UNIVERSIDADE
FEDERAL RURAL
DE PERNAMBUCO

Luiz Henrique Teixeira Carvalho

Avaliação de algoritmos multi-classe para classificação de solicitações enviadas a Ouvidoria Geral do Estado de Pernambuco

Carpina

2021

Luiz Henrique Teixeira Carvalho

Avaliação de algoritmos multi-classe para classificação de solicitações enviadas a Ouvidoria Geral do Estado de Pernambuco

Trabalho de Conclusão de Curso apresentada ao Curso de Bacharelado em Sistemas de Informação da Unidade Acadêmica de Educação a Distância e Tecnologia da Universidade Federal Rural de Pernambuco como requisito parcial à obtenção do grau de Bacharel.

Universidade Federal Rural de Pernambuco – UFRPE

Unidade Acadêmica de Educação a Distância e Tecnologia

Curso de Bacharelado em Sistemas de Informação

Orientador: Jeneffer Cristine Ferreira

Carpina

2021

Dados Internacionais de Catalogação na Publicação
Universidade Federal Rural de Pernambuco
Sistema Integrado de Bibliotecas
Gerada automaticamente, mediante os dados fornecidos pelo(a) autor(a)

C331a Carvalho, Luiz Henrique Teixeira
Avaliação de algoritmos multi-classe para classificação de solicitações enviadas a Ouvidoria Geral do Estado de Pernambuco / Luiz Henrique Teixeira Carvalho. - 2021.
44 f. : il.

Orientadora: Jeneffer Cristine Ferreira.
Inclui referências.

Trabalho de Conclusão de Curso (Graduação) - Universidade Federal Rural de Pernambuco, Bacharelado em Sistemas da Informação, Recife, 2021.

1. Classificação Automática de dados. 2. Classificação multi-classe. 3. Inteligência Artificial. 4. Ouvidoria. I. Ferreira, Jeneffer Cristine, orient. II. Título

CDD 004

LUIZ HENRIQUE TEIXEIRA CARVALHO

AVALIAÇÃO DE ALGORITMOS MULTICLASSE PARA CLASSIFICAÇÃO DE SOLICITAÇÕES ENVIADAS A OUVIDORIA GERAL DO ESTADO DE PERNAMBUCO

Trabalho de Conclusão de Curso apresentado em cumprimento às exigências do curso de Bacharelado em Sistemas de Informação da Unidade Acadêmica de Educação a Distância e Tecnologia/UFRPE para obtenção do título de Bacharel em Sistemas de Informação, sob a orientação do (a) Profª Ms. Jeneffer Cristine Ferreira.

Aprovado em 02 de março de 2021

_____

Profª Ms. Jeneffer Cristine Ferreira – UFRPE

_____

Profª Ms. Adalmeres Cavalcanti da Mota – UAEADTec/UFRPE

_____

Profª Drª. Juliana Regueira Basto Diniz – UAEADTec/UFRPE

Recife – PE, 2021

À minha família e amigos

Agradecimentos

Agradeço primeiramente a Deus, pela saúde, força de vontade e outras bençãos a mim concedidas para superar os obstáculos.

Agradeço a minha mãe, Roselia, por ser a luz que guia meus passos, por todo o amor incondicional e por nunca medir esforços para que eu alcançasse os meus objetivos.

Agradeço ao meu pai Moacy pelo apoio e por toda contribuição para formação do cidadão que eu sou hoje.

Agradeço a minha namorada Wanessa, pelo apoio e compreensão dificeis e estressantes, e por estar ao meu lado sempre que eu precisei.

Agradeço aos meus tios Abraão, João por ter me ajudado na minha caminhada.

Agradeço também aos meus outros familiares por toda a ajuda direta ou indireta.

Agradeço tambem ao meu irmão Matheus(Drago) e ao meu Primo Felipe por nessa caminhada academica, terem contribuido para minha conclusão.

Agradeço ao meu amigo Clodoalves, por além da grande amizade, me ajudou muito academicamente e me apoiando nos momentos criticos e desafiadores. Mais que um amigo um irmão.

Agradeço também a um grande amigo Rodrigo que faleceu vitima de covid-19, mas que sempre me apoio em tudo e sempre será um grande amigo.

Agradeço também a todos os professores pelos conhecimentos adquiridos que solidificaram minha formação acadêmica.

“A persistência é o caminho do êxito.”
(Charles Chaplin)

Resumo

A Ouvidoria Geral é um órgão público que abrange todo o estado de Pernambuco e todos os dias recebe diversas solicitações com os mais variados temas envolvendo todos os outros órgãos do estado, com isso em determinadas épocas do ano, essas solicitações podem chegar a onerar os recursos do estado.

O objetivo principal desse trabalho é aplicar os algoritmos de classificação multi-classe nos dados obtidos a partir do portal da transparência, e tentar prever as solicitações enviadas a Ouvidoria Geral do Estado de Pernambuco

Para obtenção dos dados da Ouvidoria Geral do Estado de Pernambuco, foi executada uma raspagem de dados no Portal da Transparência de Pernambuco. Foram obtidos os dados dos anos de 2017, 2018 e 2019.

Foi aplicado nos dados da ouvidoria os algoritmos de Arvore de Decisões(*Decision Tree*), Floresta Aleatoria(*Random Forest*), *Bagging* e *kNN*.

Os resultados mostraram que os algoritmos de classificação automática de dados, particularmente os algoritmos de *Decision Tree*(Arvore de decisões), *Random Forest*(Floresta Aleatória) e *Bagging* conseguiram de 55% e 32% nas classes de tipo e órgão respectivamente, tendo um aproveitamento de um acerto a cada duas tentativas na classe de tipo e de um acerto a cada três tentativas na classe de órgão.

Os algoritmos também foram avaliados acerca de seu desempenho em tempo de criação e treinamento do modelo, tendo o algoritmo de *Decision Tree*(Arvore de decisões) como o mais performático.

Palavras-chave: Classificação Automática de dados, Classificação multi-classe, Inteligencia Artificial, Ouvidoria.

Abstract

The Ombudsman's Office is a public agency that covers the entire state of Pernambuco and every day receives several requests with the most varied themes involving all other organs of the state, with that in certain times of the year, these requests can come to burden the resources of State.

The main objective of this work is to apply the multi-class classification algorithms to the data obtained from the transparency portal, and to try to predict requests sent to the Ombudsman's Office of the State of Pernambuco

To obtain data from the Ombudsman's Office of the State of Pernambuco, data scraping was carried out on the Pernambuco Transparency Portal of Pernambuco. Data for the years 2017, 2018 and 2019 were obtained.

The algorithms Decision Tree, Random Forest, Bagging and kNN were applied to the ombudsman data.

The results showed that the automatic data classification algorithms, particularly the Decision Tree, Random Forest, Bagging algorithms achieved 55 % and 32 % in the type and organ classes respectively, taking advantage of one hit every two attempts in the type class and one hit every three attempts in the organ class.

The algorithms were also evaluated about their performance in time of model creation and training, with the Decision Tree algorithm as the most performative.

Keywords: Classification, Automated data classification, multiclass classification, Artificial Intelligence, Ombudsman.

Lista de ilustrações

Lista de tabelas

Lista de abreviaturas e siglas

ADs	Arvore de decisão
ARPE	Agencia Reguladora de Pernambuco
DECISION TREE	Arvore de decisões
IA	Inteligência Artificial
kNN	k-nearest neighbors
SCGE	Secretaria da Controladoria-Geral do Estado
SDS	Secretaria de Defesa Social
RANDOM FOREST	Floresta Aleatória
OGE	Ouvidoria Geral do Estado de Pernambuco

Sumário

1 Introdução

Uma ouvidoria governamental é um canal de comunicação totalmente criado dentro de um setor de administração do estado, tendo como finalidade, receber denúncias, reclamações, elogios e solicitações, informações dos cidadãos, e ajudar a resolve-los.(IASBECK, 2010)

A Ouvidoria Geral do Estado de Pernambuco (OGE) é um órgão do governo estadual que tem como finalidade, fortalecer a cidadania, aproximar o cidadão da gestão pública estadual e contribuir com a melhoria contínua dos serviços públicos prestados à sociedade.(SCGE, 2020)

A OGE está vinculada à Secretaria da Controladoria-Geral do Estado (SCGE) e é responsável por receber, examinar e encaminhar sugestões, elogios, solicitações, reclamações e denúncias aos órgãos e entidades responsáveis por adotar as devidas providências, possibilitando que o cidadão se relacione de forma direta com os órgãos executores dos serviços públicos, exercendo assim um importante papel de incentivo ao controle social.(SCGE, 2020)

Toda solicitação enviada a OGE é feita diretamente pelo portal da Ouvidoria sendo recebida pela sede da ouvidoria na Controladoria Geral do Estado. A partir deste ponto, é feita uma triagem para decidir para qual órgão aquela solicitação deverá ser enviada. Após a resolução da solicitação, é enviada uma resposta ao cidadão.

Todo esse processo é feito em sistemas desenvolvidos e gerenciados pelos servidores da área de Tecnologia da Informação dentro da Controladoria Geral do Estado de Pernambuco.

Os dados da OGE são considerados dados abertos (ISOTANI; BITTENCOURT, 2015). Desta forma, são dados que podem ser livremente utilizados, reutilizados e redistribuídos por qualquer pessoa desde que seja informado a fonte.

Os dados da ouvidoria por serem abertos podem ser encontrados diretamente dentro do site do Portal da Transparência do Estado de Pernambuco dentro da área de acesso a informação e consulta pública.(TRANSPARENCIA, 2020)

Esses dados estão disponíveis no portal da transparência, mas por serem disponibilizados pela web, não estão estruturados, tendo em visto a necessidade do uso de mineração de dados para obtenção desses dados e estruturação.

Mineração de dados é a descoberta de estruturas e padrões em grandes e complexos *datasets*.

De acordo com(CAMILO; SILVA, 2009) a mineração de dados é uma das tec-

nologias mais promissoras da atualidade, pois a partir de uma grande quantidade de dados não estruturados é possível extrair bastante informações.

A mineração de dados é uma importante parte do escopo desse projeto, porque é a partir dessa técnica que será possível entender os dados da OGE e estrutura-los afim de obter um melhor entendimento dos mesmos.

1.1 Problematização

A Ouvidoria Geral é um órgão público que abrange todo o estado de Pernambuco e todos os dias recebe diversas solicitações com os mais variados temas envolvendo todos os outros órgãos do estado, com isso em determinadas épocas do ano essas solicitações podem chegar a onerar os recursos do estado.

