MÉTODO PARA CLASSIFICAÇÃO DE PEDESTRES OBEDIENTES E OPORTUNISTAS EM TRAVESSIAS SEMAFORIZADAS BASEADO EM MACHINE LEARNING

Francisco Altanizio Batista de Castro Junior
Beliza Carvalho de Oliveira
Gabriel Amorin Rabelo Nobre
Kaio Gefferson de Almeida Mesquita
Moisés Gomes de Holanda
Universidade Federal do Ceará
Programa de Pós-graduação em Engenharia de Transportes

Resumo: O aumento da taxa de motorização em países em desenvolvimento aumentou o risco de sinistros viários para os usuários vulneráveis, como os pedestres. Nesse sentido, este trabalho tem como objetivo prever o comportamento dos pedestres obedientes e oportunistas em travessias semaforizadas de Fortaleza utilizando técnicas de *machine learning*. Para realizar essa caracterização foram utilizados dados coletados de vídeos gravados pelas câmeras de monitoramento da Prefeitura. Foram selecionadas travessias estratégicas, que possuem um maior fluxo de pedestres e de veículos. Hipóteses foram desenvolvidas a fim de conhecer melhor o fenômeno a ser estudado. Os modelos de classificação supervisionada foram ajustados aos dados satisfatoriamente, revelando a possibilidade de prever o comportamento de risco dos pedestres a partir de suas características. Os algoritmos de *Nayve Bayes*, Regressão Logística e *Random Forest*, obtiveram uma acurácia de previsão de oportunismo nas travessias de 84%, 82% e 77%, respectivamente.

Palavras-chave: Pedestres; Obedientes; Oportunistas; *Aprendizado de Máquina*; Previsão.

Abstract: The increase in the rate of motorization in developing countries has increased the risk of road accidents for vulnerable users, such as pedestrians. In this sense, this work aims to predict the behavior of obedient and opportunistic pedestrians at traffic lights in Fortaleza using machine learning techniques. To carry out this characterization, data collected through videos recorded by the monitoring cameras of the Municipality were used. Strategic crossings were selected, which have a greater flow of pedestrians and vehicles. Hypotheses were developed to better understand the phenomenon to be studied. The supervised classification models were satisfactorily fitted to the data, revealing the possibility of predicting the risk behavior of pedestrians based on their characteristics. The algorithms of Nayve Bayes, Logistic Regression and Random Forest, obtained an accuracy of prediction of opportunism in the crossings of 84%, 82% and 77%, respectively.

Keywords: Pedestrians; Obedient; Opportunists; Machine Learning; Prediction.

1. INTRODUÇÃO

Metade dos sinistros viários com vítimas envolvem usuários vulneráveis, sendo 23% pedestres. No Brasil, a taxa de atropelamentos é cerca de 18%, sendo 78% em vias urbanas. Nesse contexto, a predição do comportamento de risco do pedestre oferece uma ferramenta poderosa e informativa nos regimes de planejamento urbano e pode ser aproveitada para informar o posicionamento adequado de melhorias e projetos de tratamento para melhorar a segurança dos pedestres (MURPHY *et al*., 2017).

Sinistros viários envolvendo pedestres tornaram-se um grande problema de segurança em todo o mundo, principalmente nos países em desenvolvimento, como o Brasil, segundo

Zhang (2005). Nesses países existe uma alta densidade populacional e rápida urbanização com a falta de aderência às regras de trânsito por motoristas e pedestres (WRI BRASIL, 2015).

Os pedestres em interseções semaforizadas podem ser classificados em obedientes à semaforização e oportunistas. Os oportunistas decidem realizar a travessia durante o tempo de vermelho do sinal, desrespeitando o tempo semafórico reservado à passagem de veículos e não de pedestres. Esse comportamento é influenciado por diversos fatores, os quais se configuram como importantes para a sua compreensão (Liu *et al*., 2000; Yang *et al*., 2006). Jacobsen (2011) descreve que esse comportamento violador tem forte relação na movimentação dos pedestres durante os seus deslocamentos. Fatores como a velocidade de deslocamento, tempo de travessia e travessias em grupo têm relação com esse comportamento.

A predição do comportamento dos pedestres oportunistas é essencial em um contexto em que diversas cidades, localizadas principalmente em países de Terceiro Mundo, têm experienciado consideráveis incrementos na taxa de motorização de suas populações. A predição do comportamento pode auxiliar em mecanismos capazes de evitar um sinistro viário, além de encontrar padrões no comportamento de risco dos usuários.

O objetivo deste trabalho é propor um método para classificação de pedestres entre oportunistas e obedientes em travessias semaforizadas de Fortaleza, baseado em modelos supervisionados de *machine learning*. Vale destacar que antes da etapa de modelagem e avaliação, a contextualização da região, bem como a relação entre as variáveis preditoras e o fenômeno foram levantadas. A seção dois relata os principais fatores que influenciam no comportamento dos pedestres. A seção três relata a metodologia adotada pelo trabalho. A seção quatro apresenta a contextualização dos dados seguido das hipóteses levantadas. A seção cinco apresenta a caracterização da modelagem e a seção seis apresenta os principais resultados da predição e avaliação dos modelos. Por fim, a seção sete apresenta as considerações finais do estudo.

