GOVERNO DO ESTADO DE MATO GROSSO DO SUL

SECRETARIA DE ESTADO DE ADMINISTRAÇÃO E DESBUROCRATIZAÇÃO

FUNDAÇÃO ESCOLA DE GOVERNO

APLICAÇÃO DE ALGORITMOS DE APRENDIZADO DE MÁQUINA PARA PREVISÃO DE CRIMES EM MATO GROSSO DO SUL

Pré-projeto apresentado ao XVII Prêmio Sul-Mato-Grossense de Inovação na Gestão Pública, ano 2022, na categoria Ideias Inovadoras Implementáveis como pré-requisito para concessão do prêmio.

Campo Grande, MS

2022

SUMÁRIO

1 Título da Ideia Inovadora Implementável.

Aplicação de Algoritmos de Aprendizado de Máquina para Previsão de Crimes em Mato Grosso do Sul.

2 Caracterização da situação anterior.

Os dados são cada vez mais importantes no mercado atual e se tornaram peça-chave para orientar na tomada de decisões e o planejamento estratégico de empresas de todos os segmentos (TAN et al., 2016). No entanto, para transformar essas informações em conhecimento e vantagens competitivas, é preciso que as organizações se tornem *data driven* (SORESCU, 2017). *Data Driven* se refere a processos organizacionais orientados a dados, ou seja, quando a empresa baseia a tomada de decisão e o planejamento estratégico na coleta e na análise de informações – e não em intuições ou simples experiências (SORESCU, 2017).

O *Data Driven* surgiu como extensão da ciência de dados, campo do conhecimento que utiliza métodos científicos e algoritmos para transformar dados – estruturados e não estruturados – em conhecimento. Atualmente, no ambiente corporativo, isso é feito por meio de ferramentas como *Big Data*, Inteligência Artificial e o Aprendizado de Máquina (*Machine Learning*), para obter *insights* a partir da coleta, cruzamento e interpretação de dados. O objetivo não é outro senão aumentar a competitividade da organização e promover melhores resultados.[1]

Toda corporação, tanto empresarial, quanto governamental, planeja suas ações a partir de dados qualitativos ou quantitativos gerados no dia-a-dia. Os dados coletados devem ilustrar as experiências vividas e demonstrar os erros e acertos cometidos diariamente. O planejamento de ações vem aumentando consideravelmente e sendo usadas exaustivamente no planejamento corporativo através do uso de sistemas computacionais e técnicas de análise de dados (HAN et al., 2011).

Os modelos de previsão quantitativos utilizam basicamente dados históricos para detectar padrões de comportamento e estimá-los no futuro. Tais modelos empregam técnicas computacionais e estatísticas para representar e executar ações para os quais foram criados. Assim, a aquisição de ferramentas deste tipo deve ser encarada como um diferencial organizacional, pois adicionará suporte a decisões a serem tomadas pelos gestores. Diversas áreas estão utilizando previsão para o suporte de decisões realizadas por seus gestores, por

[1] https://blog.neoway.com.br/data-driven/

exemplo o uso da previsão de preços de ações no mercado imobiliário, pontuação de jogos de futebol, tempo previsto para acontecer um novo ataque cardíaco em um paciente, ataque a uma rede de computadores ou um assalto a um domicílio residencial.

Técnicas de previsão para o suporte de decisões já vem sendo utilizada em diversos segmentos do estado de Mato Grosso do Sul. No Tribunal de Justiça de MS, por exemplo, técnicas de aprendizado de máquina vêm sendo empregada para a classificação de texto. [2] Do campo jurídico também surgiu startup que unifica jurisprudência de diversos tribunais e possibilita a pesquisa por meio de inteligência artificial. A plataforma foi desenvolvida com suporte da Fundect (Fundação de Apoio ao Desenvolvimento do Ensino, Ciência e Tecnologia de Mato Grosso do Sul), dentro do Programa Centelha, que investe em ideias inovadoras.[3]