1.2 Justificativa

Atualmente existem diversos algoritmos que podem ajudar a minerar e classificar essas solicitações, com isso esse trabalho tem como cenário estudar e analisar esses algoritmos para que a ouvidoria consiga antecipar quando essas grandes demandas irão chegar e poder classificar quais os principais temas por data.

1.3 Objetivo Geral

Esse trabalho tem como cenário aplicar os algoritmos de classificação nos dados obtidos a partir do portal da transparência, e tentar prever as solicitações enviadas a Ouvidoria Geral do Estado de Pernambuco.

1.4 Objetivo Especifico

- Relatar as principais características da Ouvidoria
- Categorizar os principais tipos de solicitações enviadas a Ouvidoria
- Relatar e estudar alguns dos principais algoritmos de classificação
- Minerar e analisar os dados da ouvidoria
- Classificar os dados da ouvidoria

Para tanto, diante destes objetivos pretende-se analisar os algoritmos de classificação de dados a partir dos dados da Ouvidoria geral do Estado de Pernambuco nos próximos capítulos.

2 Referencial Teórico

O trabalho apresentado é um projeto que visa estudar e analisar algoritmos de mineração e classificação para classificar as principais solicitações enviadas a Ouvidoria Geral do estado de Pernambuco (OGE), a partir de uma massa de dados retirada do Portal da Transparência do Estado de Pernambuco usando *Web Scraping*, que é disponibilizado pela Secretaria da Controladoria Geral do Estado de Pernambuco que esta vinculada a OGE.(TRANSPARENCIA, 2020)

2.1 Orgãos Públicos

Os órgãos públicos formam a estrutura do Estado, mas não têm personalidade jurídica, uma vez que são apenas parte de uma estrutura maior, essa sim detentora de personalidade. Como parte da estrutura maior, o órgão público não tem vontade própria, limitando-se a cumprir suas finalidades dentro da competência funcional que lhes foi determinada pela organização estatal. (ROMANO, 2020)

Os órgãos não são pessoas, nem se distinguem do Estado. São círculos de atribuições os feixes individuais de poderes funcionais repartidos no interior da personalidade estatal e expressados através dos agentes nele providos.(ROMANO, 2020)

A ouvidoria por ser um órgão público, não tem autonomia sobre as decisões dos órgãos que integram o governo do estado, ela somente redireciona as solicitações para o respectivos órgãos e depois encaminha a resposta ao cidadão.

2.2 Ouvidoria

Ouvidoria é um serviço prestado aos clientes e cidadãos por meio do qual é possível apresentar reclamações, críticas, sugestões e até mesmo elogios à qualidade das trocas empreendidas. Para tanto, atua como mídia, produzindo, reproduzindo e reformulando sentidos. Seu objetivo principal é curar vínculos estremecidos no relacionamento entre as organizações e seus públicos.(IASBECK, 2010)

A Ouvidoria Geral do Estado é um órgão do governo do estado de Pernambuco vinculado a Secretaria da Controladoria Geral do Estado e visa atender a população recebendo diversas solicitações por dia.

Os dados da Ouvidoria Geral do Estado utilizados nesse projeto são dados abertos e foram extraídos a partir do Portal da Transparência do Estado de Pernambuco na seção de Acesso a informação.

O Portal da Transparência é supervisionado pelo servidores da área de Tecnologia da Informação dentro da Secretaria da Controladoria Geral do Estado de Pernambuco, mesmo órgão onde a Ouvidoria Geral de Estado de Pernambuco esta sediada.

A ouvidoria Geral do Estado de Pernambuco fica sediada na Secretaria da Controladoria Geral do Estado de Pernambuco, entretanto cada órgão do estado de Pernambuco tem um setor de Ouvidoria onde recebem as solicitações, as tratam e depois as devolvem para a sede da ouvidoria com suas respectivas respostas.

Os tipos de solicitações enviadas a Ouvidoria:

- Denúncia: São as denuncias enviadas pela população aos órgãos do estado de Pernambuco.
- Elogio: São os elogios enviados aos órgãos do estado de Pernambuco.
- Informação: São as solicitações de informação do estado de Pernambuco.
- Reclamação: São reclamações enviadas aos órgãos do estado de Pernambuco.
- Solicitações: São as solicitações genéricas que não são solicitações de informação enviadas aos órgãos do estado de Pernambuco.
- Sugestão: São sugestões enviadas aos órgãos do estado de Pernambuco.
- Não Classificado: São solicitações que não entram em nenhuma das categorias anteriores.

Na tabela 1 são exibidos os tipos de solicitações enviadas a ouvidoria divididos por quantidade entre os anos de 2017 e 2018:

Tabela 1 – Quantidade de solicitações enviadas a ouvidoria por tipo entre os anos de 2017 e 2018

Tipo	Quantidade de Solicitações
Denúncia	16860
Elogio	1332
Informação	68
NÃO CLASSIFICADO	411
Reclamação	26337
Solicitação	52007
Sugestão	924

2.3 Inteligência Artificial

Nos últimos anos a inteligencia artificial vem surpreendendo a todos, sendo um avanço tecnológico que permite que algoritmos simulem redes cognitivas avançadas similar a humana.

Algumas definições de Inteligencia artificial segundo alguns acadêmicos:

Segundo Teixeira (TEIXEIRA, 2019), inteligência artificial é descrito como executar tarefas que necessitem algo que se pensava de exclusividade humana, realizar tarefas pensando e agindo racionalmente.

De acordo com Nilsson (NILSSON, 2014), inteligência artificial se divide em algumas áreas de aplicação como processamento de linguagem natural, programação automática, robóticas, visão de maquina, sistemas inteligentes envolvendo dados, etc.

Dentro da inteligência artificial existem diversos campos de estudo, sendo deles:

- *Machine Learning*: *Machine Learning* é um programa de computador que aprende com dados de exemplo ou execuções passadas.(ALPAYDIN, 2020)
- *Deep Learning*: O *Deep Learning* permite que modelos computacionais compostos por várias camadas de processamento aprendam representações de dados com vários níveis de abstração. Esses métodos melhoraram drasticamente o estado da arte em reconhecimento de fala, reconhecimento de objeto visual, detecção de objeto e muitos outros domínios, como descoberta de drogas e genômica(LECUN; BENGIO; HINTON, 2015)
- Processamento de Linguagem Natural: De acordo com (VIEIRA; LOPES, 2010) Processamento de Linguagem Natural ou *Natual Processing Language em ingles*(NLP) é um programa de computador que usa inteligencia artificial para compreender a linguagem dos seres humanos
- Redes Neurais: Uma Rede Neural é um maquina que é projetada para modelar a maneira como o cérebro realiza uma tarefa particular ou função de interesse. A rede é normalmente implementada utilizando-se componentes eletrônicos ou é simulada por programação em um computador digital. (HAYKIN, 2007)
- Mineração de Dados: Mineração de dados é a descoberta de estruturas e padrões em *data sets* grandes e complexos.(HAND; ADAMS, 2014)

2.4 Machine Learning

O *machine learning* ou aprendizado de máquina é um sistema ou algoritmo que aprende com uma massa de dados ou diversas execuções passadas.(ALPAYDIN,

2020)

Em *machine learning* existem três diferentes modos de aprendizagem(FERRERO, 2009):

- Supervisionado
- Não Supervisionado
- Semi-Supervisionado

A diferença entre esses três diferentes modos é a presença ou não do elemento classe que é o rotulo do exemplo do conjunto de dados fornecido ao algoritmo, denominado conjunto de treinamento. No aprendizado supervisionado essa classe é conhecida enquanto no aprendizado não supervisionado, os dados não possuem essas classes. No aprendizado semi-supervisionado consiste em muitos dados com classes e muitos sem.(FERRERO, 2009)

2.5 Classificação automática de dados

Nos últimos anos, várias técnicas têm sido desenvolvidas e testadas visando encontrar melhores resultados para sistemas inteligentes. As técnicas são direcionadas no sentido de atribuir à máquina, capacidade de aprendizagem similar a de um ser humano, sendo a classificação uma das principais tarefas que fazem parte das tais técnicas. Para melhorar a confiança no processo de classificação introduz-se o conceito de rejeição. A rejeição admite que um sistema de reconhecimento aplique uma decisão global de aceitar ou recusar uma hipótese se o classificador não estiver suficientemente certo.(ALMEIDA, 2010)

Algumas definições de classificação segundo alguns acadêmicos:

Segundo (COSTA et al., 1999). Classificação automática de dados é uma análise de problemas onde amostras são descritas como variáveis p- dimensionais.

De acordo com (SARITAS; YASAR, 2019). Classificação é uma poderosa técnica de mineração de dados com uma grande variedade de aplicações com vários tipos de dados existente em nossas vidas

Alguns dos principais Algoritmos de Classificação segundo (ALMEIDA, 2010):

- Classificação binária: Classificações binárias é o nome dado a categoria dos problemas de classificação que só possuem 2 classes.(ALMEIDA, 2010)

- Classificação multi-classes: Classificação multi-classe é o nome dado a categoria dos problemas de classificação onde existem mais de duas classes a ser inferida.(SANTOS, 2017)
- Classificação multi-classes nominal e ordinal: A diferença entre as quantidades nominais e ordinais é que os últimos apresentam uma ordem entre os diferentes valores que podem assumir.(ALMEIDA, 2010)
- Classificação binária com a opção de rejeição: A opção de rejeição é uma técnica utilizada para melhorar o desempenho de sistemas de reconhecimento de padrões. (ALMEIDA, 2010)
- Classificação multi-classes com a opção de rejeição: O problema da classificação com a opção de rejeição estende se também a problemas multi-classes. Para formalizar a rejeição usa a notação de confiança associada a uma hipótese. Considera um problema de classificação atribuído a um classificador C que fornece na saída uma medida de confiança ci para cada uma das "c" classes pertencentes ao problema. (ALMEIDA, 2010)

2.5.1 Classificadores

2.5.1.1 KNN

O algoritmo *k-Nearest Neighbor - kNN* -é um algoritmo de aprendizado supervisionado do tipo *lazy*, introduzido por (AHA; KIBLER; ALBERT, 1991). A ideia geral desse algoritmo consiste em encontrar os *k* exemplos rotulados mais próximos do exemplo não classificado e, com base no rotulo desses exemplos mais próximo, é tomada a decisão relativa à classe do exemplo não rotulado. Os algoritmos da família *kNN* requerem um pouco de esforço durante a etapa de treinamento. Em contrapartida, o custo computacional para rotular um novo exemplo é relativamente alto, pois, no pior dos casos, esse exemplo deverá ser comparado com todos os exemplos contidos no conjunto de exemplos do treinamento.(FERRERO, 2009)

