

APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA NA ANÁLISE DE OCORRÊNCIAS DE TRÂNSITO DE BELO HORIZONTE - MG

MACHINE LEARNING APPLICATION IN THE ANALYSIS OF TRAFFIC OCCURRENCES IN BELO HORIZONTE – MG

MARTINS, Isaac Emanuel Santos¹; DE ANDRADE, Michelle Hanne Soares²

PALAVRAS-CHAVE

Trânsito. Acidentes.
Aprendizado de Máquina. Previsão.

KEYWORDS

Traffic. Accidents. Machine Learning.
Prediction.

RESUMO

Dados da prefeitura de Belo Horizonte do ano de 2019 indicam que o total de vítimas fatais apresentou queda em relação a 2018, o que reforça o cenário de redução nos últimos nove anos. Dados de boletins de trânsito da cidade de Belo Horizonte, apontam uma redução de 51,61% nas ocorrências de acidentes com vítimas fatais de 2011 a 2019. Esta redução está possivelmente relacionada com resultados de ações governamentais, como: campanhas de conscientização, melhorias das vias, aumento da fiscalização, etc. Este estudo objetiva analisar as ocorrências de trânsito da cidade de Belo Horizonte entre os anos de 2011 a 2019, suas causas, taxa de mortalidade, condições de sinalização das vias, assim como a aplicação de técnicas de Estatística e Aprendizado de Máquina, para a geração de modelos de previsão sobre o óbito em casos de acidentes de trânsito.

ABSTRACT

Data from Belo Horizonte city hall in 2019 indicate that the total number of fatalities fell in relation to 2018, which reinforces the scenario of reduction in the last nine years. Data traffic announcements in the city of Belo Horizonte, point to a reduction of 51.61% in the occurrence of accidents with fatalities from 2011-2019. This reduction is possibly related to government actions, such as: awareness campaigns, improvements in ways, increased inspection, etc. This study aims to analyse the traffic accidents in the city of Belo Horizonte in the years 2011-2019, its causes, mortality rate, the pathways signalling conditions, as well as application techniques of Statistics and Machine Learning for generating predictive models of death in cases of traffic accidents.

¹ Graduando do curso de Sistemas de Informação do Centro Universitário Newton Paiva (isaacemanuel2@gmail.com).

² Professora do Centro Universitário Newton Paiva (michelle.andrade@newtonpaiva.br).

1 INTRODUÇÃO

As ocorrências de trânsito representam um problema global de saúde pública, demandando esforços contínuos dos órgãos públicos governamentais na prevenção eficaz e sustentável de medidas de combate a este problema.

De acordo com Damacena *et al.* (2016, p.30, *apud* OMS, Beber e dirigir: Manual de Segurança de Trânsito para Profissionais de Trânsito e Saúde, 2007) a cada ano, cerca de 1,2 milhões de pessoas morrem e milhões sofrem lesões ou se tornam deficientes em decorrência de acidentes de trânsito, principalmente em países de baixa e média renda. Os resultados de tantos acidentes de trânsito, geram uma sobrecarga pesada aos serviços de saúde e a economia das nações, além de limitações que as lesões causam aos indivíduos envolvidos.

Segundo ainda Damacena *et al.* (2016, p.30, *apud* OMS, Relatório Global Sobre Segurança Rodoviária: Hora da Ação, 2009) os acidentes de trânsito, em 2004, ocupavam a nona posição no ranking de causa mundial, e até 2030, deverão ocupar a quinta posição, ficando atrás apenas de doenças cardíacas, cerebrovasculares, pulmonares crônicas, e infecções respiratórias inferiores.

Segundo a Secretária de Estado de Saúde de Minas Gerais (2020), as ocorrências de violência e acidentes no trânsito, são responsáveis por cerca de 9% da mortalidade global, e na maior proporção dos acidentes terrestres, estão envolvidos jovens adultos do sexo masculino, residentes em áreas urbanas. Ainda, segundo a Secretária de Estado de Saúde de Minas Gerais (2020), entre os anos de 2010 a 2020, o maior quantitativo de internações por acidentes de trânsito foi de motociclistas com 47,7% dos casos, seguido de 21,3% de pedestres e ocupantes de automóveis com 11,7%.

