

UNIVERSIDADE FEDERAL DO RIO GRANDE DO SUL
FACULDADE DE CIÊNCIAS ECONÔMICAS
PROGRAMA DE PÓS-GRADUAÇÃO EM ECONOMIA

HENRIQUE JONER

**INFERÊNCIA PREDITIVA GEOESPACIAL DA CRIMINALIDADE
EM PORTO ALEGRE: UMA ABORDAGEM DE APRENDIZADO
DE MÁQUINA**

**PORTO ALEGRE
2020**

HENRIQUE JONER

**INFERÊNCIA PREDITIVA GEOESPACIAL DA
CRIMINALIDADE EM PORTO ALEGRE: UMA
ABORDAGEM DE APRENDIZADO DE MÁQUINA**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia.

Orientador:

Prof. Dr. Nelson Seixas dos Santos

Coorientador:

Prof. Dr. Carlos Schönerwald

**PORTO ALEGRE
2020**

Joner, Henrique.

INFERÊNCIA PREDITIVA GEOESPACIAL DA CRIMINALIDADE
EM PORTO ALEGRE: UMA ABORDAGEM DE APRENDIZADO DE
MÁQUINA / HENRIQUE JONER. -- 2020.

93 f.

Orientador: Dr. Nelson Seixas dos Santos

Coorientador: Dr. Carlos Schönerwald

Dissertação (Mestrado) - Universidade Federal do
Rio Grande do Sul, Faculdade de Ciências Econômicas,
Programa de Pós-Graduação em Economia, Porto Alegre,
BR-RS, 2020.

Aprendizagem de Máquina. Economia do Crime.
Inferência Preditiva. Predição do Crime. Segurança
Pública. I. Santos, N.S. orient. II. Schönerwald, C.
E., coorient. III. Título.

HENRIQUE JONER

**INFERÊNCIA PREDITIVA GEOESPACIAL DA
CRIMINALIDADE EM PORTO ALEGRE: UMA
ABORDAGEM DE APRENDIZADO DE MÁQUINA**

Dissertação submetida ao Programa de Pós-Graduação em Economia da Faculdade de Ciências Econômicas da UFRGS, como requisito parcial para obtenção do título de Mestre em Economia.

Aprovado em: Porto Alegre, 31 de Julho de 2020.

BANCA EXAMINADORA:

Prof. Dr. Nelson Seixas dos Santos - Orientador (UFRGS)

Prof. Dr. Carlos Eduardo Schönerwald da Silva - Coorientador (UFRGS)

Prof. Dr. Giacomino Balbinotto Neto (UFRGS)

Prof. Dr. Fernando Augusto Boeira Sabino da Silva (UFRGS)

Prof. Dr. Luiz Marcelo Berger (UFRGS)

AGRADECIMENTOS

Agradeço aos meus pais Gilberto Fernando Joner e Bernadete Maria de Oliveira Joner por incentivarem desde o início da minha jornada acadêmica a dedicação ao estudo e ao desenvolvimento profissional.

Aos meus irmãos Fernando Joner e Juliana Joner, que são grande influência direta na busca incessante por conhecimento.

À UFRGS e ao PPGE por fornecer os recursos necessários para a elaboração e o desenvolvimento deste trabalho, a partir do empenho todos os professores e servidores com quem tive contato.

Ao Instituto-Geral de Perícias por encorajar constantemente todos os seus servidores à investigação e busca pela verdade dos fatos, incentivando a participação destes em todas as áreas do conhecimento científico.

Ao Observatório de Segurança Pública, tanto ao Lucas quanto ao Maurício, mas em especial ao seu coordenador Major Roberto Donato pela disponibilidade em me receber sempre que necessário e por acreditar no potencial deste projeto, fornecendo todos os dados e recursos necessários para a sua elaboração.

Aos meus colegas do Departamento Médico-Legal, em especial a Ana Nunes, Caroline Werber, Eduardo Techera e Dorval Miele Jr. pela parceria e suporte nos momentos mais críticos do desenvolvimento deste estudo.

Ao Valmor Mantelli Jr. pelo aprendizado proporcionado ao longo de todo o mestrado nas nossas conversas dentro e fora da sala de aula.

Aos meus orientadores Nelson Seixas e Carlos Schönerwald pela dedicação na implementação desta pesquisa e na incessante busca pelo aprimoramento do trabalho desenvolvido.

Por fim, à minha esposa, Gabriella T. Schorn, pela parceria de anos, unindo sempre o amor e o desenvolvimento intelectual, marca da nossa união. Sempre estimulando o trabalho nos momentos mais difíceis. Foi a todo instante o suporte necessário sem o qual não seria possível a realização desta pesquisa.

RESUMO

Novas estratégias para o enfrentamento da criminalidade no Brasil são necessárias, haja vista o recorde dos índices de crimes violentos registrados nos últimos anos. Dessa forma, o objetivo desta pesquisa é demonstrar o potencial da utilização da inteligência artificial como ferramenta no combate à criminalidade. Foram testados quatro tipos diferentes de modelos na predição de eventos criminosos e do nível de criminalidade em cada localidade, sendo eles: regressão, classificação, redes neurais profundas e *long short-term memory*. O estudo analisou 351.980 crimes violentos ocorridos na cidade de Porto Alegre/RS entre janeiro de 2005 e outubro de 2019. O desempenho de cada algoritmo construído foi testado prevendo os eventos criminosos diários em diferentes números de clusters no qual a cidade foi subdividida. Os resultados apontam que todos os modelos utilizados tem capacidade significativa na predição de crimes, com destaque para o modelo de classificação construído, que ao utilizar 6 clusters de criminalidade obteve um erro médio absoluto (MAE) de 0.43 e a raiz quadrada do erro médio quadrático (RMSE) de 1.68, para a previsão de crimes diários em cada cluster, obtendo um coeficiente de determinação de 0.94. Quando o objetivo era prever o nível de criminalidade em cada cluster, o mesmo modelo de classificação obteve um R^2 de 0.98, MAE de 0.01 e RMSE de 0.07.

Palavras-chave: Aprendizagem de Máquina. Economia do Crime. Inferência Preditiva. Predição do Crime. Segurança Pública.

ABSTRACT

New strategies for fighting crime in Brazil are required, since the records of violent crimes keep increasing in recent years. The objective of this research is to demonstrate the potential of using artificial intelligence as a tool for crime reduction. Four different types of models were tested in the prediction of criminal events and level of crime in each location, namely: regression, classification, deep neural networks and long short-term memory. This study analyzed 351,980 violent crimes that occurred in the city of Porto Alegre/RS between January 2005 and October 2019. The accuracy of each algorithm was tested by predicting daily number of crimes in each different cluster of criminal events. Results point that all models have significant capacity in the prediction of crimes, with emphasis on the classification model, that using 6 criminality clusters obtained a MAE of 0.43 and a RMSE of 1.68, forecasting daily crimes in each cluster, in this case the coefficient of determination obtained was 0.94. When the objective was to predict the level of crime in each cluster, the same classification model registered a R^2 of 0.98, MAE of 0.01 and a RMSE of 0.07, demonstrating great prediction capacity.

Keywords: Machine Learning. Economics of Crime. Predictive Inference. Crime Prediction. Public Security

Sumário

Lista de Figuras	viii
Lista de Tabelas	x
Lista de Símbolos	xi
1 INTRODUÇÃO	10
2 CONJUNTURA DA CRIMINALIDADE NO BRASIL	13
3 TEORIA ECONÔMICA DO CRIME	19
4 METODOLOGIA	24
4.1 APRENDIZAGEM DE MÁQUINA	26
4.1.1 Classificação	28
4.1.2 Regressão	29
4.1.3 Clusterização <i>K</i> -means	30
4.1.4 Validação Cruzada	32
4.1.5 Árvores de Decisão	33
4.1.6 Redes Neurais Artificiais	35
4.1.7 Long Short-Term Memory	38
4.2 MÉTRICAS DE AVALIAÇÃO DOS RESULTADOS	41
5 RESULTADOS E DISCUSSÃO	47
5.1 ANÁLISE EXPLORATÓRIA DOS DADOS	47
5.2 PREVISÃO DA CRIMINALIDADE	57
5.2.1 Regressão	60
5.2.2 Classificação	69
5.2.3 DNN	73
5.2.4 LSTM	77
5.2.5 Comparativo entre os modelos	80
5.3 DISCUSSÃO	85
6 Conclusão	87
Referências Bibliográficas	89

Lista de Figuras

Figura 2.1	Evolução da Taxa de Homicídios para cada 100 mil Hab. no RS . . .	15
Figura 2.2	Participantes do programa RS SEGURO e sua representatividade na criminalidade do estado.	16
Figura 2.3	Crimes Contra a Vida nas imediações do Centro de Porto Alegre entre 2005 e 2019	17
Figura 4.1	Exemplo de classificação multiclasse	28
Figura 4.2	Exemplo de regressão linear	30
Figura 4.3	Exemplo de clusterização	31
Figura 4.4	Demonstração dos subconjuntos da validação cruzada	32
Figura 4.5	Exemplo de árvore de decisão sobre jogar tênis ou não	33
Figura 4.6	Exemplo de partição das regiões em um modelo de duas dimensões	34
Figura 4.7	Exemplo árvore binária	34
Figura 4.8	Funções de ativação Sigmóide e Tangente Hiperbólica	35
Figura 4.9	Demonstração de uma Rede Neural Artificial	36
Figura 4.10	Demonstração da taxa de aprendizagem na predição.	38
Figura 4.11	Demonstração genérica de uma célula LSTM	39
Figura 5.1	Todos os crimes acumulados anualmente	49
Figura 5.2	Possibilidades de clusterização	50
Figura 5.3	Proporção de homicídios por cluster	51
Figura 5.4	Proporção de roubos a pedestres por cluster	52
Figura 5.5	Proporção do roubos a comércios por cluster	53
Figura 5.6	Proporção do roubos a veículo por cluster	54
Figura 5.7	Evolução da criminalidade nos 42 clusters	55
Figura 5.8	Crimes por hora	56
Figura 5.9	Clusterização em 42 grupos	58
Figura 5.10	Acumulado dos crimes nos 30 dias anteriores	59
Figura 5.11	Comparação dos atributos Crime vs Risco	60
Figura 5.12	Treinamento do Algoritmo	61
Figura 5.13	Real vs Previsão	64
Figura 5.14	Comparativo dos últimos 6 meses da série - Modelo com 6 clusters	65
Figura 5.15	Risco Real vs Previsão do Risco	67
Figura 5.16	Comparação do Risco para 42 Clusters: Previsão vs Real	68
Figura 5.17	Risco Real vs Previsão do Risco	70
Figura 5.18	Comparativo dos últimos 6 meses da série (Continua)	72

Figura 5.19	Comparativo dos últimos 6 meses da série (Continuação)	73
Figura 5.20	Comparativo diário nos últimos três mês da série (continua)	74
Figura 5.21	Comparativo diário nos últimos três mês da série (continuação) . . .	75
Figura 5.22	Risco ao longo da série para cluster	76
Figura 5.23	Comparativo dos últimos 6 meses da série (Continua)	78
Figura 5.24	Comparativo dos últimos 6 meses da série (continuação)	79
Figura 5.25	Risco ao longo da série para cluster	80
Figura 5.26	Comparativo diário no último mês da série	82
Figura 5.27	Comparativo do risco ao longo da série	84

Lista de Tabelas

Tabela 5.1	Distribuição dos eventos antes da limpeza dos dados	48
Tabela 5.2	Distribuição dos eventos após a limpeza dos dados	48
Tabela 5.3	Evolução dos crimes de 2005 a 2018	49
Tabela 5.4	Autocorrelação e Estacionariedade do modelo de Regressão prevendo Crime	62
Tabela 5.5	Métricas do modelo de Regressão prevendo Crime	62
Tabela 5.6	Autocorrelação do modelo de Regressão prevendo Risco	66
Tabela 5.7	Métricas do modelo de Regressão prevendo Risco	67
Tabela 5.8	Métricas do modelo de Classificação prevendo Risco	69
Tabela 5.9	Métricas do modelo de Classificação prevendo Crime	71
Tabela 5.10	Métricas do modelo DNN prevendo Crime	73
Tabela 5.11	Métricas do modelo DNN prevendo Risco	75
Tabela 5.12	Métricas do modelo LSTM prevendo Crime	77
Tabela 5.13	Métricas do modelo LSTM prevendo Risco	79
Tabela 5.14	Métricas para Crime - Modelos com 6 Clusters	81
Tabela 5.15	Métricas para Risco - Modelos com 6 Clusters	83

Capítulo 1

INTRODUÇÃO

A alocação de recursos dos órgãos de segurança pública é um problema central para a prevenção à criminalidade. Estratégias adequadas quanto a organização do policiamento ostensivo, por exemplo, podem impedir que crimes ocorram ao frustrarem a ação criminosa. Ao mesmo tempo, estratégias pouco eficazes podem provocar danos significativos à sociedade. Dessa forma, compreender melhor a organização das atividades criminosas e procurar se antecipar a eventos deste tipo podem trazer grandes benefícios para todos os cidadãos.

As estratégias para otimizar a alocação do efetivo policial distribuído geograficamente podem estar fundamentadas em diversas teorias sociais e econométricas, assim como, na experiência profissional dos gestores. Este estudo procura demonstrar que o emprego de técnicas mais avançadas de estatística computacional podem auxiliar substancialmente na gestão dos recursos e por consequência aumentar significativamente a utilidade destes.

Apesar de ser uma atividade complexa, a previsão de eventos criminosos pode ser realizada se utilizados os métodos adequados. Dessa forma, este trabalho objetiva demonstrar o potencial do emprego de técnicas de aprendizado de máquina ao problema da previsão da criminalidade de modo a servir de subsídio à solução do problema de alocação da força policial.

Através da resposta a este problema, outras questões poderão ser compreendidas, como por exemplo, de que forma se dá a evolução espacial da criminalidade em

Porto Alegre ao longo das últimas décadas. A expectativa é que a compreensão deste problema possa influenciar diretamente na estratégia de combate à criminalidade na capital do estado.

A base de dados foi disponibilizada pelo Observatório da Segurança Pública (OSP), pertencente ao Gabinete de Gestão Integrada Estadual (GGI-E) do Governo do Estado do Rio Grande do Sul. O banco de dados é abastecido pelos órgãos de segurança pública, como a Polícia Civil (PC), Brigada Militar (BM), Instituto-Geral de Perícias (IGP) e Superintendência dos Serviços Penitenciários (SUSEPE). Os crimes que serão analisados são de quatro tipos: crimes violentos contra a vida, roubo a veículos, roubo a pedestres e roubo a estabelecimentos comerciais e de ensino. Estes são descritos detalhadamente mais adiante na metodologia, onde as características dos dados são apresentadas.

A evolução da criminalidade é uma preocupação de todas as nações. Cada região possui características criminais diferentes. Em algumas, os crimes contra a vida são muito representativos e exigem grande esforço dos órgãos de combate à criminalidade. Em outras, os crimes contra a propriedade, violência doméstica ou sexual são aqueles que mais exigem empenho. Seja qual for o crime, a velocidade e expansão geográfica com que ele se transforma, exigem maior capacidade de enfrentamento a cada dia.

A capital do Rio Grande do Sul, Porto Alegre, apesar de possuir aproximadamente 15% da população do estado, registra mais da metade dos casos de crimes contra a vida. Observando isso e levando em consideração a qualidade dos dados disponibilizados neste espaço geográfico, optou-se por dedicar este estudo apenas à esta cidade do estado. Dessa forma, após a devida análise dos resultados aqui apresentados será possível replicar facilmente o mesmo estudo para as demais localidades.

O difícil momento de gestão pública exige mais do que nunca o auxílio de dados, tecnologia e ferramentas estatísticas a fim de aumentar a eficácia no combate à criminalidade. Nas últimas décadas, as metodologias de análise e previsão de dados têm evoluído em grande velocidade, de forma que muitos métodos surgiram ou foram aperfeiçoados recentemente. Sendo assim, ao buscar uma solução através de métodos preditivos não só é possível fazê-lo a partir de diferentes ferramentas, como

é necessário, já que cada método pode ter um desempenho superior aos demais em diferentes cenários. Por esse motivo, este estudo optou por construir diferentes modelos de aprendizagem de máquina, possibilitando que estes tenham o seu desempenho comparado.

A previsão da criminalidade é uma ferramenta relativamente nova para a organização do combate ao crime, como aponta Gorr e Harries (2003). Prever o comportamento criminoso de um indivíduo pode ser impossível, porém compreender o comportamento de uma série de eventos criminosos em determinada área é algo muito mais fácil de ser realizado, por isso a escolha por utilizar dados geoespaciais. A qualidade dos dados disponibilizados pelos sistemas de informações geográficas (SIG) vem melhorando consideravelmente nos últimos anos (GORR; HARRIES, 2003). Dessa forma aliando essa evolução dos SIG's com a base de dados disponibilizada pelo órgão responsável pela análise dos dados de segurança pública no estado do Rio Grande do Sul, acredita-se que a presente pesquisa traz resultados importantes quanto à predição do comportamento criminoso nas áreas mais sensíveis a eventos criminosos da cidade de Porto Alegre.

Dessa forma, será apresentado a seguir um breve histórico da situação do Brasil e do Rio Grande do Sul quanto à criminalidade. Após isso, aborda-se os aspectos principais quanto à teoria econômica do crime e demais trabalhos anteriores realizados na temática. Depois será abordada a metodologia empregada neste estudo, destacando o funcionamento dos algoritmos de aprendizado de máquina e as métricas de avaliação dos resultados. Na sequência demonstra-se os resultados obtidos nesta pesquisa e discute-se a utilidade das observações realizadas e o impacto na gestão da segurança pública. Por fim, é apresentada a conclusão com as observações gerais acerca do estudo desenvolvido.

Capítulo 2

CONJUNTURA DA CRIMINALIDADE NO BRASIL

O Brasil tem assistido seus índices de criminalidade progredir a taxas aceleradas na última década. Em 2017, o país atingiu a marca de 31,6 homicídios para cada 100 mil habitantes ao superar o patamar de 65 mil mortes por ano (IPEA, 2020). Nos últimos dez anos foram mais de 553 mil vítimas de homicídios dolosos no Brasil, o que nos dá uma média anual superior à Síria, que vive uma guerra civil sangrenta noticiada amplamente pela mídia internacional.

Como demonstrado pelo Atlas da Violência 2019, as ocorrências de homicídios somaram 62.517 casos em 2017, resultando em uma taxa de 31,6 casos para cada 100 mil habitantes. Em comparação a eventos históricos e externos à realidade brasileira, pode-se dizer que o Brasil necessitaria apenas de 1 mês para produzir o número de homicídios em toda a União Europeia no ano 2016, mesmo este possuindo apenas 40% da sua população. Com isso, equivale dizer que o Brasil possui uma taxa de homicídios trinta vezes maior que a da União Européia. Em comparação a um cenário de conflito intenso, este volume de homicídios anual do Brasil equivale a cinco anos de vítimas da guerra do Iraque somados.

Em relação aos crimes contra a propriedade, em 2016, os casos de roubos e furtos de veículos são estimados em 455.617 no país (SINESP, 2018), representando uma média de 588,13 ocorrências para cada 100 mil habitantes. O estado com mais casos

por habitantes é São Paulo, com uma taxa de 914,16. Já o estado que apresenta o menor número de casos é a Paraíba, com uma taxa de 77,33 por habitantes. O Rio Grande do Sul possui um índice de 515,74 casos de roubos e furtos de veículos para cada 100 mil habitantes, portanto, ligeiramente inferior a média nacional.

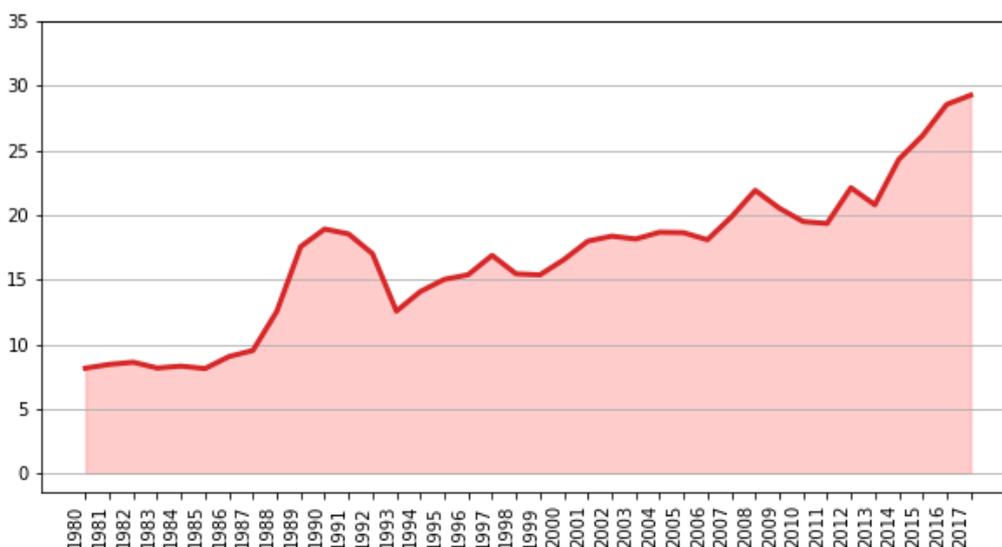
Ao mesmo tempo que os crimes anteriormente destacados têm cada vez maior frequência no cotidiano da população brasileira, a eficácia do poder público no combate a estes crimes é questionada. A mídia frequentemente apresenta séries de reportagens acompanhando o desfecho dos processos que envolvem crimes contra a vida, já que é um tema de grande interesse para a população. Algumas dessas matérias, expõe a fragilidade da eficácia do poder público no combate a criminalidade. Especificamente quanto a homicídios, segundo estudo do Tribunal de Justiça divulgado por ALVES, Shirlei (2016), 51% dos processos entre 2011 e 2015 no RS foram arquivados por falta de provas ou autoria. Outros 10% tiveram extinta a punibilidade, seja pela morte do réu ou pela a prescrição do crime. De forma que apenas 19,1% dos casos resultaram em condenação.

A situação econômica do Brasil é delicada, a restrição orçamentária que algumas unidades federativas enfrentam impede inclusive a reposição do efetivo policial de combate direto ao crime. Hoje, no Rio Grande do Sul, o número de servidores inativos na Polícia Civil e Brigada Militar supera em 7.427 matrículas o quadro de servidores ativos das duas instituições, o que representa 56,78% de inativos em relação ao total (SUL, 2019c). Esta situação impacta diretamente na atividade criminosa.

O Rio Grande do Sul têm mais de 11 milhões de habitantes e é o único estado fora das regiões Norte e Nordeste do país que apresenta crescimento da violência letal nos últimos 10 anos, quando a taxa de homicídios para cada 100 mil habitantes disparou de 19,8 para 29,3 (SINESP, 2018). Isso é um forte indicativo da relevância que este tema tem para o estado. É preciso aprimorar o combate à criminalidade com urgência.

Na figura 2.1, apresentada a seguir, percebe-se a evolução da taxa de homicídios no Rio Grande do Sul nos últimos anos. O crescimento da criminalidade é notório, evidenciando o tamanho do desafio que os gestores locais da segurança pública têm para proporcionar maior segurança aos cidadãos do estado.

Figura 2.1. Evolução da Taxa de Homicídios para cada 100 mil Hab. no RS



Fonte: Elaborado pelo autor com base nos dados do SINESP (2019).

As dificuldades na gestão da segurança pública brasileira são grandes, além da escassez de recursos, tanto em relação a bens de capital quanto a efetivo policial, e o gestor ainda enfrenta o dinamismo da criminalidade. As atividades criminosas migram de localidades, alteram sua estrutura e desenvolvem novas estratégias em função do desenvolvimento tecnológico, reinventando suas práticas.

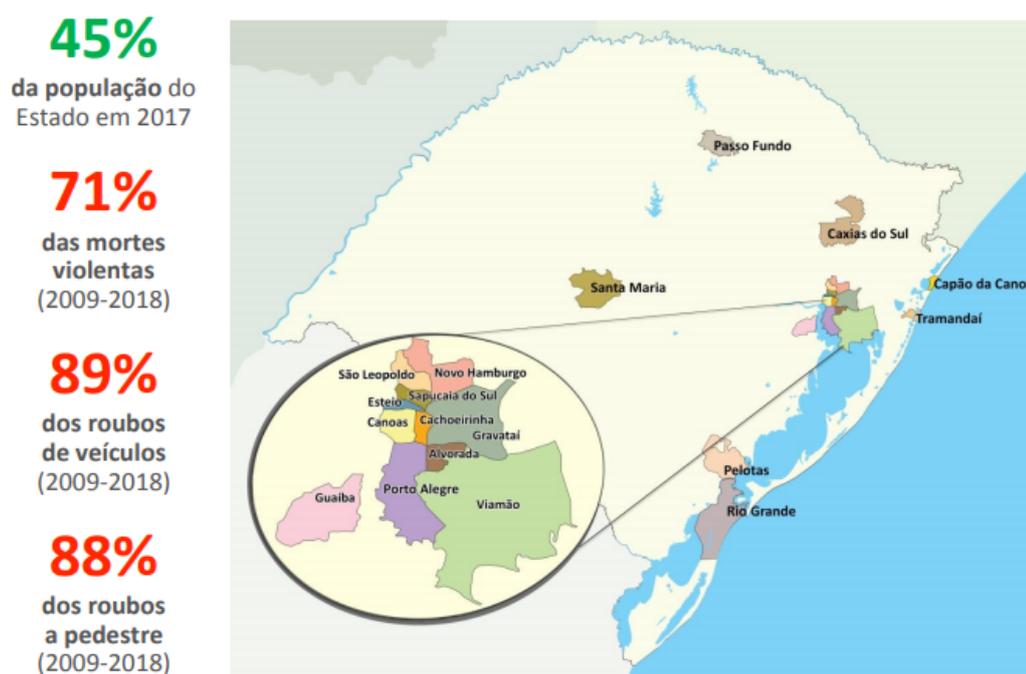
Como é de se esperar, a intensidade das atividades criminosas nas diferentes regiões do Rio Grande do Sul não é uniforme. Alguns municípios do estado tem grande representatividade na atividade criminosa. De acordo com dados divulgados pela Secretaria de Segurança do Estado do RS, dos 497 municípios do estado, 18 municípios apresentam 71% do total de mortes violentas (i. e., não naturais) no estado, 89% dos roubos a veículos e 88% dos roubos a pedestre, mesmo estes municípios possuindo apenas 45% da população do estado.