2.5.1.2 Decision Tree

O algoritmo *Decision Tree* ou arvore de decisões em português, tem como ideia geral utilizar mecanismos de categorização usando divisão hierárquica dos dados, em que um padrão desconhecido é rotulado, usando-se uma sequência de decisões. Na aplicação em dados multi espectrais, o desenho da árvore de decisão é baseado no conhecimento das propriedades espectrais de cada classe e na relação entre as classes. (DELGADO et al., 2012)

2.5.1.3 Random Forest

O algoritmo *Random Forest* ou floresta aleatória em português são conjuntos de Árvores de Decisão (ADs) criadas a partir de uma base de dados. O principal desafio ao gerar uma floresta de decisão é como obter uma boa variabilidade nas árvores que a compõem, levando a um maior poder de generalização (classificar instâncias desconhecidas) para o modelo.(RIQUETI; RIBEIRO; ZÁRATE, 2018)

2.5.1.4 Bagging

O algoritmo de *Bagging* tem como ideia geral construir os classificadores a partir de conjuntos sucessivos e independentes de amostras de dados geradas a partir do conjunto de dados original, tendo todos eles a mesma quantidade de exemplos (há, portanto, replicação e ausência de certos exemplos), criando classificadores diferentes devido à variação de exemplos nas amostras, sendo combinados através de um método de votação.(COSTA; BITTENCOURT; SOUTO, 2005)

2.5.2 Séries Temporais

Uma série temporal é uma coleção de observações feitas sequencialmente ao longo do tempo. A característica mais importante deste tipo de dados é que as observações vizinhas são dependentes e estamos interessados em analisar e modelar esta dependência. Enquanto em modelos de regressão por exemplo a ordem das observações é irrelevante para a análise, em séries temporais a ordem dos dados é crucial. Vale notar também que o tempo pode ser substituído por outra variável como espaço, profundidade, etc.(EHLERS, 2007)

2.6 Mineração de Dados

Mineração de Dados é um ramo da computação que teve início nos anos 80, quando os profissionais das empresas e organizações começaram a se preocupar com os grandes volumes de dados informáticos estocados e inutilizados dentro da empresa. Nesta época, *Data Mining* consistia essencialmente em extrair informação de gigantescas bases de dados a maneira mais automatizada possível.(AMO, 2004)

Algumas definições de mineração de dados segundo alguns acadêmicos:

Conforme (HAND; ADAMS, 2014). Mineração de dados é a descoberta de estruturas e padrões em *data sets* grandes e complexos.

Segundo (HAN; PEI; KAMBER, 2011). Explica que mineração de dados e suas ferramentas é o descobrimento de conhecimento a partir da coleta de dados.

De acordo com (AMO, 2004) a mineração de dados consiste em algumas etapas, como mostrado na figura 1:

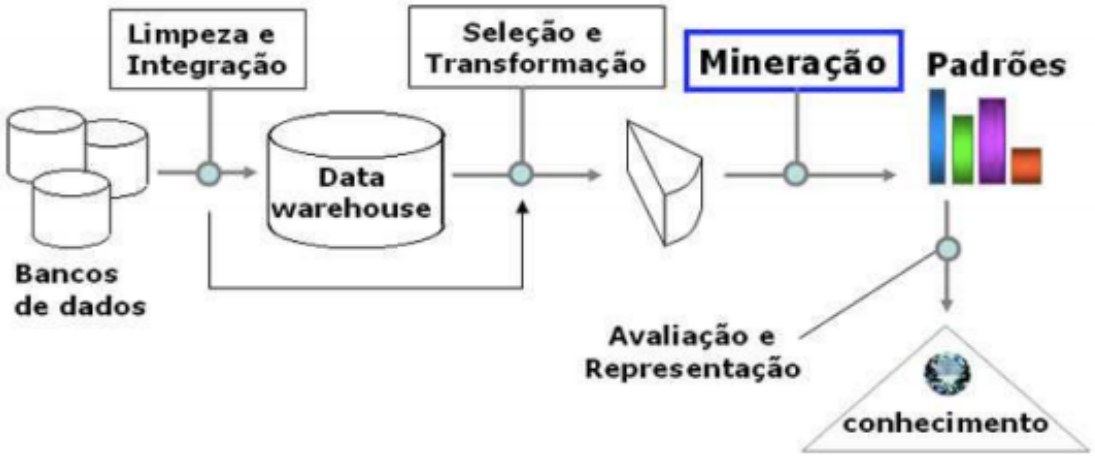

Figura 1 – Fluxo das etapas de uma mineração de dados.
Fonte:(AMO, 2004)

- Limpeza dos dados: etapa onde são eliminados ruídos e dados inconsistentes.
- Integração dos dados: etapa onde diferentes fontes de dados podem ser combinadas produzindo um único repositório de dados
- Seleção: etapa onde são selecionados os atributos que interessam ao usuário. Por exemplo, o usuário pode decidir que informações como endereço e telefone não são de relevantes para decidir se um cliente é um bom comprador ou não.
- Transformação dos dados: etapa onde os dados são transformados num formato apropriado para aplicação de algoritmos de mineração (por exemplo, através de operações de agregação).
- Mineração: etapa essencial do processo consistindo na aplicação de técnicas inteligentes a fim de se extrair os padrões de interesse.
- Avaliação ou Pós-processamento: etapa onde são identificados os padrões interessantes de acordo com algum critério do usuário.
- Visualização dos Resultados: etapa onde são utilizadas técnicas de representação de conhecimento a fim de apresentar ao usuário o conhecimento minerado

É importante distinguir o que é uma tarefa e o que é uma técnica de mineração. A tarefa consiste na especificação do que estamos querendo buscar nos dados, que tipo de regularidades ou categoria de padrões temos interesse em encontrar, ou que

tipo de padrões poderiam nos surpreender (por exemplo, um gasto exagerado de um cliente de cartão de crédito, fora dos padrões usuais de seus gastos). A técnica de mineração consiste na especificação de métodos que nos garantam como descobrir os padrões que nos interessam.(AMO, 2004)

Alguma das principais técnicas de mineração de dados segundo (AMO, 2004):

- Técnicas estatísticas
- Técnicas de aprendizado de máquina
- Técnicas baseadas em crescimento-poda-validação

As principais tarefas de mineração de dados segundo (AMO, 2004):

- Análise de Regras de Associação: Uma regra de associação é um padrão da forma X → Y , onde X e Y são conjuntos de valores (artigos comprados por um cliente, sintomas apresentados por um paciente, etc).(AMO, 2004)
- Análise de Padrões Sequenciais: Um padrão sequencial é uma expressão da forma < I1, . . . , In >, onde cada Ii é um conjunto de itens.(AMO, 2004)
- Classificação e Predição: Classificação é o processo de encontrar um conjunto de modelos (funções) que descrevem e distinguem classes ou conceitos, com o propósito de utilizar o modelo para predizer a classe de objetos que ainda não foram classificados.(AMO, 2004)
- Análise de *Clusters* (Agrupamentos): Diferentemente da classificação e predição onde os dados de treinamento estão devidamente classificados e as etiquetas das classes são conhecidas, a análise de *clusters* trabalha sobre dados onde as etiquetas das classes não estão definidas.(AMO, 2004)
- Análise de *Outliers*: Um banco de dados pode conter dados que não apresentam o comportamento geral da maioria.(AMO, 2004)

3 Metodologia

3.1 Propósito

Este trabalho teve como propósito aplicar os algoritmos de classificação multi-classe nos dados da Ouvidoria Geral do Estado de Pernambuco, e analisar o resultado.

E a partir dessa analise e classificação tentar ajudar a prever e melhorar o atedimento da Ouvidoria Geral do Estado de Pernambuco.

3.1.1 Classificação Automática de Dados

Classificação é uma poderosa técnica de mineração de dados com uma grande variedade de aplicações com vários tipos de dados existente em nossas vidas.(SARITAS; YASAR, 2019)

E este trabalho tem o objetivo aplicar alguns de seus principais algoritmos nos dados da Ouvidoria Geral do Estado de Pernambuco. Os algoritmos que serão abordados nesse trabalho:

Classificação multi-classes: Classificação multi-classe é o nome dado a categoria dos problemas de classificação onde existem mais de duas classes a ser inferida.(SANTOS, 2017)

Algoritmos de Classificação multi-classes que serão abordados nesse trabalho:

- *kNN*
- *Bagging*
- *Decision Tree*
- *Random Forest*

3.1.2 Ouvidoria

Ouvidoria é um serviço prestado aos clientes e cidadãos por meio do qual é possível apresentar reclamações, críticas, sugestões e até mesmo elogios à qualidade das trocas empreendidas. Para tanto, atua como mídia, produzindo, reproduzindo e reformulando sentidos. Seu objetivo principal é curar vínculos estremecidos no relacionamento entre as organizações e seus públicos.(IASBECK, 2010)

A Ouvidoria Geral do Estado é um órgão do governo do estado de Pernambuco vinculado a Secretaria da Controladoria Geral do Estado e visa atender a população recebendo diversas solicitações por dia.

Todos os dados da ouvidoria são dados abertos, e serão obtidos a partir do Portal da Transparência de Pernambuco.(TRANSPARENCIA, 2020)

Os dados que serão obtidos por solicitação:

- Tipo
- Protocolo
- Órgão
- Tipo de Resposta
- Assunto
- Data da Pergunta
- Pergunta
- Data da Resposta
- Resposta

3.2 Método de Pesquisa

3.2.1 Explicativa

O método de pesquisa utilizado é a explicativa, onde foram utilizadas técnicas de *web scraping* pra obtenção dos dados, e o uso de mineração de dados para análises.

Este trabalho tem como foco de pesquisa estudar os algoritmos de inteligencia artificial, e aplicá-los nos dados da Ouvidoria Geral do Estado de Pernambuco

3.3 Abordagem

3.3.1 Quantitativa

Para este trabalho sera utilizado a abordagem quantitativa, onde toda a analise a partir dos dados obtidos a partir de um *web scrapping*, serão apresentados na forma gráficos e tabelas.

Cada algoritmo de inteligência artificial utilizando as técnicas de classificação automática de dados terão resultados apresentados a partir dos dados obtidos da Ouvidoria Geral do Estado de Pernambuco

3.4 Procedimentos ou técnicas

3.4.1 Pesquisa de campo

Como pesquisa de campo deste projeto, será feito um estudo e análise dos algoritmos de inteligência artificial baseados em classificação automática de dados, utilizando as informações da Ouvidoria Geral do Estado de Pernambuco obtidas a partir do Portal da Transparência.