Na segurança pública, os órgãos da administração pública responsáveis pela segurança da população podem se beneficiar do aprendizado de máquina para tornar suas ações de combate e prevenção ao crime mais eficientes. Ao analisar os dados sobre ocorrências (local, data e hora em que ocorreram, por exemplo), é possível encontrar padrões e agir pontualmente para reduzir o número de eventos (SILVA, 2011). Um exemplo prático é o do Departamento de Polícia da Cidade de Richmond, Estados Unidos. A cada véspera de Ano Novo, Richmond, registrava um aumento no número de queixas dos cidadãos sobre os tiros aleatórios tradicionalmente associados à véspera de Ano Novo. A polícia começou a analisar os dados coletados ao longo dos anos e, com base neles, conseguiu antecipar a hora, o local e a natureza de futuros incidentes. A polícia de Richmond colocou então policiais nesses locais para responder mais rapidamente. O resultado foi uma diminuição de 47% em queixas dos cidadãos sobre tiros aleatórios e um aumento de 246% no número de armas apreendidas. O departamento também economizou US$ 15.000 em custos com pessoal.[4]

No estado de Mato Grosso do Sul, os dados sobre ocorrências criminais são armazenados no Sistema Integrado de Gestão Operacional (SIGO). [5] O estado desde 2005 vem sendo considerado pioneiro na atuação contra a criminalidade após padronizar em suas instituições de segurança. [6] O sistema SIGO é inclusive referência para outros países, até mesmo

[2] https://www.tjms.jus.br/noticia/61610/
[3] https://www.campograndenews.com.br/politica/pre-candidatos-mostram-como-inserir-inteligencia-artificial-no-servico-publico
[4] https://policeandsecuritynews.com
[5] http://www.sigo.ms.gov.br/
[6] https://www.capitalnews.com.br/reportagem-especial/sigo-ms-e-pioneiro-no-trabalho-integrado-entre-as-forcas-de-seguranca/202558

para os Estados Unidos, por ter sido o primeiro estado no Brasil a desenvolver um sistema integrado que concentrasse informações em um único banco de dados de todas as forças de segurança pública.[7] Porém, estes ainda não veem sendo explorados para auxiliar no entendimento de padrões de crimes, bem como no auxílio na tomada de decisões mais efetivas.

3 Descrição da Ideia Inovadora Implementável.

A ideia inovadora é aplicar técnicas de aprendizado de máquina para a previsão de crimes utilizando-se dos dados armazenados no sistema SIGO. O aprendizado de máquina (*machine learning*) é considerada uma subárea da Inteligência Artificial (IA), mantém relação com atividades realizadas na mineração de dados, porém atuando de forma ágil com a utilização de algoritmos específicos e recursos que possibilitam a elaboração de modelos, com base no treinamento aplicado aos dados. De acordo com SHAH et al. (2021), o número e as formas de atividades criminosas estão aumentando a um ritmo alarmante, obrigando as agências de segurança a desenvolver métodos eficientes para tomar medidas preventivas. No cenário atual de crime em rápido crescimento, as técnicas tradicionais de resolução de crimes são incapazes de produzir resultados, sendo lentas e menos eficientes. Assim, se pudermos encontrar maneiras de prever o crime, em detalhes, antes que ele ocorra, ou criar uma "máquina" que possa ajudar os policiais, isso aliviaria o fardo da polícia e ajudaria na prevenção de crimes. Para conseguir isso, ele sugere a inclusão de algoritmos e técnicas de aprendizado de máquina (ML).

No trabalho de KIM et al. (2018), as previsões de crimes foram investigadas com base no aprendizado de máquina. Dados de crimes dos últimos 15 anos em Vancouver (Canadá) foram analisados para previsão. Essa análise de crimes baseada em aprendizado de máquina envolve a coleta de dados, classificação de dados, identificação de padrões, previsão e visualização. *k-Nearest Neighbor (k-NN)* e algoritmos de árvore de decisão impulsionada também foram implementados para analisar o conjunto de dados de crimes. Em seu estudo, um total de 560.000 registros de crimes entre 2003 e 2018 foram analisados, e a previsão de crimes com uma acurácia entre 39% e 44% foi obtida pela previsão do crime usando algoritmos de aprendizado de máquina. A acurácia foi baixa como modelo de previsão, mas os autores concluíram que a acurácia pode ser aumentada ou melhorada ajustando os algoritmos e os dados do crime para aplicações específicas.

