

Projeto Semestral - Ciência de Dados
Previsão de Classificação de Segurança em Testes de Colisão Veicular
Utilizando Aprendizado de Máquina

Integrantes

Nome: Caio Rabinovich Panes Brunholi	RA: 20.01285-3
Nome: Felipe Onishi Yaegashi	RA: 20.00255-6
Nome: Jonathan Martins Gomes	RA: 20.00862-7
Nome: Matheus Marins Bernardello	RA: 20.00286-6

Github: https://github.com/CD-CrashTest/Crash_Test

Resumo

Este estudo apresenta uma abordagem de aprendizado de máquina para prever a classificação de segurança de veículos em testes de colisão (*crash tests*) com base em dados históricos de testes anteriores. Órgãos como o Latin NCAP, Euro NCAP e o NHTSA (National Highway Traffic Safety Administration) realizam esses testes, classificando veículos de 0 a 5 estrelas. Essa classificação é essencial para indicar a segurança do veículo em caso de acidente, ajudando consumidores e fabricantes a avaliar o desempenho do automóvel em situações de risco. O objetivo do estudo é prever essa classificação de segurança utilizando métodos de aprendizado de máquina, a partir de dados técnicos e de segurança veicular, como peso, tipo de veículo, presença de sistemas de alerta e proteção para o passageiro, obtidos por meio de técnicas de *web scraping*.

Introdução e Contextualização

A segurança veicular é uma prioridade na indústria automotiva. Testes de colisão são conduzidos por instituições como Latin NCAP, Euro NCAP e NHTSA para avaliar o desempenho de veículos em cenários de acidentes, atribuindo uma classificação de segurança baseada em características técnicas e de segurança dos veículos. O objetivo deste projeto é prever essa classificação de segurança com base em dados históricos de *crash tests*, utilizando aprendizado de máquina. Através da criação de um modelo preditivo, buscamos auxiliar consumidores e fabricantes a estimar a segurança de novos modelos de veículos antes da realização dos testes destrutivos.

Para o estudo de caso apresentado foi escolhido obter a base de dados do Euro NCAP. O European New Car Assessment Programme é uma organização que realiza testes de segurança em veículos novos desde 1997. Até novembro de 2024, o Euro NCAP testou mais de 1000 modelos de veículos, fornecendo avaliações detalhadas sobre a segurança de cada um. Esses testes abrangem diversos aspectos, incluindo proteção de ocupantes adultos e infantis, segurança de pedestres e sistemas de assistência ao motorista. As avaliações são atualizadas regularmente para refletir os avanços tecnológicos e as mudanças nos padrões de segurança veicular. Os dados obtidos dos veículos são sumarizados em uma classificação de 0 a 5, sendo 0 um veículo pouco seguro e 5 um altamente seguro.



5 star safety: Overall excellent performance in crash protection and well equipped with state-of-the-art crash avoidance technology.



4 star safety: Overall good performance in crash protection and prevention, but does not reach the highest standards in all key assessment areas.



3 star safety: A vehicle that provides an average level of safety performance, offering a medium standard in safety.



2 star safety: Nominal crash protection but below-average performance in one or more of the key assessment areas.



1 star safety: Marginal crash protection and minimal crash avoidance technology beyond the legal requirements.



0 star safety: Meeting type-approval standards so can legally be sold but lacking critical modern safety technology.