2. FATORES QUE INFLUENCIAM O COMPORTAMENTO DOS PEDESTRES NAS TRAVESSIAS

Diversos fatores influenciam o comportamento de risco dos pedestres nas travessias. De acordo com Jacobsen (2011) e Ariotti (2006) os atributos que podem influenciar a maneira como os pedestres se movimentam podem ser subdivididos em dois grupos: i) características dos pedestres, como idade e gênero e ii) características do ambiente, como número de faixas, fluxo de veículos e conflitos com outros usuários durante a travessia.

Um estudo analisou o comportamento de hesitação dos pedestres durante a travessia na França e no Japão. Jay *et al*. (2020) descobriram que existe uma relação entre a velocidade dos pedestres com o tamanho do grupo na travessia. Os pedestres tendem a adotarem maiores velocidades quando existem mais pessoas atravessando. Evidenciou que os japoneses têm um período mais demorado de incerteza e hesitam em violar a sinalização.

Outro estudo em Bangladesh utilizou um modelo de regressão logística para encontrar os fatores de influência do tipo de travessia e o comportamento de risco. Zafri *et al*. (2020) utilizaram os fatores relacionados com o tipo de controle da interseção, o tamanho da via, o fluxo de veículos, a brecha disponível, características dos pedestres e o fluxo de pedestres na travessia. O trabalho revelou uma relação entre a idade e os tipos de travessia, a qual pessoas mais novas têm cinco vezes mais chances de realizar uma travessia oportunista.

Bernhoft *et al* (2008) descobriram que pedestres mais velhos tendem a ser mais cautelosos que pedestres mais novos, o que vai de acordo com o estudo anterior. Outro estudo

de Ronsenbloom *et al*. (2004) mostrou que pedestres masculinos são mais propensos a violar a sinalização. Esses estudos investigaram a influência das características dos pedestres na travessia em interseções semaforizadas

Diversos estudos observaram a influência dos fatores do ambiente no comportamento do pedestre. Os fatores podem ser a densidade de pedestres, o tipo do uso do solo, as condições do tempo, a hora e o tipo de via (Mukherjee e Mitra, 2020; Vignali et al., 2020; Zhang et al., 2019). O tempo disponível para a travessia é outro fator importante que determina o comportamento do pedestre (Koh e Wong, 2014; Mukherjee e Mitra, 2019; Patra *et al*., 2019).

Bandeira (2019) verificou que a programação semafórica, fluxo de pedestres e veículos e a presença de estacionamento influenciam no comportamento dos pedestres. Yang *et al*. (2006) e Zhuping *et al*. (2011) verificaram que a obediência dos pedestres está relacionada à pressa e o fluxo dos veículos. Ressalta-se que a pressa dos usuários, fluxo veicular e a presença de veículos estacionados são características que estão relacionadas com o uso do solo.

Recentemente, foram desenvolvidos diversos algoritmos e programas computacionais de *machine learning* capazes de realizar a caracterização e a predição de forma rápida e eficiente em aplicações diversas. Nesse contexto, surgiram estudos que analisaram o comportamento dos pedestres em travessias que utilizaram técnicas de *machine learning* (Alver et al., 2021; Zhang *et al*., 2020). Alhajyaseen e Iryo-Asano (2017) analisaram a influência de fatores na velocidade adotada pelos pedestres na travessia. Outro estudo analisou a intenção de travessia dos pedestres oportunistas e suas características, utilizando algoritmos de aprendizagem profunda (Zhang et *al*., 2020).

3. METODOLOGIA

A metodologia adotada neste trabalho consistiu em três fases principais, conforme apresentado na Figura 1: (i) Contextualização; (ii) Caracterização e (iii) Avaliação dos modelos. Na fase inicial foram definidas as hipóteses norteadoras do trabalho, de modo a contribuir com a identificação das características dos pedestres oportunistas perante uma travessia semaforizada. Nesta etapa também foram apresentados o método de coleta dos dados, e as análises exploratórias que ocasionaram na escolha das variáveis para modelagem do comportamento de oportunismo.

A segunda etapa do método é a caracterização da modelagem, o qual são discutidas as variáveis selecionadas segregadas em características do usuário, espaciais do local e do comportamento. Ainda nesta etapa são discutidos os tipos de modelo adotados para classificação supervisionada, onde 1 equivale a um pedestre oportunista e 0 equivale a um pedestre obediente.

Modelos binários, apenas identificam a separação entre essas classes, enquanto modelos mais robustos baseiam a classificação em uma probabilidade de o usuário pertencer a classe, por isso a importância de se definir variáveis que representem bem o fenômeno. Partindo para a última etapa do método, são discutidas as variáveis selecionadas a partir dos seguintes critérios: (i) correlação entre as variáveis; (ii) causalidade entre as variáveis e; (iii) relação como o fenômeno. Por fim, são apresentados os resultados quanto a acurácia de previsão, erro médio, matriz de confusão e tempo de convergência pela Curva ROC. Esses critérios de avaliação serão mais bem explanados em um tópico posterior.

