



PROGRAMA DE  
RESIDÊNCIA EM  
**INTELIGÊNCIA  
ARTIFICIAL**

# SUMÁRIO

O SENAI PARANÁ	03
O SENAI TECNOLOGIA E INOVAÇÃO	04
O HUB DE INTELIGÊNCIA ARTIFICIAL SENAI	05
O PROGRAMA DE RESIDÊNCIA EM INTELIGÊNCIA ARTIFICIAL	06
EMPRESAS PARTICIPANTES	08
CASES DO PROGRAMA	10

# O SENAI PARANÁ

3

SUMÁRIO



Juntamente com o Sistema Fiep, o Senai Paraná ajuda a elevar a competitividade da indústria brasileira e a melhorar a vida das pessoas. Com dezenas de unidades distribuídas pelo estado, oferece soluções completas que vão de consultorias a pesquisas e serviços de metrologia. Além disso, o Senai Paraná promove a educação tecnológica, por meio de cursos técnicos, cursos rápidos e educação superior, atendendo às necessidades dos alunos e também da indústria.

# O SENAI TECNOLOGIA E INOVAÇÃO

4

SUMÁRIO



Com qualidade nos serviços, gestão transformadora e soluções flexíveis, o setor de Tecnologia e Inovação do Senai Paraná facilita a transição das indústrias para a era da inovação, ofertando apoio tecnológico e metodologias estruturadas para produzir melhor e com mais qualidade. E com o suporte dos Institutos Senai de Tecnologia e Inovação, as indústrias contam com modernos laboratórios para ensaios laboratoriais e serviços de certificação, consultorias especializadas, parcerias em programas de fomento, pesquisa e desenvolvimento de novos produtos ou serviços, tendo à sua disposição uma equipe preparada para encarar grandes desafios.

# O HUB DE INTELIGÊNCIA ARTIFICIAL SENAI

5

SUMÁRIO



O Hub de Inteligência Artificial do Senai Paraná permite a adoção de novas tecnologias pelo setor industrial paranaense e brasileiro, proporcionando agilidade, segurança, redução de custos, aumento da produtividade e aperfeiçoamento de produtos e serviços.

Entre os pilares de atuação do HUB, destacam-se: Inovação Corporativa (conexão entre corporações e startups), Assistente Industrial (digitalização dos processos de manufatura para a captura de dados e desenvolvimento de Inteligência Artificial) e Residência em Inteligência Artificial (formação de profissionais qualificados para o futuro da indústria).

# O PROGRAMA DE RESIDÊNCIA EM INTELIGÊNCIA ARTIFICIAL

6

SUMÁRIO



O Programa de Residência em Inteligência Artificial, impulsionado pelo Hub de Inteligência Artificial do Senai Paraná, traz um formato teórico-prático, que proporciona formação de mão de obra especializada nas tecnologias de inteligência artificial, ao mesmo tempo em que desenvolve projetos práticos para empresas patrocinadoras.

São selecionados profissionais da área de exatas que queiram se especializar em Inteligência Artificial. Os aprovados têm dedicação exclusiva de um ano para o programa, com uma bolsa-auxílio e diplomação de especialista pelo programa de pós-graduação do Sistema Fiep.



A parceria acadêmica da 2ª e da 3ª turma ocorreu com o Instituto Avançado de Inteligência Artificial, e as mentorias para o desenvolvimento técnico aconteceram com a equipe de consultores do Hub de Inteligência Artificial. O programa proporciona grande interação entre o grupo de empresas patrocinadoras e os residentes, potencializando a empregabilidade após a graduação.

7

SUMÁRIO



As empresas patrocinadoras recebem até 4 provas de conceito (POCs) ao longo do programa, podendo ser uma por trimestre por cota de patrocínio. As POCs estão relacionadas ao conteúdo programático do Programa de Residência, validando a aplicação de técnicas de IA para solucionar os problemas apresentados pela empresa. As soluções não são produtos acabados, mas dependendo das reuniões de alinhamento e mapeamento do atendimento, durante o programa podem atingir o status de mínimo produto viável (MVP). Todos os códigos desenvolvidos são de propriedade das patrocinadoras e todo o desenvolvimento está protegido por termo de confidencialidade.

# EMPRESAS PARTICIPANTES

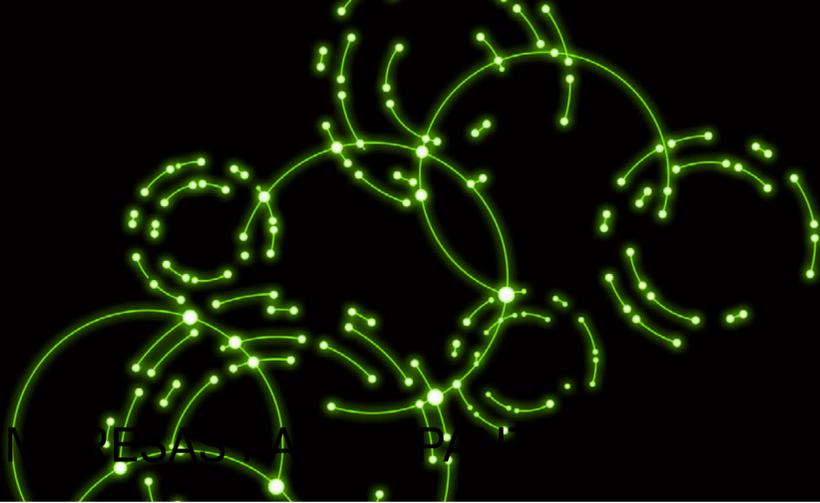
8

SUMÁRIO



A 2ª e a 3ª turma de Residência em Inteligência Artificial do Senai atingiram entregas de alto nível para as empresas patrocinadoras, que você confere ao longo deste material. Serão apresentadas as temáticas dos desafios trabalhados e as aplicações de Inteligência Artificial, demonstrando os resultados alcançados com as provas de conceito e o potencial de retorno de investimento das seguintes empresas:







# CASES DO PROGRAMA

10

SUMÁRIO



O programa teve 59 demandas desenvolvidas em um ano. Aqui iremos apresentar alguns destes projetos. Eles foram selecionados, pois podem ser divulgados por não violarem o termo de confidencialidade. Confira a seguir:

## **AKIYAMA**

CROSS-CHECKING DE BASE CADASTRAL 19

---

AGILIZAÇÃO NO PROCESSO DE BUSCA DE IMPRESSÕES DIGITAIS 21

---

PRÉ-PROCESSAMENTO PARA MELHORIA DE ACURÁCIA EM MATCHER FACIAL 23

---

## BOSCH

PREDIÇÃO DE REFUGO DAS PEÇAS BASEADO NO COMPORTAMENTO E EXECUÇÃO DA USINAGEM	25
--	----

11

## ELECTROLUX

ESTIMATIVA DE CONTRATAÇÃO DE FROTA DE CAMINHÕES PARA ESCOAMENTO DE PRODUTOS	28
---	----

SUMÁRIO



PREVISÃO DE UM MIX ÓTIMO PARA AS LINHAS DE PRODUÇÃO	30
--	----

SUMÁRIO  
CASES DO  
PROGRAMA

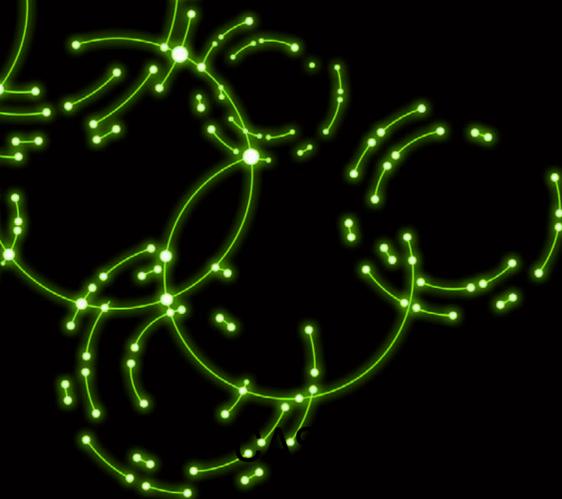


ANÁLISE DE CADÊNCIAS NAS LINHAS DE PRODUÇÃO	32
--	----

PREVISÃO DE UM MIX ÓTIMO PARA AS LINHAS DE PRODUÇÃO	34
--	----

## EMBRAPA

AUTOMAÇÃO DA METODOLOGIA DRES: DIAGNÓSTICO RÁPIDO DA ESTRUTURA DO SOLO	36
--	----



MONITORAMENTO DE MÍDIAS  
SOCIAIS 38

---

IDENTIFICAÇÃO DA INFESTAÇÃO DE  
PERCEVEJOS 41

---

CLASSIFICADOR DAS FASES DA  
MOSCA BRANCA 44

---

## **BELLAGRÍCOLA FARMBITS**

ANÁLISE DO PERFIL DOS CLIENTES  
1 47

---

ANÁLISE DO PERFIL DOS CLIENTES  
2 50

---

SISTEMA DE RECOMENDAÇÃO  
PARA IMPULSIONAR GERAÇÃO DE  
DEMANDA 53

---

PREVISÃO DE DEMANDA E  
PLANEJAMENTO COMERCIAL 56

---

12

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





---

## HEDRO

PREVISÃO DE DEMANDA E PLANEJAMENTO COMERCIAL 58

---

## HS TECHNOLOGY

CALIBRAÇÃO BAROSCAN 60

---

## INSTITUTO DE DESENVOLVIMENTO RURAL DO PARANÁ

ANÁLISE DE DADOS AGROMETEOROLÓGICOS 65

---

ANÁLISE DE ÍNDICE DE ELASTICIDADE PARA A CULTURA DO FEIJÃO 68

---

CHATBOT TELEGRAM 71

---

DETECÇÃO DE INSETOS BASEADO EM VÍDEOS E IMAGENS 73

---

13

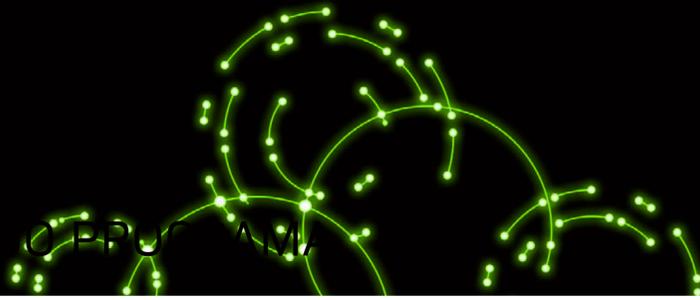
---

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





## COOPERATIVA INTEGRADA

PREVISÃO DE CAIXA PARA FIXAÇÃO  
DE SOJA 75

14

PREVISÃO DE ENTRADA EM PESO  
NOS ESTABELECIMENTOS 78

SUMÁRIO



CLASSIFICAÇÃO DE ÁREAS  
AGRICULTURÁVEIS POR SATÉLITE 81

SUMÁRIO  
CASES DO  
PROGRAMA



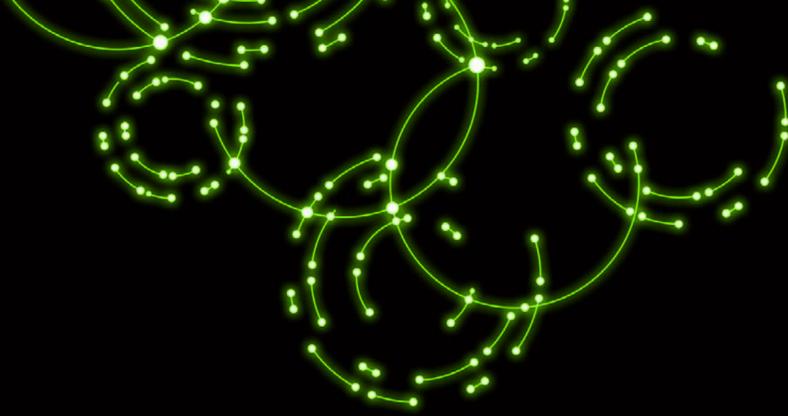
## NATOSAFE

ANÁLISE DE DIGITAIS DE RECÉM-  
NASCIDOS 84

MODELO GERADOR DE DIGITAIS  
ARTIFICIAIS E ALINHAMENTO DE  
DIGITAIS DISTORCIDAS 87

## OLIST

ANÁLISE EXPLORATÓRIA DOS  
DADOS DE VENDEDORES E  
PRODUTOS 90



---

PREVISÃO DO ÍNDICE DE REPUTAÇÃO NO MERCADO LIVRE 93

---

MINERAÇÃO DE ATRIBUTOS DAS DESCRIÇÕES DE ANÚNCIOS DE PRODUTOS 96

---

RECONHECIMENTO DE IMAGENS PARA CATÁLOGO 98

---

## **PADO**

ANÁLISE DE CRÉDITO E LIBERAÇÃO DE PEDIDO 100

---

LEITURA DE DOCUMENTOS 101

---

CONTAGEM DE MATERIAIS 103

---

IDENTIFICAÇÃO DE SEGREDO DE CHAVES 109

---

## **PRIDE**

REFATORAÇÃO DE CHATBOT 112

---

APRIMORAMENTO DE CHATBOT 115

---

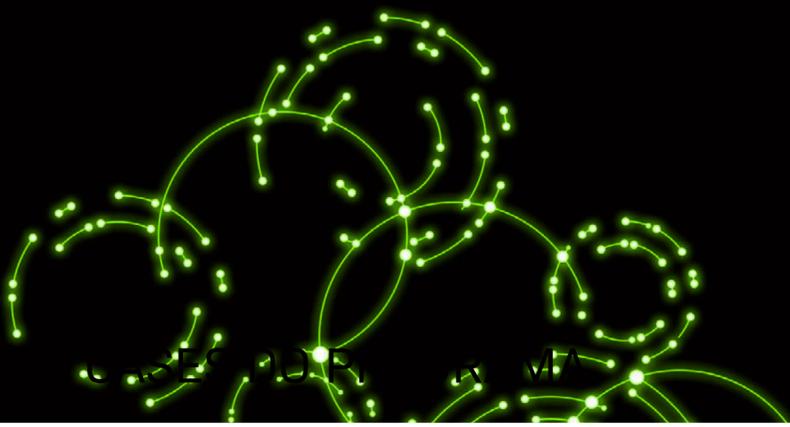
15

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





---

CHATBOT MULTIPLATAFORMA 118

---

REFATORAÇÃO DE CHATBOT 120

---

## **QUALITYSTORM**

16

---

API DE INSPEÇÃO DE IMAGENS E PROCESSOS 122

---

## **SICOOB**

---

PREVISÃO DE INADIMPLÊNCIA EM CONTAS DE PESSOA JURÍDICA E RURAL 125

---

## **TARS**

---

VALIDAÇÃO AUTOMÁTICA DE DOCUMENTOS ATRAVÉS DE OCR 130

---

ANÁLISE DE CONTRATOS 133

---

DETECÇÃO DE VAZAMENTO DE DADOS EM TEMPO REAL 136

---

DETECÇÃO DE VAZAMENTO DE INFORMAÇÕES CONFIDENCIAIS 140

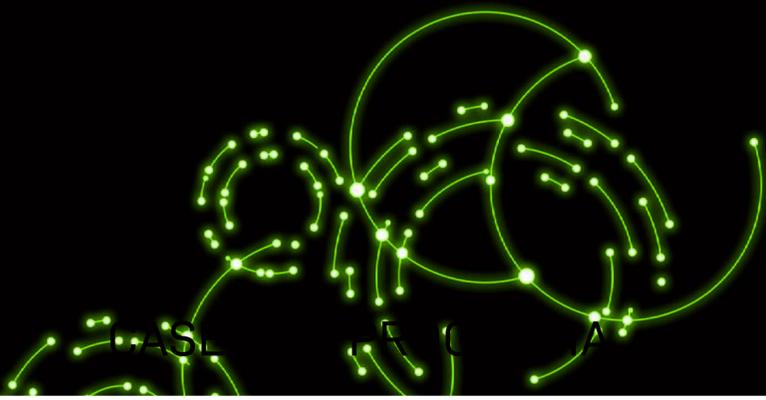
---

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





---

## TÉRMICA SOLUTIONS

SIMULAÇÃO E OTIMIZAÇÃO DE OPERAÇÕES EM TERMOELÉTRICA 143

---

## TMG - TROPICAL MELHORAMENTOS E GENÉTICA

AQUISIÇÃO DE INFORMAÇÕES SOBRE NOVOS CULTIVARES 149

---

AValiação DA INFLUÊNCIA DE FATORES CLIMÁTICOS E DO SOLO NA PRODUÇÃO DE SOJA 151

---

CATEGORIZAÇÃO DE CHAMADOS 154

---

## TRIMBLE

ANÁLISE DE EVENTOS BRUSCOS DE ACORDO COM A JORNADA DE MOTORISTAS 157

---

IDENTIFICAÇÃO DE MOTORISTAS SEM VIGÊNCIA 160

---

17

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



---

ANÁLISE DE CONSUMO DE COMBUSTÍVEIS (VELTEC) 163

---

DETECÇÃO DE SONOLÊNCIA AO VOLANTE (VELTEC) 165

---

## **VIASOFT**

MELHOR CONFIGURAÇÃO DE CAMINHÃO TANQUE 168

---

SISTEMA DE RECOMENDAÇÃO 170

---

AUTOMATIZAÇÃO DE BUSCA DE PREÇOS DE PRODUTOS DO SITE MENOR PREÇO PARANÁ 172

---

SCRAPING E CLASSIFICAÇÃO DE NOTÍCIAS 174

---

## **VOTORANTIM**

PREDIÇÃO DE ANOMALIAS 176

---

CLASSIFICAÇÃO DE REGISTROS EM POTENCIAL DE GRAVIDADE 178

---

18

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## AKIYAMA

1

## CROSS-CHECKING DE BASE CADASTRAL

**Residentes:**

Larissa Ferreira, Marcia Rodrigues, Victor Nagahama.

**Técnicas:**

Cross-checking, web-scraping, crawler, scheduler, API.

19

SUMÁRIO

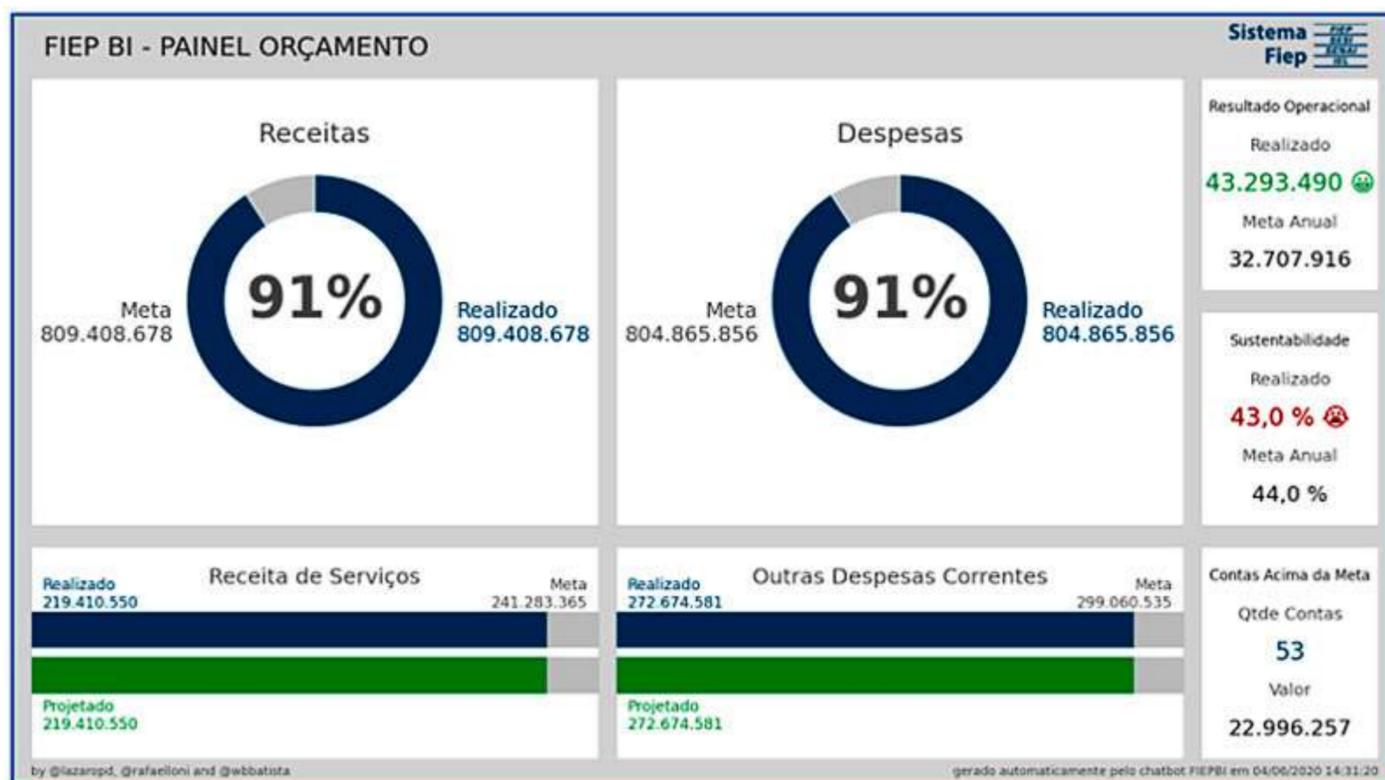
SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

A comparação entre bases de cadastro e dados públicos tem como intuito a redução de riscos a fraudes e prover maior confiabilidade nas informações fornecidas em cadastros. Sendo assim, foi desenvolvida uma ferramenta que extrai dados públicos e analisa a similaridade entre os dados de cadastro e os dados públicos da web.

**SOLUÇÃO DESENVOLVIDA**

A solução foi apresentada no formato de POC (Prova de Conceito) e usa como base

a comparação de informações entre a base cadastral do Facebook, Escavador e LinkedIn. Como resultado foi construída uma aplicação web utilizando a biblioteca streamlit, de forma que o usuário consiga selecionar o indivíduo e as redes sociais a serem utilizadas para o cross checking. Como resposta a aplicação retorna as tabelas de similaridade de cada rede social que foi solicitada na análise. Ao iniciar a aplicação, a base cadastral e a lista de redes sociais disponíveis aparecem para que o usuário possa definir os parâmetros de busca.





## AKIYAMA

2

## AGILIZAÇÃO NO PROCESSO DE BUSCA DE IMPRESSÕES DIGITAIS

**Residentes:**

Gabriel Gomes de Sousa, Karen Itsuki Taniguchi, Natalia da Silva de Sá.

**Técnicas:**

K-Means com Autoencoder Convolutacional, Redes Neurais Convolutacionais, K-Means com extração de minúscias e K-Means com Metric Embedding.

21

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Com a ampla utilização das digitais há também um aumento na base de dados, o que torna o Sistema de Identificação Automática de Digitais (AFIS) cada vez mais custoso e difícil. Busca-se então a redução do número de consultas necessárias para efetuar a busca de digitais em um banco de dados, reduzindo o tempo de busca e mantendo a assertividade inicial de reconhecimento.



## SOLUÇÃO DESENVOLVIDA

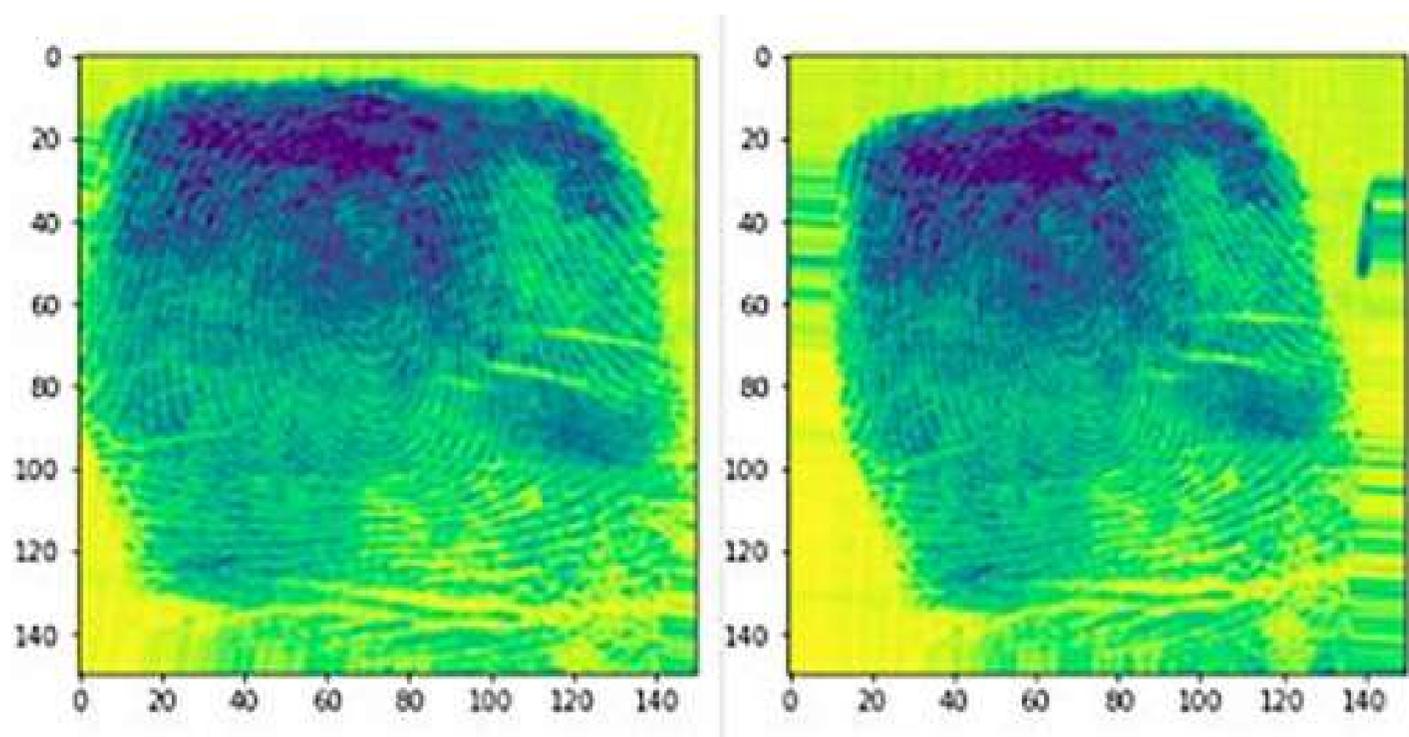
Como resultado obteve-se uma métrica para todos os modelos avaliando o número de digitais iguais (da mesma pessoa, porém em leituras diferentes); o treino de um modelo que separe toda a base de dados em classes, sendo elas separadas de acordo com três padrões visuais das digitais: loop, espiral e arco; a divisão em dois grupos utilizados para a identificação e classificação de impressões digitais; e a substituição de uma métrica extremamente difícil, que é o grau de ‘semelhança visual’ entre digitais, por uma mais simples, que é ‘a distância euclidiana entre vetores’.

22

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





## AKIYAMA

3

## PRÉ-PROCESSAMENTO PARA MELHORIA DE ACURÁCIA EM MATCHER FACIAL

**Residentes:**

Germano Yoneda, Tiago Pedrosa.

**Técnicas:**

Deep Face Align, Cosine Align, Face Recognition, DTC, WAV D8.

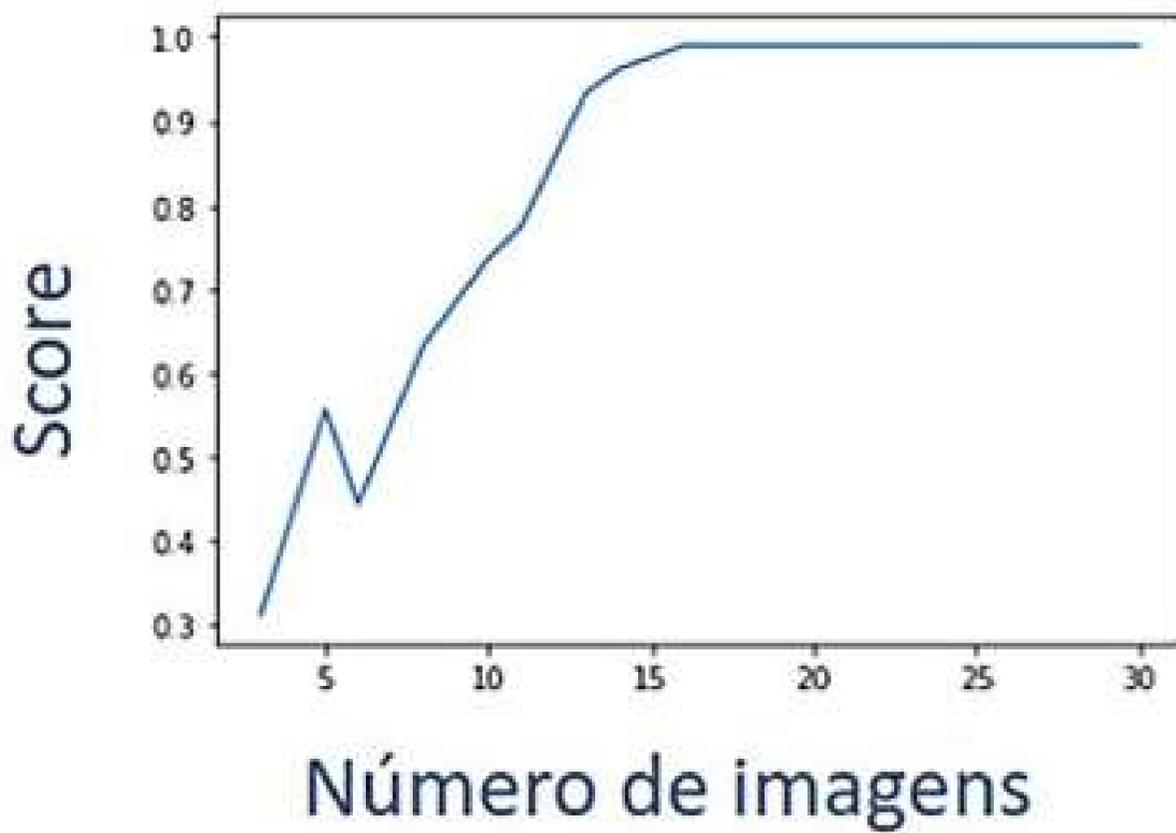
23

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**SOLUÇÃO DESENVOLVIDA**

De acordo com as técnicas escolhidas, respectivamente, temos o alinhamento utilizando a biblioteca Deep Face sem nenhum tipo de alteração adicional e utilizando um recorte da área ao redor do rosto; um método baseado na localização de nariz e olhos, bem como a face, e com base nisso gera o ponto da rotação da imagem; o alinhador facial utilizando a biblioteca Face Recognition juntamente com um recorte ao redor do rosto e a aplicação de um SVC para identificação de pessoas; redução discreta do cosseno para redução de informação e

transformação utilizando Daubechies wavelet (D8). Baseado nesses métodos obtêm-se a demonstração do acréscimo do score de acordo com o aumento do número de imagens usadas para treino.



## BOSCH

1

PREDIÇÃO DE REFUGO DAS PEÇAS BASEADO NO  
COMPORTAMENTO E EXECUÇÃO DA USINAGEM**Residentes 1:**

Ana Luisa Schimidt, André Luiz Marasca, João Victor Kawazoe Nahra.

**Residentes 2:**

André Luiz Marasca, Igor Moura, Lucas Coutinho.

**Residentes 3:**

Ana Luisa Schimidt, Lucas Coutinho.

**Técnicas:**

Data Science e Deep Learning.

25

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

O Bico Injetor é a peça responsável por regular o fluxo de combustível injetado no motor. O bom funcionamento desta peça, está atrelado à precisão e qualidade das dimensões dos furos. O processo de usinagem responsável pela fabricação dos bicos depende, dentre outras coisas, da escolha de um conjunto de parâmetros de entrada, que influenciam diretamente na qualidade do processo. Otimizar a escolha destes parâmetros é essencial para

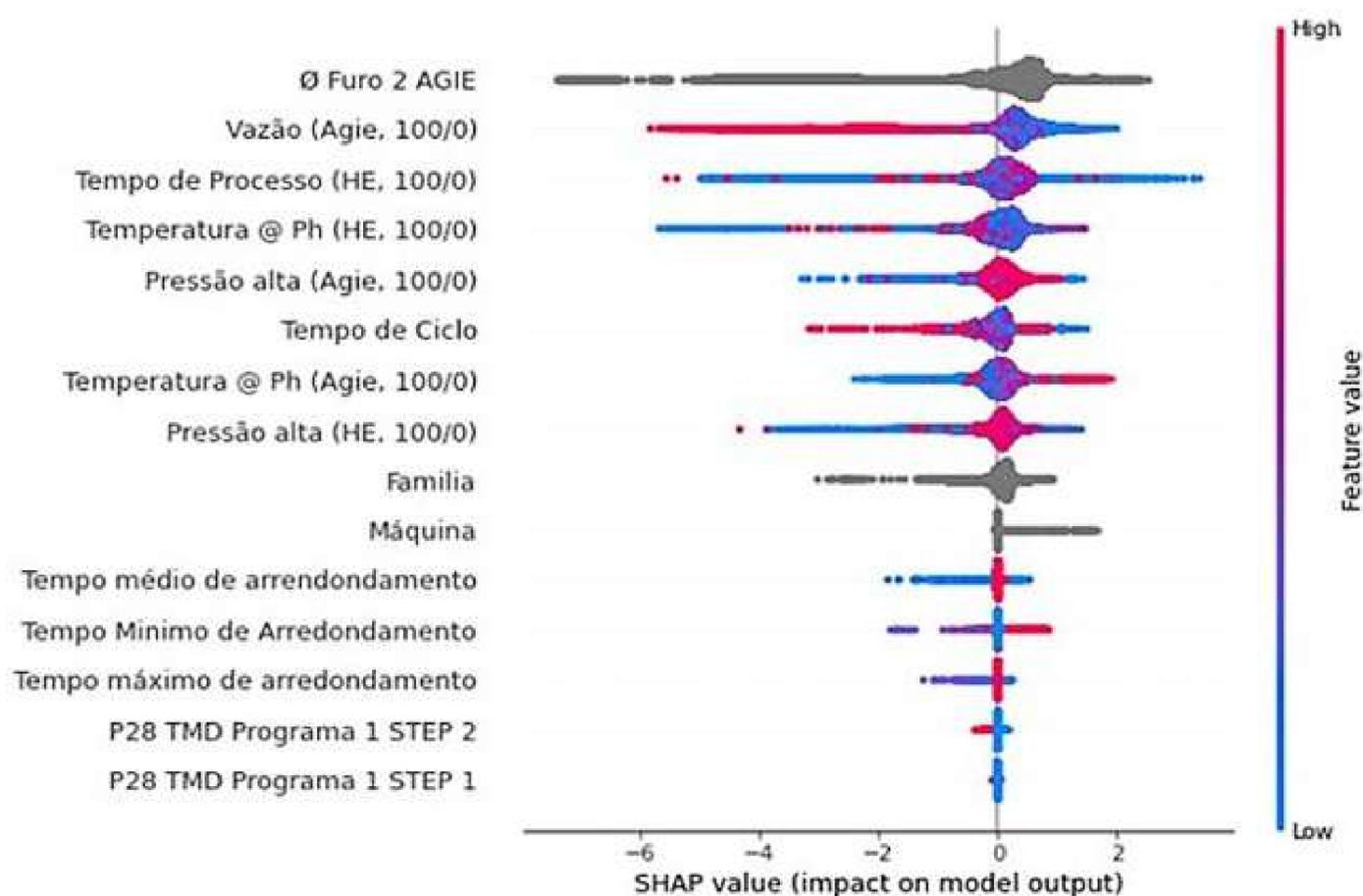
reduzir resíduos gerados por falhas, além de diminuir o tempo de produção e garantir qualidade e precisão dos furos.

### SOLUÇÃO DESENVOLVIDA

Usando ferramentas de análise de dados e machine learning, foi possível construir um dashboard interativo em streamlit responsável por sumarizar as previsões e resultados. O dashboard recebe como parâmetros informações sobre as peças, além das características das máquinas (AGIE, HE, Mycrona E Qhyd) utilizadas no processo de usinagem. E após extensa análise exploratória, a ferramenta retorna informações sobre o percentual de peças fabricadas cujos furos apresentam dimensões dentro/fora dos limites de tolerância, tempo médio de arredondamento, grau de arredondamento, angulação dos furos, temperatura do maquinário, dentre outras informações relevantes. Com esta ferramenta em mãos, é possível estimar as características do furo e sugerir modificações dos parâmetros para maximizar a probabilidade de aprovação da peça. Através disso seria possível reduzir o gasto



com falhas, diminuir os resíduos e o tempo de produção, além de melhorar a qualidade do furo.



