APLICAÇÃO DE TÉCNICAS DE APRENDIZADO DE MÁQUINA PARA CLASSIFICAÇÃO DO RISCO DE ACIDENTES EM TRECHOS RODOVIÁRIOS COM BASE NO HISTÓRICO DE ACIDENTES, CONDIÇÕES DA VIA E ÍNDICES DE CHUVA E NEBULOSIDADE

Tarcisio Costa Brum
Centro Universitário Estácio Juiz de Fora
Faculdades Integradas Vianna Junior
Afonso Celso de Castro Lemonge
Universidade Federal de Juiz de Fora
Departamento de Mecânica Aplicada e Computacional
Priscila Vanessa Zabala Capriles Goliatt
Universidade Federal de Juiz de Fora
Departamento de Ciência da Computação

RESUMO
O Brasil é um país que possui altas taxas de acidentes rodoviários. Uma forma de reduzir o número de acidentes é diminuir a exposição ao risco, como por exemplo, selecionar rotas que possuam menor probabilidade de acidentes. Face ao problema descrito, este trabalho propõe a aplicação de técnicas de aprendizado de máquina para predição do risco de acidentes em trechos rodoviários, com base no histórico de acidentes, condições da via e índices de chuva e nebulosidade. Utilizou-se como referência a malha rodoviária do estado de Minas Gerais. Os dados foram analisados através da aplicação do método PCA e os resultados foram medidos em termos da curva ROC, classificando com 80% de acerto trechos rodoviários em termos de risco de acidente. Este resultado mostra que há espaço para melhoria desta proposta de trabalho e a possibilidade de utilizar técnicas de regressão como uma alternativa aos métodos de classificação.

ABSTRACT
Brazil is a country that has high rates of road accidents. One way to reduce it is to decrease the risk of exposure, for example, to select routes that have a lower probability of accidents. In this point of view, this work proposes the use of machine learning techniques to predict the risk of accidents in road sections, based on the history of accidents, road conditions, and rainfall and cloudiness. The road network of the state of Minas Gerais was used as the reference. The data were analyzed by PCA method application and ROC curves were used to evaluate the quality of the results, classifying the road sections in terms of accident risk with an accuracy of approximately 80%. The results showed that the approach presented was considered promising and could be better explored, as well as the possibility of using regression techniques as an alternative to classification methods.

1. INTRODUÇÃO

O Brasil é um país que enfrenta altas taxas de acidentes rodoviários e consequentemente altas taxas de mortalidade. Um estudo divulgado com base em dados da Organização Mundial de Saúde (WHO, do inglês *World Health Organization*) (Sivak e Schoettle, 2014) mostra que o país ocupa a 42ª posição, com 22 mortes para cada 100 mil pessoas, em ranking que analisou a quantidade de acidentes de trânsito em 193 países. O relatório publicado pela WHO (2015) aponta tendência de aumento no número de mortes nas rodovias brasileiras desde 2009, ao contrário de muitos países, em especial os desenvolvidos, onde há queda desta tendência. Os dados da WHO (2015) mostram que em média, os acidentes de trânsito matam 43 mil pessoas por ano, representando uma das principais causas de morte no Brasil, sendo que cerca de 20% dessas mortes é oriunda de acidentes ocorridos em rodovias federais.

Existem muitos fatores que contribuem para a ocorrência de acidentes de trânsito. Naing *et al.* (2007) agrupam estes fatores em: (i) humanos (comportamento e ações pessoais); (ii) viário-ambientais (condições da via e/ou do ambiente); (iii) veiculares (forma do veículo ou falhas

mecânicas); (iv) institucionais (leis, fiscalização e investimentos em segurança, por exemplo). Muitas ações propõem e buscam formas de reduzir os índices de acidentes nas estradas ou mitigar os riscos. Lord e Washington (2018) destacam os novos e futuros desafios na mobilidade urbana, como as rodovias sustentáveis, carros autônomos e monitoramento das rodovias, e a necessidade de serem desenvolvidas novas técnicas e ferramentas voltadas à segurança das vias.

Goniewicz *et al.* (2016) e Elvik (2015), dentro da perspectiva de técnicas direcionadas ao aumento da segurança nas rodovias, destacam a abordagem de mitigação de risco de acidente rodoviário, que consiste em buscar a diminuição da exposição ao risco de acidente, verificando a quantidade de oportunidades de acidentes de determinado tipo, em um dado período de tempo e em uma dada localização. Bergel-Hayat *et al.* (2013) destacam a influência negativa das condições climáticas no fator de exposição ao risco, mostrando que condições desfavoráveis (e.g., chuva, neblina e pista molhada) aumentam o índice de acidentes rodoviários. Golob e Recker (2003) destacam, além das condições climáticas, a iluminação e o fluxo de veículo como fatores de exposição ao risco de acidentes. Adicionalmente, Jones *et al.* (2016) descrevem as características de veículos em acidentes de acordo com o tipo, dimensões, câmbio, direção, entre outros, definindo tipos de veículos mais propensos à ocorrência de acidentes.