Com isto, neste estudo, serão realizadas análises, por meio de técnicas de Estatística Aplicada em ocorrências de trânsito da cidade de Belo Horizonte, entre os anos de 2011 a 2019, com o objetivo de criar modelos de Aprendizado de Máquina capazes de gerar previsões sobre a mortalidade nos acidentes de trânsito. Os dados estão disponíveis no site da Prefeitura de Belo Horizonte, e dispõe dos boletins de ocorrências de trânsito com ou sem vítimas.

1.1 OBJETIVO GERAL

Desenvolver modelos de Aprendizado de Máquina capaz de prever a morte ou não de envolvidos em acidentes de trânsito. Para isso, foram utilizadas análises Estatísticas, como a descritiva, inferência e correlação entre as variáveis. A base de dados utilizada refere-se a todos os boletins de ocorrência de trânsito registrados na cidade de Belo Horizonte entre os anos de 2011 a 2019.

1.1.2 OBJETIVOS ESPECÍFICOS

Dentre os objetivos específicos destacam-se:

- Analisar descritivamente os dados de ocorrência de acidentes sem vítimas, com vítimas fatais e suas particularidades;
- Aplicar técnicas de Estatística para gerar visualizações, indicadores e apuração de resultados sobre ocorrências de trânsito;

1.2 ACIDENTES DE TRÂNSITO NO BRASIL

De acordo com o Pereira; Neves (2013, *apud* Denatran, 2011) o Código de Trânsito Brasileiro (CTB): *“Considera trânsito como a utilização das vias por pessoas, veículos e animais, isolados ou em grupos, conduzidos ou não, para fins de circulação, parada, estacionamento e operação de carga e descarga.”* Ainda, conforme Pereira; Neves (2013) os estudos sobre ocorrências de trânsito no Brasil ainda são escassos. E os que existem, pouco mostram a respeito do comportamento tanto do motorista quanto do pedestre.

Dados do Conselho Nacional de Trânsito (CNT) sobre a Pesquisa de Acidentes Rodoviários em 2019 mostram que em Minas Gerais houve redução de acidentes de 2014 a 2019. Em 2018, foram registrados 9.052, e em 2019, 8.716 nas rodovias federais que cortam Minas Gerais. Somente em 2019, foram 677 vidas perdidas.

Estima-se que o custo anual dos acidentes ocorridos em rodovias federais em Minas Gerais seja de R\$ 1,33 bilhão em 2019 (CNT, 2019).

Em um estudo realizado em 2012, na cidade de Belo Horizonte, buscou-se verificar, os níveis de álcool no sangue de 1.656 condutores através de blitz em pontos estratégicos em cada uma das 9 regionais (Norte, Centro-Sul, Leste, Oeste, Nordeste, Noroeste, Venda Nova, Pampulha e Barreiro) das 23h às 3h. Dos 1.656 condutores, 1.254 (76%) aceitaram responder um questionário e fazer o teste do bafômetro, o teste foi positivo para 15% dos condutores (CAMPOS; SALGADO; ROCHA, 2012).

Estudo realizado no estado do Pernambuco, com dados de ocorrências de acidentes de trânsito disponibilizados pela Polícia Federal, desenvolveu um sistema de Aprendizado de Máquina para prever acidentes. O sistema foi treinado com dados de ocorrências de acidentes, e posteriormente gerado uma ferramenta interativa de visualização dos pontos com maiores riscos no mapa de Pernambuco. Para o sistema, foram aplicados modelos de regressão para a predição do número de acidentes dado a identificação da rodovia, altura do trecho, ano, e condições meteorológicas (SOUSA; ARAUJO; AZEVEDO, 2019).

2 METODOLOGIA

Inicialmente, foi realizada análise exploratória dos dados, com o objetivo de compreender e aprofundar o estudo sobre os dados de boletins de acidentes de trânsito da cidade de Belo Horizonte. Posteriormente, foram aplicados modelos de Aprendizado de Máquina para a previsão de morte/não morte em acidentes de trânsito.

