



UNIVERSIDADE FEDERAL DE ALAGOAS - UFAL
CENTRO DE TECNOLOGIA - CTEC
CURSO DE ENGENHARIA CIVIL

**SISTEMA DE SUPORTE À DECISÃO PARA
ALTERAÇÕES NAS LINHAS DE ÔNIBUS DE MACEIÓ
COM ÊNFASE ÀS AFETADAS PELA INTERDIÇÃO DO
BAIRRO MUTANGE**

Trabalho de Conclusão de Curso

Ewerton Amorim de Oliveira

Orientador: Alexandre Lima Marques da Silva

Coorientador: Prof. Dr. Bruno Almeida Pimentel

Maceió, janeiro de 2021



UNIVERSIDADE FEDERAL DE ALAGOAS - UFAL
CENTRO DE TECNOLOGIA - CTEC
CURSO DE ENGENHARIA CIVIL

**SISTEMA DE SUPORTE À DECISÃO PARA
ALTERAÇÕES NAS LINHAS DE ÔNIBUS DE MACEIÓ
COM ÊNFASE ÀS AFETADAS PELA INTERDIÇÃO DO
BAIRRO DO MUTANGE**

TCC apresentado ao colegiado do curso de Engenharia Civil da Universidade Federal de Alagoas como parte integrante dos requisitos para obtenção de título de Engenheiro Civil.

Trabalho de Conclusão de Curso

Ewerton Amorim de Oliveira

Maceió, janeiro de 2021

Catálogo na fonte
Universidade Federal de Alagoas
Biblioteca Central
Divisão de Tratamento Técnico

Bibliotecário Responsável: Valter dos Santos Andrade – CRB-4-1251

O48s Oliveira, Ewerton Amorim de.

Sistema de Suporte à decisão para alterações nas linhas de ônibus de Maceió com ênfase às afetadas pela interdição do bairro Mutange / Ewerton Amorim de Oliveira. – 2021.
48 f.

Orientador: Alexandre Lima Marques da Silva.

Coorientador: Bruno Almeida Pimentel.

Monografia (Trabalho de conclusão de curso em Bacharelado em Engenharia Civil) – Universidade Federal de Alagoas, Centro de Tecnologia. Maceió, 2021.

Bibliografia: f. 46-48.

1. Mineração de dados. 2. Sistema Web. 3. Transporte público – Maceió. 4. Transporte coletivo - Maceió. I. Título.

CDU: 656.11

AGRADECIMENTOS

A meus pais, que me guiaram para fazer o que gosto, priorizando os estudos e que são responsáveis por eu chegar até aqui e realizar esse sonho de criança. Eles foram essenciais para me ajudar a identificar quem eu sou e onde eu posso chegar, são minha base. Ao meu irmão que mesmo atualmente distante, se fez presente em todos os momentos, trazendo seus pensamentos que sempre me ajudam no processo de desconstrução e me dando aula de respeito e humildade.

À oportunidade de estudar em uma Universidade pública e de qualidade, na qual pude aprender muito com as tantas oportunidades que ela oferece. Àqueles que lutam contra o desmonte das universidades e que me permitiram esta oportunidade.

Ao Programa de Educação Tutorial, responsável pela minha manutenção financeira na universidade. No programa tive oportunidade de viajar pela primeira vez de avião, de sair do Nordeste pela primeira vez, de conhecer outras regiões do país, de viajar horas ou até dias de ônibus para participar de congressos. Nele eu pude vivenciar o princípio da horizontalidade, organizei alguns eventos, liderei, coordenei atividades, aprendi, me desconstruí, me tornei uma pessoa melhor. O programa desenvolveu uma base sólida e vou levar o que aprendi para o profissional que serei daqui em diante. Sentirei falta de todos/as meus/minhas amigos/as que conheci graças ao programa, dos InterPET, OxePET, EnePET, EnaPET e todo o pessoal envolvido na organização e participação destes. Carregarei comigo um pouco de cada um dos 3 grupos do CTEC, dos 12 grupos da Ufal, de grupos de Brasília, da Bahia, de Santa Maria, de Fortaleza, de São Paulo, São Carlos, Campinas além de tantas outras cidades e grupos.

Ao Centro Acadêmico de Engenharia Civil, onde tive a oportunidade de lutar pela melhoria do meu curso, de me desenvolver politicamente, de ser representante discente no colegiado. Agradeço a todos das chapas Empolgação pela inspiração e às chapas Consolidação, Realiza e Avante pela ajuda e união na luta. Aos Chapocs por estarem sempre disponíveis e serem referência e ao Freitas e sua sintonia de trabalho e amizade.

Aos meus tutores, orientadores, amigos, professores que fizeram parte da minha vida acadêmica de alguma forma, principalmente os/as professores/as Roberaldo, Lucena, Christopher, Adriana, Alexandre, Bruno, Rochana, Juliane, Jessica, Andressa, Cid, Gianna, Barbirato, Roseline, Luciano, Luciana e tantos outros.

À minha amiga, e primeira chefe Olga, por ter sido a primeira pessoa a ter me apresentado a parte prática da engenharia aplicada à área de Transportes.

À minha amiga e chefe Paula, por ter me apresentado a gestão de dados aplicada ao transporte, por sempre me apoiar, por ter me ensinado tanta coisa e pela oportunidade de me fazer conhecer o que eu gosto de trabalhar e que foi essencial para este TCC, desde ideias, sugestões e reflexões.

À SMTT, por ter fornecido todo o material base para este TCC, por ter me concebido a oportunidade de estágio, com remuneração que foi muito importante neste período de pandemia.

Aos meus amigos, principalmente os presentes e que ouviram algum/uns comentário/s sobre o TCC e que foram essenciais para manutenção da minha saúde mental nessa pandemia e neste período de turbulência pré- formação. Aos meus grupinhos do BNegona, do jantar, do Dizem e do All-Stars.

À toda comunidade de desenvolvimento de código aberto como Github e Stack Overflow, como também aos desenvolvedores das bibliotecas utilizadas neste trabalho.

Ao meu orientador e coorientador, por me ajudarem a desenvolver este trabalho.

RESUMO

Maceió é uma cidade com problemas de mobilidade urbana, com demanda decrescente de Transporte Público Coletivo (TPC). Apesar de existirem sistemas inovadores de bilhetagem eletrônica e GPS eles não se comunicam, gerando a necessidade de um sistema que permita o acesso aos dados de demanda a nível espacial e temporal. Desta forma, para facilitar o trabalho da gestão, o objetivo deste trabalho foi o desenvolvimento de uma ferramenta de suporte à decisão para gerenciamento das linhas de Maceió. Como estudo de caso, foram utilizadas as linhas que foram afetadas pela interdição do bairro Mutange. Foi construída uma base para alimentação dos dados, um sistema WEB para visualização e algoritmos de classificação para ajudar na tomada de decisão. Foi possível fazer análises e encontrar pontos críticos das linhas, embora foram encontradas algumas inconsistências nos dados ao cruzar as informações. O sistema desenvolvido pode ser utilizado para melhorar a robustez e mitigar perdas de um sistema de TPC.

Palavras-chave: *Mineração de dados, Sistema WEB, Transporte Público Coletivo*

ABSTRACT

Maceió is a city with mobility problems and decreasing Public Transportation (TPC) demand. Although there are innovative electronic ticketing and GPS systems, they do not communicate with each other. It generates the need for a system that allows access to demand data at a spatial and temporal level. Thus, to facilitate the management, the work aim was to develop a decision support tool for Maceió's bus route management. The case study was the affected lines by the interdiction of the Mutange neighborhood. The data has fed at a database. A WEB system for visualization and classification algorithms helps in the decision making. It was possible to analyze and to find critical points of the lines, although some inconsistencies were found in the data when crossing the information. The developed system improves the robustness and mitigates losses of a TPC system.

Palavras-chave: Data Mining, WEB System, Public Transportation

GLOSSÁRIO

API – Application Programming Interface (Interface de programação de Aplicação): Terminologia utilizada para disponibilização de serviços, funcionalidades e informações de um aplicativo a um terceiro, sem dar detalhes da implementação, apenas dos dados resultantes.

Backend – Parte da aplicação que contém o banco de dados, controle e armazenamento de informações.

BE – Bilhetagem Eletrônica

BRT – Bus Rapid Transit (Ônibus de Trânsito Rápido): É um sistema de Transporte Público Coletivo de massa, com infraestrutura criada para agilizar o fluxo de passageiros.

Dashboard – Painel que mostra as métricas mais importantes para gestão dos dados.

Framework – Biblioteca/Aplicação genérica que tem o objetivo de simplificar processos comuns de diferentes soluções.

Frontend – Parte da aplicação que contém os componentes que interagem com o usuário (cliente da aplicação).

GPS – Global Positioning System

ML – Machine Learning (Aprendizado de Máquina)

OD – Origem-Destino

Overfitting – Quando o modelo se ajusta bem aos dados de treino, mas falha nos dados de teste/validação.

PPP – Passageiros por Ponto

Responsivo – Relativo à responsividade; adaptável a diferentes formatos de tela

SIMM – Sistema Integrado de Mobilidade de Maceió

SMTT – Superintendência Municipal de Transportes e Trânsito de Maceió

TI – Terminal Integrado de Passageiros: Terminal Rodoviário onde ocorre transbordo de passageiros com integração física e tarifária

TPC – Transporte Público Coletivo

TPP – Tempo por ponto

Sinturb-MAC – Sindicato das Empresas de Transporte Urbano de Passageiros do Município de Maceió

TV – Tempo de Viagem

VLT – Veículo Leve sobre Trilhos

SUMÁRIO

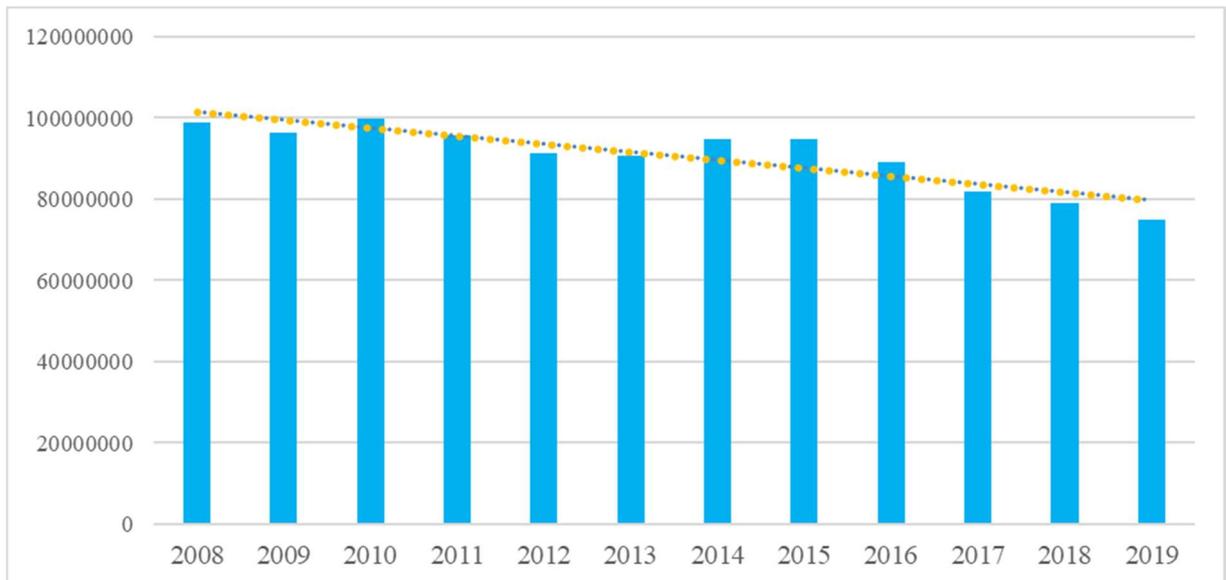
1.	INTRODUÇÃO.....	1
2.	OBJETIVOS.....	6
3.	FUNDAMENTAÇÃO TEÓRICA	7
3.1.	Leitura e unificação dos dados.....	7
3.2.	<i>Dashboard</i> de análises de dados de passageiros.....	8
3.3.	MODELAGEM DE CLASSIFICAÇÃO.....	9
3.4.	INFORMAÇÕES DO ESTUDO DE CASO	12
4.	METODOLOGIA.....	15
4.1.	Leitura e unificação dos dados.....	15
4.2.	<i>Dashboard</i> de análises de dados de passageiros.....	18
4.3.	MODELAGEM DE CLASSIFICAÇÃO.....	20
4.4.	ESTUDO DE CASO.....	22
5.	RESULTADOS e discussão	23
5.1.	Desenvolvimento do CÓDIGO.....	23
5.2.	Leitura e unificação dos dados.....	23
5.3.	<i>Dashboard</i> de análises de dados de passageiros	28
5.4.	classificação	30
5.5.	estudo de caso	32
6.	CONCLUSÕES.....	44
	REFERÊNCIAS	46

1. INTRODUÇÃO

Dentre os problemas das cidades com desenvolvimento desordenado, está a mobilidade urbana. Maceió, possui uma população de aproximadamente 1 milhão de pessoas (IBGE, 2019) e apresenta diversos problemas de mobilidade, com uma infraestrutura que prioriza os carros, exceto pela Faixa Azul (Corredor de transportes coletivos) presente em 4 avenidas da cidade.

Enquanto há um movimento crescente de ações que priorizam o transporte coletivo em grandes cidades, Maceió possui atualmente uma média diária (dias úteis) de 123 mil viagens de passageiros (Figura 1), incluindo gratuidades e integração temporal. Estes dados indicam que a demanda de passageiros tem diminuído.

Figura 1 - Demanda de passageiros ao longo dos anos



Fonte: (SMTT, 2020)

Segundo classificação do Guia do Transporte Público Coletivo (AMICCI et al., 2018) presente na Figura 2, Maceió se enquadra como cidade de grande porte, cujo projeto de intervenção recomendado varia de faixa exclusiva a trem urbano, o que mostra o potencial da cidade para grandes intervenções por parte da gestão. Para qualquer estudo de viabilidade de sistemas de intervenção, é necessário um sistema de suporte à decisão que mostre o comportamento da mobilidade e demanda de forma precisa.

Figura 2 - Compatibilidade de sistemas de TPC de acordo com o porte da cidade



Fonte: (AMICCI et al., 2018)

O Sistema de Transporte Público Coletivo (TPC) de Maceió é gerenciado pelo Sistema Integrado de Mobilidade de Maceió (SIMM), instância criada a partir do Edital de licitação do transporte coletivo da cidade (CONCORRÊNCIA NACIONAL CEL/SMG N° 001/2015) e gerenciado pela Superintendência Municipal de Transportes e Trânsito (SMTT). Conta atualmente com 105 linhas, distribuídas em 4 concessionárias, o que representa um grau de complexidade baixa, comparado a grandes cidades.

Toda a frota de ônibus coletivo é equipada com as seguintes ferramentas tecnológicas:

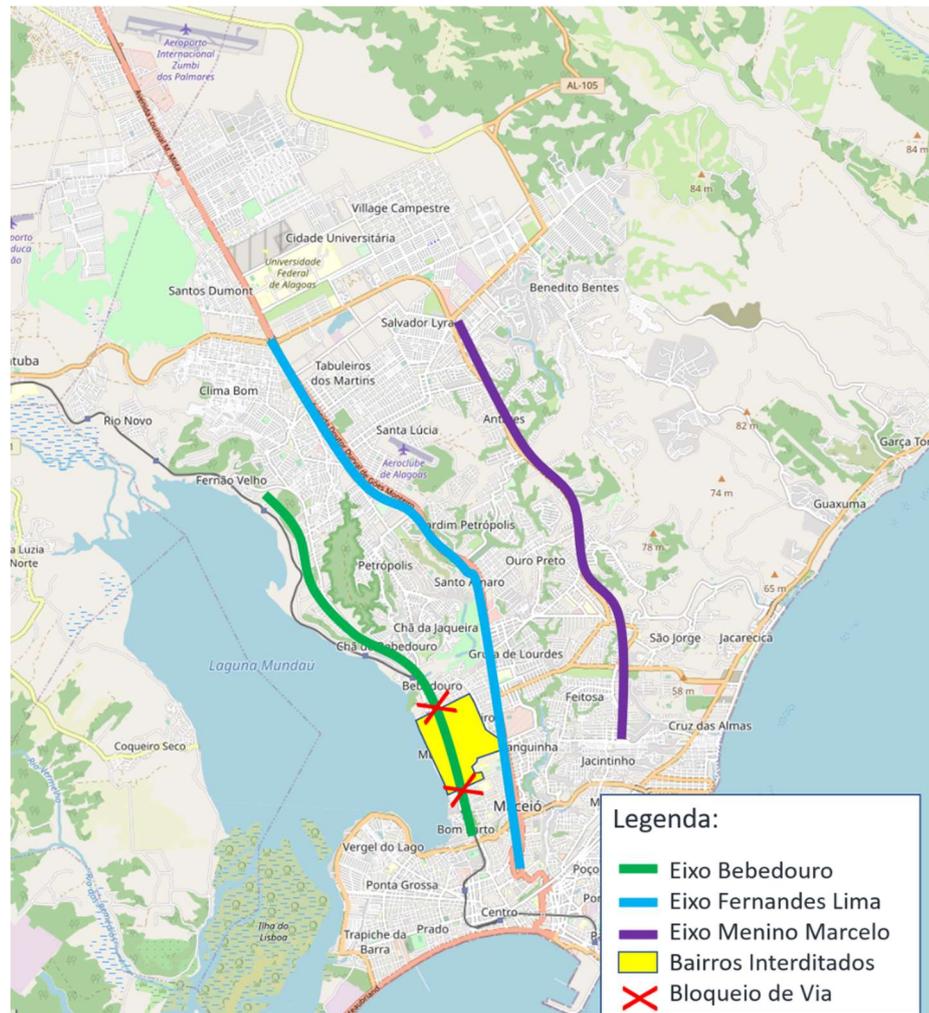
- Geolocalização e controle de toda a operação no que diz respeito a espaço e tempo através de GPS.
- Bilhetagem eletrônica com registro de todos os passageiros.