Pode-se observar que muitos autores destacam a exposição ao risco de acidente e os diferentes fatores como um elemento fundamental na redução de acidentes rodoviários e aumento da segurança das vias, e com base nesta abordagem este trabalho propõe uma forma de mensurar a exposição ao risco através da classificação de trechos rodoviários, relacionando o número de acidentes, condições físicas da via e índices de chuva e nebulosidade. A classificação foi realizada através de técnicas de aprendizado de máquina, em especial modelos de classificação, como os apresentados nos trabalhos de Chong *et al.* (2005) para analisar acidentes de trânsito urbano, e Clarke *et al.* (1998) para classificar acidentes em pontos de cruzamento. A escolha das técnicas de classificação foi realizada com base na análise dos dados pelo método de análise de componentes principais (PCA, do inglês *Principal Component Analysis*), que verifica a disposição dos dados com base nos critérios selecionados.

2. MATERIAIS E MÉTODOS

2.1. Bases de dados

2.1.1. Acidentes rodoviários

A base de dados de acidentes rodoviários em rodovias federais utilizada neste trabalho está disponível na página do Departamento Nacional de Infraestrutura de Transportes (DNIT, 2011a). A base contém a quantidade de acidentes, o uso do solo, o dia e horário, o tipo e a gravidade do acidente, bem como a quantidade de feridos e mortos para quilômetros de rodovias federais. A última atualização disponível é referente a 2011, sendo o ano de referência considerado nas demais bases de dados utilizadas neste trabalho.

2.1.2. Condições físicas das rodovias

Os dados sobre a classificação das rodovias, quanto à condição física, foram obtidos pela pesquisa realizada pela Confederação Nacional do Transporte – CNT (CNT, 2011). Esta base contém informações sobre a gestão da rodovia (pública ou concessão), a extensão pesquisada

em quilômetros, o estado geral, de pavimento, sinalização e geometria. O pavimento é analisado pela condição da superfície, velocidade possível na via (devido ao pavimento) e o pavimento do acostamento. A sinalização é avaliada pelas faixas centrais e laterais, placas de limite de velocidade, indicação e de interseção, dispositivos de proteção contínua de visibilidade, e legibilidade das placas. Para a avaliação da geometria da via considera-se o tipo e perfil da rodovia, faixa adicional de subida e condição, obras de arte e condição, curvas perigosas e condição, e acostamento. O estado geral é a média das notas recebidas nas avaliações listadas. Os resultados de cada critério são disponibilizados como classificação em ótimo, bom, regular, ruim e péssimo.

A pesquisa faz uma inferência da classificação, pela extensão pesquisada, para toda a rodovia e, portanto, todos os trechos de uma mesma rodovia possuem as mesmas classificações para as dimensões de gestão, estado geral, pavimento, sinalização e geometria.

2.1.3.Trechos rodoviários

Esta base de dados (do ano de 2011), obtida no Sistema Nacional de Viação – SNV (DNIT, 2011b), descreve os trechos rodoviários das rodovias brasileiras. Contempla as informações pelo código único do trecho, seu estado da federação, locais de início e fim e os quilômetros correspondentes. A superfície do trecho é considerada como pavimentada (pista simples, em obras de duplicação ou pista dupla) e não pavimentada (leito natural, em obras de implantação, implantada ou em obras de pavimentação).

2.1.4.Índices de chuva e nebulosidade

Os dados sobre chuva e nebulosidade foram obtidos no Instituto Nacional de Meteorologia (INMET, 2017). Foram consideradas as informações de data e hora, volume de precipitação em milímetros, e índice de nebulosidade das estações no estado de Minas Gerais que fazem este tipo de medição. A nebulosidade é uma medida de classificação da densidade de nuvens no céu, sendo menor que 1 (céu limpo), de 1 a 3 (céu pouco nublado), de 4 a 9 (céu nublado) e maior que 9 (céu encoberto).

No total, são 48 estações meteorológicas, sendo que as estações de Lambari, Coronel Pacheco e Florestal não possuem informações para 2011, sendo desconsideradas neste trabalho. Uma vez que o total de estações não expressa as medições de chuva e nebulosidade para todas as cidades do estado, foi necessário estabelecer uma regra de relação entre as cidades com estação meteorológica e as cidades que não a possuem, descrita no item 2.2 a seguir.

2.2. Relacionamento entre cidade e estação

Para relacionar as cidades que possuem estação meteorológica que realiza medições de chuva e nebulosidade com as cidades que não possuem, adotou-se como critério a distância euclidiana, conforme a equação (1) a seguir:

$$Min\left\{d_{ij} = \sqrt{\left(lat_i - lat_j\right)^2 + \left(long_i - long_j\right)^2}\right\} \quad (1)$$

em que *Min*: Mínimo;
d_{ij}: Distância entre a cidade *j* e a estação *i*;
lat: Coordenada de latitude;
long: Coordenada de longitude;
i: Índice da estação;
j: Índice da cidade;

As medidas de latitude e longitude das cidades consideradas na pesquisa foram obtidas através do Instituto Brasileiro de Geografia e Estatística – IBGE.

Desta forma, inferem-se a uma cidade que não tem estação, os dados correspondentes de chuva e nebulosidade da cidade com estação mais próxima. Por exemplo, para a cidade de Bicas-MG, que não tem estação meteorológica, foram considerados os dados da cidade mais próxima que possui estação, que no caso é Juiz de Fora - MG.