# ELECTROLUX

1

## ESTIMATIVA DE CONTRATAÇÃO DE FROTA DE CAMINHÕES PARA ESCOAMENTO DE PRODUTOS

**Residentes:**

André Luiz Marasca, Igor Iraburo Nieto, Matheus Pereira De Novaes.

**Técnica:**

Data Science.

28

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

### OPORTUNIDADE DE NEGÓCIO

A ausência de uma ferramenta eficaz que auxilie na predição da quantidade de caminhões necessários para o transporte de mercadorias até os centros de distribuição, dificulta a organização e o planejamento estratégico por parte da empresa. Acarretando contratações desnecessárias de frotas de caminhões e o fechamento temporário da fábrica por falta de caminhões. Elevando os gastos e atrapalhando a logística de distribuição dos produtos, além de gerar transtornos e má administração do tempo.

## SOLUÇÃO DESENVOLVIDA

Usando as ferramentas de ciência de dados, foi possível construir um dashboard interativo que recebe como parâmetros a quantidade de produtos novos/estocados, os limites (aceitáveis e não aceitáveis) de armazenamento da fábrica, além da capacidade de transporte dos caminhões (medida em cavalos), e retorna o número de caminhões necessários para um melhor escoamento dos produtos até os centros de distribuição. Esta solução auxilia a empresa na tomada de decisões inteligentes, visando otimizar a logística de distribuição, bem como minimizar os gastos e desperdícios.

29

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



# ELECTROLUX

2

## PREVISÃO DE UM MIX ÓTIMO PARA AS LINHAS DE PRODUÇÃO

**Residentes:**

Arthur Lins, Giovane Sylvestrin e Daniel Lima.

**Técnica:**

[DS] Análise exploratória.

**OPORTUNIDADE DE NEGÓCIO**

A determinação de mix de produção, na fase de planejamento de manufatura, tem sido objeto de amplos estudos na área de gestão da produção, abrangendo técnicas intuitivas ou gráficas, a programação matemática e o uso de abordagens heurísticas ou estratégias inteligentes. Os métodos intuitivos ou gráficos usam planilhas de cálculo e gráficos para auxiliar na elaboração de um plano agregado de produção, assemelhando-se a um processo de tentativa e erro e dificilmente resultando em um plano ótimo, ainda mais com a consideração de um número grande de variáveis para o problema (Lustosa et al., 2008). Levando em consideração a

30

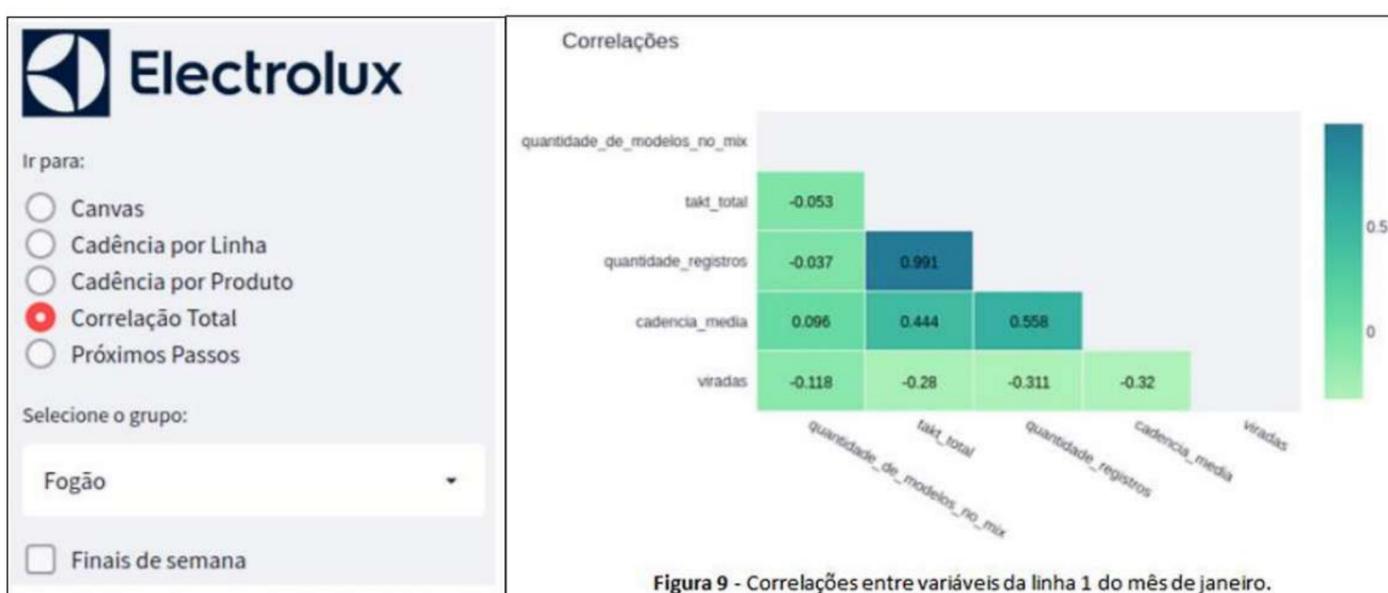
SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

dificuldade em se ter uma previsão acurada para a determinação do mix de produção, é possível utilizar aplicações para que os colaboradores consigam otimizar o processo na linha de produção e produzir de forma mais eficiente.

### SOLUÇÃO DESENVOLVIDA

O desenvolvimento de uma ferramenta interativa capaz de analisar, por linha de produto, dia a dia as combinações de produtos (mix) e que torne possível observar quais demandam mais tempo e quais tem uma performance (cadência real e cadência declarada) melhor em cada linha. Dessa forma, auxiliando os colaboradores na tomada de decisão.



# ELECTROLUX

3

## ANÁLISE DE CADÊNCIAS NAS LINHAS DE PRODUÇÃO

**Residentes:**

Arthur Lins, Giovane Sylvestrin e Daniel Lima.

**Técnica:**

[DS].

32

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

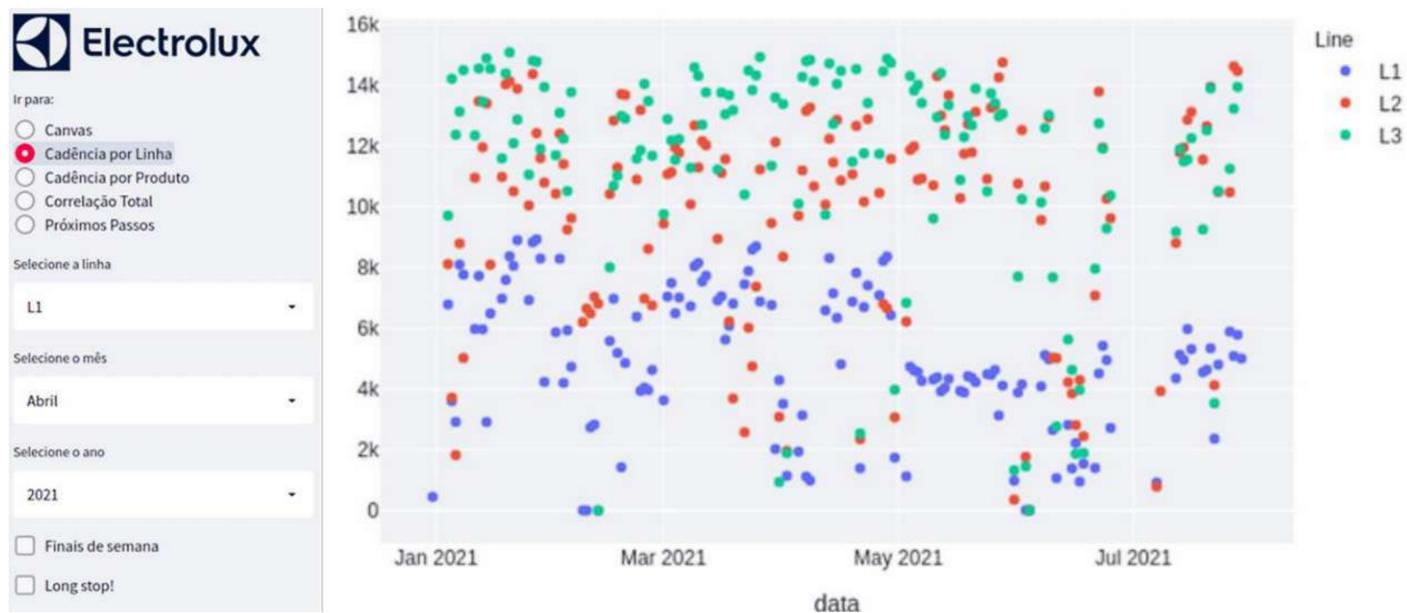
### OPORTUNIDADE DE NEGÓCIO

Uma das maiores preocupações de toda indústria é a quantidade de produtos que ela é capaz de fabricar seja por dia, por semana ou por mês. Essa informação pode ajudar no diálogo, tanto com os clientes quanto com os fornecedores no planejamento dos negócios, prevendo ações e negociando valores.

### SOLUÇÃO DESENVOLVIDA

Desenvolvimento de uma ferramenta interativa para consulta de cadência na produção de produtos por gestores quanto à performance das linhas de produção (receita, despesa, etc.) para a obtenção de outputs como cadência mensal,

quantidade de viradas e gráficos para análises e insights.



33

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





## ELECTROLUX

4

### PREVISÃO DE UM MIX ÓTIMO PARA AS LINHAS DE PRODUÇÃO.

**Residentes:**

Arthur Lins e Daniel Lima.

**Técnica:**

[DS].

**OPORTUNIDADE DE NEGÓCIO**

A determinação de mix de produção, na fase de planejamento de manufatura, tem sido objeto de amplos estudos na área de gestão da produção, abrangendo técnicas intuitivas ou gráficas, a programação matemática e o uso de abordagens heurísticas ou estratégias inteligentes. Os métodos intuitivos ou gráficos usam planilhas de cálculo e gráficos para auxiliar na elaboração de um plano agregado de produção, assemelhando-se a um processo de tentativa e erro e dificilmente resultando em um plano ótimo, ainda mais com a consideração de um número grande de variáveis para o problema (Lustosa et al., 2008). Levando em consideração a

34

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA



dificuldade em se ter uma previsão acurada para a determinação do mix de produção, é possível utilizar aplicações para que os colaboradores consigam otimizar o processo na linha de produção e produzir de forma mais eficiente.

35

### SOLUÇÃO DESENVOLVIDA

O desenvolvimento de uma ferramenta interativa capaz de analisar, por linha de produto, dia a dia as combinações de produtos (mix) e que torne possível observar quais demandam mais tempo e quais tem uma performance (cadência real e cadência declarada) melhor em cada linha. Dessa forma, auxiliando os colaboradores na tomada de decisão.

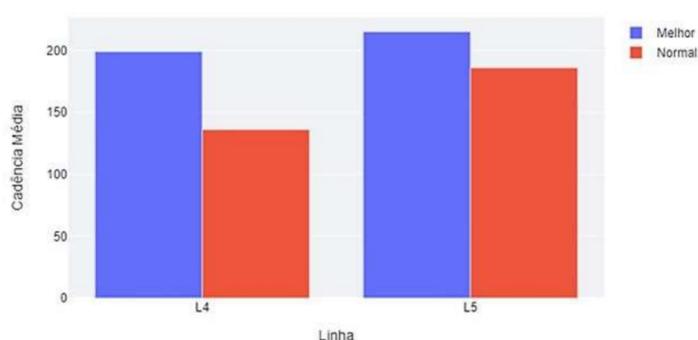
SUMÁRIO



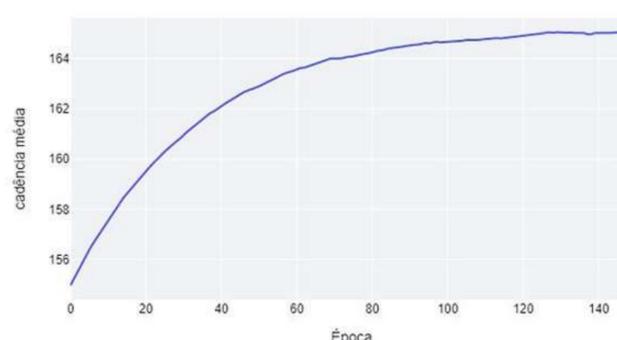
SUMÁRIO  
CASES DO  
PROGRAMA



Cadência Média por Linha



Evolução da Cadência das Populações ao longo das épocas.



## EMBRAPA

1

AUTOMAÇÃO DA METODOLOGIA DRES:  
DIAGNÓSTICO RÁPIDO DA ESTRUTURA DO SOLO**Residentes:**

Gabriel Gomes de Sousa, Melanie Gissel  
Urdangarin Gamarra.

**Técnicas:**

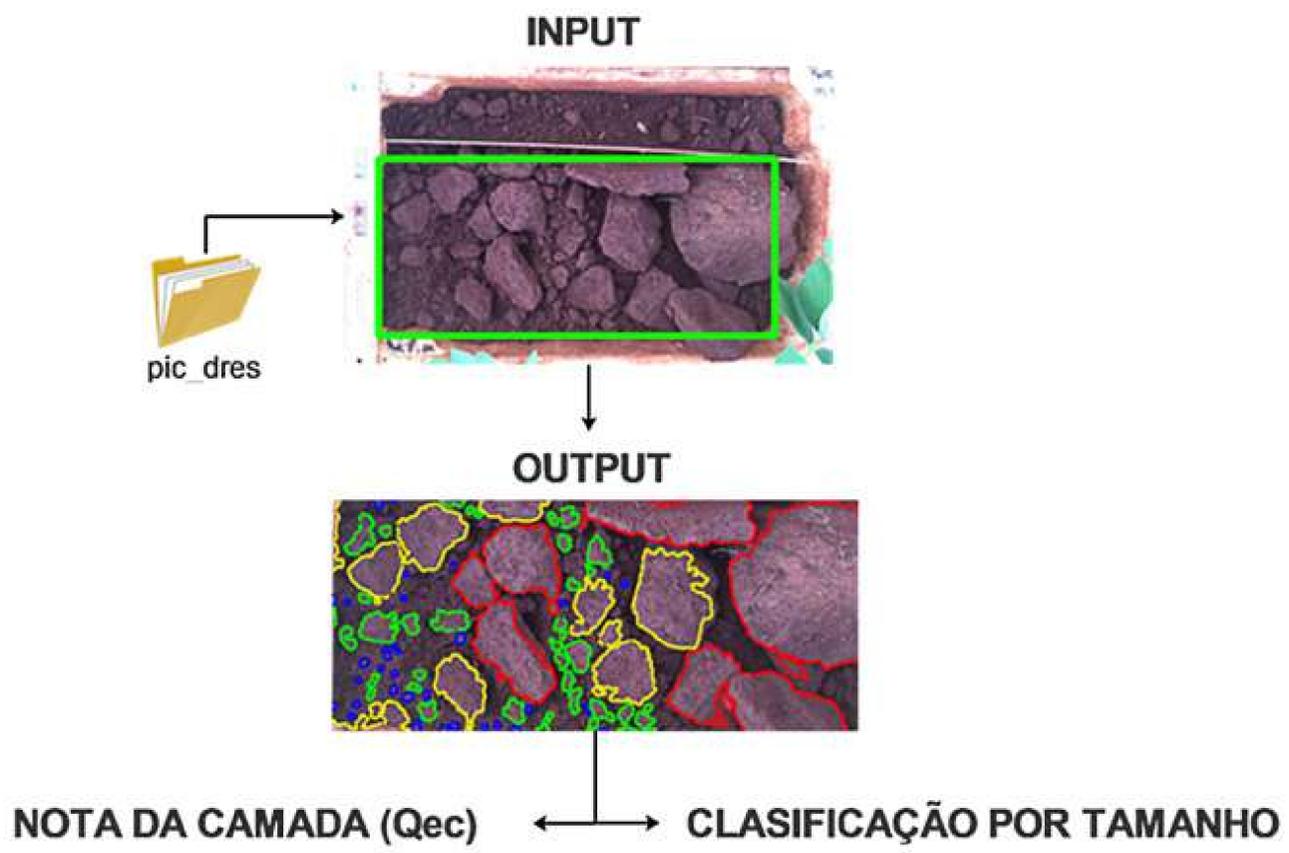
Processamento de imagens, thresholding de  
imagens e detecção de contornos.

36

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**SOLUÇÃO DESENVOLVIDA**

A empresa pretende desenvolver uma Prova de Conceito (PoC) para automatizar a classificação das amostras de solo de acordo com a metodologia DRES. Visa mais especificamente desenvolver uma interface gráfica que recebe como input uma foto de uma amostra de solo preparada de acordo com o DRES, e entrega como output a classificação dos agregados de solo por tamanho e a respectiva nota de qualidade do solo da camada.



## EMBRAPA

2

## MONITORAMENTO DE MÍDIAS SOCIAIS

**Residentes:**

Gabriel Gomes de Sousa, Germano Yoneda  
Pereira Lima, Márcia Rodrigues.

**Técnicas:**

[DS]: Abordagem estática, Abordagem Tempo-  
real, Data Science.

38

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Foi desenvolvida uma Prova de Conceito (PoC)/  
Software para Data Science que possibilitou  
extrair informações de diversas mídias sociais  
por meio de palavras-chave sugeridas (pela  
empresa), bem como o local onde ocorreu o  
maior número de pesquisas (google trends), de  
modo que essas informações foram mostradas  
em um dashboard. Mais especificamente foi  
desenvolvida uma interface gráfica que recebeu  
dados do Youtube, Facebook e LinkedIn a  
respeito do público, bem como a procura de  
matérias citando a Embrapa e da ação digital de

concorrentes e/ou temas correlatos, como por exemplo o agronegócio e inovações nessa área.

### **SOLUÇÃO DESENVOLVIDA**

A solução foi obter dados de pesquisa sobre o agronegócio e palavras-chave envolvendo a Embrapa de dois modos, um estático que possibilita o uso de dados sem o uso da internet e o outro em tempo-real, que faz uso da internet. O último método é mais vantajoso, pois permite a consulta a uma API de notícias, que fornece as principais matérias para o tema solicitado num período de 30 dias, podendo realizar até 100 requisições por dia, além de consulta a tema no google trends. Dentro das soluções desenvolvidas, encontram-se os seguintes benefícios do software para o futuro, que são uma maior agilidade na preparação de uma campanha de marketing, campanhas de marketing mais direcionadas subjetivas e suporte à tomada de decisão.

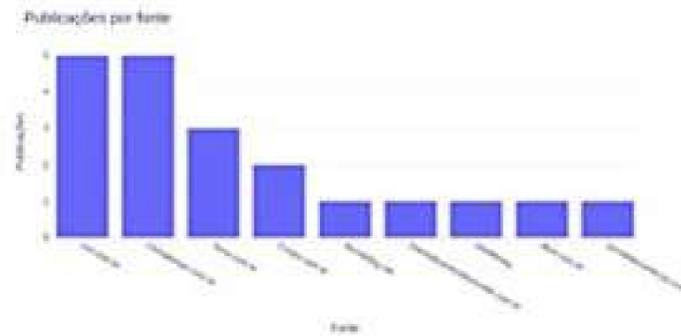
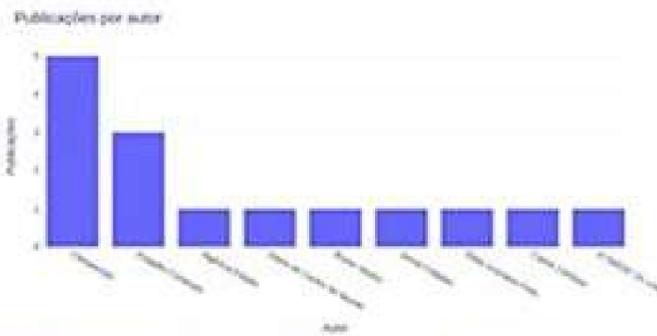
39

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## EMBRAPA

3

## IDENTIFICAÇÃO DA INFESTAÇÃO DE PERCEVEJOS

**Residentes:**

Gabriel Jonas Aguiar, Thiago Pedrosa Guilherme.

**Técnicas:**

[DS]: Análise de imagens obtidas do satélite Sentinel 2.

41

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Esse projeto tem como objetivo investigar como ocorre a proliferação de percevejos em uma plantação e buscar formas de prever como a infestação irá avançar para que seja contida com o objetivo de evitar prejuízo. Então, pretendemos entregar uma análise capaz de ajudar a entender e prever a proliferação de percevejos em plantações de soja utilizando dados coletados em campo e imagens de satélite.

Desenvolver um sistema que auxilie na identificação e proliferação de infestações por

meio da análise de dados de campo coletadas entre novembro de 2019 e fevereiro de 2020 da fazenda de Bela Vista do Paraíso, com informações do estágio da soja, infestação e produtividade. As fotos retiradas pelo satélite Sentinel 2 permitem correlacionar as imagens coletadas com a infestação e estágio da soja e com isso determinar qual local da lavoura e época tinha uma maior infestação.

### **SOLUÇÃO DESENVOLVIDA**

Durante o desenvolvimento do projeto, observamos que poderíamos explorar o problema de diferentes ângulos. Primeiramente entender a evolução da infestação dos percevejos com os dados fornecidos. Então, explorar métricas relacionadas a latitude e longitude e como a infestação se comporta ao longo dos estádios da soja. E por último procurar relações entre a evolução da infestação e imagens de satélites que registram diferentes bandas (visíveis e invisíveis).

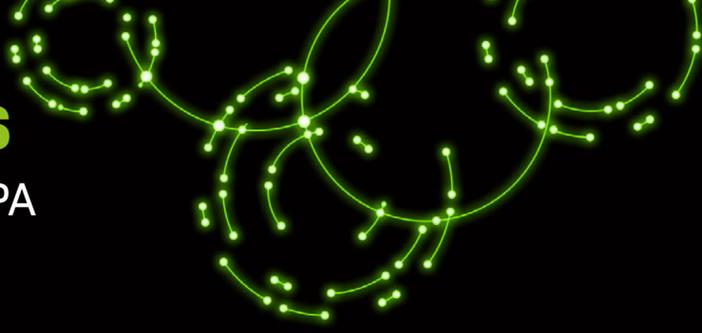
42

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





43

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## EMBRAPA

4

## CLASSIFICADOR DAS FASES DA MOSCA BRANCA

**Residentes:**

Karen Taniguchi, Pedro Humberto Nunes, Sergio Lemos Hermsdorff.

**Técnicas:**

[DP]: Função HoughCircles, OpenCV e Modelo de Predição (Faster R-CNN).

44

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Este projeto tem como objetivo o estudo de tecnologias e proposição de uma POC (Prova de Conceito – Proof of Concept) com o desenvolvimento de uma ferramenta de detecção e classificação das etapas do ciclo biológico da mosca branca em folhas de soja, avaliando a acurácia e o tempo de processamento como estudo de viabilidade para uma futura ferramenta da Embrapa.

O escopo deste projeto compreende o desenvolvimento de um conjunto de scripts de

processamento de imagem, treino e inferência para identificar diferentes estágios da mosca branca em folhas de soja. Também compreende o estudo de métricas de avaliação para diferentes modelos, considerando a acurácia de suas predições. E por último, faz parte desse escopo uma interface para auxiliar o usuário na visualização dos resultados.

### **SOLUÇÃO DESENVOLVIDA**

Para o desenvolvimento desse projeto foi aplicado um algoritmo de aprendizagem de máquina com implementação via webstreamlit para detecção de objetos que é amplamente usado nessa área, chamado Faster R-CNN. Deste modo foi desenvolvido script que roda tanto em Jupyter Notebook como também em demais plataformas e que traz como solução a possibilidade de alterar dados entre fase de treinamento e fase de teste e fazer um simples upload das imagens ou indicar qual diretórios que as imagens estão contidas para realizar a predição.

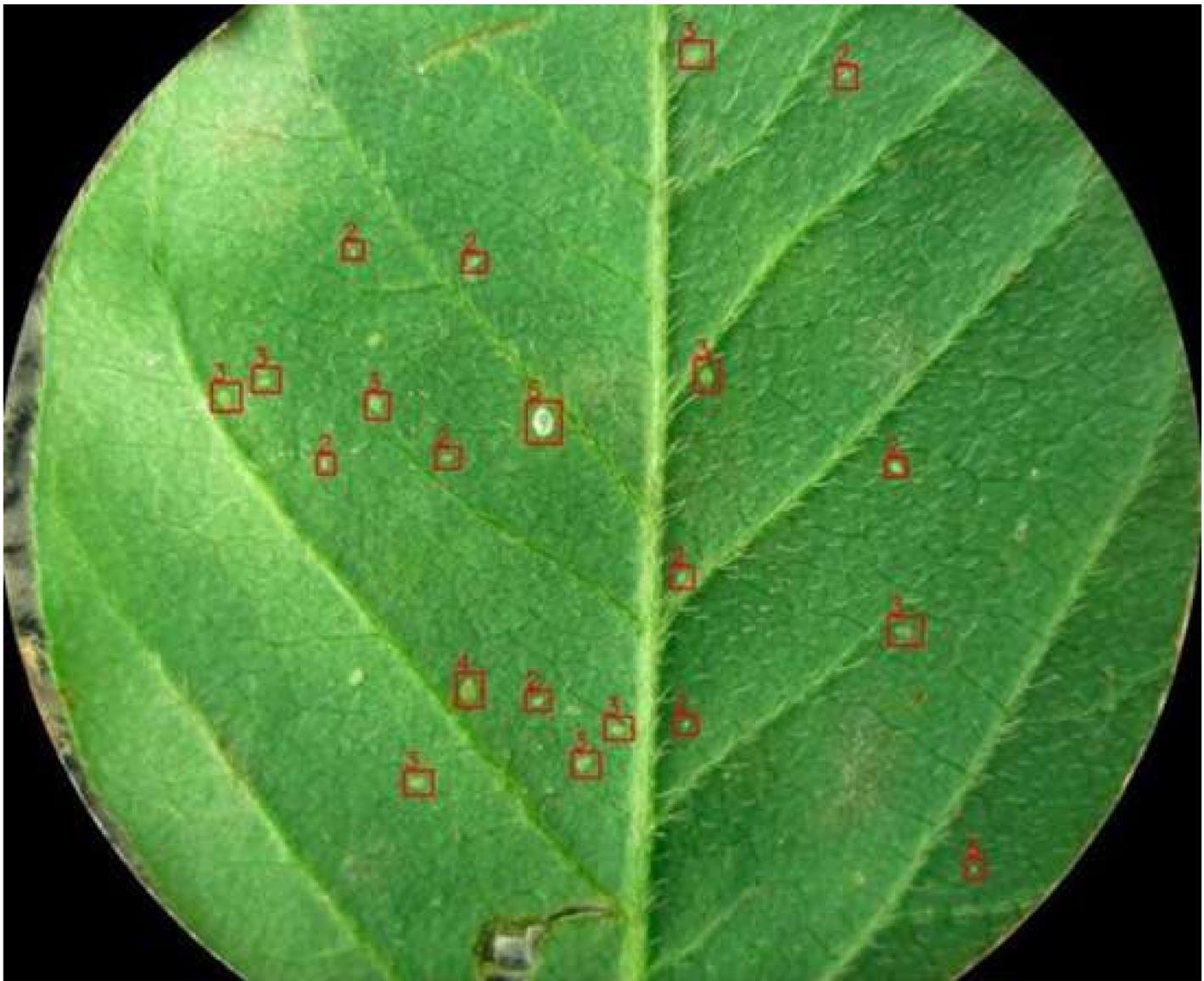
45

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





46

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



# BELLAGRÍCOLA FARMBITS

1

## ANÁLISE DO PERFIL DOS CLIENTES 1

### Residentes:

Arthur Lins, Luis Gustavo Silva e Laryssa Stephanie.

### Técnicas:

[DS] - Análise, manipulação e organização de dados, visualização de dados.

47

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

## OPORTUNIDADE DE NEGÓCIO

Os conceitos mais básicos de fidelização indicam a disputa por um padrão de compras repetidas, reforçadas por Brei e Rossi (2005) como a forma mais simples e fácil para fidelizar clientes, chegando até a perspectiva comportamental com seus conceitos de fidelidade cognitiva e afetiva (OLIVER, 1997, 1999). A fidelização concreta de clientes proporciona à instituição rendimento seguro e garantido por um bom período de tempo, razão pela qual os gestores buscam a retenção dos clientes. A longevidade de relacionamento com seus clientes influencia



favoravelmente a lucratividade da empresa, e a obtenção de um desempenho lucrativo também pode ser constituído como fator importante à satisfação das necessidades e interesses dos clientes.

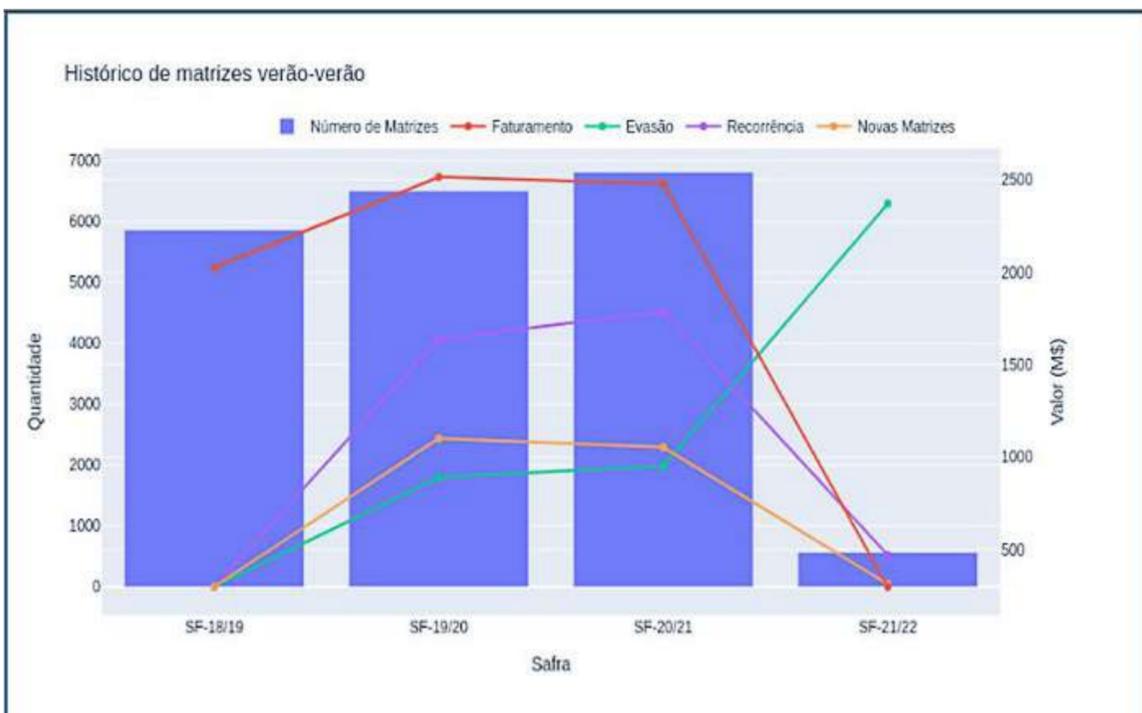
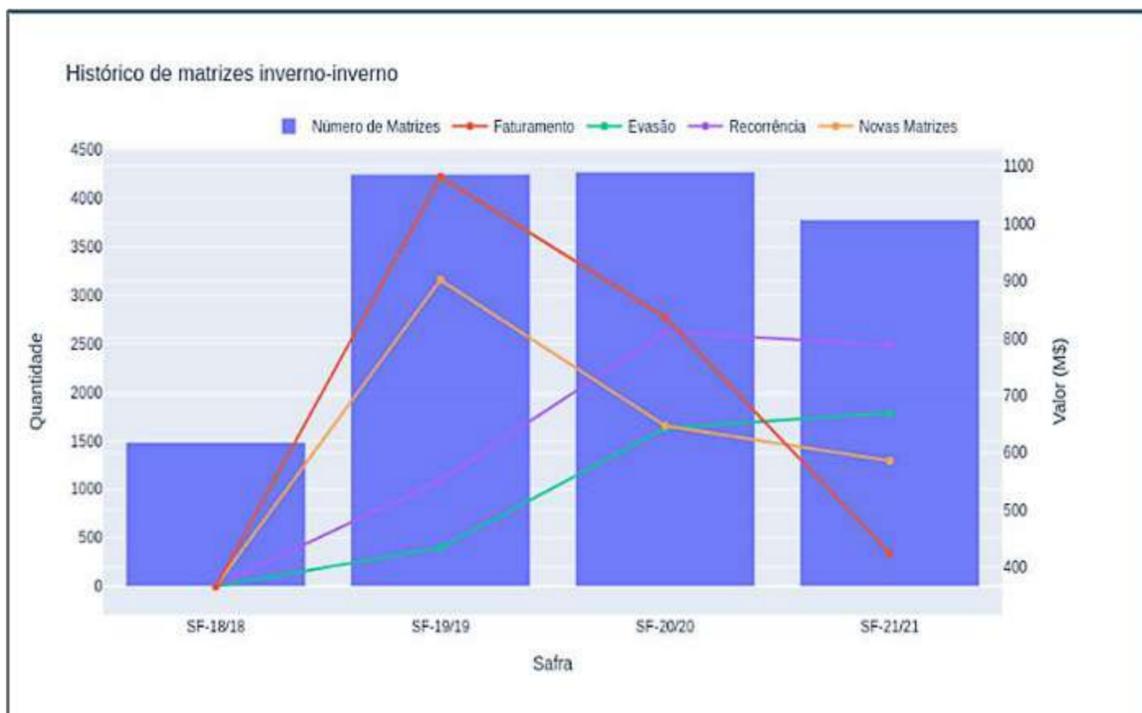
A manutenção de clientes fiéis influencia o aumento das vendas e proporciona melhores preços e menores custos (BARNES, 2002), o que é excelente em tempos globalizados e de competitividade acirrada. Esses clientes tendem a gerar maior fluxo de caixa e lucro, comprar mais e divulgar melhor a empresa. Tendo isso em vista, é muito importante avaliar como fidelizar melhor os clientes, levantar possíveis causas para sua evasão, visando a melhoria no relacionamento com os clientes.

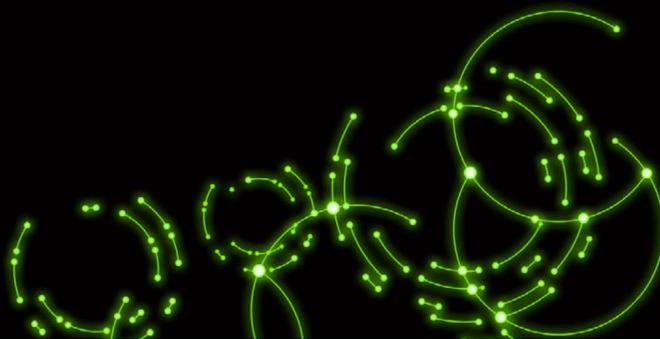
### **SOLUÇÃO DESENVOLVIDA**

O estudo de tecnologias e proposição de uma POC (Prova de Conceito – Proof of Concept) para realizar análise exploratória no histórico de interações dos clientes e possíveis dados externos, para identificar características que evidenciam a recorrência ou evasão. Durante o



período de desenvolvimento da Sprint realizou-se a exploração dos dados de movimentações da empresa Farmbits-Belagrícola durante o período de 2018 a 2021. Posteriormente foi criado um data frame listando o perfil de todos os clientes que obtivemos acesso. Nessa base de dados há features estatísticas que foram criadas para auxiliar o desenvolvimento do projeto na próxima fase.





# BELLAGRÍCOLA FARMBITS

2

## ANÁLISE DO PERFIL DOS CLIENTES 2

### Residentes:

Arthur Lins, Amanda Heino e Anna Luiza Gomes.

### Técnicas:

[DS] - Análise, manipulação e organização de dados, visualização de dados.