Fonte: Euro NCAP, 2024

Metodologia

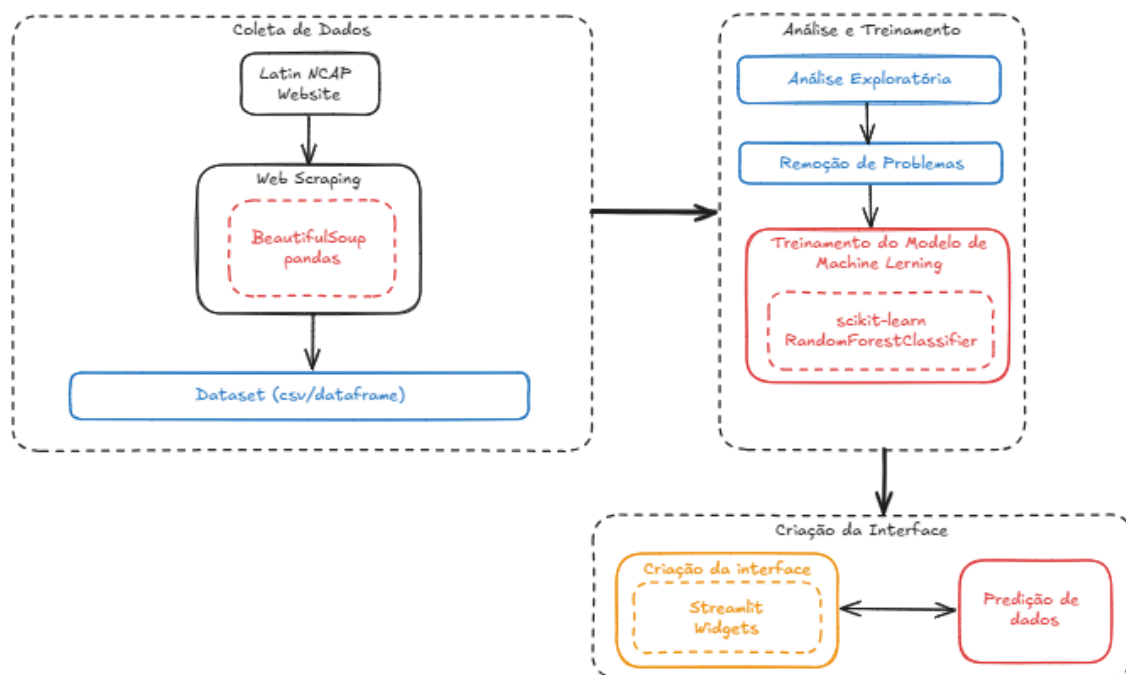
A metodologia aplicada neste estudo seguiu as seguintes etapas:

Coleta de Dados: Dados de crash tests foram extraídos via web scraping do Euro NCAP. Informações como peso, tipo de veículo, características de segurança e a classificação de segurança foram extraídas.

Pré-processamento dos Dados: Para cada variável numérica (ex., peso), foram aplicadas normalizações e preenchimento de valores ausentes com valores gerados a partir da média \pm desvio padrão. Colunas categóricas, como o tipo de veículo, foram convertidas em representações numéricas através de one-hot encoding.

Desenvolvimento do Modelo: Um modelo de rede neural foi construído utilizando o Keras, com camadas densas e funções de ativação ReLU nas camadas ocultas e função softmax na camada de saída. O modelo foi treinado com o objetivo de prever a classificação do veículo de 0 a 5 estrelas.

Interface de Usuário: A interface foi desenvolvida com o Streamlit para permitir que usuários insiram dados de veículos e visualizem a previsão de classificação de segurança. A interface também apresenta a acurácia do modelo e um histórico de treinamento.



Fonte: Próprio Autor, 2024

Foram utilizadas duas estratégias para a criação de um modelo, uma utilizando o Random Fores Classifier e outra criando uma pequena rede neural com Keras. O primeiro modelo usa o Random Forest Classifier, uma técnica de aprendizado de máquina baseada em árvores de decisão que cria múltiplas árvores para melhorar a precisão da previsão e reduzir o overfitting. Cada árvore do modelo é treinada com uma amostra aleatória dos dados, resultando em um conjunto robusto que vota para a classificação final. O Random Forest é particularmente eficaz em capturar padrões de dados tabulares e pode lidar bem com variáveis categóricas e numéricas. A construção desse modelo envolve etapas de pré-processamento, como a normalização de variáveis numéricas e a codificação one-hot das variáveis categóricas, o que permite ao modelo interpretar diferentes tipos de dados de forma mais precisa.