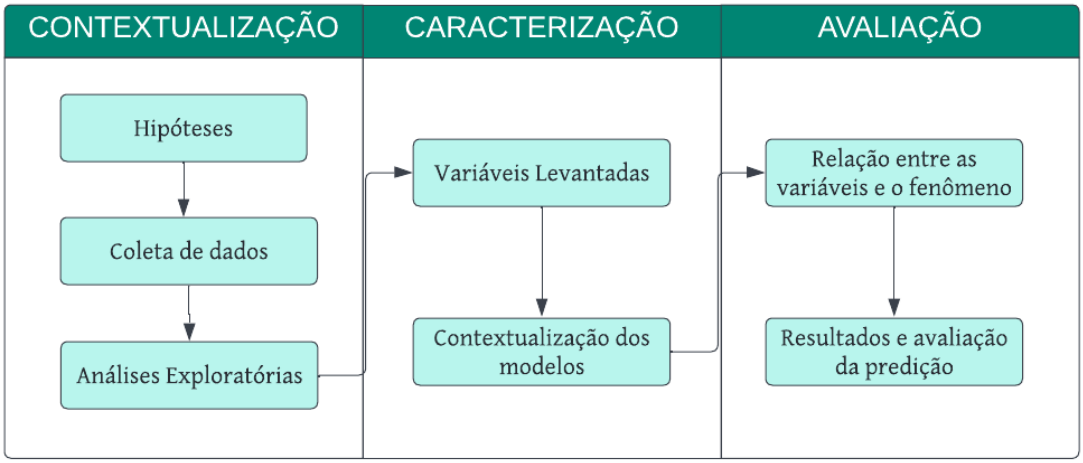

Figura 1 – Método para classificação de pedestres oportunistas e obedientes
Fonte: Autores

No aprendizado supervisionado, os dados de treinamento fornecidos ao algoritmo incluem as soluções desejadas, chamadas de rótulos. A classificação é uma tarefa típica de aprendizado supervisionado. Sendo utilizado também para prever um alvo de valor numérico, através de um modelo de regressão. Alguns algoritmos de regressão também podem ser usados para classificação e vice-versa.

Pode-se categorizar os sistemas de aprendizado de máquina por meio de generalização. Sendo a maioria das tarefas baseadas em previsões, ou seja, dada uma série de exemplos de treinamento, o sistema precisa ser capaz de generalizar em exemplos nunca vistos antes. Portanto neste trabalho foram considerados os seguintes aspectos para melhorar o índice de previsão: (i) Selecionar um modelo mais "poderoso", com mais parâmetros; (ii) Alimentar o algoritmo de aprendizado com melhores características (*feature engineering*); e (iii) Reduzir as restrições no modelo (por exemplo, reduzindo o hiperparâmetro de regularização). Todo o processo foi realizado através da linguagem *python,* em detrimento das bibliotecas *pandas, Scikit-Learn e TensorFlow* para manipulação, tratamento e modelagem dos dados.

4. CONTEXTUALIZAÇÃO

4.1. Hipóteses sobre o comportamento de pedestres em travessia de Fortaleza

As hipóteses consideradas estão relacionadas aos fatores que influenciam os pedestres serem oportunistas e a melhoria da previsibilidade baseado nestes fatores. Portanto busca-se verificar: (i) Características comportamentais tem maior influência do que características físicas dos usuários para a modelagem de classificação; (ii) Modelos menos restritivos (não lineares) se ajustam melhor aos dados indicando uma maior acurácia e generalização. A primeira Hipótese surge da necessidade de se verificar quais fatores realmente impactam no comportamento de travessia oportunista, enquanto a segunda engloba um aspecto mais metodológico sobre a modelagem com *machine learning*.

4.2. Coleta de Dados

Primeiramente foram coletados os dados relacionados ao comportamento dos pedestres na travessia, a partir de câmeras de videomonitoramento do CTAFor (instituição municipal responsável por monitorar e gerenciar o tráfego e os semáforos em Fortaleza). As informações obtidas foram referentes a demanda (fluxo de pedestres por hora) durante os intervalos de coleta, padrão de chegada, velocidade e atraso (segundos), formação de pelotão,

características dos pedestres (gênero e idade), tipo de travessia (durante verde, verde intermitente e vermelho) e se a travessia resultou em algum tipo de interação com o fluxo veicular (brecha de até 7s). A idade pode ser classificada por faixa etária, considerando os jovens como aqueles até 18 anos; os adultos entre 19 e 59 anos; e os idosos com mais de 60 anos. Os dados foram coletados com contagens nos horários fora pico (manhã: 9h - 11h; tarde: 14h-16h), durante três dias úteis de uma semana, resultando em um total de 12h de coleta em cada travessia.

Posteriormente foram levantados os dados de taxa de chegada, baseado no tempo de espera dos pedestres, em segundos, por contagem utilizando macros. Vale ressaltar que foram evidenciados apenas veículos individuais motorizados, ônibus e caminhões durante o período de vermelho para os pedestres como fator de influência de travessia.

Os dados referentes ao uso do solo e à geometria das interseções foram obtidos através da base georreferenciada de uso do solo referente ao ano de 2015 (em formato shape file) da SEFIN (Secretaria de Finanças do município de Fortaleza) e do Google Maps, onde foi possível coletar dados como largura das travessias, número de faixas e existência de canteiros, medida em metros, como largura total. As interseções em análise se localizam entre Av. Aguanambi com a R. Soriano Albuquerque e entre a Av. Santos Dumont com a Av. Desembargador Moreira (Tabela 1 e Figura 2).

Tabela 1 – Informações dos locais

Código da câmera	Logradouro I	ogradouro II	Nº de faixas	Distância	Velocidade permitida	Fluxo de pedestres (ped/h)	Fluxo veicular (veic/h)
C50	Aguanambi	Soriano Albuquerque	4	15,6 m	60 km/h	81	1656
N30	Santos Dumont	Desembargador Moreira	4	18 m	50km/h	41	961

Fonte: Autores.