Embora sejam ferramentas importantíssimas e de grande apoio, elas não se comunicam devido à distribuição de responsabilidades que faz com que cada uma seja gerenciada por uma empresa diferente. Assim sendo, o órgão gestor e fiscalizador encontra dificuldades quando é necessário fazer com que as informações de bilhetagem eletrônica e GPS precisam se comunicar.

Até o início de 2020, o TPC de Maceió se caracterizava a partir de 3 eixos principais de transportes (Figura 3), sendo eles, o eixo Menino Marcelo, o eixo Fernandes Lima/Durval de Góes Monteiro e o eixo Bebedouro. Porém, houve a constatação de instabilidade do solo fruto de desastre ambiental ocasionado por mineração. Com isso, o bairro do Mutange foi

completamente interditado e o bairro do Pinheiro parcialmente. O primeiro bairro tinha cerca de 2700 pessoas e o segundo cerca de 20 mil (DADOS AL, 2020). Os moradores ficaram sem suas casas e tiveram que abandonar os bairros, houve a interdição de parte de uma linha de VLT, e da avenida principal de ligação do eixo Bebedouro ao Centro da Cidade, o que trouxe grandes prejuízos para a mobilidade da cidade, afetando não só os moradores dos bairros.

Figura 3 - Principais eixos de Transportes de Maceió



Fonte: Autor

Dentre as linhas que atendiam a região afetada, 3 linhas foram extintas e, dentre as que restaram, existem linhas cujos itinerários ligam as mesmas regiões, com alta sobreposição entre as linhas. O fluxo destas linhas cujo itinerário fora alterado, foi direcionado para outro eixo da cidade, o eixo Fernandes Lima/Durval de Góes Monteiro (Figura 4). Com isso, houve um aumento na sobreposição das linhas e, por isso, pareceu interessante investigar a possibilidade de otimização dessas linhas.

Figura 4 - Divulgação da interdição da Avenida principal do Bairro Mutange



Fonte: Prefeitura de Maceió

Através de uma ferramenta adequada para o gerenciamento dos dados, pode-se melhorar o sistema de transportes de uma cidade. A estratégia recomendada para tomada de decisões no transporte coletivo é a matriz origem-destino (OD) da cidade, que pode ser feita através de consulta em domicílios ou terminais, mas que demanda investimento alto aos cofres públicos, sem falar que, na maioria das vezes trata de uma fotografia dos resultados associados a um momento, desconsiderando alterações sazonais.

A partir desses dados, o dimensionamento de uma linha consiste na obtenção da quantidade de passageiros no trecho crítico de cada linha. Como os dados de passageiros existentes acessíveis pelo SIMM são apenas os dados de embarque de passageiros totais por sentido, fica inviável

qualquer tomada de decisão com base nessas informações, pois desconsidera a divisão detalhada em trechos, além de nenhuma informação quanto ao desembarque dos passageiros.

Existem relatórios isolados que possuem a informação de cada embarque além de informações de GPS, separados por linha, sentido e por veículo, embora a agregação desses dados traga os seguintes problemas:

- A série temporal de embarque contém apenas a hora e não há relação com a posição espacial dos coletivos;
- Não se sabe onde cada passageiro desce;

Em contrapartida, estudos apontam para a utilização de análises com os dados que são possíveis de obter (Dados que a prefeitura já tem em mãos) (TANG; LIU; CHOUDHURY, 2020). As estimativas partem de abordagens determinísticas ou estatísticas. Uma abordagem interessante é o uso de Aprendizado de Máquina ou *Machine Learning* (ML), na qual tem sido incorporado por alguns estudos de mobilidade (JUNG; SOHN, 2017a; MA et al., 2013; QI et al., 2018). Deste modo é possível inferir sobre comportamentos de passageiros a partir de dados observados.

Vale ressaltar que todas as linhas de Maceió tiveram queda de demanda de aproximadamente 45% no ano de 2020 (SMTT, 2021), além de possíveis alterações nos períodos de pico devido ao tele trabalho ou trabalho híbrido, mundialmente adotados devido à pandemia do novo coronavírus no ano de 2020.

Parece interessante uma ferramenta na qual o operador do sistema possa ter gráficos e diagramas como resposta da análise, a fim de facilitar o entendimento e ajudar na tomada de decisão de possíveis alterações no sistema, visando sua otimização.

Diante da problemática apresentada, para facilitar o trabalho da gestão, propõe-se uma ferramenta de suporte à decisão para gerenciamento das linhas de Maceió. Como estudo de caso, serão utilizadas as linhas afetadas pelas mudanças no Mutange.

Com esta ferramenta, espera-se auxiliar na identificação de pontos problemáticos, a partir de algoritmos de aprendizado de máquina de classificação, ajudando no estudo de possíveis soluções, tais como a adoção de eixos troncais com corredores exclusivos (Corredores centrais, BRT, VLT), seccionamento e/ou ampliação de frequência ou percurso de linhas ou viagens.

2. OBJETIVOS

Este trabalho tem por objetivo geral a criação de um sistema de suporte à decisão a fim de criar uma base única e simular alterações de linhas no SIMM;

Os objetivos específicos são:

- Obtenção da quantidade de embarque dos passageiros por trecho de cada linha;
- Estimação da quantidade de desembarque dos passageiros por trecho de cada linha;
- Elaborar Matriz OD das linhas que atendem à região do Mutange;
- Identificar pontos problemáticos de linhas a partir de modelos de classificação.

3. FUNDAMENTAÇÃO TEÓRICA

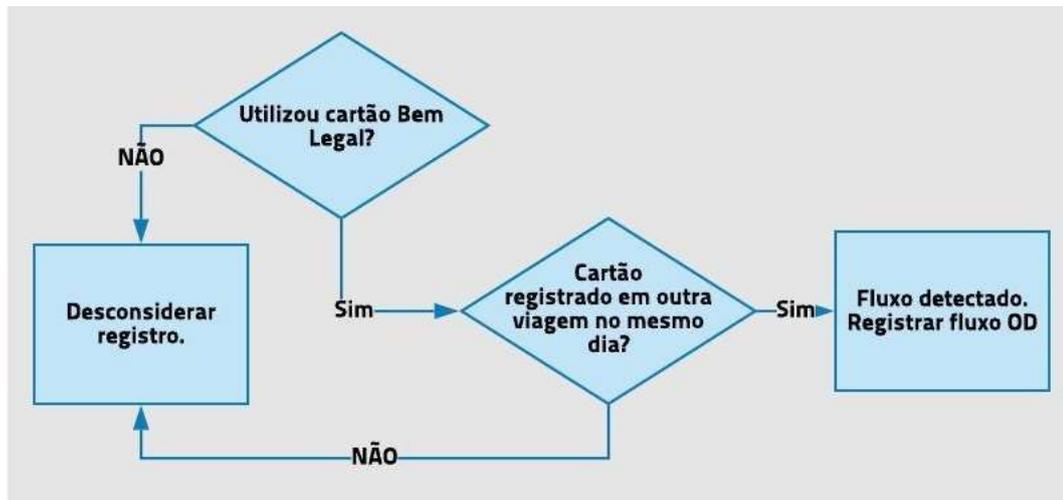
A base da construção deste trabalho passa pelas ferramentas de leitura e unificação dos dados, de construção do sistema WEB, de *Machine Learning* e informações sobre o estudo de caso.

3.1. LEITURA E UNIFICAÇÃO DOS DADOS

Uma problemática abordada por estudos brasileiros é a ausência de dados de desembarque de passageiros e a dificuldade financeira de se obter a matriz OD. Além disso, a variação ao longo do dia e sazonal Qi et al. (2018) tem sido discutida pelo fato de que a mobilidade de uma cidade sofre alterações constantes. Como alternativa, Cui (2006) propôs um modelo de estimação do comportamento temporal e espacial a partir de dados de bilhetagem eletrônica. A diferença entre este estudo e a realidade brasileira é que a cidade estudada (Pequim) conta com um sistema que armazena o local e horário de cada embarque e desembarque, o que reforça a necessidade de se conhecer o comportamento dos desembarques de passageiros, inexistente em Maceió.

A falta de conhecimento acerca dos locais e horários de desembarque são relatados por diversos trabalhos (ALSGER et al., 2016; CUI, 2006; JUNG; SOHN, 2017b; TOQUÉ et al., 2016; WANG; ATTANUCCI; WILSON, 2011), pois inúmeras cidades não possuem infraestrutura para registrar o desembarque dos passageiros. A Figura 5 ilustra a metodologia proposta inicialmente por Cui (2006). Para o caso particular de Maceió, o cartão Bem Legal é utilizado na bilhetagem eletrônica. Com base nestes dados, pode-se alimentar um banco de dados com a origem-destino, para posteriormente ser utilizado em cada viagem realizada em uma determinada linha.

Figura 5 - Metodologia para identificação dos locais de desembarque



Fonte: (ALSGER et al., 2016; CUI, 2006; JUNG; SOHN, 2017b; TOQUE et al., 2016; WANG; ATTANUCCI; WILSON, 2011) (adaptado)

Desta forma, alguns passageiros podem ser mapeados quanto aos fluxos OD. Porém, neste caso, os passageiros que validam a passagem com dinheiro e os que utilizam o cartão apenas uma vez no dia não podem ter seus fluxos diários identificados.

3.2.DASHBOARD DE ANÁLISES DE DADOS DE PASSAGEIROS

Atualmente, um sistema WEB pode ser dividido entre *backend* e *frontend*, no qual o *backend* fica no servidor e é onde as informações ficam salvas (banco de dados e arquivos). Já no *frontend* se encontram os componentes da aplicação que entra em contato com o usuário final (interface). Um *frontend* pode ser desenvolvido independentemente do *backend*, ou pode ser atrelado a ele.

Dentre as vantagens de uso de uma aplicação com *frontend* e *backend* está a liberdade de escolha de ferramentas de desenvolvimento diferentes.

Dentre as linguagens de programação mais utilizadas do mundo estão o *JavaScript* (1º) e o *Python* (2º) (GITHUB, 2020). A primeira tem como suas principais vantagens, o desenvolvimento web, já a segunda se destaca em ciência dos dados e ML.

Visto isto, visando uma aplicação WEB, com boa interação com usuário e robusta, com processamento de dados e aplicação de ML, parece interessante uma solução com *Python* como *backend* e *JavaScript* como *frontend*. A comunicação entre as duas frentes ocorre via API, que são acessadas por *url*. O *frontend* em *javascript* é responsável pela geração de gráficos e informações visuais, com a informação que vem pronta do *backend* através de uma estrutura de

dados (API) contendo apenas as informações relevantes e já processadas pelo backend em Python.

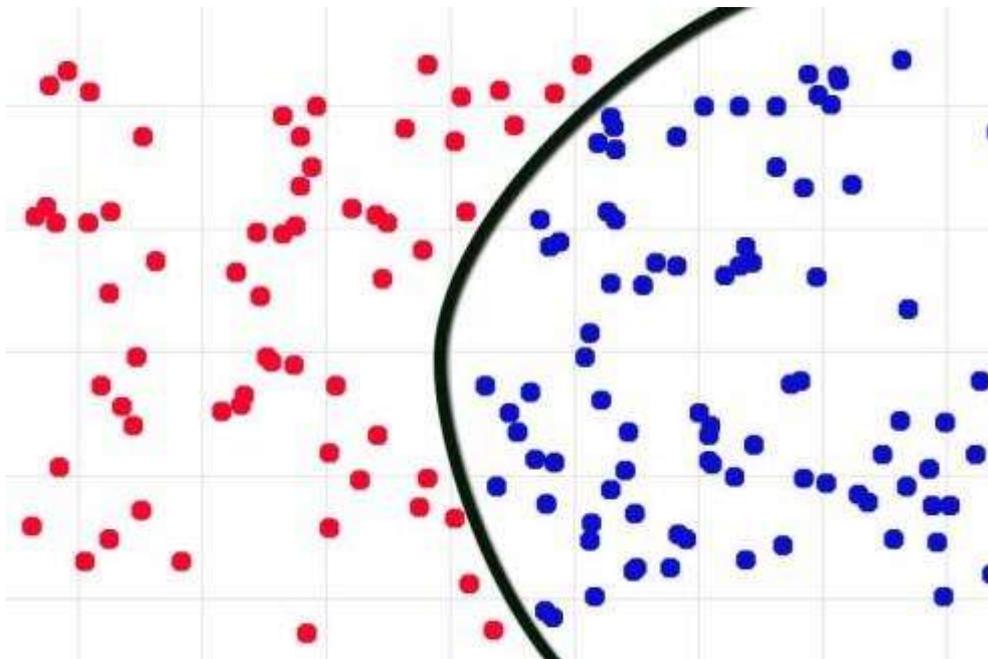
3.3.MODELAGEM DE CLASSIFICAÇÃO

A partir das demandas de passageiros devidamente estimadas, pode-se realizar um estudo de padrões que ocorrem na série de dados. Para isso tem sido utilizada inteligência artificial através de diversos algoritmos.

O aprendizado de máquina ou *Machine Learning* (ML) funciona a partir do treino de um algoritmo que visa otimizar um certo critério ou escolha, usando grandes conjuntos de dados (ALPAYDIN, 2020). Para tanto, utiliza-se de conjuntos de dados como exemplo ou observações anteriores.

Dentre os algoritmos de ML, existem os algoritmos de classificação, cujo objetivo é dividir os conjuntos de dados com base em características semelhantes entre eles. Cada dados (ou ponto) possui diversas características e uma classificação.

Figura 6 - Esquema de Classificação



Fonte: (BITREFINE, 2020)

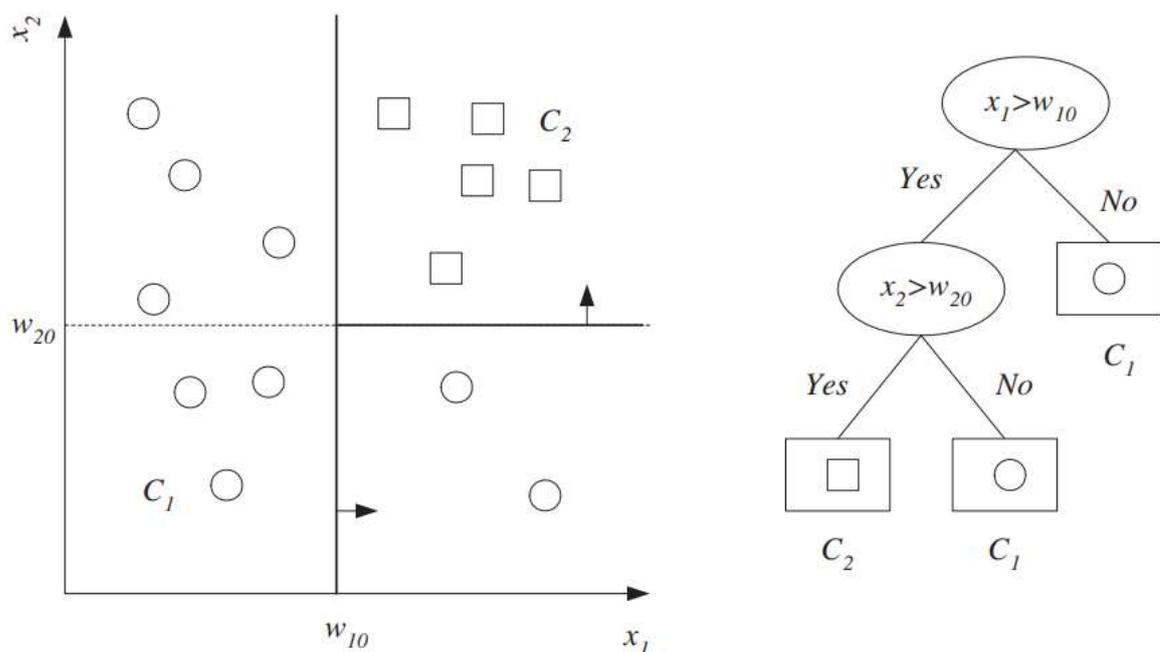
A decisão de qual classificação se dará para cada ponto ocorre a partir de algoritmos supervisionados e não supervisionados. Os supervisionados ocorrem a partir de dados cujas

classificações são conhecidas e os não supervisionados quando não se sabe a classificação de nenhum dado e busca-se alguma recomendação.

A Árvore de Decisão é um algoritmo supervisionado que pode funcionar tanto para classificação quanto para regressão. Um ponto interessante dessa abordagem é que ela permite uma resposta na forma de diagrama, que representa as escolhas do modelo. A vantagem de se ter uma visualização das características responsáveis pela classificação do conjunto é que a sugestão de decisão fica bem nítida.

Na Figura 7 pode-se observar uma visualização de uma Árvore de Decisão acompanhada de seu conjunto de dados. Os pontos ovais representam decisão e os retângulos são nós de folhas. Os valores w_{10} e w_{20} são os pesos considerados na decisão.

Figura 7 – Exemplo de uma Árvore de Decisão e sua correspondente árvore da decisão



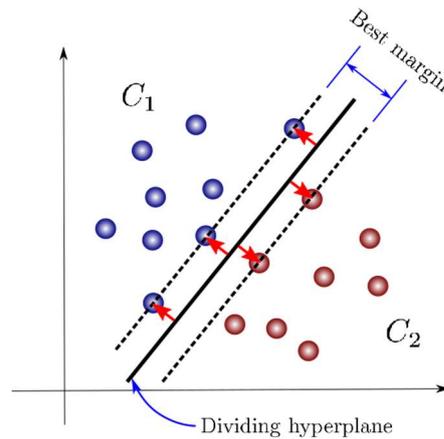
Fonte:(ALPAYDIN, 2020)

Outros métodos de classificação permitem a entrada de dados caracterizados da mesma forma que a Árvore da Decisão, com duas respectivas classificações e permitem a classificação de novos casos de entrada. Contudo, estes não fornecem uma abordagem mais visual das decisões tomadas responsáveis pela classificação.