[7] https://www.campograndenews.com.br/cidades/sistema-integrado-de-dados-surgiu-em-ms-e-virou-referencia-ate-nos-eua/

BHARATI et al. (2018) analisou em seu trabalho um conjunto de dados composto por vários crimes e previram o tipo de crime que pode ocorrer em um futuro próximo, dependendo de várias condições. O conjunto de dados do crime consiste em informações como a descrição do local do crime, tipo de crime, data, hora e coordenadas precisas do local. Diferentes combinações de modelos, como classificação KNN, regressão logística, árvores de decisão, floresta aleatória, *support-vector machine (SVM)* e métodos bayesianos foram testadas. A classificação KNN mostrou-se a melhor com uma acurácia de aproximadamente 78%. Eles também usaram diferentes gráficos que ajudaram a entender as várias características do conjunto de dados de crimes de Chicago.

No trabalho de HOSSAIN et al. (2020) é proposto um sistema que prevê o crime analisando um conjunto de dados contendo registros de crimes cometidos anteriormente e seus padrões. O sistema proposto funciona principalmente com dois algoritmos de aprendizado de máquina: uma árvore de decisão e KNN. Técnicas como o algoritmo de floresta aleatória e *Adaptive Boosting* foram utilizadas para aumentar a acurácia do modelo de previsão. O sistema proposto foi alimentado com dados de atividades criminosas por um período de 12 anos em São Francisco, Estados Unidos. Usando métodos de subamostragem e sobreamostragem juntamente com o algoritmo de floresta aleatória, a acurácia foi aumentada para 99,16%.

Foram realizados ao longo da escrita da ideia inovadora implementável alguns estudos experimentais. Apesar da base de dados ser consideravelmente pequena para um aprendizado de máquina, já foi possível obter alguns resultados com a aplicação de alguns algoritmos. A previsão do crime de roubo foi testada na área do Anhanduizinho em Campo Grande - MS. A base de dados continha a quantidade mensal de ações realizadas pela Polícia Militar de Mato Grosso do Sul (PMMS) naquela área nos anos de 2018, 2019, 2020, 2021 e 2022. As ações eram: quantidade de pessoas abordadas, quantidade de operações policiais realizadas naquela área, quantidade de armas apreendidas, quantidade de auto de prisão em flagrante, T.C.O, recuperação de foragidos, abordagens a veículos, recuperação de veículos, quantidade de rondas e por fim, a quantidade de roubos em cada mês dos anos analisados. Utilizando algoritmos de aprendizado de máquina da biblioteca *scikit-learn* foi possível prever a quantidade de roubos naquela área. Foi possível observar também que a quantidade de prisão de adolescente masculino foragido tem um grande peso para previsão do crime de roubo naquela área, ou seja, quanto mais adolescente masculino foragido a polícia militar conseguir capturar naquela área, menos roubos ocorrerão. Para avaliação do modelo foi utilizada a métrica de avaliação chamada Raiz Quadrada do Erro Médio, RMSE (da sigla em inglês *Root Mean Squared Error*). Ela é

uma métrica de avaliação que calcula a raiz quadrática média dos erros entre valores observados (reais) e predições (hipóteses). Os algoritmos que tiveram melhores resultados foram: *Random Forest Regressor, Lasso Regression, Linear Regression e o Decision Tree Regressor*. A Figura 1 mostra os resultados alcançados com o algoritmo *Random Forest Regressor*.

Figura 1 - Resultados com o algoritmo *Random Forest Regressor*

	Model	MAE	MSE	RMSE	R2	RMSLE	MAPE
0	Random Forest Regressor	6.6335	91.7076	9.5764	0.8465	0.1541	0.1087

Fonte: Colab Notebook.

A mesma técnica de previsão poderia ser aplicada para prever outros tipos de crimes. Por exemplo, atualmente, com o advento da Lei Maria da Penha (Lei 11.340/2006)[8], que cria mecanismos para coibir a violência doméstica e familiar contra a mulher, os números de registros dessa natureza tiveram um grande aumento nas delegacias do estado, gerando uma quantidade enorme de dados. Todavia, devido ao alto número de ocorrências, torna-se inviável uma análise manual por busca de padrões. As vítimas do crime de violência doméstica possuem algumas características que são cadastradas no formulário do sistema, bem como no histórico da ocorrência. Atributos como: idade da vítima, escolaridade da vítima, quanto tempo está junto com o autor, se a vítima e o autor possuem filhos juntos, se a vítima já solicitou medidas protetivas anteriormente, ou ainda, a quantidade de registros que a vítima já registrou contra o mesmo autor.