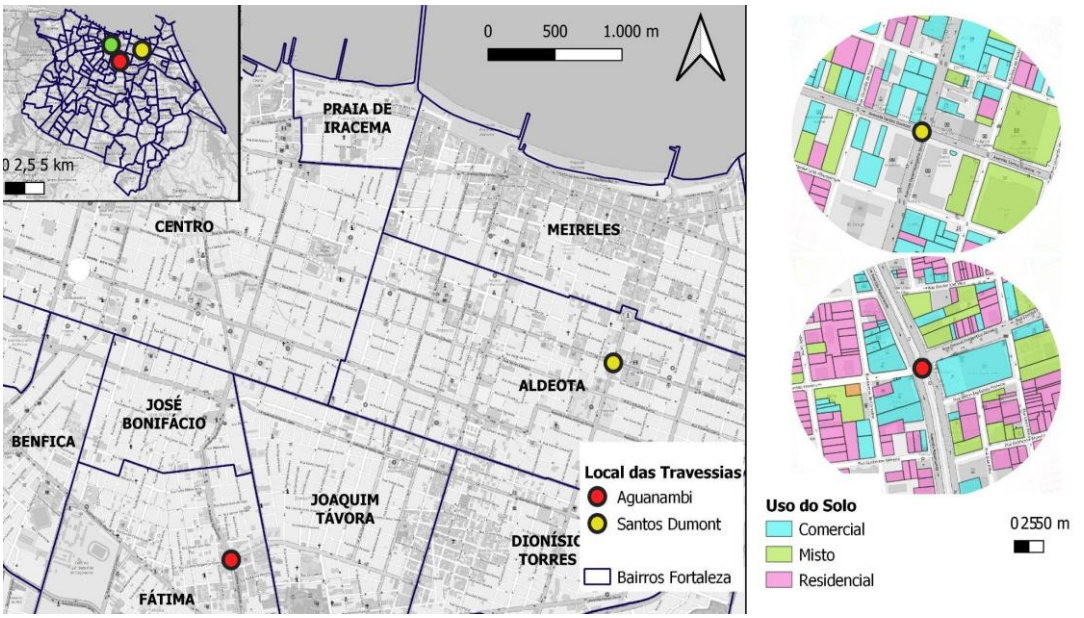

Figura 2 – Mapa de localização e uso do solo
Fonte: Autores.

4.3. Análises Exploratórias

No total foram levantados dados de 1522 pedestres, sendo classificados 57% como oportunistas e 43% como obedientes. Segregando os usuários pelo fator idade, tem-se que 11% dos oportunistas são jovens, 75% são adultos e 14% são idosos. Baseado unicamente nessa análise, as pessoas com faixa etária de 19 a 59 anos apresentam maiores tendências de avançarem o sinal, mas também sobre forte influência dessas regiões serem polos comerciais, e, portanto, apresentarem mais pedestres dessa categoria.

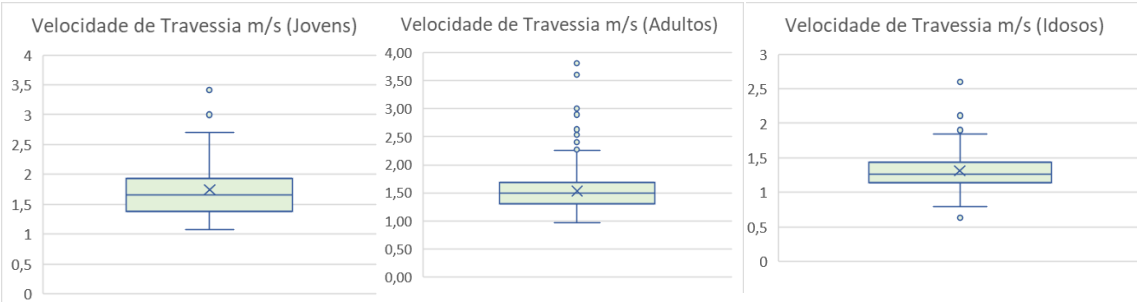

Figura 3 – *Boxplot* da velocidade de travessia por idade dos pedestres
Fonte: Autores.

A Figura 3 apresenta a distribuição das velocidades de caminhada de cada categoria de usuário, o qual é possível notar que as velocidades médias da categoria de idosos são mais reduzidas, enquanto as dos jovens são maiores que 1,6 m/s, sendo um agravante sobre o fato de atravessarem antes da sinalização adequada. O atraso médio na primeira interseção (C50) foi de 18,9 s, enquanto na interseção N30 foi de 12,5s, tendo esta segunda 43% de pedestres a menos na amostra dentro do mesmo intervalo de análise. Dessa forma o número de pessoas atravessando, bem como o tráfego veicular dão indícios de uma influência sobre o modo de travessia.

5. CARACTERIZAÇÃO DA MODELAGEM

5.1. Variáveis Levantadas

Conforme será apresentado no método de coleta dos dados, as variáveis levantadas foram definidas de forma a auxiliar na compreensão do comportamento de travessia dos pedestres e em sua posterior modelagem, sendo as seguintes: (i) Gênero; (ii) Idade; (iii) Faixa; (iv) Direção; (v) Grupo; (vi) Chegada; (vii) Início; (viii) Oportunista; (ix) Tempo de Travessia; (x) Atraso; (xi) Velocidade; e (xii) Interação.