2.1 BASE DE DADOS

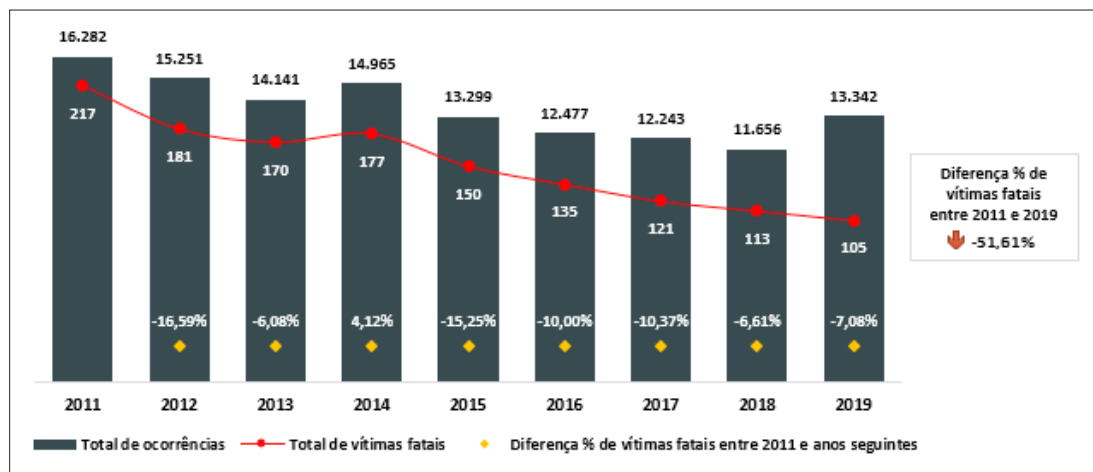
Para realização deste estudo, primeiramente foram planejadas três etapas que seriam de suma importância para o correto entendimento e refinamento dos dados, que foram: a organização, a padronização e a análise das informações, contemplando o conceito de ETL (Extração, Transformação e Carregamento - ETL, do inglês *Extract Transform Load*).

Os dados para a pesquisa, foram extraídos do Banco de Dados público da Prefeitura de Belo Horizonte (2019), na área de acesso à informação para o cidadão.

2.2 ANÁLISE EXPLORATÓRIA

Com base na análise e consolidação dos dados, percebe-se um decréscimo no número de ocorrências na cidade de Belo Horizonte, de um total de 16.282 ocorrências em 2011 para 13.342 em 2019. O percentual da diferença de vítimas fatais decresceu 51,61% entre 2011 e 2019 como mostrado no Gráfico 1.

GRÁFICO 1: Total de Ocorrências e Vítimas Fatais entre 2011 e 2019.



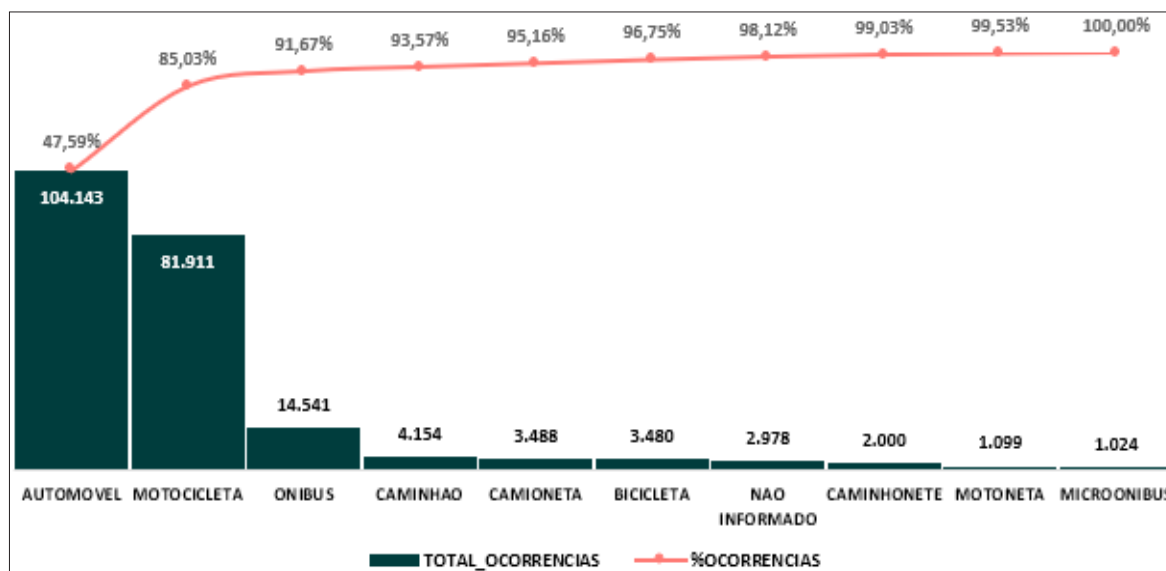
Fonte: Montenegro; Martins (2020).