O Algoritmo *Random Forest* funciona de forma semelhante à Árvore da Decisão, porém incorpora reamostragem de diversas Árvores, formando "Florestas", o que dá o nome à classificação. A ideia desta abordagem é evitar o procedimento de *overfitting* (DSA, 2020).

O algoritmo de classificação a partir de Máquinas de Vetores de Suporte (*Support Vector Machine* – SVM) (Figura 8) ocorre a partir da obtenção de uma linha de separação (também chamado de Hiperplano) entre os dados de duas classes diferentes. O modelo busca a maior distância entre dois pontos de classes diferentes (HSU; CHANG; LIN, 2003).

Figura 8 - Visualização do SVM

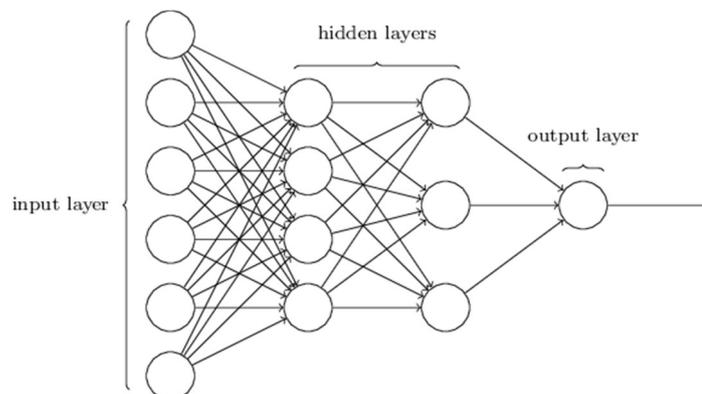


Fonte: Autor

Este modelo é sensível a números maiores que 1, por isso os dados devem ser ajustados para uma escala de -1 a 1 no momento da modelagem (CONTRERAS, 2019).

A classificação com Redes neurais usa de um algoritmo de Perceptrons de Multicamadas – *Multilayer Perceptrons* (MLP) que utiliza da relação de diferentes pontos (neurônios) em diferentes camadas a partir da atribuição de pesos (Figura 9) e contando com propagação reversa (Que trata da repetição do procedimento de estimativa a fim de otimizar a classificação) (NIELSEN, 2015).

Figura 9 - Esquema de uma rede neural



Fonte: (NIELSEN, 2015)

O algoritmo gaussiano *Naive Bayes* para classificação assume que a verossimilhança entre as características segue a distribuição Gaussiana descrita na Equação 1.

$$P(x_i | y) = \frac{1}{\sqrt{2\pi\sigma_y^2}} \exp\left(-\frac{(x_i - \mu_y)^2}{2\sigma_y^2}\right) \quad (\text{Equação 1})$$

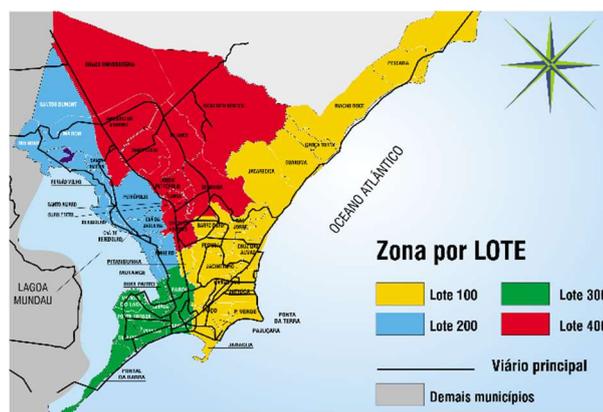
A modelagem dos diferentes estimadores ocorre a partir da separação dos dados em conjuntos de treino e teste, no qual o primeiro é onde os algoritmos aprendem com os dados, identificando as características responsáveis por dar determinada classificação aos dados. A forma que esses dados são separados no momento da modelagem interfere diretamente no resultado.

Dado um conjunto de dados, uma análise mais robusta adota o maior número possível de dados no momento de treino que implique em um bom ajuste nos dados a serem validados. Como solução para este problema, é utilizado o procedimento de Validação Cruzada, estratificando pequenos conjuntos de dados, mas mantendo as características no conjunto maior (ALPAYDIN, 2020).

3.4. INFORMAÇÕES DO ESTUDO DE CASO

O TPC de Maceió é dividido em 4 lotes, cujo lote que contém o bairro do Mutange é o lote 200 – Empresa São Francisco, conforme Figura 10.

Figura 10 - Divisão dos Lotes do TPC de Maceió

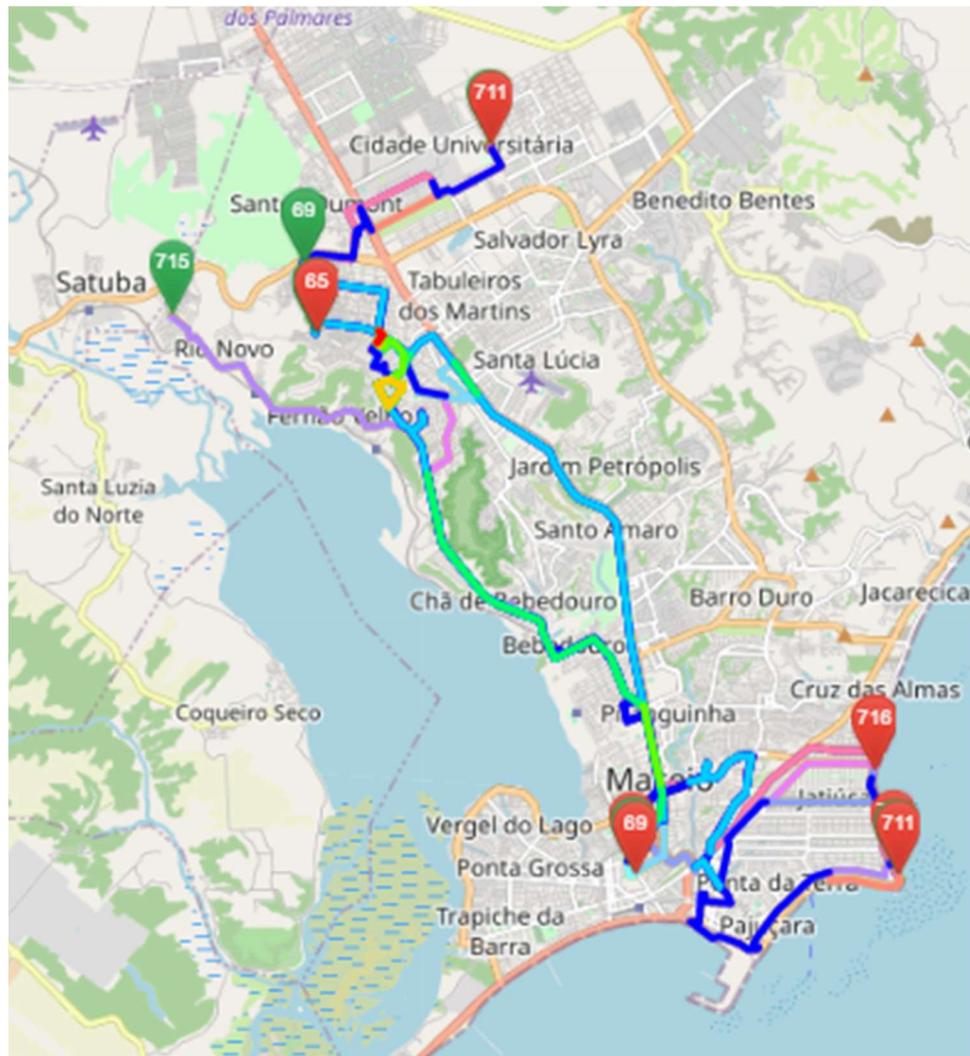


Fonte: Prefeitura de Maceió

Seguindo a Figura 11, foram selecionadas as linhas afetadas diretamente ou indiretamente pelas alterações do bairro do Mutange, ocasionadas pelo bloqueio do eixo Bebedouro. As linhas afetadas indiretamente são linhas que partem das mesmas origens de linhas afetadas, com o mesmo destino, mas utilizando de outro corredor, como o eixo da Fernandes Lima. Como

exemplo, a linha 711 – Ufal / Ponta Verde (Via Clima Bom / Osman / Bebedouro), liga a região da Colina e Tabuleiro à orla Central de Maceió, passando pelo Eixo Bebedouro, sendo diretamente afetada. Já a linha 716 – Clima Bom / Ponta Verde (Via Farol) liga a mesma região à Orla de Maceió, utilizando do Eixo Fernandes Lima, para o qual o fluxo do bebedouro foi direcionado, sendo indiretamente afetada.

Figura 11 – Mapa das linhas do lote 200 que atendiam à região do Mutange e linhas semelhantes



Fonte: Prefeitura de Maceió

Dentre as linhas do lote 200, foram selecionadas apenas as que partem de bairros da parte alta, nas proximidades da Colina dos Eucaliptos e Tabuleiro, cujos nomes encontram-se no Quadro 1.

Quadro 1 - Linhas afetadas pela interdição do Mutange

Código	Linha	Empresa
051	Santos Dumont / Centro (Via Poço)	Empresa São Francisco
057	Rio Novo / Centro (Via Farol)	Empresa São Francisco
058	Fernão Velho / Centro (Via Farol)	Empresa São Francisco
065	Rosane Collor / Centro (Via Farol)	Empresa São Francisco
068	Colina / Centro (via Bebedouro - Pç. Faculdade)	Empresa São Francisco
069	Clima Bom / Centro (Via Farol)	Empresa São Francisco
108	Clima Bom / Trapiche (Via Bebedouro)	Empresa São Francisco
711	Ufal / Ponta Verde (Via Bebedouro)	Empresa São Francisco
712	Santos Dumont / Ponta Verde (Via Farol)	Empresa São Francisco
715	Rio Novo ou Rosanne Collor / Ponta Verde	Empresa São Francisco
716	Clima Bom / Ponta Verde (Via Farol)	Empresa São Francisco

4. METODOLOGIA

A metodologia deste trabalho está dividida em 4 etapas: Leitura e unificação dos dados, *Dashboard* de análise de dados de passageiros, modelagem de classificação e por fim, as atividades desenvolvidas no estudo de caso.

4.1. LEITURA E UNIFICAÇÃO DOS DADOS

A obtenção dos dados segue as informações na listagem no Quadro 2.

Quadro 2 - Dados a serem obtidos de outras fontes

Sigla	Dados	Formato de Arquivo	Fonte
TV	Tempo de viagem de cada atendimento*	Planilha eletrônica (xlsx)	<i>GoolSystem</i>
TPP	Passagem por ponto de cada veículo em cada linha	Planilha eletrônica (xlsx)	<i>GoolSystem</i>
GP	Georreferenciamento dos pontos de ônibus de Maceió	Banco de dados da SMTT	SMTT
BE	Validações de cartões / botoeira na catraca	Arquivos Separados por vírgulas (.csv)	<i>Report Center</i>

* Uma linha em Maceió pode ter diferentes atendimentos a depender do itinerário (Avenidas ou terminais diferentes em um trecho pequeno da linha) e do serviço (convencional, semi-expressa, exclusivamente bilhetagem eletrônica).

Foram criados códigos para otimizar o download dos dados a partir dos relatórios disponíveis para o órgão de fiscalização do TPC, utilizando da biblioteca *Selenium*, que simula o comportamento de um usuário acessando uma página *WEB*.

Foram desenvolvidos códigos para leitura dos arquivos contendo os dados, utilizando a biblioteca *Pandas*. Como produto da leitura dos dados, obteve-se os seguintes subprodutos:

- a) uma série temporal de cada veículo por ponto, contendo código e endereço destes, porém sem a localização geográfica em coordenadas (Passageiros por ponto - PP).
- b) Uma série temporal de validações de cartão de passageiros (Bilhetagem Eletrônica - BE).

O embarque estimado de cada passageiro foi encontrado a partir dos dados de PP e BE, utilizando como índice a hora de passagem pelo validador e o último ponto mais próximo dessa hora registrado pelo veículo na linha utilizada pelo passageiro, visto que em ambas séries de dados, possui-se o código do veículo e a linha.

Deve-se destacar que os arquivos fornecidos pelo *GoolSystem* contêm apenas os códigos e endereços de cada ponto, com códigos diferentes por empresa para a mesma localização. Desta forma, as coordenadas de cada ponto foram obtidas a partir dos endereços, através de um código de conversão utilizando uma biblioteca do *Google Maps*. Ocorrem alguns erros nessa conversão, ocasionando uma divergência nas localidades encontradas. Para isso, é necessário comparar visualmente os dados encontrados com as posições dos pontos na página de monitoramento do *GoolSystem*.

Na Figura 12 pode observar a comparação entre as coordenadas reais do *GoolSystem* (lado esquerdo) e as coordenadas encontradas pela aplicação (lado direito). Pode-se perceber a diferença entre 2 pontos marcados em vermelho, para este trecho de uma das linhas de estudo.

Figura 12 - Identificação de erros ao obter coordenadas dos pontos



Fonte: Autor

Para fazer a correção, é utilizada a ferramenta *MyMaps* do Google, e a correção é feita de maneira manual. Na Figura 13 pode-se observar as coordenadas dos pontos após o ajuste.

Figura 13 - Correção de erro de coordenadas



Fonte: Autor

Desta forma, pode-se criar uma base de dados com as informações de cada ponto de ônibus, por empresa, com as suas coordenadas corrigidas, conforme Tabela 1.

Tabela 1 - Modelo de arquivo com informações dos pontos de ônibus

EMPRESA	PONTO	LAT	LON	TRECHO
Empresa São Francisco	TP - Clima Bom	-9,5657218	35,7905204	CLIMA BOM - B. GONZAGA
Empresa São Francisco	B.CHEIA	-9,5673387	35,7907064	CLIMA BOM - B. GONZAGA
Empresa São Francisco	CASA DE RAÇOES	-9,5688966	35,7894492	CLIMA BOM - B. GONZAGA
Empresa São Francisco	B.DE CASTRO	-9,5695016	35,7858161	CLIMA BOM - B. GONZAGA
Empresa São Francisco	GALERIA CLIMA BOM	-9,5699283	35,7826125	CLIMA BOM - B. GONZAGA

Por fim, obtém-se a informação de em que ponto de ônibus que cada passageiro embarcou, conforme Tabela 2.

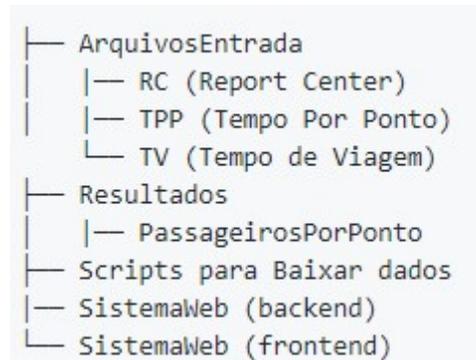
Tabela 2 - Modelo de leitura de dados de Passageiros por ponto

HORARIO	CARTAO	TIPO	LINHA	VALOR	EMPRESA	CARRO	PONTO	LINHA	PARTIDA TERMINAL	SENTIDO
2020-09-28 14:02:41	924265822	CIDADÃO	0057	3.65	SFRA	2406	QUIA MOTOR	Via Colina	2020-09-28 13:40:00	ida
2020-09-28 14:02:54	2546838849	VALE TRANSPORTE	0057	3.65	SFRA	2406	QUIA MOTOR	Via Colina	2020-09-28 13:40:00	ida
2020-09-28 14:03:47	4376689981	VALE TRANSPORTE	0057	3.65	SFRA	2406	IGREJ NOVA VIDA	Via Colina	2020-09-28 13:40:00	ida
2020-09-28 14:06:46	3793245161	VALE TRANSPORTE	0057	3.65	SFRA	2406	PN49101488	Via Colina	2020-09-28 13:40:00	ida
2020-09-28 14:06:54	1007810688	VALE TRANSPORTE	0057	3.65	SFRA	2406	PN49101488	Via Colina	2020-09-28 13:40:00	ida

4.2.DASHBOARD DE ANÁLISES DE DADOS DE PASSAGEIROS

A estrutura do projeto desenvolvido segue a estrutura abordada na Figura 14.