50

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

### OPORTUNIDADE DE NEGÓCIO

Os conceitos mais básicos de fidelização indicam a disputa por um padrão de compras repetidas, reforçadas por Brei e Rossi (2005) como a forma mais simples e fácil para fidelizar clientes, chegando até a perspectiva comportamental com seus conceitos de fidelidade cognitiva e afetiva (OLIVER, 1997, 1999). A fidelização concreta de clientes proporciona à instituição rendimento seguro e garantido por um bom período de tempo, razão pela qual os gestores buscam a retenção dos clientes. A longevidade de relacionamento com seus clientes influencia



favoravelmente a lucratividade da empresa, e a obtenção de um desempenho lucrativo também pode ser constituído como fator importante à satisfação das necessidades e interesses dos clientes.

A manutenção de clientes fiéis influencia o aumento das vendas e proporciona melhores preços e menores custos (BARNES, 2002), o que é excelente em tempos globalizados e de competitividade acirrada. Esses clientes tendem a gerar maior fluxo de caixa e lucro, comprar mais e divulgar melhor a empresa. Levando-se em conta o considerável custo para adquirir clientes novos, bem como os benefícios de reter os clientes já existentes, criar um modelo de predição de churn que auxilie a retenção do cliente é relevante para o sucesso de uma empresa num mercado competitivo (NEZZE, 2002).

### **SOLUÇÃO DESENVOLVIDA**

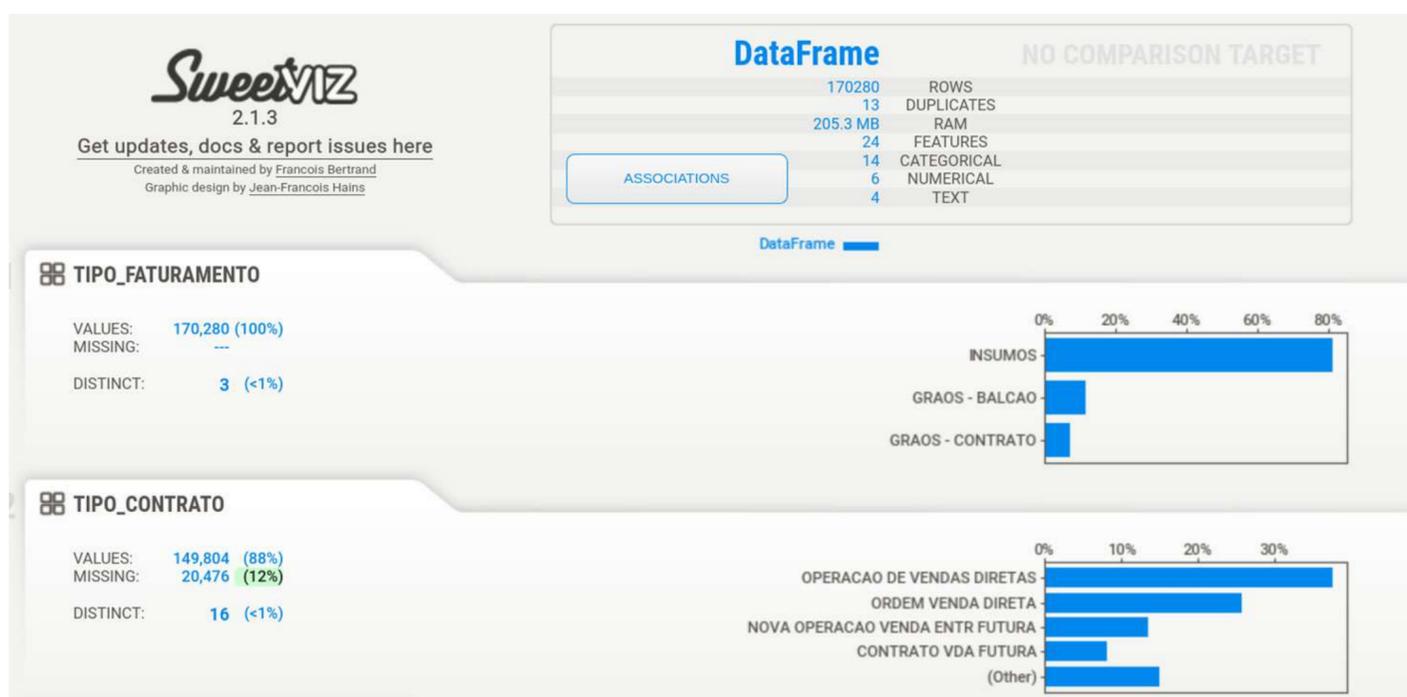
Desenvolvimento de um jupyter notebook que contempla a análise dos dados ao longo do tempo, assim como o teste de modelos de

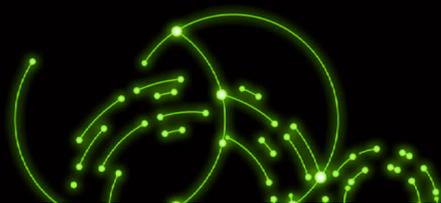
**51**

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

regressão, para obter a quantidade comprada de insumos em safras futuras para que pudesse ser comparada com o padrão de compra do cliente, e verificar se houve um aumento ou redução considerável na compra. Isso se deu através da implementação de uma API (Application Programming Interface), utilizando o microframework Flask, que permite criar aplicativos WEB em Python de forma simples e rápida gerando aplicações robustas.





# BELLAGRÍCOLA FARMBITS

3

## SISTEMA DE RECOMENDAÇÃO PARA IMPULSIONAR GERAÇÃO DE DEMANDA

### Residentes:

Ananda Heino, Daniel Lima e Rafael Kimura.

### Técnicas:

[ML e DS]: Data Science(LightFM) e Machine Learning.

53

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

### OPORTUNIDADE DE NEGÓCIO

Este projeto tem como objetivo o estudo de tecnologias e a proposição de uma POC (Prova de Conceito – Proof of Concept) de um sistema de recomendação, para a equipe de vendas da Bellagrícola, com o intuito de gerar demanda visando aumentar as vendas no período repique.

### SOLUÇÃO DESENVOLVIDA

Consta no escopo do projeto, o estudo dos dados de calendário de aplicação, nível de tecnologia, similaridades dos produtos (vendidos fora dos períodos de eventos) e dados de faturamento

dos insumos, criando um sistema de recomendação personalizado a nível de cliente. Compreende, também, o desenvolvimento de Jupyter notebooks, scripts .py e uma API. O projeto tem como proposta geral, incentivar o aumento da geração de demanda no período repique. Considerando a subdivisão do projeto em duas semanas de desenvolvimento da Sprint, temos as seguintes metas:

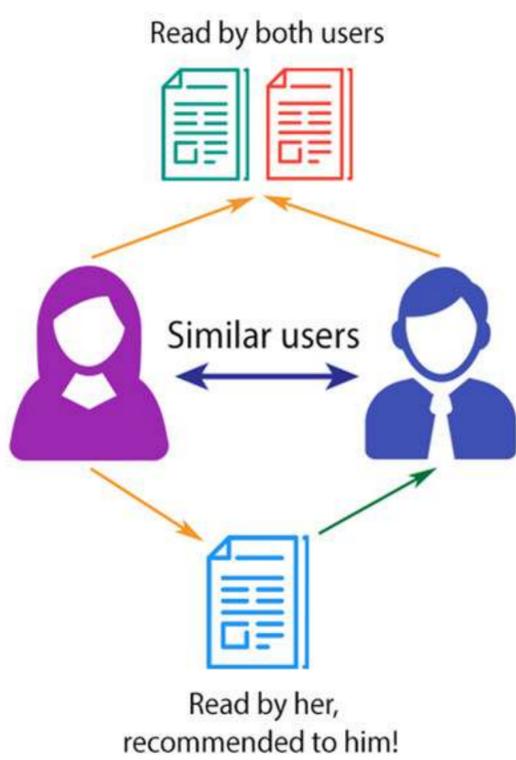
- Meta 1: Análise exploratória e preparo dos dados, criando as features para input no modelo;
- Meta 2: Treinamento do sistema de recomendação, testando métricas de validação e, posteriormente, salvando os modelos, de acordo com o tipo de safra;
- Meta 3: Implementação das classes e da API;
- Meta 4: Apresentação do funcionamento do sistema de recomendação em um Dashboard.

Entregas da Meta 1: Nessa fase foram concentrados esforços na exploração das informações disponibilizadas pela empresa patrocinadora, bem como na preparação dos dados para entrada de um sistema de recomendação. Entregas da Meta 3: Nessa

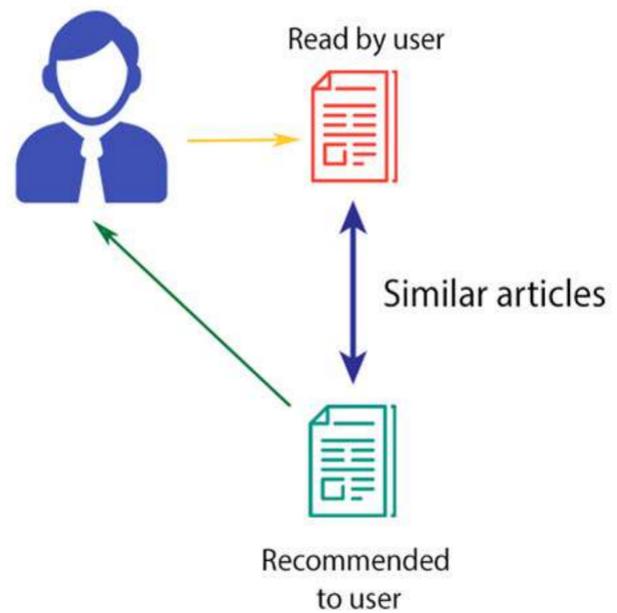


terceira fase, foram concentrados esforços no encapsulamento do código desenvolvido nas fases anteriores. Para isso, foram implementadas três classes e uma API Rest usando a biblioteca Flask.

COLLABORATIVE FILTERING



CONTENT-BASED FILTERING





# BELLAGRÍCOLA FARMBITS

4

## PREVISÃO DE DEMANDA E PLANEJAMENTO COMERCIAL

### Residentes:

Daniel Victor Teixeira Lima e Laryssa Stephanie Andrade da Costa Silva.

### Técnicas:

[Data Science] - Análise, manipulação e organização de dados, visualização de dados.

56

SUMÁRIO

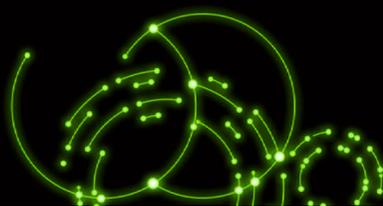
SUMÁRIO  
CASES DO  
PROGRAMA

### OPORTUNIDADE DE NEGÓCIO

O setor agrícola possui um mercado dinâmico onde decisões importantes devem ser tomadas a todo momento. A previsão de demanda de insumos agrícolas é uma ferramenta que permite especialistas do setor a fazerem previsões e planejamento com maior precisão.

### SOLUÇÃO DESENVOLVIDA

Foram realizadas análises exploratórias para verificar a correlação entre os dados e avaliados métodos de agrupamento para estudar o impacto das mudanças no planejamento atual.



Assim, uma ferramenta interativa e online para visualizar os dados da empresa na forma de um dashboard foi desenvolvida por streamlit. Os dados foram apresentados na forma de diferentes gráficos interativos onde foi simulado o funcionamento do sistema de recomendação. O método de planejamento atual da empresa foi avaliado e foram propostas sugestões de melhorias. Além disso, foram realizadas projeções de demanda em nível de safras e a distribuição em vários níveis de insumos para 5 safras posteriores.



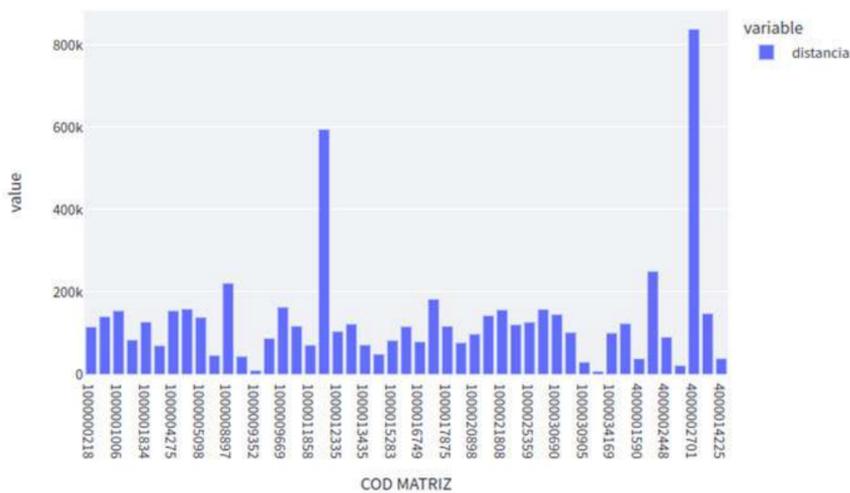
farmbits

Selecionar Página

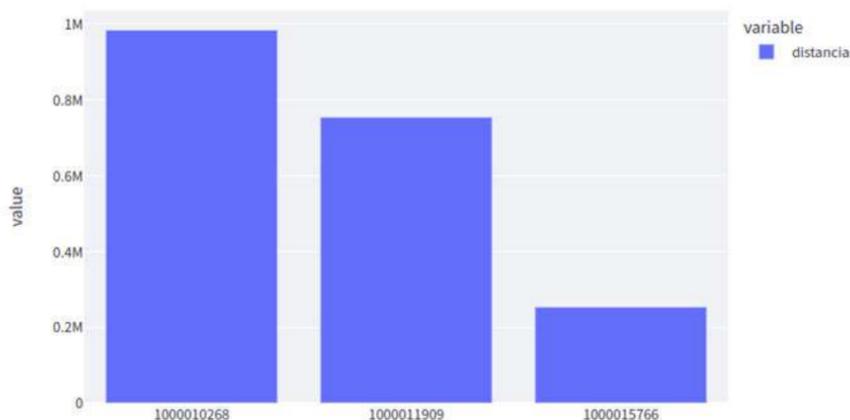
Ir para:

- Objetivos
- Avanços
- Resultados
- Próximos Passos

Atualizar base de dados



Rank 3



## HEDRO

1

PREVISÃO DE DEMANDA E PLANEJAMENTO  
COMERCIAL**Equipe 1:**

Giovane Sylvestrin, João Kawazoe e Laryssa Stephanie.

**Equipe 2:**

Giovane Sylvestrin e Rafael Kimura.

**Equipe 3:**

Igor Iraburo Nieto, Luis Gustavo Tomaz Silva e Rafael Kimura.

**Equipe 4:**

Arthur Silva Lins, Igor de Paula Moura e Igor Iraburo Nieto.

**Técnicas:**

[DS]: Análise exploratória, tratamento de dados.  
[IA/ML]: K-means, Dynamic Time Warping, Isolation Forest, Random Forest, CatBoost, Rede Neural LSTM.

58

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

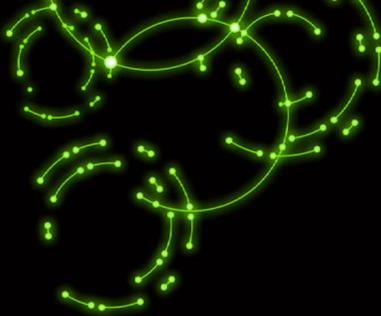
A laminação é um processo produtivo que consiste na redução da espessura de um material pela passagem e compressão entre cilindros tracionados. Esse processo tem o

objetivo de obter um material com propriedades mecânicas mais adequadas para numerosas aplicações. Como qualquer outro processo produtivo a laminação está sujeita a falhas, que neste caso podem comprometer a segurança dos operadores, além é claro do produto final e dos lucros da empresa. Quando uma falha ocorre são geradas sucatas, que à primeira vista podem parecer eventos aleatórios, mas com o estudo e a análise de sua origem podemos identificar as anomalias, diminuindo assim a reincidência de falhas ou gerando um alerta que avisa sobre um possível evento.

### SOLUÇÃO DESENVOLVIDA

Para a análise da origem das sucatas foram disponibilizados dados de sensores previamente colocados nas máquinas para detectar a assinatura normal de vibração e as assinaturas anômalas. Os sensores também detectaram inclinação e temperatura, registrando um dado a cada segundo de funcionamento. Também foram disponibilizados relatórios de parada (com falhas, geração de sucata, manutenção) anotados pelos próprios operadores.





No primeiro momento foi realizado um estudo estatístico, seguido por uma análise de dados e a compreensão das possíveis interferências no processo. O foco foi preparar um dataframe em python que poderia ser usado como entrada para um modelo preditivo.

Com os dados organizados, concentrou-se em encontrar maneiras de identificar e diferenciar os padrões de funcionamento normal e anômalo com o objetivo de prever e prevenir possíveis falhas durante o processo de laminação.

Durante esta sprint foram testadas diferentes implementações de técnicas de aprendizado não supervisionado, como K-Means, Isolation Forest e Dynamic Time Warping, que permite encontrar padrões similares em dados temporais mesmo em eventos com diferentes ritmos e velocidades. Os resultados obtidos foram utilizados para realizar uma primeira rotulação dos dados e treinar modelos de aprendizado supervisionado, em especial o RandomForest e o CatBoost, que fazem a classificação binária dos eventos analisados.

60

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

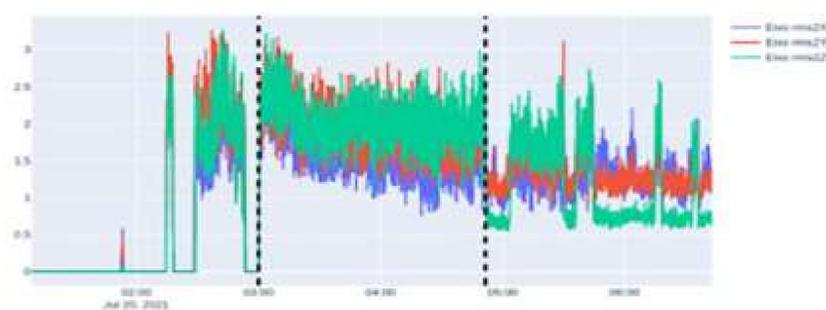
Continuando os trabalhos a rotulação dos dados foi refinada, criando indicadores que descrevem o comportamento de vibração para a modelagem e treino dos algoritmos de classificação. Modelos preditivos, especificamente redes neurais convolucionais LSTM foram implementados, com o intuito de identificar os comportamentos estranhos que antecedem os eventos de geração de sucata.

61

SUMÁRIO



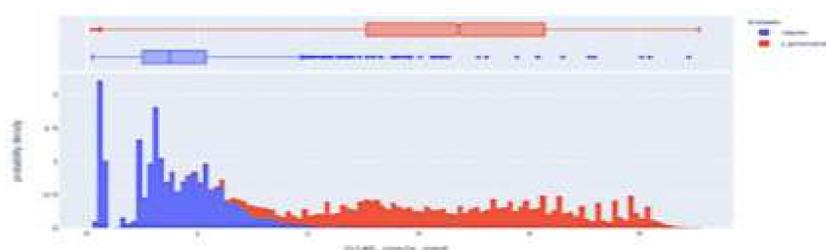
SUMÁRIO  
CASES DO  
PROGRAMA



Visualização de uma possível anomalia



Comportamentos que antecedem uma sucata



Distribuição de densidade das vibrações no eixo X de um dos sensores

## HS TECHNOLOGY

1

## CALIBRAÇÃO BAROSCAN

**Residentes:**

Gabriel Jonas Aguiar, Márcia Rodrigues e Victor Nagahama.

**Técnicas:**

Análise exploratória de dados, Regressão, ConvNet, MLP, Deep Learning.

62

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

O cliente HS Technology é uma empresa que desenvolve equipamentos eletrônicos para a área da saúde. O produto do estudo do projeto é o BaroScan, um baropodômetro baseado em uma plataforma de sensores para realizar a leitura da pressão plantar. A necessidade de nova forma para calcular a pressão se dá por conta da dificuldade de caracterização do comportamento do material, principalmente devido ao fato de não possuir propriedades elétricas uniformes, e devido interferências indesejadas entre sensores (crosstalk) que não

foram eliminadas na eletrônica e que o processo atual de calibração não permite identificar. Assim, foi proposto o desenvolvimento de um método para converter variáveis elétricas em valores de pressão de um determinado material com propriedades resistivas variáveis para aprimorar o atual algoritmo que calcula a pressão aplicada nos sensores dos produtos da HS Technology.

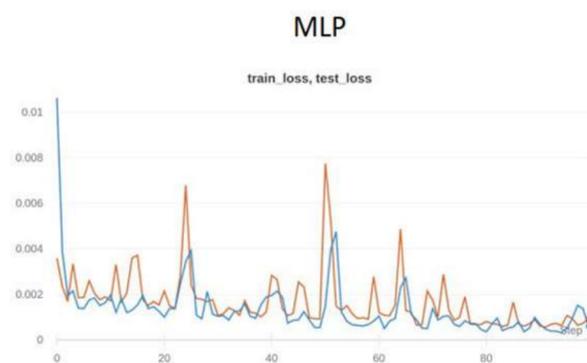
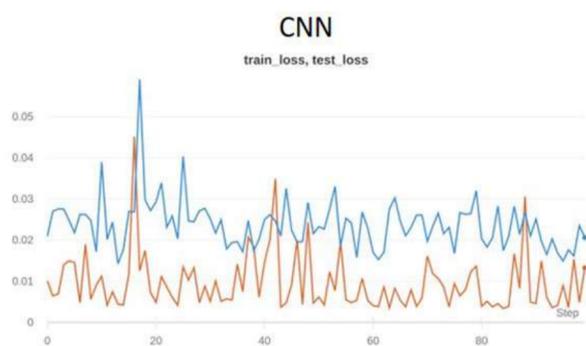
### SOLUÇÃO DESENVOLVIDA

Entrega de uma nova metodologia para realizar a calibração de sensores dos produtos da HS Technology a fim de caracterizar o comportamento do material resistivo e aprimorar o atual algoritmo que calcula a pressão aplicada.

Grupo de entregas dessa sprint: o projeto se desenvolveu em três partes, sendo estas as metas da Sprint 4: O tratamento e a preparação dos dados disponibilizados para a sprint; comparação entre a calibração proposta e existente; modelo de treinamento para calibrar lotes.



LOTE 1



64

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



# INSTITUTO DE DESENVOLVIMENTO RURAL DO PARANÁ

1

## ANÁLISE DE DADOS AGROMETEOROLÓGICOS

**Residentes:**

Tiago Pedrosa e Natália Sá.

**Técnicas:**

[DS] Dashboard Interativo e Análise de Dados.

[Estatística]: Correlação.

65

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

### OPORTUNIDADE DE NEGÓCIO

O clima e as mudanças climáticas influenciam diretamente na agricultura e na qualidade de vida das pessoas. A falta de chuva ou mudanças bruscas na temperatura podem prejudicar a qualidade da colheita de algumas culturas. A baixa umidade do ar pode estar relacionada com problemas respiratórios. Sendo assim, é interessante que seja feita uma análise confiável dos dados climáticos de uma região, que pode servir como base tanto para agricultores quanto para pesquisadores de diversas áreas.



## SOLUÇÃO DESENVOLVIDA

Neste projeto foi criada uma ferramenta de análise de dados agro meteorológicos históricos do Paraná e os resultados foram apresentados através de um dashboard interativo. As análises foram realizadas com consultas ao banco de dados convencional do Instituto de Desenvolvimento Rural do Paraná, capturando informações sobre região, assim como sua evolução climática. Também foram feitas correlações tanto entre diferentes variáveis (temperatura, umidade, insolação, etc.) quanto entre diferentes estações de coleta, exibindo padrões, tendências e correlações encontrados.

O projeto foi dividido em 4 sprints, sendo considerada uma META para cada sprint, de acordo com o objetivo proposto. As entregas da Meta 1 foram: pré-processamento da base de dados; análise de volume, quantidade e frequência de chuva por estação; análise de temperatura mínima, máxima e média; filtro de umidade relativa; correlação entre estações; correlações entre as variáveis; dashboard com o resultado das análises.





## CASES

# INSTITUTO DE DESENVOLVIMENTO RURAL DO PARANÁ

IDR

Escolha uma opção

- Home
- Chuva
- Temperatura
- Umidade
- Comparando variáveis
- Comparando regiões
- Correlação entre estações de coleta

Analise\_agrupando\_regioes.ipynb

Analise\_chuva.ipynb

Analise\_explo\_v2.ipynb

Analise\_frequencia\_chuva.ipynb

Analise\_temperatura.ipynb

Analise\_variacao\_temperatura.ipynb

Analise\_varias\_variaveis.ipynb

Apresentacao.ipynb

Comparando\_colunas\_com\_o\_fator\_chuva.i...

Correlacao\_entre\_estacoes\_de\_coleta.ipynb

Duvidas.ipynb

Filtros.ipynb

Pre-processamento.ipynb

## Dashboard de análise agrometeorológica



### IDR-Paraná

Instituto de Desenvolvimento Rural do Paraná - IAPAR-EMATER

# 67

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



# INSTITUTO DE DESENVOLVIMENTO RURAL DO PARANÁ

2

## ANÁLISE DE ÍNDICE DE ELASTICIDADE PARA A CULTURA DO FEIJÃO

### Residentes:

Larissa Luciane Ferreira e Sergio Lemos Hermsdorff.

### Técnicas:

[DS]: Análise, organização e visualização de dados. [Estatística]: Métrica, Regressão Linear, Mínimos quadrados ordinários [OLS], Mínimos quadrados recursivos [RLS].

68

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

### OPORTUNIDADE DE NEGÓCIO

O Brasil é uma grande potência tratando-se de produção de feijão, sendo o terceiro maior entre 2012 e 2016. Segundo a Secretaria de Estado de Agricultura e do Abastecimento (SEAB), nesse período, o país produziu em média mais de 2,9 milhões de toneladas do grão, representando 11% da produção mundial. Já em território nacional, quem se destaca é o estado do Paraná, que representa 22,57% e colheu 590 mil toneladas de feijão, somente em 2016, de acordo com dados da Pesquisa

Agrícola Municipal. Tendo em vista a importância do negócio para a região, é essencial avaliar os índices de elasticidade preço, que seria a sensibilidade do mercado consumidor diante das variações de preços praticados. Avaliando o histórico da cultura do feijão, considerando oferta, demanda, preço e renda será possível realizar o cálculo de índices de elasticidade para um período de tempo e utilizá-los como auxílio no planejamento, previsão e prevenção de reações de expansão ou retração por parte do mercado de consumo.

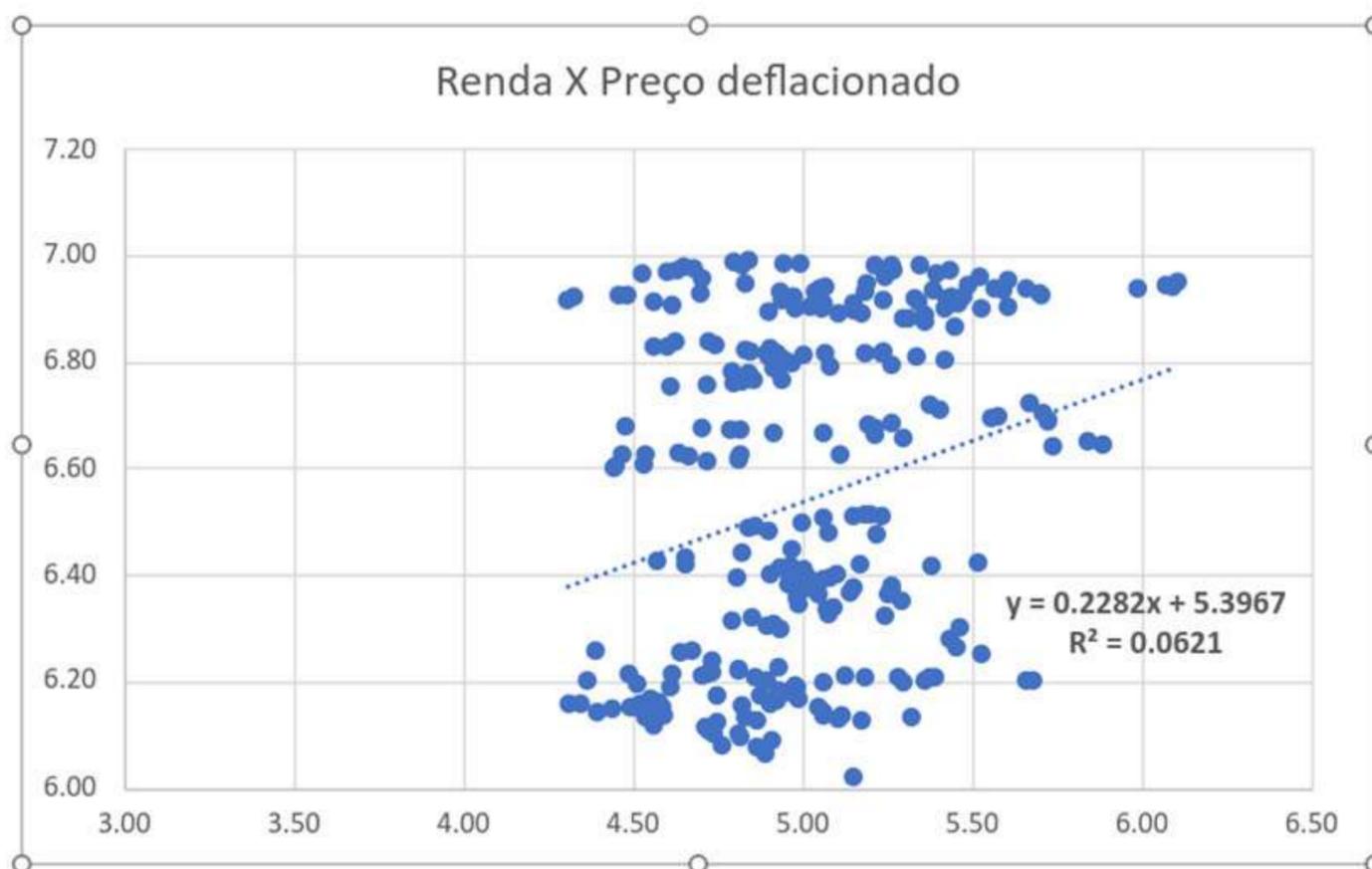
### **SOLUÇÃO DESENVOLVIDA**

O estudo de tecnologias e preposição de uma POC (Prova de Conceito–Proof of Concept) de uma ferramenta interativa online que permite ao usuário que inclua sua base de dados e que realiza o cálculo do índice de elasticidade preço demanda, oferta e renda da demanda, para um certo período de tempo, com diferentes ferramentas de regressão, e plotagem representativa para avaliação por parte do especialista.





Grupo de entregas dessa Sprint: Ferramenta (Streamlit) que permite a inclusão da base e escolha do modelo; relatório com a aplicação no conjunto de dados de feijão; apresentação da ferramenta e dos resultados da aplicação.



# INSTITUTO DE DESENVOLVIMENTO RURAL DO PARANÁ

3

## CHATBOT TELEGRAM

### Residentes:

Melanie Gissel Urdangarin Gamarra, Pedro Humberto A. P. Teixeira Nunes e Roberto Schnitzler Moure.

### Técnicas:

[IA]: BERT, Embeddings, Deep Learning, Machine Learning.

71

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

## OPORTUNIDADE DE NEGÓCIO

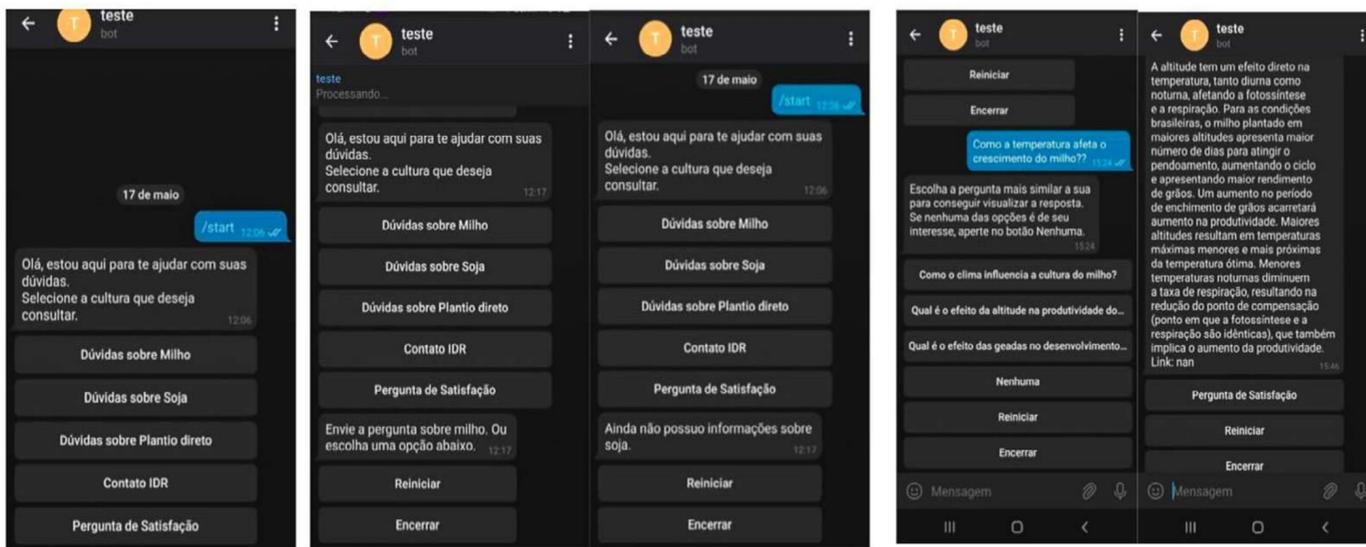
O cliente tinha uma alta demanda de atendimentos por conta de dúvidas, consultas de informações e, por isso, encontrava muita dificuldade em atender a todos.

## SOLUÇÃO DESENVOLVIDA

Melhoria da usabilidade das funções da rede neural (embeddings e similaridade) criadas e foi desenvolvido um chatbot no aplicativo Telegram que consiga processar as perguntas dos usuários e responder de acordo com o

conhecimento fornecido ao mesmo. Dessa forma, o usuário pode através do chat do Telegram se comunicar com o bot, enviando comandos e obtendo as respostas.

Grupo de entregas dessa Sprint: IA capaz de interpretar perguntas e retornar respostas adequadas utilizando BERT; estruturação de um chatbot de Telegram para a interação facilitada com o usuário para a demonstração da ferramenta; encapsulamento da IA em uma API para portabilidade da solução.





# INSTITUTO DE DESENVOLVIMENTO RURAL DO PARANÁ

4

## DETECÇÃO DE INSETOS BASEADO EM VÍDEOS E IMAGENS

**Residentes:**

Karen Itsuki Taniguchi e Victor Hugo Nagahama.

**Técnicas:**

Plataformas Make Sense e Wandb. Bounding box, Faster R-CNN, Mean Average Precision, Mean Average Recall, Web Streamlit (Front-end), Visão Computacional, Deep Learning.

73

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

### OPORTUNIDADE DE NEGÓCIO

Diversas técnicas têm sido utilizadas para combater pragas no campo, tal como novos pesticidas, rotação de culturas e feromônios para atração e eliminação de pragas. Porém, antes de aplicar qualquer processo é necessário identificá-las e tal processo pode não ser trivial para pessoas que não possuem o conhecimento dos espécimes.

Uma das principais aplicações da visão computacional consiste em identificar objetos,

que podem ser considerados como carros, pessoas, animais ou até mesmo uma radiografia pulmonar com tuberculose. Com o intuito de facilitar o processo de identificação de pragas, esta POC consistiu em trazer a metodologia visão computacional para o domínio dos insetos.

### **SOLUÇÃO DESENVOLVIDA**

Implementação de uma inteligência artificial capaz de ser treinada para detectar insetos a partir de vídeos e imagens de forma que seja disponibilizada em uma API. Bem como, uma aplicação para disponibilizar visualizações e facilitar o treinamento da IA.



# COOPERATIVA INTEGRADA

1

## PREVISÃO DE CAIXA PARA FIXAÇÃO DE SOJA

### Residentes Equipe 1:

Daniel Victor Teixeira Lima e Matheus Novaes.

### Residentes Equipe 2:

Luis Gustavo, Daniel Lima e João Kawazoe.

### Técnicas:

[IA] Machine learning (métodos de autoregressão pelo modelo SARIMAX). [DS]: Filtros da base de dados fornecida, tratamento de dados, análise exploratória e visualização estatística. [Front-End]: Aplicação com Streamlit para visualização dos resultados e interação.