Pensando nesta situação, o Governo do Estado do Rio Grande do Sul lançou em

2019 o programa RS SEGURO, um projeto que deverá ser implementado nos próximos quatro anos com a finalidade de combater a criminalidade nestes 18 municípios mais violentos. O programa será melhor detalhado na seção metodologia, já que este estudo pretende seguir as diretrizes do programa governamental, a fim de servir como ferramenta para implementação de políticas públicas na área da segurança.

Os critérios para a escolha dos municípios participantes do programa foram a taxa de homicídios para cada 100 mil habitantes (superior a 30), o número médio de mortes violentas (acima de 50 por ano) e a população (acima de 65 mil habitantes)(SUL, 2019b). Os municípios participantes do programa estão representados geograficamente na figura 2.2.

Figura 2.2. Participantes do programa RS SEGURO e sua representatividade na criminalidade do estado.



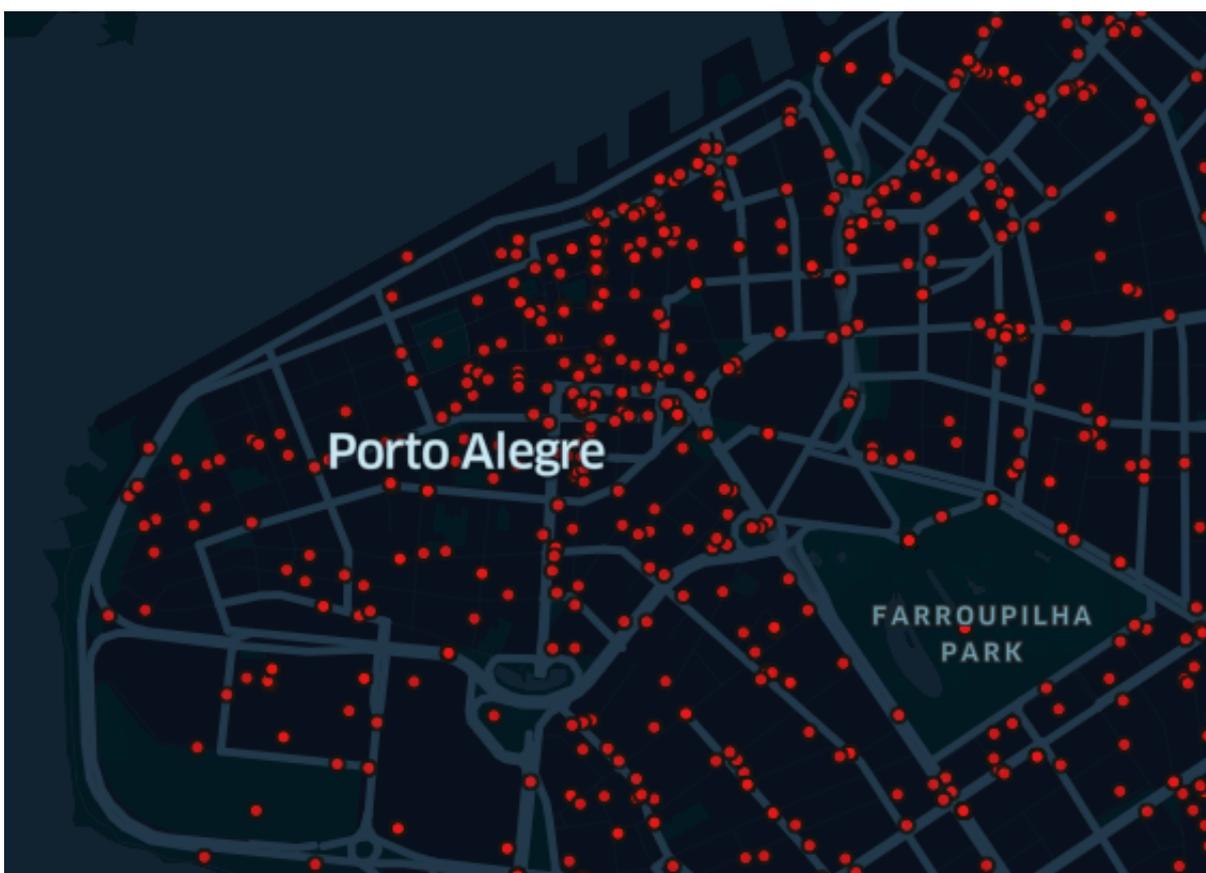
Fonte: Secretaria de Segurança Pública do RS. (SUL, 2019b).

O Rio Grande do Sul registrou 22.998 homicídios nos últimos 10 anos, sendo 5.432 destes na cidade de Porto Alegre, como apresentado pelo governo estadual no seu portal oficial(SUL, 2019a), representando 23,61% do total de homicídios do estado. Por ser a capital do estado, os dados disponíveis quanto aos eventos criminosos na

cidade estão melhores organizados na base de dados que será utilizada em comparação com os dados das demais cidades do estado. Por este motivo, e outros que serão apontados na metodologia, esta pesquisa desenvolveu uma análise aprofundada do comportamento do crime em cada uma das regiões da cidade.

A distribuição espacial da criminalidade em Porto Alegre não é homogênea. Como seria de se esperar, algumas regiões apresentam uma intensidade de atividades criminosas muito maiores que outras. Por este motivo, esta pesquisa analisa estas zonas de perigo individualmente e apresenta um modelo que prevê o comportamento criminoso separadamente para cada área selecionada, com a expectativa de fornecer melhores resultados para o gestor do que em uma análise da criminalidade da cidade como um todo. Abaixo, na figura 2.3, demonstra-se uma das possíveis áreas de perigo da cidade.

Figura 2.3. Crimes Contra a Vida nas imediações do Centro de Porto Alegre entre 2005 e 2019



Fonte: Elaborado pelo autor com base nos dados da SSP/RS (2019).

Com a alteração na organização e estrutura do crime, o combate eficaz da criminalidade deve contar com metodologias também inovadoras. As técnicas de análise de dados que são utilizadas atualmente pelos departamentos de gestão da segurança pública no Brasil traçam estratégias de enfrentamento da criminalidade limitadas, já que não possuem ferramentas de predição de eventos criminosos.

Neste sentido este trabalho pretende colaborar com o atual cenário, uma vez que contará com o avanço tecnológico e a evolução dos métodos de aprendizado de máquina nos últimos anos, elementos que podem obter resultados satisfatórios a ponto de serem utilizados no combate ao crime da cidade de Porto Alegre. Uma segunda fase do projeto será a implementação do método aqui verificado para todas as demais cidades que compõe o programa governamental RS SEGURO para posterior aplicação em outras áreas do estado.

Após esta breve apresentação da sensível situação e do desafio do combate à criminalidade no Brasil e no estado do Rio Grande do Sul. Serão introduzidos os estudos que orientam esta pesquisa, assim como também os trabalhos correlatos que têm destaque na atualidade.

Capítulo 3

TEORIA ECONÔMICA DO CRIME

No que se refere à criminalidade, a teoria econômica que tradicionalmente explica a atividade criminosa foi exposta em Becker (1968). Desde então muitos estudos têm se dedicado a compreensão deste tipo de evento a partir de um arcabouço econômico. Esta pesquisa é uma tentativa de mostrar a importância do estudo proposto em Becker (1968) ao procurar demonstrar que é possível compreender a ação criminosa a partir de um aparato lógico matemático.

Como demonstrado em Becker (1968) muitas pessoas se tornam criminosas não necessariamente por possuírem motivações diferentes das demais pessoas, mas sim por que seus benefícios e custos são diferentes. Infelizmente ainda não temos dados suficientes para testar todo arcabouço teórico de Becker (1968). Destaca-se a defesa que o autor faz do uso de ferramentas da ciência econômica para áreas que até então não estavam diretamente relacionadas, como a atividade criminosa.

A principal contribuição deste ensaio, como percebo, é demonstrar que a otimização de políticas de combate ao comportamento criminoso fazem parte de uma otimização da alocação de recursos. Como a ciência econômica desenvolveu-se na sentido de melhor alocar recursos, uma abordagem econômica torna-se então aplicável e ajuda a enriquecer a análise do comportamento ilegal.(BECKER, 1968, p. 209, tradução nossa)

Para Becker (1968) o desenvolvimento de uma atividade ilegal se assemelha ao de uma atividade lícita qualquer, através de benefícios e custos a serem analisados para a tomada de decisão. Neste caso o agente decide incorrer em um ato ilegal após

verificar as chances de sucesso e fracasso, comparando o custo da atividade com o possível prêmio pelo ato criminoso. Segundo Becker (1968), a função oferta de crimes pode ser representada como demonstrado abaixo:

$$O_j = O_j(p_j, f_j, u_j) \quad (3.1)$$

Como destacado por Becker (1968), apresenta O_j como sendo o número crimes que um determinado agente cometeria num dado período; enquanto p_j representa a probabilidade de ser condenado pelo ato criminoso; f_j a punição que pode ser aplicada em função do crime e u_j corresponde as demais variáveis que podem influenciar na conduta criminosa. Aponta também que um aumento na probabilidade de ser condenado e uma elevação na punição do agente aumentariam o preço a ser pago pela atividade criminosa, inibindo este tipo de conduta.

Becker (1968) demonstra também a ideia de maximização da utilidade dos recursos no combate à criminalidade. Os recursos necessários para implementar políticas públicas necessárias para anular a criminalidade seriam exorbitantes, tão significativos que a sociedade não estaria disposta a pagar. Já a ausência total de políticas com objetivo de conter essas atividades ilícitas da mesma forma cobriam um preço elevado da sociedade, ameaçando a própria sobrevivência humana. Assim, é preciso otimizar a utilização dos recursos e alocá-los de forma eficiente, reduzindo a criminalidade a níveis aceitáveis.

A partir de Becker (1968), outros trabalhos importantes surgiram, analisando também economicamente as atividades criminosas. Ehrlich (1973) expõe forte correlação positiva entre desigualdade de renda e crimes contra a propriedade, além de apresentar um modelo útil para avaliar a eficiência do gasto público no combate a criminalidade. Já Carr-Hill e Stern (1973) apresentam um modelo estatístico de avaliação do desempenho da força policial em diferentes regiões do Reino Unido, indicando também aspectos importantes da conduta dos criminosos e como sua relação com a sociedade influencia seu comportamento.

Além dos estudos que procuraram compreender a atividade criminosa em

si, outras pesquisas estudam a eficácia das políticas públicas de segurança e como melhorá-las. Hann (1972), apresenta o crime como um fenômeno econômico e procura demonstrar as dificuldades de compreender os custos diretamente envolvidos no crime e no combate à criminalidade, ressaltando que qualquer modelo que procure fazer isso precisaria derivar suas funções diretamente de um sistema capaz de compreender a atividade como um todo. Porém, mesmo analisando apenas a ação policial, Hann (1972) destaca que a tarefa é complexa, já que a força policial não é determinada por um sistema restrito de conduta policial, mas sim, um sistema muito mais amplo, que envolve não só a justiça, mas é influenciado também pelo legislativo.

Outro estudo com grande contribuição para a avaliação de políticas públicas voltadas para a área da segurança é apresentado em Downs William M. Landes e Hoffman (1969), onde destaca-se a análise da produtividade da força policial, principalmente apontando o cenário onde os custos operacionais aumentam ao longo das décadas sem resultar em um aumento de produtividade. A relevância do tema fez surgir muitas outras publicações que contribuem para a compreensão dos desafios envolvidos no combate à criminalidade. Cabe ressaltar trabalhos mais recentes e mais próximos da realidade desta pesquisa que foram publicados recentemente. Balbinotto Neto et al. (2018) demonstra o impacto do investimento público no combate a crimes contra a propriedade na região metropolitana de Porto Alegre. A análise econométrica avalia o desempenho das políticas, tanto no curto, como no longo prazo e os resultados apontam que o aumento no gasto com políticas de segurança pública tem impacto na redução da criminalidade, principalmente no longo prazo.

Como apontado, ao longo do tempo muitas pesquisas foram desenvolvidas a fim de compreender as ações que envolvem a atividade criminosa e o seu enfrentamento. Neste momento, são apresentados estudos pioneiros na predição da criminalidade ou que tenham relativa importância no tema.

A evolução tecnológica permitiu que surgissem novas formas de interpretar a evolução da criminalidade nos diversos lugares do mundo. Somando-se a isto, o importante espaço conquistado pela estatística computacional como ferramenta para a tomada de decisão, fez com que surgissem novas ferramentas de análise de dados para a indústria. Por sua vez, essa ascensão provocou uma demanda cada vez maior

por serviços capazes de solucionar os mais diversos problemas.

Até pouco tempo a predição da criminalidade não era considerada factível, como lembram Gorr e Harries (2003). A polícia sempre esteve acostumada a traçar estratégias para solucionar crimes específicos, analisando o comportamento de indivíduos. Ao pensar dessa forma, a predição da criminalidade realmente se torna pouco eficaz, afinal prever a conduta de uma pessoa específica é uma tarefa um tanto difícil. Porém, ao analisar a criminalidade através de *clusters* distribuídos geograficamente a chance de sucesso na previsão aumenta consideravelmente. Foi exatamente isso que propuseram Corcoran e Ware (2003), que obtiveram bons resultados ao prever a criminalidade treinando redes neurais artificiais utilizando dados clusterizados.

Pioneiro em agrupamento de pontos para análise da criminalidade, Sherman et al. (1989) influenciou diversas pesquisas. O estudo avaliou a distribuição espacial de 323.979 atendimentos policiais em mais de 115 mil endereços diferentes na cidade de *Minneapolis*, demonstrando que 3% das localidades (*hotspots*) eram responsáveis por 50% das ocorrências. Já Felson e Poulsen (2003), compararam as características temporais de roubos em diferentes cidades americanas e concluíram que se pode obter melhores resultados ao prever os delitos quando os dados são analisados levando em consideração o horário em que estes frequentemente ocorrem, ao contrário do método tradicional de se analisar o desempenho mensal, trimestral ou anual.

Mais recentemente, Stalidis et al. (2018) aplicou um modelo de redes neurais convolucionais, normalmente utilizadas em problemas de visão computacional (como a identificação de pessoas ou objetos em imagens), para a predição da localização de eventos criminosos através da análise de mapas de calor que representavam o histórico dos eventos em determinada localidade.

Muitas dessas metodologias que surgiram podem ser implementadas com a finalidade de otimizar os recursos públicos ou pelo menos para entender melhor os eventos que interferem no dia a dia da sociedade. Com o avanço do desenvolvimento tecnológico, muitos estudos direcionados a análise da criminalidade de diversas localidades podem ser encontrados. Destacam-se os trabalhos de Short et al. (), McClendon e Meghanathan (2015), Blackmore (2003) e Trumbull e N. (), como estudos que procuram compreender o desenvolvimento da criminalidade e os impactos desta na sociedade.

Outros trabalhos mais interessados na identificação de focos de criminalidade e previsão foram recentemente publicados, com destaque para Berk e Bleich (2013), Liao et al. (2010), assim como Dahbur e Muscarello (2003) e Kianmehr e Alhadj (2008). Dessa forma, demonstra-se que o século XXI reúne não só as condições para enfrentar a criminalidade a partir de uma abordagem tecnológica, mas possui também grande interesse da academia neste tópico.

Após demonstrada a evolução dos estudos que norteiam esta pesquisa, a seguir, é apresentada a metodologia escolhida para a realização da análise exploratória dos dados, assim como as características dos modelos de previsão utilizados, como também as métricas escolhidas para mensurar o desempenho destes. Depois de sucinta descrição da metodologia, os resultados são apresentados.

Capítulo 4

METODOLOGIA

Para Friedman (1966), apenas a capacidade preditiva de uma teoria pode determinar a sua validade e só faz sentido testar suas previsões com a própria experiência que ela pretende explicar. Friedman (1966) ainda aponta que uma teoria econômica não pode ser caracterizada por uma estrutura tautológica que apenas descreve consequências de determinadas ações, ela precisa prever.

Uma teoria deve ser julgada de acordo com o seu poder preditivo em relação ao fenômeno que ela pretende explicar. Apenas evidências factuais podem demonstrar se está correta ou não, ou melhor, se ela deve ter sua validade aceita ou rejeitada.(FRIEDMAN, 1966, p. 57, tradução nossa.)

Como aponta Friedman (1966), neste caso não se deve entender previsões apenas quanto a fenômenos futuros, mas também quanto a eventos pretéritos. A ideia central é conseguir explicar um evento inédito, independente do espaço de tempo em que ele ocorre, seja no passado ou no futuro.

Friedman (1966) ressalta que evidências empíricas jamais podem provar uma teoria, apenas podem não rejeitá-la. Isto é, uma hipótese é rejeitada caso suas previsões não sejam confirmadas, e é aceita caso não seja negada. Como apontado pelo autor "evidências empíricas nunca podem provar uma hipótese, apenas pode falhar em rejeitá-la"(FRIEDMAN, 1966, p. 57, tradução nossa).

Ao se escolher entre várias hipóteses que explicam determinado fenômeno, Fri-

edman (1966) esclarece que deve-se optar pela hipótese baseada na teoria mais simples e assertiva possível. Neste caso, uma teoria é mais simples quanto menor for o nível de conhecimento pré-estabelecido necessário para prever um determinado evento, e mais assertiva quanto melhor e mais abrangente for o seu desempenho.

A fim de atingir os resultados esperados, o presente estudo utiliza a base de dados disponibilizada pela Secretaria de Segurança Pública do Estado do Rio Grande do Sul. Esta base consiste de lançamentos efetuados pelos órgãos de segurança pública do estado. Como a base é extensa em função do grande número de eventos e principalmente da ampla variedade de tipicidade penal, optou-se por realizar um recorte.

Este recorte foi realizado também em função do interesse do próprio Governo do Estado do Rio Grande do Sul, que lançou em 2019, como destacado anteriormente, o programa "RS SEGURO - Programa Transversal e Estruturante de Segurança Pública"(SUL, 2019b). Este programa está estruturado em quatro eixos: "Combate ao crime", "Políticas sociais preventivas e transversais", "Qualificação do atendimento ao cidadão" e "Sistema Prisional". Ele será implementado nos 18 municípios que concentram a maior parte da criminalidade do estado.

De acordo com as diretrizes do programa supracitado o eixo combate ao crime formará estratégias de ação policial através da análise de quatro indicadores determinados, sendo três deles globais e um indicador local. Os três indicadores globais são: crimes violentos letais intencionais (homicídio doloso, lesão corporal seguida de morte e roubo seguido de morte), roubo de veículos e roubo a pedestre. Estes indicadores serão utilizados em todos os 18 municípios selecionados pelo programa, além de um indicador a mais que cada município pode escolher segundo interesses locais. Os outros indicadores que poderiam ser selecionados pelos municípios eram: roubo a estabelecimento comercial e de ensino, roubo a residência, roubo a passageiro ou motorista de transporte coletivo ou lotação e furto de veículos. Porto Alegre, que é o nosso objeto de pesquisa, optou pelo indicador de roubos a estabelecimentos comerciais e de ensino.

Dessa forma, nossa análise está concentrada nos crimes violentos letais intencionais, roubo de veículos, roubo a pedestre e roubo a estabelecimentos comerciais e de ensino, na cidade de Porto Alegre, utilizando todos os registros a partir de 2005 até

outubro de 2019, últimos dados disponibilizados pela secretaria. Os dados fornecidos apresentam a seguinte estrutura para cada evento: data, hora, logradouro, número, bairro, cidade, ponto de referência, fato e tipo de fato.

Estes dados são obtidos diretamente da base de dados da Polícia Civil, estando registrados exatamente da mesma forma que foram informados pelo escrivão policial no momento do registro da ocorrência policial. Foi necessário realizar uma análise dos dados fornecidos a fim de descartar informações que não sejam suficientes para fornecer a geolocalização do fato descrito (detalhes são informados na seção resultados). Além dos campos relativos a região geográfica que são fornecidos, o campo "fato" descreve o crime ocorrido no evento, e.g., homicídio doloso e culposo, feminicídio, latrocínio, lesão corporal seguida de morte, entre outros. Já o campo "tipo fato" fornece a tipificação do evento, seja consumado ou tentado.

Através desses dados fornecidos pela SSP/RS obteve-se as coordenadas geográficas junto a serviços de geolocalização, de forma que foi possível construir os modelos de previsão a partir dessas informações. Estes dados foram analisados de acordo com a geolocalização de cada evento na cidade de Porto Alegre e a sua relação com os eventos subsequentes. Invocando aqui a primeira lei da geografia de Tobler que reza que "todas as coisas estão relacionadas com todas as outras, mas coisas próximas estão mais relacionadas do que coisas distantes"(TOBLER, 2011, p. 234, tradução nossa). Os detalhes da análise estão demonstrados na seção resultados.

Para analisar estes dados foram utilizados diversos métodos de inferência preditiva, tanto modelos de aprendizagem de máquina conhecidos como *Shallow Learning*, quanto modelos robustos difundidos mais recentemente que ficaram conhecidos como *Deep Learning*. Ambos serão descritos a seguir.

4.1 APRENDIZAGEM DE MÁQUINA

A aprendizagem de máquina consiste na utilização de métodos computacionais capazes de aprimorar sua capacidade de previsão através da análise de uma base de dados disponível, conforme esclarece Mori (1967). Neste caso, o computador não é ex-

plicitamente programado para realizar cada operação necessária, mas sim programado para aprender com suas previsões de forma que possa aprimorar o seu resultado de forma autônoma até obter uma estimativa ótima em relação às especificações informadas. De forma mais geral, lembram os autores supracitados, "técnicas de aprendizagem são métodos orientados por dados que combinam conceitos fundamentais de ciência da computação, estatística, probabilidade e otimização"(MORI, 1967, p.414 ,tradução nossa).

Conforme destaca Murphy (2012), pode-se segmentar a aprendizagem de máquina em dois propósitos: preditivo e descritivo. Caso o interesse na construção do algoritmo seja a previsão, os modelos mais indicados são os de aprendizagem supervisionada, como é o caso desta pesquisa. Já se o interesse for descrever objetos ou fenômenos, a abordagem mais indicada é o modelo de aprendizagem não supervisionada, este modelo não será utilizado nesta pesquisa.

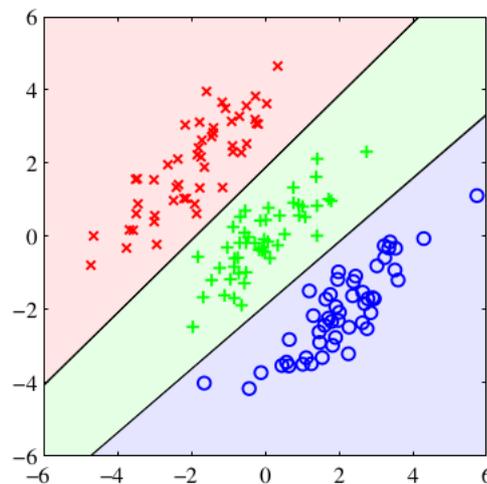
As abordagens identificadas como de aprendizagem supervisionada recebem informações sobre a variável a ser prevista e estabelecem pesos para identificar quais fatores dentro dos disponíveis interferem na variável e em que medida impactam. Neste caso, a base de dados é dividida em duas partes: uma é utilizada para treinar o algoritmo e a outra para testar os resultados obtidos no treinamento. Então, pode-se utilizar métricas de verificação dos resultados para adaptar o modelo e melhorar a sua performance.

Métodos computacionais como o *Machine Learning* tem sido cada vez mais utilizados tanto na academia como na indústria, seja na simples classificação de e-mails como *spam* ou *não-spam*, na realização de diagnósticos médicos, detecção de fraudes e até mesmo nos mecanismos de buscas virtuais que todos utilizam diariamente. Alguns dos principais problemas que a aprendizagem de máquina é utilizada e que estão relacionados de alguma forma com os métodos utilizados nesta pesquisa são: classificação, regressão, clusterização, redução de dimensionalidade, assim como modelos relacionados a aprendizagem profunda. Estes tópicos verificaremos a seguir com maiores detalhes.

4.1.1 Classificação

Classificação é um método utilizado para atribuir uma categoria para cada item analisado (MORI, 1967). Pode ser utilizada com muitas finalidades, como para identificar através da imagem de um animal a qual espécie ele pertence, ou também para determinar se um paciente possui ou não determinada patologia com base nos exames utilizados no modelo. Enfim, a possibilidade de utilização desse método é infindável. Como demonstrado por Mori (1967), o objetivo da classificação é aprender a mapear as saídas y de uma entrada x quando $y \in \{1, \dots, C\}$, onde C é o número de classes. Teremos uma classificação binária caso $C = 2$, porém se $C > 2$ então a classificação será definida como multiclasse. Há ainda os casos onde as categorias não excluem umas as outras (como, por exemplo, no caso de um objeto que pode ser classificado como pequeno e pesado ao mesmo tempo), estes casos são conhecidos como classificação multirótulo.

Figura 4.1. Exemplo de classificação multiclasse



Fonte: Bishop (2006).

