

Redes Neurais Convolucionais

Arthur Abrahão Santos Barbosa
Universidade Federal de Pernambuco
Centro de Informática
Pernambuco, Brasil
aasb2@cin.ufpe.br

Arthur Henrique
Universidade Federal de Pernambuco
Centro de Informática
Pernambuco, Brasil
ahac@cin.ufpe.br

Filipe Samuel da Silva
Universidade Federal de Pernambuco
Centro de Informática
Pernambuco, Brasil
fss8@cin.ufpe.br

Vinicius Bastos Moreira Principe
Universidade Federal de Pernambuco
Centro de Informática
Pernambuco, Brasil
vbmp@cin.ufpe.br

I. INTRODUÇÃO

Acidentes de trânsito são inesperados e causam diversas perdas. Existem diversas variáveis que contribuem com a gravidade de um acidente. O intuito deste estudo é desenvolver um modelo que ajude na predição do chamado em casos de acidentes em vias terrestres.

O trabalho envolve também, a partir de uma base de dados, determinar um sistema de apoio à decisão, que a partir dos dados coletados, possui capacidade de apontar caminhos a serem seguidos. E, com base na experiência do domínio do problema, o stakeholder pode definir as decisões a serem tomadas.

Além disso, houve um avanço do conhecimento a partir da noção de IA explicável. E, com esse objetivo, usamos três métodos para, além de desenvolver um bom sistema, entender com propriedade o que está lá dentro, que determina o porquê e quais caminhos os algoritmos estão tomando

II. BASE DE DADOS

A base de dados vieram da junção das bases de acidentes de trânsito [1] ocorridos no Condado de Montgomery - Maryland, EUA e das informações dos motoristas envolvidos neste acidente [2].

Estas informações foram registradas pelo “Sistema automatizado de Relatórios de Acidentes da Polícia estadual de Maryland”.

A. Escopo e Seleção dos Dados

Como definido pela base de dados, o escopo é dado apenas pelos acidentes de trânsito que ocorreram no Condado de Montgomery. Nele, temos informações sobre estado da via, trânsito, clima e detalhes sobre o impacto (caso tenha ocorrido). Entretanto, grande parte desses dados são informações a posteriori, se fazendo necessário então a remoção dessas colunas.

B. Definição do Objetivo

Com o conhecimento gerado a partir dos modelos treinados é possível aplicar a predição em diversos segmentos do mercado, como por exemplo: Seguradoras, empresas de locação

de veículos, planos de saúde e outros prestadores de serviços relacionados a carro ou a saúde do condutor.

C. Pré Processamento dos Dados

Inicialmente a ideia girava em torno de tentar prever a gravidade do acidente, qual a chance de existirem vítimas graves (sendo esse o alvo binário). Neste sentido, foram encontradas algumas dificuldades.

Com a junção dos datasets, foram obtidas 77 colunas de atributos. Dentre essas colunas, 51 eram de valores que precisariam ser removidos (dados a posteriori) ou tratados, como por exemplo colunas com muitos valores nulos. Considerando também as restrições de captação de dados do veículo, foram incluídos apenas os atributos mais significativos e pertinentes para a análise, enquanto outras colunas foram agrupadas, para melhor organizar os dados, restando 26 features para serem analisadas.

A principal técnica utilizada para tratamento foi a criação de variáveis “dummies” [3]. Variáveis categóricas ou variáveis binárias foram alteradas para variáveis “flags”. Um bom exemplo de uso dessa técnica é a coluna de local de impacto: variável categórica que indicava a posição do impacto no veículo. Possuía valores baseados na técnica militar que usa o relógio para indicar posição, ou seja, 9 horas, 1 horas, 6 horas. Agrupando esta variável, chegamos aos valores “colisão frontal”, “colisão lateral”, “colisão traseira” ou “outro”. Assim, reduzimos de 15 (cima, baixo e desconhecido também eram valores) possíveis valores para apenas 4.

Por recomendações do professor, evitamos deixar variáveis com valores ordenados, pois para o modelo pode passar a ideia de “prioridade” ou de “ordem”, tornando a coluna enviesada. Sendo assim, usamos a técnica de variáveis “dummies” [3] e experimentamos usar a técnica “Frequency Encoding” [4], onde codificamos uma variável categórica de forma que seu valor passe a ser representado pela frequência de vezes em que o seu valor aparece no conjunto de dados.

D. Definição do Alvo

No primeiro momento, o alvo binário tratava-se de prever se havia chance de ter vítimas graves ou não em um acidente. Entretanto, devido a dificuldade em obter bons resultados com o dataset, foi estudada a possibilidade de trocar para a coluna de dano ao veículo e o impacto que isso teria na ideia principal do projeto.

O alvo da classificação binária ficou então definido em relação ao dano do veículo, sendo então o objetivo descobrir se houve ou não dano significativo ao mesmo.

III. EXTRAÇÃO DE DADOS, RESULTADOS E DISCUSSÃO

Para extrair o conhecimento inserido na base de dados, foram utilizados três métodos: regressão logística, árvore de decisão, e indução de regras

A. Regressão Logística

Após treinar o modelo de regressão logística, foram analisadas as features com maior coeficiente beta, e que possuísem maior significância de acordo com o p-valor, onde os coeficientes de maior módulo tem mais relevância ao definir a classe alvo.