Como a análise de acidentes ocorre a cada trecho rodoviário e é baseada nos índices de chuva e nebulosidade e que, por sua vez, estão atrelados às cidades, os trechos rodoviários também são atrelados à determinada região geográfica de uma cidade, conforme base de dados do SNV. Portanto, os índices de chuva e nebulosidade medidos em determinada estação valem para toda a região geográfica de uma cidade e, consequentemente, para os trechos rodoviários contidos nesta região. Em alguns casos da base SNV há trechos rodoviários pertencentes a duas ou mais cidades e, neste caso, foram considerados aleatoriamente para somente uma destas cidades visto que, devido à indisponibilidade dos dados, não há como subdividir um trecho rodoviário e definir qual parte do trecho pertence a uma ou a outra cidade. Esta abordagem não impacta na análise dos dados e resultados, uma vez que os trechos que pertencem a duas ou mais cidades permanecem na mesma área de uma estação meteorológica.

Os dados das estações meteorológicas foram registrados somente no horário de 12h00min. Uma restrição imposta neste trabalho foi contabilizar as ocorrências de acidente que ocorreram entre os horários de 10h00min às 14h00min como forma de considerar, com menor margem de erro, os acidentes que provavelmente ocorreram em período de chuva e nebulosidade consideradas.

2.3. Análise de componentes principais e técnicas de classificação binária

A Análise de Componentes Principais, do inglês *Principal Components Analysis* (PCA) é, como descrevem Abidi e Williams (2010), uma técnica de análise multivariada que tem por finalidade básica a análise dos dados visando sua redução, a eliminação de sobreposições e a escolha das formas mais representativas de dados a partir de combinações lineares das variáveis originais. Foi proposta por Karl Pearson em 1901 e, em síntese, seu objetivo é condensar a informação contida nas variáveis originais em um conjunto menor de variáveis estatísticas (chamadas de componentes) com uma menor perda possível de informação. É bastante útil quando uma representação gráfica não é possível, uma vez que pode representar somente as combinações de parâmetros com maior parcela na variância dos dados.

As técnicas de aprendizado de máquina utilizam o princípio da indução onde, a partir de um conjunto de exemplos, infere-se conclusões genéricas sobre a população (Lorena e Carvalho, 2007). O aprendizado por indução é divido em supervisionado e não supervisionado. Como descreve Haykin (1999), no aprendizado supervisionado o algoritmo de classificação atua a partir de um conhecimento prévio do ambiente externo, expresso por conjuntos de exemplos com os dados de entrada e os resultados esperados, enquanto no aprendizado não supervisionado não existem exemplos prévios já rotulados (com os resultados esperados) e o algoritmo de classificação tenta agrupar as medias segundo uma medida de qualidade.

Existem muitos trabalhos que relacionam PCA e técnicas de aprendizagem de máquina. Haykin (2009) e Johnson e Wichem (2004), por exemplo, descrevem o uso de técnicas de

análise multivariada para avaliar a relação dos dados onde há muitos atributos para classificação. Schölkopf *et al*. (1997) mostram um trabalho de formação de Kernels (uma das áreas de aprendizado de máquina) pela utilização de PCA, e Heba *et al*. (2010) fazem um trabalho de implementação conjunto de PCA e máquinas de vetor suporte para separação de dados.

3. DESENVOLVIMENTO

3.1. Resumo dos dados

Pelos dados disponíveis das bases de dados utilizadas, foram consideradas nesta pesquisa as rodovias federais BR-040, BR-050, BR-116, BR-135, BR-146, BR-153, BR-251, BR-262, BR-267, BR-354, BR-356, BR-364, BR-365, BR-381, BR-452, BR-459, BR-460, BR-464 e BR-474, totalizando 19 rodovias que percorrem o estado de Minas Gerais, representadas na Figura 1.

Figura 1: Rodovias federais no estado de Minas Gerais. Fonte: DNIT (2018).

O estado de Minas Gerais possui a maior malha rodoviária brasileira, o que equivale a 269.546 km (16% do total), sendo que 7.689 km (29%) são rodovias federais (Governo de Minas Gerais, 2017).

Algumas estatísticas nestas rodovias (considerando o total de acidentes ao longo do ano) foram levantadas como forma de mapear, de forma geral, a ocorrência de acidentes (Figura 2) e os tipos de acidentes registrados (Figura 3).