As ocorrências com vítimas fatais estão em sua maioria relacionadas ao sexo masculino. Um total de 166 óbitos de pessoas do sexo masculino em 2011 para 51 de pessoas do sexo feminino no mesmo ano, e um total de 18 óbitos de pessoas do sexo feminino em 2019, para 87 do sexo masculino. O percentual da diferença de vítimas fatais do sexo feminino entre os anos de 2011 a 2019, teve uma redução de 64,71%, enquanto que nas vítimas do sexo masculino houve redução de 47,59%.

A análise efetuada sobre a taxa de mortalidade por tipo de ocorrência, apontou o número de mortalidade maior para os acidentes de “Atropelamento de Pessoa” com 485 óbitos que equivale a 35,43%. Já o menor índice ficou com o tipo “Queda de Pessoa de Veículo” com um total de 18 óbitos equivalente a 1,31%. Já o acidente do tipo “Abalroamento” (“Abalroamento é uma colisão leve nas laterais dos veículos. Em geral acontece em cruzamentos ou devido a manobras inesperadas” (APROVA DETRAN, 2020)), teve um total de 245 óbitos, equivalente a 17,90% entre os anos de 2011 a 2019.

O Gráfico 2 identifica quais os tipos de veículos que mais se envolveram em acidentes entre os anos de 2011 a 2019. O veículo automóvel possui um total de 104.143 ocorrências em acidentes. O tipo de veículo com o menor índice foi o micro-ônibus com um total de 1.024 ocorrências em acidentes.

GRÁFICO 2: Total de Ocorrências por Tipo de Veículo entre 2011 e 2019.



Fonte: Montenegro; Martins (2020).

Os acidentes avaliados de 2011 a 2019, encontram-se distribuídos em 10 regionais da cidade de Belo Horizonte. A regional Centro-Sul foi a que registrou o maior número de ocorrências com um total de 23.624, equivalente a 19,10%. A regional que registrou o menor número foi a de Venda Nova com um total de 8.920 (7,21%) ocorrências.

Ao relacionar as variáveis das condições climáticas, não foi encontrado nenhuma relação da causa da ocorrência de acidentes com fatores climáticos ruins, sendo que 48,87% não informaram a característica climática. Porém, ao analisar os dados, verificou-se o maior percentual de ocorrências de trânsito em condições climáticas estáveis, as quais se passaram em vias não sinalizadas.

3 TÉCNICAS DE APRENDIZADO DE MÁQUINA

O Aprendizado de Máquina pode ser classificado em três categorias básicas: aprendizagem supervisionada, aprendizagem não supervisionada e aprendizagem por reforço. No aprendizado supervisionado os exemplos são rotulados. O aprendizado é dito supervisionado pois, além de conter a entrada (vetor de atributos), possui

também o rótulo ou saída (classe). Os exemplos rotulados são repassados ao algoritmo, que constrói através do treinamento, uma função de mapeamento que tem como objetivo determinar corretamente a classe de novos exemplos não rotulados (SANTOS 2016, *apud* KONAR, 1999).

Neste estudo, foram utilizados três algoritmos supervisionados de classificação: **Árvore de Decisão** (*Decision Tree*), **Floresta Randômica** (*Randon Forest*) e **Regressão Logística** (*Logistic Regression*).

Segundo Dantas; Donadia (2013, *apud* Breiman *et al.*, 1984) Árvores de decisão configuram métodos que utilizam uma apresentação gráfica baseada em árvores, cujo objetivo é identificar grupos de indivíduos com características de interesse em comum.

Já o *Random Forest* é um método que combina a ideia de *Bagging* (uma técnica usada para reduzir a variância das previsões, combina o resultado de vários classificadores, modeladores em diferentes subamostras do mesmo conjunto de dados (ADMINVOO, 2016)) com a seleção randômica de variáveis explicativas no processo de indução da árvore. (Dantas; Donadia, 2013 *apud* Breiman, 2001).