De acordo com (BRAZ et al., 2009) os dados volumosos cadastrados nos formulários das ocorrências policiais são uma fonte riquíssima de conhecimento relativo às vítimas, aos criminosos e à natureza dos crimes, sendo armazenados, visando a recuperação de dados para análise estatística e descoberta de conhecimento.

Algoritmos de aprendizado de máquina (*machine learning*) poderiam ser aplicados na base de dados contendo esses atributos. Seria possível, por exemplo, após a etapa de extração de padrões (*data mining*), detectar o perfil das vítimas com maior chance de permanecer com o agressor e prever novos registros de boletins de ocorrências por essa vítima ou uma nova vítima com perfil semelhante (*crime prediction*). O modelo aprenderia com experiências passadas (no caso dados de registros de crimes de violência doméstica), extraindo o padrão das vítimas com

[8] http://www.planalto.gov.br/ccivil_03/_ato2004-2006/2006/lei/l11340.htm

mais ou com menos quantidade de registros anteriores, por isso o atributo **quantidade de registros anteriores** seria tão importante, pois seria o atributo de saída **(atributo alvo).** O modelo de inteligência artificial seria capaz de aprender e a prever a quantidade **(tarefa de regressão)** de boletins de ocorrência futuros que uma nova vítima irá registrar. A partir do conhecimento novo, o estado poderia aplicar políticas públicas focada às vítimas com perfil mais vulnerável, tornando possível ainda, o interrompimento do ciclo de violência doméstica.

Um dos benefícios que o conhecimento novo poderia trazer também, seria a sua utilização pelo PROMUSE - Programa Mulher Segura. O PROMUSE é um programa da Polícia Militar do Estado de Mato Grosso do Sul, instituído por meio da Portaria PMMS nº 032/2018, que faz monitoramento e proteção das mulheres em situação de violência doméstica e familiar. Policiais Militares devidamente capacitados realizam policiamento orientado com objetivo de promover o enfrentamento à violência doméstica contra mulheres, por meio de ações de prevenção, visitas técnicas, conversas com vítimas, familiares e até mesmo com os agressores, fazendo os encaminhamentos pertinentes aos órgãos da rede municipal de atendimento à mulher em situação de violência. Com o conhecimento novo, vítimas com perfil mais vulnerável poderiam ser priorizadas no atendimento, auxiliando na preservação a vida e o patrimônio por meio de políticas integradas de segurança pública. A Figura 2 ilustra um atendimento do PROMUSE a uma vítima de violência doméstica.

Figura 2 - Atendimento do Programa Mulher Segura – PROMUSE

Fonte: PMMS.

A Ideia Inovadora Implementável apresenta uma contribuição moderna na gestão pública de Mato Grosso do Sul no tocante ao **eixo social**, pois auxilia na preservação a vida e o patrimônio por meio de políticas integradas de segurança pública.

4 Objetivos propostos.

Objetivo geral:

Aplicar algoritmos de aprendizado de máquina para previsão de crimes no estado de Mato Grosso do Sul.

Objetivos específicos:

- Aplicar algoritmos de aprendizagem de máquina supervisionado para tarefa regressão ou classificação em dados de crimes registrados no estado de Mato Grosso do Sul, armazenados no sistema SIGO;
- Utilizar o conhecimento obtido como ferramenta de apoio ao processo de tomada de decisão na área de segurança pública;
- Contribuir para o debate sobre a aplicação de técnicas de descoberta de conhecimento em diferentes áreas da segurança pública.

5 Resultados esperados.

- Obtenção de conhecimentos novos a partir dos dados;
- Obtenção de modelos acurados;
- Fomentar medidas preventivas para diminuição da criminalidade no estado;
- Criar uma cultura de importância dos dados na segurança pública do estado de MS, cultura *Data Driven*.

6 Público-alvo.

O público alvo da Ideia Inovadora Implementável seriam todos os Sul-mato-grossenses, pois a Polícia Militar está presente nos 79 (setenta e nove) municípios de MS. A população contaria com um policiamento mais efetivo e inteligente.