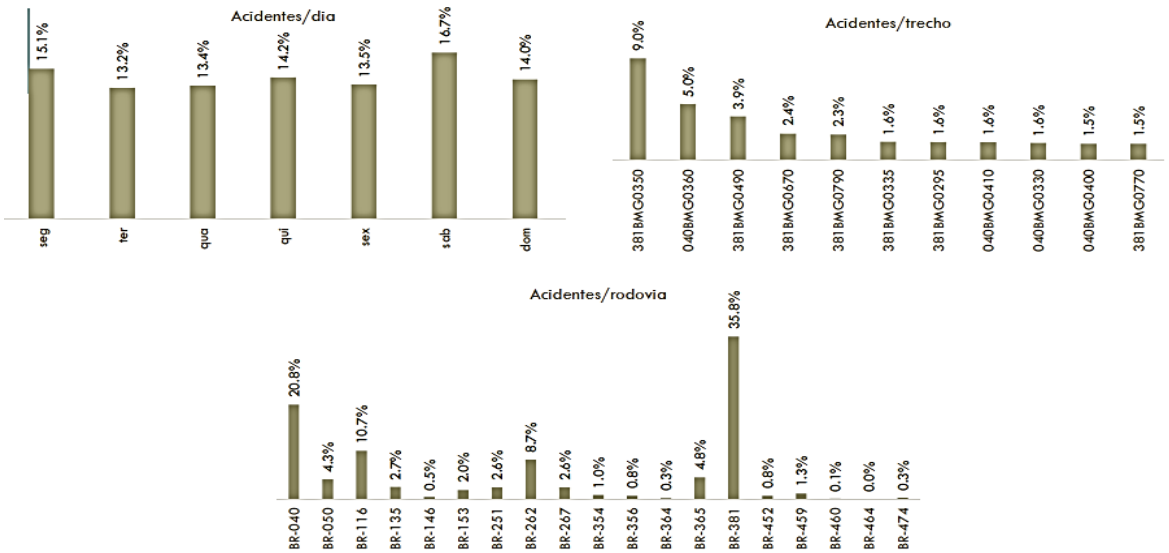

Figura 2: Incidência de acidentes por dia, por trecho rodoviário e por rodovia. Fonte: DNIT (2011a).

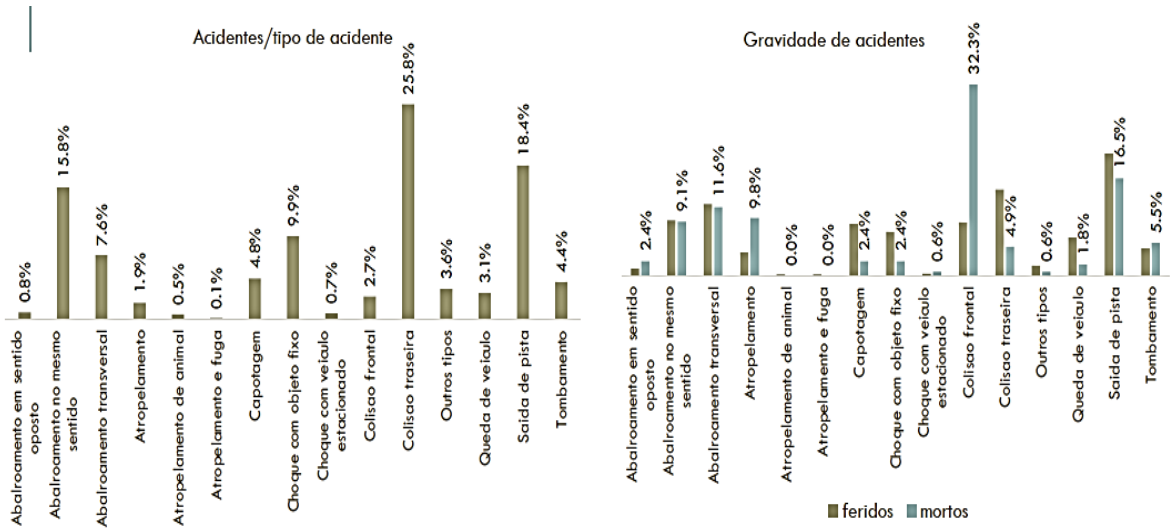

Figura 3: Incidência de tipos de acidentes e gravidade. Fonte: DNIT (2011a).

O gráfico de acidentes/dia (Figura 2) mostra que os acidentes ocorreram, em sua maioria, nos dias de sábado e segunda-feira. Os gráficos acidentes/trecho (Figura 2) e acidentes/rodovia (Figura 2) destacam as rodovias BR-040 e BR-381 como as que mais registraram acidentes em 2011 e, em especial, o trecho 381BMG0350 da BR-381 localizado no município de Caeté, apresentou o maior índice de acidentes dentre todos os trechos rodoviários analisados.

A Figura 3 resume os tipos de acidentes, mostrando que os acidentes de colisão traseira, saída de pista e abalroamento no mesmo sentido respondem pelos maiores índices de ocorrência. Quanto à gravidade (Figura 3), os acidentes do tipo colisão frontal e saída de pista são os que ocasionam maior índice de mortos e feridos, respectivamente.

3.2. Análise e resultados

3.2.1. Análise de componentes principais

Neste trabalho, utilizou-se a linguagem R para proceder à análise de PCA, sendo a variável dependente dada pelo resultado da classificação do trecho rodoviário (1=acidente e 0=não acidente), e as variáveis independentes dadas por: uso do solo, dia da semana, condição geral da via, condição do pavimento, da sinalização, geometria e os índices de chuva e de nebulosidade. A Figura 4 resume a variância total dos dados do problema pela quantidade de variáveis independentes consideradas, ou seja, com quantas variáveis independentes a variância total é explicada. Observa-se que o problema não pode ser resumido a poucas variáveis independentes (com 7 variáveis que há uma representação considerável da variância total do problema) e consequentemente, como mostrado na Figura 5, é mais difícil diferenciar os grupos de "acidente" e de "não acidente".