75

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

## OPORTUNIDADE DE NEGÓCIO

Devido ao crescimento do número de cooperados e interação com a produção de soja, a análise e preparação de caixa para atender às solicitações de fixação do produto se tornou pouco escalável e controlável. Assim, busca-se analisar o histórico de interação dos cooperados e fatores externos que possam auxiliar na predição da fixação e realizar um controle de fluxo de caixa

mais assertivo, reduzindo prejuízos, destravando operações e melhorando a experiência do cooperado. É proposta uma ferramenta de análise de dados que auxilie fornecendo uma previsão que auxilie no acompanhamento das tendências de mercado, observando tendências e fazendo relações importantes para o processo de fixação.

### **SOLUÇÃO DESENVOLVIDA**

Foi realizada a integração das principais tabelas de dados que foram disponibilizadas pela cooperativa Integrada, sendo elas: o conjunto histórico de dados individuais de fixações por ano (2003-2021) e os dados climáticos de regiões que possuem estabelecimentos da cooperativa. Adicionalmente a essas tabelas, foram utilizados dados da cotação do dólar no mesmo período obtidos por APIs. Os dados foram analisados individualmente e passaram por análise exploratória. Inicialmente foram gerados conjuntos de DataFrames com features permitindo sua utilização em modelos de machine learning. Foi desenvolvido uma ferramenta para a regressão dos valores totais





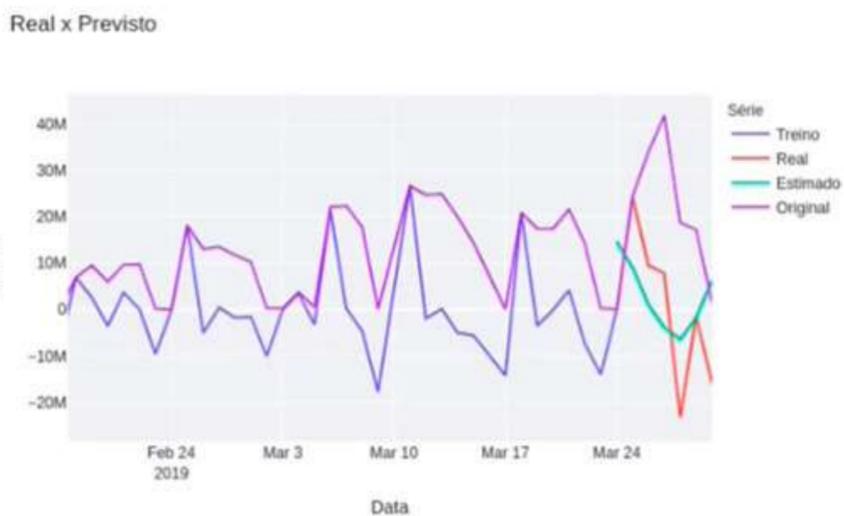
de fixação, fazendo uma estimativa para um intervalo de tempo específico, utilizando o modelo autoregressivo SARIMAX, devido ao seu bom desempenho para séries temporais e médias móveis com comportamento sazonal.

Parâmetros utilizados e estatísticas gerados pelo SARIMAX (esquerda); modelo de previsão de valores para um período de 7 dias (direita).

Dep. Variable:	y	No. Observations:	103
Model:	SARIMAX(3, 0, 0)x(0, 1, [1], 4)	Log Likelihood	-1789.222
Date:	Thu, 21 Oct 2021	AIC	3588.443
Time:	15:50:36	BIC	3601.419
Sample:	0	HQIC	3593.693
	- 103		
Covariance Type:	opg		

	coef	std err	z	P> z	[0.025	0.975]
ar.L1	0.8392	0.135	6.234	0.000	0.575	1.103
ar.L2	0.4035	0.156	2.579	0.010	0.097	0.710
ar.L3	-0.3955	0.139	-2.855	0.004	-0.667	-0.124
ma.S.L4	-0.8488	0.156	-5.454	0.000	-1.154	-0.544
sigma2	3.953e+14	1.03e-16	3.83e+30	0.000	3.95e+14	3.95e+14
Ljung-Box (L1) (Q):	0.12	Jarque-Bera (JB):	6.34			
Prob(Q):	0.73	Prob(JB):	0.04			
Heteroskedasticity (H):	0.45	Skew:	0.52			
Prob(H) (two-sided):	0.02	Kurtosis:	3.69			



# COOPERATIVA INTEGRADA

2

## PREVISÃO DE ENTRADA EM PESO NOS ESTABELECIMENTOS

### Residentes:

Igor de Paula Moura, Giovane Ronei Sylvestrin e João K. Nahra.

### Técnicas:

Machine learning (método de regressão Prophet, supervisionado por ExtraTreesRegressor e LGBMRegressor). [DS]: Análise exploratória de dados e visualização estatística. [Front-End]: Aplicação com Streamlit para visualização dos resultados e interação. Jupyter Notebook para treinos e inferências.

78

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

### OPORTUNIDADE DE NEGÓCIO

A predição dos volumes de entrega de uma determinada produção dos cooperados em diferentes unidades de uma cooperativa é capaz de auxiliar na logística de planejamento, uma vez que possibilita um melhor dimensionamento da contratação de mão de obra. A contratação ocorre atualmente com base em experiência vivenciada ou histórico, apresentando baixa assertividade



para a cooperativa. Dessa forma, é proposto um modelo preditor para a entrega dos volumes, a partir da análise de dados de abastecimento nos estabelecimentos e o cruzamento com os dados históricos de previsão do tempo.

### **SOLUÇÃO DESENVOLVIDA**

Foi desenvolvida uma ferramenta para a regressão dos dados de pesagens dos produtos agrícolas da cooperativa Integrada, fornecendo uma previsão que auxilia no acompanhamento dos estabelecimentos, observando tendências e fazendo relações importantes para o processo de pesagem. A partir dos dados disponibilizados foram realizados testes de diferentes modelos, sendo o modelo de regressão Prophet o que apresentou melhor ajuste e a implementação do dashboard foi feita em Python. Para realizar o treinamento desse modelo foram utilizados como dados de entrada os históricos de pesagens e os históricos de previsão de tempo (chuva, temperatura mínima e máxima) da cooperativa para diferentes períodos de tempo. A abordagem supervisionada para machine learning na previsão de séries temporais foi





utilizando algoritmos baseados em árvore de decisão. Os algoritmos ExtraTreesRegressor e LGBMRegressor apresentaram desempenhos satisfatórios para o horizonte de previsão, sendo ambos consideravelmente mais rápidos em processamento que os outros algoritmos.

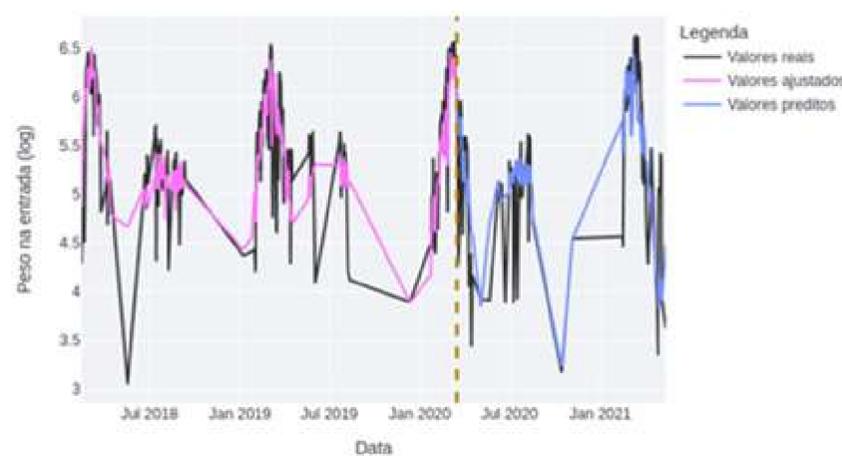
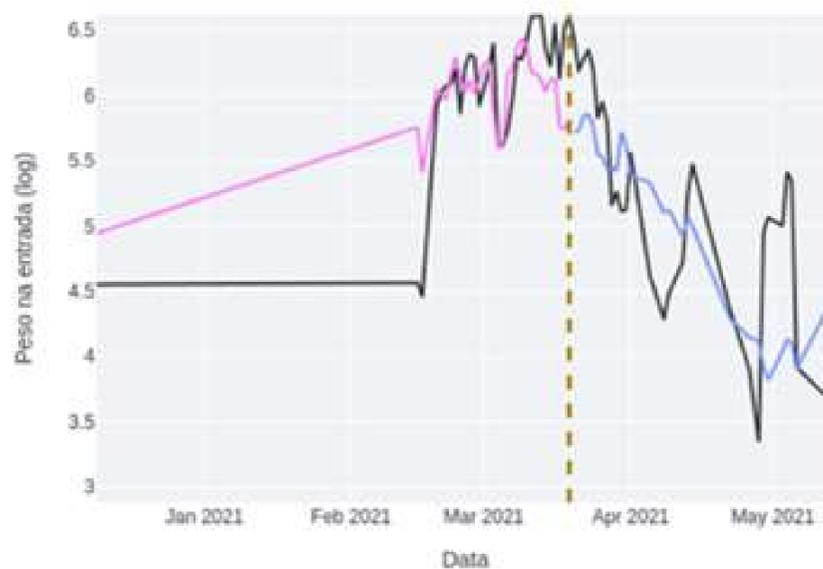
80

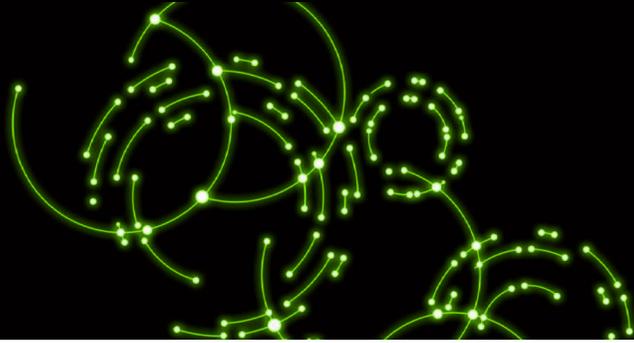
Previsão da entrada em peso da soja em grão, para um estabelecimento específico da cooperativa para um período de 30 dias (direita) e 120 dias (esquerda).

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





## COOPERATIVA INTEGRADA

3

### CLASSIFICAÇÃO DE ÁREAS AGRICULTURÁVEIS POR SATÉLITE

**Residentes:**

Ananda H. Coelho, Luis Gustavo Silva e Rafael H. Kimura.

**Técnicas:**

[IA]: Visão computacional (data augmentation) e Deep learning (MLP e CNN). [DS]: Análise exploratória e correlação de dados, extração de parâmetros e visualização de dados. [Front-End]: Jupyter com acesso aos dados.

81

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

### OPORTUNIDADE DE NEGÓCIO

O processamento digital de imagens de satélite permite uma visão completa de grandes áreas agriculturáveis e é uma necessidade recorrente na busca pela máxima produtividade. A identificação de diferentes culturas permite a tomada de atitudes corretivas, evitando perdas na produtividade da safra. Dessa forma, desenvolver um classificador de alta acurácia, sendo representativo para cada classe das diferentes culturas, pode ser uma ferramenta

estratégica para a cooperativa no sentido da definição de novos caminhos.

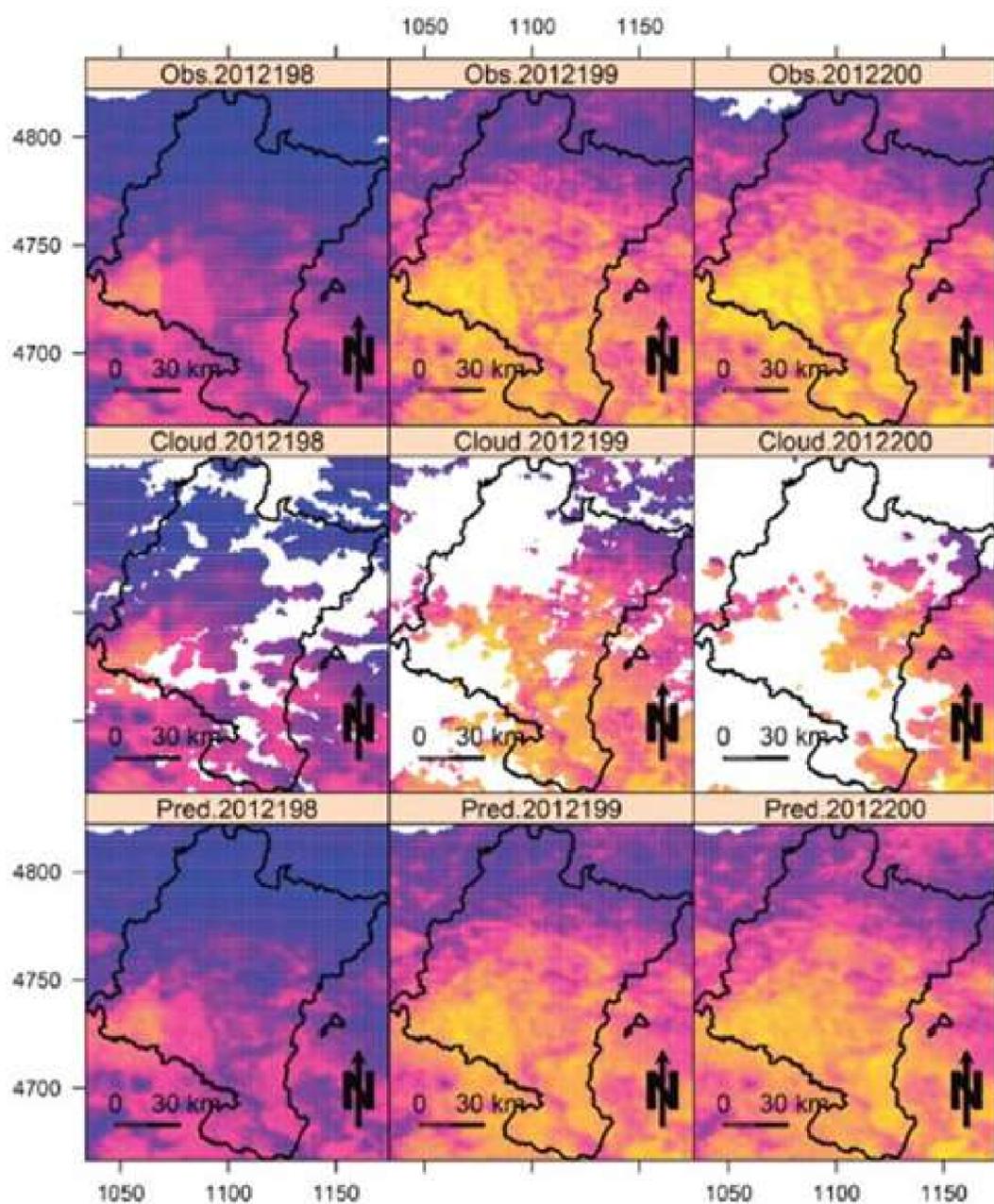
## SOLUÇÃO DESENVOLVIDA

O classificador de culturas foi desenvolvido através do estudo das imagens de satélite de diversas cidades do estado do Paraná, abrangendo culturas de verão e de inverno. Os dados foram explorados por meio de testes de novos modelos de classificação - tanto de pixels quanto de imagens segmentadas - em Jupyter Notebooks e scripts .py com o objetivo de melhorar a acurácia na classificação das culturas. A equipe adaptou a modelagem para um problema de classificação binária, testando a implementação de modelos especializados na identificação de uma única cultura por vez. Para uma melhor abordagem exploratória foram adicionados os índices de vegetação (NDVI, GNDVI e EVI), informações de localização (latitude e longitude) e valores de HSV. Foi analisado o comportamento dos índices de NDVI e GNDVI de cada imagem temporal para as cidades e culturas distintas. As imagens foram segmentadas em pedaços menores para





selecionar imagens com informações relevantes, como o desenvolvimento de uma cultura a partir do valor do NDVI, e passando por uma etapa de data augmentation. Após a segmentação de imagens em tamanhos uniformes, selecionando imagens baseadas no valor médio do NDVI, os dados foram modelados considerando o componente temporal em uma Rede neural Multilayer Perceptron (MLP) simples e em uma Rede Neural Convolutacional (CNN).





# NATOSAFE

1

## ANÁLISE DE DIGITAIS DE RECÉM-NASCIDOS

### Residentes:

Gabriel Jonas Aguiar, Karen Itsuki Taniguchi, Pedro Humberto Augusto Paz Teixeira Nunes.

### Técnicas:

[Data Science] - Análise exploratória, manipulação e organização de dados, processamento digital de imagens (binarização, filtragem e esqueletização). [IA] - Extração de características.

84

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

## OPORTUNIDADE DE NEGÓCIO

A impressão digital é um método de identificação amplamente utilizado por ser único para cada indivíduo, de fácil coleta e não invasivo. Na identificação de recém-nascidos essas características são extremamente importantes, pois evita adulteração ou violação em braceletes utilizados no método de identificação tradicional. Diferentemente da impressão digital de um adulto, recém-nascidos possuem a digital bastante compactada, não

completamente formada e durante o processo de coleta são mais agitados o que dificulta sua posterior identificação. Desta forma, foram desenvolvidas ferramentas para auxiliar o processo de identificação de recém-nascidos utilizando impressões digitais.

### **SOLUÇÃO DESENVOLVIDA**

A solução, entregue na forma de POC, consistiu no desenvolvimento de um algoritmo de pré-processamento para separar regiões conhecidas como vales e cristas das digitais, evidenciar as regiões de vale, extrair minúcias das digitais (terminações e bifurcações) e singularidades (formas de loops, espirais e deltas). O processo técnico consistiu em remover ruídos das imagens pela aplicação de filtro da mediana, técnica de binarização por limiarização adaptativa gaussiana e filtro Gabor para destacar os vales e esqueletização para extração das características de minúcias e singularidades.

85

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





86

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



A Figura mostra uma digital processada com o algoritmo em que os pontos verdes são bifurcações e os pontos azuis as terminações. As regiões de loops estão centradas nos quadrados laranjas e a região de delta nos quadrados vermelhos.

## NATOSAFE

2

MODELO GERADOR DE DIGITAIS ARTIFICIAIS E  
ALINHAMENTO DE DIGITAIS DISTORCIDAS**Residentes Equipe 1:**

Gabriel Jonas Aguiar, Germano Yoneda e Melanie Urdangarín.

**Residentes Equipe 2:**

Gabriel Sousa, Germano Yoneda, Victor Hugo Nagahama.

**Residentes Equipe 3:**

Gabriel Sousa e Tiago Pedrosa.

**Técnicas:**

[Data Science] - Processamento digital de imagens (ORB e Image Stitching). [IA] - Redes Neurais Convolucionais (CNN), Redes Neurais Generativas (GANs) e Random Sample Consensus (RANSAC).

87

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

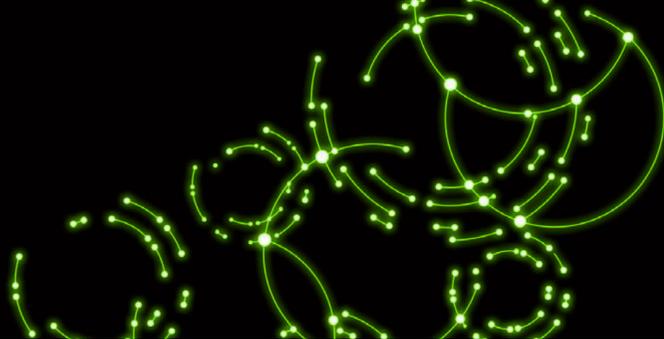
A biometria por impressão digital é um método de identificação único e extensivamente utilizado desde documentos pessoais de identificação como RG e Carteira de Trabalho até autenticação para transações bancárias em caixas eletrônicos

e uso de equipamentos eletrônicos pessoais como notebooks e celulares. Por isso, otimizar o processo de reconhecimento de impressões digitais é uma tarefa de extrema relevância nos dias atuais. Um dos métodos de melhorar o processo de reconhecimento é a criação de uma base de dados de impressões digitais geradas artificialmente.

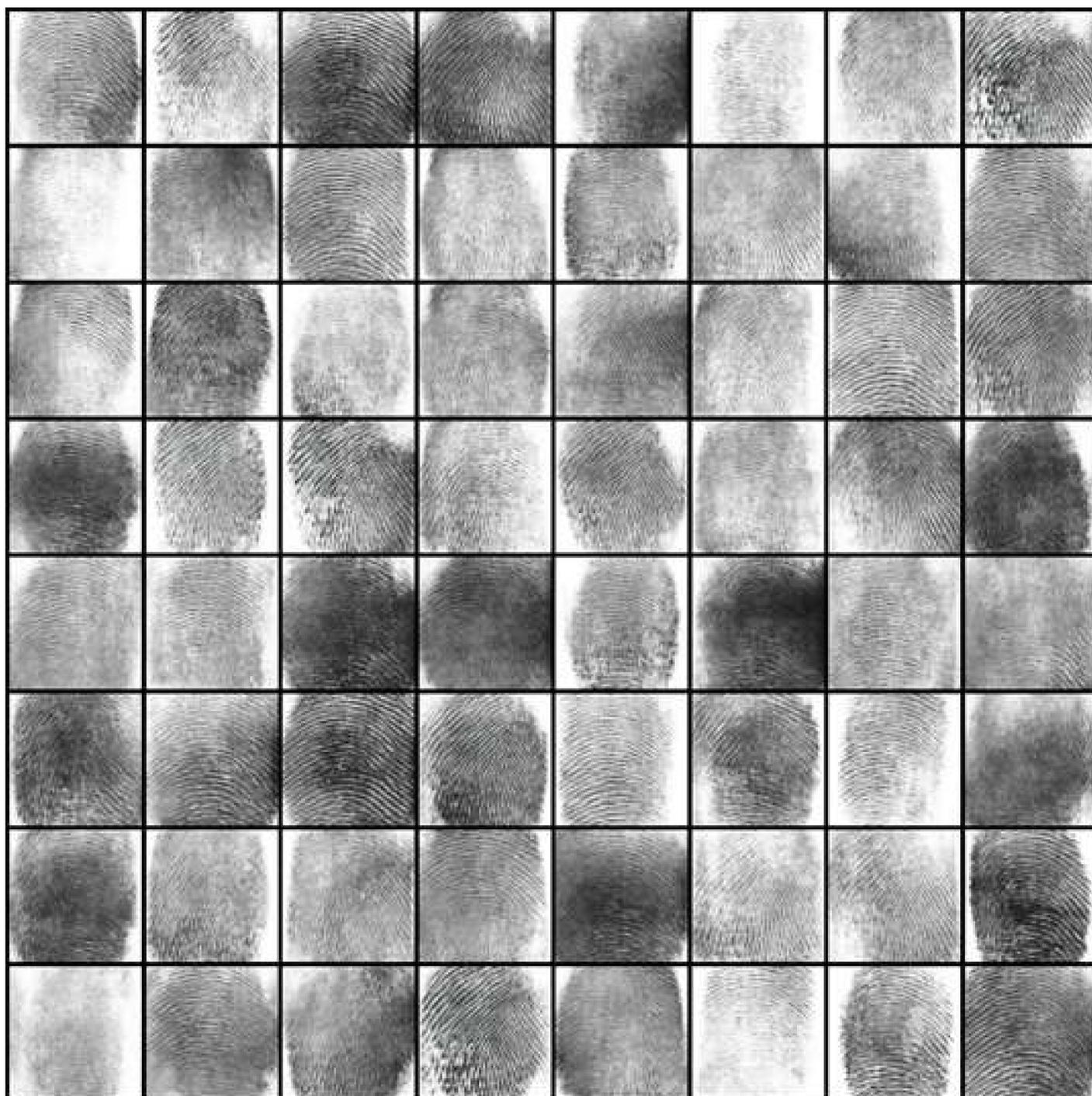
### **SOLUÇÃO DESENVOLVIDA**

Com o objetivo de aumentar o número de amostras de impressões digitais, foram geradas impressões digitais artificiais por Redes Neurais Generativas (GANs) e para que as imagens criadas tivessem alta resolução foi criado um modelo de super-resolução baseado em Redes Neurais Convolucionais (CNN). Nesse processo, o modelo recebeu uma imagem 64x64 e retornou a mesma imagem com resolução de 256x256. Além disso, a base de dados fornecida continha digitais distorcidas e, para isso, foi criado um algoritmo de alinhamento das imagens distorcidas utilizando as técnicas de processamento digital de imagens Oriented FAST and Rotated BRIEF (ORB) e Image Stitching com





## Random Sample Consensus (RANSAC).



89

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



A figura apresenta alguns exemplos de digitais geradas artificialmente ao fim do treinamento.



## OLIST

1

## ANÁLISE EXPLORATÓRIA DOS DADOS DE VENDEDORES E PRODUTOS

**Residentes:**

Anna Luiza Gomes e Igor Moura.

**Técnicas:**

[DS]: Análise exploratória, tratamento de dados.

90

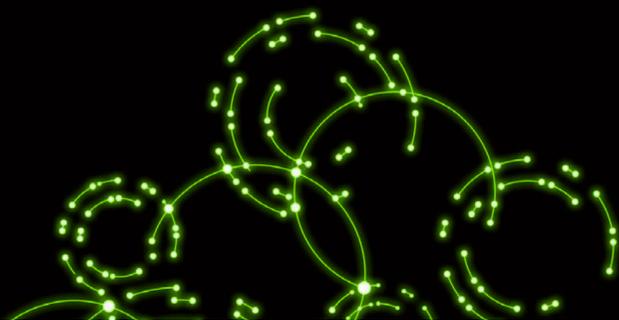
SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Atualmente a internet é o canal de vendas predominante de grande parte das empresas, assim sendo, os dados de usuário coletados podem revelar características importantes se analisados da forma correta. Por meio das informações disponíveis é possível entender o perfil dos clientes e propor soluções direcionadas que podem gerar grande impacto na utilização das plataformas, na criação de novas ferramentas e por fim nos lucros das empresas.

**SOLUÇÃO DESENVOLVIDA**

O foco inicial do projeto foi na organização dos

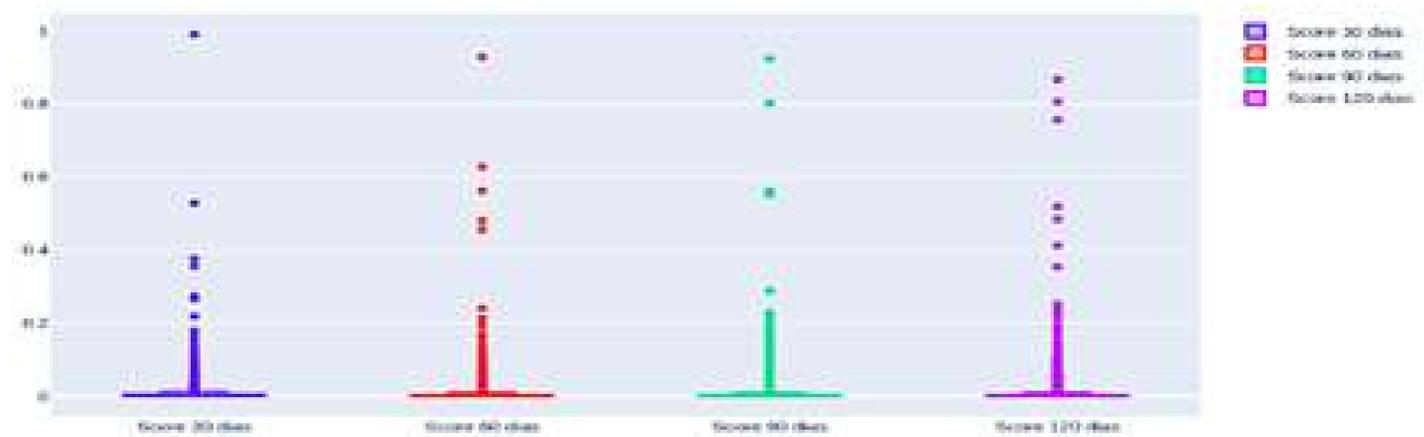


dados fornecidos pela empresa. Foram utilizadas nas análises duas tabelas de dados, uma tabela com registros de vendas, dados dos produtos vendidos, data, quantidade de reclamações; e uma tabela com informações detalhadas sobre os vendedores cadastrados na Olist. Inicialmente foi realizada a análise exploratória dos dados. Após a organização em pandas dataframes, foram realizadas as primeiras análises de correlação, produtos com mais reclamações e outros aspectos importantes. Em seguida foram desenvolvidos pipelines para organizar as informações mais relevantes baseando-se em todos os dados coletados. Ainda foram disponibilizados Jupyter notebooks para visualização gráfica dos resultados obtidos. Os algoritmos foram preparados de modo a possibilitar a inserção de tabelas no mesmo formato, sendo possível a fácil utilização com novos dados. O objetivo final do desenvolvimento foi a obtenção de insights para direcionar os negócios e as ações tomadas pela Olist. A partir da compreensão do comportamento dos usuários e da utilização de ferramentas de análise e direcionamento, é possível propor

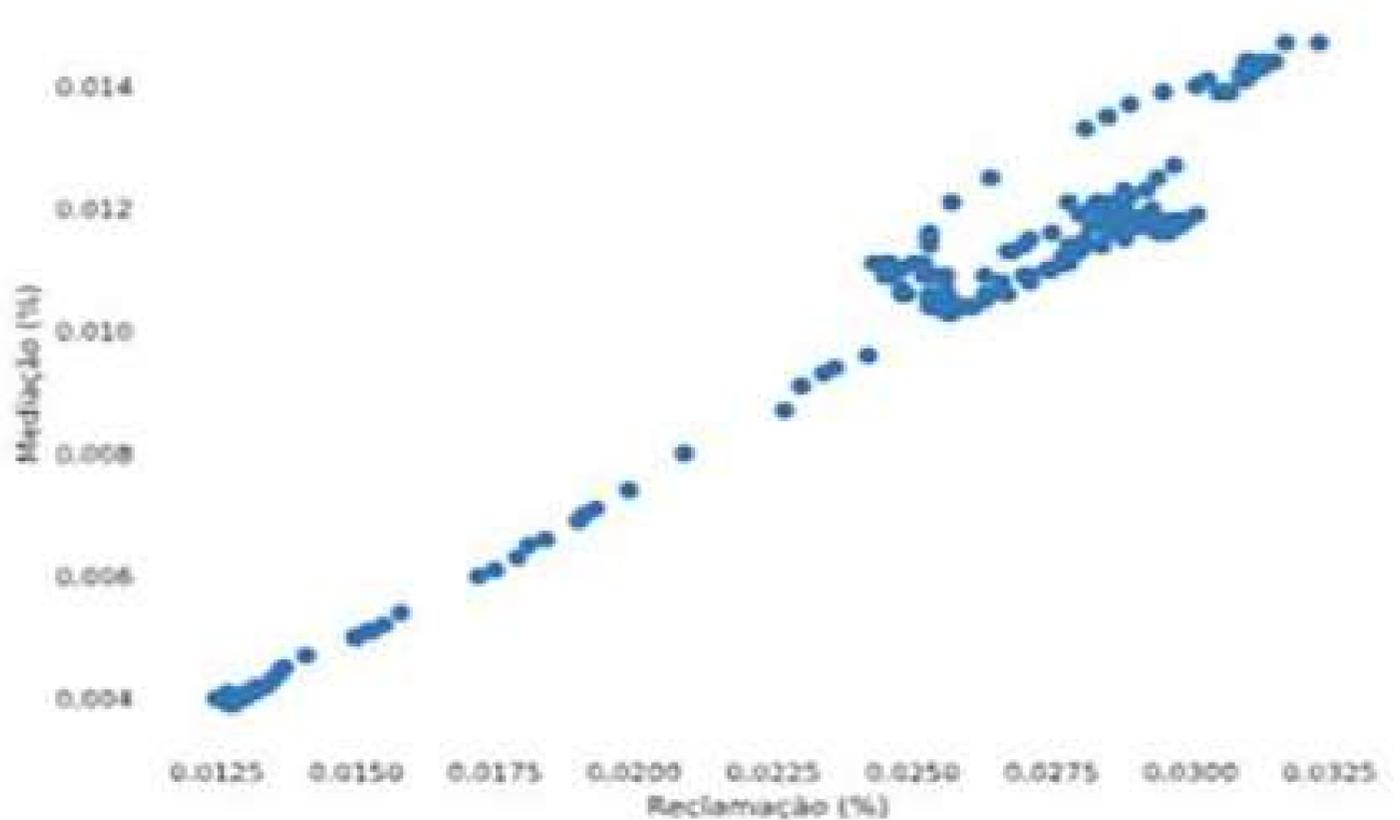




soluções que gerem maior competitividade nos anúncios, tendo uma maior retenção e captação de clientes.



Boxplot da quantidade de vendas dos sellers, por período de tempo.



Interação entre os indicadores de mediação e reclamação.



## OLIST

2

PREVISÃO DO ÍNDICE DE REPUTAÇÃO NO  
MERCADO LIVRE**Residentes:**

Ana Luisa Schmidt e Matheus Novaes.

**Técnicas:**

[DS]: Análise exploratória, tratamento de dados.

[ML]: Random Forest Regressor.

93

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

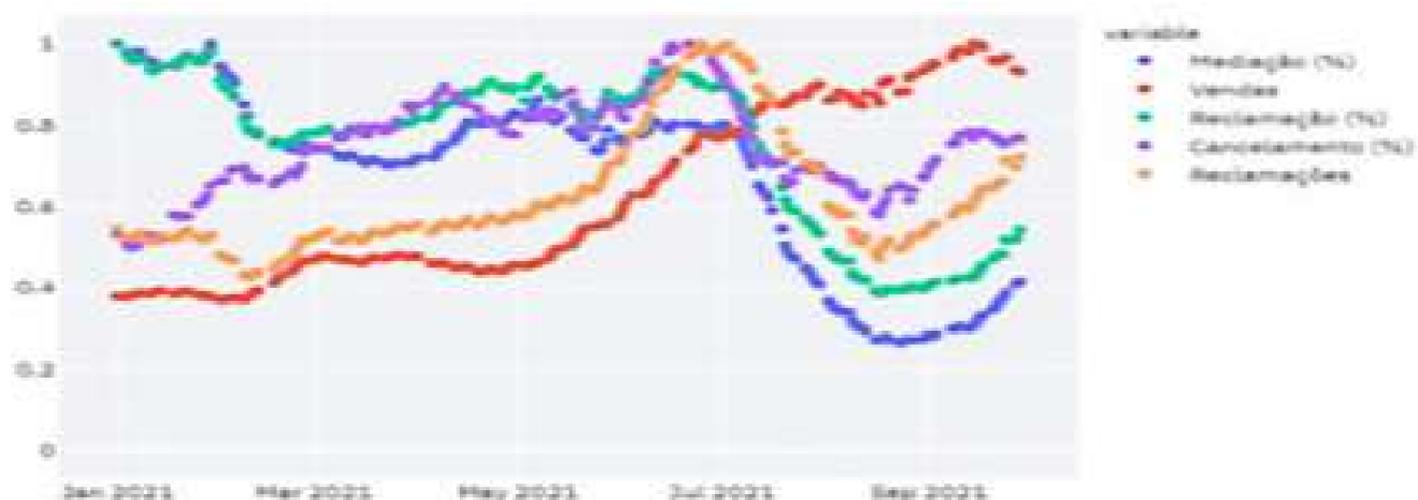
A possibilidade da tomada de ações preventivas por parte das empresas é de grande valor para a manutenção dos indicadores de reputação nos marketplaces. Pensando nisso, a Olist, que conta com um extenso histórico de vendas trouxe como demanda o desenvolvimento de uma ferramenta de previsão baseada em técnicas de machine learning. Uma ferramenta de visualização, ou Dashboard, também agrega valor ao produto, uma vez que permite a melhor identificação dos padrões de correlação, a entrada de novos dados, e a atualização

das previsões de acordo com o interesse da empresa.