3.4.2 Obtenção dos dados

Para obtenção dos dados da Ouvidoria Geral do Estado de Pernambuco, foi necessário realizar um *Web scrapping*. Segundo (MITCHELL, 2018), *web scraping* é um programa automatizado que consegue requisitar dados de servidores web para extrair as informações necessárias.

Os dados sobre a Ouvidoria Geral do Estado de Pernambuco se encontram na seção de acesso a informação e consulta pública no portal da transparência.(TRANSPARENCIA, 2020)

Foi observado que o link para visualização das informações das solicitações enviadas a Ouvidoria Geral do Estado de Pernambuco, tem um padrão mostrado abaixo:

- web.transparencia.pe.gov.br
- /ModuloCidadao
- </atendimento_detail.xhtml?atendimentoId=>
- <Protocolo da solicitação enviada a Ouvidoria Geral do Estado>

Os Protocolos da Ouvidoria Geral do Estado de Pernambuco também seguem um padrão, como mostrado abaixo:

- Ano da solicitação
- Sequencial Numérico começando a partir de 1

Por exemplo um atendimento realizado no ano de 2019:

- 2019207

Neste protocolo de atendimento mostrado acima, vemos que o o ano de atendimento é 2019 e o número sequencial dele é 207.

O *script* então tenta abrir a pagina e se ela for publica ele obtém os dados a partir do *HTML* da pagina.

Seguindo essa logica dos protocolos de atendimento, foi realizado um *web scrping* na pagina do Portal da Transparência, tentando obter 50000 protocolos por ano.

Foi obtido dados dos anos de 2017,2018 e 2019

A automação iterou entre os anos de 2017,2018 e 2019 e para cada um deles tentou abrir a pagina de 1 a 50000, os que estavam públicos o *script* conseguiu baixar, os que não eram, foram ignorados.

O procedimento para realização desses dados foi executado entre os dias 31/05/2020 e 08/06/2020

3.4.3 Análise dos dados

Foi realizado a aplicação dos algoritmos e analise utilizando as seguintes linguagens de programação:

- Python

Todos os resultados das analises foram apresentados usando as seguintes ferramentas:

- Jupyter Notebook (Python)
- Plotly

Os algoritmos aplicados, foram os algoritmos de inteligencia artificial baseados em classificação automática de dados multi-classe.

Algoritmos estudados e analisados utilizando os dados obtidos da Ouvidoria Geral do Estado de Pernambuco:

- Classificação multi-classes usando os classificadores *kNN*, *Decision Tree*, *Random Forest* e *Bagging*

Os algoritmos foram aplicados usando as classes de Tipo e Órgão e tendo outro eixo a serie temporal de mês. O experimento tentou prever a probabilidade de cada classe para cada mês.

4 Resultados

Neste capítulo são apresentados e discutidos os resultados obtidos no experimento descrito no Capítulo 3. A Seção 4.1 apresenta uma análise acerca dos resultados encontrados obtidos por cada um dos algoritmos em análise. A Seção 4.2 mostra o desempenho computacional obtido pelos algoritmos em termos de tempo de geração do modelo e tempo de predição do modelo. Por fim, a Seção 4.3 sumariza e discute os resultados obtidos.

4.1 Análise

Como resultado geral, observando os seguintes algoritmos:

- *kNN*
- *Decision Tree*(Arvore de Decisões)
- *Random Forest*(Floresta Aleatória)
- *Bagging*

4.1.1 Tipo

Foi constado que os algoritmos de *Random Forest*(Floresta Aleatoria), *Decision Tree*(Arvores de Decisão) e *Bagging* tiveram melhor desempenho que o *kNN* em prever os possíveis principais tipos dos dados da ouvidoria para cada mês.

Os algoritmos Random Forest(Floresta Aleatoria), Decision Tree(Arvores de Decisão) e Bagging tiveram resultados muito parecidos, demonstrando eficiência de 55% na predição dos resultados, enquanto *kNN* demonstrou eficiência somente de 38% como mostrado na tabela 2.

Tabela 2 – Porcentagem de acerto por Algoritmo utilizando a classe tipo

Algoritmo	Porcentagem
Bagging	55%
Decision Tree	55%
kNN	38%
Random Forest	55%

Na figura 2 é possível ver o gráfico da probabilidade de cada tipo ocorrer em cada mês utilizando o algoritmo de *Bagging*.

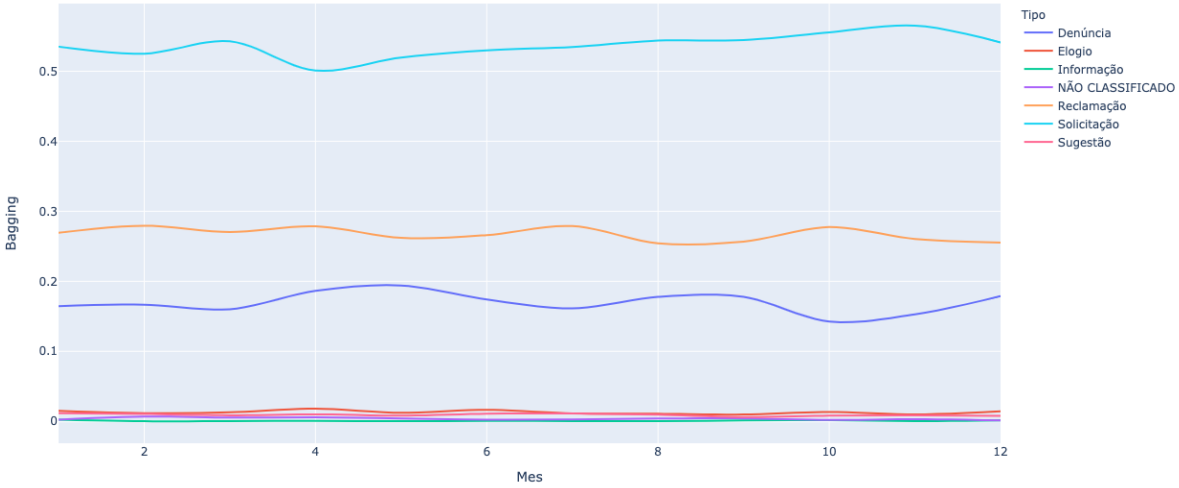

Figura 2 – Probabilidade de cada tipo por mês usando o algoritmo de Bagging.
Fonte: O autor

Na figura 2 o gráfico descreve a probabilidade de cada tipo acontecer por mês utilizando o algoritmo de *Bagging*, sendo o eixo horizontal os meses escritos por numero do mês, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

Na figura 3 é possível visualizar a probabilidade de cada tipo ocorrer por mês tendo uma visão por tipo, utilizando o algoritmo de *Bagging*:

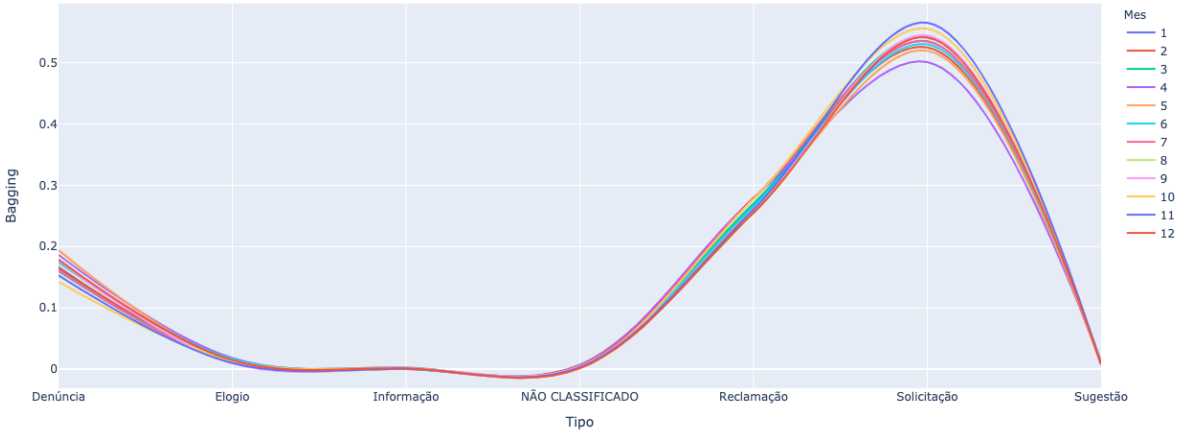

Figura 3 – Probabilidade de cada tipo por mês com visão por tipo usando o algoritmo de Bagging.
Fonte: O autor

Na figura 3 o gráfico descreve a probabilidade de cada tipo acontecer por mês utilizando o algoritmo de *Bagging*, sendo o eixo horizontal os tipos, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

Nas figuras 2 e 3 é possível perceber que solicitação é o tipo mais solicitado a ouvidoria com picos durante o mês de outubro e novembro.

O segundo tipo com mais ocorrência é a reclamação com picos durante o mês de fevereiro.

Os algoritmos de *Random Forest*(Floresta Aleatória) e *Decision Tree*(Arvores de decisão) possuem resultados muito parecidos ao algoritmo de *Bagging* como mostrado nas figuras 4, 5, 6 e 7.