As variáveis de gênero e idade correspondem às características dos usuários. Enquanto a variáveis de existência de faixa de pedestres é uma característica física do local. Todas as outras variáveis são relações entre o pedestre e o meio, como à exemplo a direção da travessia e o tempo de travessia. As variáveis categóricas como gênero (0 – F; 1 - M), idade (0 – Jovem; 1 – Adulto; 2 -Idoso), existência de faixa (0 – Não; 1 - Sim), direção (0 – Leste; 1- Oeste; 2 – Norte; e 4 – Sul), grupo (0 – Sozinho; 1 – 2 ou +), chegada (0 – Verde; 1 – amarelo, 2 - vermelho) e interação (0 – Não; 1 – Sim) foram transformadas em variáveis *dummy*, sendo o primeiro termo igual a 0, para possibilitar a inclusão de todos os fatores.

5.2. Modelagem e discussão da avaliação dos modelos

Posteriormente foi definido os parâmetros para cada modelo a fim de compreender e prever o padrão de oportunismo das travessias semaforizadas. Optou-se por uma modelagem baseada em *machine learning*, pois é capaz de prever com mais precisão possibilitando usar uma vasta quantidade de dados.

Foi definido um grupo de treinamento da base de dados, e outro para teste seguindo a seguinte proporção 7:3. Foram definidos modelos para prever a probabilidade do pedestre ser oportunista. Para a probabilidade, foram comparados os seguintes modelos de classificação supervisionados, por atenderem a necessidade para compreensão das relações entre as variáveis independentes e a dependente: *Random Forest (RF), Naive Bayes (NB)* e Regressão Logística (RL) (Géron, 2019). Os parâmetros utilizados para avaliar o nível de previsão dos modelos de classificação foram o *precision*, *recall* e *f1-score*.

A *precision* mede quantos elementos selecionados são relevantes (relação entre verdadeiros positivos e o total de observações da amostra), enquanto a *recall* mede quantos elementos relevantes foram selecionados (relação entre verdadeiros positivos e o total de positivos). Positivos são os usuários que atravessaram no momento do sinal vermelho (oportunistas). O *f1-score* é um parâmetro que indica a qualidade geral do modelo (quanto mais próximo de 1, melhor o nível de predição).

O modelo *Random Forest* faz parte do conjunto de modelos *ensemble,* que se beneficiam da combinação de diferentes modelos de árvore de decisão para se obter um único resultado, necessitando de um maior custo computacional. Para problemas de classificação é utilizado o valor com maior moda. Para a árvore de decisão a escolha das variáveis que comporão cada nó é aleatória. A premissa principal é a de que não é possível definir uma relação linear entre variáveis e que a curva de separação entre os grupos não é linear.

O algoritmo *Naive Bayes* é um classificador probabilístico. Sua escolha se deve pelo desempenho ser superior ao de outros modelos de classificação, necessitando de um pequeno conjunto de dados para obter resultados precisos. O modelo baseia-se na premissa de independência das variáveis, prevendo a classe que pertence a variável de acordo com probabilidades predefinidas baseada nas características da amostra.

A regressão logística trata-se de um caso particular dos modelos lineares generalizados. Seu objetivo é produzir, a partir de um conjunto de observações, um modelo que permita a predição de valores tomados por uma variável categórica, sendo sua curva de probabilidade não linear. O *logit* é o único dentre os 3 modelos que separa as classes com uma relação linear, ou seja, é um modelo mais restritivo que os outros dois (Gerón, 2019).

Após a modelagem também foi avaliado a capacidade preditiva dos modelos de classificação com os dados de teste, verificando se o modelo realmente “adaptou-se”, ou seja, consegue expandir aceitavelmente seus resultados de predição para conjuntos de dados desconhecidos. O modelo também pode ser classificado como ocorrendo um *underfitting*, quando não consegue adaptar-se aos padrões dos dados para predição ou *overfitting*, quando aprende o padrão apenas dos dados de treinamento, não conseguindo expandir para novos conjuntos. Portanto, foram avaliados com base nos seguintes indicadores de qualidade de previsão: a matriz de confusão, na Raiz do Erro Quadrático Médio (RMSE - *Root Mean Squared Error*) e no indicador *Area Under the ROC Curve* (AUC). A *ROC Curve* relaciona o número de falsos positivos com os verdadeiros positivos de diferentes amostras de testes de dados.

6. AVALIAÇÃO DOS MODELOS

6.1. Discussão da relação entre as variáveis e o fenômeno de oportunismo na travessia

Baseado na literatura levantada na análise de correlação e de sensibilidade dos parâmetros foram definidas as variáveis preditoras dos modelos de *machine learning*. A Figura 4 apresenta a correlação entre todas as variáveis, de modo que os coeficientes negativos indicam que quanto maior o valor do parâmetro menor a probabilidade do usuário ser oportunista.

Tomando como exemplo a variável grupo, o coeficiente negativo indica que quanto mais aglomerado os usuários, menor a chance de ele ser oportunista, tendendo a ocorrer quando ele está sozinho (baseado apenas nesta análise). Dessa forma a variável a ser prevista foi a probabilidade do usuário ser oportunista, e as preditoras foram direção, grupo, chegada e interação. As análises iniciais indicam forte relação do modo semafórico e do fato do usuário estar em grupo, como o fato dele ser oportunista. Apenas com essa análise não foi possível identificar se as características dos usuários interferem nesse fenômeno. Vale destacar que a variável "início" equivale ao tempo semafórico vermelho quando o pedestre iniciava a caminhada, por isso sua correlação foi de quase 100%, causando um sobreajuste dos dados em todos os modelos calibrados, portanto ela foi desconsiderada para maior veracidade e capacidade de predição.