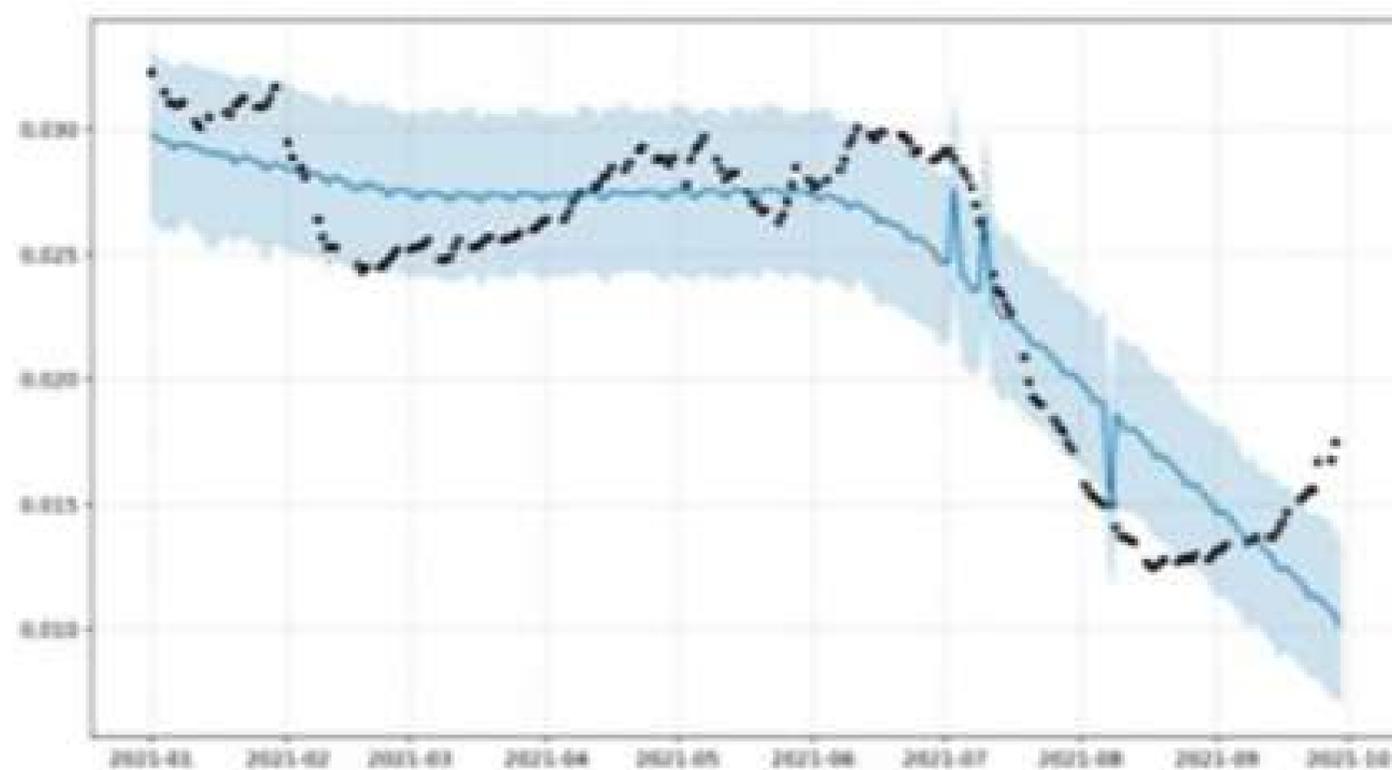
### SOLUÇÃO DESENVOLVIDA

O foco inicial foi na análise exploratória dos dados e na construção e organização de dataframes com as informações mais relevantes. Em seguida foram feitos testes com modelos de regressão utilizando algoritmos do Scikit-learn como RandomForestRegressor e realizou-se uma seleção de features inteligente com o objetivo de otimizar o modelo e encontrar o impacto das features selecionadas nos três indicadores: Reclamação, Mediação e Cancelamento. Com os resultados foram gerados instrumentos gráficos para uma visualização mais clara do que foi analisado. Finalmente desenvolveu-se um dashboard utilizando o streamlit, onde foi possível acompanhar a previsão do comportamento de cada um dos indicadores de reputação, para 1, 7, 15 ou 30 dias a partir dos dados de histórico analisados.





Comportamento dos indicadores ao decorrer do tempo.



Valores reais (em preto) e a predição (em azul) para o indicador de reclamação.

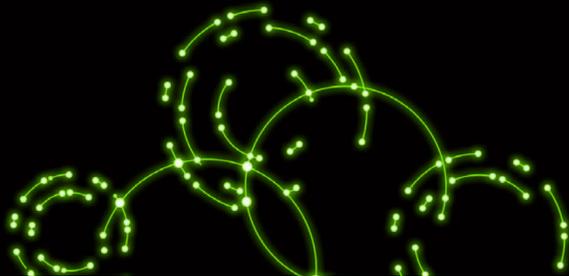
95

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





## OLIST

3

MINERAÇÃO DE ATRIBUTOS DAS DESCRIÇÕES  
DE ANÚNCIOS DE PRODUTOS**Residentes:**

Arthut Lins e Luis Gustavo.

**Técnicas:**

[IA] spaCy e NLTK.

96

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Modelos matemáticos podem ser utilizados como ferramentas consistentes para auxílio a tomada de decisão no ambiente organizacional. Neste sentido é importante proporcionar um procedimento organizado e consistente que auxilia na tarefa de gestão de recursos humanos, materiais e financeiros de uma organização, abrangendo diversas áreas, modelos e algoritmos que permitem ao gestor tomar decisão em problemas complexos. Assim, a otimização e padronização dos atributos e seus valores em um modelo gera um ganho de conhecimento do produto. Tornando o processo de mineração de atributos e seus valores mais

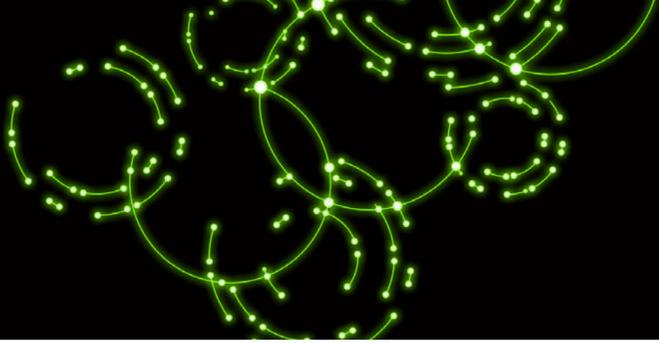


claros e mutáveis.

### **SOLUÇÃO DESENVOLVIDA**

Foi possível criar uma tabela que para cada anúncio, os atributos contidos nele foram identificados de acordo com seus valores. Comparamos o resultado com os dados brutos fornecidos e encontramos um resultado com boa confiabilidade. Outro resultado importante é um dicionário que, pelos parâmetros da função criada, pode ser usado tanto para um produto, para um tipo de produto, para um grupo de produto ou um dicionário geral, dependendo dos parâmetros escolhidos na função. Assim, a capacidade do algoritmo de fornecer uma mineração mais confiável que os fornecidos por terceirizadas. Existem algumas dificuldades quanto a confirmação de uma métrica absoluta, uma vez que nosso minerador pode ter assimilado palavras equivocadas também, uma vez que o dicionário que criamos e usamos no algoritmo tem origem nos atributos da empresa.





## OLIST

4

## RECONHECIMENTO DE IMAGENS PARA CATÁLOGO

**Residentes:**

Ana Luisa Schmidt e Matheus Novaes.

**Técnicas:**

[IA] CNN e [Plataforma] Streamlit.

98

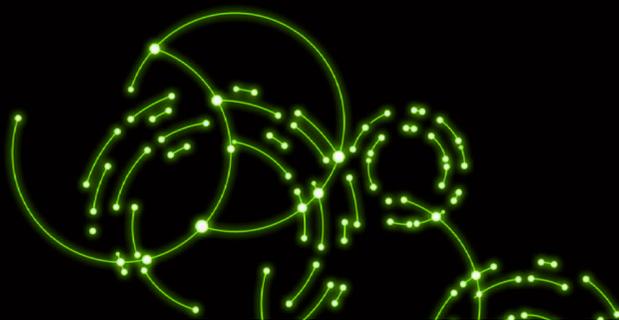
SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Cadastro do produto é uma parte muito significativa da venda em plataformas de e-commerce, o que engloba tanto a descrição, quanto foto e seleção correta do tipo de produto. Visando facilitar o pré-cadastro do produto realizado pelo vendedor, a Olist optou por desenvolver, uma ferramenta capaz de, a partir de uma imagem do produto, identificar o código do mesmo.

**SOLUÇÃO DESENVOLVIDA**

Foi possível criar uma ferramenta que ao receber uma imagem indica a qual produto do catálogo ela pertence. No futuro irá auxiliar no pré-



cadastro dos produtos realizado pelo vendedor. Nesse projeto, a inteligência artificial está presente em quase todos os passos, desde o reconhecimento do produto contido na imagem, a classificação do objeto e a correlação do produto da imagem com o do catálogo. Além disso, como parte do produto, foi entregue uma visualização do produto na plataforma Streamlit onde é possível simular o uso da mesma.



## PADO

1

## ANÁLISE DE CRÉDITO E LIBERAÇÃO DE PEDIDO

**Residentes:**

Germano Yoneda, Sergio Lemos Hermsdorff e Roberto.

**Técnicas:**

KMeans, DBSCAN, Gradient Boosting Classifier.  
[Estatística]: Análise exploratória em Boxplot, Mapa de calor. [Main-Frame]: Streamlit

100

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

A empresa pretende alterar seu fluxo manual de análise de crédito para um modelo automatizado do processo de liberação ou bloqueio de crédito utilizando um modelo de aprendizado de máquina, a fim de diminuir o número de pedidos analisados individualmente pela auditoria manual, realizando a inferência de aprovações manuais para aprovações automáticas que aumente ou mantenha a qualidade da avaliação de crédito e diminua assim gastos com recursos humanos.



## SOLUÇÃO DESENVOLVIDA

Primeiramente foi realizada uma análise exploratória dos dados iniciais da PADO para identificar assim possíveis correlações entre os dados. Após isso, aplicou-se as regras de negócio para assim gerar um score e posteriormente aplicação de modelos não supervisionados como KMeans e DBSCAN que apresentaram resultados não satisfatórios devido a ausência de uma correlação positiva entre os dados disponíveis.

Após isso fez-se a remoção de dados vazados e uma nova tentativa de modelo utilizando agora um modelo de aprendizado supervisionado nos novos dados variando as classes. O novo modelo, o Gradient Boosting Classifier, apresentou melhores resultados onde foi realizado o agrupamento das classificações fornecidas de acordo com o esperado para cada um dos casos, separando-os em Aprovados Automaticamente e os demais casos. Depois, aplicou-se técnicas de otimização sendo observada uma curva de aprendizado do sistema que demonstra melhora na predição dos casos conforme é aumentada

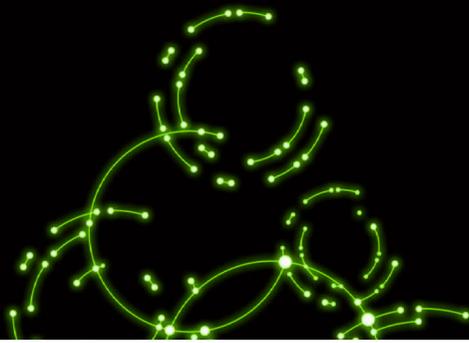
# 101

SUMÁRIO

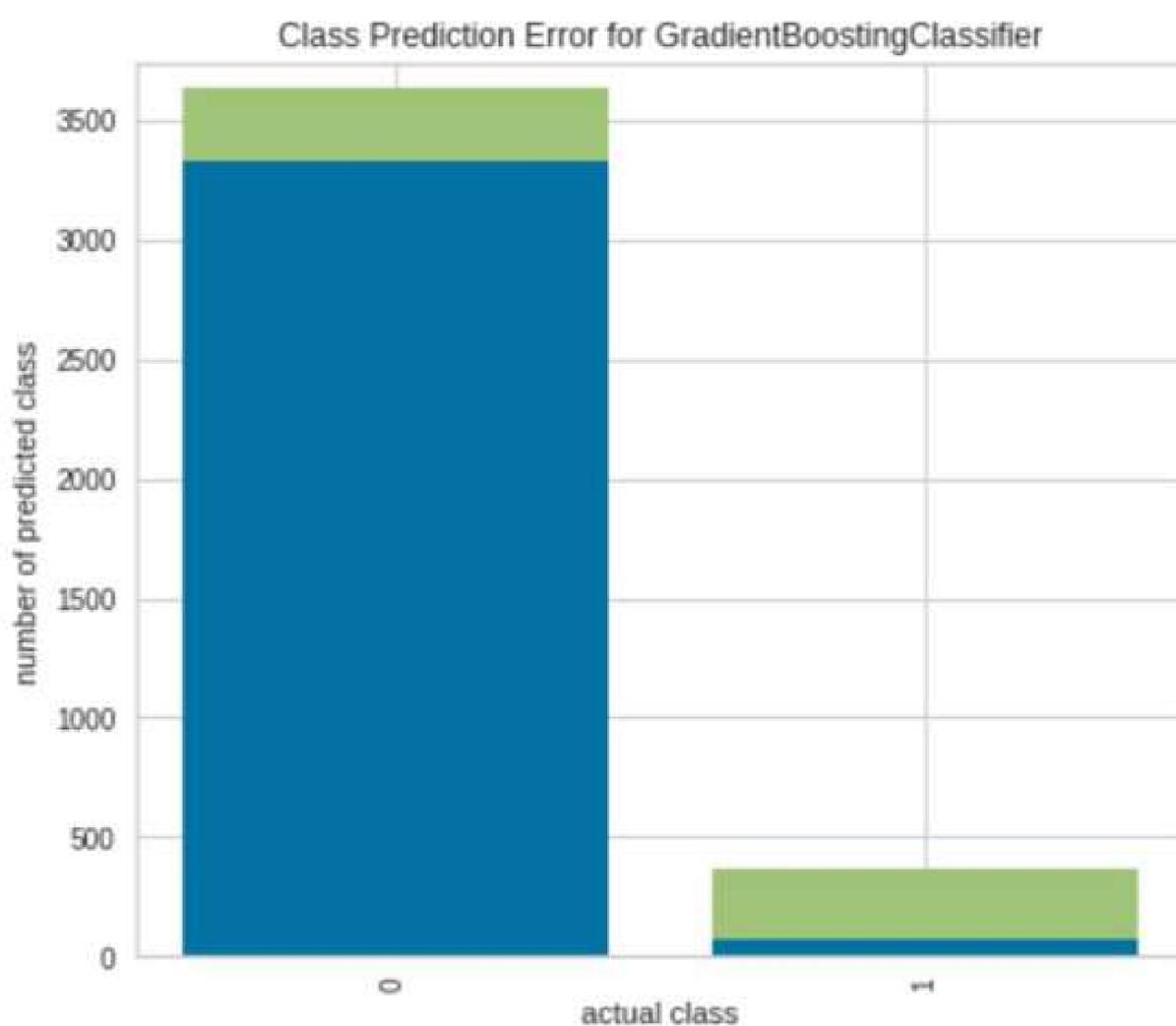


SUMÁRIO  
CASES DO  
PROGRAMA





a base de dados fornecida, sugerindo que o sistema possa chegar a melhores resultados também utilizando a base de dados completa em sua predição.



102

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## PADO

2

## LEITURA DE DOCUMENTOS

**Residentes:**

Marcia Rodrigues, Pedro Nunes e Tiago Pedrosa.

**Técnicas:**

[IA]: OCR, ROI, Flash, chamada POST. [Main-Frame]: python

103

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

A empresa visa criar uma ferramenta que objetiva tornar o processo de RH mais simples e rápido. Hoje todos os documentos digitalizados e enviados pelos futuros funcionários da empresa precisam ser analisados e as informações enviadas pelos candidatos precisam se transcritas de forma manual para o sistema da empresa sendo este demorado e passível à erros. Com o novo processo será possível automatizar esta etapa através de uma ferramenta interativa que retorna um json com dados estruturados a partir do

desenvolvimento de um OCR que realiza a extração e a estruturação de informações de documentos pessoais. A POC está relacionada ao desenvolvimento de um uma ideia viável, funcional, e que pode ser escalável para um produto, mas sua entrega não está condicionada ao desenvolvimento de softwares com interfaces.

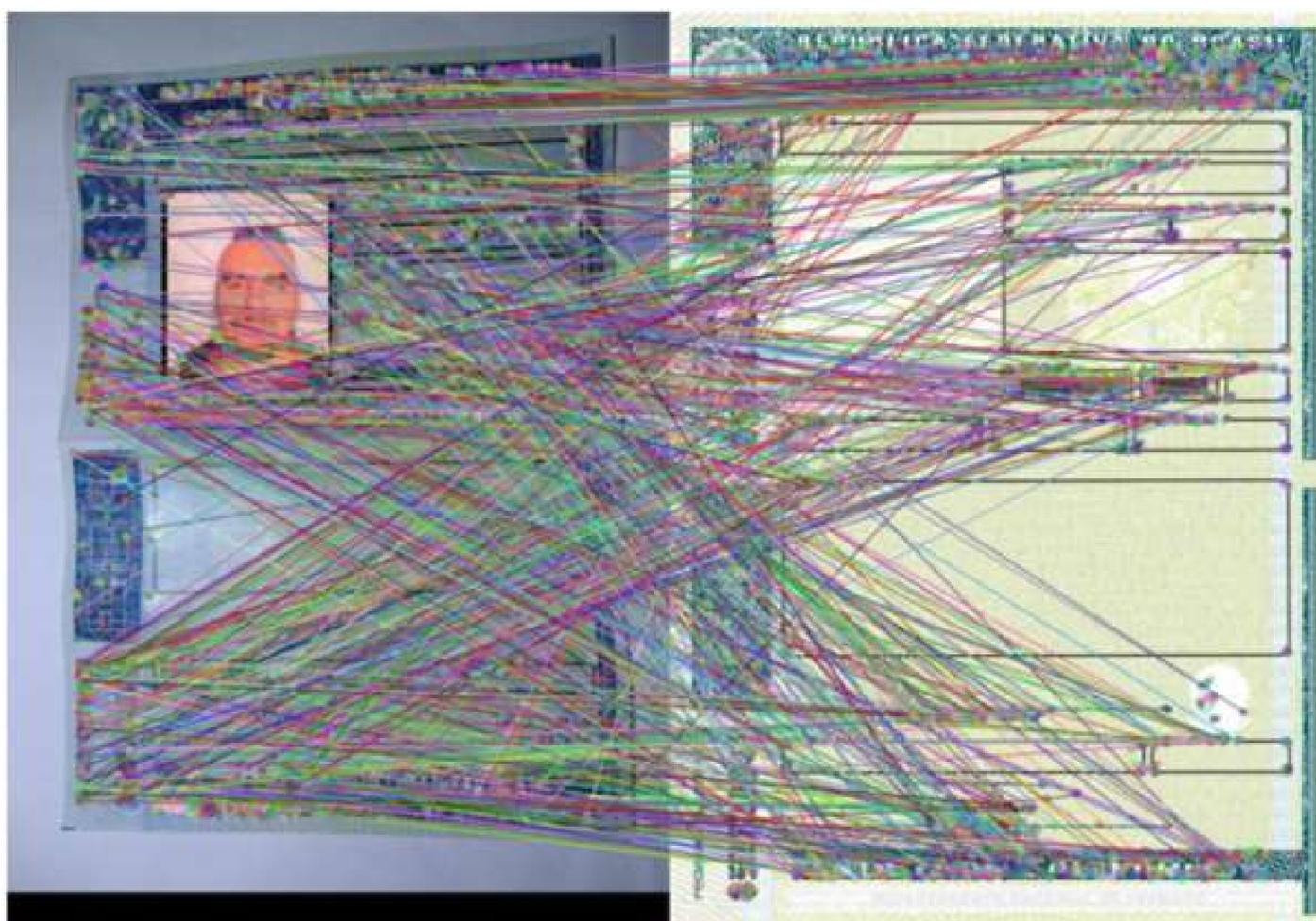
### **SOLUÇÃO DESENVOLVIDA**

Primeiramente era necessário realizar a extração de informações em documentos digitalizados, realizando assim uma adequação do documento à algumas normas da empresa, como o alinhamento ao plano de fundo para posteriormente utilizando o método de OCR e de adequação a ROIs realizar a extração das informações do documento. Posteriormente a isso, foi criado um .json que é enviado para uma API para posterior armazenamento em banco de dados.

**104**

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA



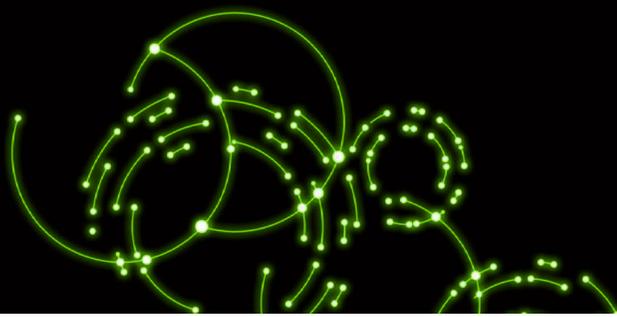
105

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





## PADO

3

## CONTAGEM DE MATERIAIS

**Residentes:**

Karen Taniguchi, Larissa Luciane Ferreira e Márcia Rodrigues.

**Técnicas:**

[IA]: Fastrer R-CNN, Pytorch Save, HoughCircles e Canny. [Main-Frame]: python e bot do Telegram.

106

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

A empresa visa criar um processo de contagem de peças (vergalhões ou perfis de cadeados) automatizado que, atualmente é feita de maneira mecânica e visual por funcionários sendo um processo moroso e muito suscetível a erros.

**SOLUÇÃO DESENVOLVIDA**

Para desenvolver a solução para esse problema foi necessário utilizar funções de contagem de objetos com fotos contendo os materiais de interesse para assim realizar a contagem dos perfis de cadeados e vergalhões. Para a



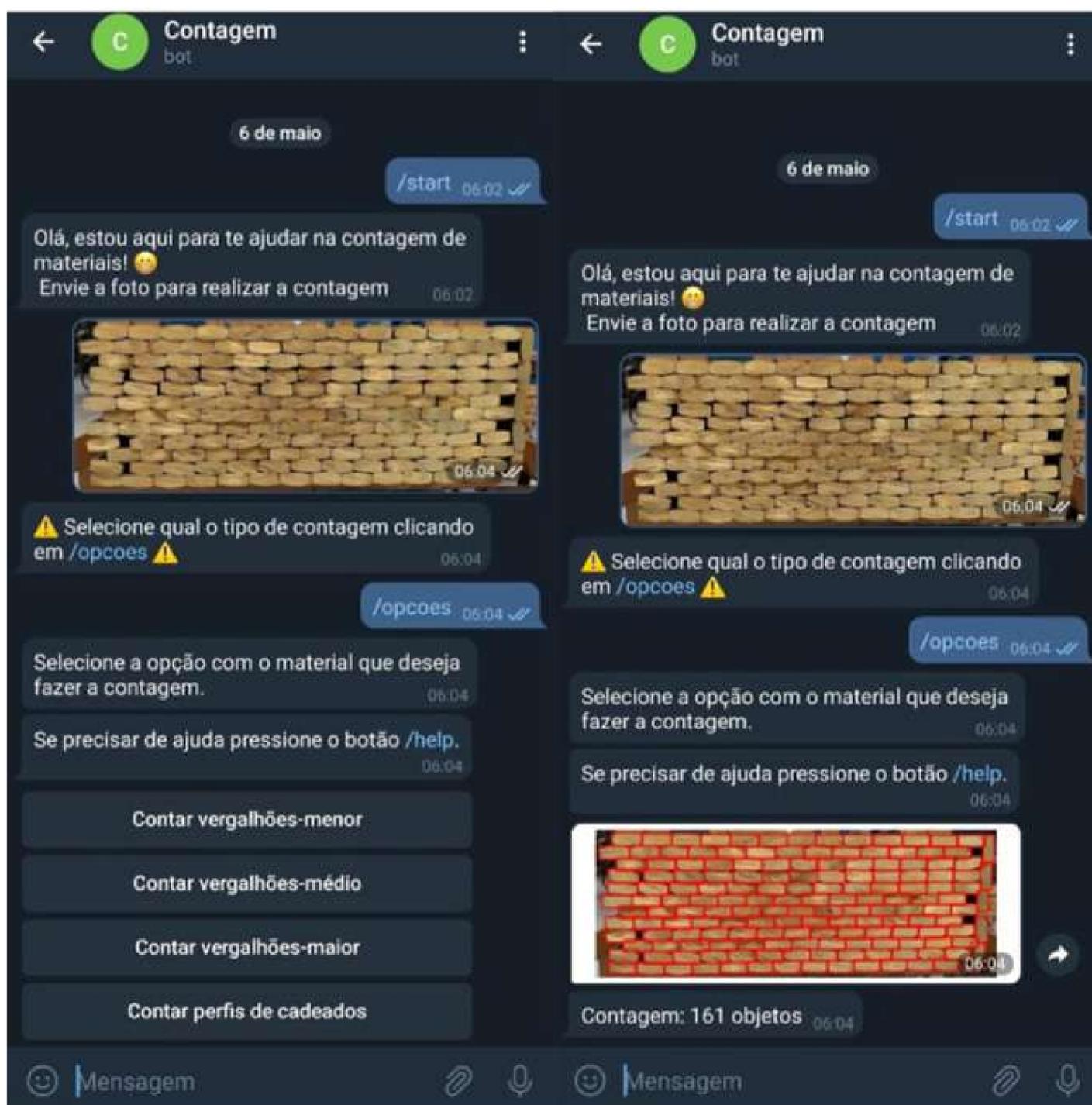
contagem de perfil de cadeado foi utilizada uma ferramenta de deep learning chamado Faster R-CNN salvando o modelo com o Pytorch Save para que não seja necessário fazer o processo de treinamento novamente. Além disso, foi realizada a remoção de bounding boxes sobrepostas e com área maior ou menor que 40% da média das demais.

Já para a contagem de vergalhões a metodologia abordada foi baseada na função HoughCircles implementada pela biblioteca OpenCV, com o objetivo de detectar formatos circulares em uma imagem utilizando também uma função de detecção de bordas denominada de Canny. Em virtude da natureza paramétrica da função HoughCircles, em contraste aos diferentes tamanhos e alinhamentos das imagens de vergalhões; o grupo de vergalhões foi dividido em três subgrupos, classificados de acordo com o diâmetro do perfil de suas peças e com a distância mínima entre os centros dos círculos. Após isso fez-se a integração das funções de contagem com um bot do Telegram sendo criada assim uma interface entre o operador e o POC





desenvolvido.



108

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## PADO

4

## IDENTIFICAÇÃO DE SEGREDO DE CHAVES

**Residentes:**

Melanie Gissel Urdangarin Gamarra e Natalia da Silva de Sá.

**Técnicas:**

[IA]. [Main-Frame]

109

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

As chaves confeccionadas pela empresa possuem um certo número de vales com possíveis alturas diferentes, sendo assim, existem muitas possibilidades de combinação de segredo. A identificação de um segredo de chave pode ser solicitada pelo cliente a fim de gerar uma nova chave ou um novo cadeado/fechadura com o mesmo segredo. A detecção automática dos segredos de chaves facilita a geração de novas chaves ou cadeados/fechaduras com um mesmo segredo e pode ser implementada para uso de clientes e sua identificação pode ser feita



utilizando foto da chave e gabarito.

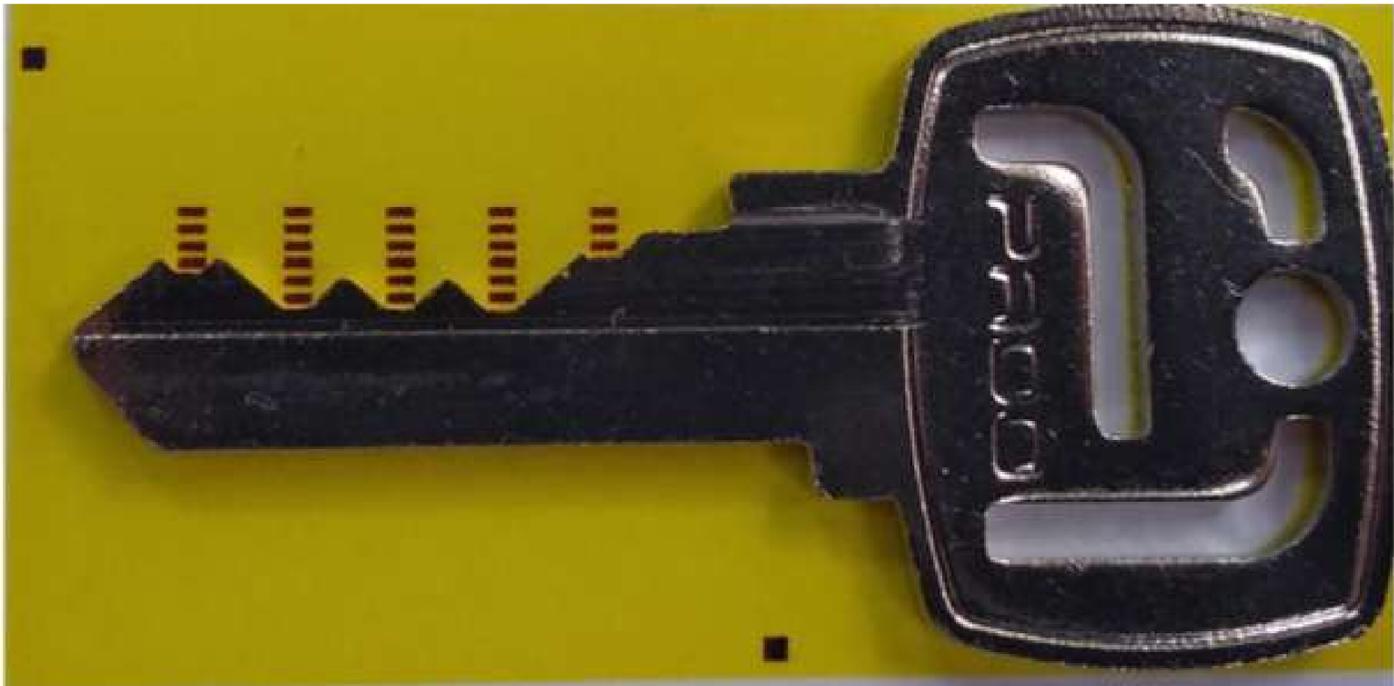
### **SOLUÇÃO DESENVOLVIDA**

Para desenvolver a entrega solicitada foram necessárias 4 etapas: reconhecer e recortar o gabarito na imagem, contar marcações, reconhecer a partir da canaleta se a chave é de lado esquerdo ou direito e retornar o número do segredo.

Para a primeira etapa a imagem é transformada para o espaço de cores HSV que, passa por processos de conversão de cores a cinza e binarizações para que posteriormente sejam detectados os contornos da imagem. Por último, foi utilizado um filtro de áreas. Para a contagem de marcações foi necessário identificar e recortar a área da foto contendo as marcações utilizando equações que levavam em consideração a padronização da disposição das chaves no gabarito e, em seguida realizar um processo de tratamento de imagem similar ao realizado na primeira etapa. Para a identificação da canaleta foi necessário utilizar a ferramenta HoughLinesP da biblioteca OpenCV e alguns



tratamentos como alteração de contraste e brilho, binarização e erosão a fim de encontrar as linhas da canaleta.



111

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## PRIDE

1

## REFATORAÇÃO DE CHATBOT

**Residentes:**

Ana Luisa Schmidt, Ananda Coelho e Igor Moura.

**Técnicas:**

[IA] Arvore de Decisão, NLP e [Plataformas] BLiP e Watson IBM Cloud.

112

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Tendo em vista que a utilização de chatbots para o atendimento ao cliente é um grande facilitador para a empresa no aspecto de economizar o atendimento por razão das respostas rápidas de informações triviais contidas em FAQs e similares, o objetivo da Sprint foi a melhoria dos chatbots existentes por meio de refatoração, com melhora no desempenho, engajamento dos clientes, redução no tempo e no índice de evasão de clientes nessa fase de atendimento.

**SOLUÇÕES DESENVOLVIDAS**



Foi desenvolvido a refatoração com base em Arvore de Decisão e NLP com o uso da plataforma BLiP e o uso da plataforma Watson da IBM para o armazenamento em nuvem dessas melhorias. Tendo sido construído uma melhoria no fluxo das mensagens, desde o contato inicial até o momento de término da conversa ou que foi encaminhado para o serviço de atendimento ao cliente da empresa. Dentro desse fluxo, foram usadas algumas melhorias para reconhecer, construir novas palavras chaves e exclusão de palavras que não melhoram o desempenho do bot. Para além disso também foram disponibilizadas informações novas, como envio de endereços e imagens para melhor comunicação com o cliente.

Foi utilizada Arvore de Decisão e NLP (Processamento de Linguagem Natural, em português) para melhorar as decisões no momento em que a conversa acontece com o bot, com intuito de otimizar o fluxo de conversa, o que por si tem os efeitos de melhorar o tempo de atendimento, índice de evasão, eliminação de bugs e menos desgaste

**113**

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

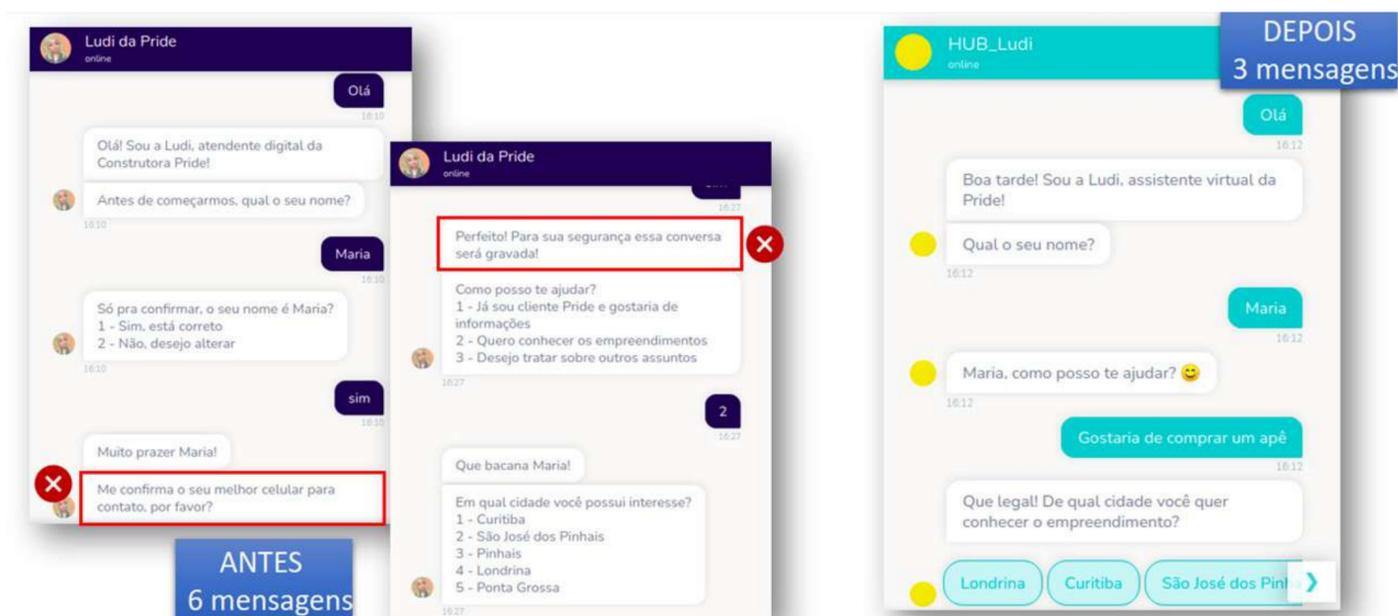
do cliente e conseqüentemente menor evasão. Após uma pesquisa, foi selecionada o Watson da IBM como provedor de IA pelo desempenho nos parâmetros: datas de lançamento, plano do provedor e artigo acadêmico relacionado ao desempenho do mesmo. Com isso, foi detectada partes no fluxo da conversa com maior urgência de atendimento pessoal e outras que imagens e endereços seriam o suficiente. E também, atrás da Arvore de decisão foi possível entender quais palavras deveriam ser chaves para alguns fluxos de conversa, fazendo que o sistema pudesse sempre retreinar com novas palavras e continuamente melhorando o desempenho do bot. Já as palavras com baixo desempenho no papel de palavra-chave, com o tempo podem ser excluídas do banco de dados.

114

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





# PRIDE

2

## APRIMORAMENTO DE CHATBOT

### Residentes:

Anna Luiza Gomes, João Victor Kawazoe e Igor Nieto.