As features com maior valor positivo tem maior contribuição para definir que houve dano significativo ao veículo, enquanto as de valores mais negativos possuem uma importância maior para definir se não houve dano significativo.

Se o carro está se movendo ou é particular há uma maior chance de possuir dano significativo após o acidente, enquanto se tiver algum não motorista participando do acidente (pedestre ou ciclista), se o carro estiver acelerando ou se a colisão for na mesma direção, a probabilidade de não haver dano significativo é bem menor.

TABLE I: características mais Relevantes de Acordo com A Regressão Logística

Feature	Beta	p-valor
is_moving	1.259	0
is_particular	1.255	0
RelatedNon-Motorist	-2.652	0
is_accelerating	-1.212	0
CollisionType=SameDir	-1.175	0

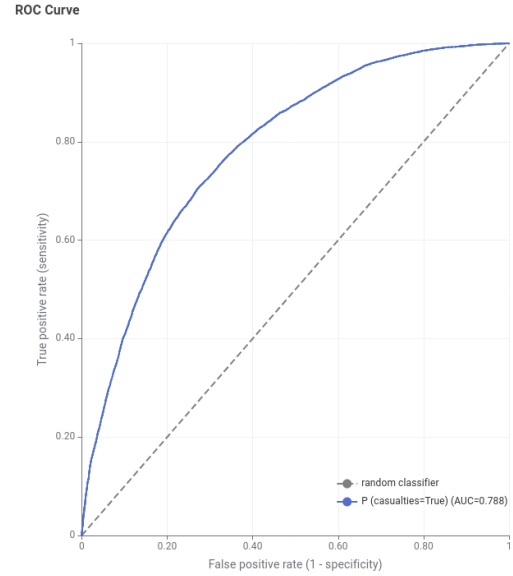


Fig. 1: Curva ROC Para o Regressor Logístico

B. Árvore de Decisão

A árvore de decisão é um dos modos mais simples de visualizar o conhecimento presente em uma base de dados de modo compreensível. As variáveis mais importantes para definir a classe alvo foi principalmente o tipo de colisão, sendo a colisão em ângulo a com maior probabilidade de causar dano ao veículo. Outras características importante foi a divisão da estrada.

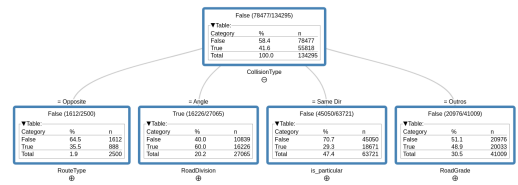


Fig. 2: Árvore de decisão gerada usando o Knime

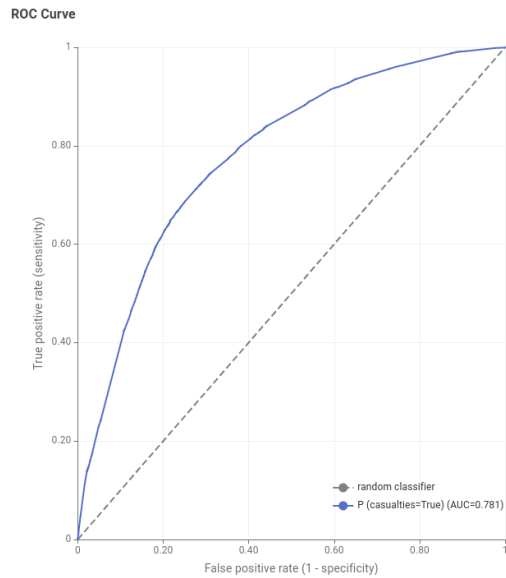


Fig. 3: Curva ROC Para a Árvore de Decisão

C. Indução de Regras de Classificação

TABLE II: Indução de Regras Com o Algoritmo RIPPER

Feature	Cobertura	Confiança	Lift
---------	-----------	-----------	------

TABLE III: labeltab:tab2 HERE

IV. CONCLUSÃO

REFERENCES

- [1] M. Montgomery County. (2024) Crash reporting - incidents data. [Online]. Available: https://data.montgomerycountymd.gov/Public-Safety/Crash-Reporting-Incidents-Data/bhju-22kf/about_data
- [2] —. (2024) Crash reporting - drivers data. [Online]. Available: https://data.montgomerycountymd.gov/Public-Safety/Crash-Reporting-Drivers-Data/mmzv-x632/about_data
- [3] S. Date. What are dummy variables and how to use them in a regression model. [Online]. Available: <https://timeseriesreasoning.com/contents/dummy-variables-in-a-regression-model/>
- [4] M. Halford. (2018, Oct.) Target encoding done the right way. [Online]. Available: <https://maxhalford.github.io/blog/target-encoding/>
- [5] W. H. Organization. (2024, Dec.) Road traffic injuries. [Online]. Available: <https://www.who.int/news-room/fact-sheets/detail/road-traffic-injuries>