O modelo paramétrico de Regressão Logística, segundo Dantas; Donadia (2013, *apud* Mccullagh; Nelder, 1989), trata-se de um caso particular dos modelos lineares generalizados, definido pela distribuição binomial com função de ligação canônica (*logit*) apropriado para a modelagem de resposta binária ou categórica.

Nesta análise utilizou-se o software *Orange Data Mining*, que possibilitou a unificação das bases de dados, filtragem das colunas e triagem das informações. Além do uso de algoritmos de Aprendizado de Máquina.

No conjunto de dados tem-se 13 atributos (*Features*), e o atributo *Target* sendo a DECLARACAO_OBITO. Assim, a *Target* será SIM ou NÃO, o que significa MORTE ou NÃO MORTE, conforme Tabela 1.

Entre os anos de 2011 a 2019 temos um total de número de vítimas fatais (Morte) de 1.369 pessoas, sendo que mais de 24 mil vítimas envolvidas em acidentes estão relacionadas com a *Target* NÃO (Não morte).

TABELA 1: Lista de Atributos e sua descrição

DESC_TIPO_ACIDENTE	Contém todos os tipos de acidentes
DESC_TEMPO	Informa a descrição do tempo
PAVIMENTO	Contém o tipo do pavimento da via
DESC_REGIONAL	Informa o nome da regional
LOCAL_SINALIZADO	Informa se o local é sinalizado ou não
VELOCIDADE_PERMITIDA	Informa a velocidade permitida da via
DESC_UPS	Informa se o acidente teve ou não vítima fatal
CONDUTOR	Informa quem causou a ocorrência do trânsito se foi o condutor ou não
SEXO	Informa o sexo da pessoa envolvida na ocorrência de trânsito
CINTO_SEGURANCA	Informa se o envolvido estava ou não de cinto de segurança
EMBREAGUES	Informa se o envolvido estava ou não embriagado
DECLARACAO_OBITO	Informa se houve morte ou não na ocorrência de trânsito
ESPECIE_VEICULO	Informa o tipo do veículo envolvido na ocorrência de trânsito
PEDESTRE	Informa se houve ou não algum pedestre envolvido na ocorrência de trânsito
PASSAGEIRO	Informa se teve ou não algum passageiro envolvido na ocorrência

Fonte: o próprio autor.

Visando equilibrar a base de dados, foi realizada a separação aleatória de 3.000 registros com a *target* NÃO, desse modo, os algoritmos apresentados foram executados em uma base de dados totalizando 4.369 registros.

A técnica de treinamento utilizada nos algoritmos supervisionados foi a Validação Cruzada ou *Cross*

Validation. “A validação cruzada é um método de reamostragem e tem como objetivo avaliar a capacidade de generalização do seu modelo. Em outras palavras, verificar o quão pronto seu modelo está para receber novos dados” (MELGAREJO, 2020).

As métricas de apuração dos resultados utilizadas foram: AUC (*Area Under Curve*), CA (*Classification Accuracy*), F1, *Precision* (Precisão) e *Recall* (Revocação). Segundo Degani (2019), a métrica CA é utilizada para medir a performance geral do modelo analisado. A *Recall*, mede a capacidade do modelo para encontrar exemplos corretamente classificados. Já a *Precision* mede o quão preciso o modelo previu os valores positivos. A métrica F1 faz a combinação das métricas *Precision* e *Recall*, utilizada para avaliar conjuntos de dados desproporcionais do modelo analisado.

Conforme Rodrigues (2018), a métrica AUC, nada mais é que uma maneira de resumir a curva ROC (*Receiver Operating Characteristic* - Curva Característica de Operação do Receptor) em um único valor, agregando todos os similares da ROC, calculando a “área sob a curva”.

3.1 RESULTADOS

A falta de correlação dos atributos, identificado na análise exploratória, sinalizam uma não linearidade dos dados. Atributos como condições climáticas não interferiram diretamente nos acidentes com morte. Porém, foi identificado que a falta de sinalização das vias pode contribuir para os acidentes com morte. Com isso, foram usados todos os atributos na classificação.