As variáveis selecionadas correspondem a direção de travessia, grupo, chegada e interação. As características dos usuários (idade e gênero) se mostraram pouco relevantes para a predição, e em segundo lugar de relevância as características físicas dos cruzamentos. Dessa forma as variáveis utilizadas correspondem as características do comportamento de travessia dos pedestres.

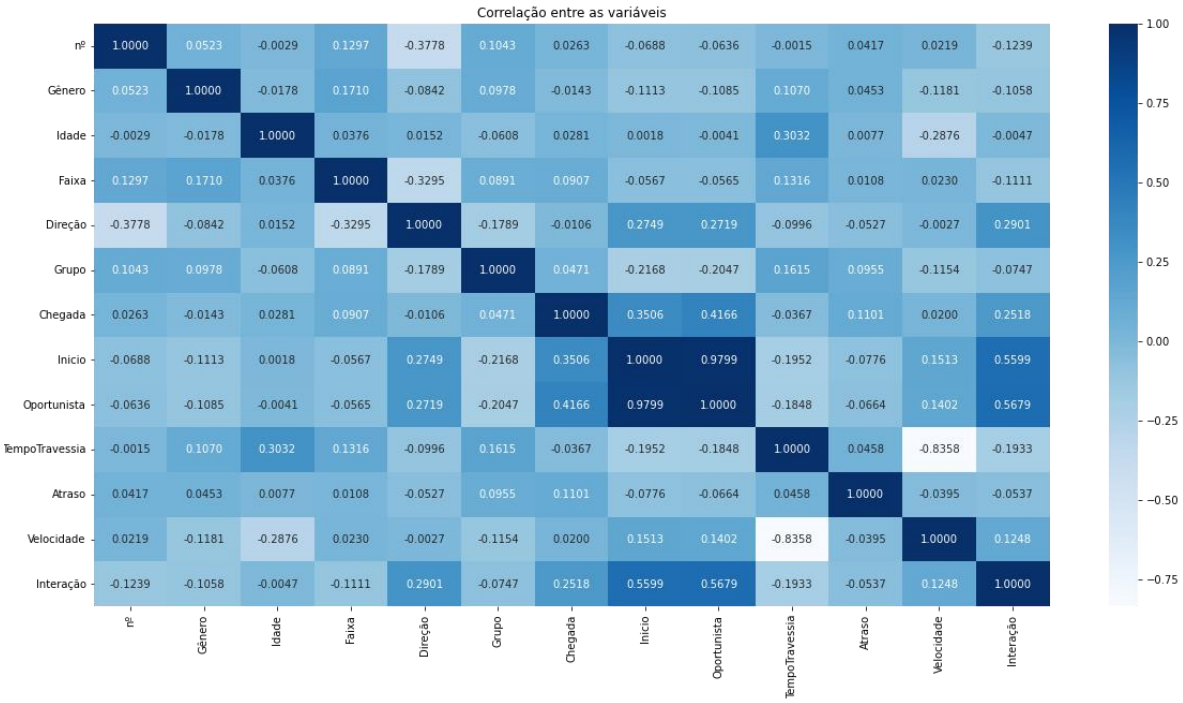

Figura 4 – Matriz de Correlação entre as variáveis
Fonte: Autores.

6.2. Resultados da modelagem e análise dos modelos

Cada modelo de probabilidade foi treinado utilizando dados de treino. Posteriormente os dados independentes para teste foram utilizados de forma a realizar previsões e por fim calculados o nível de acurácia de cada modelo (Tabela 2). A acurácia do modelo NB pelo parâmetro *f1-score* foi o mais baixo com 0,77 enquanto o RF e o RL foram acima do desejado (0,8) com 0,84 e 0,82, respectivamente. Levando-se em consideração os parâmetros em conjunto, o modelo mais regular é o RF, pois obteve praticamente todos os parâmetros acima de 0,80 (excetuando-se o recall dos usuários obedientes). Isto indica alguma não linearidade na relação entre as variáveis.

Tabela 2 – Resultados da modelagem da probabilidade de oportunismo ao atravessar.

	Regressão Logística			**Naive Bayes**			**Random Forest**		
	1	0	accuracy	1	0	accuracy	1	0	accuracy
precision	0,89	0,72	-	0,90	0,64	-	0,86	0,83	-
recall	0,76	0,87	-	0,64	0,89	-	0,88	0,79	-
f1 - socore	0,82	0,79	0,82	0,75	0,74	0,77	0,87	0,81	0,84

Fonte: Autores.

A Figura 5 apresenta as curvas ROC para medir o aprendizado dos modelos. A curva ROC do modelo RL apresentou valor de acurácia (AUC) = 0,82, enquanto o modelo RF obteve um AUC = 0,84, e por fim o *Naive Bayes* obteve AUC = 0,77. Considerando o mínimo de acurácia em 0,8, o modelo de nayve bayes não se mostra eficiente para este tipo de predição.