Figura 14 - Estrutura Geral do Projeto



Fonte: Autor

Para iniciar a análise, é necessário utilizar os scripts na pasta “Scripts para baixar dados”, ou baixá-los manualmente. Os arquivos necessários são os do Quadro 2. Em seguida, deve-se rodar o arquivo `main.py` (que vai fazer a leitura dos arquivos relacionados à operação e localização). Em seguida, deve-se rodar o arquivo `open_mrc_detail.py` (que vai fazer a leitura dos dados de Bilhetagem eletrônica) e, o arquivo `consolidate_mrc_detail.py` (que vai fazer a unificação dos dados de GPS). Por fim, será utilizado o arquivo `find_local_pass.py`, que vai realizar o procedimento de encontrar a localização (código do ponto de ônibus) de embarque de cada passageiro registrado pela bilhetagem eletrônica,

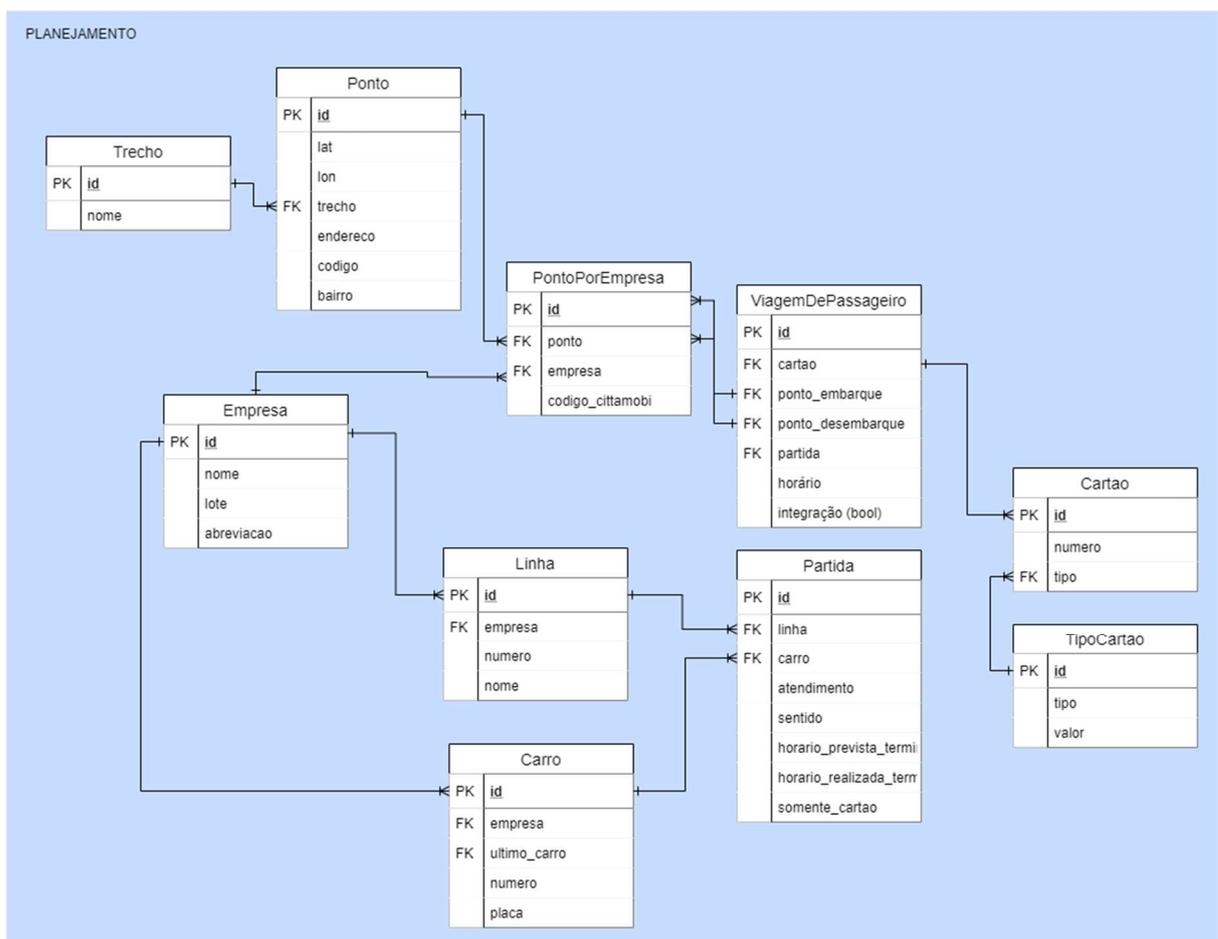
Estas rotinas criam e acessam automaticamente os arquivos na pasta Resultados, onde cada arquivo gerado fica salvo. Estes scripts iniciais servem para o processamento exclusivo da base de dados em questão (combinação do *GoalSystem* com o *Mercury ReportCenter*). Estes subprodutos ficam disponíveis em arquivos binários no formato `.pickle` e são comprimidos no

formato .zip para ocupar menos espaço e permitir uma leitura posterior mais rápida das informações.

Foi montado um banco de dados relacional utilizando o banco de dados *Postgresql* para armazenamento dessas informações e permitir o processamento a partir da linguagem de programação *Python*. Ambas ferramentas são de uso gratuito para não agregar custos aos municípios, podendo ser replicado por terceiros.

Foi criada uma base de dados a partir do diagrama na Figura 15. Assim, o sistema ficou pronto para fazer as análises posteriores, além de preparar as séries de dados a serem disponibilizadas nos gráficos e tabelas da aplicação. Para acessar esta base de dados, foi criado um sistema *WEB*

Figura 15 - Diagrama do banco de dados



Fonte: Autor

A entrada de dados no sistema WEB se dá a partir da deve-se rodar o arquivo main.py na pasta Sistema *WEB* (backend). A entrada de dados se dá a partir de um arquivo em Excel ou .zip (pickle) cujas colunas padrão seguem o modelo na Tabela 2.

Para estudos futuros cuja base de dados seja outra, deve-se alterar este arquivo para fazer a leitura dos dados necessários.

O *backend* (base de dados do sistema *WEB*) está montado utilizando *Python* e *Django*, disponibilizando os dados em API, por meio da Biblioteca *Django Rest Framework*. Para processamento dos dados, é utilizada as bibliotecas *Pandas* e *Numpy*, que são focadas em ciência dos dados e permitem uma execução simples e eficiente, contendo métodos prontos para uso.

O *frontend* (interface) foi criado utilizando Quasar Framework, que utiliza internamente de *VueJS* e *Javascript*. O Quasar permite a criação de sites responsivos (SILVA, 2018; ZEMEL, 2015) de forma bem estruturada e organizada, a partir da criação e utilização de componentes de página criados por terceiros.

Ambas metodologias de desenvolvimento do *frontend* e *backend* utilizam de frameworks de terceiros, com utilização de padrões de projeto (GIRIDHAR, 2013) e Programação Orientada a Objetos (COX, 1986), ajudando a reutilização e manutenção do código.

Os gráficos gerados foram gerados a partir das bibliotecas: Gráficos *Google*, *Plotly* e *ApexCharts*. Ambas são utilizadas no *frontend*, utilizando *JavaScript*.

Para melhor entender o comportamento dos dados, é essencial uma boa ferramenta que permita selecionar parte da base de dados, selecionando subconjuntos dos dados a partir de características tais como: quais linhas; qual faixa horária; quais dias da semana.

Os filtros a serem aplicados nos dados ocorrer de forma intuitiva, por meio de formulários e contam com as seguintes possibilidades:

- a) Filtro por hora da partida do terminal;
- b) Filtro por hora de embarque do passageiro;
- c) Sentido de embarque (ida ou Volta)

O *Dashboard* tem como objetivo facilitar a visualização por meio de gráficos adequados para cada tipo de informação a ser apresentada.

4.3.MODELAGEM DE CLASSIFICAÇÃO

As classificações dos conjuntos de dados, tratam de identificar viagens críticas da seguinte forma:

- a) Viagens críticas com prejuízo na operação: Viagens cujo custo supera a receita. Para isso é necessário a estimativa de custo por quilometragem, além da quilometragem de cada linha.
- b) Viagens Críticas Cheias: Viagens que contém embarques acumulados em 3 trechos seguidos maior do que 60 passageiros.

Foi utilizada a biblioteca *Scikit-Learn* para todas as ferramentas necessárias para desenvolvimento desta etapa (PEDREGOSA et al., 2011).

Inicialmente foi criado um “modelo enganoso” cujo ajuste se dá apenas a classificação de maior intensidade para o conjunto de dados, que servirá de referência a uma classificação completamente tendenciosa. Este modelo é conhecido como classificador enganoso.

Em seguida, será feita a modelagem de cada um dos modelos: SVM, *Naive Bayes*, MLP (redes neurais), *Árvore da Decisão* e *Random Forest*.

Cada modelagem ocorre por via de validação cruzada, com dados estratificados, avaliando diferentes hiper parâmetros seguindo a Quadro 3.

Quadro 3 - Análise de sensibilidade dos Hiperparâmetros

Modelo	Hiperparâmetros
SVM	Kernel: Linear, Polinomial ou RBF Grau polinomial: 2, 3, 4 e 5
Naive Bayes	<i>Smoothing</i> : 100 valores de 10^{-9} a 10^1
MLP	Dimensões das camadas ocultas: 3 camadas (de 50 neurônios cada); 3 camadas (50, 100 e 50 neurônios) e, 1 camada de 100 neurônios
<i>Árvore da Decisão</i> e <i>Random Forest</i>	Profundidade Máxima: 3 a 20 Número mínimo de dados em cada separação: 2, 64, 256, 512, 1000 Número mínimo de dados por folha: 2, 64, 256, 512, 1000

A partir dos melhores modelos, foi feita a estimativa de acurácia 30 vezes, para obter o resultado médio de cada modelo.

Finalmente, foi estimada a árvore da decisão para o melhor conjunto de parâmetros e obter o diagrama representativo da árvore, utilizando a biblioteca *Graphviz*.

4.4. ESTUDO DE CASO

Foram baixados os dados referentes às linhas de estudo. Como metainformações, foram adicionados qual o eixo de transporte de cada linha e o tipo de viagem quanto à bilhetagem (viagens de bilhetagem plena e viagens convencionais). Desta forma, foi feita a análise de consistência das informações nas diferentes bases de dados. Foi selecionado um período de estudo de 28/09/2020 a 10/10/2020.

Foram obtidos os arquivos de passageiros em cada ponto de ônibus (embarque e desembarque, conforme item 4.1)

Os pontos foram agrupados em trechos da linha. Para isso, cada linha terá seu itinerário dividido em trechos. Por fim, foi obtida a quantidade de passageiros por trecho.

Para estimação da matriz OD, foi criado um script para modelar os fluxos de passageiros, aplicando inicialmente a metodologia utilizada inicialmente por CUI (2006), a partir do fluxograma na Figura 5.

Foi alimentada a base de dados de um sistema *WEB*, que permite analisar os dados de cada linha, acompanhar a demanda de passageiros por trecho, com embarque e desembarque, identificação de fluxos OD. O sistema conta com gráficos intuitivos e filtros, de modo a facilitar a visualização e ajudar no processo de tomada de decisão.

A partir dos trechos de cada linha, foi construída uma tabela de sobreposição de trechos da linha.

Com base nos resultados, são avaliadas algumas possibilidades de alterações nas linhas:

- Seccionamento ou ampliação de linhas ou viagens;
- Verificar possibilidades de viagens expressas ou semi-expressas.
- Alterações nos horários de saída;
- Agrupamento de linhas;

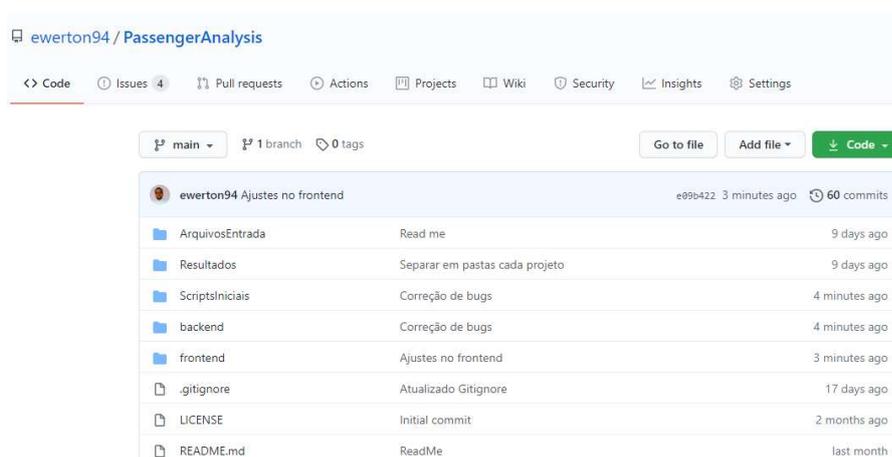
5. RESULTADOS E DISCUSSÃO

Os resultados deste trabalho são: código desenvolvido, ferramentas para leitura e unificação, *Dashboard* para análise dos dados dos passageiros, modelos de classificação e resultados dos estudos de caso. Em cada ponto encontra-se a discussão associada.

5.1. DESENVOLVIMENTO DO CÓDIGO

O código foi hospedado no GitHub. Ele está documentado e pode ser acessado na página do usuário [ewerton94 \(https://github.com/ewerton94/PassengerAnalysis\)](https://github.com/ewerton94/PassengerAnalysis) (Clique para acessar o repositório). O código hospedado é de livre acesso e vai permitir que outras pessoas utilizem o código ou possa fazer contribuições.

Figura 16 - Repositório online



Fonte: captura de tela do *GitHub*

O código foi todo gerenciado pelo *GitHub*, o que facilita o registro de pendências e de etapas de construção do código. A documentação do repositório foi construída visando a reutilização do código por terceiros.

5.2. LEITURA E UNIFICAÇÃO DOS DADOS

Durante a leitura dos arquivos, foram encontradas inconsistências nos dados de gps, apresentando duplicidade de dados e registro indevido de itinerários, registrando viagens das linhas em período ocioso (Deslocamento à garagem, terminal ou viagem vazia). Para isso, foram utilizados os registros de saída de terminal e ponto de retorno para corrigir essas falhas.

Com o cruzamento de informações e unificação da informação, pode-se identificar problemas de cadastro tais como linha cadastrada equivocadamente no sistema, prefixo de veículo que fora atualizado em uma base, e não fora em outra, dificultando a unificação da informação.

Foi necessário o preenchimento de falhas nos dados de GPS, devido a falhas no sistema de georreferenciamento. Mesmo com esse preenchimento, não foi possível identificar o embarque de alguns passageiros. Na Tabela 3 observa-se os percentuais de sucesso das metodologias, quantidades em que se pode encontrar a localização do embarque ou do desembarque dos passageiros, com base no número total de passageiros de cada linha, no período estudado.

Tabela 3 - Percentual de locais de embarques e desembarques encontrados

Cód.	% de bilhetagem eletrônica	% de embarques encontrados	% de desembarques encontrados
051	82,06	58,39	7,15
057	75,53	94,95	20,63
058	88,95	93,30	16,91
065	86,43	87,19	16,98
068	91,96	28,75	3,16
069	78,91	91,00	19,39
108	82,61	71,05	10,59
711	83,40	42,55	3,74
712	84,25	86,14	20,68
715	77,06	68,49	8,21
716	72,19	88,24	16,91

Os problemas no GPS são piores quando os veículos não registram passagem pelo ponto de origem ou destino, que fez com que todos os passageiros destas viagens ficassem sem identificação, pois sem saber o limite de tempo da viagem (início e fim) não é possível filtrar os passageiros neste intervalo. Todas as linhas tiveram passageiros de viagens inteiras que não puderam ter suas localizações de embarque identificadas.

O principal problema encontrado em todas as linhas foi a presença de viagens sem a identificação de horários realizados, que ocorre quando há alta divergência entre os horários programados e realizados pelos veículos, ou quando os veículos que foram programados para a linha, foram substituídos na operação, sem que o sistema fosse atualizado.

Na Figura 17 há um print de um dos relatórios utilizados, nas primeiras viagens da linha 711, no dia 01/10/2020, onde pode-se perceber diversos espaços em branco no horário realizado. O veículo de prefixo 2613 merece destaque, pois, foram planejadas duas viagens para o mesmo veículo, uma saindo de 06:00 (Viagem com duração prevista de 2h) e outra saindo de 06:30. Certamente a primeira previsão ocorreu na operação com outro veículo e o sistema não foi atualizado.

Figura 17 - Problemas no planejamento da operação

Programação		Veículo		Saída				Chegada			
Sentido	Atendimento	Prefixo	Funcionário	Ponto	Prev.	Real.	Dif.	Ponto	Prev.	Real.	Dif.
Ida	UFAL/VIA PAULO OLANDA	2109 - ESF	3371	TP Village	04:30	04:26	-4	sant	05:30	05:35	5
Volta	UFAL/VIA PAULO OLANDA	2109 - ESF	3371	sant	05:30	05:35	5	TP Village	06:50	06:53	3
Ida	UFAL / PONTA VERDE (VIA	2207 - ESF	3758	TP Village	05:00	04:58	-2	sant	06:20	06:05	-15
Volta	UFAL / PONTA VERDE (VIA	2207 - ESF	3758	sant	06:20	06:05	-15	TP Village	07:10	07:18	8
Ida	UFAL/VIA PAULO OLANDA	2105 - ESF	3192	TP Village	05:30	05:44	14	sant	06:50	07:07	17
Volta	UFAL/VIA PAULO OLANDA	2105 - ESF	3192	sant	06:50	07:08	18	TP Village	08:10	08:21	11
Ida	UFAL/VIA PAULO OLANDA	2613 - ESF	3048	R.S.FRANCIS	06:00			sant	07:20		
Volta	UFAL/VIA PAULO OLANDA	2613 - ESF	3048	sant	07:20			TP Village	08:50		
Ida	UFAL/VIA PAULO OLANDA	2613 - ESF	3254	TP Village	06:30	06:29	-1	sant	07:50		
Volta	UFAL/VIA PAULO OLANDA	2613 - ESF	3254	sant	07:50			TP Village	09:30	09:49	19
Ida	UFAL/VIA PAULO OLANDA	2109 - ESF	3371	TP Village	07:00	07:11	11	sant	08:20		
Volta	UFAL/VIA PAULO OLANDA	2109 - ESF	3371	sant	08:20			TP Village	09:10		
Ida	UFAL / PONTA VERDE (VIA	2207 - ESF	3758	TP Village	07:40	07:41	1	sant	08:50	09:18	28
Volta	UFAL / PONTA VERDE (VIA	2207 - ESF	3758	sant	08:50	09:18	28	TP Village	09:50	10:34	44
Ida	UFAL/VIA PAULO OLANDA	2105 - ESF	3192	TP Village	08:20	08:40	20	sant	09:30		
Volta	UFAL/VIA PAULO OLANDA	2105 - ESF	3192	sant	09:30	10:07	37	TP Village	10:30		

Fonte: captura de tela de relatório do *GoolSystem*

Este mesmo problema ocorreu na linha 068, que apresentou o percentual mais baixo de embarques encontrados. Porém, nesta linha, operada por 2 veículos, teve um dos veículos com problema no GPS nos dias analisados, respondendo como Sem Detecção em Ponto de Controle (SDPC), conforme Figura 18.