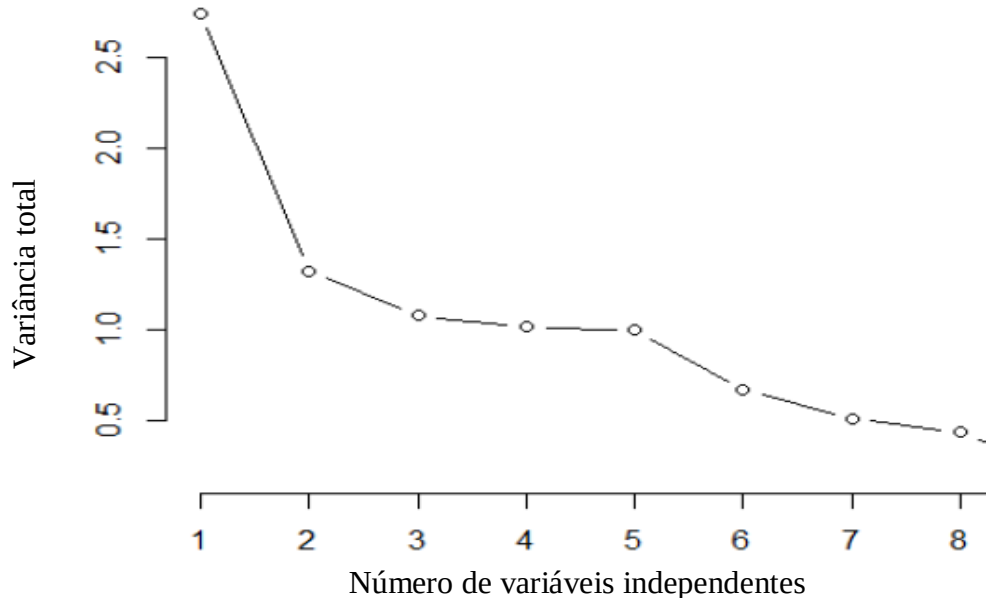

Figura 4: Variância total explicada pela quantidade de variáveis consideradas.

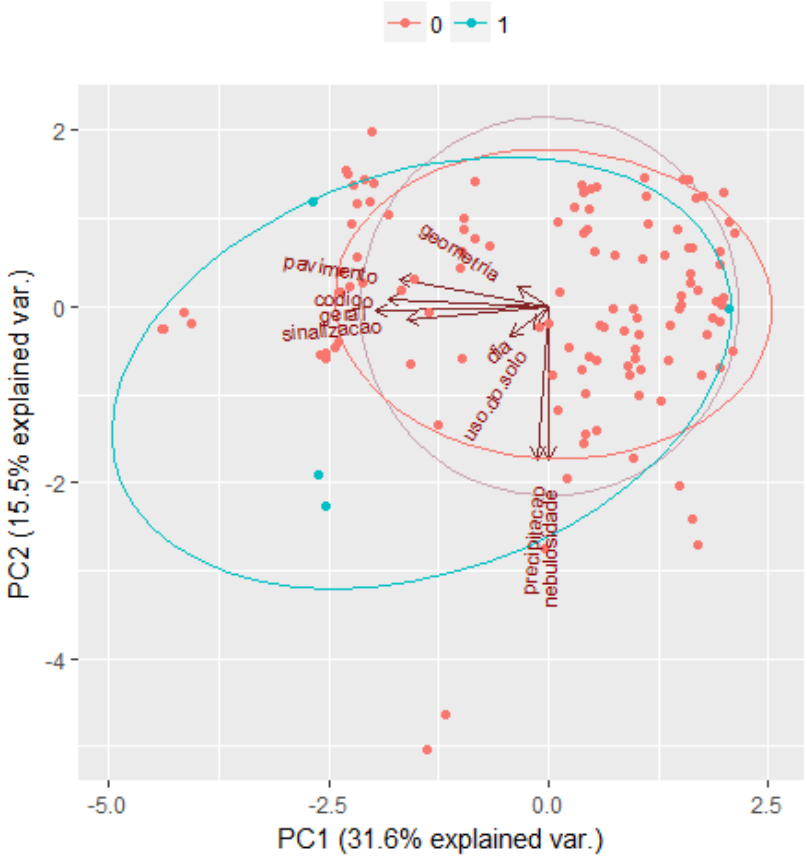

Figura 5: Análise de componentes principais da ocorrência de acidentes.

A dificuldade em separar os grupos de "acidente" e "não acidente", reside no fato da abordagem do problema analisar as ocorrências de acidente por trecho e por dia, ou seja, há mais ocorrências de "não acidente" quando se analisa diariamente.

Como pode ser observada na Figura 5, a quantidade de eventos de "acidente" é consideravelmente inferior aos eventos de "não acidente", fato que direciona a serem utilizadas técnicas de classificação para dados considerados desbalanceados, ou seja, quando há muitas observações de um evento comparado a outro.