### Técnicas:

[IA] Arvore de Decisão, NLP e [Plataformas] BliP e Watson IBM Cloud.

115

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

### OPORTUNIDADE DE NEGÓCIO

Pelas mesmas razões da Sprint passada, ainda são necessárias melhorias para atingir um melhor engajamento nos chatbots e nessa, foi aprimorada algumas funcionalidades e foi desenvolvido para o ambiente de WhatsApp.

### SOLUÇÕES DESENVOLVIDAS

A fim de aumentar o engajamento dos clientes, foram implementadas melhorias para além das plataformas vistas e dos sites das empresas.

A implementação de bots no WhatsApp trazem novas dificuldades, foi visto que palavras como

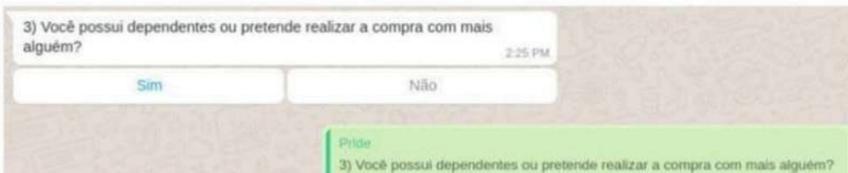
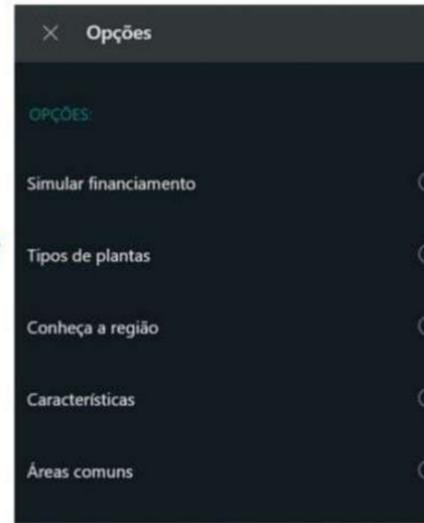
“bom dia” e “me chamo” aparecem com muita frequência, o que poderia fazer com que elas se tornassem, erroneamente, palavras-chave. Foram desenvolvidas soluções relacionadas a remoção de palavras de cumprimento, questões relacionadas a ortografia e recuperação de e-mail do usuário. Também houveram funcionalidades implementadas no aplicativo, como: menu inicial para conversa com bot em áreas específicas, botões para direcionar a conversa durante o fluxo, opção de “voltar” para não ter que reiniciar a conversar por uma escolha errada e caso mesmo assim o usuário não se sinta confortável, após 3 tentativas sem boa resposta no fluxo da conversa, é oferecida ajuda pelo bot.

Foram usadas a plataforma BLiP e Watson para as finalidades despostas na Sprint passada, com o uso de diversas APIs foram melhoradas as interações durante o fluxo, tanto com a solicitação de mensagens passadas quanto para redirecionamento, sem falar da API que auxiliou nas informações para desenvolvimento da produção de chatbot no ambiente do WhatsApp.

116

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA





## PRIDE

3

## CHATBOT MULTIPLATAFORMA

**Residentes:**

João Victor Kawazoe, Matheus Novaes.

**Técnicas:**

[IA] Arvore de Decisão, NLP e Redes Neurais.  
[Plataformas] BliP e Watson IBM Cloud.

118

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Seguindo a mesma ideia das Sprints passadas, o projeto chegou em um novo patamar que é observar como os chatbots podem se integrar em diversas plataformas, não só no site, mas Whatsapp (na Sprint passada) e agora no Telegram também. Com isso surge a ideia de construir uma proposta multiplataforma para integrar as informações e alcançar um maior público e engajamento com os clientes.

**SOLUÇÕES DESENVOLVIDAS**

Foram aprimorados os chatbots existentes



e não só em relação a desempenho, mas a escalabilidade, capturando as informações de usuários que solicitam atendimento fora do horário e documentado o fluxo. Assim, dando a empresa, oportunidade de uma expansão organizada com conhecimento mais profundo das demandas que surgem por essa via.

**119**

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## PRIDE

4

## REFATORAÇÃO DE CHATBOT

**Residentes:**

Ana Luisa Schmidt, Ananda Coelho e Igor Moura.

**Técnicas:**

[IA] Arvore de Decisão, NLP. [Plataformas] BliP e Watson IBM Cloud.

120

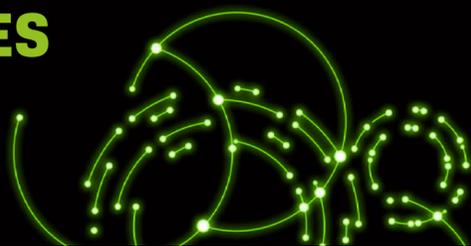
SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Nesse estágio do projeto, a oportunidade em si foi a entrega do projeto. Que dado a complexidade e o desenvolvimento abordado nas Sprints passadas e nessa, fica claro que atingiu o objetivo de melhorar fluxo, aparência, usabilidade e integração com os sistemas mencionados.

**SOLUÇÕES DESENVOLVIDAS**

Foram feitos ajustes no ambiente de produção focados em captura de leads e melhora na experiência de usuário, foi revertido todas



as ações relacionadas a inteligência artificial e foram também implementados menus estáticos. E para os atendimentos feitos em horário comercial, foi implementado um fluxo para atendimento humano. Também foram construídos diversos fluxos para situações diferentes e observação de inatividade para desligamento do bot.

**121**

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

# QUALITYSTORM

1

## API DE INSPEÇÃO DE IMAGENS E PROCESSOS

**Residentes:**

Vitor Nagahama e Roberto Moure

**Técnicas:**

Processamento de Imagens, Data Augmentation

122

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

### OPORTUNIDADE DE NEGÓCIO

A empresa tem um produto que é capaz de criar linhas de inspeção para um determinado objeto fazendo-se assim testes de inspeção, tendo em cada teste um conjunto de itens de inspeção onde o operador manualmente realiza a inspeção e tira uma foto do objeto para comprovação de sua decisão que poder ser aprovado, aprovado com ressalvas, reprovado e liberado. Este processo é extremamente demorado e, para reduzir o tempo de análise é possível criar uma IA para detecção de reprovações em itens de inspeção.

## SOLUÇÃO DESENVOLVIDA

Para criar a IA fez-se necessário inicialmente tratar a base de dados recebida, sendo eliminadas fotos testes e inspeções sem fotos anexadas. Após isso fez-se a análise exploratória para escolher o melhor candidato para o desenvolvimento da solução para que ela seja genérica para qualquer item de inspeção.

Para o Desenvolvimento da IA foi feito o treinamento de um detector de imagem utilizando o SOTA com treinamento em GPU capaz de dizer se a inspeção foi aprovada ou reprovada. Este treinamento se dividiu em pré-processamento dos dados, separação treino e teste, Data-augmentation para ajuste de classificação desbalanceada e treino. Já na Construção da API foi feito um desenvolvimento orientado a objeto sendo disponibilizado em 3 end-points: train\_evaluation, train e predict onde, o primeiro treina a IA e tem sua acurácia avaliada, o segundo treina todos os conjuntos de dados e, o terceiro é o ambiente de produção onde, é enviada a imagem e a inspeção é

123

SUMÁRIO

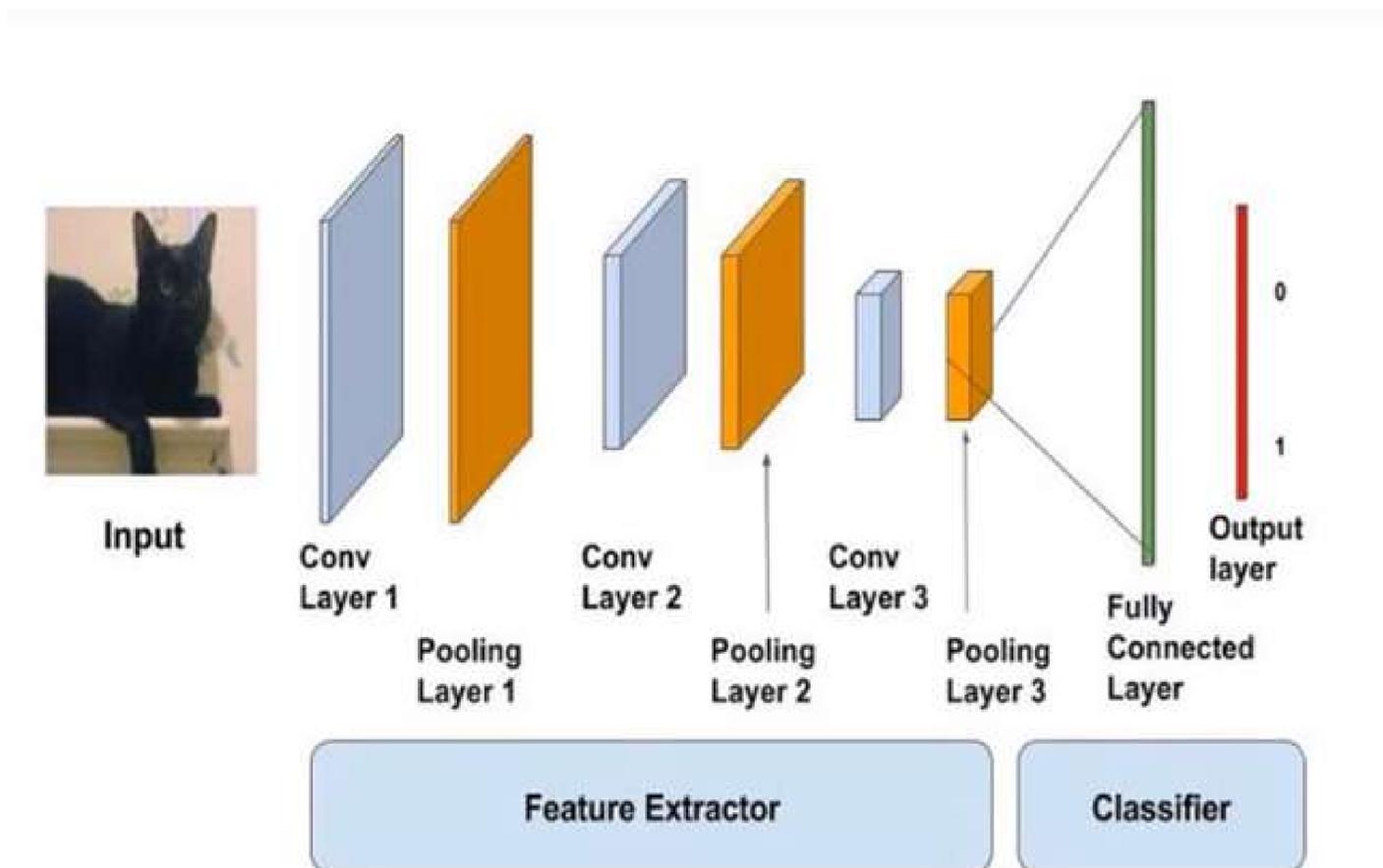


SUMÁRIO  
CASES DO  
PROGRAMA





realizada.



124

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## SICOOB

1

## PREVISÃO DE INADIMPLÊNCIA EM CONTAS DE PESSOA JURÍDICA E RURAL

**Residentes Equipe 1:**

Lucas Coutinho Freitas, Rafael Hideki Kimura.

**Residentes Equipe 2:**

Arthur Silva Lins, Lucas Coutinho Freitas, Matheus Novaes.

**Residentes Equipe 3:**

Igor de Paula Moura, Laryssa S. A. da Costa Silva, Lucas Coutinho Freitas.

**Residentes Equipe 4:**

João Victor Kawazoe Nahra, Lucas Coutinho Freitas.

**Técnicas:**

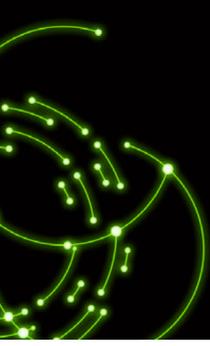
Data Science: Feature Extraction (PCA, t-DSNE), Clustering (K-means Elbow), SHAP (SHapley Additive exPlanations). Machine Learning: Training (Decision Tree, RandomForest, LightGBM, XGBoost, CatBoost), Validation (k-fold Cross-Validation). API: Docker, Flask, Streamlit.

125

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Atualmente, as análises de crédito de contas rurais e pessoas jurídicas são feitas por agentes especialistas, observando métricas tradicionais



e de pouca diversidade de fonte de dados. As análises muitas vezes se restringem ao conhecimento prévio do analista, tornando esse tipo de abordagem pouco escalável e auditável. Com a implantação da cultura data-driven e técnicas de inteligência artificial, as observações históricas de centenas de variáveis geradas pelos clientes diariamente tomam o foco da análise que, de maneira automatizada, extrai e explicam as informações latentes na alta cardinalidade de dimensões desses dados. Desta forma, permite-se uma alta escalabilidade e transparência na forma como são tomadas as decisões.

### **SOLUÇÃO DESENVOLVIDA**

O primeiro passo foi a utilização de técnicas de visualização e extração de características durante a análise exploratória para obter métricas relevantes com o intuito de filtrar as variáveis que teriam o maior impacto no modelo desenvolvido. Em seguida, a próxima etapa foi determinada pela análise da importância das características extraídas em relação à classificação de inadimplências nas contas por

**126**

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

meio de um estudo de correlação, bem como a separação dos dados relevantes em grupos de treinamento e validação do modelo a ser desenvolvido.

Feito isso, foram selecionados alguns modelos de predição baseados em árvores de decisão e seus desempenhos foram comparados com o objetivo de encontrar a melhor abordagem para o problema proposto. Tendo em vista que um erro de previsão de falso-negativo para inadimplência (cliente inadimplente classificado como adimplente) acarreta danos financeiros ao banco Sicoob, o comparativo de desempenho foi ponderado para que houvesse o menor número de ocorrências de falso-negativo. Portanto, como resultado, obteve-se o modelo de Catboost como vencedor, capturando cerca de 80% dos clientes inadimplentes em sua previsão.

Por fim, desenvolveu-se uma aplicação web utilizando os frameworks Flask e Streamlit via Docker para que o projeto pudesse ser executado de maneira isolada, distribuída e descentralizada dentro dos servidores do Sicoob.

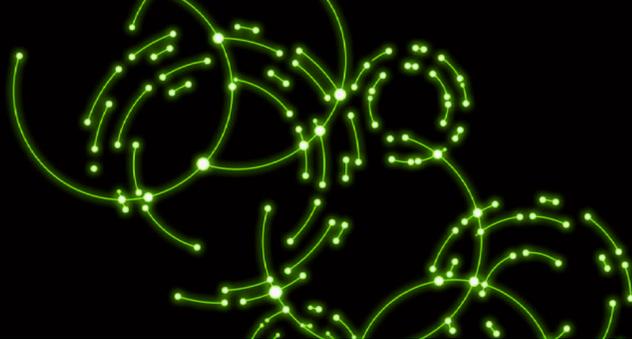
127

SUMÁRIO



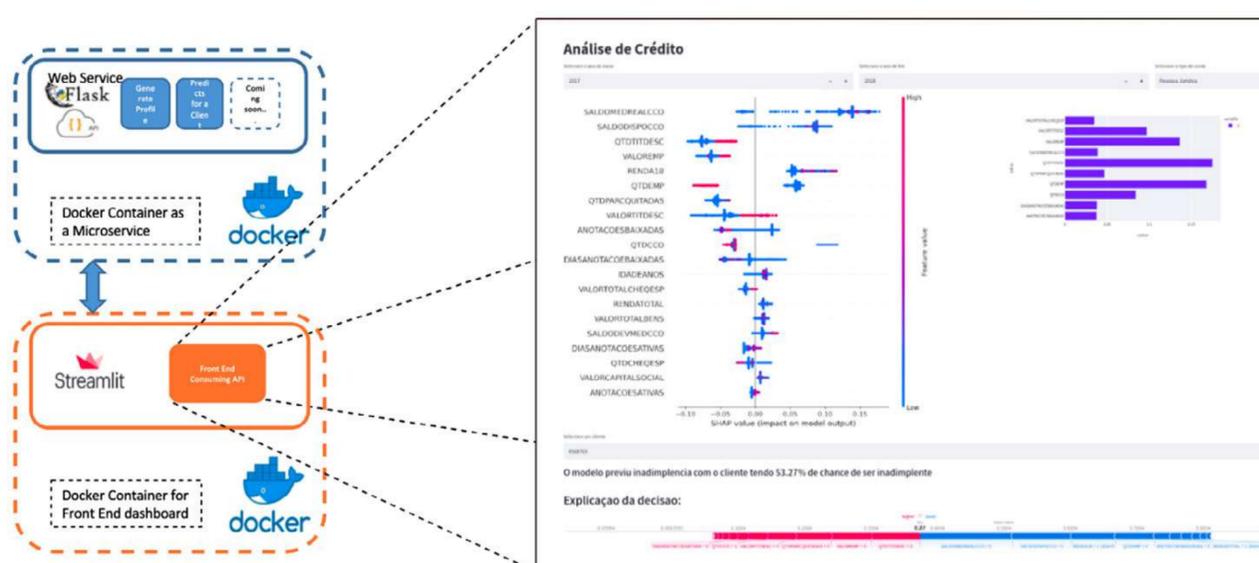
SUMÁRIO  
CASES DO  
PROGRAMA





Por meio de tal aplicação é possível consultar as informações dos clientes contidas na base de dados e utilizar o modelo de previsão já treinado para classificar uma conta como inadimplente de maneira simplificada e intuitiva.

**MELHORES VISUALIZAÇÕES DO TRABALHO:**



SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



Figura 01 - Ilustração da aplicação para tela de análise de clientes.

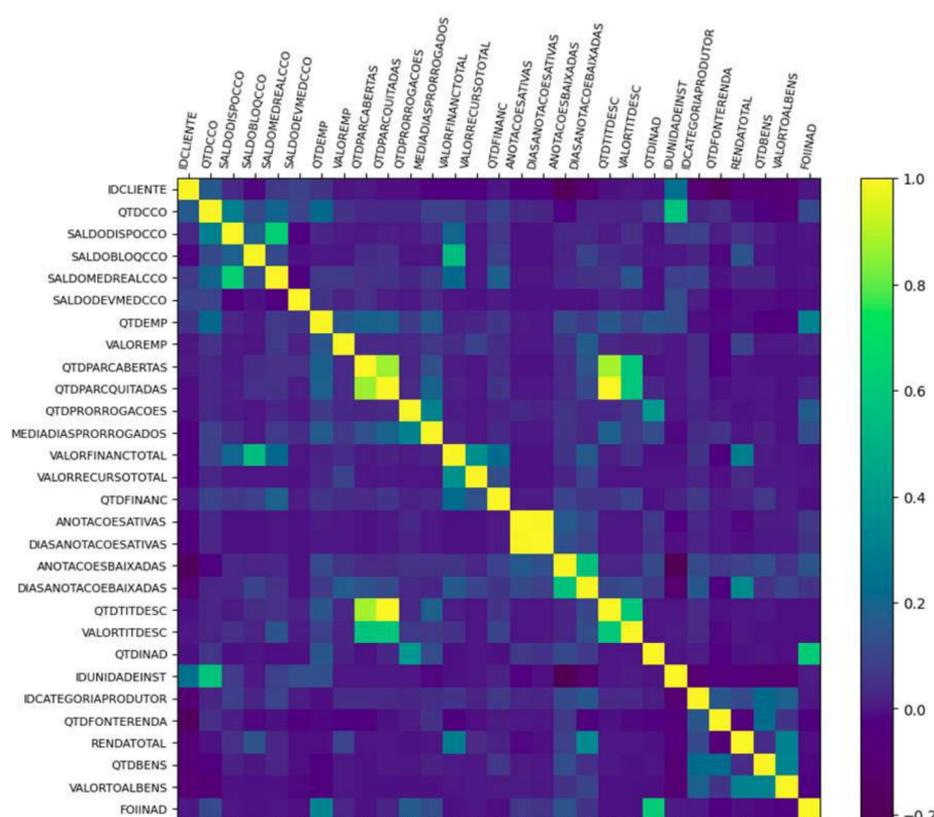
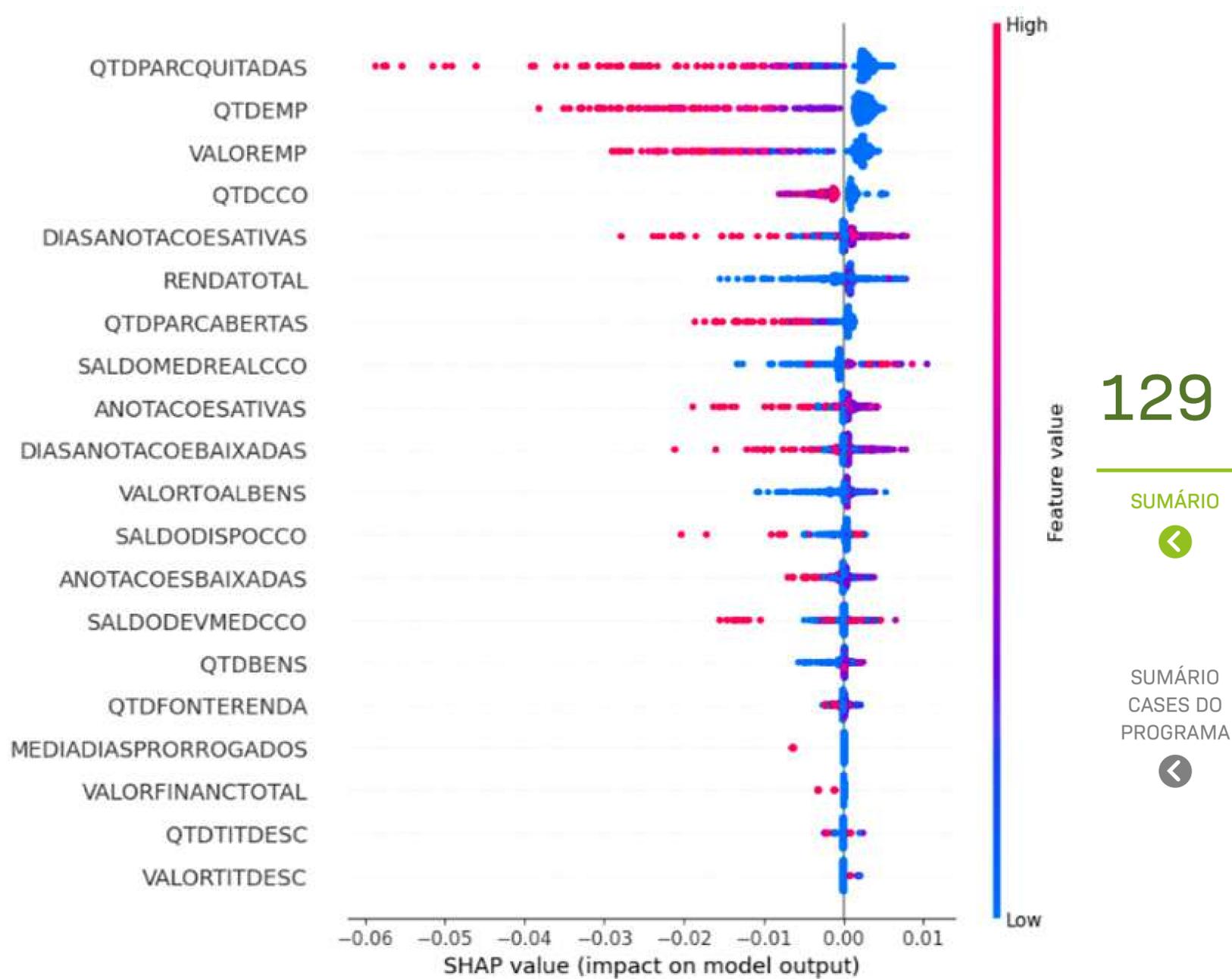
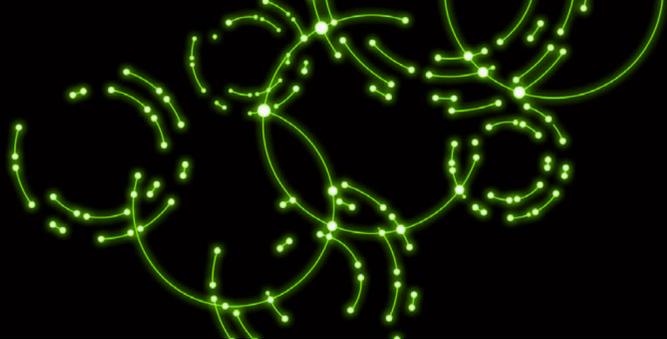


Figura 02 - Matriz de correlação (heatmap) das features de conta Rural.



129

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## TARS

1

VALIDAÇÃO AUTOMÁTICA  
DE DOCUMENTOS ATRAVÉS DE OCR**Residentes:**

Gabriel Jonas Aguiar, Marcia Rodrigues e Sergio Lemos Hermsdorff.

**Técnicas:**

[IA]: Tesseract OCR (Reconhecimento Óptico de Caracteres) incrementada com o Pytesseract e PIL (Python Image Library). [Front-End]: Web App construído com React.js. [Back-End]: Flask RESTful API

130

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

A implementação de novas técnicas e tecnologias nos processos produtivos auxiliam na redução de tempo na execução de tarefas rotineiras e entregas com maior qualidade, uma vez que o trabalho que era exclusivamente de elaboração humana, passa a ser automatizado por uma máquina. Dessa forma, há significativa redução no tempo de trabalho manual, e maior dedicação a atividades que requerem maior intelecto e assertividade.



Neste projeto, objetivou-se o reconhecimento textual de documentos do tipo pdf editáveis, e estabelecer, por evidência objetiva, que os dados são consistentemente compatíveis aos requisitos especificados no modelo de controle disponibilizado pela empresa, com verificação e validação documental automatizadas.

**131**

SUMÁRIO



### **SOLUÇÃO DESENVOLVIDA**

Foi desenvolvido inicialmente o sistema de verificação de dois tipos de documentos, boletos e notas fiscais. Os modelos de padronização foram fornecidos pelo cliente. Essa verificação busca informações mínimas obrigatórias nos campos do documento, e classifica em três grupos, dependendo das informações encontradas: Aprovado, Aprovado com Ressalvas e Reprovado.

Para a interface, foi desenvolvida uma API em React JS, com a biblioteca open-source Tesseract incrementada com o Pytesseract, e com argumentos do tipo PIL (Python Image Library). A interface permite o upload do arquivo a ser verificado e validação, de acordo com

SUMÁRIO  
CASES DO  
PROGRAMA

o tipo: Nota Fiscal ou Boleto. Foi utilizada a biblioteca Flask, em Python, para a verificação dos documentos.

## DocEval

132

The screenshot shows the DocEval web application interface. At the top, there are four PDF document icons with the following titles: 'WT SPEED - NOTA MODELO 21 - ERRO.pdf', 'Nota Fiscal 21 - CONECTA ITAPE - errada.pdf', 'Nota Fiscal 21 - DZ DA ROCHA - errada.pdf', and 'Nota Fiscal 21 - NETMONTES - errada.pdf'. Below the icons, there are radio buttons for 'NF' (selected) and 'Boleto'. A checkbox labeled 'Empresa optante pelo simples nacional.' is checked. At the bottom, there are two buttons: 'VOLTAR' (blue) and 'VALIDAR' (green).

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## TARS

2

## ANÁLISE DE CONTRATOS

**Residentes:**

Pedro Nunes e Roberto Moure.

**Técnicas:**

[IA]: Tesseract OCR (Reconhecimento Óptico de Caracteres) incrementada com o Pytesseract e PIL (Python Image Library). [Front-End]: Web App construído com React.js. [Back-End]: Flask RESTful API e REGEX.

133

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Em continuidade da demanda anterior, este projeto visa a otimização e automação de processos rotineiros, gerando economia no tempo de execução de tarefas. Tendo em vista um aumento considerável na geração de volumes de dados, é importante instaurar sistemas que visem não apenas o armazenamento adequado de dados, como também maior rapidez na maneira que estas informações são classificadas e utilizadas. Para isso, existem processos de automação que reduzem consideravelmente o

tempo de análise e processamento destes dados gerados diariamente em uma empresa.

## SOLUÇÃO DESENVOLVIDA

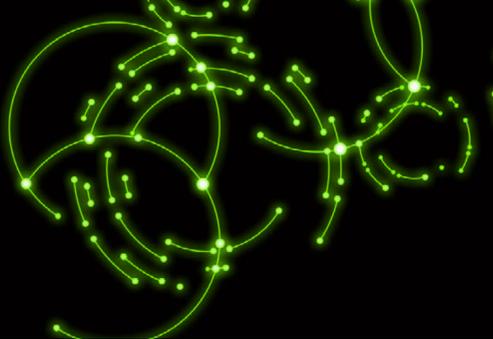
Este projeto é continuidade da demanda anterior, em que foi desenvolvido um sistema de leitura, extração e validação de informações de NF's e Boletos, com o uso da biblioteca open-source Tesseract incrementada com o Pytesseract. A solução proposta nesta etapa faz uma varredura nos termos contidos no documento e transforma em expressões regulares, também denominado REGEX. Com base nestas expressões, é possível definir 'termos-chave' para busca textual, além de permitir a edição e manipulação destes dados. Também é possível classificar se algum dado no documento está incompleto no momento da validação, e indicá-lo.

Para garantir a usabilidade pelos usuários do sistema, foram empregadas técnicas, unificação de funcionalidades e aprimoramento da API, garantindo a disponibilidade das novas funções através do sistema previamente desenvolvido (API). Também foram incrementados os

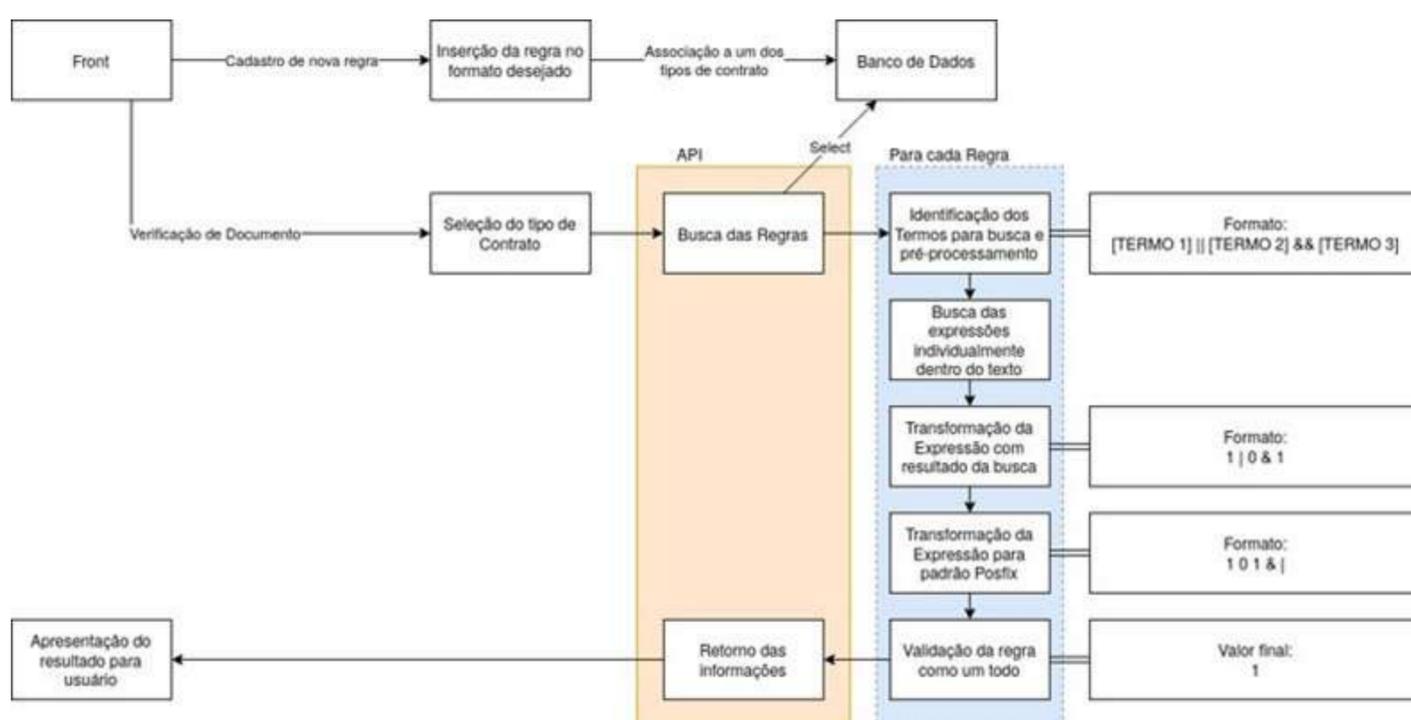
134

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA



documentos que podem ser validados, incluindo os seguintes tipos: contrato de link dedicado para consumo, contrato de link dedicado para revenda e contrato STFC - Telefonia Fixa.



135

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## TARS

3

## DETECÇÃO DE VAZAMENTO DE DADOS EM TEMPO REAL

**Residentes:**

Natalia da Silva de Sá e Sergio Lemos Hermsdorff.

**Técnicas:**

[IA]: Tesseract OCR (Reconhecimento Óptico de Caracteres) e EasyOCR, utilizando Python e bibliotecas PyTorch, MSS e Tkinter. [Back-End]: Biblioteca Pyinstaller para a criação do programa em Ubuntu e Windows.

136

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Em 14 de agosto de 2018 foi sancionada a Lei Geral de Proteção de Dados (LGPD), sob o número 13.709, que entrou em vigor em agosto de 2020. Com isso, muitas empresas tiveram que se adequar e atualizar ou criar uma política de proteção e tratamento dos dados dos clientes (pessoa física). Essa lei possui alguns parâmetros principais, como o consentimento do uso pelo proprietário, armazenamento adequado dos dados, tratamento e sigilo, por

parte da empresa, protegendo informações consideradas confidenciais, como nome, sexo, orientações políticas, etnia, condições de saúde, etc. Tendo em vista a proteção dos dados, este projeto propôs a implementação de medidas de segurança para prevenir fraude e vazamento de dados considerados sensíveis, de forma automatizada, mitigando possíveis multas pelo uso indevido deste conteúdo e garantindo maior segurança aos clientes.

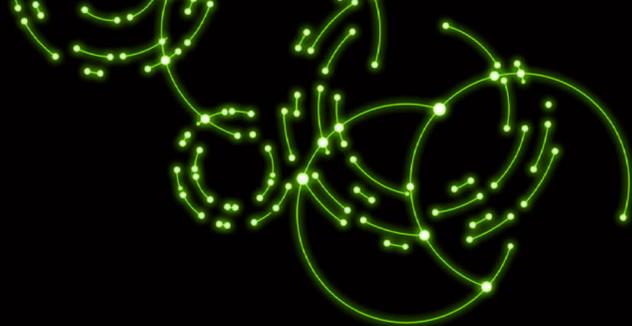
### SOLUÇÃO DESENVOLVIDA

A solução desenvolvida busca evitar o vazamento de dados sigilosos e atender as normativas da empresa e legislação atual, garantindo proteção no uso e distribuição de dados dos clientes. Com o uso da biblioteca MSS, objetivou-se o escaneamento de tela do computador, no momento da utilização do sistema interno da TARS, e a transformação de caracteres em texto através do Reconhecimento Ótico de Caracteres – Optical Character Recognition (OCR). Após a transformação de imagem em texto, o sistema realiza uma varredura textual, buscando palavras-chaves sensíveis, sendo parte delas

137

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA



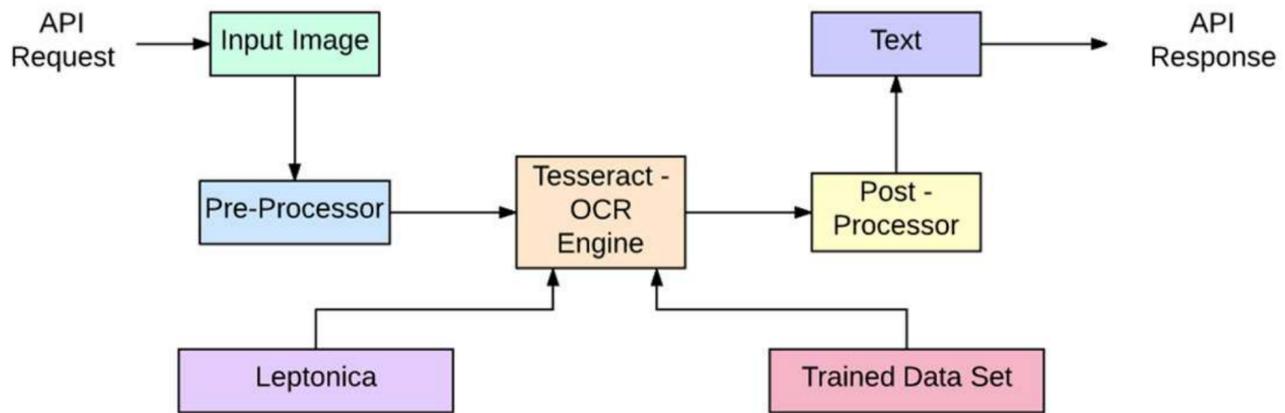
dados como: CPF, e-mail, skype, filiação, data de nascimento. Esta varredura no banco de dados é sincronizada com as preferências de uso de dados dos clientes, ou seja, o sistema verifica quais dados o cliente permitiu armazenamento ou não e compara com os valores presentes no banco de dados.