Após a execução dos três algoritmos: Árvore de Decisão (*Decision Tree*), Floresta Randômica (*Random Forest*) e Regressão Logística, obteve-se os resultados apresentados na Tabela 2. Percebe-se que as métricas Acurácia, F1, *Precision* e *Recall* são similares entre os três métodos, somente a AUC (Área sobre a Curva) é que difere, mostrando que mesmo com resultados próximos, o método de Regressão Logística foi o mais assertivo e eficiente nos resultados.

TABELA 2: Comparação entre Métricas dos Algoritmos de Aprendizado de Máquina

Modelo	AUC	CA	F1	<i>Precision</i>	<i>Recall</i>
Árvore de Decisão	0,994	0,994	0,994	0,994	0,994
Floresta Randômica	0,998	0,994	0,994	0,994	0,994
Regressão Logística	0,999	0,994	0,994	0,994	0,994

Fonte: o próprio autor.

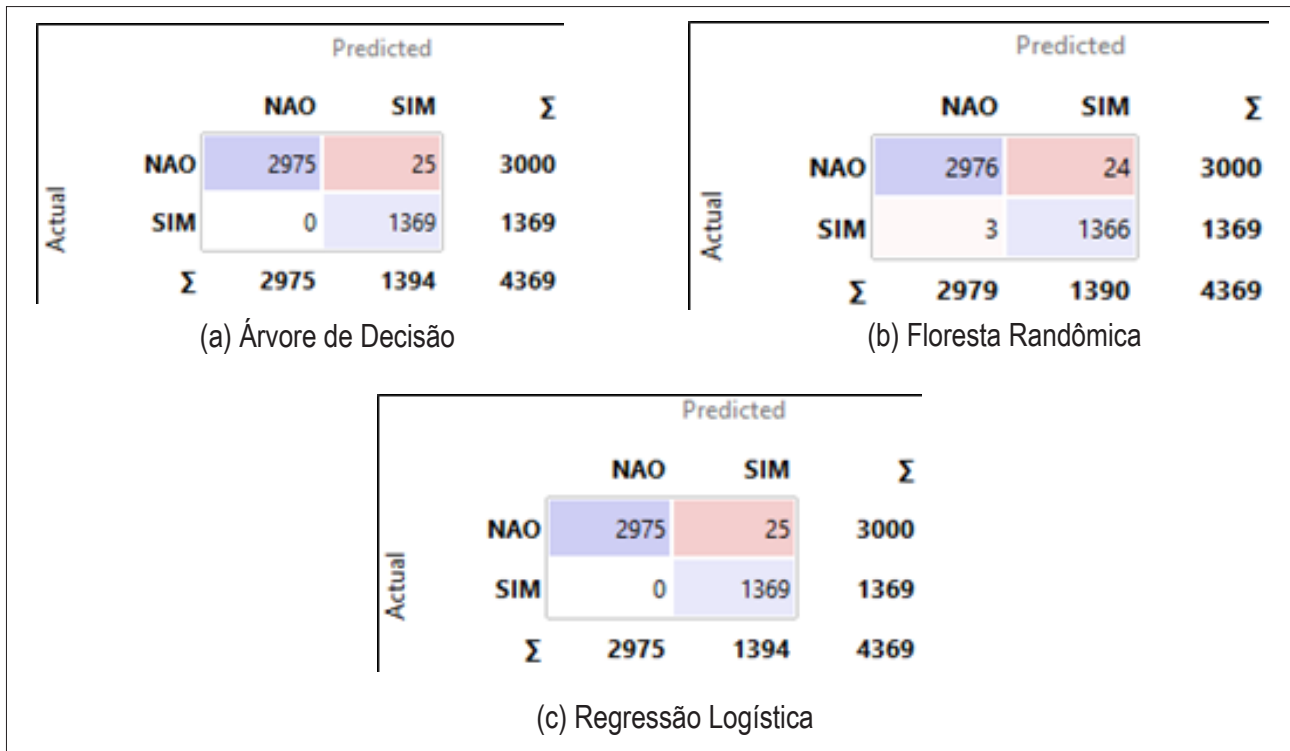
Uma métrica comumente adota na análise de algoritmos de classificação é a denominada Matriz de Confusão. A Figura 1 mostra as três matrizes geradas nos experimentos dos algoritmos.

No método Árvore de Decisão, o algoritmo classificou corretamente com a *Target* “NÃO”, 2.975 pessoas que realmente não tiveram óbito, e errou 25 classificações. Já para a *Target* “SIM”, o algoritmo classificou corretamente todos os registros, ou seja, acertou todas as previsões de óbito.