3.2.2. Aprendizado de máquina: Métodos de classificação binária

Para a abordagem utilizada neste trabalho, devido à disposição dos dados, o aprendizado de máquina é do tipo supervisionado, onde para um dado conjunto de exemplos rotulados como 0 (sem acidente) e 1 (com acidente) para os atributos uso do solo, dia da semana, condição geral da via, condição do pavimento, da sinalização, geometria e os índices de chuva e de nebulosidade, deve-se produzir um classificador, ou um modelo preditor, com capacidade de predizer o rótulo de novos dados, processo este denominado de treinamento.

A base de entrada de dados foi obtida pelo relacionamento entre as bases de dados já mencionadas anteriormente, construindo um banco de dados na linguagem MySQL, consistindo em um total de 152.120 observações. Verificando a proporção de rótulos, observou-se que 96,61% são de "não acidente" contra 3,39% de "acidentes". Desta forma, utilizou-se algoritmos de classificação do tipo *sampling* com o propósito de transformar dados

desbalanceados em dados balanceados.

Como descrevem Rahman e Davis (2013), as principais formas de transformar dados desbalanceados em balanceados são através das técnicas de *over sampling* (replica as observações da classe minoritária), *under sampling* (reduz o número de observações da classe majoritária), *Synthetic data generation* – SMOTE (gera observações artificiais) e *Cost Sensitive Learning* – CSL (avalia o custo associado a uma classificação errônea).

Neste trabalho, utilizou-se o pacote ROSE – *Random Over Sampling Examples* (Lunardon *et al*., 2014), na linguagem R, que gera observações artificiais para dados desbalanceados conforme a técnica SMOTE. Além da técnica SMOTE, utilizou-se os métodos de *over sampling, under sampling* e ambos combinados (*over* e *under sampling)*.

Como classificadores, foram utilizados métodos contidos no pacote ROSE: Árvore de decisão, *Naive Beyes* e Rede Neuronal. Para avaliação da eficácia dos modelos de predição utilizados, adotou-se a medida ROC – *Receiver Operating Characteristics*, que é útil para avaliar modelos em domínio com grande desbalanceamento entre as classes (Prati e Flach, 2005; Flach e Wu, 2005). A medida ROC sugere que, quanto maior a área sobre a curva, melhor a acurácia do modelo. Lunardon *et al*. (2014) sugerem as seguintes classificações de medidas ROC: 0,5 e 0,6 (falhas); 0,6 a 0,7 (ruim); 0,7 a 0,8 (fraco); 0,8 a 0,9 (bom); 0,9 a 1,0 (excelente). A Tabela 1, resume os resultados (medidas ROC) obtidos nos testes computacionais empregados no presente trabalhos, e a Figura 6 apresenta as curvas ROC obtidas.

Tabela 1: Medidas ROC dos métodos de classificação utilizados.

Técnicas de balanceamento de dados	Métodos de Classificação		
	Árvore de Decisão	*Naive Beyes*	Rede Neural
Under Sampling	0,771	0,758	0,747
Over Sampling	0,782	0,758	0,748
Over e *Under Sampling* combinados	0,793	0,757	0,758
SMOTE	0,801	0,756	0,759

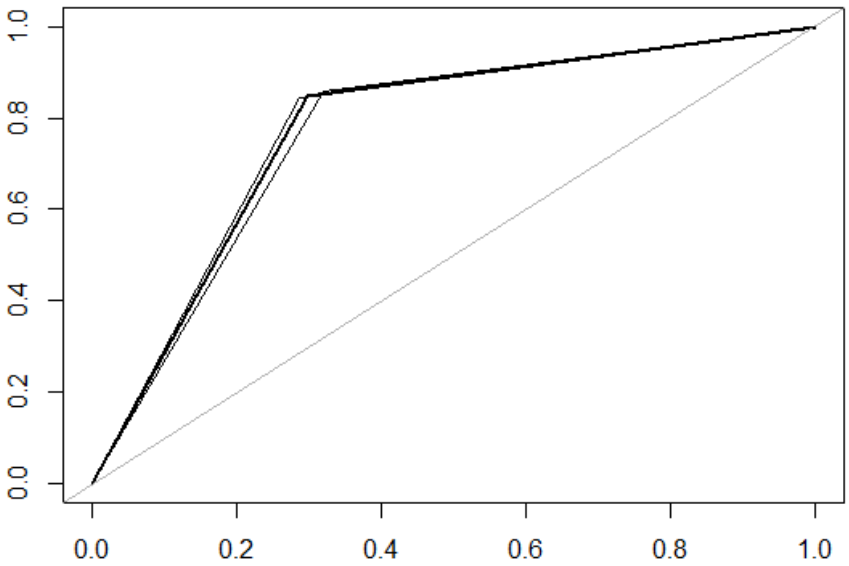

Figura 6: Curvas ROC dos métodos de classificação utilizados.