Figura 18 - Veículo Sem Detecção em Ponto de Controle

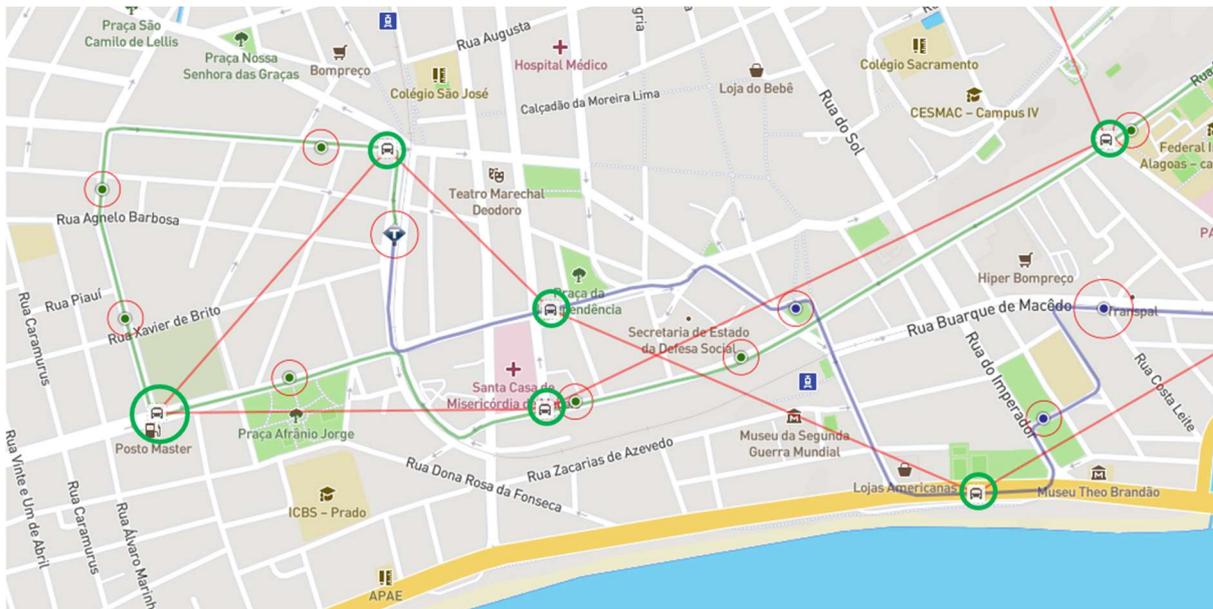
Atendimento		Veículo		Saída				Chegada			
	Prefixo	Funcionário	Ponto	Prev.	Real.	Dif.	Ponto	Prev.	Real.	Dif.	
Colina / Poço (via Sanatório)	2412 - ESF	3101	TP-Colina	06:00	05:56	-4	TS	06:40	SDPC		
Colina / Poço (via Sanatório)	2412 - ESF	3101	TS	06:40	SDPC		TP-Colina	07:30			
Colina CENTRO/EMBARQUE	2432 - ESF	0724	TP-Colina	07:00	06:59	-1	TS	07:55	08:01	6	
Colina CENTRO/EMBARQUE	2432 - ESF	0724	TS	07:55	08:01	6	TP-Colina	08:40	09:02	22	
Colina / Poço (via Sanatório)	2412 - ESF	3101	TP-Colina	08:10	08:11	1	TS	08:47			
Colina / Poço (via Sanatório)	2412 - ESF	3101	TS	08:47			TP-Colina	10:25	SDPC		
Colina CENTRO/EMBARQUE	2432 - ESF	0724	TP-Colina	09:20	09:30	10	TS	09:56	10:27	31	
Colina CENTRO/EMBARQUE	2432 - ESF	0724	TS	09:56	10:27	31	TP-Colina	11:00	11:28	28	
Colina / Poço (via Sanatório)	2412 - ESF	3101	TP-Colina	10:30	10:29	-1	TS	11:10	11:23	13	
Colina / Poço (via Sanatório)	2412 - ESF	3101	TS	11:10	11:23	13	TP-Colina	12:10	12:20	10	
Colina CENTRO/EMBARQUE	2432 - ESF	0724	TP-Colina	11:40	11:38	-2	TS	12:20	12:27	7	
Colina CENTRO/EMBARQUE	2432 - ESF	0724	TS	12:20	12:27	7	TP-Colina	13:10	13:23	13	
Colina / Poço (via Sanatório)	2412 - ESF	3101	TP-Colina	12:50	SDPC		TS	13:10	SDPC		
Colina / Poço (via Sanatório)	2412 - ESF	3101	TS	13:10	SDPC		TP-Colina	14:20	14:30	10	
Colina CENTRO/EMBARQUE	2432 - ESF	3978	TP-Colina	14:00	14:03	3	TS	14:45	14:49	4	
Colina CENTRO/EMBARQUE	2432 - ESF	3978	TS	14:45	14:49	4	TP-Colina	15:40	15:48	8	

Fonte: captura de tela de relatório do *GoolSystem*

Foi analisado o registro do veículo ponto a ponto, e verificou-se que o mesmo apresentou problemas de registro de TPP conforme Figura 19. Os círculos verdes são os pontos com coordenadas registradas, mas os círculos vermelhos são os pontos registrados (pontos de controle) onde há falhas, o que inclui o ponto de retorno na maioria das viagens deste veículo,

prejudicando o registro de intervalo de viagem e, conseqüentemente, prejudicando a identificação de embarque dos passageiros desta viagem.

Figura 19 - Registro de coordenadas por ponto



Fonte: captura de tela de sistema “Replay” do GoolSystem

Estes problemas relacionados a GPS prejudicam a comunicação da operação para os usuários do *Cittamobi*, utilizado pelo cidadão para obter previsões de horários das linhas de Maceió. Desta forma, identificar os problemas neste sistema, além de tornar a análise mais robusta, fornece um melhor sistema de informação para a população.

A divergência de informações nas diversas fontes de dados é um problema presente em sistemas isolados e sem comunicação entre si. Dentre os problemas encontrados estão a falta de atualização de itinerários e a troca de prefixos dos veículos.

A falta de atualização nas rotas pré-definidas no sistema GPS ocasionou a atribuição de passageiros a localidades que não são reais, devido à interpolação programada com base nos pontos atribuídos no sistema para cada itinerário. Essa interpolação é necessária para corrigir erros de falhas no GPS, mas ocasionou este erro eventual. Um exemplo deste problema foi a identificação de passageiros da linha 711 – Ufal / Ponta Verde nos pontos da Universidade Federal de Alagoas, que estava fechada no período analisado.

Alguns veículos possuem prefixos diferentes no sistema de GPS e no de Bilhetagem eletrônica, o que ocasionou erro em todos os passageiros de alguns veículos. As divergências se deram a partir da atualização de prefixos de veículos a fim de cumprir determinações do edital de

licitação do TPC de Maceió, porém os prefixos foram atualizados em um sistema e em outro não, gerando divergências.

Para resolver este problema, foi necessário obter os metadados dos veículos no sistema da SMTT, e fazer a relação de prefixos diferentes para um mesmo veículo, a partir de suas respectivas placas. Desta forma, o sistema pode prosseguir com as análises.

Na linha 715 – Colina / Ponta Verde, que é dividida em 2 itinerários, cujo cadastro no *GoolSystem* é: 715 – Rio Novo / Ponta Verde e 7151 – Rosane Collor / Ponta Verde. Já no sistema de BE a linha possui apenas um prefixo: 715, sendo necessário preparar o algoritmo para fazer esta substituição e dificultando a análise de passageiros das linhas separadamente.

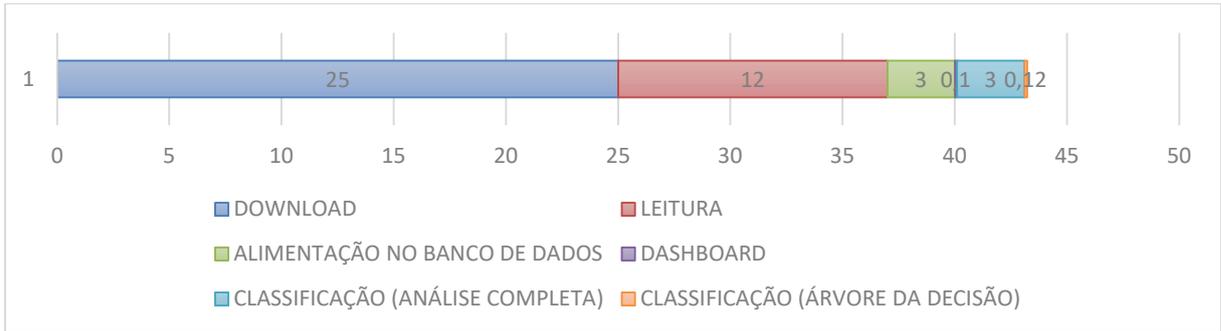
Outro problema semelhante foi com a linha 2058 – Fernão Velho / Colina, onde não foi encontrada a posição geográfica de nenhum passageiro, embora encontra-se a validação de passageiros. A linha 2058 é um seccionamento da linha 058 – Fernão Velho / Centro (Via Colina) e só opera aos domingos, em substituição desta. No *GoolSystem* a linha 2058 é um itinerário da linha 058, ao invés de ser tratada como uma linha isolada, o que prejudicou o cruzamento de dados.

Diante dessas divergências, fica evidente a necessidade de unificação das informações em uma base de dados a ser utilizada pelos diversos sistemas. Em não sendo possível a unificação dos dados em uma base principal e única, vale ressaltar a importância do cuidado ao manter a informação atualizada em todos os sistemas para permitir o uso de diversas fontes para promover análises consistentes a partir do cruzamento dessas informações.

Devido à grande demanda de processamento de dados envolvida nesta metodologia, não é possível a obtenção dos dados em tempo real, visto que é necessário baixar os relatórios a partir do dia seguinte da operação, na forma de planilhas, além de executar a leitura de arquivos e não de bases de dados em tempo real, além de precisar do próximo embarque no dia para obter o desembarque.

A Figura 20 apresenta o tempo gasto por etapa da etapa de obtenção dos dados. Destaca-se o tempo gasto para baixar os relatórios e lê-los, pois é necessário acessar a internet e abrir arquivos com alta demanda de processamento. Uma solução alternativa seria a obtenção os dados a partir de API, em formatos que facilitem a leitura de forma otimizada, direto da internet, sem a necessidade de baixar arquivos, enaltecendo a necessidade de investimento e incentivo à sistemas unificados e API por parte dos órgãos públicos, ganhando agilidade ao cruzar informações de diversas fontes.

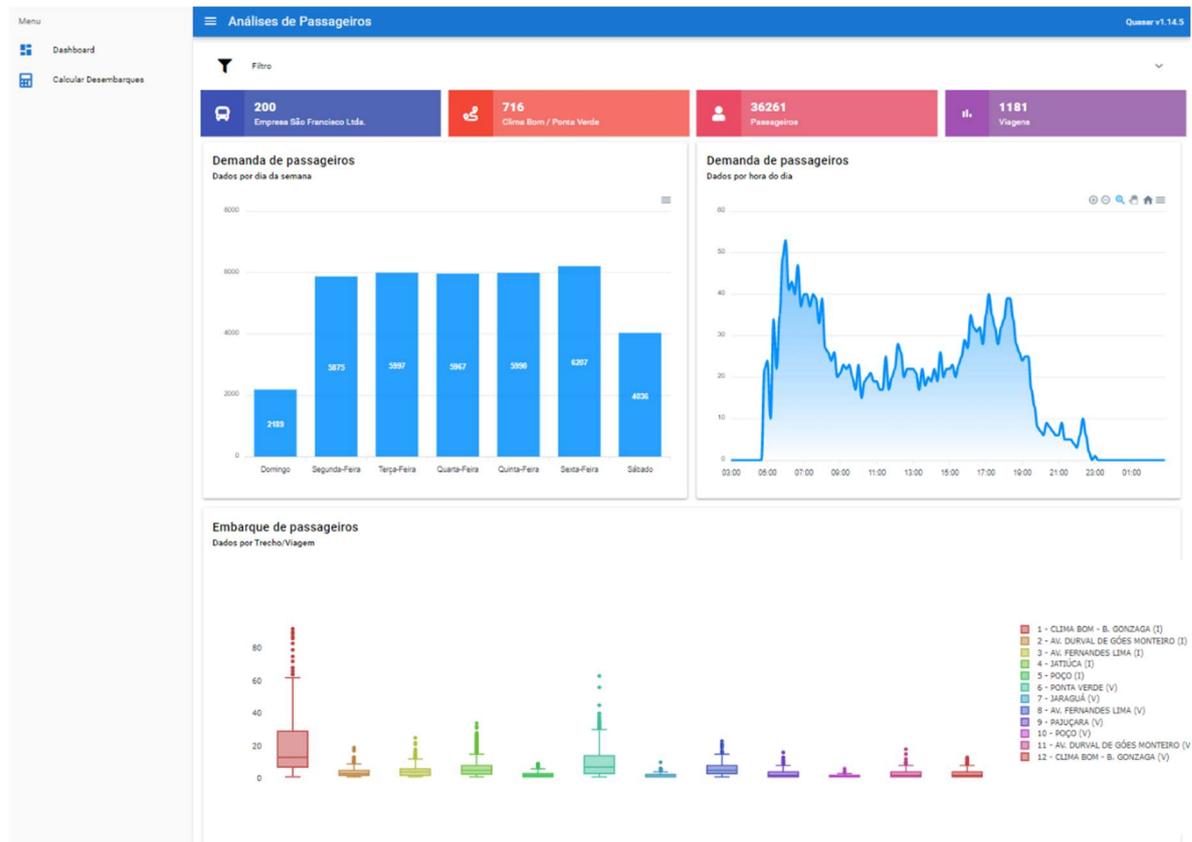
Figura 20- Tempo de execução por etapa (em minutos)



5.3.DASHBOARD DE ANÁLISES DE DADOS DE PASSAGEIROS

A Figura 21 traz uma captura de tela da base do sistema web em desenvolvimento. A visualização no formato de dashboard permite decisões direcionadas. Nesse sistema será possível emitir gráficos nos formatos .png e .svg em alta qualidade. Como exemplo ilustrado, apenas a linha 716 com sua respectiva demanda de passageiros por dia da semana, no período analisado.

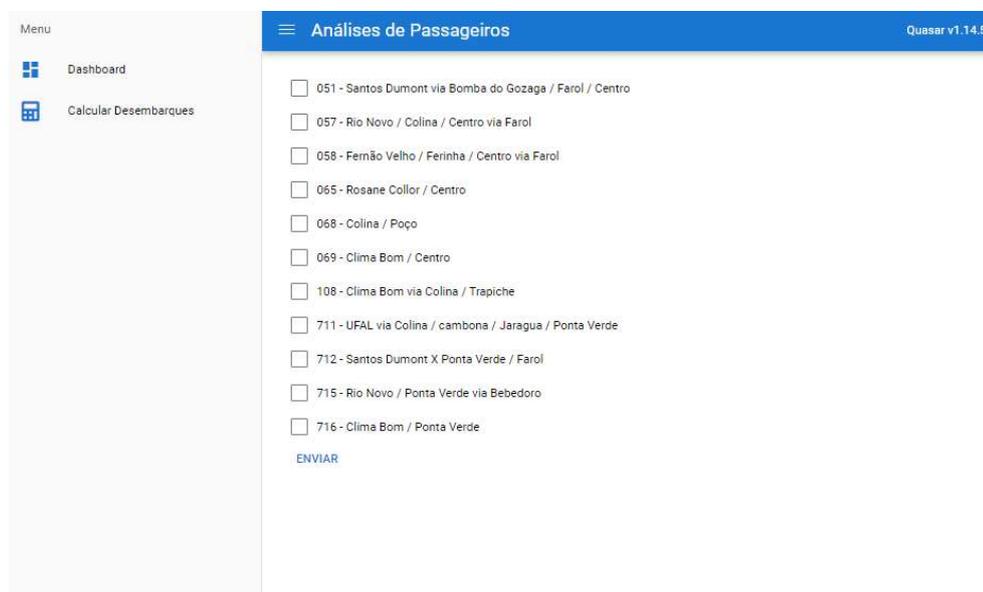
Figura 21 - Captura de tela do *Dashboard* em funcionamento



Fonte: Autor

Esta página conta com os dados alimentados a partir da metodologia, e seus respectivos gráficos interativos. As linhas a serem estudadas são escolhidas na página principal conforme Figura 22.

Figura 22 - página de seleção de linhas a serem estudadas



Fonte: Autor

A possibilidade de se ter filtros nos dados permitiu as análises do estudo de caso.

Os códigos do sistema *WEB* utilizam de diversos padrões de desenvolvimento de código tanto no *frontend* quanto no *backend*. A Figura 23 Apresenta o código página principal do *Dashboard*, com os principais gráficos. Nota-se a reutilização de código a partir de componentes, ajudando na reutilização e manutenção do código.