7 Ações e etapas da implementação.

Para aplicação das técnicas de aprendizado de máquina sobre os dados de crimes armazenados no SIGO, um Termo de Convênio entre a SEJUSP e uma instituição de ensino superior ou médio técnico, voltada para área de tecnologia da informação, onde as atividades seriam elaboradas no formato de estágio supervisionado ou na disposição de bolsa de estudo pela FUNDECT seria celebrado. Ressalta-se que a Universidade Federal de Mato Grosso do

Sul (UFMS) possui programas de pós-graduação *stricto sensu* com linhas de pesquisa na área de Inteligência Artificial. Também poderia ser implantado um **departamento especializado para análise dos dados.** Um laboratório poderia ser instalado nas dependências físicas da própria SEJUSP, onde ficariam os analistas e os acadêmicos. Os analistas seriam os profissionais da segurança pública com conhecimento na área.

Uma parceria semelhante já vem acontecendo no estado do Ceará e tem rendido excelentes resultados. O estado tem aplicado tecnologias de inteligência artificial no combate à criminalidade e com isso vem alcançado reduções expressivas nos Crimes Violentos Letais Intencionais (CVLI) e nos Crimes Violentos contra o Patrimônio (CVP). Isso foi possível graças à parceria entre a Secretaria da Segurança Pública e Defesa Social do Ceará (SSPDS) e a Universidade Federal do Ceará. Um Sistema de Visualização e Mineração de Dados para Análise de Criminalidade foi desenvolvido lá a partir desta parceria. O sistema mapeia pontos com altos índices de criminalidade na cidade, além de apontar estatisticamente quais modalidades de crimes podem vir a ser cometidas em determinados locais. O mecanismo também permite que sejam realizados planejamentos estratégicos e redirecionamento de policiamento, baseando-se em manchas criminais. [9] A parceria entre a UFC e a SSPDS rendeu ainda outra parceria, foi a com o Ministério da Justiça e Segurança Pública (MJSP). O MJSP por meio da SENASP firmou uma parceria, onde foram desenvolvidas novas ferramentas que serão repassadas para unidades da federação: o Sinesp Big Data, o Sinesp Geo, o Sinesp Tempo Real e o Sinesp Busca.[10]

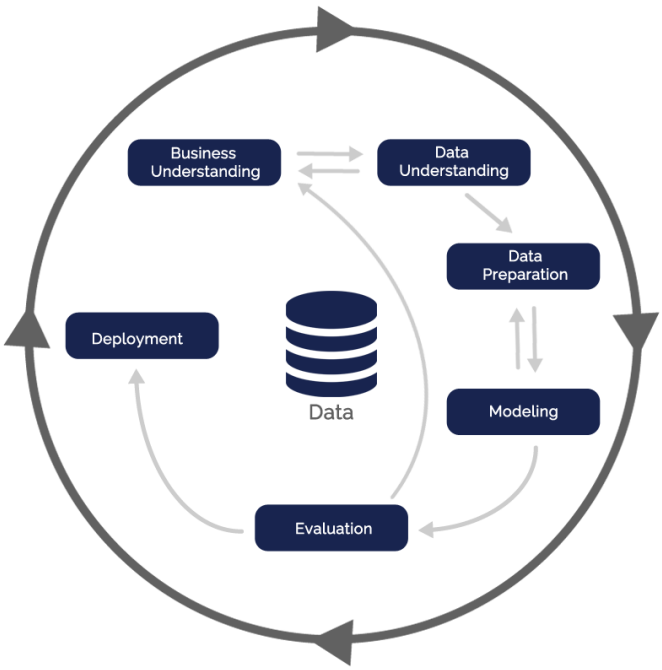
Para atingir os objetivos propostos nessa Ideia Inovadora Implementável, será utilizada uma instanciação do projeto de mineração de dados definido pelo método CRISP-DM *(Standard Process for Data Mining)*, a metodologia CRISP-DM é uma metodologia de mineração de dados composta por 6 etapas (Compreensão do Negócio, Compreensão dos Dados, Preparação dos Dados, Modelagem, Avaliação, e Desenvolvimento), as quais guiam o processo de extração de padrões e conhecimento a partir de bases de dados (PADUA; SOUZA, 2018).

