Acesso em: 2026.



PANDAS. Python Data Analysis Library. 2026. Disponível em: <https://pandas.pydata.org>. Acesso em: 2026.

PREOSTI, Elettra. Data analysis for Formula 1. Berkeley Scientific Journal, v. 25, n. 2, 2021. Disponível em: <https://doi.org/10.5070/BS325254500>. Acesso em: 24 Mar. 2026.

PROVOST, Foster; FAWCETT, Tom. Data Science and Its Relationship to Big Data and Data-Driven Decision Making. Big Data, v. 1, n. 1, p. 51–59, 2013. Disponível em: <https://doi.org/10.1089/big.2013.1508>. Acesso em: 12 Abr. 2026.

PYTHON SOFTWARE FOUNDATION. Python (versão 3.11.9). 2001-2026. Disponível em: <https://www.python.org>. Acesso em: 23 maio 2026.

PYTORCH. PyTorch 2.12 Documentation. [2026?]. Disponível em: <https://docs.pytorch.org/docs/2.12/index.html>. Acesso em: 2026.

THE MATPLOTLIB DEVELOPMENT TEAM. Matplotlib: Visualization with Python. 2012-2026. Disponível em: <https://matplotlib.org>. Acesso em: 2026.

ZHAO, Zhixuan. Deep Neural Network-Based Lap Time Forecasting of Formula 1 Racing. Applied and Computational Engineering, v. 47, n. 1, p. 61–66, 15 mar. 2024. Disponível em: <https://doi.org/10.54254/2755-2721/47/20241191>. Acesso em: 19 Abr. 2026.