Figura 23 - parte do código da página de *Dashboard*

```

<template>
  <div >
    <q-page class="q-pa-sm">
      <q-card class="bg-transparent no-shadow no-border">
        <q-card-section class="q-pa-none">
          <Filtro/>
        </q-card-section>
      </q-card>
      <div :key="geralKey">
        <q-card class="bg-transparent no-shadow no-border">
          <q-card-section class="q-pa-none">
            <div class="row q-col-gutter-sm">
              <CardResumo itemColor="#5064b5" iconColor="#3e51b5" iconName="fas fa-bus" :title="linha.lote" :subtitle="linha.empresa"/>
              <CardResumo itemColor="#f37169" iconColor="#f34636" iconName="fas fa-route" :title="linha.numero" :subtitle="linha.linha"/>
              <CardResumo itemColor="#ea6a7f" iconColor="#ea4b64" iconName="fas fa-user" :title="linha.passageiros" subtitle="Passageiros"/>
              <CardResumo itemColor="#a270b1" iconColor="#9f52b1" iconName="bar_chart" :title="linha.viagens" subtitle="Viagens"/>
            </div>
          </q-card-section>
        </q-card>
        <div class="row">
          <div class="col-12 col-md-6">
            <GráficoGeral
              class="my-card q-ma-sm"
              tipoGráfico = "DadosPorDiaDaSemana"
              modeloGráfico = "bar"
              title = "Demanda de passageiros"
              subtitle = "Dados por dia da semana"
              :timeout="1"
            >>/GráficoGeral
          </div>
          <div class="col-12 col-md-6">
            <GráficoGeral
              class="my-card col-12 col-md-6 q-ma-sm"

```

Fonte: Autor

A Figura 24 apresenta o código do Gráfico por hora do dia, modularizado em uma classe *DadosPorHoraDoDia*, que é o padrão de classe para gráficos no *backend*. Esta estrutura de *frontend* e *backend* permite a criação de novos gráficos de forma ágil.

Figura 24 - Código do gráfico por hora do dia

```

class DadosPorHoraDoDia(BaseData):
    columns = ['id', 'horario']

    def calcular(self):
        df = self.get_df()
        df.index = pd.to_datetime(df['horario'])
        df = df.groupby(pd.Grouper(freq='10Min')).count()
        df['time'] = (pd.to_datetime(df.index) + pd.Timedelta(hours=-3)).time
        df = df.groupby('time').mean()[['id']]
        df.index = ('2010-01-01'+df.index.to_series().astype(str)).apply(pd.Timestamp)
        df.columns = ['Passageiros']
        df = self.ordenar_index(df, lambda x: x)
        return self.grafico_de_area(df.astype(int))

```

Fonte: Autor

5.4. CLASSIFICAÇÃO

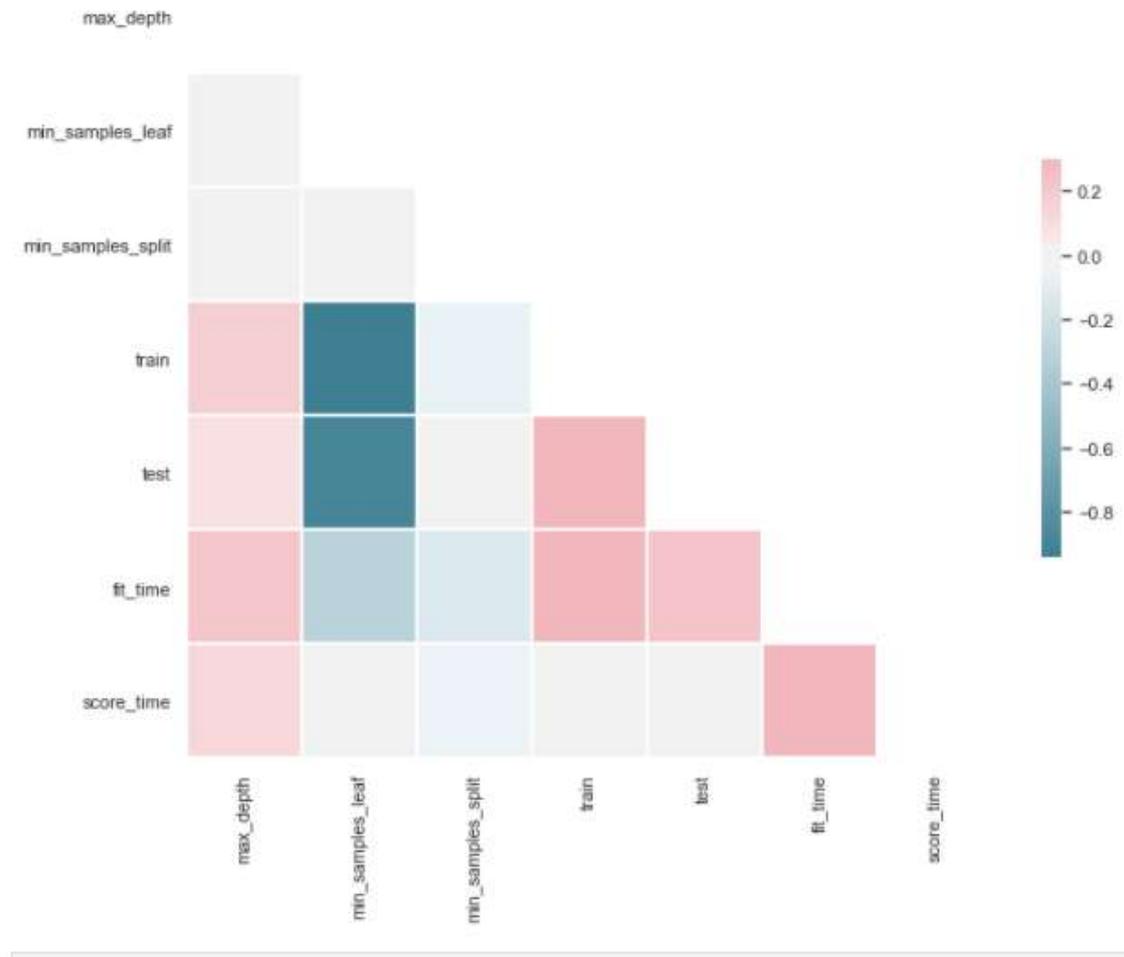
O modelo de árvore da decisão se mostrou eficaz, dentro da média do intervalo de confiança do melhor modelo (Média \pm 2 desvios padrão), conforme Tabela 4.

Tabela 4 - Acurácia dos ajustes dos modelos

	Modelo	Treino		Teste		Parâmetros
		Média	Desvio Padrão	Média	Desvio Padrão	
Viagens Vazias nos dias úteis	Dummy	65,06	0,01	65,06	0,07	-
	Naive Bayes	74,16	0,35	72,99	3,62	Smoothing: $2,85 * 10^{-5}$
	MLP (Redes neurais)	80,55	0,99	78,52	4,07	Dimensões das camadas ocultas: 3 camadas (de 50 neurônios cada) Ativação: tanh
	SVM	82,51	0,48	78,45	5,84	RBF
	Random Forest	84,67	0,27	80,95	3,89	Profundidade Máxima: 11 Nº mínimo/separação: 32 Nº mínimo/folha: 128
	Árvore da Decisão	82,43	0,61	80,24	3,35	Profundidade Máxima: 11 Nº mínimo/separação: 128 Nº mínimo/folha: 32
	Modelo	Treino		Teste		Parâmetros
		Média	Desv. P.	Média	Desv. P.	
Viagens Vazias nos Sábados	Dummy	72,30	0,02	72,30	0,20	-
	Naive Bayes	78,43	0,89	77,57	5,25	Smoothing: $8,11 * 10^{-7}$
	MLP (Redes neurais)	80,03	3,35	76,25	3,60	Dimensões das camadas ocultas: 3 camadas (de 50 neurônios cada) Ativação: tanh
	SVM	79,32	2,42	77,38	4,40	Linear
	Random Forest	81,11	0,58	80,18	4,38	Profundidade Máxima: 15 Nº mínimo/separação: 32 Nº mínimo/folha: 32
	Árvore da Decisão	81,21	0,72	80,62	6,82	Profundidade Máxima: 5 Nº mínimo/separação: 32 Nº mínimo/folha: 128
	Modelo	Treino		Teste		Parâmetros
		Média	Desv. P.	Média	Desv. P.	
Viagens Vazias nos Domingos	Dummy	50,08	0,09	49,24	0,25	-
	Naive Bayes	62,95	1,56	58,76	12,51	Smoothing: $4,32 * 10^{-6}$
	MLP (Redes neurais)	62,59	3,71	57,74	11,31	Dimensões das camadas ocultas: 1 camada de 100 neurônios Ativação: tanh
	SVM	63,72	1,71	51,12	13,00	Linear
	Random Forest	69,46	1,62	57,60	15,47	Profundidade Máxima: 5 Nº mínimo/separação: 32 Nº mínimo/folha: 2
	Árvore da Decisão	71,01	1,85	58,07	13,35	Profundidade Máxima: 5 Nº mínimo/separação: 32 Nº mínimo/folha: 2
	modelo	Treino		Teste		Parâmetros
		Média	Desv. P.	Média	Desv. P.	
Viagens cheias	Dummy	91,86	0,01	91,86	0,08	-
	Naive Bayes	94,01	0,38	93,35	2,44	Smoothing: $5,33 * 10^{-7}$
	MLP (Redes neurais)	94,11	0,54	93,95	2,22	Dimensões das camadas ocultas: 3 camadas (de 50 neurônios cada) Ativação: tanh
	SVM	94,61	0,25	93,93	2,91	Polinômio de grau 3
	Random Forest	96,21	0,18	94,17	2,41	Profundidade Máxima: 17 Nº mínimo/separação: 2 Nº mínimo/folha: 32
	Árvore da Decisão	95,69	0,17	94,90	1,73	Profundidade Máxima: 7 Nº mínimo/separação: 2 Nº mínimo/folha: 128

A análise de sensibilidade dos parâmetros permitiu que fosse escolhido os melhores intervalos possíveis dos valores de cada parâmetro a fim de encontrar o ajuste ideal. O intervalo dos parâmetros foi definido a partir do gráfico de correlação na Figura 25.

Figura 25 - Correlação entre parâmetros na Árvore da Decisão



Fonte: Autor

Para este exemplo, conclui-se que o parâmetro *max_depth* se ajusta melhor para valores mais altos e o *min_samples_leaf* para valores mais baixos, por isso foi colocado valores até 20 de profundidade máxima e até 1000 para o número mínimo de amostras por folha.

5.5. ESTUDO DE CASO

A Figura 26 apresenta a demanda de passageiros das linhas em estudo ao longo do dia, no qual podemos ver basicamente 2 períodos de pico, um no início da manhã (05:30 às 07:30) e um no fim da tarde (16:30 às 19:00).

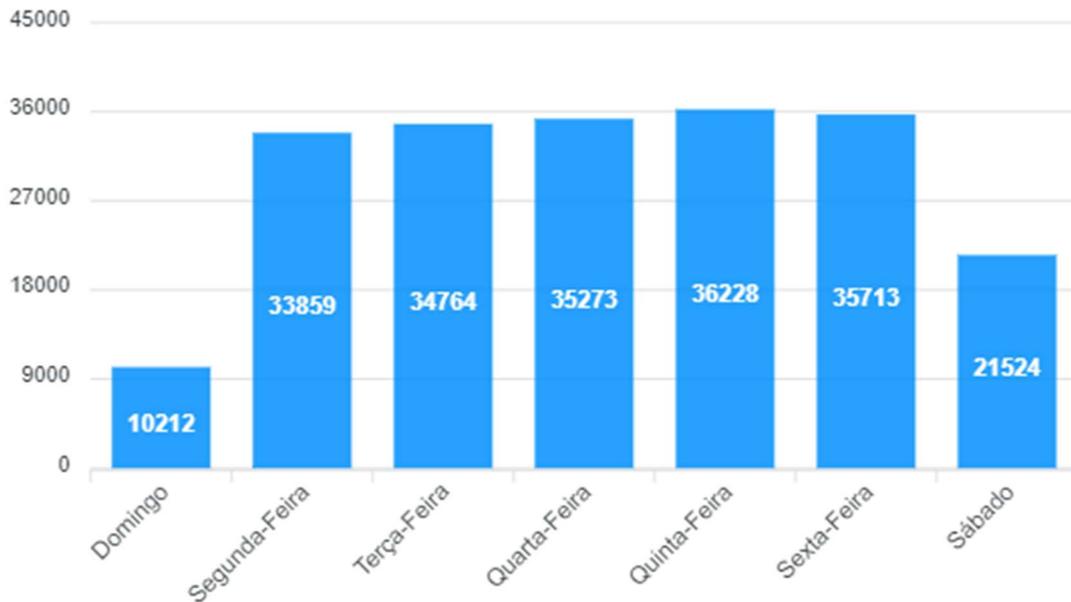
Figura 26 - Demanda de passageiros ao longo do dia



Fonte: Autor

Analisando a demanda por dia da semana (Figura 27), pode-se corroborar com a operação dividida entre dias úteis, sábado e domingo. A demanda de segunda a sexta (dias úteis) apresenta valores próximos.

Figura 27 - Demanda de passageiros por dia da semana



Fonte: Autor

A Tabela 5 mostra a sobreposição das linhas em estudo, quanto aos trechos atendidos, na qual podemos perceber 67% de sobreposição entre linhas ligando os bairros de origem ao Centro de Maceió, e de 71% entre as linhas entre os bairros de origem ao litoral Central de Maceió, utilizando o eixo Bebedouro.

Tabela 5 - Sobreposições dos trechos das linhas

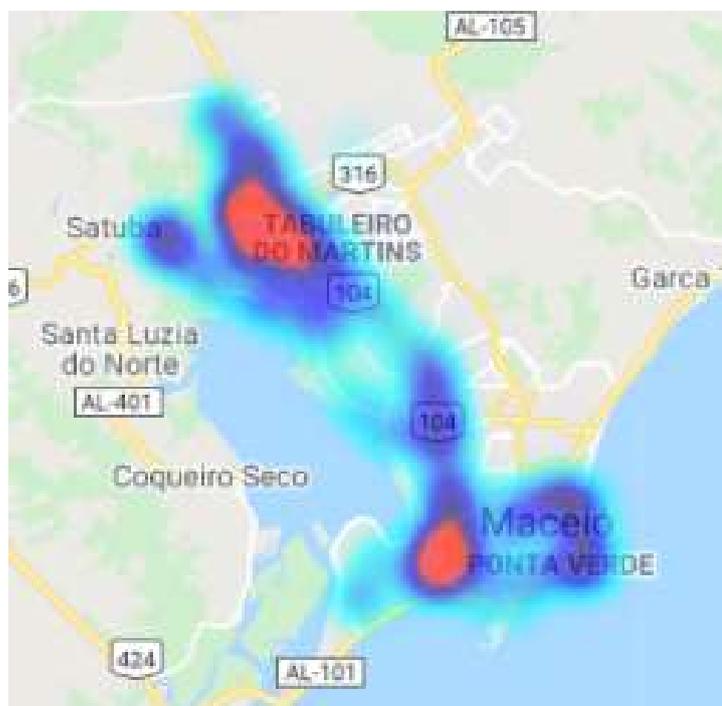
	051	057	058	065	068	069	108	711	712	715
051										
057	44%									
058	40%	86%								
065	44%	71%	63%							
068	33%	25%	23%	25%						
069	57%	67%	57%	67%	18%					
108	16%	24%	29%	24%	41%	19%				
711	25%	14%	14%	14%	35%	10%	32%			
712	38%	13%	13%	13%	18%	15%	4%	38%		
715	19%	20%	25%	20%	35%	10%	38%	71%	32%	
716	50%	27%	25%	27%	21%	33%	10%	37%	58%	37%

Vale ressaltar que esses valores são para trechos em comum e não quilometragem e que todas as linhas estudadas compartilham 100% de um trecho intermediário na Av. Fernandes Lima, Fruto do desvio do Bairro do Mutange. Por este motivo, é essencial o estudo de um corredor troncal passando pela Avenida Fernandes Lima e investimento de um corredor de alta capacidade e priorização do TPC, pois é um problema prestes a virar caótico.

Quanto à demanda de passageiros por localização geográfica, o mapa na Figura 28 indica os pontos com maiores intensidades de embarque. Como pode-se perceber, há uma grande demanda nos 2 polos da cidade, o que dá a ideia de movimento pendular, porém, neste caso não se pode inferir sobre a demanda de destino destes passageiros. O movimento pendular é um dos problemas mais enfrentados pelas cidades (MOURA; CASTELLO BRANCO; FIRKOWSKI, 2005; NASCIMENTO NETO, 2020), que ocorre quando as pessoas partem de uma mesma região (residencial) para uma mesma região (empresarial) no período da manhã e fazem o sentido contrário no fim da tarde, criando uma disputa em busca do equilíbrio do sistema: de

um lado a população necessita de mais viagens devido à superlotação nos horários de pico e do outro as empresas não podem aumentar tanto o número de viagens devido ao prejuízo gerado com as viagens nos sentidos contrários os deslocamentos das pessoas.