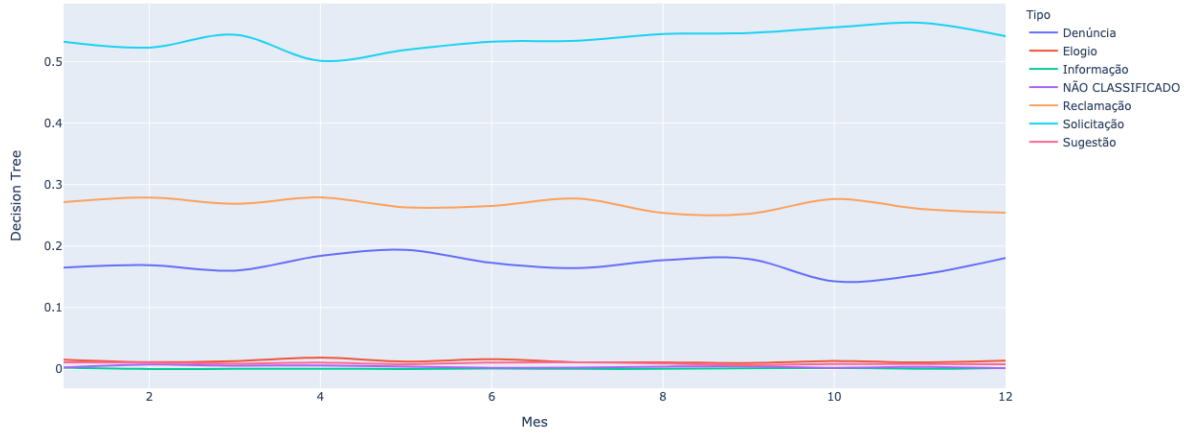

Figura 4 – Probabilidade de cada tipo por mês usando o algoritmo de Decision Tree.
Fonte: O autor

Na figura 4 o gráfico descreve a probabilidade de cada tipo acontecer por mês utilizando o algoritmo de *Decision Tree*, sendo o eixo horizontal os meses escritos por numero do mês, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

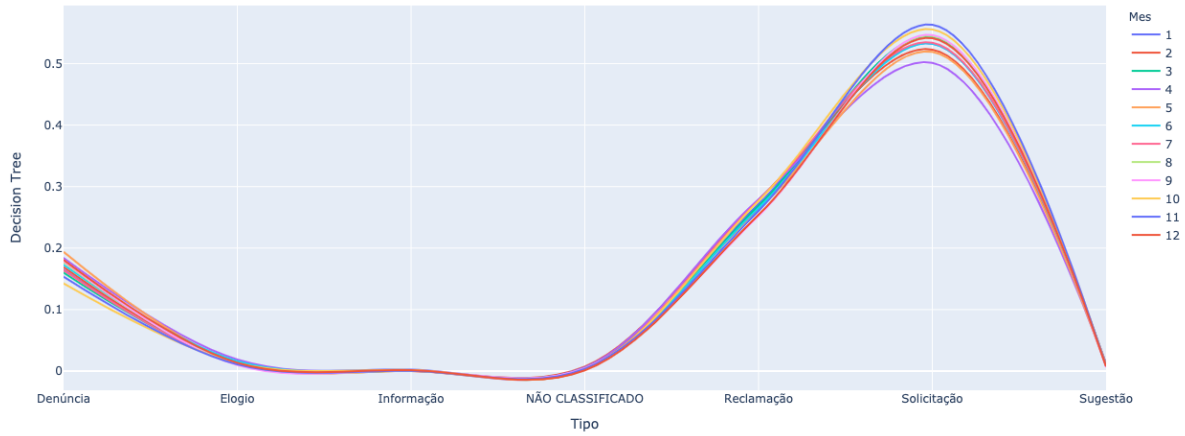

Figura 5 – Probabilidade de cada tipo por mês com visão por tipo usando o algoritmo de Decision Tree.
Fonte: O autor

Na figura 5 o gráfico descreve a probabilidade de cada tipo acontecer por mês utilizando o algoritmo de *Decision Tree*, sendo o eixo horizontal os tipos, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

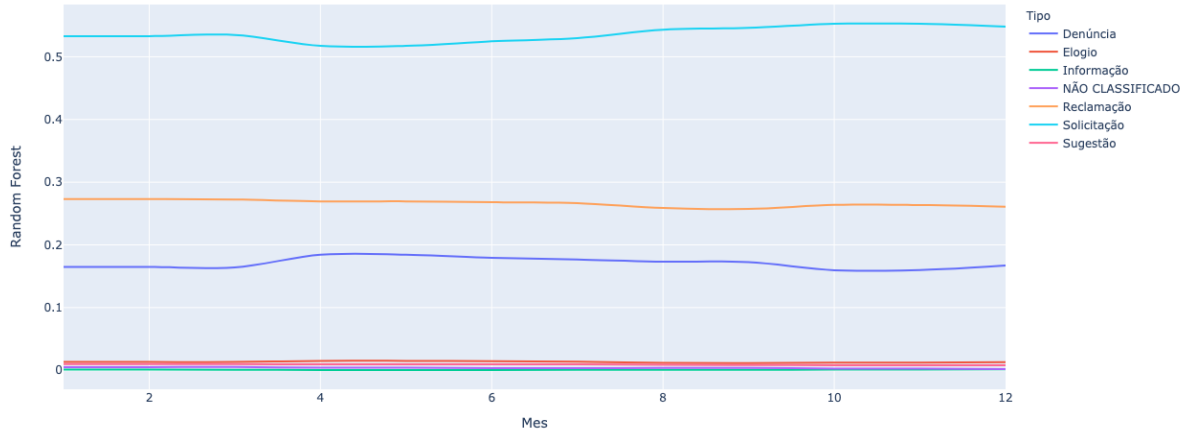

Figura 6 – Probabilidade de cada tipo por mês usando o algoritmo de Random Forest.
Fonte: O autor

Na figura 6 o gráfico descreve a probabilidade de cada tipo acontecer por mês utilizando o algoritmo de *Random Forest*, sendo o eixo horizontal os meses escritos por numero do mês, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

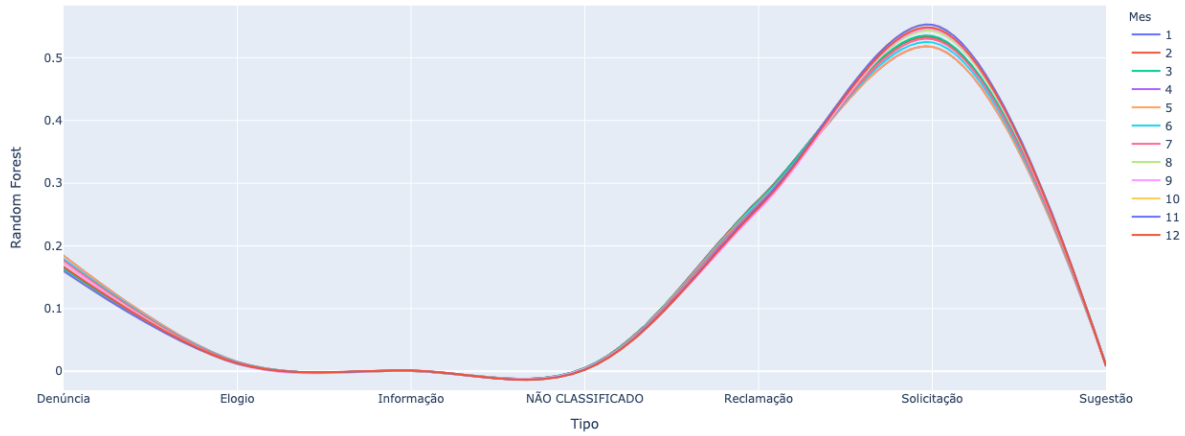

Figura 7 – Probabilidade de cada tipo por mês com visão por tipo usando o algoritmo de Random Forest.
Fonte: O autor

Na figura 7 o gráfico descreve a probabilidade de cada tipo acontecer por mês utilizando o algoritmo de *Random Forest*, sendo o eixo horizontal os tipos, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

O algoritmo *kNN* que obteve a menor porcentagem teve resultado bem diferente dos anteriores, como mostrado nas figuras 8 e 9.

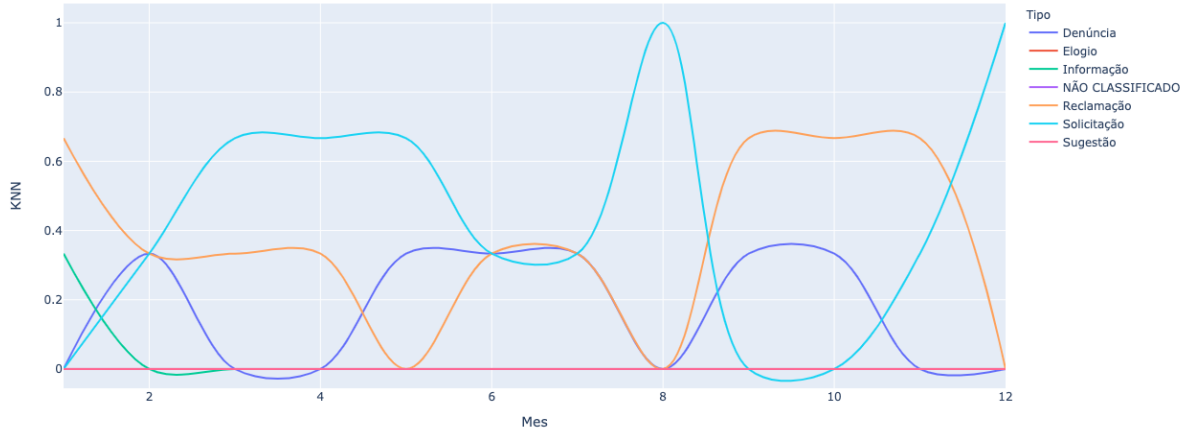

Figura 8 – Probabilidade de cada tipo por mês usando o algoritmo de kNN.
Fonte: O autor

Na figura 8 o gráfico descreve a probabilidade de cada tipo acontecer por mês utilizando o algoritmo de *kNN*, sendo o eixo horizontal os meses escritos por numero do mês, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

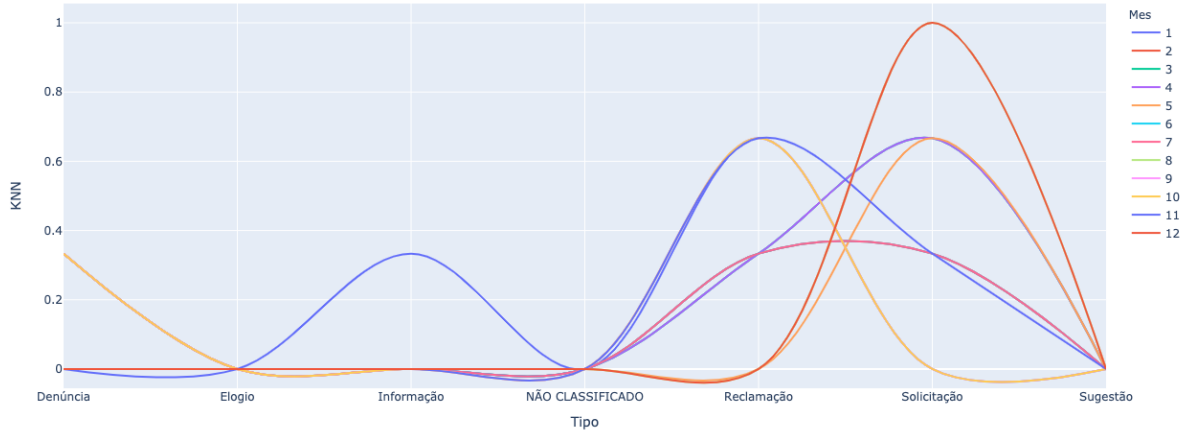

Figura 9 – Probabilidade de cada tipo por mês com visão por tipo usando o algoritmo de kNN.
Fonte: O autor

Na figura 9 o gráfico descreve a probabilidade de cada tipo acontecer por mês utilizando o algoritmo de *kNN*, sendo o eixo horizontal os tipos, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

É possível ver na figura 8 que solicitação continua sendo o tipo mais solicitado a ouvidoria, porém os meses onde esse pico acontece muda para agosto e dezembro.