Com a finalidade de evitar possíveis riscos, como falhas na formatação (texto com letra maiúscula no banco de dados e minúscula na tela, ou formatação distinta entre o banco de dados e tela do computador) foram abrangidas duas modalidades de escaneamento, parcial – parte das informações sigilosas encontradas, ou total – todo conteúdo sigiloso encontrado, de forma idêntica ao banco de dados. O programa é instalado diretamente no sistema operativo, sendo eles Ubuntu ou Windows, e gera alertas no caso de alteração, cópia vazamentos de informações em diferentes meios como e-mail, documentos de texto, chamada de vídeo etc.





### OCR Process Flow



## TARS

4

DETECÇÃO DE VAZAMENTO DE INFORMAÇÕES  
CONFIDENCIAIS**Residentes:**

Germano Yoneda Pereira Lima e Larissa Luciane Ferreira.

**Técnicas:**

[IA]: Tesseract OCR (Reconhecimento Óptico de Caracteres), biblioteca difflib e flashtext.

[Back-End]: Nvidia-Docker para arquivo de configuração de servidor, Python – com biblioteca Pyinstaller para criação de arquivo executável/instalável.

140

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Além de brindar maior transparência com os consumidores e clientes, possuir sistemas que viabilizem a proteção de dados sigilosos é de suma importância para as empresas, uma vez que o não cumprimento de ações de segurança podem inclusive gerar prejuízos financeiros, como multas governamentais e processos por parte dos clientes. Outro prejuízo gerado pelo vazamento de dados é a perda de credibilidade, uma vez que a empresa que teve o sistema

invadido, deve, muitas vezes, publicar em meios de comunicação sobre o incidente. Para acionar as autoridades nacionais no caso de vazamento de dados, é necessário possuir uma política de uso e armazenamento de dados instaurados numa organização, devido à documentação que deverá ser apresentada nestes casos. Tendo em vista a importância na proteção de informações sigilosas, a TARS solicitou um sistema que auxiliasse na detecção de vazamento de dados.

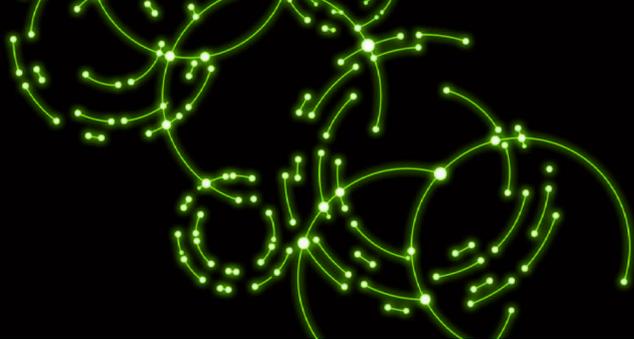
### SOLUÇÃO DESENVOLVIDA

Em continuidade da demanda anterior, a TARS solicitou um sistema para mitigar o vazamento de informações confidenciais dos clientes, atendendo as legislações em vigor. Essa etapa do projeto visou a melhoria do sistema entregue, com foco na otimização da arquitetura cliente-servidor e redução de riscos para casos de falha na identificação das informações a serem capturadas em tela, e erros de identificação se há ou não um vazamento de dados. Após testes em algumas bibliotecas para verificar rapidez, otimização na transferência de informações e maior rapidez no processamento, foi definido o

141

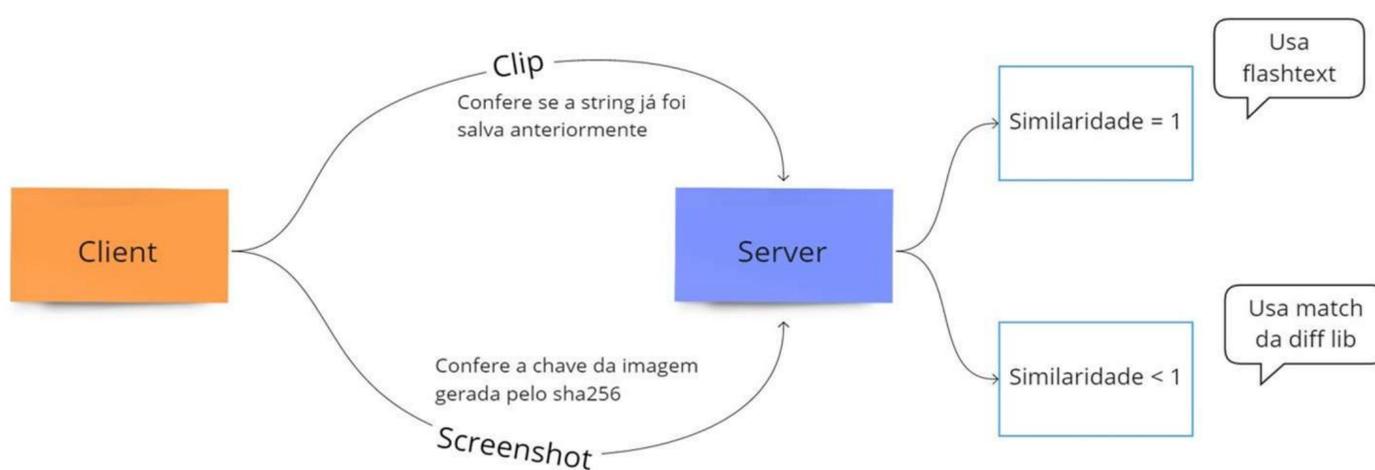
SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA



uso da biblioteca Tesseract OCR, que também permitia configurações que facilitaram o reconhecimento de caracteres especiais (ex: @, #, \$, %).

Também foram utilizadas as bibliotecas difflib e flashtext, para reconhecimento de termos parcialmente ou exatos, respectivamente. É possível que o usuário selecione o grau de similaridade dos termos buscados, para abranger mais dados ou facilitar a filtragem. Em caso de verificação de possíveis vazamentos de informações, o sistema gera um arquivo em formato CSV com informações sobre o acesso dos dados, como por exemplo o IP do cliente, hora de acesso e outros dados de log.



# TÉRMICA SOLUTIONS

1

## SIMULAÇÃO E OTIMIZAÇÃO DE OPERAÇÕES EM TERMOELÉTRICA

### Residentes Equipe 1:

Melanie Gissel Urdangarin Gamarra, Natalia Silva de Sá e Victor Hugo Nagahama.

### Residentes Equipe 2:

Karen Taniguchi, Larissa Luciane Ferreira e Melanie Gissel Urdangarin Gamarra.

### Residentes Equipe 3:

Gabriel Gomes de Souza, Larissa Luciane Ferreira e Pedro Humberto Nunes.

### Residentes Equipe 4:

Gabriel Jonas Aguiar, Marcia Regina Ramalho Rodrigues e Roberto Schnitzler Moure.

### Técnicas:

[DS] Tratamento de dados, visualização e exploração de dados, análise exploratória.

[Estatística] Métodos de regressão para predição, teste e acurácia de modelos preditivos.

[IA] Machine learning, redes neurais, deep Learning.

143

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

## OPORTUNIDADE DE NEGÓCIO

Durante o processo siderúrgico, quatro

subprodutos são produzidos: gás de coqueria (coke oven gas - COG), gás de aciaria (Linz-Donawitz gas - LDG), gás de alto forno (blast furnace gas - BFG) e alcatrão (TAR). Esses subprodutos apresentam elevada capacidade de produção de energia elétrica, quando direcionados a centrais termoelétricas. A central termoelétrica é uma instalação industrial que transforma a energia térmica, obtida a partir da queima de combustíveis nas caldeiras, em energia elétrica. A queima desses subprodutos na presença de ar disponibiliza o conteúdo energético do combustível e gera gases de exaustão, que em certos momentos apresentam coloração visível (fumaça escura). A determinação da combinação de gases de exaustão que geram as emissões visíveis e os momentos em que são produzidas na termoelétrica é de grande importância para a empresa contratante, no intuito de posteriormente conseguir determinar as causas do fenômeno e reduzi-lo, contribuindo para a diminuição dos impactos ambientais.

## **SOLUÇÃO DESENVOLVIDA**

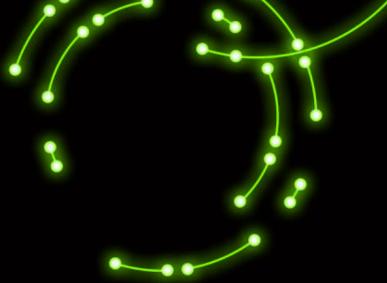
144

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





A demanda inicial do projeto consistia em, a partir da base de dados da caldeira 2 (CTE2), repassada pela empresa requerente, identificar comportamentos atípicos de gases emitidos durante processos siderúrgicos. Foi realizado o tratamento dos dados, eliminando outliers, e em sequência, foi feita a exploração e a visualização dos dados. Para a melhor visualização/ investigação dos resultados da emissão dos gases em estudo, foram produzidos alguns dashboards. Em seguida, foi realizada uma análise estatística, indicando comportamentos incomuns para determinados gases, conduzindo a algumas hipóteses para este fenômeno. Foram identificados padrões de emissões visíveis e falsos negativos e, foram levantadas as possíveis causas de emissões visíveis.

Dando continuidade à demanda, foi utilizado um tipo de modelagem, denominado por regressão. A regressão, além de auxiliar no entendimento de como determinadas variáveis podem influenciar no comportamento de outras variáveis (associadas ao final do processo siderúrgico, no caso desta demanda), pode ajudar na

145

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA



construção de modelos de predição. Os modelos de regressão utilizados foram o Random Forest e o Extra Trees, visando inicialmente classificar as emissões e posteriormente explorar o impacto causado nas emissões quando ocorrer a mudança nas variáveis de entrada. Foi proposta uma POC (prova de conceito) de modelagem de dados de forma a predizer as quantidades produzidas dos poluentes, óxido de nitrogênio (NOx), óxido de enxofre (SOx), monóxido de carbono (CO) e material particulado (MP), assim como, predizer a quantidade de saída de oxigênio da caldeira. Testes e avaliação da acurácia dos novos modelos preditivos também foram realizados.

Dando sequência à demanda, com a adição de novas variáveis de entrada, foi aplicado novamente o modelo de regressão, que desta vez, retorna como saída os parâmetros ambientais, maximizando a potência e a eficiência. Há uma boa aproximação entre os dados obtidos a partir do modelo preditivo com os dados reais (medidos pelos sensores), como mostra a figura a seguir.

146

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



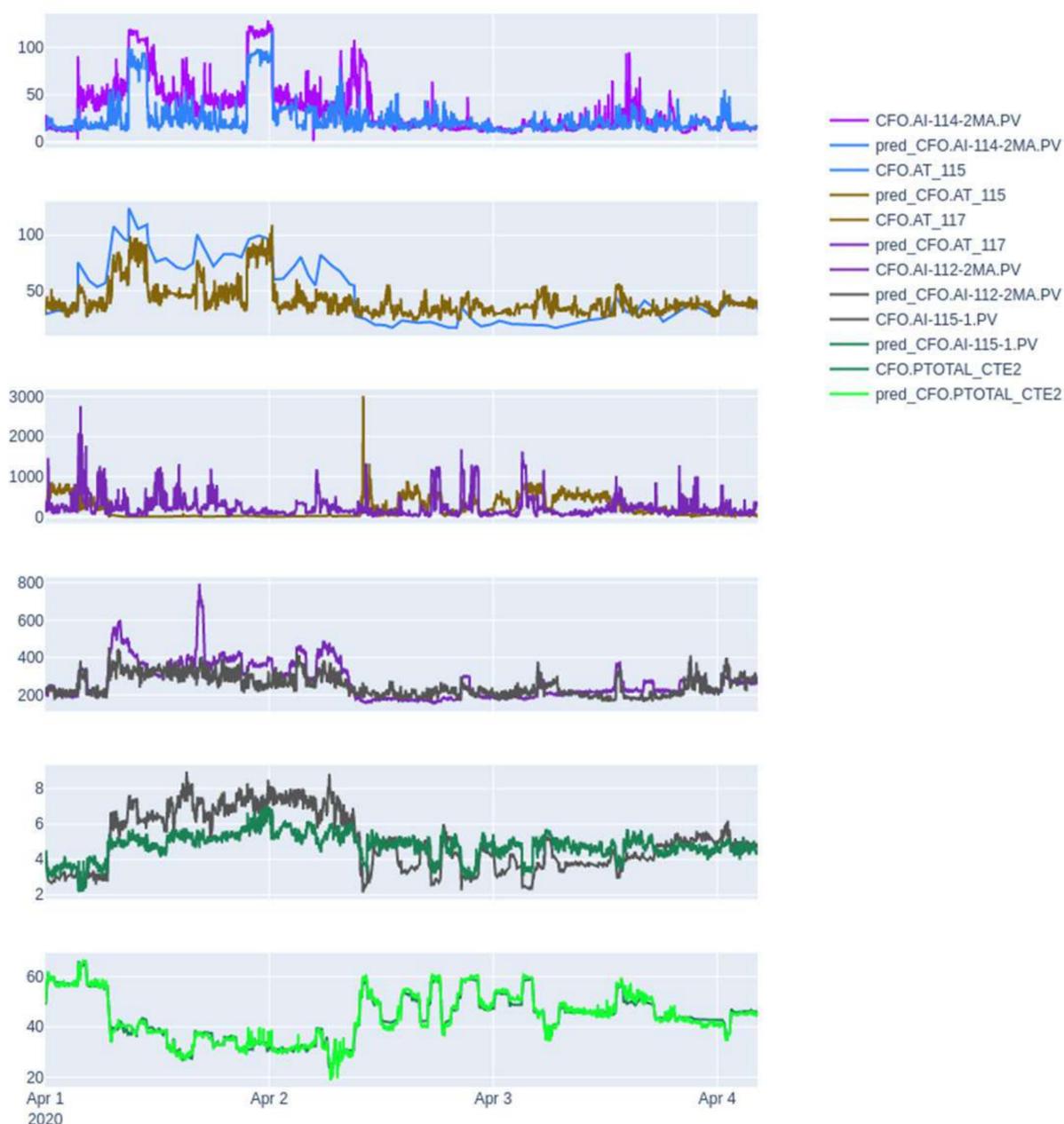
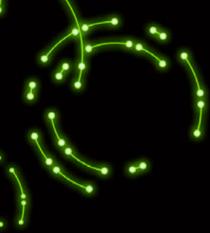


Figura 1: Comparação entre os dados - modelo de regressão X reais.

Posteriormente, foi desenvolvida uma ferramenta em que é possível manipular os parâmetros de entrada, propiciando a otimização do comportamento da termoelétrica, com a utilização do chamado algoritmo genético (AG). Ao longo desta etapa da demanda, foi proposto a implementação de tipos distintos de processos de otimização:

- Um em que é maximizada a potência;





- Um em que é minimizado o material particulado (se estiver acima de 60);
- Um em que é maximizada a eficiência;
- Um em que, é minimizado o material particulado (se estiver acima de 40), maximizada a eficiência e a potência é pré-definida em torno de um certo valor.

Ao final da demanda, foram reproduzidas as análises realizadas ao longo das sprints 1, 2 e 3, para um novo conjunto de dados, da caldeira 1 (CTE 1). Os resultados obtidos pelo modelo preditivo para a caldeira 1 se ajustam bem os dados reais (obtidos pelos sensores), validando os modelos aplicados à caldeira 2, ao longo do projeto.



# TMG - TROPICAL MELHORAMENTOS E GENÉTICA

1

## AQUISIÇÃO DE INFORMAÇÕES SOBRE NOVOS CULTIVARES

### Residentes:

Ananda Coelho, Giovani Sylvestrin, Igor Nieto.

### Técnicas:

[DS]: Análise exploratória dos dados. [IA]: Web Crawler, Processamento de Linguagem Natural (NLP), Modelo BERT. [Main-Frame]: Power BI

149

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

### OPORTUNIDADE DE NEGÓCIO

Por ano, chegam a ser lançadas até 400 novas cultivares de soja, sobre os quais a empresa busca se manter atualizada. O que requer a coleta de 42 características de cada uma dessas cultivares. Nesse contexto, a automação da coleta de dados traria à empresa uma vantagem competitiva no mercado.

### SOLUÇÃO DESENVOLVIDA

O problema foi abordado de duas maneiras. Uma abordagem clássica de scraping, em que é programado o caminho exato dentro de cada site

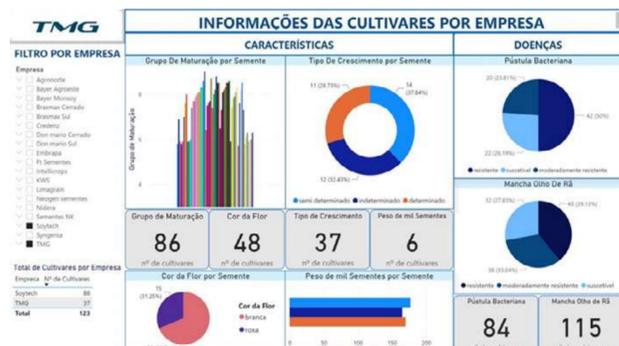
de onde as informações serão retiradas – o que confere a máxima precisão para obtenção do dado, às custas de flexibilidade e abrangência, no sentido de que só funciona no site para o qual foi programado e para de funcionar caso algum caminho seja modificado. E uma abordagem de processamento de linguagem natural (NLP), com um método de Machine Learning chamado BERT Q&A. Que ao contrário da primeira abordagem, não possui 100% de precisão, porém uma vez treinado o modelo ele é altamente escalável pode abranger centenas a milhares de sites. A solução final foi a criação de uma web crawler para navegar em sites de interesse e recolher as informações relevantes de lançamentos de variedades de soja e apresentá-las em um dashboard.

150

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



# TMG - TROPICAL MELHORAMENTOS E GENÉTICA

2

## AVALIAÇÃO DA INFLUÊNCIA DE FATORES CLIMÁTICOS E DO SOLO NA PRODUÇÃO DE SOJA

### Residentes:

Igor Nieto, Laryssa Stephanie.

### Técnicas:

[DS]: Utilização de APIs de solo e clima para mineração de dados e plotagem de mapas com os resultados. [IA]: Krigagem e Inverse Distance Weighting (IDW). [Estatística]: Interpolação. [Main-Frame]: Jupyter.

151

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

### OPORTUNIDADE DE NEGÓCIO

Para garantir uma boa produtividade e rentabilidade com a produção de soja, se faz necessária a combinação de diferentes procedimentos, observância do clima e do solo. Logo, é imprescindível conhecer tanto o mapeamento do solo, quanto acompanhar as previsões meteorológicas e fenologia da soja, visto que estas condições são insumos determinantes para o sucesso de atividades agrícolas. Portanto, a empresa busca compreender melhor a influência de fatores

que possam afetar o desempenho produtivo das cultivares de acordo com o clima ou tipo de solo da região em que será feito o plantio, para que assim, possa focar em cultivares mais competitivas no mercado.

### **SOLUÇÃO DESENVOLVIDA**

Para obter informações do solo, foi utilizado como fonte um programa brasileiro chamado PronaSolos - Programa Nacional de Solos do Brasil. Por ser um programa nacional ele possui todas as informações do solo brasileiro. Quanto aos dados meteorológicos, diversas fontes foram consultadas, porém a solução mais viável encontrada foi o Instituto Nacional de Meteorologia (INMET), que diferente de outras APIs, disponibiliza gratuitamente seus dados e possui um maior período de dados históricos para download em arquivos CSV.

Foi realizada a análise exploratória dos dados de clima e solo encontrados. A partir das informações coletadas e da aplicação de premissas geoestatísticas, foi avaliada tanto a suficiência dos dados quanto à adequação

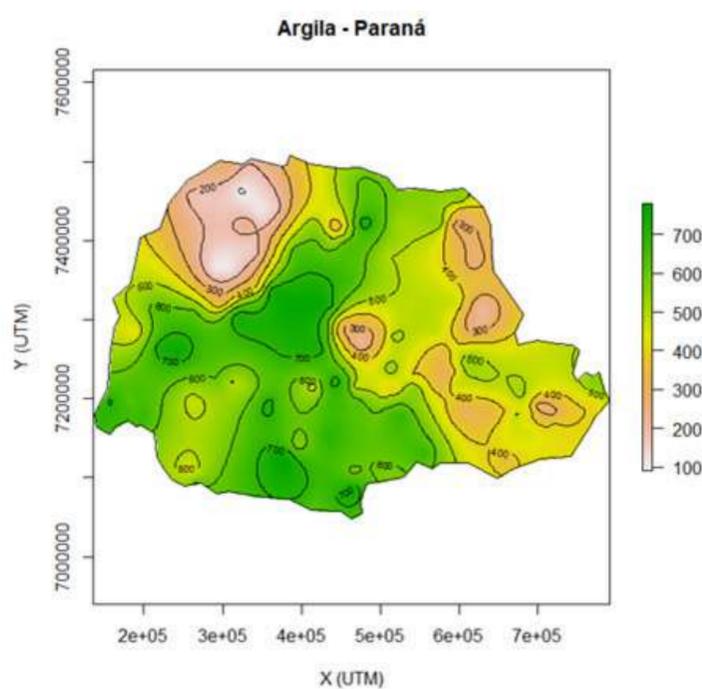
**152**

SUMÁRIO

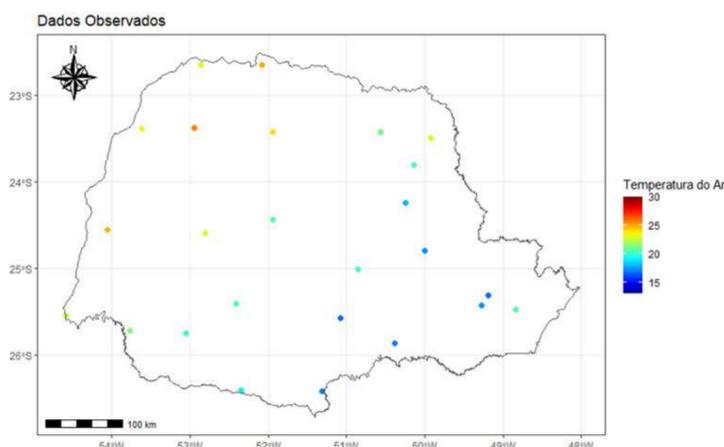
SUMÁRIO  
CASES DO  
PROGRAMA



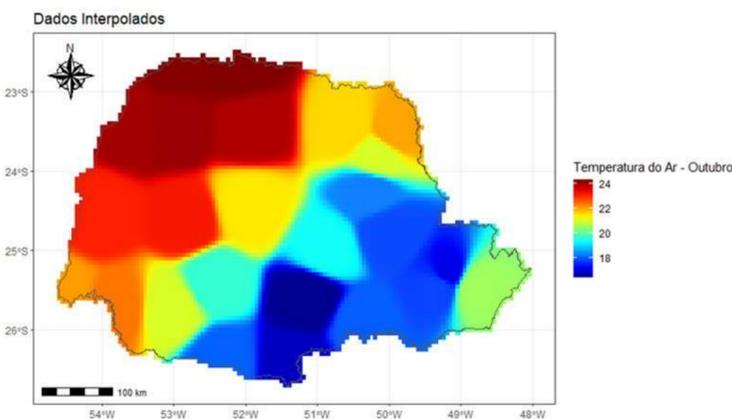
de distintos modelos de interpolação. Modelos estes que foram testados e produziram mapeamentos da distribuição espacial das características de interesse para o cultivo da soja. O resultado deste trabalho foi entregue em forma de POC.



Plotando dados observados



Plotando dados interpolados



# TMG - TROPICAL MELHORAMENTOS E GENÉTICA

3

## CATEGORIZAÇÃO DE CHAMADOS

**Residentes:**

Anna Luiza Gomes, Matheus Novaes, Igor Nieto.

**Técnicas:**

[DS]. [IA]: Utilização de técnicas de processamento de linguagem natural (NLP) e algoritmos de agrupamento de textos.

[Estatística]: Interpolação. [Main-Frame]: Jupyter.

154

SUMÁRIO

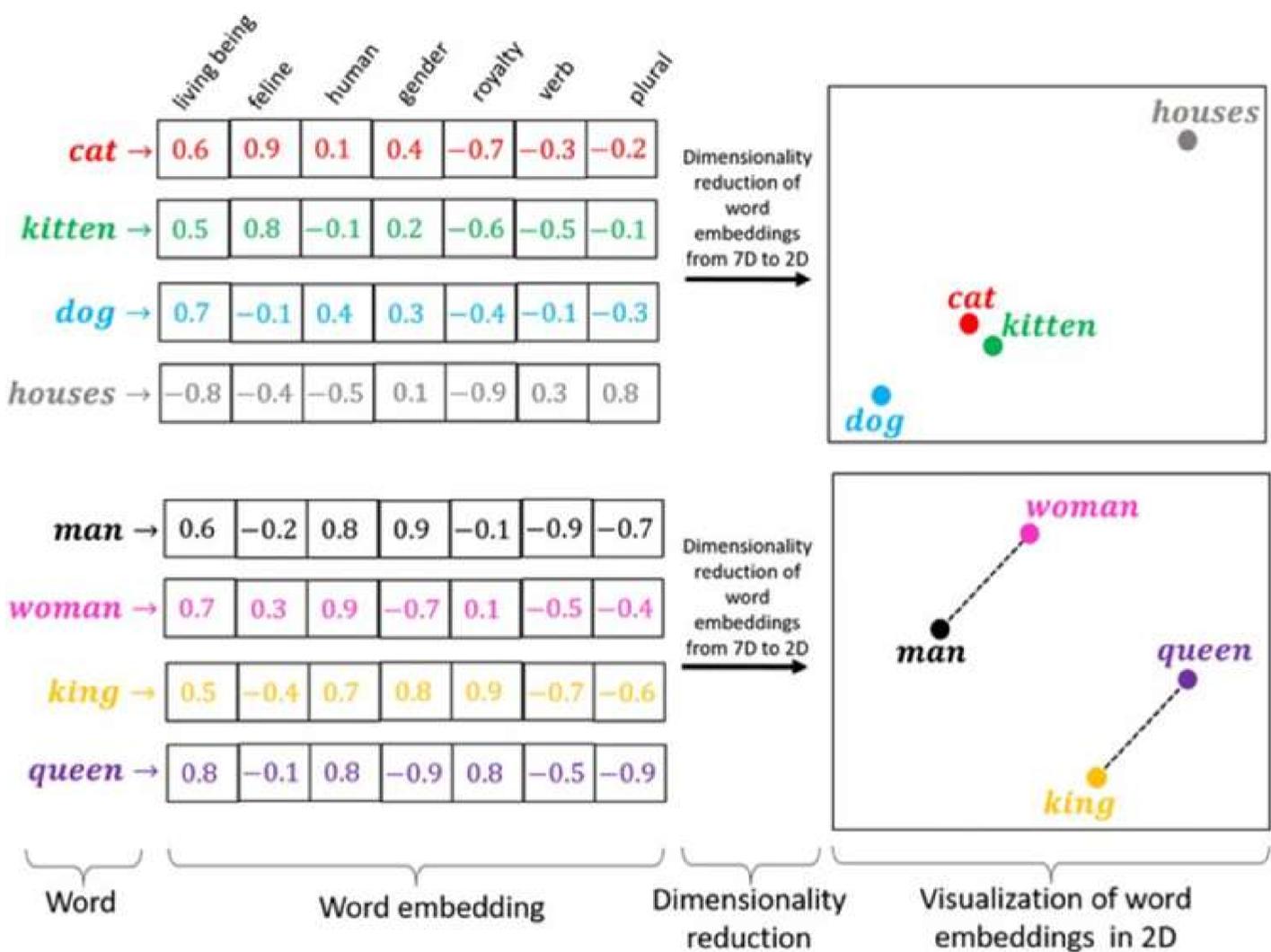
SUMÁRIO  
CASES DO  
PROGRAMA

### OPORTUNIDADE DE NEGÓCIO

A empresa visa realizar um redirecionamento mais rápido de seus clientes que entram em contato via help desk. Help desk se refere a um serviço de atendimento aos clientes que procuram por soluções, esclarecimentos sobre dúvidas e outras solicitações para problemas técnicos. Além disso, um serviço de help desk centraliza o recebimento das solicitações dos clientes, o que facilita as respostas e resolução dos problemas.

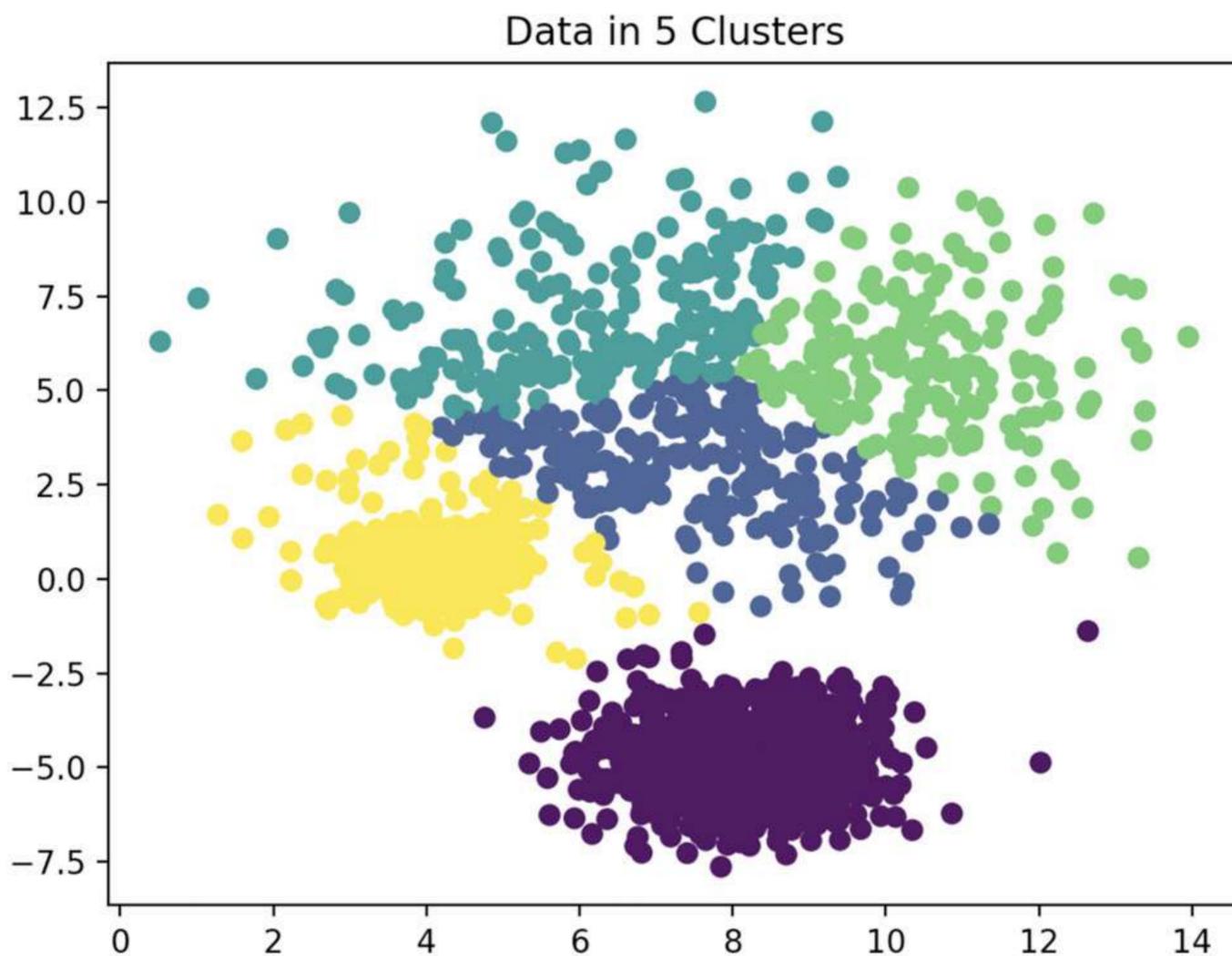
**SOLUÇÃO DESENVOLVIDA**

Foi realizada a análise exploratória dos dados de chamados dos clientes e suas possíveis categorias, baseadas em resolução do problema. Após tratados os dados, foram utilizados algoritmos para atribuir valores às palavras dos chamados, segundo múltiplas dimensões de classificação, baseado em seu significado.



Em seguida foram utilizadas técnicas de clusterização de textos para realizar o agrupamento de chamados em categorias. E foi desenvolvida uma interface que permite ao

usuário inserir um chamado que deseje consultar as possíveis categorias nas quais se pode enquadrá-lo.



156

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## TRIMBLE

1

ANÁLISE DE EVENTOS BRUSCOS DE ACORDO  
COM A JORNADA DE MOTORISTAS**Residentes:**

Germano Yoneda Pereira Lima, Larissa Luciane Ferreira e Pedro Humberto Augusto Paz Teixeira Nunes.

**Técnicas:**

[DS] Análise exploratória referente à ocorrência de eventos, visualização de dados. [Estatística]: Interpolação. [Main-Frame]: Jupyter.

157

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Lesões e mortes por acidentes nas rodovias têm se tornado uma epidemia em todo o mundo. Segundo a Polícia Rodoviária Federal (PRF) apenas no Brasil em 2014, houve 168.593 acidentes nas rodovias, totalizando 100.396 feridos e 8.227 mortos. A falta de atenção e o sono ao volante causaram 32,3% e 6% dos acidentes fatais, respectivamente. Em um estudo realizado pelo Ministério dos Transportes, Portos e Aviação, foi constatado que em 53,7% dos acidentes a causa é a imprudência dos



motoristas. Diante desses dados alarmantes, torna-se importante conhecer e entender as causas ou os fatores de risco relacionados a esses acidentes.

Esse projeto visa explorar um possível indicador de imprudência no trânsito, que é a ocorrência de eventos considerados bruscos, tais como: Frenagem Brusca, Curva Brusca, Aceleração Brusca, Possível Colisão, Fadiga, Distração e Excesso de velocidade. A abordagem do projeto visa expor a distribuição da ocorrência desses eventos durante a jornada de trabalho do motorista, identificando dessa maneira, os picos de ocorrências de eventos considerados bruscos. Também visa avaliar características associadas a esses eventos.