Conforme aponta Murphy (2012), a classificação se dá através de predição probabilística. Assim, será necessário verificar a probabilidade daquele item pertencer a cada uma das categorias analisando a entrada como um determinado vetor x e a base

de dados D . Então o melhor palpite sobre a classificação será encontrado utilizando

$$\hat{y} = f'(x) = \underset{c=1}{\operatorname{argmax}} p(y = c|x, D) \quad (4.1)$$

Dessa forma conseguiremos otimizar a predição, identificando a categoria do item através da moda da distribuição $p(y|x, D)$.

4.1.2 Regressão

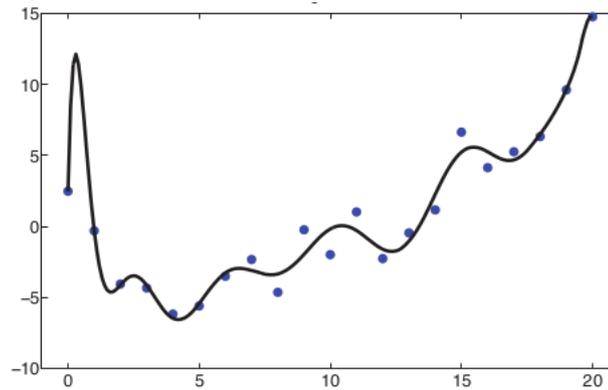
Diferentemente da classificação a regressão realiza a previsão de qualquer número real. Sendo x a entrada e y a saída a ser mensurada do subconjunto \mathbb{R} . A regressão linear é o modelo de regressão mais utilizado e pode ser entendida da seguinte forma:

$$y(\mathbf{x}) = \mathbf{w}^T \mathbf{x} + \epsilon = \sum_{j=1}^D w_j x_j + \epsilon \quad (4.2)$$

Conforme esclarece Murphy (2012), $\mathbf{w}^T \mathbf{x}$ representa o produto escalar entre o vetor \mathbf{x} e o vetor de pesos do modelo \mathbf{w} , onde ϵ caracteriza o erro residual entre as previsões obtidas e a resposta verdadeira. Este erro residual pode ser reduzido através da minimização de uma função custo, que reduzirá o erro deixando o modelo mais eficaz. A função custo para uma regressão linear pode ser entendida da seguinte forma:

$$E(\mathbf{w}) = \frac{1}{2} \sum_{n=1}^N \{y(x_n, \mathbf{w}) - t_n\}^2 \quad (4.3)$$

Onde podemos minimizar cada imprecisão da função $y(x, \mathbf{w})$ para cada valor de \mathbf{w} em relação a base de dados. Como aponta Bishop (2006), a soma dos quadrados dos erros entre a predição $y(x_n, \mathbf{w})$ para cada ponto x_n correspondente aos valores alvo t_n .

Figura 4.2. Exemplo de regressão linear

Fonte: Murphy (2012).

4.1.3 Clusterização *K*-means

A clusterização é utilizada geralmente para analisar base de dados muito grandes. Como apresenta Murphy (2012, p. 1067, tradução nossa) "a clusterização é o processo de agrupamento de objetos similares". Espera-se que através da utilização deste método seja possível identificar as principais áreas de ocorrência dos crimes estudados na cidade de Porto Alegre, permitindo o melhor desempenho na previsão da criminalidade nestas áreas.

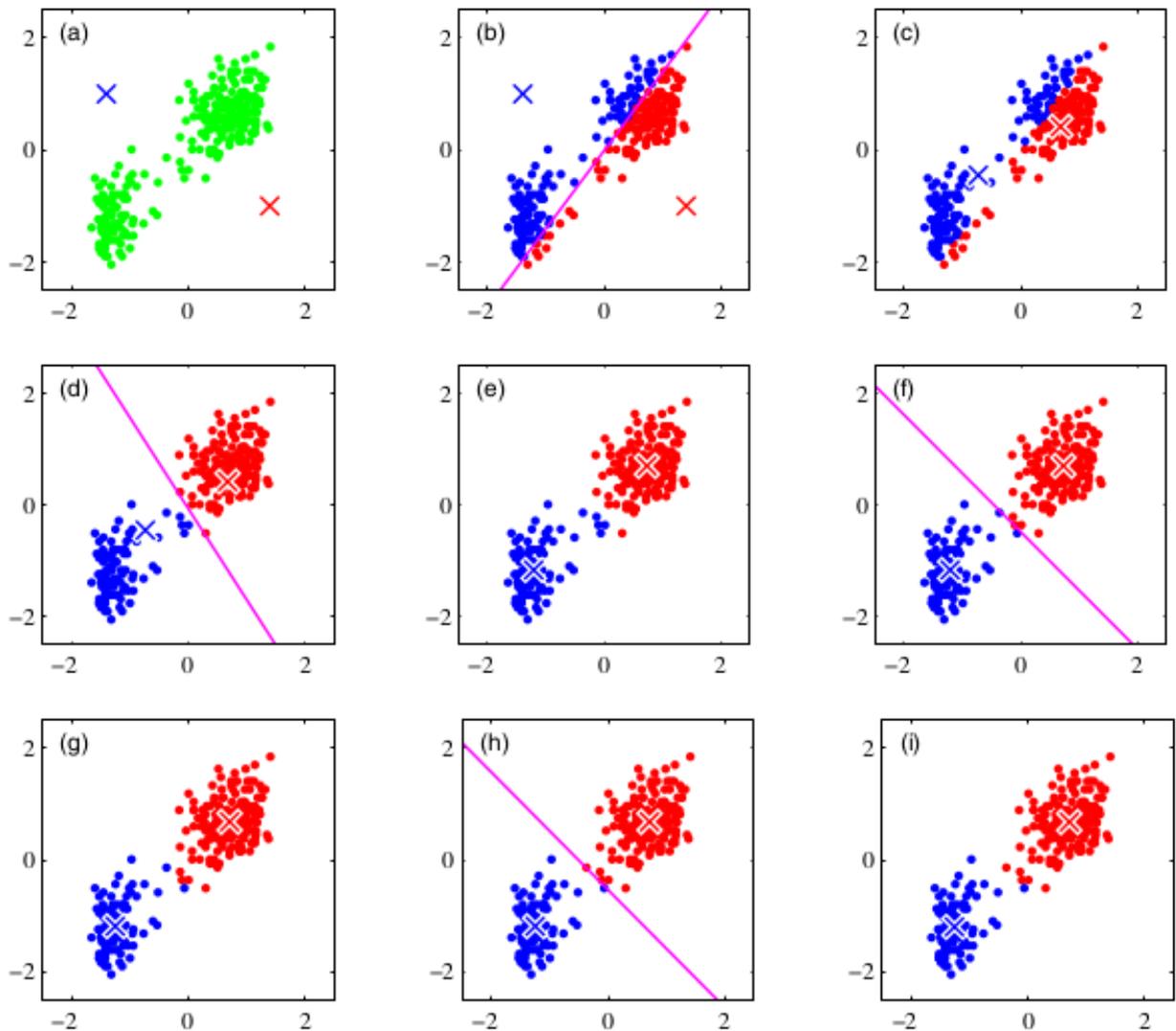
Pode-se pensar um cluster também como um agrupamento de pontos num espaço D -dimensional onde a distância entre os pontos dentro do cluster é menor que a distância entre estes pontos com os demais fora do cluster, como denota Bishop (2006). O autor ainda demonstra que analisando os vetores μ_k de um determinado espaço D -dimensional onde $k = 1, \dots, K$ e μ seria o padrão associado ao K -ésimo cluster, podemos entender μ_k como o centro de cada cluster. De forma que pode ser expresso na seguinte forma:

$$J = \sum_{n=1}^N \sum_{k=1}^K r_{nk} \|x_n - \mu_k\|^2 \quad (4.4)$$

Neste caso o objetivo é minimizar J , isso ocorre através de um processo dinâ-

mico de otimização em série tanto de r_{nk} e de μ_k . Como explica Bishop (2006), após escolhidos os valores iniciais de μ_k é preciso minimizar J em relação a r_{nk} mantendo μ_k fixo, logo após isso minimizamos J em relação a μ_k mantendo agora r_{nk} fixo. Este processo é repetido até que os clusters estejam adequadamente definidos. Este processo pode ser visualmente compreendido como apresentado na figura 4.3:

Figura 4.3. Exemplo de clusterização

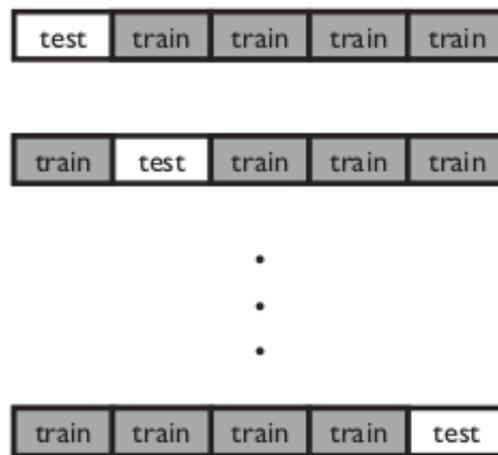


Fonte: Bishop (2006).

4.1.4 Validação Cruzada

O tamanho da base de dados pode ser um problema para se obter uma boa validação dos resultados. Uma base de dados pequena pode trazer incertezas quanto aos erros estimados, lembra Goodfellow et al. (2016). Essa situação pode ser contornada utilizando um método conhecido como validação cruzada¹. Neste caso a base de dados é fragmentada em n subconjuntos, então o algoritmo é treinado em todas estas partições exceto em uma que servirá como teste para avaliar o seu desempenho, como demonstrado na figura 4.4:

Figura 4.4. Demonstração dos subconjuntos da validação cruzada



Fonte: Mori (1967).

Segundo Mori (1967) inicialmente uma amostra S de m variáveis classificadas é subdividida em n partes. De forma que a i -ésima parte seja uma amostra classificada $((x_{i1}, y_{i1}), \dots, (x_{im}, y_{im}))$ de tamanho m_i . Então, o algoritmo será treinado para todo $i \in [1, n]$ exceto na i -ésima partição, gerando uma hipótese h_i . Esta hipótese será testada na partição selecionada, então o valor do parâmetro θ é avaliado em relação ao erro médio da hipótese h_i . Este processo pode ser expresso da seguinte forma:

$$\hat{R}_{CV}(\theta) = \frac{1}{n} \sum_{i=1}^n \frac{1}{m_i} \sum_{j=1}^{m_i} L(h_i(x_{ij}), y_{ij}) \quad (4.5)$$

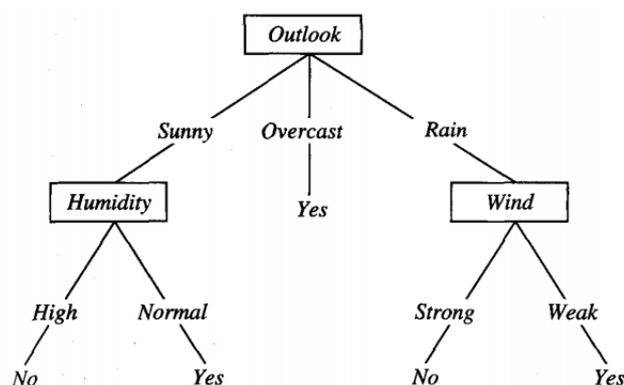
¹cross-validation

4.1.5 Árvores de Decisão

Um método muito utilizado em função da sua fácil visualização é o modelo de árvore de decisão, seja para atividades de classificação ou regressão. Este método consiste na subdivisão de um problema geral em vários problemas menores. Cada árvore possui uma raiz onde estão as informações disponíveis para a solução do problema. A partir da raiz surgem os nós da árvore que representam uma decisão a ser tomada pelo modelo. Cada nó poderá ter duas ou mais saídas, o último nó é identificado como folha, onde encontram-se os atributos que serão utilizados como parâmetro para a tomada de decisão ou predição do modelo. (BISHOP, 2006).

Este método é muito utilizado em inferência preditiva, com ótimos resultados tanto na área médica com a previsão de diagnósticos, quanto no mercado financeiro analisando o risco envolvido nas operações de crédito (MITCHELL, 1997). Conforme esclarece o autor citado anteriormente, uma árvore de decisão pode ser entendida como uma disjunção de conjunções de restrições dos atributos das instâncias, de forma que "cada caminho da raiz da árvore até as folhas desta correspondem a uma conjunção de testes de atributos e a árvore em sí é a disjunção destas conjunções"(MITCHELL, 1997, p. 53, tradução nossa).

Figura 4.5. Exemplo de árvore de decisão sobre jogar tênis ou não

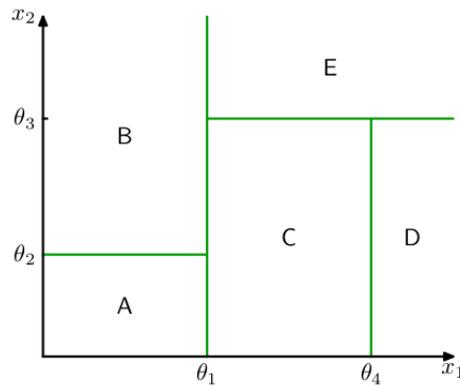


Fonte: Mitchell (1997).

A figura 4.6, apresentada abaixo, demonstra um plano de duas dimensões fracionado em cinco regiões. Em cada região há um modelo independente para prever a variável dependente, conforme descrito por Bishop (2006) no caso de uma problema

de regressão nós iremos prever uma constante para cada região, já em um problema de classificação cada região estaria atrelada a uma classe.

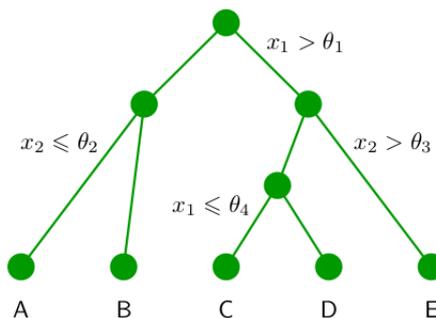
Figura 4.6. Exemplo de partição das regiões em um modelo de duas dimensões



Fonte: Bishop (2006).

Assim, aponta Bishop (2006), primeiramente x_1 é avaliado em relação ao parâmetro θ , então a entrada é dividida em duas regiões quando $x_1 \leq \theta_1$ ou $x_1 > \theta_1$. Agora estas duas regiões podem ser analisadas de forma independente, onde por fim as folhas da árvore representaram as regiões finais do modelo, conforme pode ser percebido na figura 4.7:

Figura 4.7. Exemplo árvore binária



Fonte: Bishop (2006).

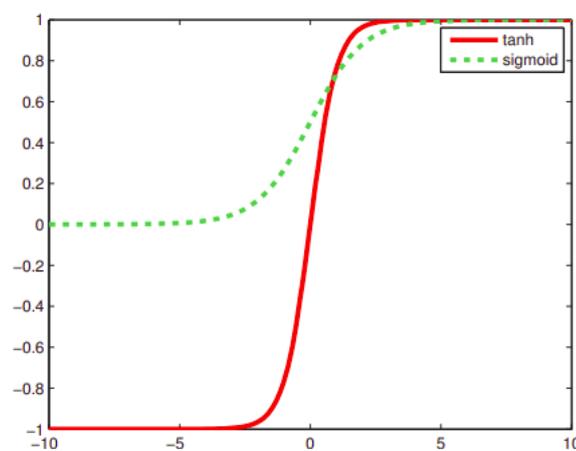
Encontrar a estrutura da árvore de decisão que minimiza a soma do quadrado dos erros pode não ser fácil em função do grande número de possibilidades de estruturas diferentes. Bishop (2006) aponta que após iniciar a árvore pela raiz devemos ir adicionando um nó de cada vez, em cada etapa haverá um número de regiões que o

nó poderá ser fragmentado. A otimização do número de regiões dependerá da decisão de qual variável separar assim como qual será o limite estabelecido para cada nova região. O autor supracitado lembra que este resultado pode ser obtido por um processo exaustivo onde a escolha ótima das variáveis a separar é dada pela média local dos dados. Este processo será repetido para todas as possibilidades e aquela que apresentar o menor resíduo da soma do quadrado dos erros é escolhida. A árvore então pode ser podada quando reduzir o erro encontrado para níveis inferiores aos limites estabelecidos.

4.1.6 Redes Neurais Artificiais

Seguindo adiante nas metodologias utilizadas nesta pesquisa abordaremos brevemente a construção e o funcionamento de uma rede neural artificial (RNA). Segundo Specht (1991) uma RNA é "normalmente definida como uma rede composta de um grande número de neurônios que estão massivamente interconectados operando paralelamente e que aprendem através da experiência"(SPECHT, 1991, p. 568, tradução nossa). Uma rede neural artificial com realimentação² pode ser entendida como uma encadeação de regressões logísticas empilhadas onde a saída pode ser tanto um modelo de regressão linear quanto novamente uma regressão logística, dependendo do resultado que se espera obter (MURPHY, 2012).

Figura 4.8. Funções de ativação Sigmóide e Tangente Hiperbólica



Fonte:Murphy (2012).

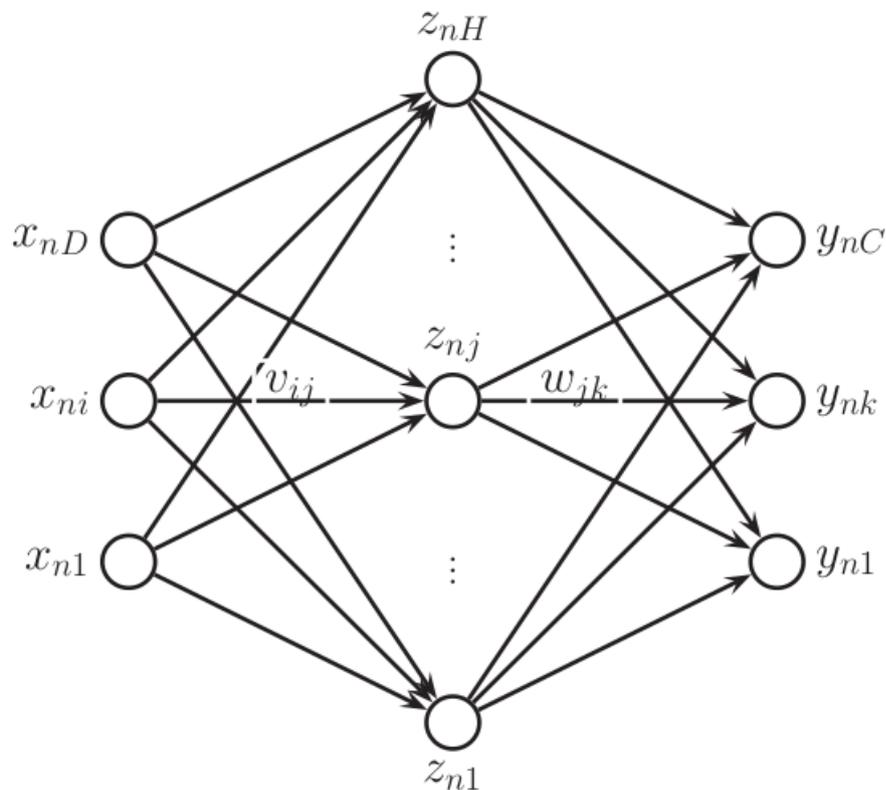
²feedforward neural network

As regressões mais utilizadas como função de ativação em cada neurônio são as funções sigmóide e tangente hiperbólica (\tanh), demonstradas na figura acima. A função sigmóide retorna um valor entre 0 e 1 para qualquer número real, já a função \tanh retorna um valor entre -1 e 1. Dependendo do que está se está tentando aprender será conveniente a utilização de cada uma das funções. Outras funções podem ser utilizadas para treinamento dos neurônios, porém estas são as mais utilizadas nesta pesquisa. A seguir, demonstra-se as funções sigmóide (4.6) e \tanh (4.7).

$$f(x) = \frac{1}{1 + e^{-x}} \quad (4.6)$$

$$\tanh(t) = \frac{e^t - e^{-t}}{e^t + e^{-t}} \quad (4.7)$$

Figura 4.9. Demonstração de uma Rede Neural Artificial



Fonte:Murphy (2012).

Na figura anterior (4.9) percebe-se uma rede neural com apenas uma camada, onde x e y são as entradas e saídas respectivamente e Z_nH , Z_{nj} e Z_{nl} são os neurônios que compõe a camada oculta. As redes neurais podem ter muitas camadas, esses modelos com mais de uma camada são conhecidos como redes neurais profundas (DNN), são especificamente estes modelos de redes neurais que foram usados neste estudo. Como esclarece Murphy (2012) caso se queira resolver um problema de regressão com uma rede neural não regressiva de duas camadas teria:

$$p(y|\mathbf{x}, \boldsymbol{\theta}) = N(y|\mathbf{w}^T \mathbf{z}(\mathbf{x}), \sigma^2) \quad (4.8)$$

$$\mathbf{z}(\mathbf{x}) = g(\mathbf{V}\mathbf{x}) = [g(\mathbf{v}_1^T \mathbf{x}), \dots, g(\mathbf{v}_H^T \mathbf{x})] \quad (4.9)$$

Neste caso a função de ativação g não deve ser linear para que o sistema todo se transforme em uma grande regressão linear, afirma Murphy (2012). O autor indica que na camada oculta ($\mathbf{z}(\mathbf{x}) = \phi(\mathbf{x}, \mathbf{V})$) H é o número de neurônios nas camadas ocultas e \mathbf{V} os pesos das variáveis de entrada (\mathbf{x}) para a camada oculta, enquanto w representa os pesos da camada oculta para a saída (y).

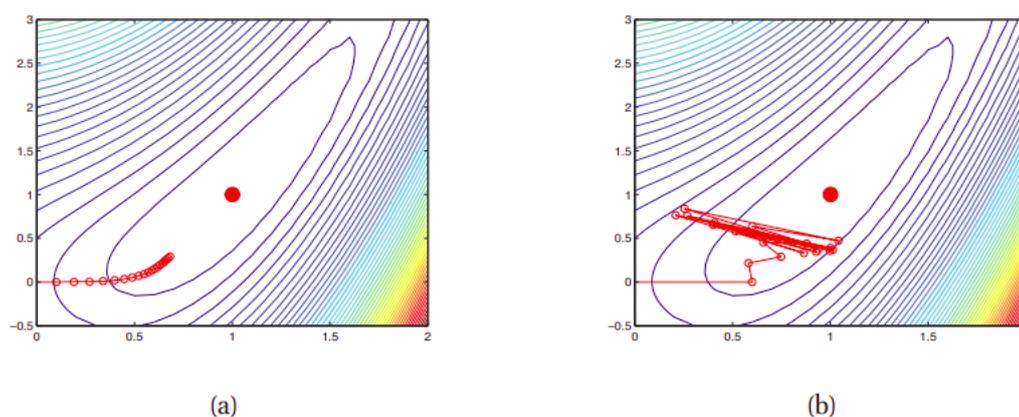
A forma que esta rede neural consegue aprender está relacionada a minimização do erro entre os valores obtidos para y no dataset de treinamento e os valores esperados. Este processo é conhecido como descida do gradiente, onde o algoritmo utiliza inicialmente um valor aleatório para os pesos usados para a previsão. Após isso, esses pesos são redefinidos a partir da minimização da função custo determinada, conforme cada neurônio em cada camada vai sendo utilizado. Esses pesos precisam ser ajustados também no sentido oposto ao do fluxo de informações da rede neural, processo conhecido como retropropagação³, pois após a chegada da saída a rede informa aos neurônios o desempenho da rede para que estes possam então ajustar os pesos, fazendo com que o algoritmo aprenda a resolver o problema solicitado.

A velocidade que ocorre essa adequação dos pesos é chamada taxa de aprendizagem e esta interfere diretamente no desempenho do algoritmo. Uma taxa de aprendi-

³*Backpropagation*

dizagem muito pequena pode não propiciar uma aprendizagem adequada assim como uma taxa muito grande não permite o aprendizado do algoritmo, como demonstrado na imagem abaixo. No cenário (a) a taxa de aprendizagem é muito pequena, no cenário (b) a taxa é muito alta, o objetivo é atingir o ponto a ser previsto, neste caso o ponto (1,1).

Figura 4.10. Demonstração da taxa de aprendizagem na predição.



Fonte: Murphy (2012).

4.1.7 Long Short-Term Memory

Outro tipo de rede neural muito utilizada para análise de séries de tempo é um tipo de rede neural recorrente conhecida como memória de longo e curto prazo⁴(LSTM). Essas redes neurais possuem células capazes de esquecer uma informação irrelevante de forma que ela não seja passada adiante para a próxima célula, mantendo no modelo apenas as informações mais importantes.