Na Figura 3 são apresentadas as etapas definidas pelo Modelo CRISP-DM, bem como a iteração entre as etapas. As setas indicam as dependências mais importantes e frequentes entre

[9] https://www.sspds.ce.gov.br/2017/10/03/title9532/
[10] https://www.sspds.ce.gov.br/2019/08/20/secretaria-da-seguranca-do-ceara-desponta-como-inovadora-em-seminario-de-boas-praticas-do-ministerio-da-justica/

as fases. A sequência das fases não é rigorosa. Na verdade, a maioria dos projetos se move para frente e para trás entre as fases conforme necessário.

Figura 3 - CRISP-DM



Fonte: Otaris, 2018.

Os detalhes de cada passo do método definido para o processo em questão são:

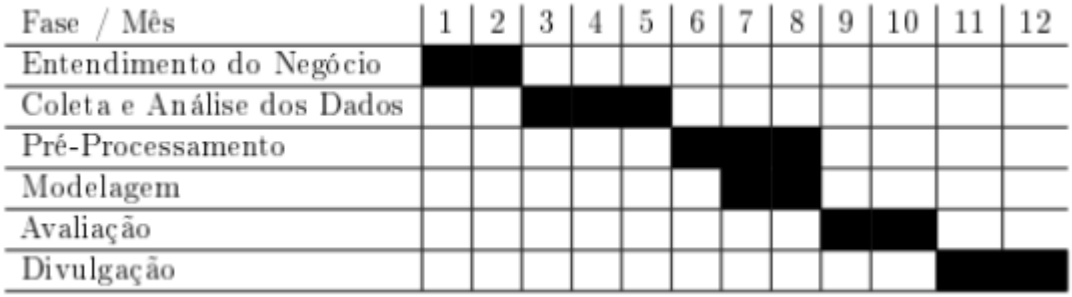
1. **Compreensão do negócio.** Será adquirida, juntamente com a Secretaria de Justiça e Segurança Pública de MS e a Compnet/CADG, a base de dados contendo registros de crimes dos quais pretende-se prever. Posteriormente, serão elencadas quais as técnicas de aprendizado de máquina serão mais adequadas para atender o objetivo e para os tipos de dados disponíveis, bem como será feita a avaliação dos mesmos.
2. **Entendimento dos dados.** Com o objetivo de se obter conhecimento do conjunto de dados e identificar atributos importantes, irrelevantes e correlacionados, além de identificar valores ausentes, exemplos redundantes, bem como técnicas para o tratamento destes. Além disso, o conjunto de dados será explorado de forma a identificar a distribuição de valores dos atributos de forma a escolher técnicas mais apropriadas nas próximas etapas.
3. **Preparação dos dados.** Serão realizados procedimentos de imputação ou remoção de atributos ausentes, padronização ou remoção de atributos

considerados irrelevantes ou prejudiciais para o processo de extração de conhecimento, e conversão de tipos de atributos para tipos mais adequados de acordo com as técnicas de modelagem a serem empregadas.

4. **Modelagem.** Serão aplicados algoritmos de aprendizado de máquina preditivos nos dados da base SIGO para previsão de crimes.
5. **Avaliação.** Serão aplicados esquemas de validação como o *K-fold cross-validation,* e serão aplicadas métricas tradicionais, como acurácia, precisão, revocação, F1 e AUC-ROC em caso do emprego de algoritmos de classificação. Já no emprego de regressão serão usadas métricas de avaliação como: RMSE, MAPE e R-quadrado.
6. **Implantação.** A última etapa do modelo CRISP-DM consiste na implementação da solução trabalhada ao longo de todo o processo, podendo variar de um relatório simples com as informações e conhecimentos extraídos dos dados, até a implementação de um processo de mineração de dados nas rotinas da empresa (SHEARER; CHAPMAN et al., 2000). O conhecimento extraído por meio das análises será disponibilizado às autoridades competentes para fins de conhecimento em relação aos padrões de crimes, bem como para o suporte à tomada de decisões.

A Figura 4 mostra o cronograma com a previsão para implantação seguindo a metodologia CRISP-DM. O prazo para a implantação do projeto será de 12 meses.

Figura 4 - Cronograma para implantação do projeto



Fonte: Autoria Própria.