Na figura 10 é possível ver uma comparação entre todos os algoritmos para a classe de tipo.

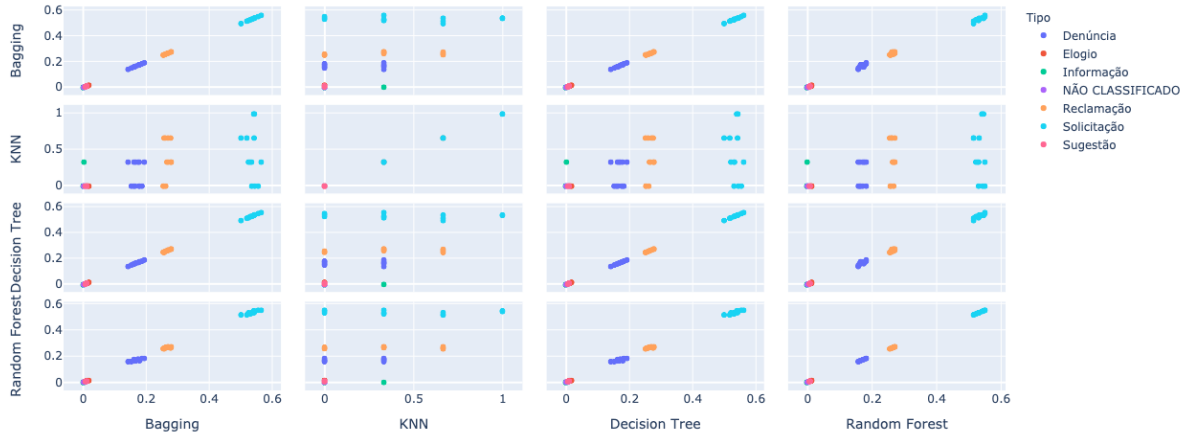

Figura 10 – Comparação de todos os algoritmos usando a classe de tipo.
Fonte: O autor

Na figura 10 o gráfico descreve uma comparação das probabilidades entra cada algoritmo utilizando a classe de tipo.

4.1.2 Orgão

Foi constado que os algoritmos de *Random Forest*(Floresta Aleatória), *Decision Tree*(Arvores de Decisão) e *Bagging* tiveram melhor desempenho que o *kNN* em prever os possíveis principais órgãos dos dados da ouvidoria para cada mês.

Os algoritmos Random Forest(Floresta Aleatória), Decision Tree(Arvores de Decisão) e Bagging tiveram resultados muito parecidos, demonstrando eficiência de 32% na predição dos resultados, enquanto *kNN* demonstrou eficiência somente de 20% como mostrado na tabela 3.

Tabela 3 – Porcentagem de acerto por Algoritmo utilizando a classe órgão

Algoritmo	Porcentagem
Bagging	32%
Decision Tree	32%
kNN	20%
Random Forest	32%

Na figura 11 é possível ver o gráfico da probabilidade de cada tipo ocorrer em cada mês utilizando o algoritmo de *Bagging*.

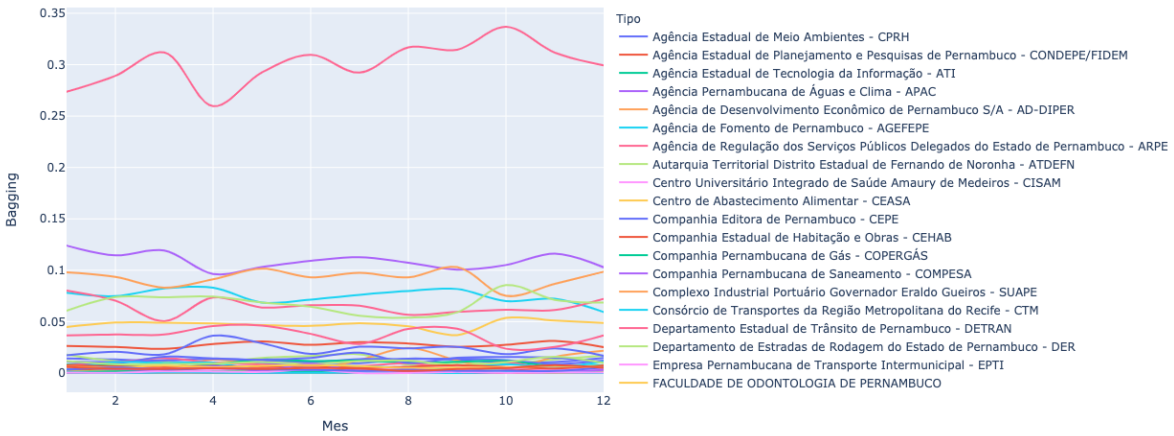

Figura 11 – Probabilidade de cada órgão por mês usando o algoritmo de Bagging.
Fonte: O autor

Na figura 11 o gráfico descreve a probabilidade de cada órgão receber uma solicitação por mês utilizando o algoritmo de *Bagging*, sendo o eixo horizontal os meses escritos por numero do mês, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

Na figura 12 é possível visualizar a probabilidade de cada órgão ocorrer por mês tendo uma visão por tipo, utilizando o algoritmo de *Bagging*:

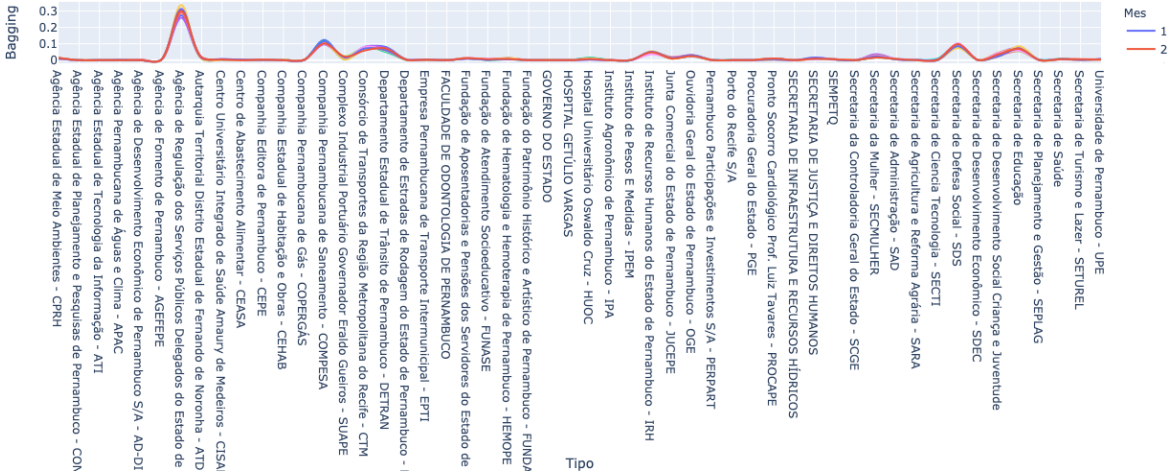

Figura 12 – Probabilidade de cada órgão por mês com visão por órgão usando o algoritmo de Bagging.

Fonte: O autor

Na figura 12 o gráfico descreve a probabilidade de cada órgão receber uma solicitação por mês utilizando o algoritmo de *Bagging*, sendo o eixo horizontal os órgãos, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

Nas figuras 11 e 12 é possível perceber que ARPE(Agência Reguladora de Pernambuco) é o órgão que mais recebe solicitação na ouvidoria com picos durante o mês de agosto e outubro.

O segundo órgão com mais ocorrência é a COMPESA(Companhia Pernambucana de Saneamento) com picos em janeiro e março, sendo superado pela SDS(Secretaria de Defesa Social) em setembro.

Os algoritmos de *Random Forest*(Floresta Aleatória) e *Decision Tree*(Arvores de decisão) possuem resultados muito parecidos ao algoritmo de *Bagging* como mostrado nas figuras 13, 14, 15 e 16.

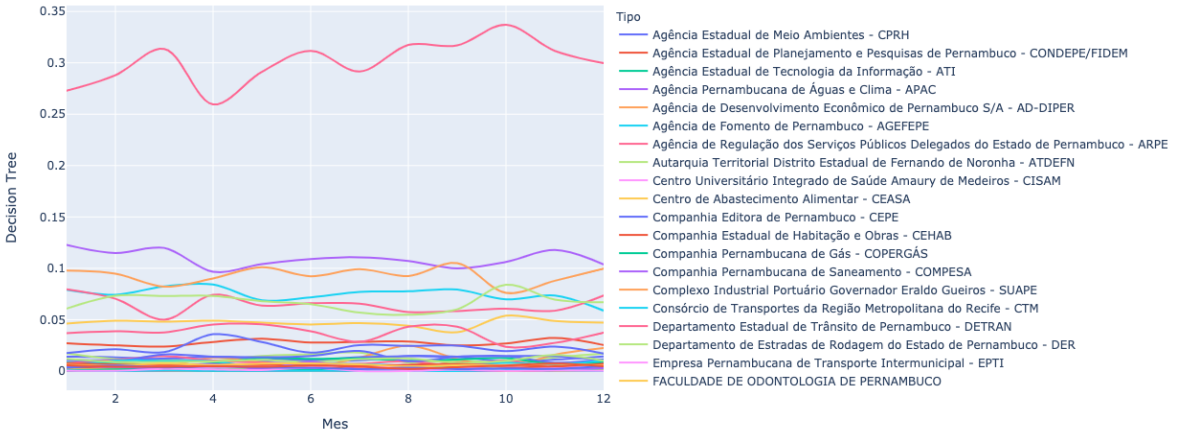

Figura 13 – Probabilidade de cada orgão por mês usando o algoritmo de Decision Tree.
Fonte: O autor

Na figura 13 o gráfico descreve a probabilidade de cada órgão receber uma solicitação por mês utilizando o algoritmo de *Decision Tree*, sendo o eixo horizontal os meses escritos por numero do mês, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