Observa-se que o melhor resultado obtido foi pelo método ROSE utilizando o classificador de árvore de decisão, com ROC = 0,801, indicando que este é um bom modelo preditor. Este resultado indica que o modelo acerta a classificação do evento (acidente ou não acidente) em 80,1% dos casos. Como exemplo, um próximo evento não registrado na base de dados (dia 25/08, trecho = 381BMG0350, chuva = 0 mm, nebulosidade = 1 (céu limpo), histórico de acidentes = 1, estado geral = bom, pavimento = bom, sinalização = ruim, geometria = bom), se o modelo classificar como "não acidente" este resultado possui um nível de confiança de 80,1%, ou seja, de realmente não ter acidente.

Observa-se também que, com exceção do resultado anterior, os demais acertaram a classificação entre 74,7% e 79,3%, sendo considerado como fraco. Este fato mostra três vertentes possíveis de oportunidades de melhoria: (1) Análise dos modelos: O teste de novos classificadores nos mesmos dados para validar melhor os resultados obtidos e verificar se é possível aumentar o índice de acerto; (2) Disponibilidade dos dados: Utilizar os mesmos parâmetros para bases de dados de mais anos, o que pode melhorar a acurácia do modelo, uma vez que a quantidade de dados aumenta e um histórico de tempo de todos os parâmetros são considerados. Um histórico de dados de acidentes pode proporcionar refinar o modelo e a opção de trabalhar com classificadores de regressão, ao invés de binário, torna-se mais atraente devido a possibilidade de classificar um evento pela sua probabilidade de ocorrência (um valor entre 0 e 1), que incorpora mais aplicabilidade à proposta descrita neste trabalho; (3) Conteúdo das bases de dados: Além de escassas, as bases de dados brasileiras possuem poucos parâmetros para análise. Como exemplo, uma base de acidentes rodoviários do Reino Unido (disponível na plataforma Kaggle.com) possui mais atributos registrados em ocorrências de acidentes, como tipo de veículo, dados do condutor (idade, sexo, condições físicas, por exemplo), iluminação da via e outros.

4. CONSIDERAÇÕES FINAIS

A proposta consistiu em introduzir algumas técnicas da área da Inteligência Computacional, em especial as técnicas e modelos de aprendizado de máquina, na área de transportes. Mais especificamente, o propósito foi o de prever possíveis acidentes em determinados trechos

rodoviários de Minas Gerais, com base nos princípios descritos por Goniewicz *et al.* (2016) e Elvik (2015), que consistem na redução à exposição ao risco de acidente rodoviário que consiste na classificação do risco em trechos rodoviários analisados. Em síntese, um conhecimento sobre determinada rodovia ou trecho pertencente a ela, através do dia específico e das condições climáticas deste dia, pode influenciar na escolha de se utilizar uma rodovia em detrimento à outra devido ao risco de acidente, sendo uma forma de diminuir a exposição ao risco.

Neste trabalho, podemos observar que as técnicas adotadas nas bases de dados utilizadas podem propor resultados satisfatórios, devido à aplicabilidade computacional das técnicas e a possibilidade dos resultados serem utilizados como fonte de informação para outros serviços computacionais, como *softwares* de roteirização de veículos e/ou sistema de gerenciamento de tráfego em organizações públicas e/ou privadas. Como exemplo, a construção de uma rota de veículos se dá pela distância e custos envolvidos, mas também pode considerar o risco de acidente envolvido em cada rota selecionada e influenciar a escolha de caminhos com menor risco de acidente a um custo aceitável.

Entende-se que a ideia proposta encontra-se ainda em fase inicial e que necessita de adaptações para mensurar com melhor exatidão as probabilidades de acidentes, podendo vir a se tornar um modelo de regressão, e não de classificação como descrito, onde o trecho rodoviário possa vir a ser rotulado com uma probabilidade de acidente (valores entre 0 e 1) e não de forma binária como desenvolvido neste trabalho. Isto proporcionaria uma melhor mensuração do risco de acidente e consequentemente resultados mais próximos ao real. Além disso, a necessidade de se obter bases de dados disponibilizadas com maior frequência e, principalmente, que possam considerar mais atributos quando da ocorrência de um acidente é fundamental visto que, em aprendizado de máquina, quanto maior a quantidade e variedade de dados melhor a assertividade dos métodos de classificação.

O Brasil, como já descrito, possui índices crescentes de acidentes e precisa estudar e propor melhorias para o sistema rodoviário atual objetivando diminuir esta triste estatística e, uma das formas possíveis, é melhorar as bases de dados disponíveis para fornecer subsídios para pesquisadores da área proporem melhorias em todo o sistema, através de modelos como os propostos neste trabalho.

Agradecimentos

Ao Conselho Nacional de Desenvolvimento Científico e Tecnológico – CNPq (Processo 06186/2017-9) e Fundação de Amparo à Pesquisa de Minas Gerais - FAPEMIG (Processos TEC PPM 528/11 e TEC PPM 388/14).

REFERÊNCIAS BIBLIOGRÁFICAS

Abdi. H., & Williams, L.J. (2010). Principal component analysis. *Wiley Interdisciplinary Reviews: Computational Statistics*,.v. 2, n. 4, p. 433-459.

Bergel-Hayat, R., Debbarh, M., Antoniou, C., & Yannis, G. (2013). Explaining the road accident risk: weather effects. *Accident Analysis & Prevention*, v. 60, n.1, p. 456-465.