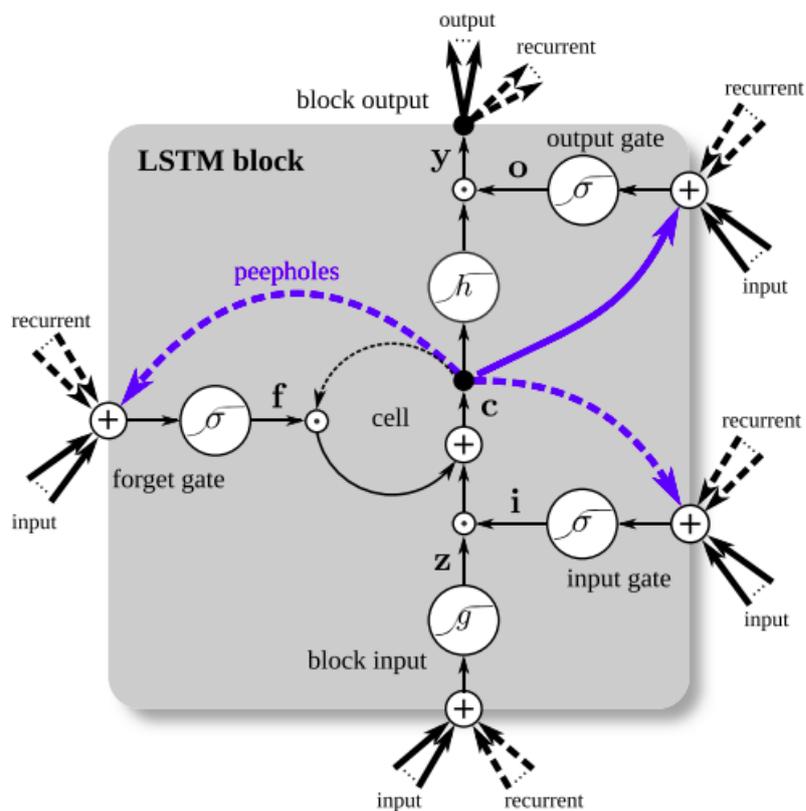
Este tipo de célula foi demonstrada pela primeira vez no trabalho de Hochreiter e Schmidhuber (1997) e desde então esse tipo de rede neural tem sido utilizada com grande intensidade. As suas principais aplicações estão relacionadas ao reconhecimento de discursos (GRAVES; JAITLEY, 2014) e aprendizagem da escrita manual (GRAVES, 2013). Como estas redes utilizam métodos de aprendizados baseados em gra-

⁴Long Short-Term Memory

dientes e estes podem assumir valores muito elevados ou muito pequenos de forma que não seja possível haver aprendizagem alguma, estas redes conseguem manter os valores dos gradientes em intensidades que possibilitem a aprendizagem, utilizando derivadas como demonstrado por Hochreiter e Schmidhuber (1997).

As células LSTM utilizam o conceito de portões, estes controlam o fluxo de informações pela cadeia de células. Ao utilizarem funções sigmóide os portões retornam valores entre 0 e 1 indicando a relevância da informação da célula anterior para os dados introduzidos na célula atual. A multiplicação da informação pelo resultado da função sigmóide retornando 0 faz com que aquela informação seja descartada, caso a função retorne 1 a multiplicação da informação faz com que ela continue sendo transmitida para as células seguintes, possibilitando o aprendizado. As células são colocadas umas após as outras substituindo o formato usual de camadas ocultas em uma rede neural tradicional.

Figura 4.11. Demonstração genérica de uma célula LSTM



Fonte: Greff (2015).

Como pode-se observar na imagem anterior, a célula LSTM possui portões de entrada, saída e de esquecimento. Como esclarece Goodfellow et al. (2016), cabe ao portão de esquecimento calcular a representatividade da informação passada. Assim, sendo $\mathbf{x}^{(t)}$ o vetor atual do portão de entrada e $\mathbf{h}^{(t)}$ o vetor da camada oculta no mesmo instante t , \mathbf{b}^f , \mathbf{U}^f e \mathbf{W}^f então seriam respectivamente o viés, os pesos no momento da entrada e os pesos da série recorrente, como segue:

$$f_i^{(t)} = \sigma \left(b_i^f + \sum_j U_{i,j}^f x_j^{(t)} + \sum_j W_{i,j}^f h_j^{(t-1)} \right) \quad (4.10)$$

A condição da célula então é classificada de acordo com o portão de esquecimento e atualizada a cada novo loop de informações, o que é esperado em qualquer rede neural recorrente, como demonstrado a seguir:

$$s_i^{(t)} = f_i^{(t)} s_i^{(t-1)} + g_i^{(t)} \sigma \left(b_i + \sum_j U_{i,j} x_j^{(t)} + \sum_j W_{i,j} h_j^{(t-1)} \right) \quad (4.11)$$

Goodfellow et al. (2016) lembra que o portão de entrada $g_i^{(t)}$ funciona da mesma forma que o portão de esquecimento, porém utilizando seus próprios parâmetros. Da mesma forma, o portão de saída $q_i^{(t)}$ também utiliza uma função sigmóide para verificar o impacto do estado dessa célula no desempenho do modelo. As expressões dos portões de entrada e saída são apresentadas respectivamente a seguir:

$$g_i^{(t)} = \sigma \left(b_i^g + \sum_j U_{i,j}^g x_j^{(t)} + \sum_j W_{i,j}^g h_j^{(t-1)} \right) \quad (4.12)$$

$$q_i^{(t)} = \sigma \left(b_i^o + \sum_j U_{i,j}^o x_j^{(t)} + \sum_j W_{i,j}^o h_j^{(t-1)} \right) \quad (4.13)$$

4.2 MÉTRICAS DE AVALIAÇÃO DOS RESULTADOS

Para avaliar os resultados de cada modelo foram utilizadas quatro medidas. Estas são as medidas de erros mais utilizadas para avaliar métodos de aprendizagem de máquina. Foram escolhidos como método de avaliação o erro absoluto médio, erro quadrático médio, raiz do erro quadrático médio e coeficiente de determinação, conhecido como R^2 .

- **Erro Médio Absoluto**

$$MAE = \frac{1}{n} \sum_{i=1}^n |y_i - \hat{y}_i| \quad (4.14)$$

O erro médio absoluto apresenta a diferença em unidades entre o estimador Y e o parâmetro \hat{Y} . Dessa forma podemos verificar em quantas unidades o modelo analisado erra em relação à média. Ele mede a magnitude do erro da previsão em relação ao dado original sem levar em consideração o seu viés, já que o dado analisado é na forma de módulo. Esta métrica demonstra cada erro com a mesma intensidade, não penalizando observações mais distantes em relação a média com maior peso, diferente das que serão apresentadas a seguir.

- **Erro Médio Quadrático**

$$MSE = \frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2 \quad (4.15)$$

O erro quadrático médio mede a diferença quadrática entre o estimador Y e o parâmetro \hat{Y} . Diferentemente do erro médio absoluto, neste caso as observações que se dispersam de forma mais evidente da média são penalizadas em relação aquelas que estão mais próximas. Dessa forma o MSE mede mais apropriadamente a variabilidade

do estimador. Essa penalização dos erros mais discrepantes pode ser muito interessante em vários casos, principalmente quando queremos analisar o potencial de predição de um modelo onde um erro de grande dimensão pode implicar em problemas relevantes para a aplicação do modelo. Assim, esta métrica nos permite deixar mais evidente a dispersão dos resíduos em relação ao parâmetro observado. Um problema dessa métrica é que o valor verificado não está na mesma dimensão das unidades observadas, já que o resultado é um erro quadrático e muitas vezes não faz sentido falar em unidades quadráticas, como em relação a crimes.

- **Raiz Quadrada do Erro Médio Quadrático**

$$RMSE = \sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2} \quad (4.16)$$

A fim de corrigir o problema de interpretação ocorrido com o MSE utiliza-se a raiz do erro quadrático médio. Assim, consegue-se manter a penalização dos resíduos mais dispersos em relação à média e ganha-se em termos de interpretação, já que o resultado está na mesma dimensão dos dados observados, facilitando a compreensão do impacto do erro que a predição pode ter em relação ao dado.

- **Coefficiente de Determinação**

$$R^2 = 1 - \frac{SS_{RES}}{SS_{TOT}} = 1 - \frac{\sum_{i=1}^n (y_i - \hat{y}_i)^2}{\sum_{i=1}^n (y_i - \bar{y}_i)^2} \quad (4.17)$$

Todas as métricas utilizadas até aqui são negativamente orientadas, ou seja, quanto menor o seu valor, melhor. Não é o caso, do R^2 , neste caso quanto maior o valor obtido melhor, os valores podem variar entre 0 e 1. O R^2 apresenta quanto um modelo consegue explicar um problema analisado através de uma reta ajustada. Se todas as observações incidirem sobre a reta ajustada então a soma do quadrado dos

resíduos será nula de forma que R^2 será 1, indicando total correlação entre as variáveis previsoras e a variável dependente. Já se as observações estiverem muito dispersas em relação à reta ajustada, então o valor do quadrado dos resíduos será grande, dessa forma o valor do coeficiente de determinação será 0, indicando que o modelo não consegue explicar a variável dependente. Vale lembrar que essa métrica de explicabilidade do modelo é relativa a um modelo linear.

- **Durbin-Watson**

O teste Durbin-Watson (D-W) é muito utilizado em modelos de regressão para testar se há autocorrelação entre os resíduos. Quando houver a presença de autocorrelação a previsão pode parecer mais significativa do que de fato é em função da subestimação dos erros. A hipótese nula do teste é de que não existe correlação. O resultado do modelo utilizado pode variar entre 0 e 4, o valor esperado para garantir a não existência de autocorrelação é próximo a 2, caso o resultado seja mais próximo de 0 ele indica a existência de correlação serial positiva, caso seja próximo de 4 ele estará demonstrando a presença de correlação serial negativa. O método utilizado é descrito abaixo:

$$Durbin - Watson = \frac{\sum_{t=2}^T ((e_t - e_{t-1})^2)}{\sum_{t=1}^T e_t^2} \quad (4.18)$$

- **Augmented Dickey-Fuller**

O teste de Dickey-Fuller permite verificar se uma série de tempo é estacionária ou não. Este tipo de teste é fundamental para verificar a qualidade de uma previsão, já que muitos modelos apesar de apresentarem grande potencial preditivo, são construídos sobre séries não estacionárias, portanto possuindo uma tendência. Uma série estacionária possui características estatísticas constantes ao longo do tempo, ainda assim é uma série que se desenvolve aleatoriamente. O resultado do teste é mais negativo para séries mais estacionárias. Após calculada a diferença entre os elementos ao longo

da série, o Dickey-Fuller (DF) é testado para $\gamma = 0$ como hipótese nula, e $\gamma < 0$ como hipótese alternativa. A hipótese nula é de que o processo é não-estacionário.

$$\Delta y_t = \alpha + \beta t + \gamma y_{t-1} + \delta_1 \Delta y_{t-1} + \dots + \delta_{p-1} \Delta y_{t-p+1} + \varepsilon_t \quad (4.19)$$

$$DF_\tau = \frac{\hat{\gamma}}{SE(\hat{\gamma})} \quad (4.20)$$

- **Erro Relativo Absoluto**

O erro relativo absoluto (RAE) é utilizado para comparar a performance de um modelo predictor em relação a um modelo muito simples que apenas prevê a média dos eventos. Quanto maior o RAE menos interessante é o modelo, resultados próximos a zero indicam um modelo mais eficaz.

$$RAE = \frac{\sum_{i=1}^n |\hat{y}_i - y_i|}{\sum_{i=1}^n |\bar{y}_i - y_i|} \quad (4.21)$$

- **Raiz Quadrada do Erro Quadrático Relativo**

Da mesma forma que o RAE, a raiz quadrada do erro quadrático relativo (RRSE) compara o desempenho do modelo com outro muito simples que apenas prevê a média dos eventos. A diferença entre os dois é que o RRSE compara os modelos a partir da

raiz quadrada do erro médio quadrático (RMSE), enquanto o RAE compara a partir do erro absoluto médio (MAE).

$$RRSE = \sqrt{\frac{\sum_{i=1}^n (\hat{y}_i - y_i)^2}{\sum_{i=1}^n (\bar{y}_i - y_i)^2}} \quad (4.22)$$

- **Mediana do Erro Absoluto Percentual**

Outra métrica interessante para avaliar modelos é utilizar a mediana do erro absoluto percentual (MdAPE). Com essa medida é possível verificar qual a mediana dos erros percentuais, de forma a evitar o problema de *outliers* influenciarem a média, por isso pode ser mais interessante analisar a mediana dos erros percentuais do que a média dos erros percentuais.

Essa métrica tem uma particularidade que pode impossibilitar a sua utilização em alguns casos. Quando o erro percentual é muito pequeno ela não consegue identificar a mediana e o MdAPE acaba ficando zerado, o contrário também pode ser um problema, quando o erro percentual é muito grande o valor do MdAPE explode e então não é possível identificar a mediana da mesma forma. Mas esses problemas ocorrem especificamente em modelos com erros extremamente pequenos ou muito grandes, de certa forma essa condição do MdAPE já indica particularidades dos modelos analisados.

$$MdAPE = \left| \frac{y_i - \hat{y}_i}{y_i + \hat{y}_i} \right| = \left| \frac{e_i}{y_i + \hat{y}_i} \right| \quad (4.23)$$

- **Raiz Quadrada do Erro Médio Quadrático Normalizado**

Muito utilizado para avaliar modelos de previsão a raiz quadrada do erro médio quadrático normalizado compara o desempenho do RMSE com a média dos eventos.

Essa métrica pode ser interessante pois permite comparar modelos que estão em diferentes escalas, caso dessa pesquisa já que pretende-se verificar o desempenho dos modelos para diferentes números de clusters.

$$NRMSE = \frac{\sqrt{\frac{1}{n} \sum_{i=1}^n (y_i - \hat{y}_i)^2}}{\bar{Y}} \quad (4.24)$$

Capítulo 5

RESULTADOS E DISCUSSÃO

Inicia-se agora a demonstração dos resultados obtidos nesta pesquisa e posteriormente a discussão sobre o impacto desses na gestão da segurança pública. Faz-se necessário ressaltar as características do ambiente computacional onde estes resultados foram obtidos. Todas as análises foram realizadas em compilador Jupyter Notebook (versão 6.0.1); as análises de regressão e classificação foram obtidas através do framework XGBoost (versão 1.2.0-SNAPSHOT), que utiliza o método da descida do gradiente, como ressaltado anteriormente. Quando necessário realizar transformações na escala dos dados o framework utilizado foi o Sklearn (versão 0.21.3). Já para os modelos de redes neurais, tanto o modelo DNN quanto LSTM, foram utilizadas as bibliotecas Tensorflow (versão 1.14.0) e Keras (versão 2.3.1).

5.1 ANÁLISE EXPLORATÓRIA DOS DADOS

Os dados fornecidos pela SSP/RS foram analisados afim de permitir uma melhor compreensão da dimensão da criminalidade na cidade de Porto Alegre ao longo do período estudado. Como descrito anteriormente os crimes foram divididos em quatro grupos. Para facilitar a compreensão os crimes classificados como CVLI (crimes violentos letais intencionais) serão doravante denominados como homicídios, já que estes representam significativamente (95,51%) o total de eventos deste tipo. O crime descrito como "Roubo a Comércio" refere-se a roubos a todos os tipos de estabeleci-

mentos empresariais, incluindo escolas, lembrando que não estão presentes os furtos, já que o objetivo do programa RS SEGURO é combater crimes violentos, que possam oferecer risco a vida da vítima. Na tabela 5.1, logo abaixo, estão representados os dados fornecidos pela SSP/RS.

Tabela 5.1. Distribuição dos eventos antes da limpeza dos dados

Crime	Eventos	%
Homicídios	16.720	4,15
Roubo de Veículo	105.982	26,29
Roubo a Pedestre	247.122	61,31
Roubo a Comércio	33.270	8,25
Total	403.094	100,00

Fonte: Elaborado pelo autor.

Os eventos representam cada ocorrência registrada entre janeiro de 2005 e outubro de 2019, o total de 403.094 ocorrências não pôde ser utilizado na sua totalidade pois muitos registros não continham informações mínimas que possibilitassem identificar as coordenadas geográficas. Infelizmente 4.994 homicídios não possuíam endereço informado, o que representa praticamente um terço do total de crimes deste tipo. Nos demais crimes a proporção de dados desprezados foi significativamente menor, o que nos permitiu analisar 351.980 eventos do total fornecido. A distribuição final dos dados observados está apresentada a seguir na tabela 5.2.

Tabela 5.2. Distribuição dos eventos após a limpeza dos dados

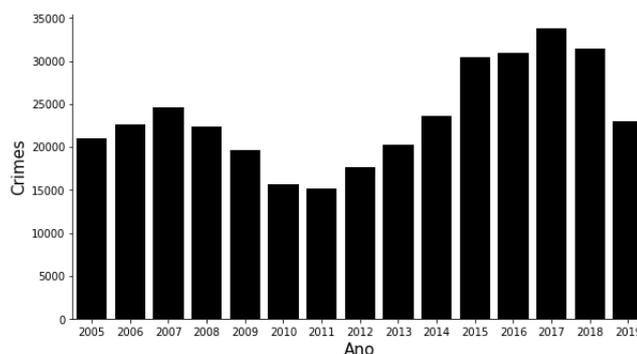
Crime	Eventos	%
Homicídios	11.726	3,33
Roubo de Veículo	105.916	30,09
Roubo a Pedestre	201.315	57,20
Roubo a Comércio	33.023	9,38
Total	351.980	100,00

Fonte: Elaborado pelo autor.

Ao analisar o comportamento da atividade criminosa anualmente, percebe-se que esta se dá de forma escalonada, indicando ondas de criminalidade ao longo dos anos estudados. O número de ocorrências em 2005 era de pouco mais de 20 mil eventos, em 2018 esse número foi 49,70% maior. Ao verificar a redução da criminalidade no

ano de 2019 no gráfico seguir, representado na figura 5.1, vale lembrar que os dados estão limitados a outubro de 2019, portanto faltam dois meses de informação no último ano analisado.

Figura 5.1. Todos os crimes acumulados anualmente



Fonte: Elaborado pelo autor.

Os crimes estudados apresentam variação nada uniforme no período analisado. Abaixo, compara-se os eventos no primeiro ano e no último ano completo dos dados disponibilizados. Enquanto os homicídios e os roubos a pedestres tiveram aumento de 72,15% e 84,01% respectivamente, os roubos a veículo cresceram 42,92% e os roubos a comércios reduziram 59,96%, como demonstrado pela tabela 5.3:

Tabela 5.3. Evolução dos crimes de 2005 a 2018

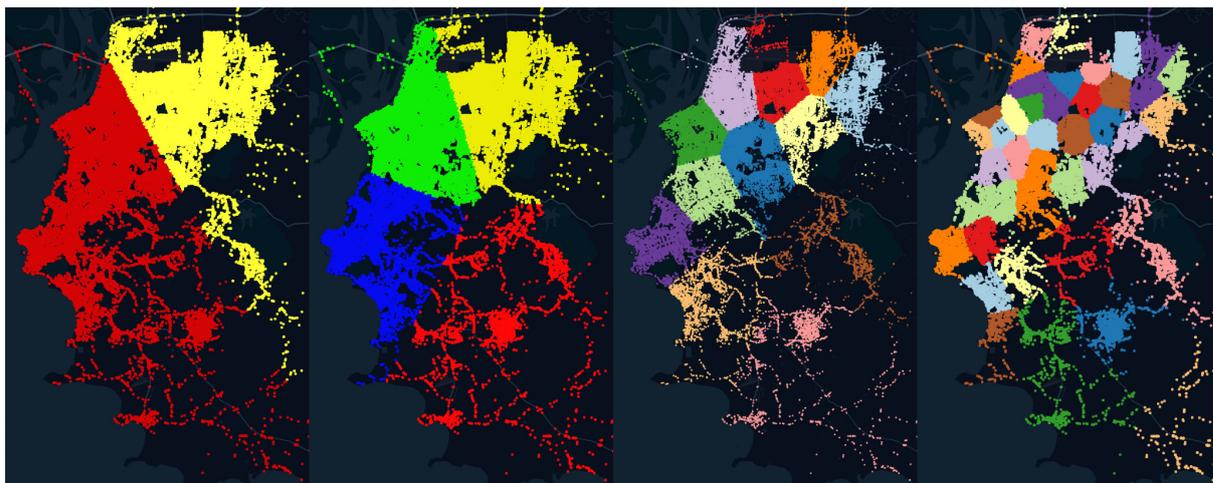
Crime	2005	2018	%
Homicídios	596	1.026	72,15
Roubo de Veículo	5.950	8.504	42,92
Roubo a Pedestre	11.166	20.547	84,01
Roubo a Comércios	3.249	1.301	-59,96
Total	20.961	31.378	49,70

Fonte: Elaborado pelo autor.

Como o principal objetivo desta pesquisa é fornecer ferramentas que possibilitem a previsão dos eventos criminosos na cidade de Porto Alegre, é preciso analisar a distribuição desses eventos pelo espaço geográfico da cidade. Para isso, faz-se necessário fragmentar a distribuição da criminalidade da cidade. Ao invés de utilizar os 84 bairros de Porto Alegre já convencionados nas análises territoriais, optou-se por criar

clusters de eventos. Essa repartição pode ser feita de infinitas formas, por isso no momento de criar os modelos de previsão estes foram testados com diversas combinações diferentes. Para ilustrar algumas possibilidades demonstra-se na figura 5.2 uma comparação da distribuição dos crimes no espaço geográfico fragmentado em 2 clusters, 4 clusters, 12 clusters e 42 clusters, respectivamente.

Figura 5.2. Possibilidades de clusterização

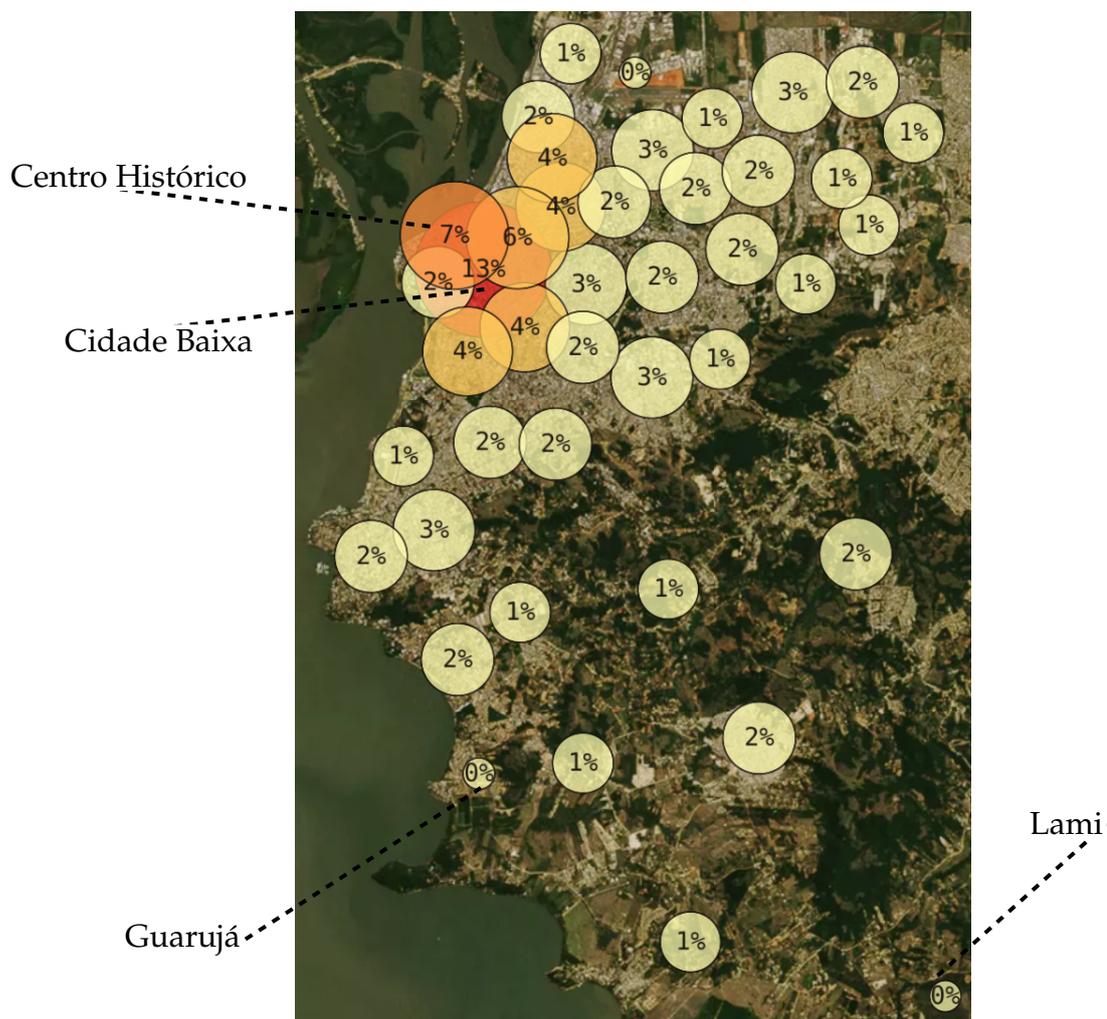


Fonte: Elaborado pelo autor.

Para compreender melhor a distribuição de cada crime na cidade optou-se por utilizar o modelo com 42 clusters. A seguir, é apresentada a organização geográfica dos dados obtidos segundo essa clusterização em 42 grupos. Os percentuais nas imagens foram arredondados para permitir uma melhor visualização.

Na figura 5.3 pode-se verificar a proporção de homicídios na cidade de Porto Alegre. De todos os 42 clusters, o que teve mais registros de homicídios foi o cluster que representa a região do bairro Restinga, com 8,00% do total da série. O segundo cluster com mais homicídios foi o que representa a região do bairro Santa Tereza com 5,17% do total. O cluster com menor registro de homicídio foi o que representa a região dos bairros Boa Vista e Três Figueiras com 0,42% dos registros, o segundo menos violento registrou 0,53% dos homicídios na série e representa a região do bairro Auxiliadora. Do total de 42 clusters, 27 deles possuem proporção menor que 3% do total de eventos.