O segundo modelo, uma rede neural construída com Keras, adota uma abordagem mais complexa, ideal para capturar interações profundas entre as variáveis de entrada. A rede neural é composta por múltiplas camadas densas com funções de ativação ReLU nas camadas ocultas, que ajudam a capturar não linearidades nos dados, e uma função de ativação softmax na camada de saída, adequada para classificação multiclasse. Ao contrário do Random Forest, a rede neural exige uma normalização rigorosa dos dados de entrada para que as variáveis estejam em uma escala uniforme, aumentando a eficiência do treinamento. Enquanto o Random Forest fornece uma interpretação mais intuitiva das variáveis importantes, a rede neural, por sua vez, é mais adequada para identificar padrões complexos e interações entre as variáveis que não são facilmente observáveis em modelos baseados em árvores de decisão.

As colunas escolhidas para os treinamentos com o Random Forest e a rede neural incluíram principalmente variáveis relacionadas às características de segurança e especificações técnicas do veículo. Para ambos os modelos, as colunas selecionadas foram:

- *Kerb Weight* (Peso do Veículo): Essa variável é importante pois o peso do veículo afeta diretamente sua estabilidade e capacidade de absorver impacto em caso de colisão.
- *Safety Features* (Equipamentos de Segurança): Esta coluna reflete a presença de tecnologias e dispositivos que aumentam a segurança, como airbags e sistemas de frenagem, que influenciam positivamente na classificação de segurança.
- *CRS Installation Check* (Facilidade de Instalação de Cadeirinha Infantil): A facilidade de instalação de cadeirinhas reflete a segurança oferecida para crianças, um fator que contribui diretamente para a pontuação de segurança.
- *Class* (Classe do Veículo): A classe ou tipo do veículo (por exemplo, SUV, sedan, hatchback) é uma variável relevante, pois diferentes classes apresentam variações estruturais e de segurança que podem impactar o desempenho em testes de colisão.
- *Seatbelt Reminder* (Lembrete de Cinto de Segurança): O lembrete do cinto é um sistema de alerta simples, mas eficaz, que promove a segurança e pode influenciar a avaliação do veículo.

Resultados e Discussão

Os dados foram adequadamente processados e utilizados para treinar o modelo Random Forest com uma acurácia por volta de 76% nos testes e o modelo Keras com uma acurácia por volta de 75%, demonstrando a capacidade do modelo de generalizar as características dos dados de *crash test* para prever a classificação de segurança. A interface Streamlit simplificou a interação com o sistema, permitindo que usuários insiram os parâmetros como peso do veículo, nota de segurança e características de instalação de cadeirinhas para crianças, e recebam uma previsão da classificação de segurança.

Os resultados obtidos indicam que o modelo possui uma boa capacidade de generalização, tornando-o adequado para prever classificações de segurança para novos modelos de veículos. A aplicação de uma rede neural permitiu capturar interações complexas entre as variáveis, que seriam difíceis de modelar utilizando métodos de aprendizado de máquina mais simples. No entanto, limitações foram encontradas, como o tamanho restrito do conjunto de dados e a falta de dados categóricos adicionais que poderiam contribuir para a acurácia do modelo.

Conclusão

Este projeto desenvolveu uma ferramenta de previsão de segurança veicular baseada em aprendizado de máquina, com uma interface amigável para o usuário e acurácia satisfatória. O modelo preditivo tem o potencial de apoiar a indústria

automotiva na fase de desenvolvimento e de auxiliar consumidores a fazerem escolhas informadas. Futuros aprimoramentos no modelo podem incluir o aumento da base de dados e a inclusão de variáveis que incorporem novos aspectos de segurança. A ferramenta possui um potencial significativo para tornar o processo de avaliação de segurança mais acessível e eficiente, fornecendo uma análise de risco inicial com base em características técnicas do veículo.

Referências

Latin NCAP. (2023). "Crash Test Results". Disponível em:
<https://www.latinncap.com/po/resultados>.

NHTSA. (2023). "5-Star Safety Ratings". Disponível em:
<https://www.nhtsa.gov/ratings>.

EURO NCAP. (2024). *About Euro NCAP*. Disponível em:
<https://www.euroncap.com/en/about-euro-ncap/>