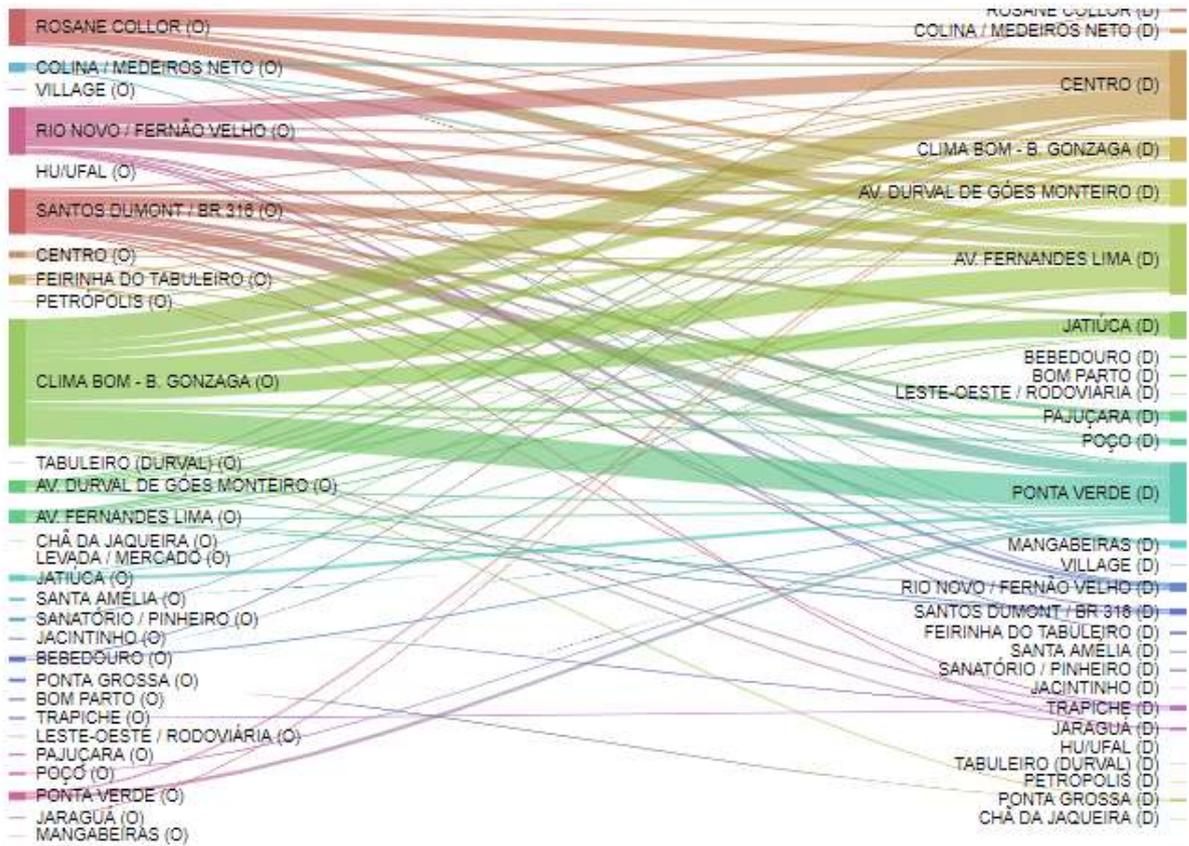
Figura 28 - mapa de calor de intensidade de embarque de passageiros



Fonte: Autor

Já a partir da Figura 29, constata-se que grande parte dos passageiros com origem nos bairros da parte alta (Rosanne Collor, Rio Novo, Santos Dumont, Clima Bom) tem como destino os bairros da Ponta Verde, Jatiúca e Centro, o que corrobora com a premissa de movimento pendular destas linhas, baixos índices de renovação e consequente perda de capital em viagens cuja receita não é suficiente para atender os gastos de locomoção.

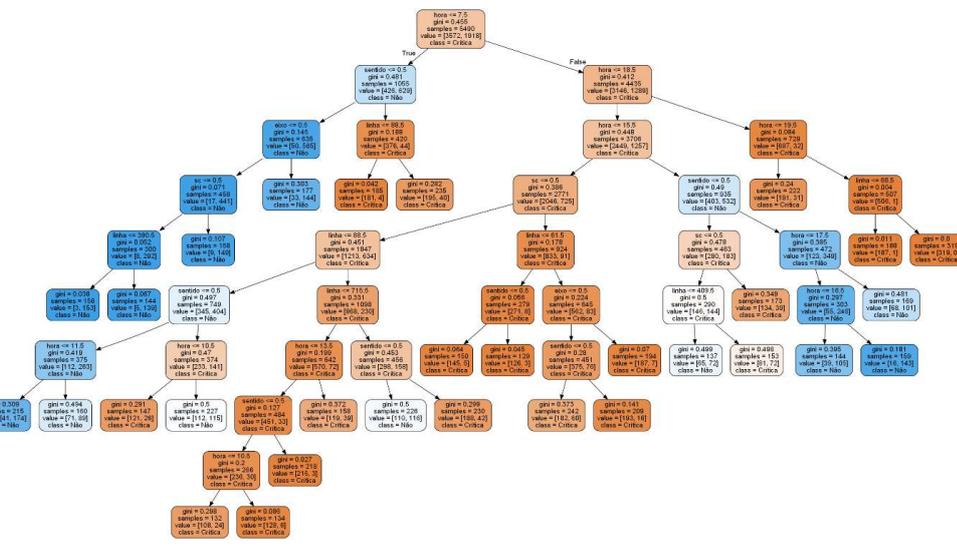
Figura 29 - Origem-Destino por trecho das linhas



Fonte: Autor

Foi constatado o custo financeiro de 6,23 R\$/Km na Empresa São Francisco a partir de relatório de receita do Sinturb-MAC. A partir disso foi encontrado o esquema de classificação de viagens que representam prejuízo financeiro, que pode ser constatado na Figura 30.

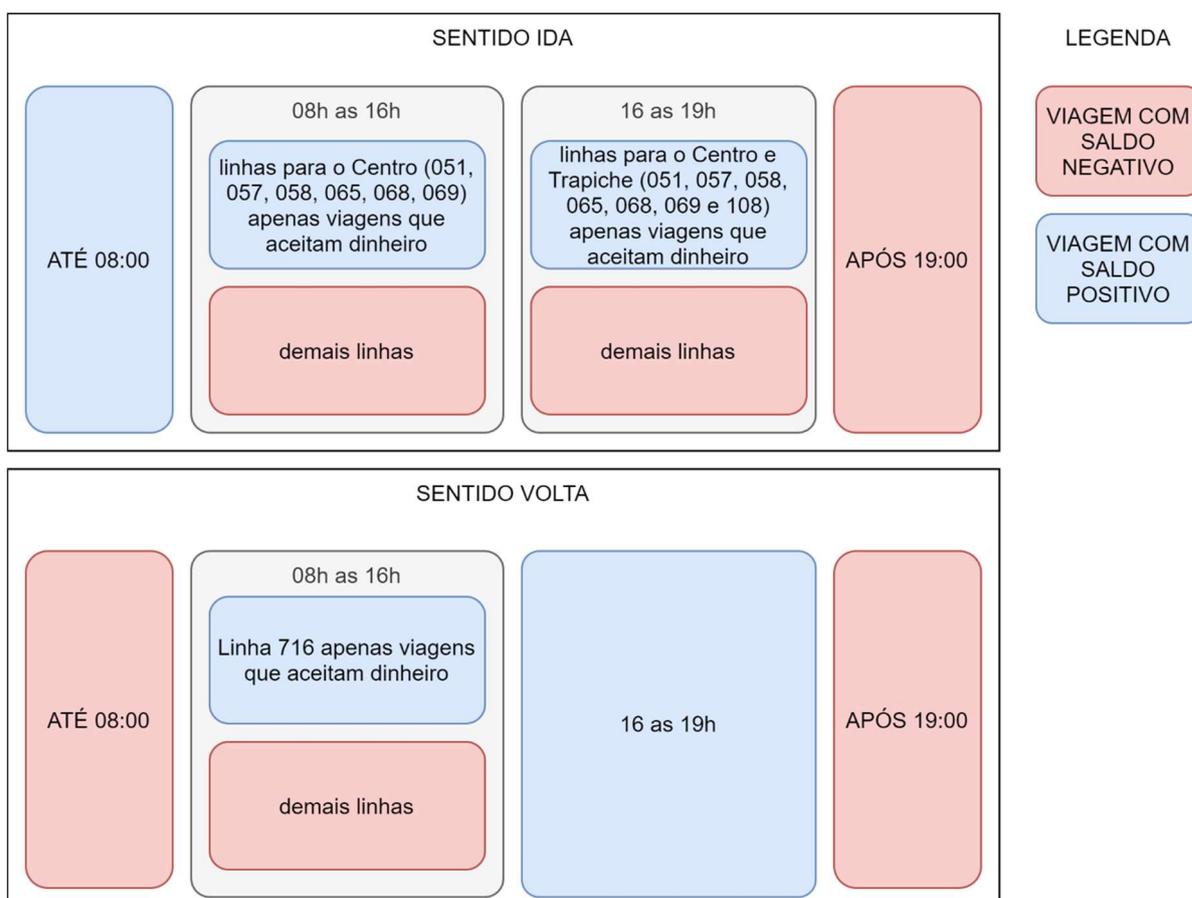
Figura 30 - Árvore da decisão para viagens com prejuízo financeiro



Fonte: Autor

Esta visualização padrão do programa deve ser interpretada, pois a mesma apresenta os resultados de forma numérica. A Figura 31 representa a mesma informação esquematizada de forma mais intuitiva. Pode-se perceber que as viagens noturnas são um grande gargalo na operação do sistema em tempo de pandemia, com escolas e faculdades fechadas. Por isso seria interessante uma operação otimizada no período noturno e aproveitar mais do Terminal Integrado da Colina dos Eucaliptos para mitigar os prejuízos no período da noite com uma operação mais enxuta.

Figura 31 - Esquema de classificação de viagens de todas as linhas estudadas (Dias úteis)



Fonte: Autor

O movimento pendular é um fator importante em todas as linhas estudadas, nas quais há um movimento no sentido de ida no pico da manhã e no sentido da volta no período do pico da tarde. Uma possível alternativa a ser estudada a fundo é a operação de algumas viagens no período da manhã que fiquem em um ponto mais próximo aos destinos das linhas aguardando o retorno no pico da tarde. Em contrapartida a operação e distribuição de pessoal para estas viagens e aumento de período ocioso é um ponto negativo desta abordagem e por isso deve ser feita uma análise mais apurada da possibilidade.

A Figura 32 identifica as intensidades de embarque por coordenada geográfica, caracterizando uma alta demanda de passageiros nos dois polos da cidade (nas origens das linhas no pico da manhã e nos destinos das linhas no pico da tarde).

Figura 32 - Mapa de calor de embarque de passageiros nos horários de pico



Fonte: Autor

Uma característica importante é o predomínio de viagens críticas entre as com bilhetagem plena. Tal resultado, por um lado, pode direcionar para uma validação da escolha atual razoável quanto às viagens escolhidas para serem de bilhetagem plena, visto que são em horários que há pouco número de passageiros.

Por outro lado, deve-se ter uma atenção ao tempo de espera da população que não possui cartão, visto que a maioria das linhas possui percentual de passageiros com cartão acima de 10% e chegando a 28%. Visto isso, é importante a intensificação de campanhas de divulgação para utilização de bilhetagem eletrônica, além de investimentos por parte da prefeitura e empresas visando facilitar o pagamento por outros meios que não sejam dinheiro. Muitas cidades apontam para cartões de crédito/débito e pagamentos via *QR Code* (BAZANI; MOREIRA, 2020; MARQUES, 2020; SOUZA, 2020).

Outro período em destaque é o período de 8h às 16h, onde podem ser aproveitadas viagens de duplo itinerário, ligando os bairros ao Centro da Cidade e retornarem passando pela Ponta Verde e assumindo o itinerário da linha 716 no sentido de volta.

Quanto às viagens no período após às 19h, foi feita uma análise usando como limite de classificação as viagens mais críticas, com saldo abaixo do percentil 25 para este conjunto de dados, o que representa um prejuízo de R\$ 72 por viagem.

De acordo com a Figura 33, as linhas 051, 711, 712 e 715 merecem estudos de ajustes de horários e são responsáveis pelas viagens com maior prejuízo, com ênfase à linha 715, que é a única com prejuízo nos dois sentidos. Vale salientar que os itinerários das linhas 711 são sobrepostos em maioria do percurso (67%, conforme Tabela 5) e que ambas passam por um terminal integrado antes de subdividir seus itinerários, o que é um fator positivo na otimização destas linhas.

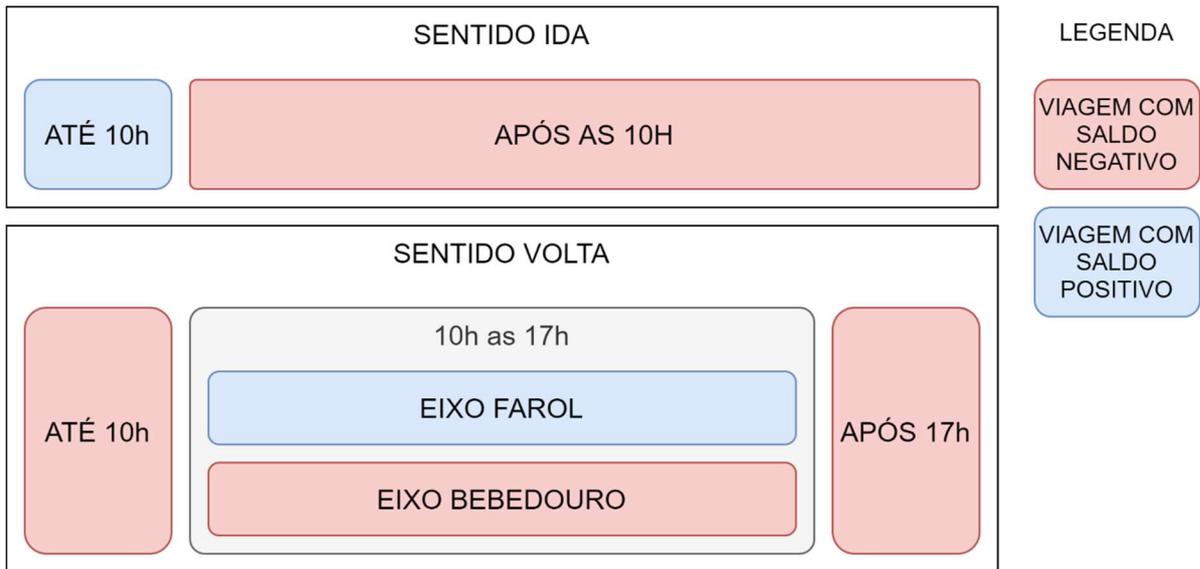
Figura 33 - Classificação de viagens após às 19h quanto ao prejuízo



Fonte: Autor

A operação aos sábados (Figura 34) possui saldo positivo apenas no período da manhã no sentido de ida e no sentido de volta das 10h às 17h, apenas para as linhas do eixo Farol. Desta forma, a operação das linhas do eixo Bebedouro precisam ser otimizadas aos Sábados, aproveitando-se mais do TI da Colina.

Figura 34 - Esquema de classificação de viagens de todas as linhas estudadas (Sábado)

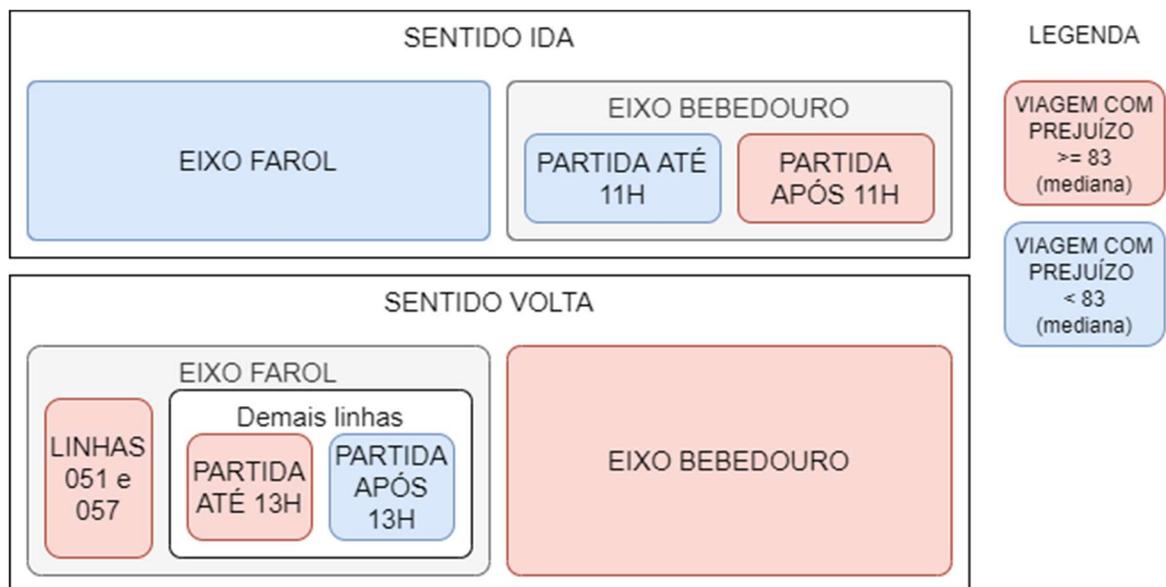


Fonte: Autor

Aos domingos, a classificação encontrada com melhor ajuste foi de viagem crítica para todas as viagens. Consequentemente, foi feita a análise para a mediana do conjunto de dados de saldo de viagem aos domingos, com saldo de R\$ 83,42 de prejuízo. A utilização da mediana indica que metade das viagens aos domingos têm prejuízo maior que R\$ 83,42.

Ao classificar as viagens pela mediana, de forma semelhante à operação aos sábados, as viagens mais críticas são as que transitam pelo eixo do Bebedouro, cujas soluções partem do mesmo princípio (Figura 35).