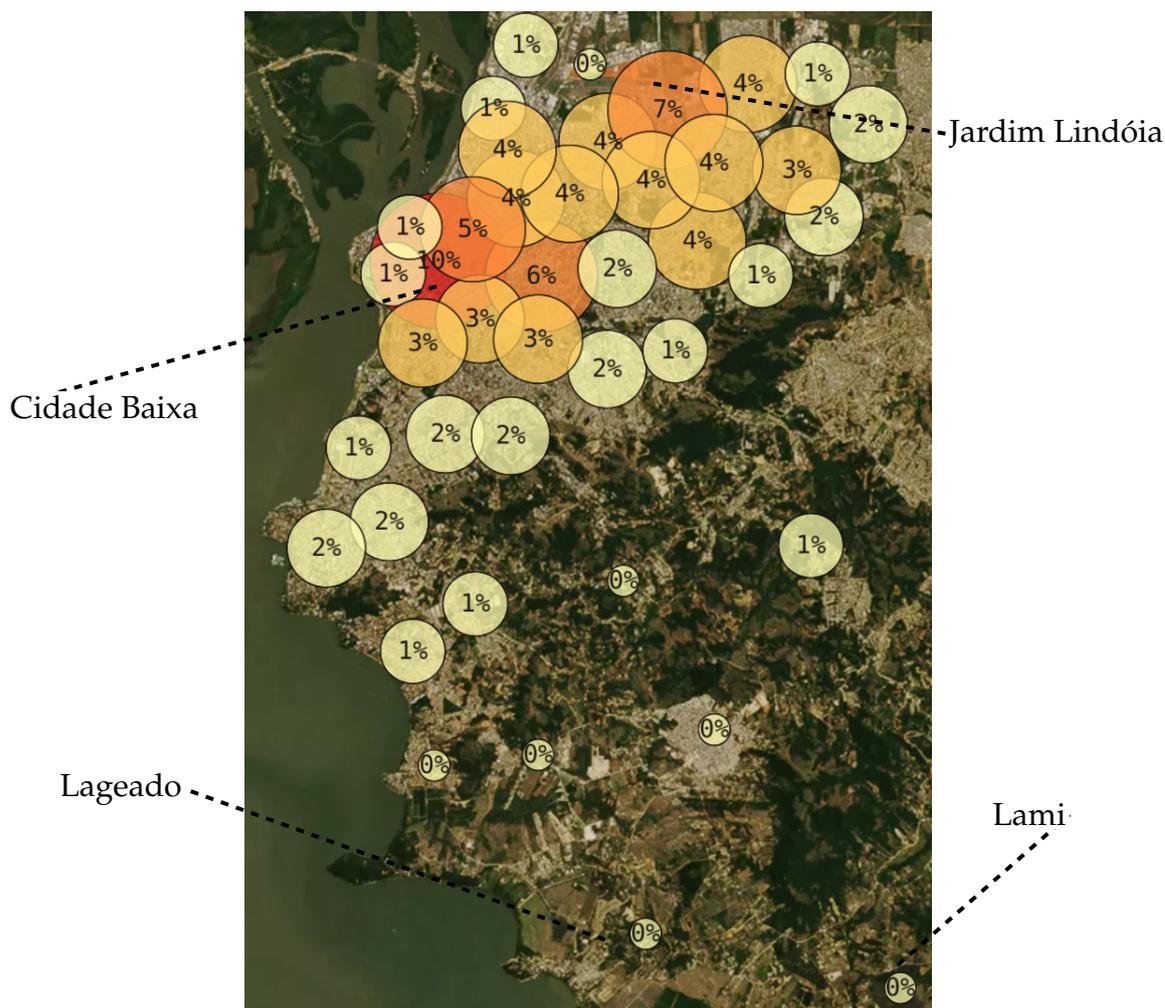
Figura 5.5. Proporção do roubos a comércios por cluster



Fonte: Elaborado pelo autor.

Os dois clusters que apresentam maior proporção dos roubos de veículos são os que representam as proximidades dos bairros Cidade Baixa e Jardim Floresta/Lindóia, representando respectivamente 10,04% e 6,82% do total. Os clusters que representam as regiões dos bairros Lami e Lageado, com 0,10% e 0,23% respectivamente, são os dois clusters com menor número de ocorrências, ambos clusters representam regiões da zona sul, como demonstrado na figura 5.6.

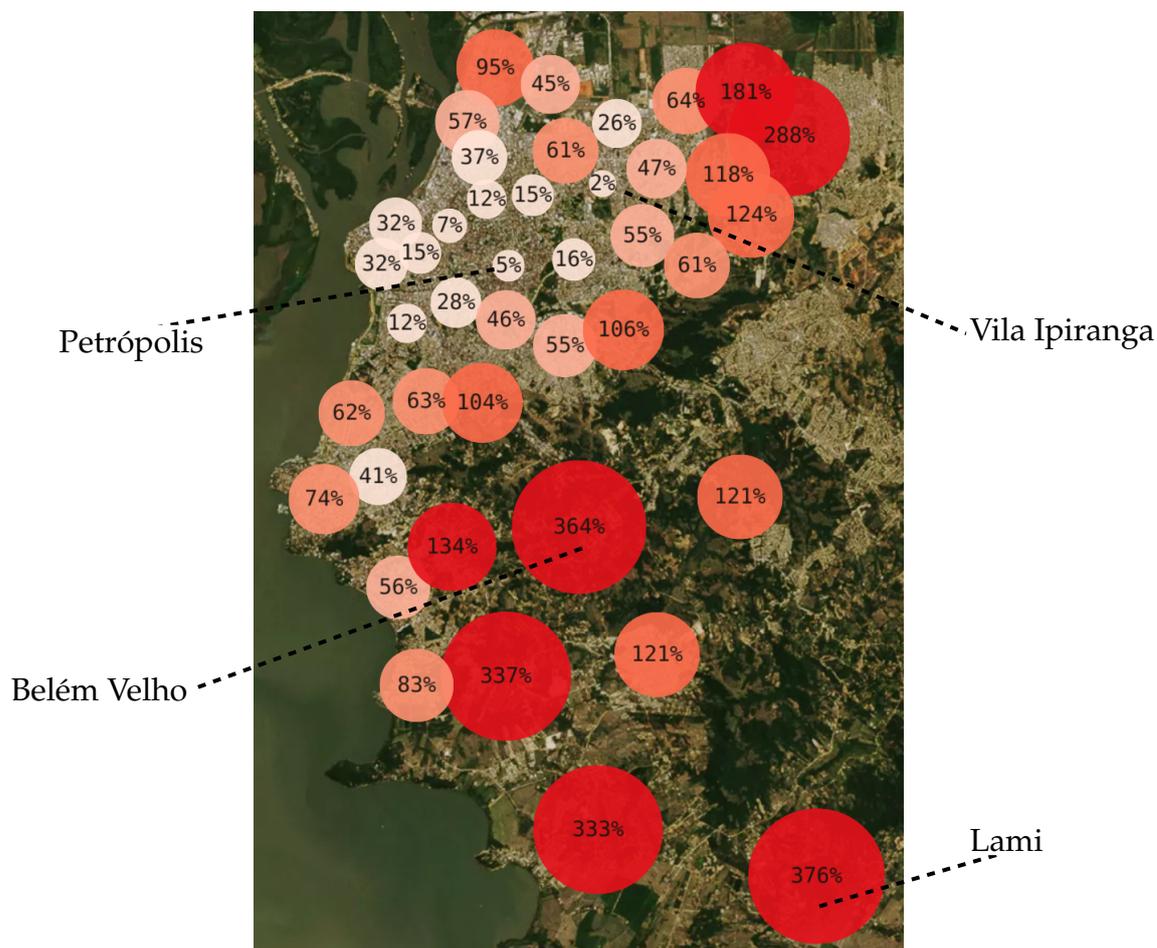
Figura 5.6. Proporção do roubos a veículo por cluster



Fonte: Elaborado pelo autor.

Percebe-se nos mapas apresentados que os clusters mais distantes da região central da cidade geralmente têm uma participação menor na criminalidade da cidade, excluindo-se o bairro Restinga que no caso dos homicídios tem destaque entre todos clusters. Dessa forma se faz interessante observar a evolução da criminalidade em cada cluster ao longo dos anos analisados. No mapa representado na figura 5.7 é apresentada a evolução do número de eventos criminosos em cada cluster entre os anos de 2005 e 2018 (último ano inteiro dos dados fornecidos).

Figura 5.7. Evolução da criminalidade nos 42 clusters



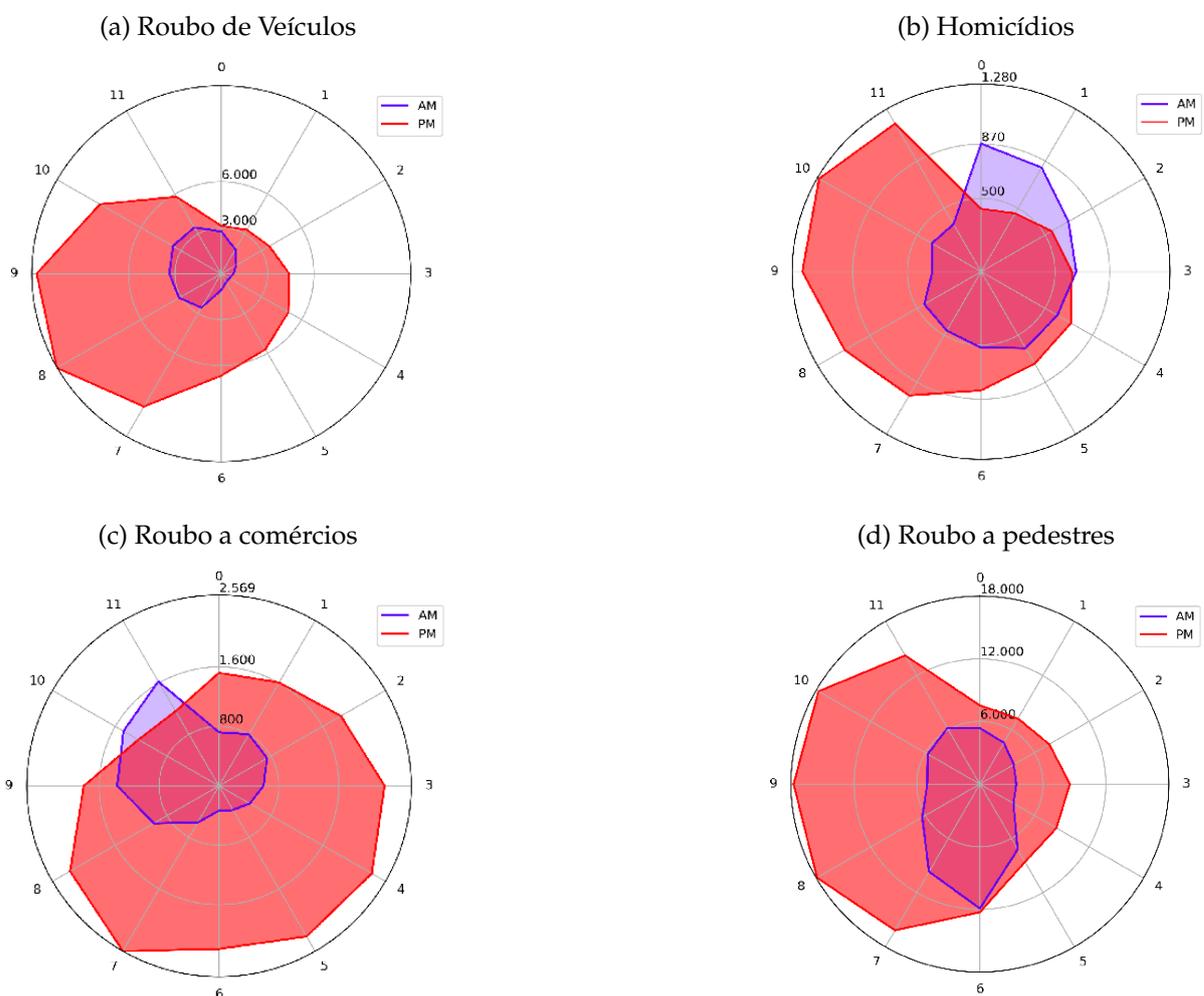
Fonte: Elaborado pelo autor.

Neste mapa é possível verificar que os clusters mais afastados da região central da cidade são aqueles que tiveram maior aumento da criminalidade no período. Os clusters que apresentam as maiores taxas de aumento da criminalidade são os que representam as proximidades dos bairros Belém Velho e Lami, com aumentos respectivos de 364,4% e 376,4%. Já as regiões que tiveram maior estabilidade no número de eventos dentro da série são as regiões dos bairros Vila Ipiranga, com aumento de apenas 2,37% e a do bairro Petrópolis, que registrou crescimento de 4,79% no número de crimes nos 14 anos analisados.

Além da região onde ocorrem os crimes, outro fator relevante para a gestão da segurança pública é identificar em quais momentos do dia os crimes geralmente ocor-

rem, para assim melhor construir uma estratégia de atuação. Abaixo pode-se verificar o comportamento de cada crime analisado durante as horas do dia. Percebe-se que o crime de homicídio (b) ocorre em grande parte entre 7pm e 1pm, já o crime de roubo de veículos (a) ocorre sobretudo entre 7pm e 10pm. No caso do roubo a pedestres (d) não há muita diferença, esses crime ocorrem geralmente entre 7pm e 11pm e o crime de roubo a estabelecimentos comerciais (c) parece ter um comportamento diferente dos demais, esses crimes ocorrem com maior frequência entre as 2pm e 9pm.

Figura 5.8. Crimes por hora



Fonte: Elaborado pelo autor.

Após compreender algumas características dos crimes da cidade na análise até aqui apresentada, a seguir será possível verificar a descrição dos resultados utilizados na previsão dos eventos criminosos e as particularidades dos modelos adotados.

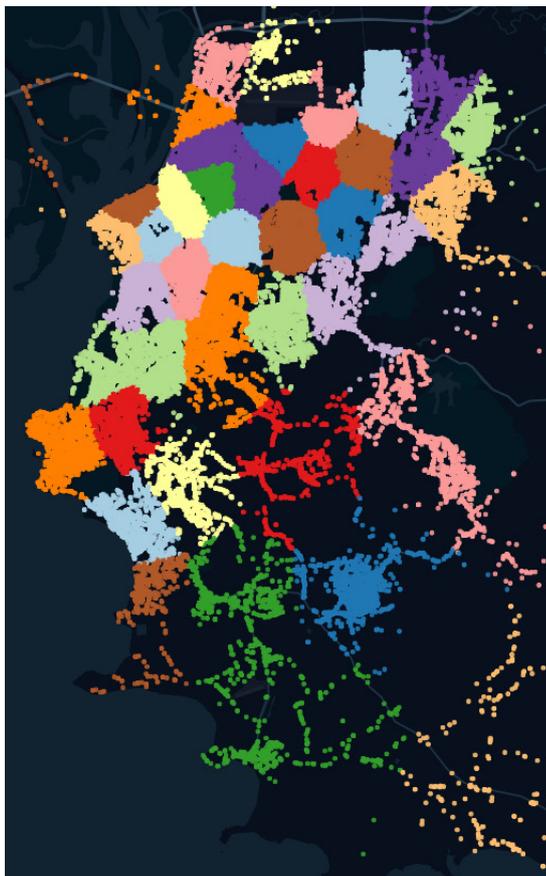
5.2 PREVISÃO DA CRIMINALIDADE

Na subseção anterior foi possível verificar que as regiões da cidade têm distintas realidades quanto a dimensão que a criminalidade assume no dia a dia e a sua evolução ao longo da série fornecida pela SSP/RS. Ao pensar nisso, opta-se por montar modelos que pudessem gerar previsões para cada região da cidade de forma independente.

Prever o comportamento da cidade de forma unificada pode ter algumas vantagens em termos de gestão pública do Estado do Rio Grande do Sul, porém, limitando este trabalho aos interesses da cidade de Porto Alegre, já que os dados são exclusivamente desta, parece mais interessante para o gestor obter previsões de fragmentos da cidade e assim conseguir alocar melhor os recursos.

Assim como na subseção anterior, foi utilizado um processo de clusterização dos eventos e não os bairros da cidade para compreender os dados da série. Agora será utilizado este mesmo método para a construção das regiões a serem compreendidas para montar os modelos preditivos. Os clusters foram construídos de acordo com o método conhecido por *K-means*, descrito na subseção 4.1.3.

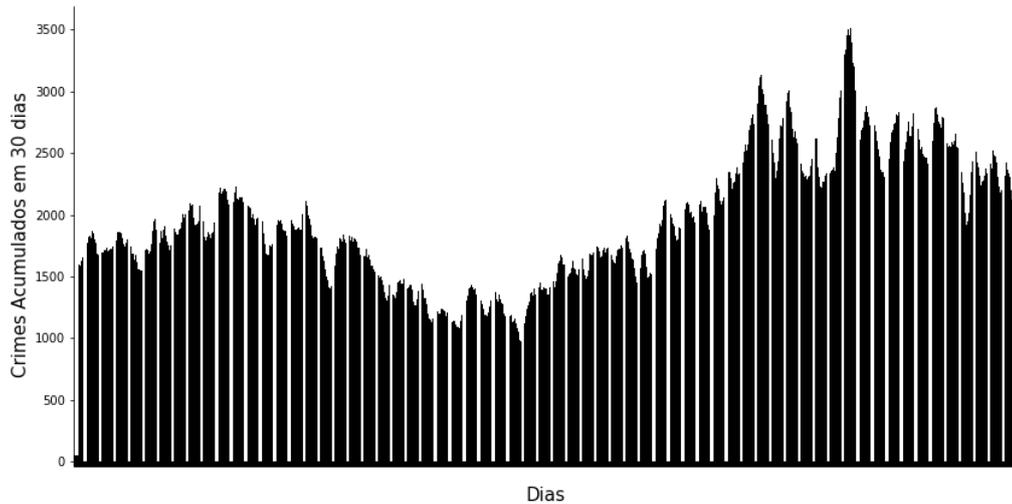
Dessa forma, os 351.980 eventos foram agrupados de diferentes formas. Como a cidade possui 84 bairros, um dos formatos escolhidos foi a clusterização em 84 regiões distintas. Entre tantas outras possibilidades de segmentação, apresenta-se a seguir os resultados obtidos para a fragmentação dos eventos em 42,24,12 e 6 clusters. De maneira que é possível comparar a capacidade dos modelos em prever eventos em diferentes cenários. A figura 5.9 apresenta dos eventos clusterizados em 42 grupos.

Figura 5.9. Clusterização em 42 grupos

Fonte: Elaborado pelo autor.

Para gerar estes clusters foi necessário construir um algoritmo capaz de verificar junto a serviços de geolocalização as coordenadas geográficas de cada evento. Infelizmente essas informações só constam na base de dados fornecida a partir de 2017. Sendo assim todos os eventos anteriores tiveram que ter seus atributos geográficos requisitados ao serviço de geolocalização utilizado.

A fim de construir os atributos necessários para utilizar nos modelos previsores foi necessário analisar detalhadamente o comportamento da série para além do que foi apresentado na subseção anterior. Um atributo que mostrou grande importância para o modelo foi a utilização dos crimes acumulados nos últimos 30 dias em cada cluster para compreender os eventos diários. Como apresentado na figura 5.10, é possível verificar o comportamento dos eventos em ondas de criminalidade.

Figura 5.10. Acumulado dos crimes nos 30 dias anteriores

Fonte: Elaborado pelo autor.

Também foram utilizados como atributos previsores o número de crimes ocorridos no dia anterior em cada cluster, o dia da semana, dia do mês, o mês, ano e o risco. Este último atributo que denomina-se *risco*, foi construído pensando na utilidade que o modelo precisa entregar a gestão pública. Mesmo que em termos de resultado do modelo faça diferença, a sua capacidade de prever o número exato de eventos criminosos para fins operacionais da pasta da segurança pública, basta prever um determinado nível de atividade criminosa em cada cluster.

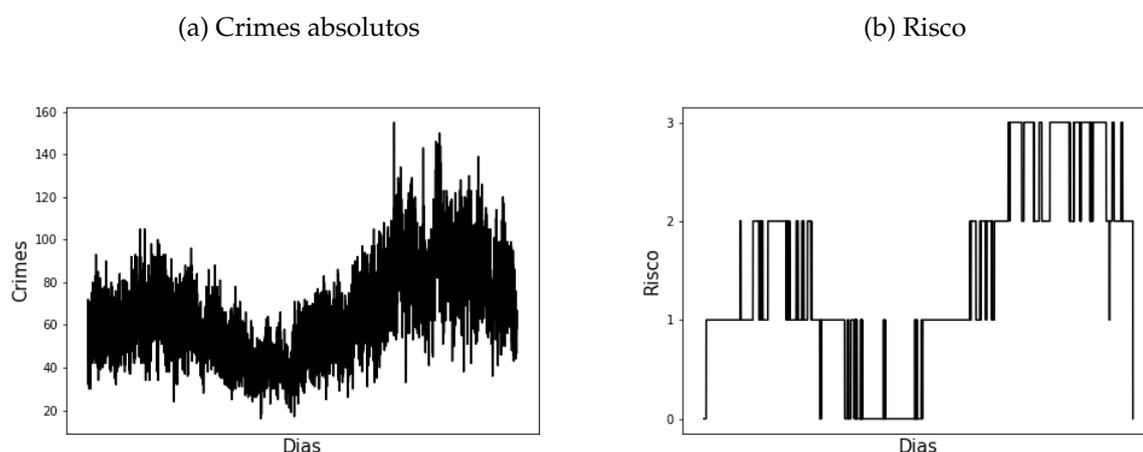
Dessa forma, foram criados 4 níveis de risco: fraco, moderado, alto e muito alto. Para classificar cada nível da variável risco foi utilizado o atributo de crimes acumulados nos últimos 30 dias. Foi analisada a distribuição deste atributo em todos os clusters e separada em cada nível de risco pelo desvio padrão da distribuição. Assim, seja Z os crimes acumulados nos últimos 30 dias:

$$\sigma = \sqrt{\sum_{i=1}^n \frac{(Z_i - \hat{Z}_i)^2}{n - 1}} \quad (5.1)$$

Então, caso um cluster apresentar em determinado dia $Z < -\sigma$ ele será clas-

sificado como de risco fraco. Se apresentar $-\sigma < Z < \mu$ será classificado com risco moderado, já se $\mu < Z < +\sigma$ então será considerado de risco alto, por fim se apresentar $Z > +\sigma$ será classificado como de risco muito alto. Este atributo será utilizado principalmente nos modelos preditivos que utilizam métodos de classificação e serão demonstrados posteriormente. A figura abaixo ilustra como os dados podem ser classificados do número absoluto de crimes para a variável risco utilizando a metodologia acima descrita. No gráfico (b) da figura 5.11 o risco está demonstrado no eixo y da seguinte forma: 0 para risco fraco, 1 para moderado, 2 para alto e 3 para risco muito alto.

Figura 5.11. Comparação dos atributos Crime vs Risco



Fonte: Elaborado pelo autor.

5.2.1 Regressão

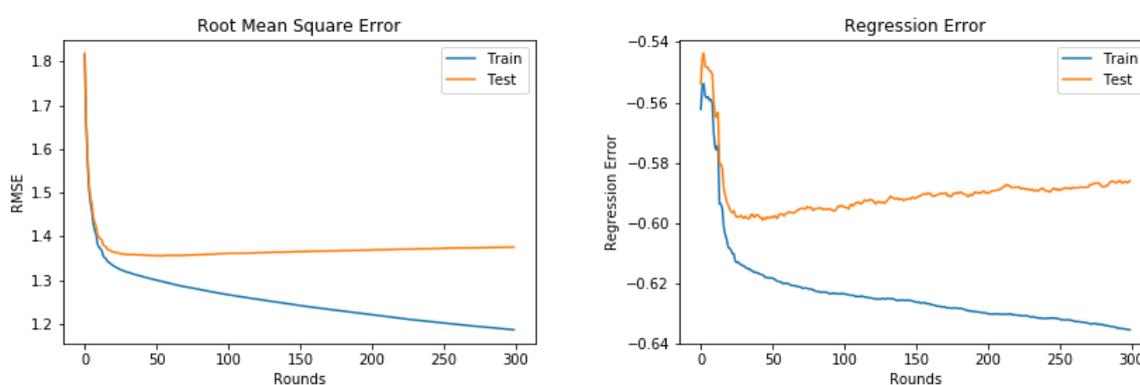
Antes de iniciar as previsões, todos os modelos foram treinados para atingirem um alto desempenho na predição dos eventos. Abaixo, é demonstrada a evolução de algumas métricas ao longo do treinamento do modelo de regressão. Outras métricas foram utilizadas, porém, para não estender esta apresentação não serão demonstradas.

Na figura 5.12 pode-se observar a evolução das métricas a cada rodada de análise do algoritmo, tanto na base de dados utilizada, como treinamento quanto na parte da base separada para testar o algoritmo. É possível comparar o desempenho do mo-

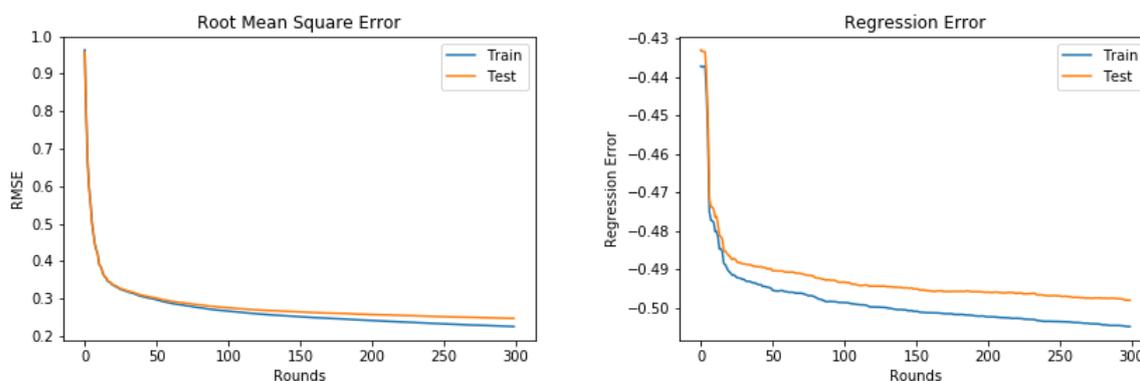
delo prevendo os eventos criminosos diários de cada cluster com a previsão do risco de violência naquele instante para cada cluster. Percebe-se que conforme avançam as rodadas de análise o algoritmo vai diminuindo o erro das previsões, assim como também o RMSE. No caso das previsões para os eventos criminosos é possível verificar que a base de teste não tem um desempenho tão bom ao longo das rodadas quanto a previsão do risco diário em cada cluster.