8 Recursos necessários.

Os recursos humanos empregados no projeto serão compostos por servidores públicos da área de segurança com saber em ciência de dados, além de acadêmicos da área de tecnologia da informação. Os acadêmicos receberiam bolsa de estudo pela FUNDECT e os servidores públicos já possuem a remuneração do estado. Todavia, como forma de incentivo ao crescimento profissional, e até mesmo o interesse em trabalhar no **departamento especializado para a análise de dados** que seria criado, um adicional poderia ser pago a esses servidores com o saber comprovado. Enfim, para a implantação do projeto utilizaríamos de parcerias, servidores próprios, equipamentos de informática e instalações físicas do próprio estado.

9 Mecanismos de avaliação.

O projeto utilizará como método de avaliação, o *k-fold cross-validation* (CUNHA, 2019), onde a amostra de dados D é dividida em *k* partes de tamanhos semelhantes e a cada iteração os dados de da amostra D_k são utilizados para o teste e o restante para treinamento, de modo que, ao final do processo, todos os dados são utilizados como teste. A escolha do valor de *k* é uma importante etapa para este processo, a escolha de um valor de *k* baixo pode diminuir em muito o número de exemplo de treinamento, enquanto um valor de *k* alto pode ocasionar um alto gasto computacional e consequentemente demandar maior tempo para processamento, sendo mais comum o uso de $k = 10$ (CUNHA, 2019). Quanto às medidas de avaliação do modelo, um subconjunto de dados rotulado em duas diferentes classes: fluxo provável (positivo) ou fluxo improvável (negativo) e submetidos previamente ao treinamento do modelo para a construção da matriz de confusão, serão avaliados através das seguintes métricas (ROSSI, 2011):

- **Precisão:** avaliação dos acertos para a classe positiva em relação a todos os documentos que o modelo classificou como pertencente à classe. Seguindo a fórmula:

$$precisão = \frac{VP}{VP+FP},$$

no qual *VP* são os elementos que o modelo classificou como positivos e que são verdadeiramente positivos, e *FP* são elementos que o modelo classificou como positivos, mas que foram rotulados como negativos (ROSSI, 2011).

- **Revocação:** avaliação dos acertos para uma determinada classe em relação a todos os documentos que verdadeiramente pertenciam à classe. Seguindo a fórmula:

$$revocação = \frac{VP}{VP+FN},$$

no qual *FN* são elementos que o modelo classificou como negativos, mas que foram rotulados como positivos (ROSSI, 2011).

- **Acurácia:** A acurácia (*accuracy* ou ACC) é considerada uma das métricas mais simples e importantes. Ela avalia simplesmente o percentual de acertos, ou seja, ela pode ser obtida pela razão entre a quantidade de acertos e o total de entradas:

$$acurácia = \frac{Total\ de\ Acertos}{Total\ de\ Itens}$$

Utilizando como base a matriz de confusão, podemos obter a acurácia pela fórmula (FERRARI, 2017):

$$acurácia = \frac{VP+VN}{VP+FN+VN+FP}$$

- **F-score:** *F–measure, F-score* ou *score* F_1 é uma média harmônica calculada com base na precisão e na revocação. Ela pode ser obtida com base na equação (FERRARI, 2017):

$$f1 = 2 * \frac{Precisão * Sensibilidade}{Precisão + Sensibilidade}$$

Para a avaliação dos modelos de aprendizado de máquina baseados em **tarefas de regressão**, além das medidas como **MAPE, MAE e MSE**, seria usado também o R-quadrado, ele é uma medida estatística de quão próximos os dados estão da linha de regressão ajustada. Ele também é conhecido como o coeficiente de determinação ou o coeficiente de determinação múltipla para a regressão múltipla. O R-quadrado é bastante simples: é a porcentagem da variação da variável resposta que é explicada por um modelo linear.