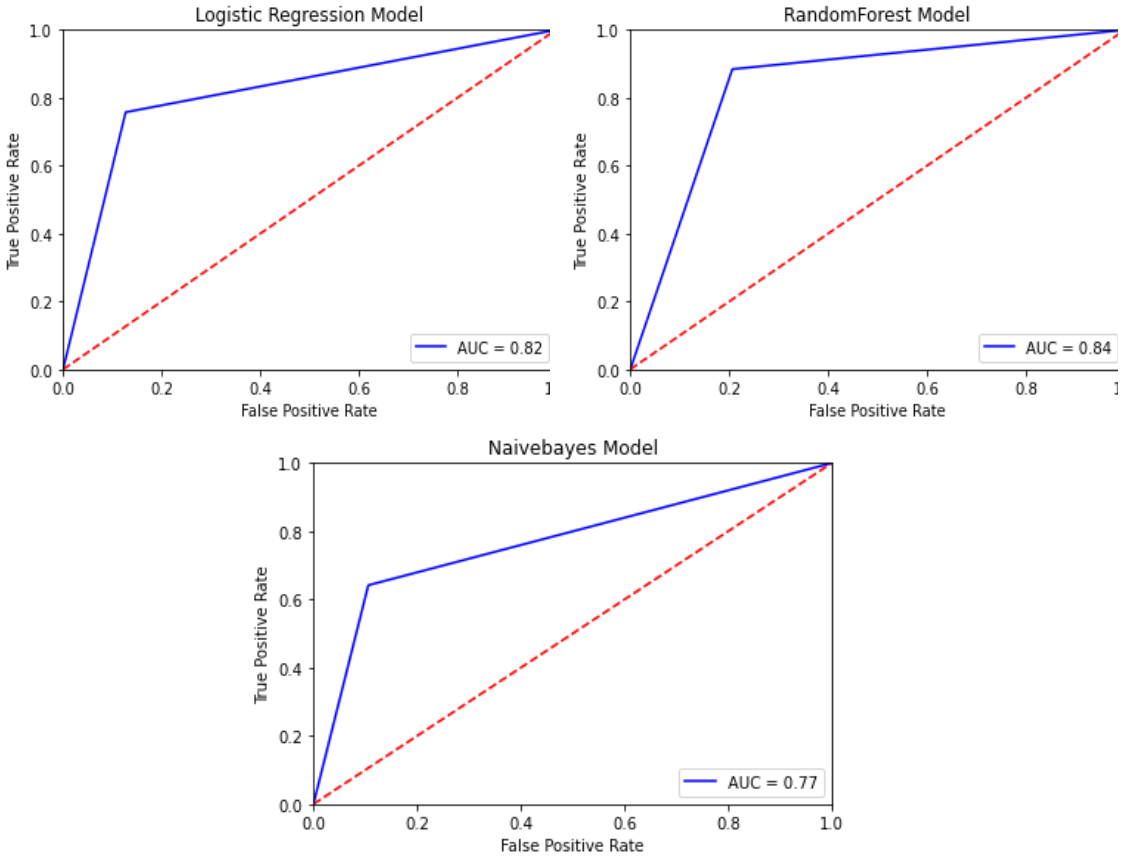

Figura 5 – Curva ROC para os modelos RL, RF e NB

Fonte: Autores.

A área sobre a curva ROC do modelo RL foi superior ao modelo NB, indicando que seu conjunto consegue generalizar melhor as predições, porém o modelo NB consegue classificar corretamente mais rápido (com menos amostras), visível pela mudança na inclinação da curva ROC, se comparado ao primeiro modelo. Por fim, o erro quadrático médio (RMSE) dos modelos RL, NB e RF foram de 0,44, 0,50 e 0,39, respectivamente. O modelo RF por conter acurácia aceitável e todos os parâmetros regulares, foi escolhido para realizar a previsão da variável de probabilidade do pedestre ser oportunista durante uma travessia em vias semaforizadas.

7. CONSIDERAÇÕES FINAIS

Diante dos resultados obtidos, houve-se uma maior compreensão do fenômeno do comportamento dos pedestres na região de estudo da cidade de Fortaleza. Os resultados podem auxiliar os tomadores de decisão na elaboração e implementação de ferramentas capazes de prever o comportamento de risco dos pedestres.

Foram desenvolvidos três modelos de *machine learning* para prever o comportamento obediente do pedestre. O modelo de *random forest* apresentou a maior acurácia, sendo o modelo escolhido para prever o comportamento dos pedestres nas travessias. Além de ter tido desempenho superior para predição das categorias quando analisado de forma segregada também. O modelo de *naive bayes* teve a melhor taxa de generalização, enquanto o modelo *logit* mesmo sendo um modelo linear, apresentou resultados satisfatórios. De acordo com as hipóteses levantadas, ao final do trabalho acredita-se que características da travessia tem maior influência do que características físicas dos pedestres, embora essas corroborem o fenômeno. Outro ponto importante é que quanto menos restritivo for o modelo, há indícios de que melhor é seu nível de previsão e menor será o erro médio.

Uma das limitações deste trabalho é não incluir formas mais diversas de uso do solo no estudo, uma vez que a quase totalidade das interseções está em áreas centrais que concentram muito uso do solo não residencial. Portanto, é recomendável que haja uma análise mais desagregada quanto ao uso do solo, pois esse é um atributo relevante para explicar o comportamento dos pedestres.

Referências

Alhajyaseen, W. K. M., e Iryo-Asano, M. (2017) Studying critical pedestrian behavioral changes for the safety assessment at signalized crosswalks. Safety Science, 91, 351–360. doi:10.1016/j.ssci.2016.09.002.

Alver, Y., Onelcin, P., Cicekli, A., e Abdel-Aty, M. (2021) Evaluation of pedestrian critical gap and crossing speed at midblock crossing using image processing. Accident Analysis and Prevention, 156(April), 106127. doi:10.1016/j.aap.2021.106127.

ARIOTTI, Paula. Análise do padrão de comportamento de pedestres em travessias semaforizadas. 2006.