No método Floresta Randômica, o algoritmo classificou corretamente com a *Target* “Não”, 2.976 pessoas que realmente não tiveram óbito, e errou 24 classificações. As previsões para a *Target* “SIM”, tiveram 3 registros classificados erroneamente e 1.366 classificados corretamente.

Já o método Regressão Logística teve o seu comportamento na matriz de confusão similar ao método Árvore de Decisão. Já a Curva ROC teve ligeira sensibilidade em função da proporção de falsos positivos com relação aos métodos Árvore de Decisão e Floresta Randômica.

FIGURA 1: Matriz de Confusão dos Métodos Supervisionados de Classificação



Fonte: o autor.

Após a execução e análise dos três métodos supervisionados, foi executada a predição com 100 registros escolhidos aleatoriamente da base de dados de acidentes de trânsito. Retirou-se dos dados selecionados a *Target* **DECLARACAO_OBITO**, e posteriormente efetuou-se a predição com os modelos: Árvore Randômica e Regressão Logística, previamente executados em ambiente de treinamento e teste. O resultado apresentou 97% de acurácia para o *Random Forest* e 94% de acurácia para o *Logistic Regression*. Demonstrou-se assim, a possibilidade de prever a morte ou não em caso de acidentes de trânsito, utilizando os dados extraídos de boletins de ocorrência.

4 CONCLUSÃO

É provável que a atuação de entidades governamentais na prevenção de acidentes ao longo do tempo, tenha contribuído para a redução de vítimas fatais, de 2011 a 2019.

Considerando que o objetivo deste trabalho foi analisar as ocorrências de trânsito na cidade de Belo Horizonte, constatou-se que as bases de dados disponibilizadas pela Prefeitura de Belo Horizonte apresentam informações incompletas ou faltantes, no que tange os dados de boletins de ocorrência. Como por exemplo: número elevado de idades fornecidas iguais a zero e campos com informações relevantes não preenchidas. Após os devidos tratamentos e uso de técnicas de Estatística, apurou-se a falta de correlação direta entre os atributos para prever o óbito de uma vítima de acidente. O que reflete nos modelos apresentados, que ao utilizar todos os atributos mostrou-se satisfatório na tarefa de classificação supervisionada. Assim, conclui-se que as condições climáticas isoladamente não interferiram nos acidentes com morte, porém, a falta de sinalização pode ser um fator que interferiu nos acidentes com vítimas fatais. Além de, notoriamente, o maior número de ocorrências serem causados por pessoas do sexo masculino.

Percebeu-se ainda, que é necessário um maior estudo na área de prevenção de acidentes, para que mais ações preventivas possam ser implementadas em regionais específicas na cidade Belo Horizonte. Por isso, os próximos passos deste estudo estão relacionados com o estudo aprofundado de correlação entre as variáveis e a previsão de locais (bairro e rua) que teriam maior probabilidade de acidentes com vítimas fatais.

Concluindo, o presente estudo, pode colaborar para um planejamento de políticas públicas visando a prevenção de acidentes de trânsito.