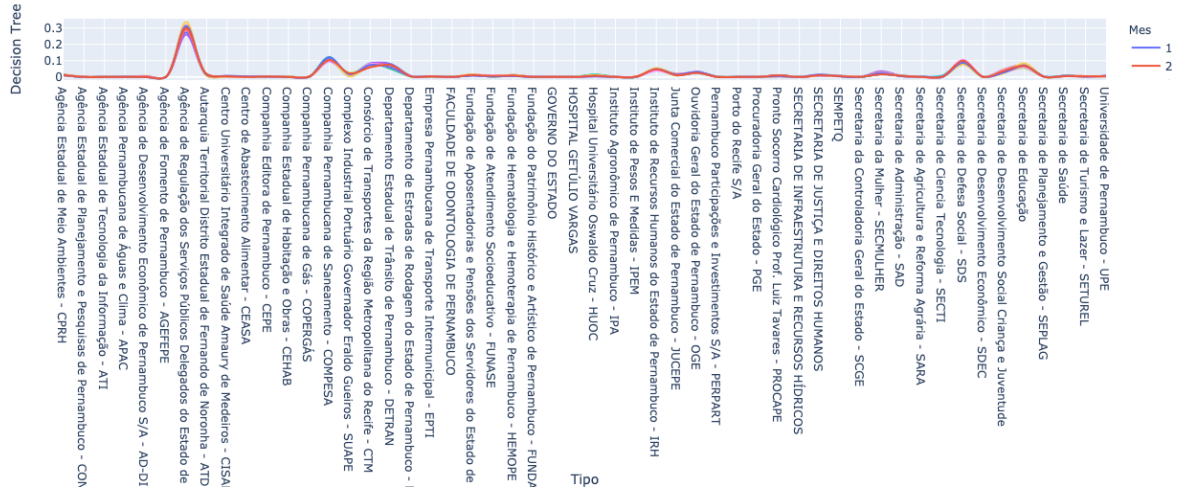

Figura 14 – Probabilidade de cada orgão por mês com visão por orgão usando o algoritmo de Decision Tree.
Fonte: O autor

Na figura 14 o gráfico descreve a probabilidade de cada órgão receber uma solicitação por mês utilizando o algoritmo de *Decision Tree*, sendo o eixo horizontal os órgãos, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

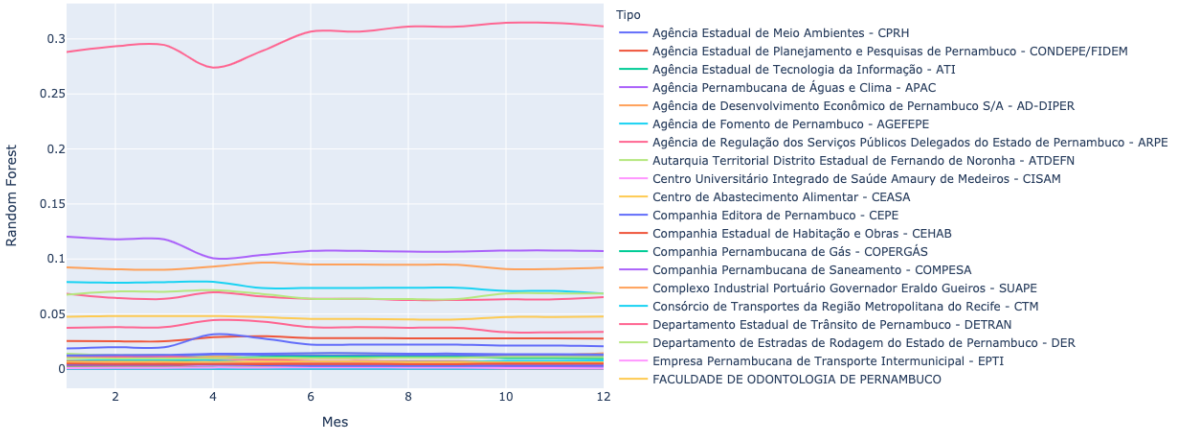

Figura 15 – Probabilidade de cada orgão por mês usando o algoritmo de Random Forest.

Fonte: O autor

Na figura 15 o gráfico descreve a probabilidade de cada órgão receber uma solicitação por mês utilizando o algoritmo de *Random Forest*, sendo o eixo horizontal os meses escritos por numero do mês, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

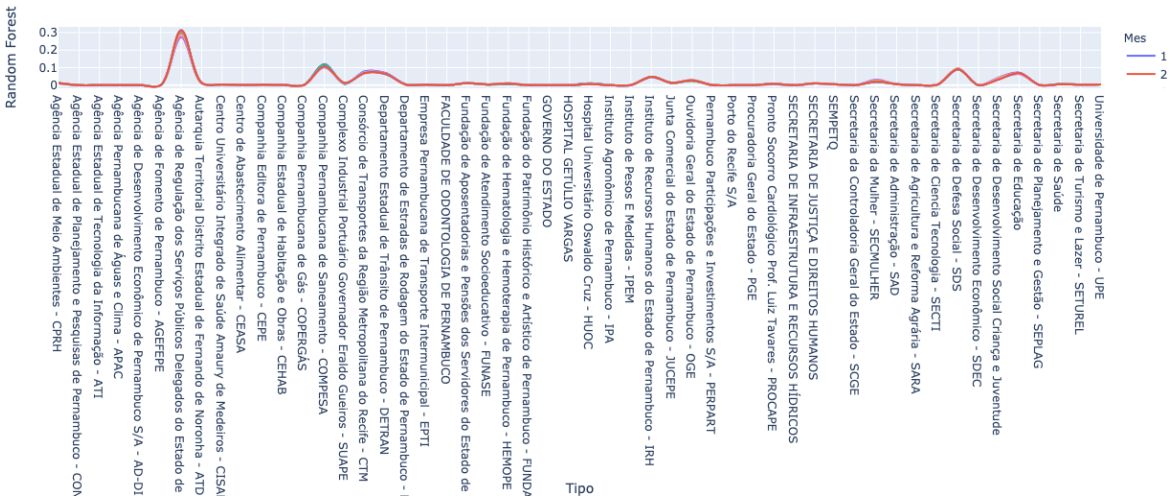

Figura 16 – Probabilidade de cada orgão por mês com visão por orgão usando o algoritmo de Random Forest.

Fonte: O autor

Na figura 16 o gráfico descreve a probabilidade de cada órgão receber uma solicitação por mês utilizando o algoritmo de *Decision Tree*, sendo o eixo horizontal os órgãos, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

Dessa vez o algoritmo *kNN* que obteve a menor porcentagem teve resultado não tão diferentes dos anteriores, como mostrado nas figuras 17 e 18, porém continua tendo linhas de probabilidades diferente dos outros algoritmos, contudo os picos continuaram os mesmos.

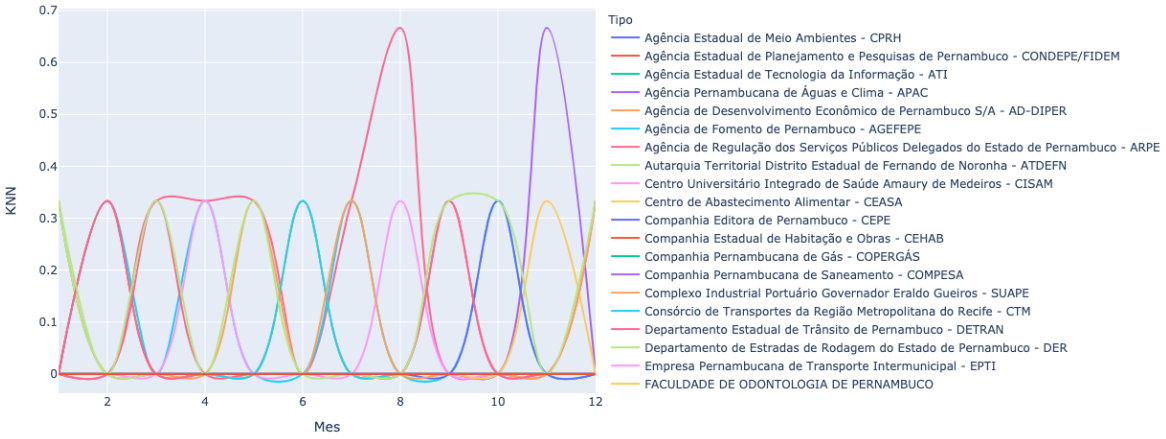

Figura 17 – Probabilidade de cada orgão por mês usando o algoritmo de kNN.
Fonte: O autor

Na figura 17 o gráfico descreve a probabilidade de cada órgão receber uma solicitação por mês utilizando o algoritmo de *kNN*, sendo o eixo horizontal os meses escritos por numero do mês, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

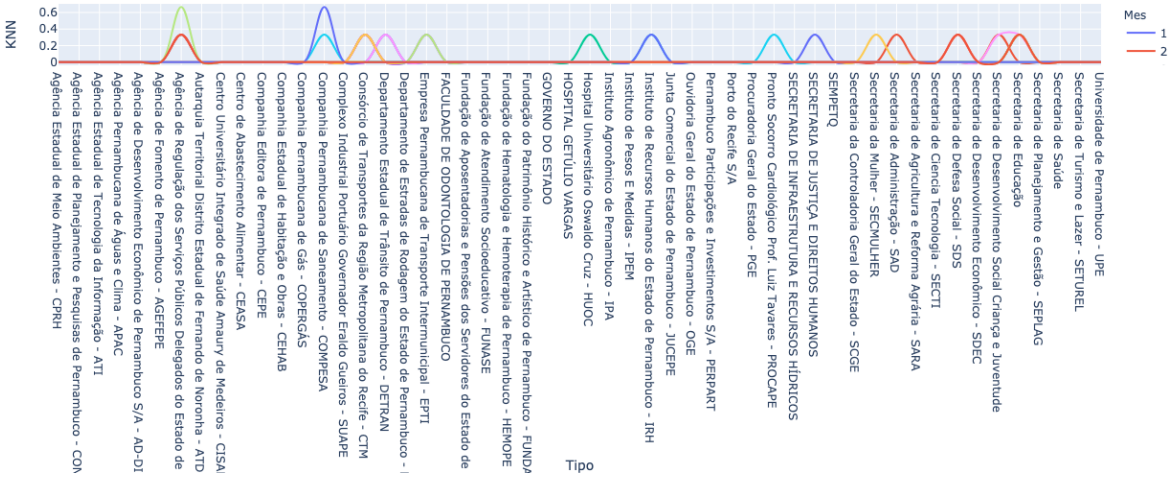

Figura 18 – Probabilidade de cada orgão por mês com visão por orgão usando o algoritmo de kNN.
Fonte: O autor

Na figura 18 o gráfico descreve a probabilidade de cada órgão receber uma

solicitação por mês utilizando o algoritmo de *kNN*, sendo o eixo horizontal os órgãos, e o eixo vertical a probabilidade em porcentagem/100 (por exemplo 0.5 equivale a 50%)

É possível ver na figura 17 que ARPE(Agência Reguladora de Pernambuco) continua sendo o tipo mais solicitado a ouvidoria, porém os meses onde esse pico acontece muda para agosto.