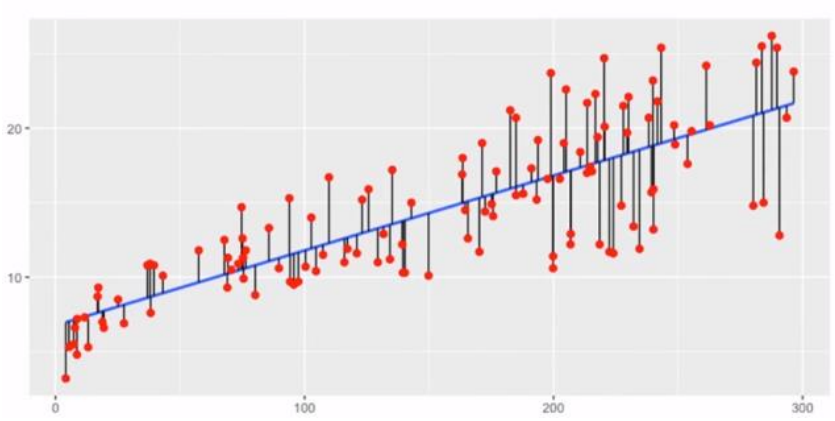
O R-quadrado está sempre entre **0 e 100%:**

- **0%** indica que o modelo não explica nada da variabilidade dos dados de resposta ao redor de sua média.

- **100%** indica que o modelo explica toda a variabilidade dos dados de resposta ao redor de sua média.

Em geral, quanto maior o R-quadrado, melhor o modelo se ajusta aos seus dados. A melhor regressão tem a capacidade de minimizar a distância entre todos os pontos da distribuição de dados, cada distância dos pontos vermelhos para a reta azul conforme a Figura 5 abaixo representa o erro que esta regressão acumula.

Figura 5 - Regressão Linear



Fonte: Gutelvam, 2020.

10 Obstáculos na realização da Ideia Inovadora Implementável.

Os principais obstáculos na execução dessa Ideia Inovadora Implementável são:

1. O preenchimento incorreto e a falta de preenchimentos dos dados nos boletins de ocorrências.
2. Celebração do Termo de Convênio com instituição de ensino superior voltada para o campo da tecnologia da informação.

11 Referências Bibliográficas ou de Projetos Catalogados ou Validados.

BHARATI, A. e SARVANAGURU, R. (2018). **Crime prediction and analysis using machine learning.** *International Research Journal of Engineering and Technology*. páginas 1037–104.

BRAZ, L. M.; FERREIRA, R.; DERMEVAL, D.; VÉRAS, D.; LIMA, M.; TIENGO, W. (2009). **Aplicando mineração de dados para apoiar a tomada de decisão na segurança pública do estado de alagoas.**

CHAPMAN, P. et al. (2000). **CRISP-DM 1.0: Step-by-step data mining guide.** [S.l.: s.n.]. 13 p.

HAN, J., PEI, J., e KAMER, M. (2011) **Data mining: concepts and technique.**

HOSSAIN, S., ABTAHEE, A., KASHEM, I., HOQUE, M. M., e SARKER, I. H. (2020). **Crime prediction using spatio-temporal data.** *In International Conference on Computing Science, Communication and Security.* páginas 277–289. Springer.

KIM, S., JOSHI, P., KALSI, P. S., TAHERI, P. (2018). **Crime analysis through machine learning.** *In* 2018 IEEE 9th *Annual Information Technology, Electronics and Mobile Communication Conference (IEMCON),* páginas 415–420. IEEE.

NUCCI, H. H. P. (2019). **Classificação do grau de acabamento de gordura da carcaça de bovinos de corte usando aprendizado de máquina.** *Master's thesis,* UFMS.

PADUA, A. F. L. D. O.; SOUSA, F. A. (2018). **Methodology crisp-dm: Potentialities in the discovery of knowledge in education data.**

ROSSI, R. G. (2011). **Representação de coleções de documentos textuais por meio de regras de associação.** PhD *thesis*, Universidade de São Paulo.

ROSSI, R. G. (2015). **Classificação automática de textos por meio de aprendizado de máquina baseado em redes.** PhD *thesis,* Universidade de São Paulo.

SHEARER, C. (2000). **The crisp-dm model: The new blueprint for data mining.** *Journal of data warehousing*. v. 5, p. 13–22.

SILVA, E.; ROVER, A. (2012). **O processo de descoberta do conhecimento como suporte à análise criminal: minerando dados da segurança pública de Santa Catarina.** *In International Conference on Information Systems and Technology Management,* volume 8, páginas 3144–3174.

SORESCU, A. (2017). *Data-driven business model innovation. Journal of Product Innovation Management*, v. 34, n. 5, p. 691-696.

TAN, P. N., STEINBACH, M., e KUMAR, V. (2016). **Introduction to data mining.** *Pearson Education India.*