Figura 5.12. Treinamento do Algoritmo

(a) Prevendo Crimes



(b) Prevendo Risco



Fonte: Elaborado pelo autor.

A tabela 5.4, apresentada a seguir, exhibe alguns resultados obtidos ao utilizar as métricas descritas na metodologia para avaliação dos modelos. Como pode-se observar o teste Durbin-Watson (D-W) mostra que para qualquer modelo de clusterização analisado a presença de correlação serial é mínima e quanto maior o número de clus-

ters utilizado no modelo menos autocorrelação existe.

Tabela 5.4. Autocorrelação e Estacionariedade do modelo de Regressão prevendo Crime

Nº de Clusters	Durbin-Watson	ADF	Valores Críticos (1%)	valor-p
6	1.77	-5.54	-3.43	0.00
12	1.60	-28.31	-3.43	0.00
24	1.82	-49.24	-3.43	0.00
42	1.96	-74.01	-3.43	0.00
84	1.98	-106.85	-3.43	0.00

Fonte: Elaborado pelo autor.

Já o teste Augmented Dickey-Fuller (ADF), mostra que todas as séries analisadas são estacionárias. O valor crítico de 1% do teste mostra a pontuação máxima necessária para recusar a hipótese nula ao nível de 1%. Todos os modelos de clusters apresentam valores inferiores ao necessário para recusar a hipótese nula ao nível estipulado e todos apresentam valor-p menor que 0.00.

Observando a tabela 5.5, logo abaixo, verifica-se os resultados do coeficiente de determinação (R^2) obtidos, pode-se perceber que modelos que utilizam um número menor clusters possuem um potencial de explicação da predição maior. Enquanto o modelo de regressão utilizando 84 clusters obteve um coeficiente de 0,42, o modelo que usou apenas 6 clusters apresentou um R^2 de 0,87, um resultado 45 centésimos superior. Este resultado mostra uma correlação de 0,93 para o modelo com 6 clusters e 0,64 para 84 clusters.

Tabela 5.5. Métricas do modelo de Regressão prevendo Crime

Nº de Clusters	MAE	MSE	RMSE	RAE	RRSE	MdAPE	NRMSE	R^2
6	1.76	6.53	2.55	0.30	0.34	0.14	0.23	0.87
12	1.66	5.54	2.35	0.40	0.41	0.25	0.36	0.79
24	1.13	2.32	1.52	0.58	0.59	0.41	0.56	0.62
42	0.87	1.38	1.17	0.64	0.65	0.59	0.76	0.55
84	0.62	0.74	0.86	0.72	0.74	—	1.11	0.42

Fonte: Elaborado pelo autor.

O erro médio absoluto (MAE) para 6 clusters indica que o modelo erra pouco

menos de 2 crimes em relação ao número absoluto de crimes diários em cada cluster. Esse número cai conforme aumenta o número de clusters utilizados no modelo, influenciados também pelo fato de que um número de clusters maior indica uma frequência de crimes menor em cada cluster. Afinal, os eventos criminosos ficam mais distribuídos no espaço geográfico da cidade.

Utilizando a raiz quadrada do erro quadrático médio (RMSE) pode-se penalizar o modelo que apresenta erros maiores em relação ao dado real a ser previsto. Em se tratando da construção de um modelo para auxiliar na gestão dos recursos para segurança pública esta métrica mostra-se importante. Dessa forma, observa-se que o RMSE é 36,54% maior que o MAE usando um modelo com 6 clusters e 39,68% maior que o MAE no modelo com 84 clusters. Ainda assim, mesmo usando a estrutura com maior RMSE, o modelo erra em média em 3 eventos criminosos por dia em cada cluster.

Analisando tanto o erro relativo absoluto (RAE) quanto a raiz quadrada do erro quadrático relativo (RRSE), percebe-se que o modelo que utiliza menos clusters possui um desempenho melhor em relação a modelos simples que utilizam apenas a média como um previsor. Valores próximos a 1 indicam um desempenho muito similar ao modelo simples que usa a média como previsão. Portanto, os algoritmos que utilizaram seis e doze clusters tiveram um desempenho muito superior aos demais.

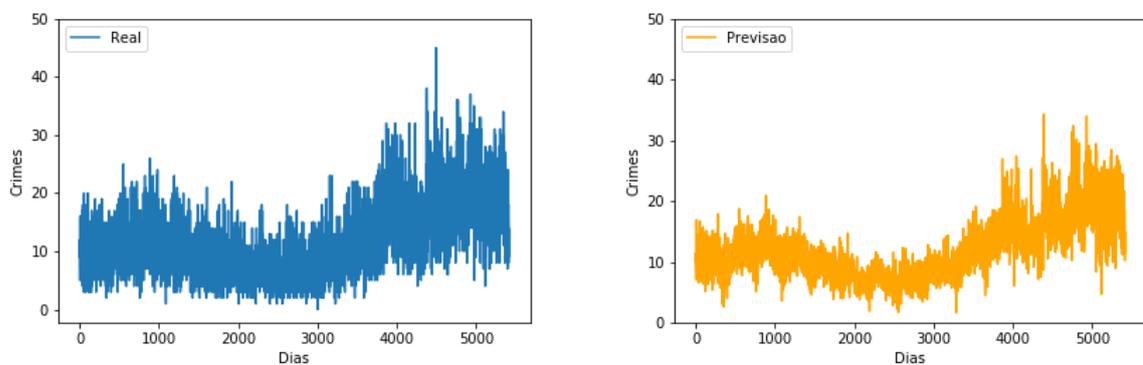
Verifica-se através do MdAPE que a mediana dos erros absolutos percentuais para os modelos de regressão que utilizam menos clusters é significativamente menor em relação aos modelos com mais clusters. O algoritmo que utilizou 6 clusters obteve um erro percentual mediano de 0,14, isso nos diz que metade dos erros percentuais está abaixo e metade acima desse valor. Já utilizando 84 clusters, a mediana nem consegue ser encontrada em função do alto percentual dos erros. Como informado na metodologia este é um problema desse tipo de erro pois quando os percentuais do erro médio são muito grandes o MdAPE explode, impossibilitando averiguar a mediana.

Através da análise do RMSE normalizado, pode-se perceber que o erro para o algoritmo que utilizou 6 clusters representa em média 23% da média dos crimes diários por cluster. Esse valor sobe muito conforme o número de clusters utilizados aumenta. O modelo com mais clusters chega a ter erros 111% superiores a média.

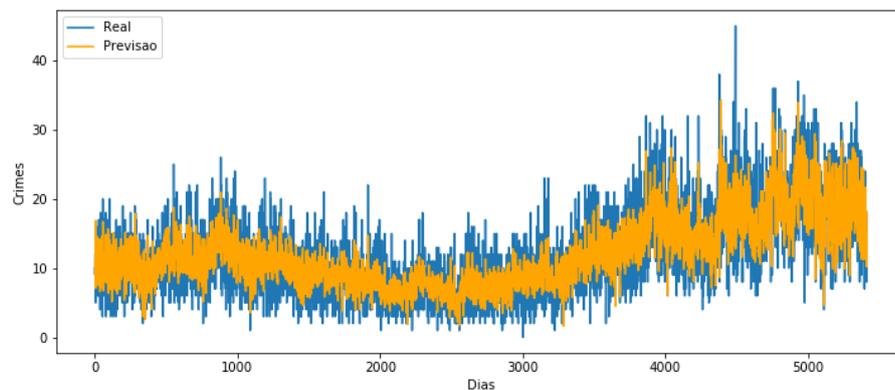
A fim de demonstrar visualmente os resultados das previsões diárias dos modelos passa-se a apresentar a seguir, na figura 5.13, exemplos de resultados obtidos. O gráfico abaixo expõe a comparação entre os dados diários reais de criminalidade para o modelo com 6 clusters e a previsão diária de eventos criminosos para o mesmo modelo. Percebe-se que apesar de existir uma diferença entre a realidade e a previsão, o modelo capturou bem a evolução da criminalidade ao longo da série.

Figura 5.13. Real vs Previsão

(a) Análise individual



(b) Análise conjunta

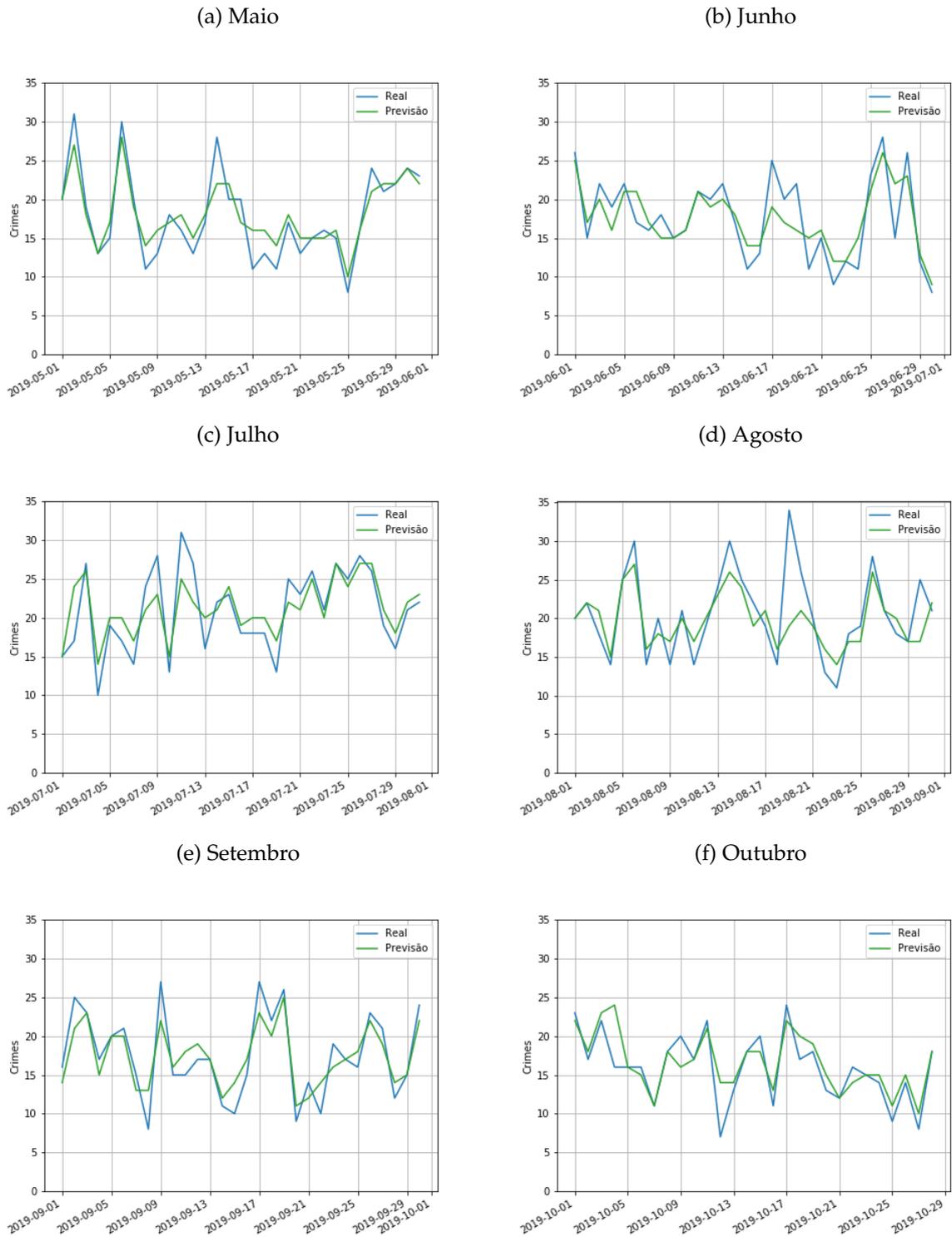


Fonte: Elaborado pelo autor.

Para demonstrar a capacidade de previsão do modelo de regressão, apresenta-se a na figura 5.14 o resultado para os últimos seis meses de toda a série histórica analisada. Neste caso as previsões são relativas a um dos 6 clusters em que os eventos foram segmentados. Aqui é possível perceber que apesar de o modelo não ser preciso em muitos dias, ele compreende muito bem a dinâmica em que os eventos ocorrem dia

após dia. Como pode-se observar na figura a seguir:

Figura 5.14. Comparativo dos últimos 6 meses da série - Modelo com 6 clusters



Fonte: Elaborado pelo autor.

A seguir passa-se a analisar o potencial de previsão do modelo de regressão para o atributo risco. Da mesma forma que os modelos de regressão para prever os eventos absolutos, o resultado do teste Durbin-Watson (D-W) indica a não existência de autocorrelação ou efeito mínimo em alguns casos, como o modelo que utilizou 42 clusters. Ainda assim, o resultado é satisfatório.

Tabela 5.6. Autocorrelação do modelo de Regressão prevendo Risco

Nº de Clusters	Durbin-Watson
6	2.09
12	1.93
24	2.00
42	2.17
84	2.02

Fonte: Elaborado pelo autor.

Como aponta a tabela 5.7, é possível verificar que os erros analisados (MAE, MSE e RMSE) são muito pequenos, mostrando que o modelo consegue prever o risco de violência em cada cluster de forma muito satisfatória. O coeficiente de determinação mostra que modelos utilizando um número de cluster menor possuem um nível de explicação maior que os modelos com muitos clusters, apesar de o coeficiente ser consideravelmente alto até mesmo para o modelo com 84 clusters. A figura 5.15 mostra bem o potencial de previsão do atributo risco. Para exemplificar foi utilizada a previsão para o último ano da série de um dos clusters avaliados no modelo com um total de 6 clusters.

As métricas RAE e RRSE demonstram que ao prever o nível de risco e não o número absoluto de crimes o modelo de regressão consegue obter uma performance muito superior em relação a aplicação de um modelo preditivo simplificado que tem a média como previsão. Novamente, modelos com menos clusters apresentam potencial superior aos demais.

Tabela 5.7. Métricas do modelo de Regressão prevendo Risco

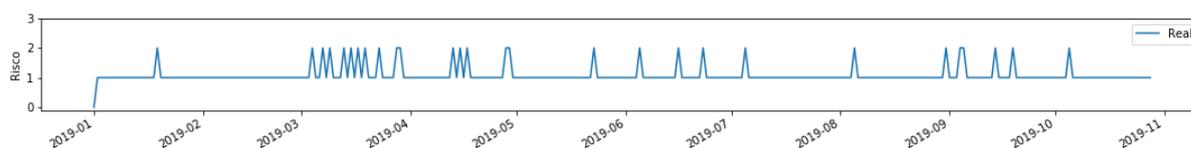
Nº de Clusters	MAE	MSE	RMSE	RAE	RRSE	MdAPE	NRMSE	R ²
6	0.05	0.01	0.10	0.06	0.10	0.03	0.06	0.97
12	0.06	0.01	0.12	0.11	0.17	0.03	0.08	0.97
24	0.10	0.03	0.17	0.14	0.21	0.03	0.11	0.96
42	0.16	0.06	0.24	0.23	0.30	0.07	0.17	0.94
84	0.19	0.08	0.28	0.30	0.37	0.09	0.19	0.94

Fonte: Elaborado pelo autor.

Analisando a mediana do erro absoluto percentual (MdAPE) verifica-se que todos os algoritmos têm erro mediano inferior a 10%, ou seja metade dos erros está abaixo e acima desse percentual, o que parece ser um ótimo resultado. Da mesma forma, ao analisar o NRMSE, percebe-se que o erro do algoritmo de regressão ao prever o nível de risco está muito próximo da média dos erros. O modelo que utiliza 6 clusters erra em média 6% em relação ao nível de risco médio de cada cluster. Já o modelo com mais clusters tem um erro normalizado de até 19% em relação a média do risco de cada cluster.

Figura 5.15. Risco Real vs Previsão do Risco

(a) Análise individual



(b) Análise conjunta

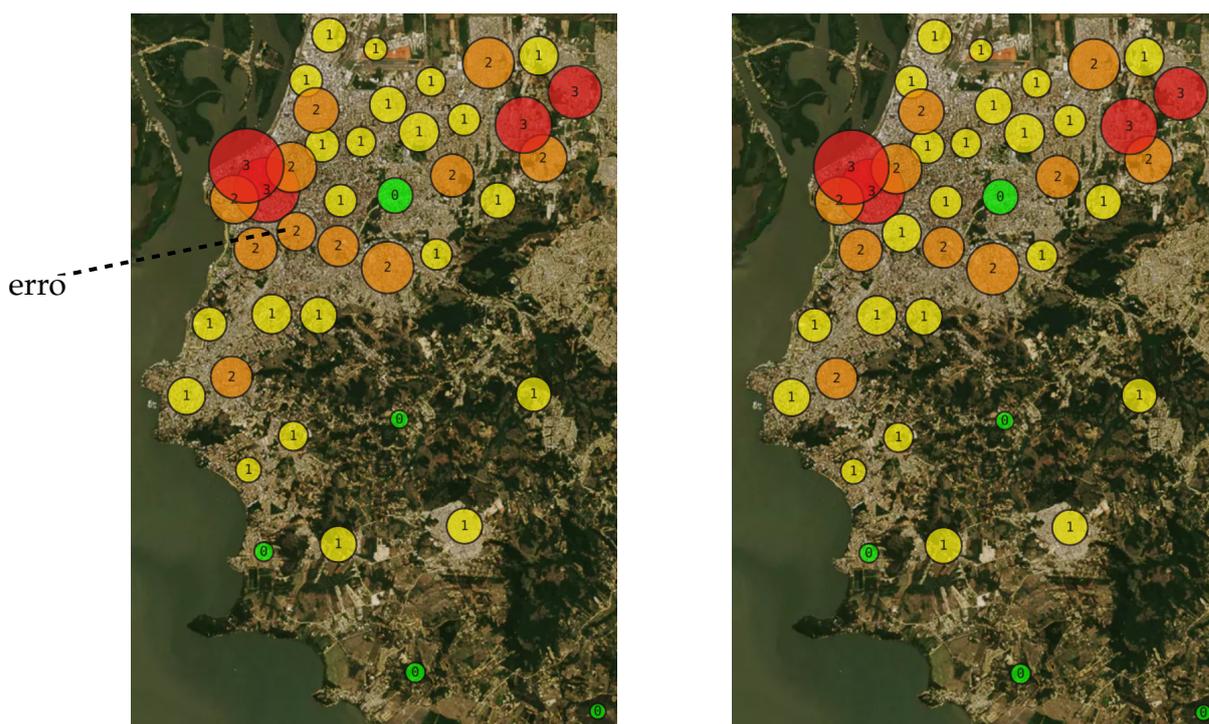


Fonte: Elaborado pelo autor.

Os resultados dos modelos para este atributo mostram que é possível prever adequadamente o risco estipulado para cada região da cidade. Essa ferramenta pode auxiliar a identificar pontos críticos da cidade e melhorar a distribuição dos recursos segundo as previsões dos modelos.

Na figura 5.16, apresentada a seguir, foi comparado o risco real de cada cluster com a previsão do modelo, neste caso o modelo escolhido foi o de 42 clusters. Os dados são para o último dia da série de dados: dia 28 de outubro de 2019. É possível verificar que a previsão erra o risco para apenas um cluster, para todos os demais a previsão é precisa. O único cluster com previsão equivocada, que está situado na região central da cidade, errou por uma unidade de risco, o dado real era nível de risco 1 e o modelo previu risco de nível 2.

Figura 5.16. Comparação do Risco para 42 Clusters: Previsão vs Real



Fonte: Elaborado pelo autor.

5.2.2 Classificação

Os próximos modelos a serem analisados são aqueles que utilizaram métodos de classificação para prever tanto o número de crimes diário em cada cluster, quanto o risco de violência diário em cada região. Os algoritmos de classificação utilizados nesta pesquisa foram construídos com base no modelo de árvores de decisão apresentados anteriormente. A tabela 5.8 apresenta os resultados das métricas escolhidas para avaliar estes modelos prevendo o atributo risco.

Tabela 5.8. Métricas do modelo de Classificação prevendo Risco

Nº de Clusters	MAE	MSE	RMSE	RAE	RRSE	MdAPE	NRMSE	R ²
6	0.01	0.01	0.07	0.01	0.08	0.00	0.05	0.98
12	0.01	0.01	0.07	0.01	0.11	0.00	0.05	0.95
24	0.02	0.02	0.13	0.02	0.16	0.00	0.09	0.94
42	0.04	0.04	0.21	0.06	0.26	0.00	0.15	0.90
84	0.08	0.08	0.27	0.12	0.36	0.00	0.19	0.85

Fonte: Elaborado pelo autor.

Percebe-se que os coeficientes de determinação (R²) de todos os modelos são muito similares aos dos modelos que utilizaram o método de regressão para prever o risco diário de cada cluster. Apenas o algoritmo que usa 84 clusters obteve R² inferior a 0,90. Outra métrica que apresentou resultados similares ao modelo de regressão foi a NRMSE, que indicou que os modelos que utilizam menos clusters tem um menor erro em relação ao valor médio dos eventos. No caso do modelo com 6 clusters, o RMSE representa apenas 5% do risco médio diário de cada cluster, já o modelo que usa 84 clusters tem um RMSE de 0,27, representa 19 % da média do risco diário de cada cluster, como demonstra o NRMSE.

Nestes modelos é possível verificar a situação onde o MdAPE não consegue encontrar a mediana dos erros médios percentuais em função do valor muito baixo que esses erros assumem. Comportamento contrário ao apresentado por esta métrica no modelo de regressão com 84 clusters, demonstrado anteriormente.

Os modelos de classificação apresentam potencial preditivo muito superior a modelos simplistas que poderiam prever apenas o comportamento médio a cada dia, é

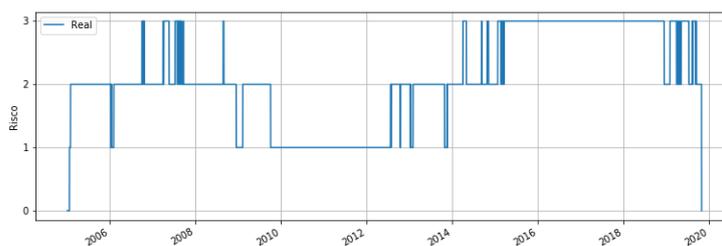
o que evidencia tanto o RAE quanto RRSE. Ambos erros comparam o desempenho dos modelos com estes modelos que tem a média como predictor. Quanto mais próximo de zero, mais diferenciada é a capacidade preditiva do modelo, isso mostra que os modelos com menos clusters tem um potencial preditivo maior, principalmente o algoritmo que utilizou apenas 6 clusters para realizar as previsões.

O erro absoluto médio (MAE) indica que estes modelos tem potencial preditivo elevado, pois o MAE é muito próximo de zero, principalmente nos modelos que utilizam 6,12 e 24 clusters. Mesmo que se utilize o RMSE a fim de penalizar mais severamente os erros maiores, os erros de modelos com poucos clusters parecem satisfatórios.

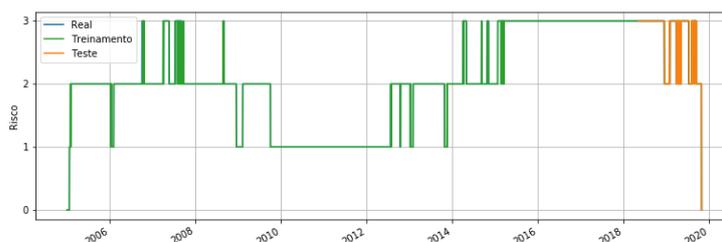
Para demonstrar a capacidade de previsão dos modelos de classificação, a seguir, na figura 5.17, são apresentadas as previsões de um único cluster para toda a série de dados. Desta vez, é demonstrado um modelo onde todo o set de treinamento é representado pelas primeiras observações e o set de teste representado pelas últimas observações. Percebe-se que o modelo prevê adequadamente o nível de risco diário da série.

Figura 5.17. Risco Real vs Previsão do Risco

(a) Real



(b) Previsão



Fonte: Elaborado pelo autor.

Os modelos de classificação também podem ser utilizados para prever o número de eventos criminosos em cada cluster. Neste caso percebe-se que o potencial de previsão destes modelos de classificação são muito interessantes. O erro absoluto médio (MAE) para o modelo que utiliza 6 clusters possui uma divergência de menos de 1 crime em relação ao dado real, o que representa uma capacidade preditiva muito boa.

Utilizando qualquer uma das estruturas de clusters o MAE obtido é inferior a 1, mostrando consistência nos resultados. Analisando o RMSE percebe-se que a penalização de erros mais dispersos em relação ao dado esperado elevam bastante o erro médio dos modelos. O NRMSE indica que o modelo com 6 clusters apresenta um RMSE que representa 16% da média dos crimes diários por cluster. Este valor sobe consideravelmente conforme mais clusters são introduzidos nos modelos. O modelo que utiliza 84 clusters apresenta NRMSE de 1,31 indicando que o RMSE representa 131% da média dos crimes diários em cada cluster, evidenciando um baixa capacidade preditiva.