Bandeira, T.P. A *Method for Microscopic Modeling of Pedestrian Delay at Signalized Crossings*. Dissertação de Mestrado. Universidade Federal do Ceará. Fortaleza:2019.

Bernhoft, I. M.; Carstensen, G. *Preferences and behaviour of pedestrians and cyclists by age and gender*. Transportation Research Part F: Traffic Psychology and Behaviour 11.2 (2008): 83-95. 16.

Demir, B., Özkan, T., e Demir, S. (2019) *Pedestrian violations: Reasoned or social reactive? Comparing theory of planned behavior and prototype willingness model*. Transportation Research Part F: Traffic Psychology and Behaviour, 60, 560–572. doi:10.1016/j.trf.2018.11.012

Géron A. (2019) Hands-On Machine Learning with Scikit-Learn, Keras, and Tensorflow: Concepts, Tools, and Techniques to Build Intelligent Systems. O'Reilly Media. v 1. 856 p.

Jacobsen, A. C. *Microssimulação da Travessia de Pedestres*. Dissertação de Mestrado. Universidade Federal do Rio Grande do Sul: 2011.

Jay, M., Régnier, A., Dasnon, A., Brunet, K., e Pelé, M. (2020) *The light is red: Uncertainty behaviours displayed by pedestrians during illegal road crossing*. Accident Analysis and Prevention, 135(November 2019), 105369. doi:10.1016/j.aap.2019.105369

Koh, P. P., e Wong, Y. D. (2014) *Gap acceptance of violators at signalised pedestrian crossings*. Accident Analysis and Prevention, 62, 178–185. doi:10.1016/j.aap.2013.09.020

Liu, R.; Silva, J.P.;Seco, A.J.*A bi-modal micro-simulation tool for the assessment of pedestrian delas and traffic management*.9th International Association of Travel Behavior Research Conference., Gold Coast, Australia: 2000.

Mukherjee, D., e Mitra, S. (2019) *A comparative study of safe and unsafe signalized intersections from the view point of pedestrian behavior and perception*. Accident Analysis and Prevention, 132(October 2018), 105218. doi:10.1016/j.aap.2019.06.010

Mukherjee, D., e Mitra, S. (2020) *A comprehensive study on factors influencing pedestrian signal violation behaviour: Experience from Kolkata City, India*. Safety Science, 124(October 2019), 104610. doi:10.1016/j.ssci.2020.104610

Murphy, Brendan; Levinson, David M.; Owen, Andrew. *Evaluating the Safety In Numbers effect for pedestrians at urban intersections*. Accident Analysis & Prevention, v. 106, p. 181-190, 2017.

Patra, M., Perumal, V., e Rao, K. V. K. (2019) *Modelling the effects of risk factor and time savings on pedestrians' choice of crossing facilities at signalised intersections*. Case Studies on Transport Policy, (October), 0–1. doi:10.1016/j.cstp.2019.10.010

Ronsenbloom, T.; Nemrodoy, D.; Barkan H. *For heaven's sake follow the rules: pedestrians' behavior in an ultra-orthodox and a non-orthodox city*. Transportation Research Part F: Traffic Psychology and Behaviour 7.6 (2004): 395-404.

Vignali, V., Pazzini, M., Ghasemi, N., *Lantieri, C., Simone, A., e Dondi, G. (2020) The safety and conspicuity of pedestrian crossing at roundabouts: The effect of median refuge island and zebra markings.* Transportation Research Part F: Traffic Psychology and Behaviour, 68, 94–104. doi:10.1016/j.trf.2019.12.007

WRI BRASIL. (2015) Impactos da redução dos limites de velocidade em áreas urbanas. Embarq Brasil.

Yang, J. *et al. Modeling pedestrians' road crossing behavior in traffic system micro-simulation in China*. Transportation Research Part A, v. 40, p. 280-290. 2006

Zafri, N. M., Sultana, R., Himal, M. R. H., e Tabassum, T. (2020) *Factors influencing pedestrians' decision to cross the road by risky rolling gap crossing strategy at intersections in Dhaka, Bangladesh*. Accident Analysis & Prevention, 142(December 2019), 105564. doi:10.1016/j.aap.2020.105564

Zhou, H., Romero, S. B., e Qin, X. (2016) *An extension of the theory of planned behavior to predict pedestrians' violating crossing behavior using structural equation modeling*. Accident Analysis and Prevention, 95, 417–424. doi:10.1016/j.aap.2015.09.009

Zhuping, Z *et al. Pedestrian crossing behaviors at signalized intersections: observational study and survey in China*. In: TRANPORTATION RESEARCH BOARD 90th ANNUAL MEATING, 2011, Washington, DC. ANAIS…Washington, DC.:TRB, 2011. DVD-ROMAssociação Brasileira de Normas Técnicas, Rio de Janeiro.

Zhang, D. *Evaluation System of Road Traffic Safety Management*. China Communication Press, Beijing, 2005.

Zhang, C., Chen, F., e Wei, Y. (2019) *Evaluation of pedestrian crossing behavior and safety at uncontrolled mid-block crosswalks with different numbers of lanes in China*. Accident Analysis and Prevention, 123(January 2018), 263–273. doi:10.1016/j.aap.2018.12.002

Zhang, S., Abdel-Aty, M., Yuan, J., e Li, P. (2020) Prediction of Pedestrian Crossing Intentions at Intersections Based on Long Short-Term Memory Recurrent Neural Network. Transportation Research Record, 2674(4), 57–65. doi:10.1177/0361198120912422.