Chong, M., Abraham, A., & Paprzycki, M. (2005). Traffic accident analysis using machine learning paradigms. *Informatica*, v. 29, n. 1, p. 89-98.

Clarke, D. D., Forsyth, R., & Wright, R. (1998). Machine learning in road accident research: decision trees describing road accidents during cross-flow turns. *Ergonomics*, v. 41, n. 7, p. 1060-1079.

CNT (2011). *Relatório gerencial: pesquisa CNT de rodovias 2011*. Confederação Nacional do Transporte, Brasília, DF.

DNIT (2018). *Condições das Rodovias*. Departamento Nacional de Infraestrutura de Transportes. Disponível em:

http://servicos.dnit.gov.br/condicoes/mg.htm. Acesso em Setembro de 2018.
DNIT (2011a). *Estatística de acidentes – Ano de 2011*. Departamento Nacional de Infraestrutura de Transportes. Disponível em: http://www.dnit.gov.br/rodovias/operacoes-rodoviarias/estatisticas-de-acidentes. Acesso em: Janeiro de 2018.
DNIT (2011b). *Sistema Nacional de Viação 2011*. Departamento Nacional de Infraestrutura de Transportes. Disponível em: http://www.dnit.gov.br/sistema-nacional-de-viacao/sistema-nacional-de-viacao. Acesso em: Janeiro de 2018.
Elvik, R. (2015). Some implications of an event-based definition of exposure to the risk of road accident. *Accident analysis & prevention*, v. 76, p. 15-24.
Flach, P.; Wu, S. (2005) Repairing concavities in roc curves. *Proceedings of the nineteenth international joint conference on artificial intelligence (IJCAI 2005*), Edinburgh , UK, v.1, p. 702–707.
Golob, T. F., & Recker, W. W. (2003). Relationships among urban freeway accidents, traffic flow, weather, and lighting conditions. *Journal of transportation engineering*, v. 129, n. 4, p. 342-353.
Goniewicz, K., Goniewicz, M., Pawłowski, W., & Fiedor, P. (2016). Road accident rates: strategies and programmes for improving road traffic safety. *European journal of trauma and emergency surgery*, v. 42, n. 4, p. 433-438.
Haykin, S (1999*). Neural networks - A compreensive foundation*. Ed. Prentice-hall, New Jersey, 2nd edition.
Haykin, S. S. (2009*). Neural networks and learning machines*. Ed. Pearson Prentice-hall, Upper Saddle River, NJ, USA.
Heba, F. E., Darwish, A., Hassanien, A. E., & Abraham, A. (2010). Principle components analysis and support vector machine based intrusion detection system. *In Intelligent Systems Design and Applications (ISDA), 2010 10th International Conference* on. IEEE, v.10, p. 363-367.
INMET (2011). *BDMEP - banco de dados meteorológicos para ensino e pesquisa.* Instituto Nacional de Metereologia. Disponível em: http://www.inmet.gov.br/portal/index.php?r=bdmep/bdmep. Acesso em: Novembro de 2017.
Johnson, R. A., & Wichern, D. W. (2004). *Multivariate analysis.* Ed. Pearson Prentice-hall, Upper Saddle River, NJ, USA.
Jones, I. S. (2016). *The effect of vehicle characteristics on road accidents*. Ed. Elsevier, Rio de Janeiro, Brazil.
Lord, D., & Washington, S. (2018*).* Introduction. In Safe Mobility: Challenges, Methodology and Solutions. *Emerald Publishing Limited*, p. 1-10.
Lorena, A. C.; de Carvalho, A (2007). Uma introdução às máquinas de vetor suporte. *Revista de informática teórica e aplicada*, v. 14, n. 2, p. 43-67.
Lunardon, N.; Menardi, G.; Torelli, N. (2014). ROSE: a package for binary imbalanced learning. *R journal*, v. 6, n.1, p. 82–92.
Menardi, G; Torelli, N. (2014). Training and assessing classification rules with imbalanced data. *Data mining and knowledge discovery*, v. 28, n.1, p. 92–122.
Naing, C. et al (2007). *Which factors and situations for human functional failures? developing grids for accident causation analysis.* Longhborough University, Longhborough, UK.
Prati, R. C.; Flach, P. R. (2005). An algorithm for rule learning based on roc analysis. *Proceedings of the nineteenth international joint conference on artificial intelligence (IJCAI 2005)*, Edinburgh, UK, v.1, p. 823–828.
Rahman, M.M.; Davis,D. N. (2013) Addressing the class imbalance problem in medical datasets. *International journal of machine learning and computing* , v. 3, n. 2, p. 224–228.
Schölkopf, B., Smola, A., & Müller, K. R. (1997). Kernel principal component analysis. *In International Conference on Artificial Neural Networks*. Springer, Berlin, Heidelberg, p. 583-588.
Sivak, M.; Schoettle, B (2014). *Mortality from road crashes in 193 countries: a comparison with other leading causes of death*. University of Michigan, Transportation Research Institute, Michigan, EUA.
WHO (2015). *Global status report on road safety 2015*. World Health Organization. Disponível em: http://www.who.int/iris/handle/10665/189242. Acesso em: Agosto de 2018.