REFERÊNCIAS

- ABREU AMM.; LIMA JMB.; GRIEP RH. Acidentes de trânsito e a frequência dos exames de alcoolemia com vítimas fatais na cidade do Rio de Janeiro. Esc Anna Nery Rev Enfermagem. Rio de Janeiro, v.13, n.1, jan/mar. 2009. Disponível em: <<http://www.scielo.br/pdf/ean/v13n1/v13n1a07.pdf>>. Acesso em: 01 out. 2020.
- ADMINVOO. Um tutorial completo sobre modelagem baseada em árvores de decisão (códigos R e Python), dez.2016. Disponível em: <<https://www.vooo.pro/insights/um-tutorial-completo-sobre-a-modelagem-baseada-em-tree-arvore-do-zero-em-r-python/>> Acesso em: 08 dez. 2020.
- ANDRADE, SM.; SOARES, DA.; BRAGA, GP.; MOREIRA JH; BOTELHO, FMN. Comportamentos de risco para acidentes de trânsito: um inquérito entre estudantes de medicina na região sul do Brasil. Rev Assoc Med Bras. Londrina, v.49, n.4, p.439-44, ago.2003. Disponível em: <<http://www.scielo.br/pdf/ramb/v49n4/18346.pdf>>. Acesso em: 01 out. 2020.
- APROVA DETRAN. O que é e como evitar um abalroamento? 2020. Disponível em: <<https://www.aproवादetran.com.br/direcao-defensiva/colisoes-no-transito> > Acesso em: 02 out. 2020.
- CAMPOS VR.; SALGADO RS.; ROCHA MC. Bafômetro positivo: correlatos do comportamento de beber e dirigir na cidade de Belo Horizonte, Minas Gerais, Brasil em 2012, 2020. Disponível em: <<https://www.scielo.org/article/csp/2013.v29n1/51-61/>> Acesso em: 01 out. 2020.
- CNT/SEST SENAT. Pesquisa CNT de rodovias 2019. Brasília, p. 236, 2019. Disponível em <<https://pesquisarodovias.cnt.org.br/downloads/ultimaversao/gerencial.pdf>>. Acesso em: julho 2020.
- DAMACENA et al. Consumo Abusivo de Álcool e Envolvimento em Acidentes de Trânsito na População Brasileira em 2013, 2020. Disponível em: <<https://www.scielo.org/article/csc/2016.v21n12/3777-3786/pt/#>> Acesso em: 30 set. 2020.
- DANTAS D.; DONADIA EA. Comparação entre as técnicas de regressão logística, árvore de decisão, bagging e random forest aplicadas a um estudo de concessão de crédito, 2013. Disponível em: <http://www.coordest.ufpr.br/wp-content/uploads/2018/12/TCC_DanielEricson.pdf>. Acesso em: 03 dez. 2020.
- DEGANI GL. Análise de sentimentos dos clubes paranaenses de futebol na rede social Twitter, 2019. Disponível em: <<https://acervodigital.ufpr.br/bitstream/handle/1884/64986/GABRIEL%20LUIS%20DEGANI.pdf?sequence=1&isAllowed=y>>. Acesso em: 06 dez. 2020.
- MELGAREJO JL. Validação Cruzada (Cross Validation) - Avaliando seu modelo de forma clara e concisa, 2020. Disponível em: <<https://medium.com/@joorge.lm/valida%C3%A7%C3%A3o-cruzada-cross-validation-avaliando-seu-modelo-de-forma-clara-e-concisa-10d27fa4efb8>> Acesso em: 05 dez. 2020.
- MONTENEGRO, Bruna; MARTINS, Isaac. Estudo sobre acidentes de trânsito no município de Belo Horizonte - MG, 2020. Disponível em <<http://ormuztech.com.br/blog/?p=677>> Acesso em 07 dez. 2020.
- PEREIRA, I. G.; NEVES, F. S. Comportamentos de Risco: Uma Revisão Bibliográfica Sobre as Causas de Acidentes de Trânsito nos Últimos 10 Anos, 2020. Disponível em: <<http://ojs.unirg.edu.br/index.php/1/article/view/429/176>> Acesso em: 29 set. 2020.
- RODRIGUES V. Entenda o que é AUC e ROC nos modelos de Machine Learning, 2018. Disponível em: <<https://medium.com/bio-data-blog/entenda-o-que-%C3%A9-auc-e-roc-nos-modelos-de-machine-learning-8191fb4df772>> Acesso em: 06 dez. 2020.
- SANTOS RMM. Técnicas de Aprendizado de Máquina Utilizadas na Previsão de Desempenho Acadêmico, 2016. Disponível em: <http://acervo.ufvjm.edu.br/jspui/bitstream/1/1327/1/rodrigo_magalhaes_mota_santos.pdf>. Acesso em: 03 dez. 2020.
- SECRETARIA DE ESTADO DE SAÚDE DE MINAS GERAIS. Vida no trânsito em 2020, 2020. Disponível em: <<https://www.saude.mg.gov.br/vidanotransito>> Acesso em: 29 set. 2020.
- SOUSA, R. S; ARAUJO, D. R. B; AZEVEDO, V. M; Um Sistema para Predição de Risco de Acidentes em Rodovias de Pernambuco, 2019. Disponível em: <<http://revistas.poli.br/index.php/rep/article/view/1328/608>> Acesso em: 05 out. 2020.