Na figura 19 é possível ver uma comparação entre todos os algoritmos para a classe de tipo.

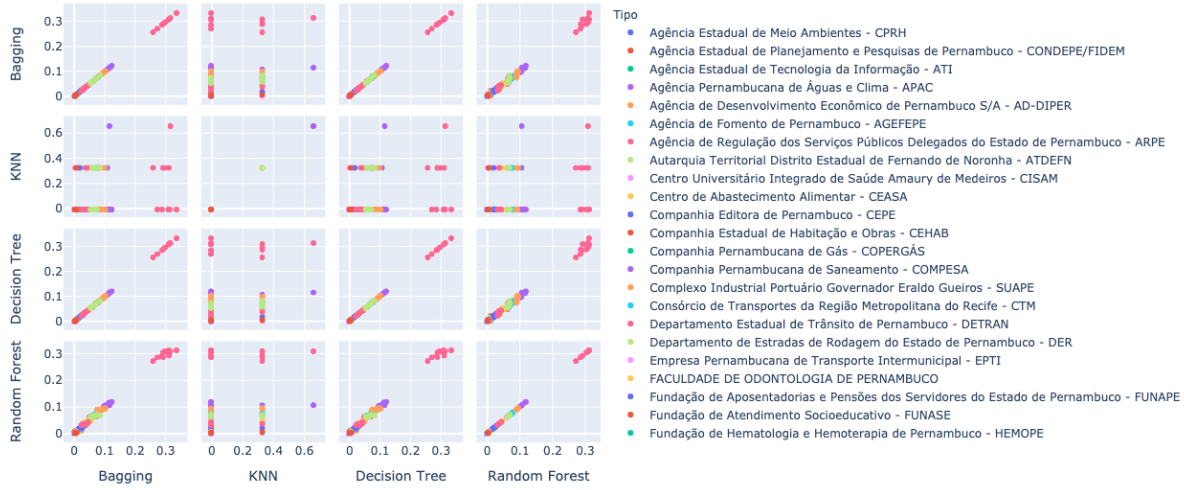

Figura 19 – Comparação de todos os algoritmos usando a classe de orgão.
Fonte: O autor

Na figura 19 o gráfico descreve uma comparação das probabilidades entra cada algoritmo utilizando a classe de órgão.

4.2 Desempenho

Além da analisar a precisão e probabilidade dos modelos de cada algoritmo, foi analisado também o tempo de criação e treinamento para cada modelo.

É possível analisar na tabela 4 que o algoritmo do *Random Forest*(Floresta Aleatória) é o algoritmo que tem o maior tempo de criação e treinamento sendo o pior nesse quesito.

O algoritmo com o melhor tempo de criação e treinamento é o *Decision Tree*(Arvore de decisões), como mostrado na tabela 4

Tabela 4 – Tempo de criação e treinamento por modelo em segundos

Algoritmo	Máximo	Mínimo	Média
Bagging	0.46	0.39	0.41
Decision Tree	0.18	0.16	0.16
kNN	1.33	1.29	1.30
Random Forest	2.02	1.64	1.78

4.3 Discussão

Após a analise dos resultados para cada algoritmo, foi observado que os algoritmos de *Decision Tree*(Arvore de decisões), *Random Forest*(Floresta Aleatória) e *Bagging* possuem resultado iguais, porém o algoritmo de Decision Tree(Arvore de decisões) foi mais performático em relação a criação do modelo e treinamento, como mostrado na tabela 4.

O algoritmo de *kNN* foi o menos preciso em relação aos outros algoritmos e seu tempo de criação e treinamento foi o segundo pior perdendo somente para o *Random Forest*(Floresta Aleatória), como mostrado na tabela 4.

Sobre os resultados das probabilidades, foi possível encontrar padrões entre as series temporais de mês e as classes de Tipo e Órgão, como mostrado no seção 4.1.

5 Conclusão

O presente trabalho promoveu o ensaio em relação a assertividade dos algoritmos de classificação automática de dados multi-classes, utilizando os dados da Ouvidoria Geral do Estado de Pernambuco. A avaliação teve foco em tentar prever as possíveis classe baseado na serie temporal de mês.

Os resultados mostraram que os algoritmos de classificação automática de dados, particularmente os algoritmos de *Decision Tree*(Arvore de decisões), *Random Forest*(Floresta Aleatória) e *Bagging* conseguiram de 55% e 32% nas classes de tipo e órgão respectivamente, tendo um aproveitamento de um acerto a cada duas tentativas na classe de tipo e de um acerto a cada três tentativas na classe de órgão.

Os algoritmos também foram avaliados acerca de seu desempenho em tempo de criação e treinamento do modelo, tendo o algoritmo de *Decision Tree*(Arvore de decisões) como o mais performático.

Contudo, deve-se apontar que, devido a natureza de um trabalho de conclusão de curso, as análises fazem um recorte espacial e temporal bastante limitado. Desta forma, deve-se ressaltar que as conclusões aqui obtidas não podem ser generalizadas de forma simples, devendo este resultado ser corroborado ou contradito por trabalhos futuros que realizem avaliações de maior alcance.

Também deve se levar em consideração as limitações técnicas encontradas, tais como as limitações de recursos computacionais.

Devido aos resultados alcançados neste trabalho, um novo experimento utilizando os dados da Ouvidoria Geral do Estado de Pernambuco será efetuado futuramente. Também sera avaliado o desempenho de outros algoritmos de classificação automática de dados.

Referências

AHA, D. W.; KIBLER, D.; ALBERT, M. K. Instance-based learning algorithms. *Machine learning*, Springer, v. 6, n. 1, p. 37–66, 1991.

ALMEIDA, E. D. *Algoritmos de Classificação Com a Opção de Rejeição*. Tese (Doutorado) — Dissertação de Mestrado) Porto: Faculdade De Engenharia Da Universidade Do Porto, 2010.

ALPAYDIN, E. *Introduction to machine learning*. [S.l.]: MIT press, 2020.

AMO, S. de. Técnicas de mineração de dados. *Jornada de Atualizaçao em Informatica*, 2004.

CAMILO, C. O.; SILVA, J. C. d. Mineração de dados: Conceitos, tarefas, métodos e ferramentas. *Universidade Federal de Goiás (UFC)*, p. 1–29, 2009.

COSTA, J. A. F.; BITTENCOURT, V. G.; SOUTO, M. C. de. Aplicação de multi-classificadores no reconhecimento de classes estruturais de proteínas. In: *Anais do Congresso Nacional de Matemática Aplicada e Computacional (CNMAC)*. [S.l.: s.n.], 2005. v. 6.

COSTA, J. A. F. et al. Classificação automática e análise de dados por redes neurais auto-organizáveis. [sn], 1999.

DELGADO, R. C. et al. Classificação espectral de área plantada com a cultura da cana-de-açúcar por meio da árvore de decisão. *Engenharia Agrícola*, SciELO Brasil, v. 32, n. 2, p. 369–380, 2012.

EHLERS, R. S. Análise de séries temporais. *Laboratório de Estatística e Geoinformação. Universidade Federal do Paraná*, v. 1, p. 1–118, 2007.

FERRERO, C. A. *Algoritmo kNN para previsão de dados temporais: funções de previsão e critérios de seleção de vizinhos próximos aplicados a variáveis ambientais em limnologia*. Tese (Doutorado) — Universidade de São Paulo, 2009.

HAN, J.; PEI, J.; KAMBER, M. *Data mining: concepts and techniques*. [S.l.]: Elsevier, 2011.

HAND, D. J.; ADAMS, N. M. Data mining. *Wiley StatsRef: Statistics Reference Online*, Wiley Online Library, p. 1–7, 2014.

HAYKIN, S. *Redes neurais: princípios e prática*. [S.l.]: Bookman Editora, 2007.

IASBECK, L. C. A. Ouvidoria é comunicação. *Organicom*, v. 7, n. 12, p. 14–24, 2010.

ISOTANI, S.; BITTENCOURT, I. I. *Dados Abertos Conectados: Em busca da Web do Conhecimento*. [S.l.]: Novatec Editora, 2015.

LECUN, Y.; BENGIO, Y.; HINTON, G. Deep learning. *nature*, Nature Publishing Group, v. 521, n. 7553, p. 436–444, 2015.

MITCHELL, R. *Web scraping with Python: Collecting more data from the modern web*. [S.l.]: ”O’Reilly Media, Inc.”, 2018.

NILSSON, N. J. *Principles of artificial intelligence*. [S.l.]: Morgan Kaufmann, 2014.

RIQUETI, G.; RIBEIRO, C.; ZÁRATE, L. Classificando perfis de longevidade de bases de dados longitudinais usando floresta aleatória. 2018.

ROMANO, R. T. *ÓRGÃO PÚBLICO: CONCEITO E RELAÇÃO ENTRE ELES*. 2020. Disponível em: <https://jus.com.br/artigos/54248/orgao-publico-conceito-e-relacao-entre-eles>.

SANTOS, V. Avaliacao de um classificador multiclasse com abordagem one-vs one usando diferentes classificadores binarios. 2017.

SARITAS, M. M.; YASAR, A. Performance analysis of ann and naive bayes classification algorithm for data classification. *International Journal of Intelligent Systems and Applications in Engineering*, v. 7, n. 2, p. 88–91, 2019.

SCGE. *Ouvidoria*. 2020. Disponível em: <https://www.scge.pe.gov.br/ouvidoria-geral/>.

TEIXEIRA, J. *O que é inteligência artificial*. [S.l.]: E-Galáxia, 2019.

TRANSPARENCIA, P. da. *Portal da Transparencia*. 2020. Disponível em: <http://web.transparencia.pe.gov.br/>.

VIEIRA, R.; LOPES, L. Processamento de linguagem natural e o tratamento computacional de linguagens científicas. *EM CORPORA*, p. 183, 2010.