### **SOLUÇÃO DESENVOLVIDA**

Inicialmente foram realizadas análises exploratórias para identificar os dados que seriam utilizados no projeto. Depois disso, foi calculada a porcentagem decorrida da jornada dos trabalhadores até a ocorrência dos eventos de interesse (bruscos). Uma outra

**158**

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





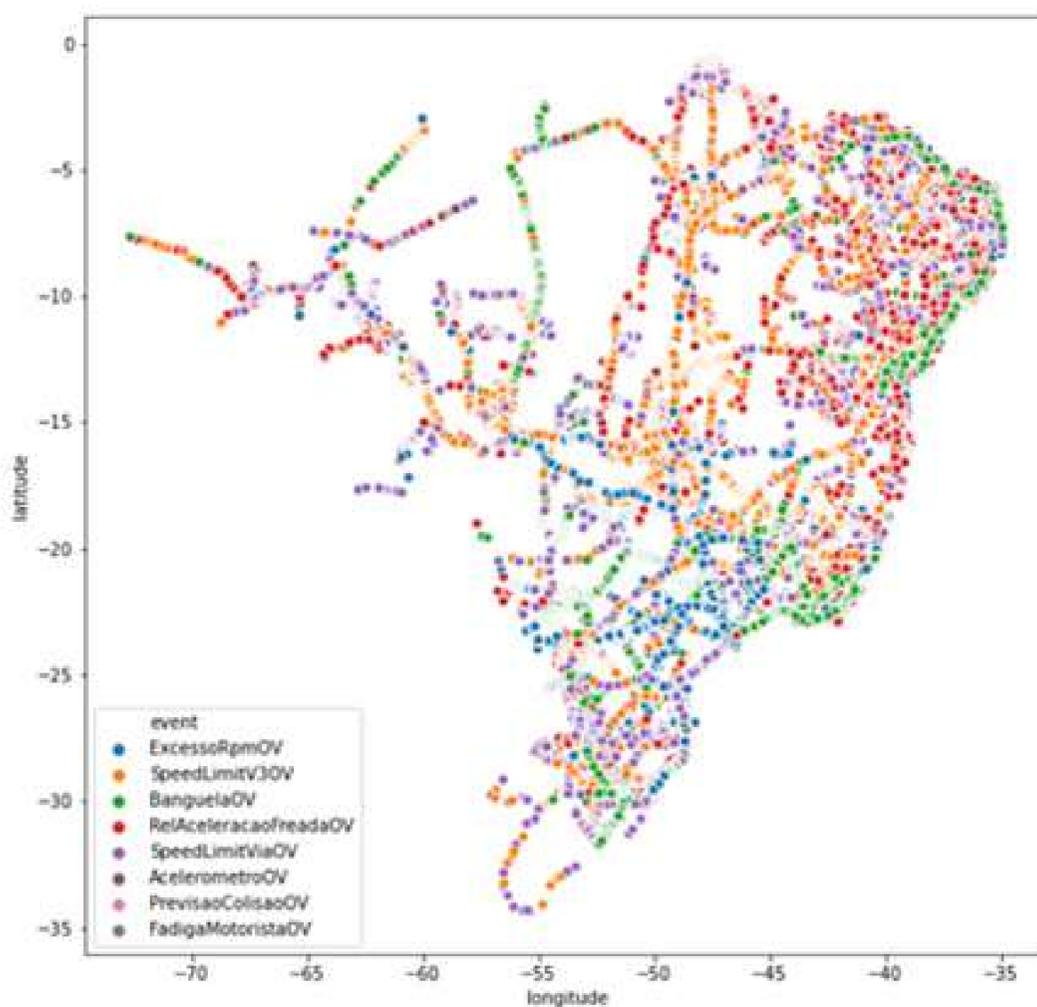
exploração realizada foi a de mapas de calor da porcentagem da jornada do motorista até a ocorrência do evento e, da quantidade de ocorrência de eventos por período do dia (manhã, tarde e noite). Análises da quantidade de eventos de interesse por estado também foram feitas. Outra análise realizada foi a quantidade de eventos de interesse ocorridos no meio urbano e não urbano. Foram analisadas a maior parte dos eventos separadamente, conduzindo a gráficos da distribuição de cada evento pelo Brasil. Por fim, foram analisadas a quantidade de eventos por ano e marca do veículo.

159

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## TRIMBLE

2

## IDENTIFICAÇÃO DE MOTORISTAS SEM VIGÊNCIA

**Residentes:**

Gabriel Jonas Aguiar e Natalia da Silva de Sá.

**Técnicas:**

[IA] Machine Learning.

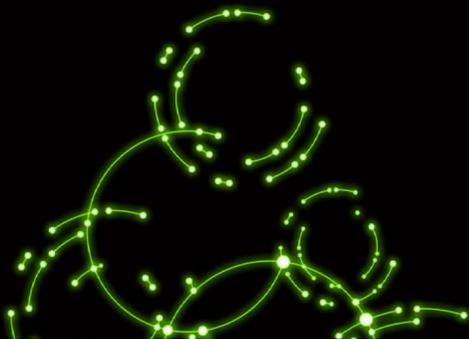
160

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Identificar os motoristas durante uma vigência é essencial, principalmente pelo fato da empresa requerente poder avaliar possíveis riscos e identificar infrações cometidas pelo motorista devido a sua conduta. Ao iniciar uma vigência, existem motoristas que não utilizam os meios primários de identificação, como por exemplo a biometria. Portanto, se faz necessário a identificação do motorista de uma outra forma. Modelos de machine learning podem ser empregados para modelar o comportamento de motoristas e consequentemente identificá-los.

**SOLUÇÃO DESENVOLVIDA**



Foi recebida uma base de dados com as vigências, features e identificação do motorista. Inicialmente foi realizado um tratamento da base de dados. Em seguida foram empregados modelos de machine learning, tanto de classificação não-supervisionada como de classificação supervisionada. Para a classificação supervisionada foram empregados os algoritmos Decision tree, Random forest e XGboost. Ao utilizar esses algoritmos foi possível encontrar alguns padrões de direção dos motoristas, mas não foi possível identificar o motorista. Para o método de classificação não-supervisionada foram aplicados os algoritmos DBscan e KMeans. Neste caso, foi possível encontrar padrões de conduta dos motoristas e conseqüentemente identificar os motoristas.

161

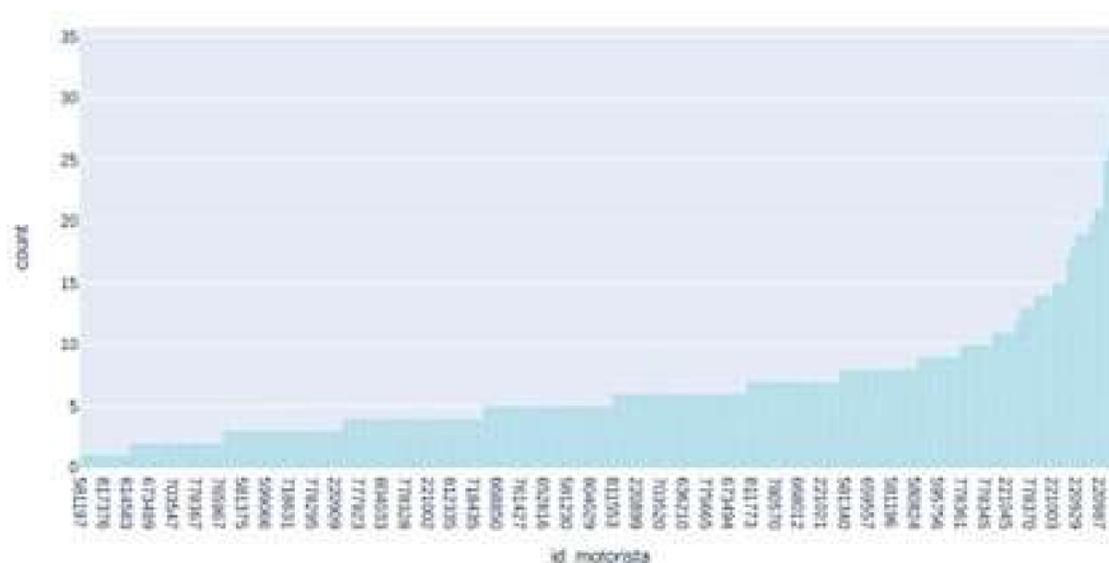
SUMÁRIO

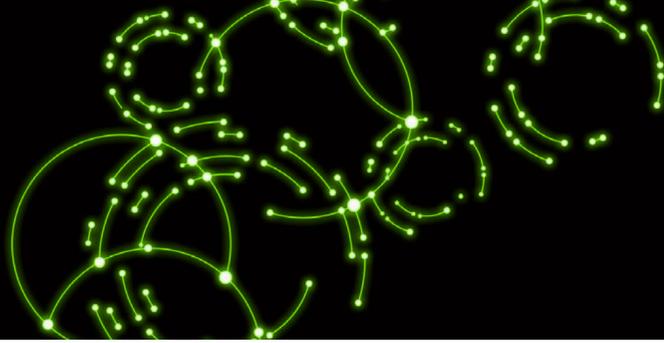


SUMÁRIO  
CASES DO  
PROGRAMA



Número de vigências por id motorista





## TRIMBLE

3

ANÁLISE DE CONSUMO DE COMBUSTÍVEIS  
(VELTEC)**Residentes:**

Melanie Gissel Urdangarin Gamarra, Sergio Lemos Hermsdorff.

**Técnicas:**

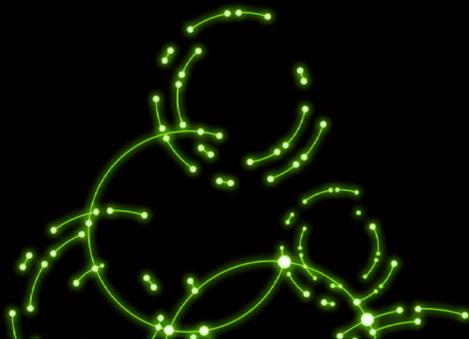
f\_regression, RandomForest, XGBRegressor, SVR, RandomForestRegressor.

162

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

O consumo de combustíveis é uma das variáveis que podem ser altamente correlacionadas com custos em uma empresa de transporte, além de agravar questões ambientais. Logo, o aumento na eficiência do combustível tende a ser financeiramente benéfico para a empresa e, ao mesmo tempo, menos nocivo ao meio ambiente. Mesmo que essa eficiência esteja inicialmente relacionada ao como o combustível é produzido (isto é, à qualidade do mesmo) e à eficiência dos veículos individualmente, em largas escalas da utilização do produto existem fatores



emergentes que também podem contribuir. Fatores estes que podem ser identificados por uma inteligência artificial, com o objetivo de facilitar tomadas de decisões, uma vez que as causas do consumo elevado de combustível estarão mais evidentes.

### **SOLUÇÃO DESENVOLVIDA**

Uma abordagem possível para redução de custos neste contexto é estabelecer um padrão de comportamento de grupo e checar a existência de casos pontuais destoantes que, em geral, são os que podem estar gerando custos extras. Para esta finalidade, a variável relevante a ser estudada é a quilometragem por litro de combustível. A partir dos dados providos pela empresa, calculou-se o consumo de combustível padrão. Com isso, foi possível estabelecer um parâmetro que mede a diferença do consumo em uma viagem com a média e, com isso, classificar dados considerados destoantes. Destes dados, foram usados cinco métodos distintos para estabelecer quais variáveis são mais correlacionadas com o consumo elevado de combustível para que, assim, possamos



compará-las com os dados do grupo padrão.

Essa comparação é feita na figura 1, onde é possível ver que, em média, os motoristas/veículos do grupo padrão ficam parados em média 7% da viagem, enquanto os destoantes ficam parados 63%. Na figura 2, podemos ver uma prévia da interface do software desenvolvido.

164

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



Número de variáveis para comparação

20

Ver resultado do feature importance

	Mediana Padrão	Mediana Destoam
tempo_estado_movimento %	0.9327	0.3630
TEMPO DE VIAGEM	6,495.5000	653
tempo_rpm_marcha_lenta	0.0552	0.4919
tempo_estado_parado %	0.0671	0.6350
tempo_rpm_verde_economica	0.0823	0.2299
tempo_rpm_abaixo_verde	0.0214	0.1445
tempo_rpm_verde_potencia	0.6706	0.1351
veiculo_ano	2018	2017
tempo_rpm_amarelo	0.0313	0
tempo_motor_ocioso %	0	0
tempo_banguela %	0	0

Ver gráficos boxplot das variáveis

Figura 1

**veltec**  
A TRIMBLE COMPANY

**Análise de Consumo de Combustíveis**

Selecione qual algoritmo será utilizado

Abordagem I

Abordagem II

Insira o caminho da pasta onde se encontra o arquivo base: 'vigencias\_viagem.csv'

data\_processada/vigencias\_viagem.cs

Figura 2

## TRIMBLE

4

DETECÇÃO DE SONOLÊNCIA AO VOLANTE  
(VELTEC)**Residentes:**

Tiago Pedrosa Guilherme, Gabriel Gomes Sousa,  
Roberto Schnitzler Moure.

**Técnicas:**

RandomForest, LSTM.

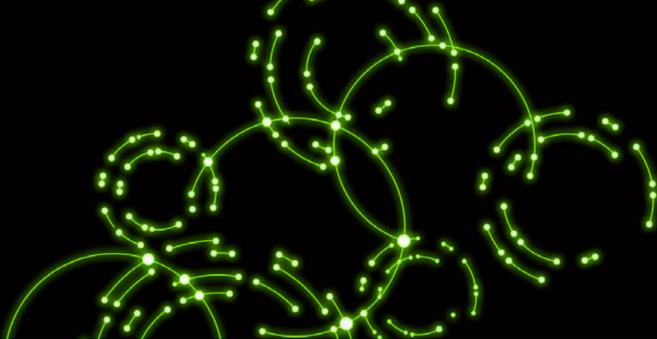
165

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

A exaustão por sono ao volante é tão perigosa quanto o consumo de álcool. No entanto, sua detecção é mais difícil pois os sinais que o indivíduo apresenta são mais sutis do que observado em pessoas embriagadas. Apesar disso é possível utilizar imagens de câmeras de bordo para treinar uma inteligência artificial que possa detectar esses sinais sutis e acionar algum tipo de alerta aos motoristas, evitando assim, possíveis acidentes.

**SOLUÇÃO DESENVOLVIDA**



Inicialmente, utilizou-se da biblioteca Holistic para identificar e extrair landmarks de rostos e de expressões corporais. Assim, é possível criar um dataset com os landmarks relacionados às classes: atento, desatento e sonolento.

Uma vez definida as classes, foi utilizado um classificador RandomForest e uma rede neural LSTM para comparar o desempenho e o modelo classificador.

Na figura 1, é possível visualizar um exemplo de como a inteligência artificial consegue diferenciar uma pupila aberta de uma fechada, retornando um parâmetro que é utilizado nas linhas de código para efetuar a classificação. Na figura 2, temos uma representação do modelo retornando este parâmetro como função do tempo. Deste modo, o mesmo pode agora ser implementado nas cabines dos veículos para monitoramento em tempo real do motorista e associado à um alerta, caso o mesmo seja detectado em alguma das classes pré-definidas.

166

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





Figura 1

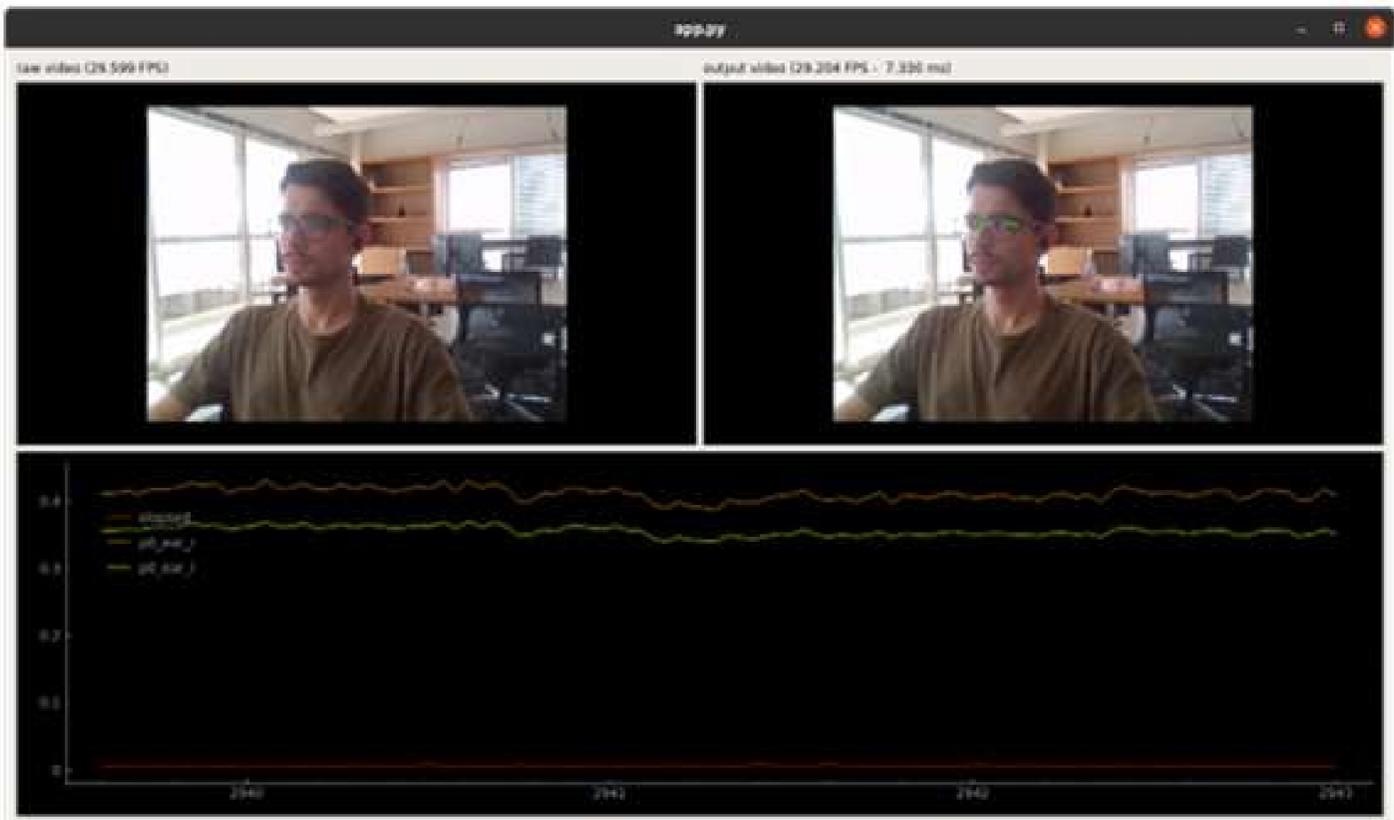


Figura 2

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## VIASOFT

1

MELHOR CONFIGURAÇÃO DE CAMINHÃO  
TANQUE**Residentes:**

Anna Luiza Gomes, Lucas Coutinho e Luis Gustavo T. Silva.

**Técnicas:**

[DS]: Agrupamento de pedidos utilizando a técnica K-Means e otimização de rotas pela API Google Directions. [Front-End]: Aplicação com Streamlit para visualização dos resultados.

168

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

A proposta de automatizar um processo tem como resultado final a otimização e aprimoramento do fluxo de trabalho dentro de uma empresa, além disso se torna possível reduzir custos, tempo, desperdícios, expandir a produtividade e reduzir possíveis falhas, através de um monitoramento mais preciso das atividades. Compreendendo os benefícios, o objetivo dessa sprint foi a criação de um sistema para automatizar as escolhas dos pedidos e configuração interna de um caminhão-tanque

através da tecnologia, e assim minimizar o tempo gasto pelos colaboradores da empresa e melhorar o aproveitamento de espaço interno do caminhão.

## SOLUÇÃO DESENVOLVIDA

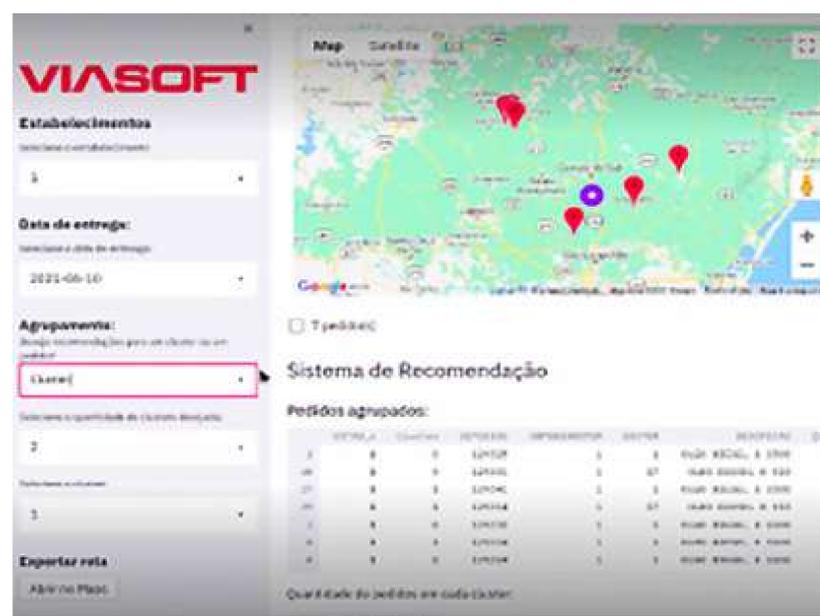
Foram desenvolvidos 3 algoritmos em Python, o primeiro foi para definir o melhor caminhão para entrega de um determinado pedido, onde foi utilizado uma árvore de possibilidades dentro da lógica. O segundo foi para agrupamento dos pedidos com base em sua localização e utilizando os dados de pedidos cadastrados no dataset. E o terceiro para definição da melhor rota para um determinado pedido, utilizando uma API Google Directions para solicitação de rotas. Para aplicação e interface front-end foi utilizada o Streamlit.

169

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## VIASOFT

2

## SISTEMA DE RECOMENDAÇÃO

**Residentes:**

André Luiz Marasca, Ananda Helena Heino Coelho e Daniel Victor Teixeira Lima.

**Técnicas:**

[DS]: Utilização de SQL para manipular os dados e flask para desenvolver uma API integrativa ao sistema do cliente. [IA]: LightFM.

170

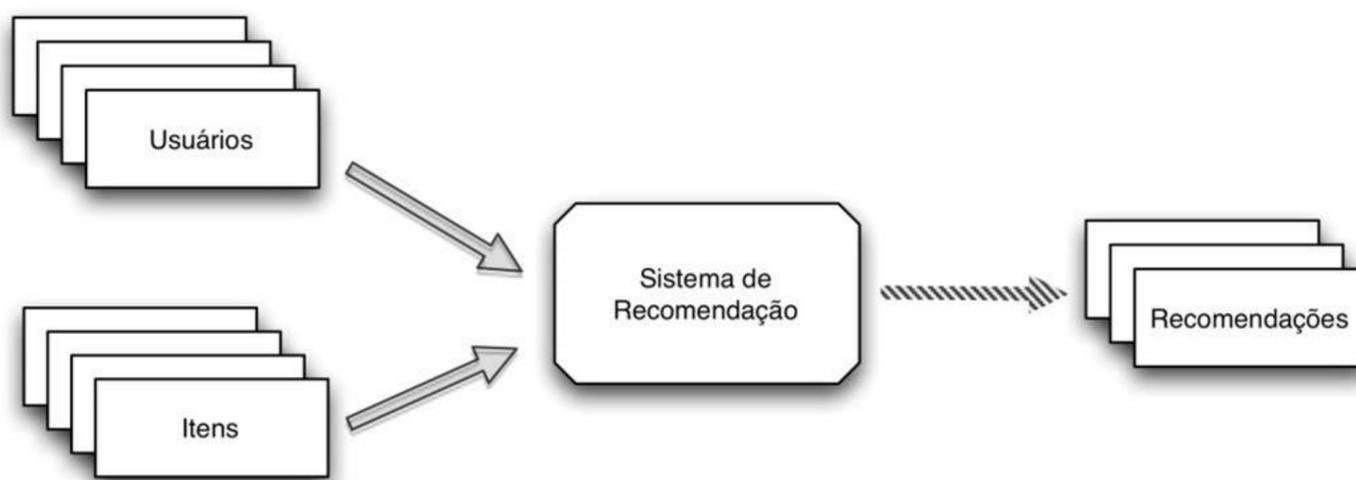
SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Um sistema de recomendação é um método de filtragem capaz de prever a melhor classificação ou preferência de um item para um usuário. O projeto proposto nessa Sprint foi o desenvolvimento de um sistema de recomendações capaz de criar um ranking de itens para um usuário, isso foi possível através de duas tabelas SQL, uma que continham históricos de compras de usuários e outra com a similaridade entre os itens.

## SOLUÇÃO DESENVOLVIDA

O algoritmo desenvolvido utiliza uma biblioteca chamada LightFM, ela utiliza uma técnica onde é possível se relacionar o usuário com o item e assim poder gerar um modelo que quando recebe os parâmetros de treino é capaz de criar um ranking de itens recomendados a um usuário, ordenado do mais relevante ao menos relevante. Para poder integrar o sistema de recomendações com o sistema da Viasoft, foi desenvolvida uma API em flask, assim ficaria de mais fácil acesso. A Inteligência Artificial é utilizada na LightFM que realiza previsões que apresentam recomendações de itens a um usuário, isso ocorre através do treino do modelo com os dados de compra do usuário e itens relacionados.



171

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



## VIASOFT

3

## AUTOMATIZAÇÃO DE BUSCA DE PREÇOS DE PRODUTOS DO SITE MENOR PREÇO PARANÁ

**Residentes:**

Anna Luiza Gomes e Matheus Novaes.

**Técnicas:**

[DS]: POC (Prova de Conceito–Proof of Concept), Selenium Id e WebDriver. [API]: Flask.

172

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Atualmente temos facilidade em realizar compras online e por conta disso grande parte desse consumo ocorre de formas impulsiva. Em vista desse fato, criar o hábito de realizar um comparativo de valores antes de adquirir o produto, além de fazer você encontrar o melhor preço, também evita de realizar compras impulsivas. Conhecendo essa boa prática, o objetivo da Sprint 3 é automatizar esse processo fazendo a busca automática por preço de uma lista de produtos a partir do site “Menor Preço”.

## SOLUÇÃO DESENVOLVIDA

Primeiramente foi desenvolvido um script para ler os dados que são códigos de barras, utilizou-se uma classe chamada DataReader para ler planilhas Excel no formato .xlsx onde supostamente estariam os dados. Após essa etapa, foi criado um script de Webscrapping do site Menor Preço onde é inserido o código do produto ou descrição, preenchido informações de filtragem e retornado o preço calculado e informações do detalhamento do cálculo. Por fim, foi desenvolvido uma API usando Flask para atender as requisições de busca de preços de produto.

173

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



The screenshot displays a web application interface for product price calculation. At the top, a progress bar shows four steps: 'Produtos' (Products), 'Filtros' (Filters), 'Preço' (Price), and 'Certificado' (Certificate). The 'Produtos' step is currently active. Below the progress bar, there is a search area with several filter options: 'Descrição Utilizada pelo Fabricante', 'Descrição Informada na NF-e', 'Quantidade de NF-e com esta Descrição', 'Possui outras Descrições Informadas', and 'Informação da Tabela da ANVISA'. There are also checkboxes for 'Medicamentos' and 'Combustíveis'. A search bar contains the barcode '7899674017086' and the product description 'DEO ANT ABOVE COUNTRY MASC 24X100ML 90G'. Below the search bar, there are two buttons: 'Limpar Produtos' (Clear Products) and 'Continuar: Informe os Filtros' (Continue: Inform Filters). The 'Preço' step is highlighted in the progress bar. Below the search area, there is a message: 'O preço foi calculado de acordo com os produtos e filtros abaixo apresentados, levando em consideração os preços praticados nas operações de compra e venda, autorizadas por meio das Notas Fiscais Eletrônicas - NF-e, no Estado do Paraná. Por favor, revise os dados abaixo e consulte o preço calculado.' The calculated price is displayed as 'Preço Calculado: R\$ 4,45 / Unidade (UN)'.

## VIASOFT

4

## SCRAPING E CLASSIFICAÇÃO DE NOTÍCIAS

**Residentes:**

Giovane Sylvestrin e João Kawazoe.

**Técnicas:**

[DS]: POC (Prova de Conceito–Proof of Concept) e Scrapping. [IA]: Machine Learning.

174

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Milhares de informações são geradas diariamente no mundo, onde a grande maioria se encontra em plataformas digitais. Diante desse cenário o ideal é que ocorra uma filtragem dessas notícias para se manter sempre atualizado. O objetivo dessa Sprint está conectado a essa mesma lógica, um desenvolvimento de uma ferramenta para coleta e classificação de notícias mais recentes, realizando assim uma prova de conceito a partir de uma amostra de 378 notícias fornecidas pela VisaSoft.



## **SOLUÇÃO DESENVOLVIDA**

A partir das amostras fornecidas foi possível fazer uma divisão em nove classes, com quantidades próximas, para que fosse treinando o modelo de classificação. Inicialmente foi realizado o Scrapping, que é a “ação” de coletar informações de fontes da internet, para coleta de notícias de determinados portais. Após a extração de notícias foi feito o Processamento de Linguagem Natural que realiza a conversão do conteúdo textual em um valor numérico para que seja possível utilizá-lo no modelo de classificação. Por fim foi criado o modelo de classificação de notícias, desenvolvendo o mesmo utilizando Machine Learning. Para se utilizar o modelo deve-se fazer a instalação das tecnologias dependentes, e sua saída é um arquivo Excel formato .csv com a listagem das notícias.

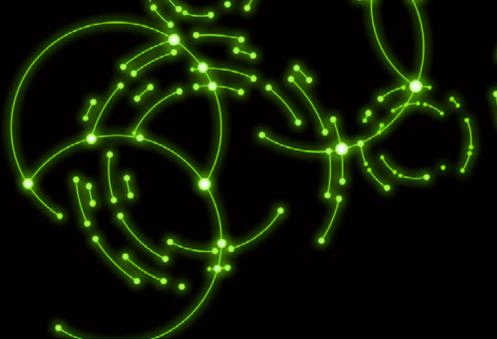
**175**

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





# VOTORANTIM

1

## PREDIÇÃO DE ANOMALIAS

**Residentes:**

Karen Itsuki Taniguchi, Marcia Regina Ramalho Rodrigues, Victor Hugo Nagahama.

**Técnicas:**

Regressão Linear.

176

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA

### OPORTUNIDADE DE NEGÓCIO

O processo de obtenção de semente fina pode ser complexo. De maneira geral, existem 3 etapas distintas (precipitação, classificação e filtração) onde, em cada uma delas, técnicas diferentes e específicas são aplicadas de maneira sequencial para a obtenção de um bom produto final. Como toda produção em massa, podem existir falhas durante o processo de obtenção do produto final, o que, em geral, não tem uma boa correlação com custos e/ou lucros. Apesar da complexidade do processo, a utilização de inteligência artificial pode ser capaz

de enxergar padrões no mesmo, com o objetivo de conseguir prever anomalias no resultado final e suas possíveis causas, para que assim, seja possível tomar medidas mais assertivas e diretas para o aumento na eficiência da produção.

177

### SOLUÇÃO DESENVOLVIDA

A principal meta do projeto é o desenvolvimento de um modelo preditivo da relação 25/44 da semente fina de 3 a 7 dias. Para esta finalidade, foi realizada uma análise exploratória dos dados disponibilizados pela empresa a fim de estabelecer quais são as variáveis relevantes. Nestas variáveis, foram aplicados modelos de Regressão Linear Robusta e Extra Trees. A acurácia do modelo é mostrada na figura a seguir. Os valores de erro quadrático médio (RMSE) mostra que as previsões possuem uma precisão muito boa, uma vez que quando menor o valor de RMSE menos incerteza existe na previsão.

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA

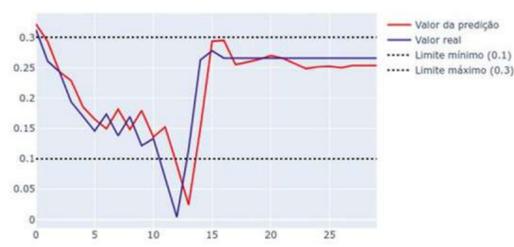


Comparação valor real e predito - Extra Trees



RMSE: 0.0415

Comparação valor real e predito - Robust model



RMSE: 0.0398

## VOTORANTIM

2

## CLASSIFICAÇÃO DE REGISTROS EM POTENCIAL DE GRAVIDADE

**Residentes:**

Germano Yoneda Pereira Lima, Larissa Luciane Ferreira, Natália da Silva de Sá.

**Técnicas:**

NLP, K-Means, KNN, BERT.

178

SUMÁRIO

SUMÁRIO  
CASES DO  
PROGRAMA**OPORTUNIDADE DE NEGÓCIO**

Apesar de uma situação não-ideal, acidentes de trabalho tendem a ocorrer na grande maioria das organizações. Por esse motivo, empregar técnicas e métodos para a redução do número de acidentes é essencial para promover a segurança e bem-estar de todos os funcionários e colaboradores da empresa, em geral.

Entretanto, por onde começar? Como identificar quais setores requerem atenção imediata e quais métodos para prevenir acidentes?

O primeiro passo para responder as perguntas



acima vem da classificação de acidentes em um banco de dados para que, assim, a análise exploratória dos mesmos possa identificar padrões e sugerir quais setores requerem alguma ação imediata, dado seu grau de gravidade. E, nesse ponto, é onde inteligências artificiais podem ser implementadas com maestria. Isto é, na classificação de eventos e procura por padrões.

### **SOLUÇÃO DESENVOLVIDA**

A partir dos registros de acidentes, foram utilizadas ferramentas do tipo NLP, que, por sua vez, consiste na habilidade do computador compreender a língua humana como ela é escrita/falada. Deste modo, uma inteligência artificial foi treinada no banco de dados provido pela empresa para que a mesma fosse capaz de identificar, através de palavras chaves, qual foi o tipo de acidente. Um exemplo é mostrado a seguir.

179

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA





Frase 1

Ao fechar o porta luvas do veículo (caminhão pipa), o motorista prendeu ACIDENTE o dedo ATINGIDA causando leve escoriação RESULTADO do indicador da mão ATINGIDA esquerda.

Frase 2

O empregado estava realizando treinamento quando a cadeira veio a entortar ACIDENTE os pés, ocasionando a queda RESULTADO dele no solo.

Frase 3

O funcionario teve um desequilíbrio ACIDENTE no ultimo degrau, o mesmo estava segurando o corrimão e bateu RESULTADO o joelho ATINGIDA no chão.

Com isso, foi possível a utilização de técnicas a fim de atribuir um valor à cada acidente, com o objetivo de classificá-los por gravidade. Este resultado permitiu a visualização gráfica destas classificações e, de maneira clara, pôde ser observado quais setores eram mais propícios à acidentes, de acordo com a gravidade do mesmo. Conforme o objetivo do projeto, este resultado facilita a identificação de setores que requerem algum ajuste de segurança, listados por prioridade.

180

SUMÁRIO



SUMÁRIO  
CASES DO  
PROGRAMA



# FICHA TÉCNICA

181

SUMÁRIO



**AUTORIA:**

Leonardo Goshi Sanches

**ORGANIZAÇÃO TÉCNICA:**

Éric Leopoldino e Leonardo Goshi Sanches

**REVISÃO:**

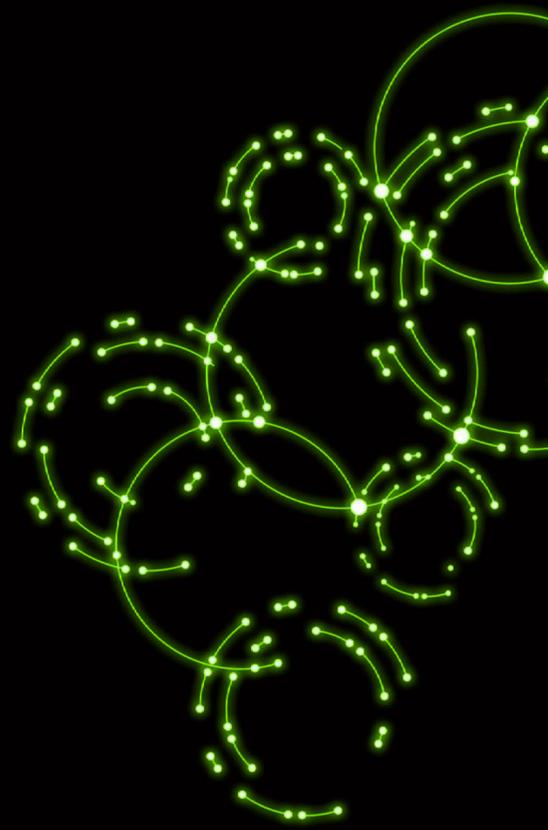
Paula Graziela de Carvalho

**PROJETO GRÁFICO DE DIAGRAMAÇÃO:**

Ricardo Hurmus

**DEPARTAMENTO REGIONAL DO PARANÁ**

*Qualquer parte desta obra pode ser reproduzida, desde que citada a fonte.*



QUER FAZER  
PARTE DO  
PROGRAMA DE  
RESIDÊNCIA  
EM INTELIGÊNCIA  
ARTIFICIAL DO  
SENAI?

**ACESSE O SITE**

[senaipr.com.br/tecnologiaeinovacao](http://senaipr.com.br/tecnologiaeinovacao)