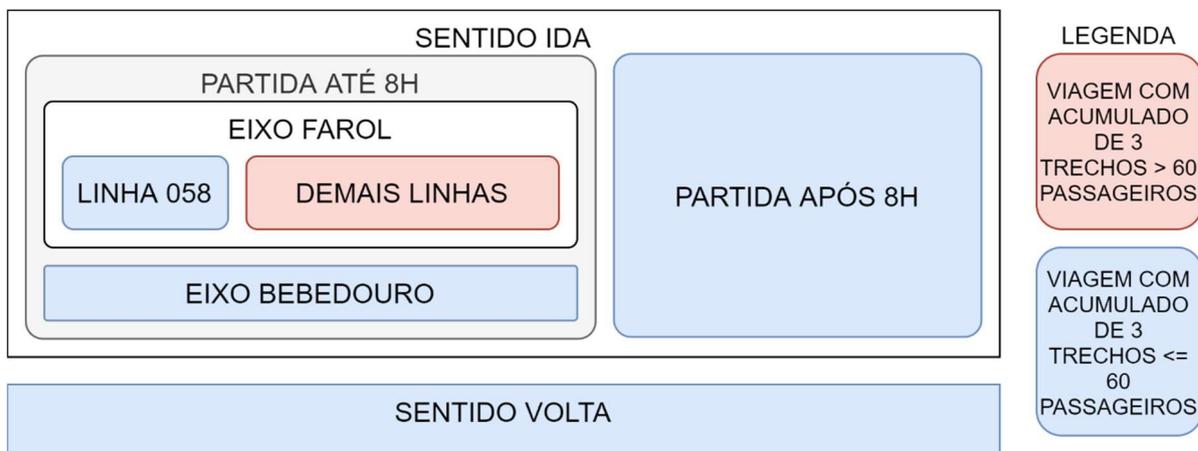
Figura 35 - Esquema de classificação de viagens de todas as linhas estudadas (Domingo)



Fonte: Autor

Para a análise de viagens cheias (Figura 36), o padrão encontrado foi de viagens críticas no pico da manhã para as linhas do eixo Farol, com exceção à 058.

Figura 36 - Esquema de classificação de viagens cheias de todas as linhas estudadas



Fonte: Autor

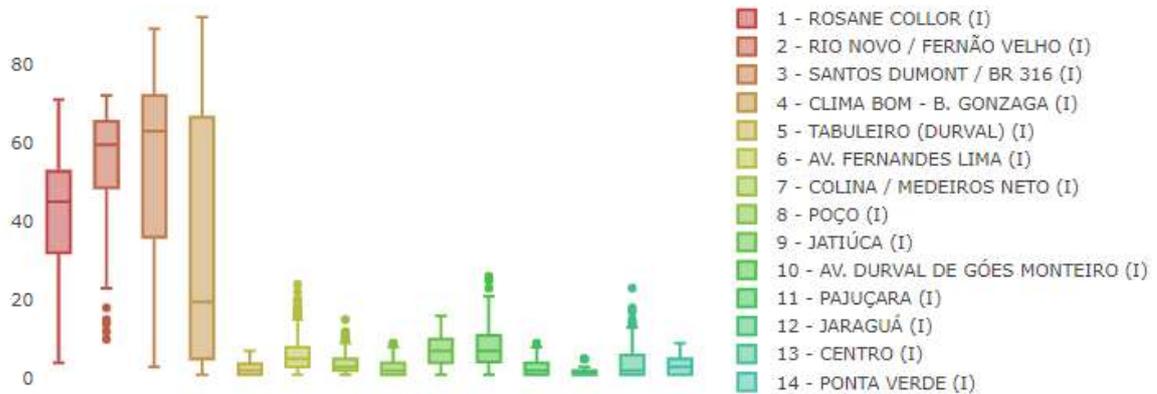
Percebe-se que não há classificação de viagens no período da tarde como viagens cheias. Há a possibilidade deste problema não relatar a realidade com base em algumas possibilidades:

- Alguns passageiros podem ser identificados no sentido ida, por causa da proximidade entre as paradas no ponto de retorno;
- Algumas linhas, como as linhas da Ponta Verde fazem um trajeto Circular na parte baixa da Cidade, local de alta demanda de passageiros conforme Figura 11. Como se tratam de muitos trechos na parte baixa, o acumulado de 3 pode diluir estes embarques e não haver muitos desembarques, de acordo com o movimento pendular das linhas.

Para análise de viagens cheias, é recomendado uma metodologia que permita uma estimativa assertiva do desembarque dos passageiros e que consiga assim estimar com alta precisão a lotação de passageiros por trecho.

Visando soluções para resolver viagens cheias seguindo a classificação da Figura 36, foi obtido o embarque por trecho destas linhas apresentado na Figura 37. Pode-se afirmar que a grande demanda destas viagens parte dos trechos iniciais das linhas e possuem pouca demanda no decorrer das linhas. Dentre estas, apenas as linhas 057 e 065 atendem ao TI da Colina e merecem atenção quanto ao embarque/desembarque neste terminal.

Figura 37 - Demanda de passageiro por trecho por viagem no pico da manhã

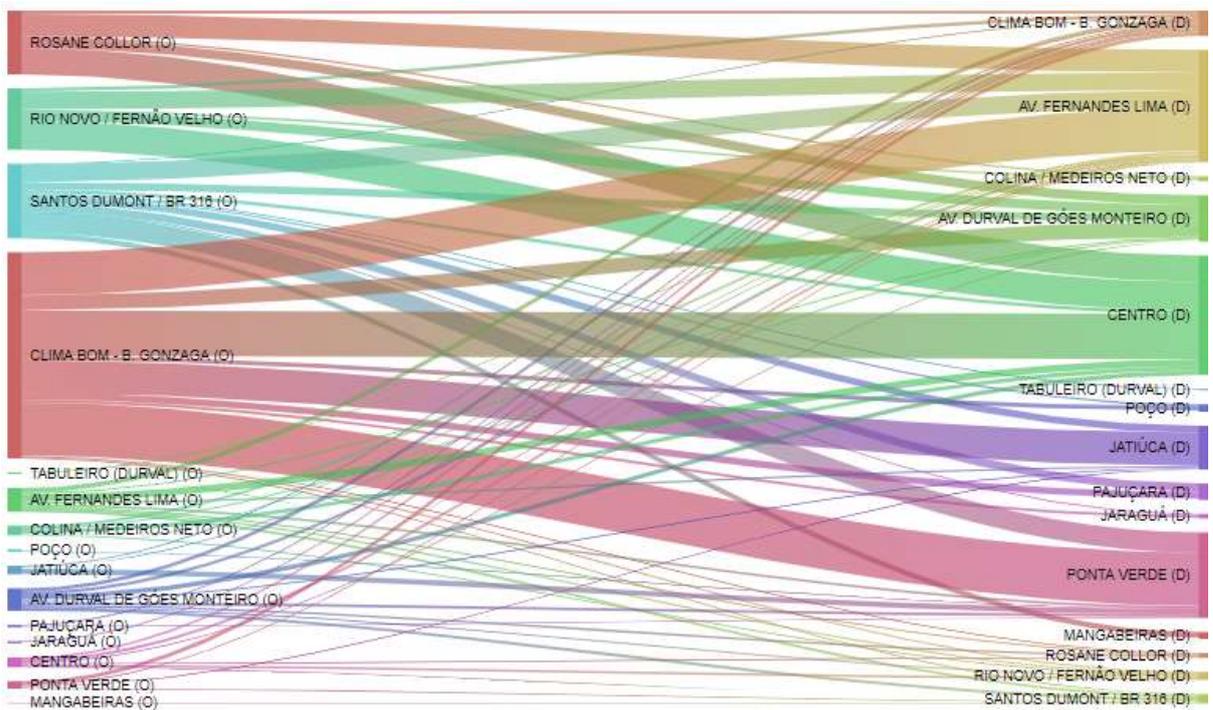


Fonte: Autor

Na Figura 38 fica evidente que nestes 4 trechos os fluxos, destaca-se o trecho Clima Bom / Bomba do Gonzaga, com alta intensidade de demanda para a Av. Fernandes Lima, Centro, Ponta Verde e Jatiúca. Desta forma, parece interessante estudar a viabilidade de algumas alternativas, tais como:

- Viagens extras expressas entre Clima Bom e Ponta Verde e Jatiúca (Há demanda para que estas viagens partam do Santos Dumont);
- Viagens extras semi-expressas entre Clima Bom e Centro ou Ponta Verde, com paradas na Avenida Fernandes Lima.
- Reforço nas linhas;

Figura 38 - Fluxo OD no pico da manhã



Fonte: Autor

Muitas das possibilidades de melhoria das linhas estudadas partem de unificação de linhas, investimentos em sistemas troncais e serviços expressos, porém, como sugerido pelo Guia do Transporte Público Coletivo (AMICCI et al., 2018), seria necessário a instalação de corredores centrais ou sistemas de BRT ou VLT. É a partir destes sistemas que seria possível diminuir o tempo de deslocamento e aumentar a demanda do TPC a partir do incentivo pelo tempo de viagem.

6. CONCLUSÕES

A comunidade de código aberto foi crucial para o desenvolvimento deste trabalho, pois foram atreladas diversas frentes de trabalho a partir da adoção de bibliotecas já desenvolvidas por terceiros e utilizadas por milhares de pessoas.

Por este motivo, foi possível alcançar os objetivos do trabalho, construindo um sistema interativo e intuitivo para análise dos dados. Porém, espera-se com este trabalho incentivar a utilização de sistemas unificados, disponibilização dos dados via API e cruzamento de informações do TPC, pois, foram constatadas diversas divergências ao unificar as informações. Estes problemas dificultaram a eficiência do sistema ao estimar embarque e desembarque de passageiros.

A partir do cruzamento de dados, foi possível identificar problemas de cadastro nos sistemas. Esses apontamentos contribuem para uma operação de forma mais efetiva e com controle de informação produzindo um sistema mais robusto a cada análise realizada.

O modelo de árvore de decisão se ajustou bem à problemática, ficando em 1º lugar entre os modelos estudados, além de permitir uma visualização acerca das características que implicam nas viagens críticas. Com isso, esta abordagem pode ser utilizada pela operação a fim de melhorar a qualidade operacional dos sistemas de TPC.

Os dados apontam para uma problemática do movimento pendular, que merece atenção por parte das empresas e da prefeitura. As empresas precisam mitigar os prejuízos e a prefeitura precisa direcionar ações que visem alterar a ocupação de áreas da cidade a fim de melhorar a mobilidade da cidade e diminuir o movimento pendular.

A partir das classificações das Árvores da Decisão, foi possível identificar pontos críticos que podem ter estudos direcionados por parte das empresas, visando mitigá-los, melhorando assim a operação. Dentre os principais pontos estão a identificação de alta de prejuízo nas linhas do eixo Bebedouro com relação ao eixo Farol e viagens cheias no pico da manhã, principalmente entre o Clima Bom e a parte central de Maceió.

Para tentar solucionar os problemas encontrados, propõe-se estudar o agrupamento de linhas, a operação de veículos alternando a linha conforme altera o sentido, a implantação de linhas expressas e semi-expressas e, transporte de massa na Avenida Fernandes Lima.

Em trabalhos futuros, sugere-se a adição de métodos que possam melhorar o percentual de embarques encontrados, a partir de um melhor preenchimento de falhas, como também

identificar pontos problemáticos em um trecho, corrigi-los na operação e fazer análise dos dados na semana seguinte.

Quanto à melhoria das estimativas de desembarque, sugere-se a melhora a partir da inserção de outros dados relacionados, como dados de uso e ocupação do solo, ou, fazer a análise considerando os pontos de transbordo, como terminais de integração. Desta forma, pode-se estudar a lotação de passageiros em trechos críticos como critério de avaliação de viagens cheias.

REFERÊNCIAS

ALPAYDIN, E. **Introduction to machine learning**. 4. ed. Cambridge, Massachusetts: The MIT Press, 2020.

ALSGER, A. et al. Validating and improving public transport origin-destination estimation algorithm using smart card fare data. **Transportation Research Part C: Emerging Technologies**, v. 68, p. 490–506, 1 jul. 2016.

AMICCI, A. G. N. et al. **Guia TPC: orientações para seleção de tecnologias e implementação de projetos de transporte público coletivo**. Rio de Janeiro: Banco Nacional de Desenvolvimento Econômico e Social, 2018.

BAZANI, A.; MOREIRA, W. **CPTM e Metrô apresentam novo aplicativo de compra de QR Code para pagamento de passagens**. Disponível em: <<https://diariodotransporte.com.br/2020/12/11/cptm-e-metro-apresentam-novo-aplicativo-de-compra-de-qr-code-para-pagamento-de-passagens/>>. Acesso em: 8 jan. 2021.

BITREFINE. **Machine Learning**. Disponível em: <<https://bitrefine.group/machine-learning/>>. Acesso em: 8 jan. 2021.

CONTRERAS, O. **Support Vector Machines for Classification | by Oscar Contreras Carrasco | Towards Data Science**. Disponível em: <<https://towardsdatascience.com/support-vector-machines-for-classification-fc7c1565e3>>. Acesso em: 4 jan. 2021.

COX, B. **Object-oriented programming: an evolutionary approach**. 1986.

CUI, A. **Bus Passenger Origin-Destination Matrix Estimation Using Automated Data Collection Systems** by. Tese de Doutorado. Massachusetts Institute of Technology.

DADOS AL. **Bairros de Alagoas - Bairros de Alagoas [XLSX] - Portal Brasileiro de Dados Abertos**. Disponível em: <<https://dados.gov.br/dataset/bairros-de-alagoas/resource/b353b006-766d-4a0c-9f94-824fab6adc20>>. Acesso em: 8 jan. 2021.

DSA, D. S. A. **Capítulo 19 - Overfitting e Regularização - Parte 1 - Deep Learning Book**. Disponível em: <<http://deeplearningbook.com.br/overfitting-e-regularizacao-parte-1/>>. Acesso em: 8 jan. 2021.

GIRIDHAR, C. **Learning Python Design Patterns Second Edition Learning Python Design Patterns Second Edition Credits**.

GITHUB. **The State of the Octoverse | The State of the Octoverse explores a year of change with new deep dives into developer productivity, security, and how we build communities on GitHub.** Disponível em: <<https://octoverse.github.com/>>. Acesso em: 8 jan. 2021.

HSU, C.-W.; CHANG, C.-C.; LIN, C.-J. **A Practical Guide to Support Vector Classification.** [s.l: s.n.]. Disponível em: <<http://www.csie.ntu.edu.tw/~cjlin>>. Acesso em: 4 jan. 2021.

IBGE. **IBGE - Panorama de Maceió.** Disponível em: <<https://cidades.ibge.gov.br/brasil/al/maceio/panorama>>. Acesso em: 25 jan. 2020.

JUNG, J.; SOHN, K. Deep-learning architecture to forecast destinations of bus passengers from entry-only smart-card data. **IET Intelligent Transport Systems**, v. 11, n. 6, p. 334–339, 1 ago. 2017a.

JUNG, J.; SOHN, K. Deep-learning architecture to forecast destinations of bus passengers from entry-only smart-card data. **IET Intelligent Transport Systems**, v. 11, n. 6, p. 334–339, 1 ago. 2017b.

MA, X. et al. Mining smart card data for transit riders' travel patterns. **Transportation Research Part C: Emerging Technologies**, v. 36, p. 1–12, 1 nov. 2013.

MARQUES, J. **Ônibus de Curitiba passam a aceitar cartões de débito e crédito para pagamento de tarifa.** Disponível em: <<https://diariodotransporte.com.br/2020/07/13/onibus-de-curitiba-passam-a-aceitar-cartoes-de-debito-e-credito-para-pagamento-de-tarifa/>>. Acesso em: 8 jan. 2021.

MOURA, R.; CASTELLO BRANCO, M. L. G.; FIRKOWSKI, O. L. C. DE F. Movimento pendular e perspectivas de pesquisas em aglomerados urbanos. **São Paulo em Perspectiva**, v. 19, n. 4, p. 121–133, dez. 2005.

NASCIMENTO NETO, P. A dimensão esquecida da política habitacional: reflexões a partir do caso da Área Metropolitana de Curitiba (PR). **Cadernos Metrópole**, v. 22, n. 47, p. 215–246, abr. 2020.

NIELSEN, M. A. **Neural Networks and Deep Learning** Determination Press, 2015. Disponível em: <<http://neuralnetworksanddeeplearning.com>>. Acesso em: 8 jan. 2021

PEDREGOSA, F. et al. **Scikit-learn: Machine Learning in Python** Gaël Varoquaux Bertrand Thirion Vincent Dubourg Alexandre Passos PEDREGOSA, VAROQUAUX,

GRAMFORT ET AL. Matthieu Perrot *Journal of Machine Learning Research*. Disponível em: <<http://scikit-learn.sourceforge.net.>>. Acesso em: 3 jan. 2021.

QI, G. et al. Analysis and prediction of regional mobility patterns of bus travellers using smart card data and points of interest data. **IEEE Transactions on Intelligent Transportation Systems**, 2018.

SILVA, M. *Web Design Responsivo: Aprenda a criar sites que se adaptam automaticamente a qualquer dispositivo, desde desktops até telefones celulares*. 2018.

SMTT. **SMTT - Passageiros transportados por ano**. Disponível em: <<http://www.smttmaceio.com.br/portal/smttonline/grafico.html>>. Acesso em: 25 jan. 2020.

SOUZA, S. **Cartão de débito C6 Bank começa a ser aceito no transporte público; conheça a novidade**. Disponível em: <<https://fdr.com.br/2020/10/05/cartao-de-debito-c6-bank-comeca-ser-aceito-no-transporte-publico-conheca-novidade/>>. Acesso em: 8 jan. 2021.

TANG, T.; LIU, R.; CHOUDHURY, C. Incorporating weather conditions and travel history in estimating the alighting bus stops from smart card data. **Sustainable Cities and Society**, v. 53, p. 101927, 1 fev. 2020.

TOQUÉ, F. et al. **Forecasting dynamic public transport origin-destination matrices with long-short term memory recurrent neural networks**. IEEE Conference on Intelligent Transportation Systems, Proceedings, ITSC. *Anais...Institute of Electrical and Electronics Engineers Inc.*, 22 dez. 2016. Acesso em: 31 jan. 2020

WANG, W.; ATTANUCCI, J.; WILSON, N. Bus Passenger Origin-Destination Estimation and Related Analyses Using Automated Data Collection Systems. **Journal of Public Transportation**, v. 14, n. 4, p. 131–150, dez. 2011.

ZEMEL, T. *Web Design Responsivo: páginas adaptáveis para todos os dispositivos*. 2015.