Tabela 5.9. Métricas do modelo de Classificação prevendo Crime

Nº de Clusters	MAE	MSE	RMSE	RAE	RRSE	MdAPE	NRMSE	R ²
6	0.43	2.84	1.68	0.07	0.22	0.00	0.16	0.94
12	0.61	2.89	1.70	0.14	0.30	0.00	0.26	0.90
24	0.82	2.24	1.49	0.41	0.57	0.00	0.55	0.67
42	0.73	1.63	1.27	0.54	0.70	0.00	0.82	0.50
84	0.54	1.02	1.01	0.63	0.87	0.00	1.31	0.23

Fonte: Elaborado pelo autor.

Neste caso, o MdAPE também não conseguiu identificar a mediana dos erros absolutos percentuais em função dos erros representarem percentualmente valores muito baixos. Já o RAE e o RRSE, apontam que o modelo de 6 clusters detém um potencial preditivo muito superior a simples aplicação da média como variável previsora, principalmente se considerar o RAE, que detém um valor muito próximo de zero, ponto ótimo das medidas RAE e RRSE.

O coeficiente de determinação evidencia que os modelos que utilizam um número de clusters menor, possuem uma capacidade maior de explicação da variável dependente. É possível notar que, ao mesmo tempo que o R² do modelo de classi-

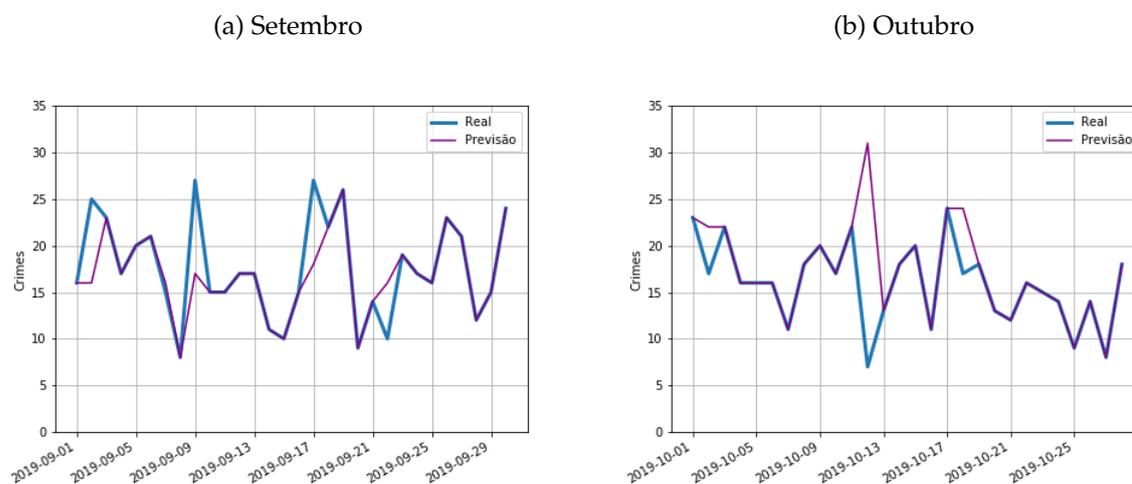
ficação com 6 clusters é superior ao R^2 do modelo de regressão (0,87), o modelo que utiliza 84 clusters tem um resultado bem inferior neste quesito em relação ao modelo de regressão (0,42).

Da mesma forma que foi apresentado anteriormente, será possível verificar o potencial de previsão dos modelos de classificação a partir da demonstração das previsões dos últimos seis meses da série para o mesmo cluster antes observado. Como pode-se perceber na figura 5.18 e figura 5.19, este modelo possui grande capacidade preditiva mesmo em relação a um dado com grande variabilidade, como o número de crimes diários em cada cluster.

Figura 5.18. Comparativo dos últimos 6 meses da série (Continua)



Fonte: Elaborado pelo autor.

Figura 5.19. Comparativo dos últimos 6 meses da série (Continuação)

Fonte: Elaborado pelo autor.

5.2.3 DNN

Os modelos de redes neurais profundas apresentaram resultados inferiores aos modelos antes descritos, como demonstrado na tabela 5.10. O coeficiente de determinação indica um potencial explicativo menor das variáveis independentes, especialmente para os modelos com mais clusters. O R^2 do modelo que utilizou 6 clusters é 2,15 vezes superior ao modelo que utilizou 84 clusters.

Tabela 5.10. Métricas do modelo DNN prevendo Crime

Nº de Clusters	MAE	MSE	RMSE	RAE	RRSE	MdAPE	NRMSE	R^2
6	3.03	15.92	3.99	0.50	0.53	0.25	0.37	0.71
12	2.46	11.64	3.41	0.59	0.60	0.35	0.52	0.61
24	1.43	3.70	1.92	0.73	0.74	0.48	0.71	0.43
42	1.00	1.86	1.36	0.74	0.75	0.63	0.88	0.42
84	0.68	0.90	0.95	0.80	0.82	—	1.23	0.33

Fonte: Elaborado pelo autor.

Percebe-se que o erro médio absoluto indica um erro de mais de três crimes diários por cluster no modelo que utiliza seis clusters. Esse erro chega a quase quatro crimes diários quando analisado o RMSE. Ao normalizar o RMSE pela média dos

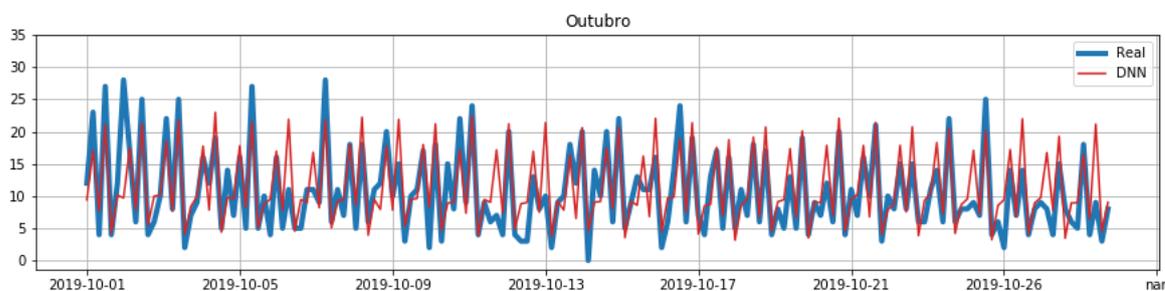
eventos criminosos é possível perceber que os erros do modelo representam um percentual considerável da média de eventos diários por cluster. Mesmo o modelo que usa apenas 6 clusters obteve um NRMSE de 0,37, muito mais alto do que o obtido nos modelos anteriormente discutidos. Já o algoritmo que utilizou 84 clusters obteve um RMSE de 123% da média dos eventos diários.

O MdAPE indica que o modelo com 6 clusters apresenta uma mediana dos erros absolutos percentuais de 25%, evidenciando que metade da amostra erra por, pelo menos, 1/4 do valor dos crimes diários em cada cluster. Já o modelo com 42 clusters essa mediana percentual atinge 63%, um valor consideravelmente alto. Os erros do modelo com 84 clusters foram tão elevados que o MdAPE não conseguiu encontrar a mediana dos erros absolutos percentuais em função da dimensão dos valores neste modelo.

Estes modelos também apresentam um resultado pouco satisfatório em relação a um modelo simples que usa a média como a previsão para os crimes diários. Como evidenciado no RAE e no RRSE, todos os modelos apresentam valores mais próximos a 1 do que a zero, estando assim mais próximos do modelo de simples aplicação da média do que de modelos preditivos mais eficazes.

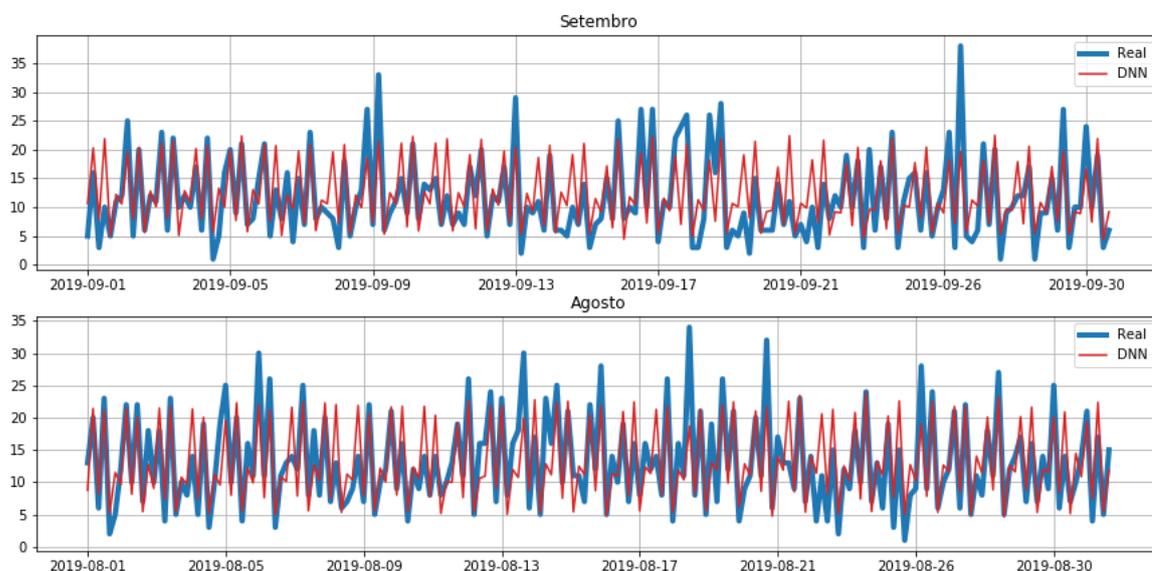
Na figura 5.20 e figura 5.21, apresentada a seguir, é demonstrado o potencial preditivo do modelo DNN que utiliza 6 clusters para os últimos 6 meses da série analisada. Neste caso, não foi escolhido um cluster em específico para apresentar, o gráfico demonstra a previsão do modelo para todos os 6 clusters em conjunto.

Figura 5.20. Comparativo diário nos últimos três mês da série (continua)



Fonte: Elaborado pelo autor.

Figura 5.21. Comparativo diário nos últimos três mês da série (continuação)



Fonte: Elaborado pelo autor.

Os modelos de redes neurais artificiais utilizados parecem obter um resultado superior em relação à previsão dos eventos criminosos. Ao serem aplicados para prever o risco de violência diário em cada cluster o desempenho é bastante superior, conforme demonstrado na tabela 5.11:

Tabela 5.11. Métricas do modelo DNN prevendo Risco

Nº de Clusters	MAE	MSE	RMSE	RAE	RRSE	MdAPE	NRMSE	R ²
6	0.07	0.07	0.26	0.08	0.27	0.00	0.18	0.98
12	0.01	0.01	0.11	0.02	0.16	0.00	0.08	0.96
24	0.01	0.01	0.09	0.01	0.12	0.00	0.06	0.95
42	0.02	0.02	0.14	0.03	0.17	0.00	0.10	0.97
84	0.00	0.00	0.02	0.00	0.04	0.00	0.02	0.99

Fonte: Elaborado pelo autor.

O coeficiente de determinação (R²) mostra que a capacidade de explicação das variáveis independentes são superiores para esta tarefa de prever o risco, nenhum dos modelos de DNN apresentou um R² inferior a 0,95. É interessante notar que, estes modelos construídos de redes neurais artificiais possuem uma performance superior utilizando muitos clusters, como percebe-se ao analisar as métricas do modelo que utilizou 84 clusters.

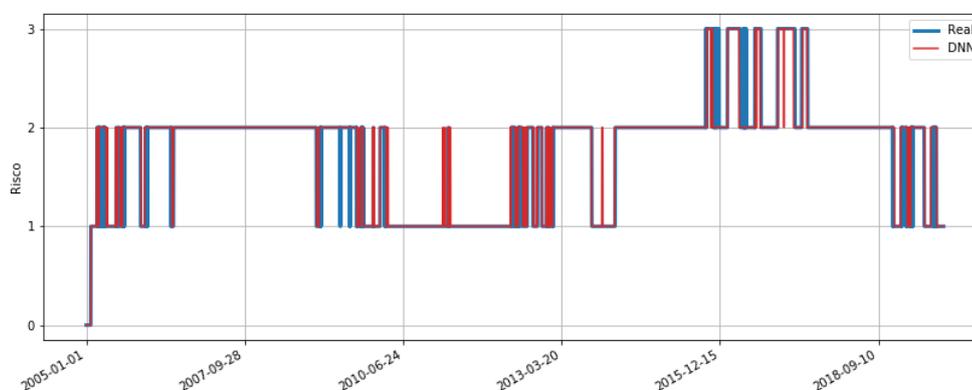
Verificando o NRMSE percebe-se que a raiz do erro médio quadrático representa apenas 2% do risco médio diário de cada cluster, um erro bem baixo. Já o modelo que emprega 6 clusters possui um NRMSE muito superior, de 0,18, indicando um erro proporcional 9 vezes superior em relação ao modelo com mais clusters.

Não foi possível medir o MdAPE em função da característica dos erros percentuais deste modelo. Um problema que se mostra recorrente desta técnica.

Ao contrário do desempenho dos modelos de redes neurais artificiais para prever os crimes, estes modelos ao preverem o risco de criminalidade obtêm um desempenho muito superior a simples aplicação da média como um atributo predictor. É o que denota o RAE, onde todos os modelos apresentaram valores muito próximos de zero. Se é interessante penalizar os erros mais discrepantes da média então o RRSE mostra que o modelo com 84 clusters tem um desempenho muito superior em relação a simples aplicação da média do risco diário em cada cluster.

Abaixo, na figura 5.22, apresenta-se a comparação entre o risco real de um determinado cluster ao longo da série e o risco previsto para este cluster. O modelo utilizado para demonstrar, ao exemplo dos anteriormente apresentados, é o que utiliza 6 clusters. Como antes destacado, este algoritmo tem um desempenho consideravelmente inferior ao que emprega 84 clusters, porém opta-se por seguir a mesma forma de apresentação dos modelos anteriores.

Figura 5.22. Risco ao longo da série para cluster



Fonte: Elaborado pelo autor.

5.2.4 LSTM

O último modelo a ser analisado é o modelo de redes neurais artificiais que utilizam células de memória de curso e longo prazo para verificar quais informações são relevantes a ponto de serem transmitidas para os próximos neurônios e quais são irrelevantes a ponto de serem esquecidas, e assim não perpetrar na rede neural uma informação irrelevante para o modelo. A tabela 5.12, exibida a seguir traz as métricas para o modelo LSTM.

Tabela 5.12. Métricas do modelo LSTM prevendo Crime

Nº de Clusters	MAE	MSE	RMSE	RAE	RRSE	MdAPE	NRMSE	R ²
6	3.23	19.59	4.43	0.55	0.59	0.28	0.40	0.70
12	2.79	16.72	4.08	0.66	0.71	0.39	0.63	0.63
24	1.42	3.79	1.94	0.72	0.75	0.49	0.72	0.46
42	1.02	2.09	1.45	0.76	0.80	0.69	0.94	0.36
84	0.74	1.05	1.03	0.87	0.88	—	1.32	0.26

Fonte: Elaborado pelo autor.

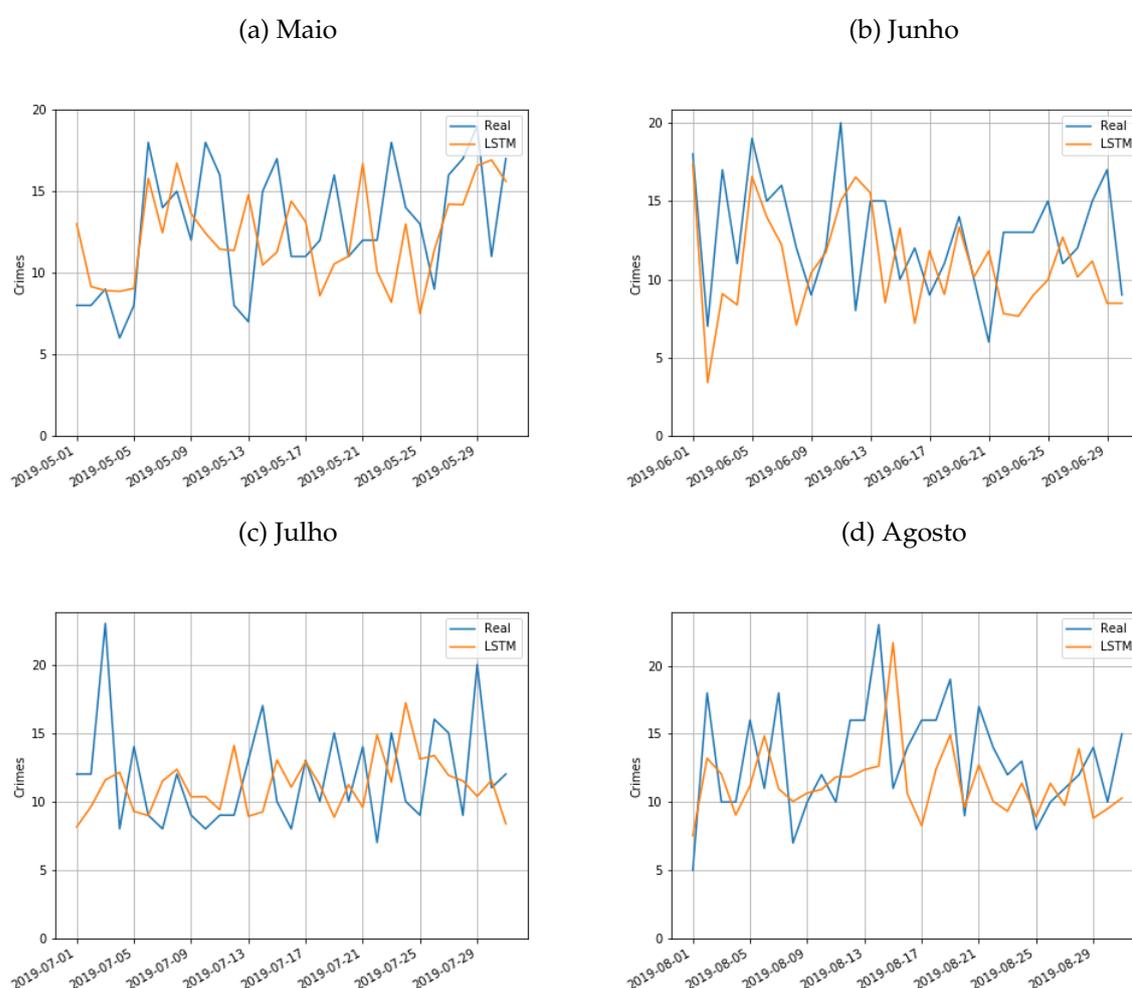
Ao verificar o coeficiente de determinação percebe-se que estes modelos, a exemplo dos algoritmos DNN, não possuem um grande potencial de explicação das variáveis pevisoras, pelo menos quanto à previsão dos crimes diários em cada cluster. O erro médio absoluto (MAE) mostra que os modelos LSTM erram a previsão em uma grandeza de pouco mais de 3 crimes diários por cluster, para o algoritmo com 6 clusters, e 0,74 crimes diários, para o modelo que utiliza 84 clusters. Já o RMSE varia entre 1,03 e 4,43.

Essa discrepância no RMSE pode ser melhor compreendida utilizando a normalização apresentada pelo NRMSE, nesse caso percebe-se que o modelo com 6 clusters apresenta um RMSE que representa 40% da média dos crimes diários por cluster, um valor que pode ser entendido como alto, e este indicador aumenta conforme novos clusters vão sendo introduzidos.

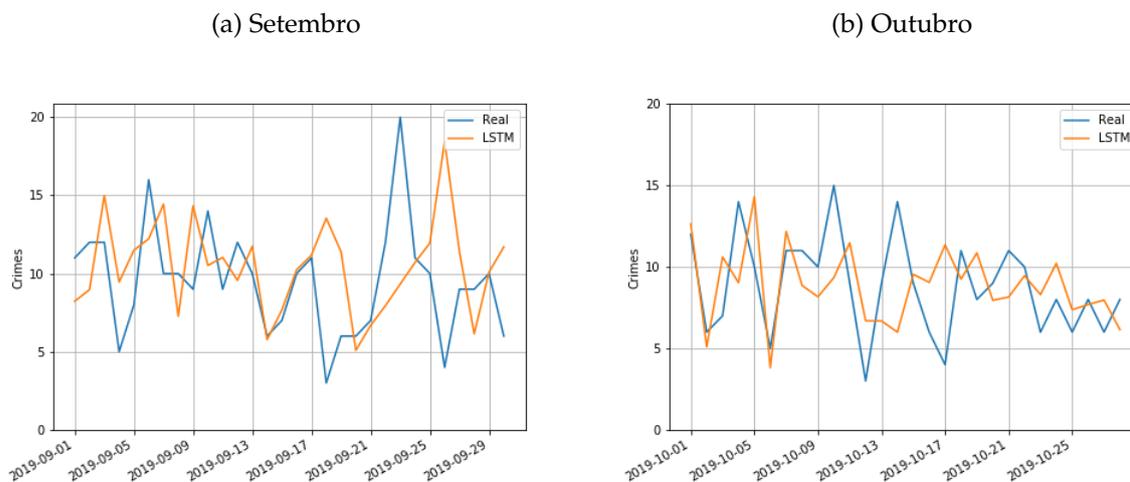
Aparentemente os modelos LSTM oferecerem erros percentuais mais adequados à aplicação do MdAPE, visto que para a maioria dos modelos ele apresenta resultado interpretável. O modelo que denota a menor mediana é aquele que utiliza menos clus-

ters, onde o MdAPE denota um erro mediano de 28%, um resultado que parece pouco interessante. Essa capacidade preditiva inferior dos modelos LSTM nesta aplicação fica mais evidente quando verificados o RAE e o RRSE, ambos apresentam valores próximos a 1, mostrando que o algoritmo não se diferencia muito de um modelo simples de utilização da média como atributo predictor. A seguir, na figura 5.23 e figura 5.24, como nos modelos anteriores, é apresentado o desempenho do modelo que utiliza 6 clusters ao prever os últimos seis meses da série.

Figura 5.23. Comparativo dos últimos 6 meses da série (Continua)



Fonte: Elaborado pelo autor

Figura 5.24. Comparativo dos últimos 6 meses da série (continuação)

Fonte: Elaborado pelo autor.

Para encerrar a análise individual de cada tipo de modelo utilizado será verificado o potencial preditivo dos modelos LSTM para prever o atributo risco. As métricas são apresentadas na tabela 5.13:

Tabela 5.13. Métricas do modelo LSTM prevendo Risco

Nº de Clusters	MAE	MSE	RMSE	RAE	RRSE	MdAPE	NRMSE	R ²
6	0.20	0.12	0.35	0.25	0.37	0.10	0.10	0.94
12	0.16	0.10	0.32	0.27	0.46	0.04	0.22	0.80
24	0.20	0.10	0.31	0.30	0.38	0.07	0.21	0.86
42	0.34	0.22	0.47	0.49	0.58	0.20	0.33	0.69
84	0.33	0.21	0.46	0.50	0.61	0.16	0.31	0.59

Fonte: Elaborado pelo autor.

Ao contrário do modelo de redes neurais artificiais analisado anteriormente, os modelos LSTM não tem um desempenho tão bom quando levado em consideração o coeficiente de determinação R². Nesse caso apenas o modelo com 6 clusters possui um R² superior a 0,90, todos os demais apresentam coeficientes inferiores. O menor deles é observado no modelo que utiliza 84 clusters, quando o R² alcança 0,59.

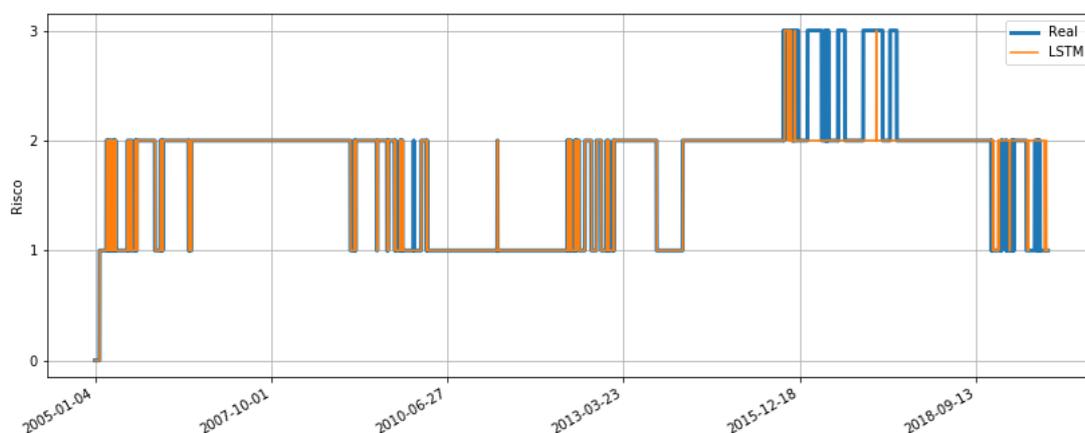
Tanto o MAE quanto o RMSE apresentam valores elevados para a análise do risco, muito superiores aos modelos antes demonstrados. Normalizando o RMSE, verifica-se que o modelo com 6 clusters tem um erro menor em relação a média do

risco diário em cada cluster, apresentando uma proporção de 10% deste. Já o MdAPE mostra que o modelo que utiliza 12 clusters possui a menor mediana dos erros, 0,04, indicando que metade dos erros deste modelo são inferiores a este valor.

Mesmo assim, através da análise do RAE e do RRSE, percebe-se que estes modelos não estão tão afastados da simples aplicação da média como forma de prever o nível de risco diário dos clusters. O modelo que usa menos clusters novamente denota um resultado superior aos demais.

Na figura 5.25 demonstra-se o desempenho do modelo LSTM para prever o nível de risco em um cluster ao longo de toda a série. Pode-se perceber claramente a ineficácia deste modelo que se mostra evidente mais ao final da série. Novamente o algoritmo escolhido foi o que compreende 6 clusters.

Figura 5.25. Risco ao longo da série para cluster



Fonte: Elaborado pelo autor.

5.2.5 Comparativo entre os modelos

Agora será apresentada a comparação dos algoritmos utilizados nas previsões. São demonstrados os modelos de previsão para 6 clusters já que estes apresentam um desempenho melhor levando em consideração a métrica R^2 . Na tabela 5.14, abaixo, é possível verificar que o modelo de classificação tem um desempenho significativa-

mente melhor que os demais para prever os crimes diários em cada cluster.

Tabela 5.14. Métricas para Crime - Modelos com 6 Clusters

Modelo	MAE	MSE	RMSE	RAE	RRSE	MdAPE	NRMSE	R²
Regressão	1.76	6.53	2.55	0.30	0.34	0.14	0.23	0.87
Classificação	0.43	2.84	1.68	0.07	0.22	0.00	0.16	0.94
DNN	3.03	15.92	3.99	0.50	0.53	0.25	0.37	0.71
LSTM	3.23	19.59	4.43	0.55	0.59	0.28	0.40	0.70

Fonte: Elaborado pelo autor.

O erro médio absoluto (MAE) indica um desempenho muito superior do modelo de classificação, mostrando que este modelo erra em média por menos de 1 crime (0,43) a previsão em relação ao dado real. Já o modelo de regressão, erra por quase 2 crimes em cada dia e os modelos DNN e LSTM erram por mais de 3 crimes por dia em cada um dos 6 clusters. Isso mostra um desempenho 75,56% melhor do modelo de classificação em relação ao modelo de regressão, 85,80% melhor que o modelo DNN e 86,68% melhor que o modelo LSTM.

É interessante penalizar os erros mais dispersos de cada modelo, já que uma previsão que está muito distante do dado real pode afetar significativamente a estratégia de gestão dos recursos usados a fim de inibir a criminalidade. Dessa forma, analisando a raiz o erro médio quadrático (RMSE) verificamos que o erro do modelo de classificação é 34,11% menor que o modelo que utiliza regressão, 57,89% menor que o modelo DNN e 62,07% menor que o modelo LSTM. Ainda um resultado muito significativo do modelo de classificação.

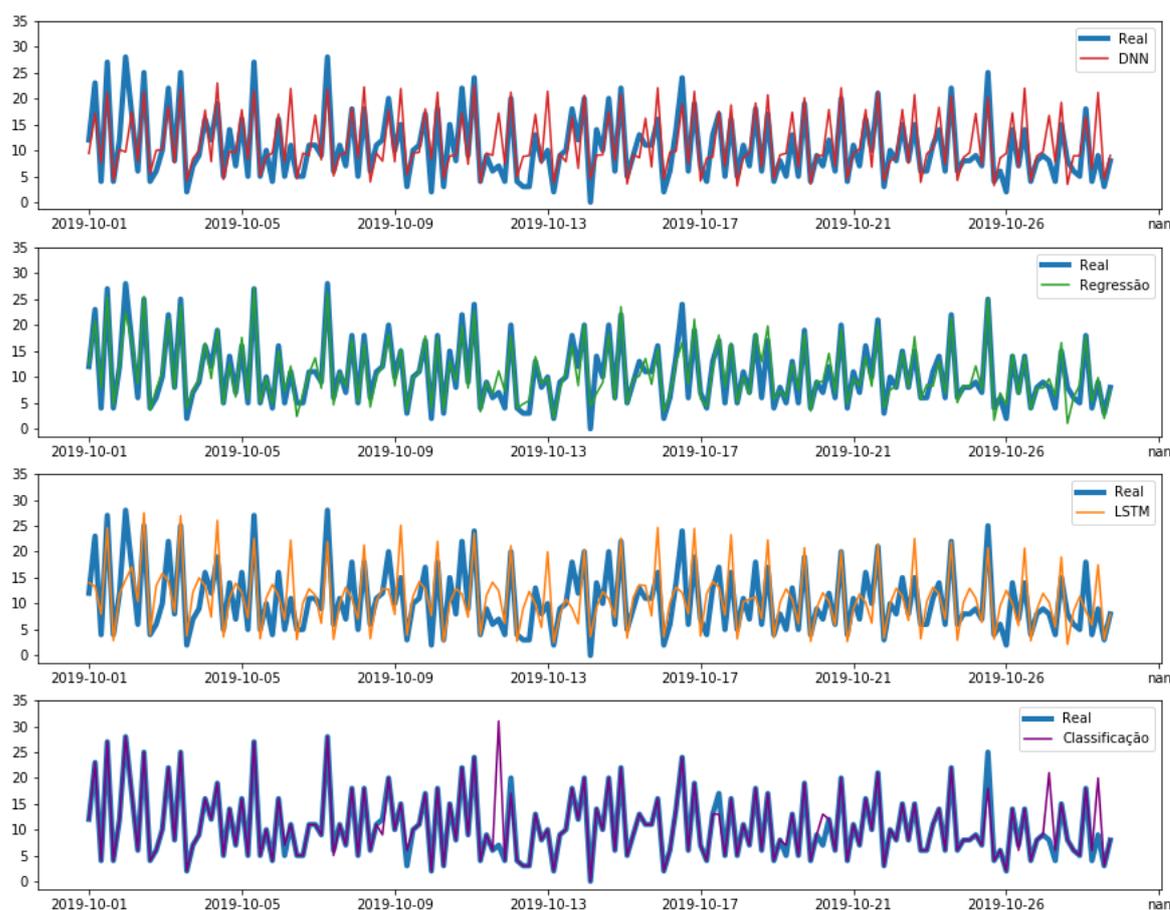
A normalização do RMSE confirma esse desempenho do modelo de classificação para prever os eventos criminosos diários em cada cluster. Neste caso, o NRMSE aponta que o RMSE representa 16% da média dos crimes cometidos diariamente em cada cluster, um resultado 60% menor que o modelo com o menor potencial de previsão dos crimes absolutos, o modelo LSTM.

O modelo de classificação também possui um erro absoluto relativo (RAE) inferior aos demais, mostrando que é o modelo que melhor se distingue de modelos básicos de previsão. Da mesma forma, este foi o modelo que apresentou o menor RRSE,

evidenciando que este modelo também tem uma dispersão dos erros menor que os demais.

O coeficiente de determinação denota mais claramente essa superioridade do modelo de classificação, indicando um R^2 de 0,94 para este. Como demonstra a tabela, os modelos LSTM e DNN obtiveram um R^2 de 0,71 e 0,70, respectivamente. Já o modelo de regressão não ficou muito abaixo do modelo de classificação neste quesito, mesmo tendo obtido um R^2 inferior a 0,90. Na figura 5.26, podemos comparar o desempenho de cada modelo prevendo a criminalidade diária para cada um dos 6 clusters que a cidade foi segmentada. Neste caso, os 6 clusters são apresentados graficamente em conjunto e não separadamente. O período utilizado foi o último mês da série de dados.

Figura 5.26. Comparativo diário no último mês da série



Fonte: Elaborado pelo autor.

Pode-se perceber na figura anterior que o modelo de classificação realmente erra muito pouco em relação aos demais, como destacado pelas métricas anteriormente apresentadas. O modelo de regressão também mostra boa previsibilidade dos eventos, já os modelos DNN e LSTM demonstram compreender a evolução da criminalidade ao longo do mês, porém com uma precisão muito menor em relação ao número de crimes absoluto diários em cada cluster.

Por fim, passa-se a comparar o desempenho dos modelos com relação ao atributo risco. A tabela 5.15 mostra que, neste caso, o modelo de classificação possui destaque em relação aos demais, porém com uma ênfase menor, haja vista que tanto o modelo de regressão quanto o modelo DNN apresentam coeficiente de determinação similar. Neste caso, o modelo LSTM se mostra menos preciso em relação aos demais.

Tabela 5.15. Métricas para Risco - Modelos com 6 Clusters

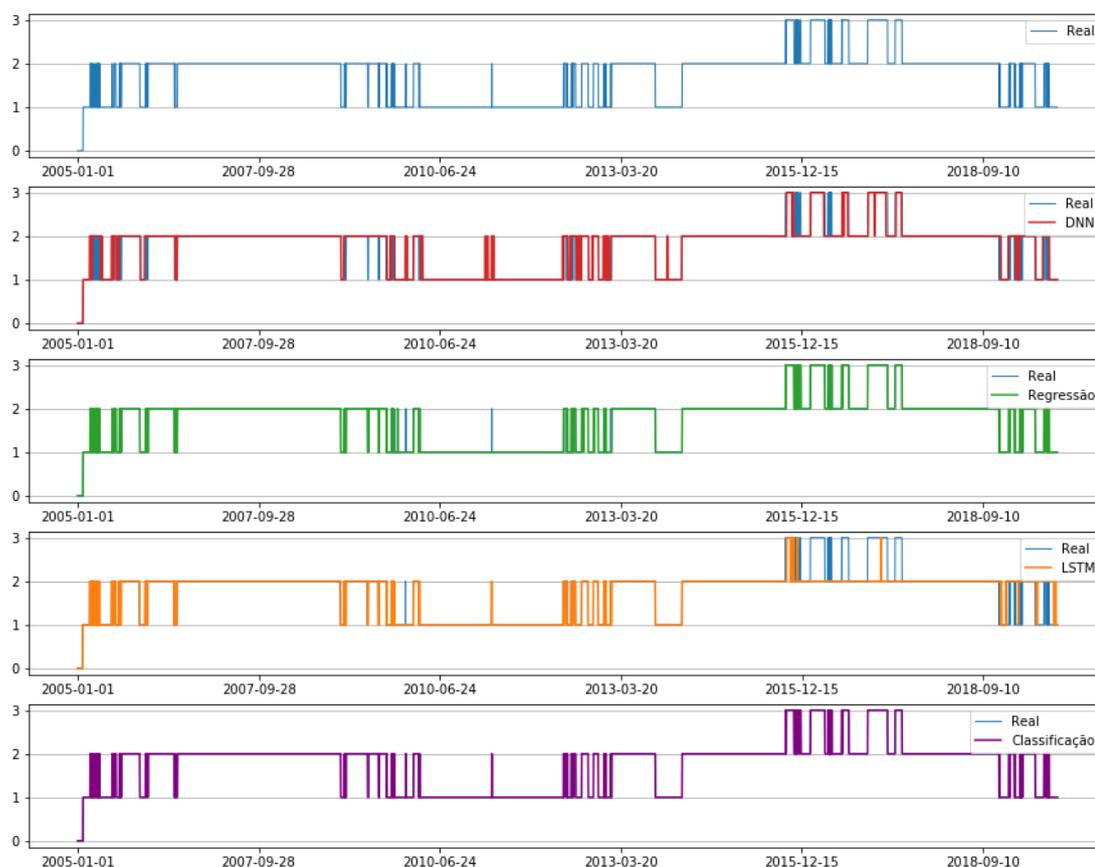
Modelo	MAE	MSE	RMSE	RAE	RRSE	MdAPE	NRMSE	R²
Regressão	0.05	0.01	0.10	0.06	0.10	0.03	0.06	0.97
Classificação	0.01	0.01	0.07	0.01	0.08	0.00	0.05	0.98
DNN	0.07	0.07	0.26	0.08	0.27	0.00	0.18	0.98
LSTM	0.20	0.12	0.35	0.25	0.37	0.10	0.10	0.94

Fonte: Elaborado pelo autor.

O erro absoluto médio (MAE) denota bem a superioridade do modelo de classificação, o erro de 0,01 representa um desempenho 95% melhor que o obtido pelo modelo LSTM (0,20). Mesmo analisando o RMSE, o modelo de classificação ainda apresenta um MAE 80% menor. Verificando a forma normalizada do RMSE, a métrica NRMSE destaca a superioridade dos modelos de classificação e regressão em relação aos demais.

O erro absoluto relativo (RAE) e a raiz quadrada do erro quadrático relativo (RRSE) também corroboram essa percepção. Nesse caso, os modelos de classificação e regressão têm destaque da mesma forma, já que seus valores se aproximam muito de zero, o que indicaria um algoritmo extremamente preciso. De fato, o erro do modelo de classificação neste quesito é tão pequeno que foi o único dos modelos aqui destacados onde a mediana dos erros absolutos percentuais (MdAPE) não foi capaz de ser identificada em função do seu valor pouco representativo.

Figura 5.27. Comparativo do risco ao longo da série



Fonte:

Fonte: Elaborado pelo autor.

Na figura 5.27 está demonstrado a evolução do atributo risco para um mesmo cluster específico da série com 6 clusters ao longo de todo o período analisado. Percebe-se que em relação ao risco real, os algoritmos que utilizaram modelos de classificação e regressão conseguem prever muito bem a evolução do atributo ao longo da série. Já os algoritmos que utilizaram modelos DNN e LSTM, deixam evidentes as previsões que não se concretizaram ao longo do período analisado.

Neste caso, como foi utilizada a série de dados completa para classificar o risco diário de violência em cada cluster, pode-se pensar ser mais fácil identificar as ondas de criminalidade em função disso, já que a média de crimes anuais aumenta muito nos últimos anos em relação ao período inicial. Porém, ao aplicar os mesmos modelos para classificar o risco utilizando um recorte qualquer da série temporal, o desempenho dos modelos segue o demonstrado anteriormente, por isso não há qualquer prejuízo em apresentar a previsão do risco para toda a série de dados.

Após demonstrados os resultados obtidos na análise exploratória dos dados e dos modelos preditivos, a seguir são discutidos estes resultados e indicadas as potencialidades desses modelos. Após isso, a conclusão final deste estudo é apresentada.

5.3 DISCUSSÃO

Como demonstrado anteriormente, os modelos preditivos têm condições de auxiliar os gestores na tomada de decisão sobre a melhor alocação da força policial no espaço geográfico da cidade. O desempenho dos modelos indica que é possível se antecipar à criminalidade, o que pode trazer um ganho significativo no seu combate.

Se aplicados de forma contínua, os modelos aqui demonstrados podem ser utilizados como ferramenta para impulsionar a utilidade do gasto público em segurança e melhorar a eficácia do trabalho da força policial. A inteligência artificial é uma grande aliada das organizações e colabora para que essas possam tomar melhores decisões. Logo, com o Estado não poderia ser diferente.

A formulação de estratégias de enfrentamento da criminalidade precisa incorporar o desenvolvimento e a utilização de novas ferramentas, como estas expostas aqui. Dessa forma, o Estado poderá conduzir suas políticas no sentido da otimização dos recursos, respeitando os recursos dos cidadãos e entregando um serviço mais eficaz.

O estudo aqui apresentado limitou-se a análise de apenas um município do estado do Rio Grande do Sul, porém é possível entender que a sua aplicação pode ser estendida para todas as regiões do estado. Os modelos podem ser construídos para analisar individualmente as microregiões do estado, auxiliando na tomada de decisão sobre a necessidade de deslocamento da força policial dentro do estado.

Estes modelos podem colaborar para identificar batalhões da Brigada Militar que necessitarão de mais recursos no futuro próximo, identificando aqueles que precisam de uma atenção urgente em relação aos demais. Também podem indicar à Polícia Civil a previsão de atividade de um determinado tipo de crime particular, fazendo com que o órgão consiga traçar operações estratégicas de combate a crimes específi-

cos. Ao Instituto Geral de Perícias estas abordagens podem fornecer auxílio na melhor estruturação das localidades atendidas por cada regional e na necessidade de recursos a serem viabilizados no curto prazo. Já atendendo à Superintendência dos Serviços Penitenciários, estes modelos podem indicar a necessidade de liberação de vagas no sistema prisional de cada região e auxiliar o órgão na tomada de decisões que envolvam a alocação dos recursos distribuídos pelo estado todo.

Haja vista o que foi demonstrado aqui, é preciso mencionar o potencial da estatística computacional como ferramenta para a melhor estruturação do Estado. Não apenas grandes empresas possuem departamentos voltados unicamente para a construção de modelos preditivos, mas pequenas e médias empresas já contratam consultorias a fim de compreender melhor o seu negócio e se antecipar a tendências do mercado que não são perceptíveis sem a utilização de técnicas computacionais complexas.

Se empresas de todos os ramos utilizam este tipo de estratégia com sucesso, por que não deveria o Estado também utilizar? O que foi demonstrado nesta pesquisa reforça a urgente necessidade de estruturação de projetos computacionais em todos os poderes do Estado. A sociedade espera que o poder público consiga desempenhar suas atribuições de forma competente na alocação de recursos e eficaz na obtenção dos resultados. A obtenção de resultados que maximizam o interesse público, parecem estar relacionados a união entre a experiência humana e a inteligência artificial.

Capítulo 6

Conclusão

A importância desta pesquisa está relacionada à urgência com que se busca uma resposta para o enfrentamento da criminalidade no Brasil. Mesmo apresentando os índices antes demonstrados, não possui publicações a respeito da colaboração da inteligência artificial para o combate à criminalidade. Ao mesmo tempo a capacidade computacional aumenta a cada dia de forma que os modelos podem demonstrar desempenho significativamente superior com o passar do tempo.

Como foi demonstrado, os modelos preditivos apresentam bom desempenho para prever tanto o número de crimes que ocorrerão nos dias seguintes nas mais diversas regiões da cidade, como desempenham uma excelente eficácia quanto a previsão do risco de criminalidade em cada região da cidade. Também verifica-se que modelos que apresentam menos clusters demonstraram compreender melhor os dados e obtiveram os melhores resultados. Com exceção do modelo de redes neurais profundas (DNN) que obteve desempenho melhor com 84 clusters do que com um número de clusters menor, todos os demais modelos tiveram desempenho superior com um número de clusters reduzido.

Foi atingido o objetivo de construir modelos que tenham capacidade de prever a criminalidade de forma eficaz, superior a simples análise de uma média da criminalidade ou a continuidade da tendência indicada pelo período anterior. Todos os modelos aqui apresentados têm desempenho superior a estas metodologias, sem exceção, porém com destaque para o modelo de classificação, que apresentou os melhores

resultados.

Este estudo mostra o potencial desta metodologia para gestão da segurança pública do estado do Rio Grande do Sul. A atual crise vivida pelos estados só fortalece a necessidade de estudos a fim otimizar a eficiência das atividades de Estado, principalmente em relação a problemas tão críticos como este. Já que os recursos são escassos, é preciso medir cada passo dado, ainda mais em se tratando de uma área da magnitude da segurança pública e o impacto que uma gestão exitosa pode causar na qualidade de vida da população.

Referências Bibliográficas

- ALVES, Shirlei. **Retrato da impunidade: RS arquiva metade dos processos de homicídios** Zero Hora, Porto Alegre, 26 nov. 2016. Disponível em: <<https://gauchazh.clicrbs.com.br/seguranca/noticia/2016/11/retrato-da-impunidade-rs-arquiva-metade-dos-processos-de-homicidios-8489962.html>>.
- BALBINOTTO NETO, G.; SCHUCH, R. C.; UHR, D.; UHR, J. Z. Crimes against Property and Public Security Spending: A Dynamic Panel Data Analysis for the Metropolitan Area of Porto Alegre. *The Empirical Economics Letters*, v. 17, p. pp. 611–618, 2018.
- BECKER, G. S. Crime and Punishment: An Economic Approach. *Journal of Political Economy*, v. 76, n. 2, p. 139–146, 1968.
- BERK, R. A.; BLEICH, J. Statistical Procedures for Forecasting Criminal Behavior. *Criminology & Public Policy*, [s.l.], v. 12, n. 3, p. 513–544, 2013.
- BISHOP, C. M. *Pattern Reconigtion and Machine Learning*. Singapore: Springer, 2006. 761 p. ISBN 9780387310732.
- BLACKMORE, F. L. E. A panel data analysis of crime in South Africa. *I South African Journal of Economic and Management Sciences*, v. 6, n. 3, p. 439–558, 2003.
- CARR-HILL, R.; STERN, N. **An Econometric Model of the Supply and Control of Recorded Offenses in England and Wales**. *Journal of Public Economics*, 1973.
- CORCORAN, I. D. W. J. J.; WARE, J. A. Predicting the geo-temporal variations of crime and disorder. *International Journal of Forecasting*, v. 19, p. 623–634, 2003.
- DAHUR, K.; MUSCARELLO, T. Classification system for serial criminal patterns. *Artificial Intelligence and Law*, v. 11, n. 4, p. 251–269, 2003. ISSN 09248463.
- DOWNS WILLIAM M. LANDES, S. R. A.; HOFFMAN, R. B. **Round Table on Allocation of Resources in Law Enforcement**. *The American Economic Review*, v. 59, n. 2, p. pp. 504–512, 1969.
- EHRlich, I. Participation in Illegitimate Activities: A Theoretical and Empirical Investigation. *Journal of Political Economy*, v. 81, n. 3, p. pp. 521–565, 1973.

- FELSON, M.; POULSEN, E. Simple indicators of crime by time of day. *International Journal of Forecasting*, v. 19, p. 623–634, 2003.
- FRIEDMAN, M. The Methodology of Positive Economics. *Essays In Positive Economics*, n. 1, p. pp. 3–16, 30–43, 1966.
- GOODFELLOW, I.; BENGIO, Y.; COURVILLE, A. *Deep learning*. [S.l.]: MIT press, 2016.
- GORR, W.; HARRIES, R. Introduction to crime forecasting. *International Journal of Forecasting*, v. 19, p. 551–555, 2003.
- GRAVES, A. Generating sequences with Recurrent Neural Networks. *Technical Report*, 2013.
- GRAVES, A.; JAITLY, N. Towards end-to-end Speech Recognition with Recurrent Neural Networks. *ICML*, 2014.
- GREFF, K. e. a. LSTM: A Search Space Odyssey. *IEEE Transactions on Neural Networks and Learning Systems*, v. 28, n. 10, p. pp. 2222–2232, 2015.
- HANN, R. G. Crime and the Cost of Crime: An Economic Approach. *Journal of Research in Crime and Delinquency*, p. pp. 12–30, 1972.
- HOCHREITER, S.; SCHMIDHUBER, J. Long short-term memory. *Neural Computation*, v. 9, n. 8, p. pp. 1735–1780, 1997.
- IPEA, I. de P. E. A. Atlas da Violência 2019. *IPEA - Instituto de pesquisa econômica aplicada*, n. 1, p. 21, 2020. Disponível em: <<https://www.ipea.gov.br/atlasviolencia/download/19/atlas-da-violencia-2019>>.
- KIANMEHR, K.; ALHAJJ, R. Effectiveness of support vector machine for crime hot-spots prediction. *Applied Artificial Intelligence*, v. 22, n. 5, p. 433–458, 2008. ISSN 08839514.
- LIAO, R.; WANG, X.; LI, L.; QINH, Z. A novel serial crime prediction model based on Bayesian learning theory. *2010 International Conference on Machine Learning and Cybernetics*, v. 4, n. July, p. 1757–1762, 2010.
- MCCLENDON, L.; MEGHANATHAN, N. Using Machine Learning Algorithms to Analyze Crime Data. *2018 Fifth International Conference on Internet of Things: Systems, Management and Security*, v. 2, n. 1, p. 1–12, 2015. ISSN 23940840.
- MITCHELL, T. M. *MACHINE LEARNING*. [S.l.: s.n.], 1997. ISSN 03600300. ISBN 0070428077.
- MORI, A. R. e. A. T. M. *Foundations of Machine Learning*. London: The MIT Press, 1967. v. 20. 414 p.
- MURPHY, K. P. *Machine Learning A Probabilistic Perspective*. London: The MIT Press, 2012. 1067 p. ISBN 9780262018029.

- SHERMAN, L. W.; GARTIN, P. R.; BUERGER, M. E. Hot Spots of Predatory Crime: Routine Activities and the Criminology of Place. *Criminology*, v. 27, n. 1, p. 27–56, 1989. ISSN 17459125.
- SHORT, M. B.; D'ORSOGNA, M. R.; PASOUR, V. B.; TITA, G. E.; BRANTINGHAM, P. J.; BERTOZZI, A. L.; CHAYES, L. B. A statistical model of criminal behavior. *Mathematical Models and Methods in Applied Sciences*, n. supp01, p. 1249–1267. ISSN 0218-2025.
- SINESP. **Sistema Nacional de Informação de Segurança Pública**. 2018. Disponível em: <<https://www.justica.gov.br/sua-seguranca/seguranca-publica/sinesp-1>>.
- SPECHT, D. F. A General Regression Neural Network. *IEEE Transactions on Neural Networks*, v. 2, n. 6, p. 568–576, 1991.
- STALIDIS, P.; SEMERTZIDIS, T.; DARAS, P. **Examining Deep Learning Architectures for Crime Classification and Prediction**. p. 1–12, 2018. Disponível em: <<http://arxiv.org/abs/1812.00602>>.
- SUL, R. G. D. **Indicadores Criminais**. 2019. Disponível em: <<https://ssp.rs.gov.br/indicadores-criminais>>.
- SUL, R. G. D. **RS SEGURO**. 2019. Disponível em: <<https://ssp.rs.gov.br/rs-seguro>>.
- SUL, R. G. D. **Secretaria da Fazenda do Estado do RS**. 2019. Disponível em: <<<https://tesouro.fazenda.rs.gov.br/lista/3759/despesa-de-pessoal>>>.
- TOBLER, W. R. A Computer Movie Simulating Urban Growth in the Detroit Region. *Economic Geography*, v. 46, p. 234–240, 2011.
- TRUMBULL, C. C.; N., W. Estimating the economic model of crime with panel data. *The Review of Economics and Statistics*, n. 2, p. 360–366.