



UFOP

Universidade Federal
de Ouro Preto

**Universidade Federal de Ouro Preto
Instituto de Ciências Exatas e Aplicadas
Departamento de Computação e Sistemas**

Sistema para classificação de malha viária baseado em Smartphones por meio de aprendizado supervisionado

Davidson Esteves Nunes

TRABALHO DE CONCLUSÃO DE CURSO

ORIENTAÇÃO:

Talles de Medeiros

COORIENTAÇÃO:

Vinícius Fernandes Soares Mota

Erik de Britto e Silva

**Fevereiro, 2018
João Monlevade–MG**

Davidson Esteves Nunes

**Sistema para classificação de malha viária
baseado em Smartphones por meio de
aprendizado supervisionado**

Orientador: Talles de Medeiros

Coorientadores: Vinícius Fernandes Soares Mota
Erik de Britto e Silva

Monografia apresentada ao curso de Sistemas de Informação do Instituto de Ciências Exatas e Aplicadas, da Universidade Federal de Ouro Preto, como requisito parcial para aprovação na Disciplina “Trabalho de Conclusão de Curso II”.

Universidade Federal de Ouro Preto

João Monlevade

Fevereiro de 2018

N972s

Nunes, Davidson Esteves.

Sistema para classificação de malha viária baseado em smartphones por meio de aprendizado supervisionado [manuscrito] / Davidson Esteves Nunes. - 2018.

82f.:

Orientador: Prof. Dr. Talles Henrique de Medeiros.

Coorientador: Prof. Dr. Vinícius Fernandes Soares Mota.

Monografia (Graduação). Universidade Federal de Ouro Preto. Instituto de Ciências Exatas e Aplicadas. Departamento de Computação e Sistemas de Informação.

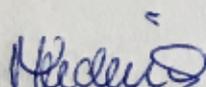
1. Sistemas de informação. 2. Inteligência artificial. 3. Rodovias. I. Medeiros, Talles Henrique de. II. Mota, Vinícius Fernandes Soares. III. Universidade Federal de Ouro Preto. IV. Título.

FOLHA DE APROVAÇÃO DA BANCA EXAMINADORA

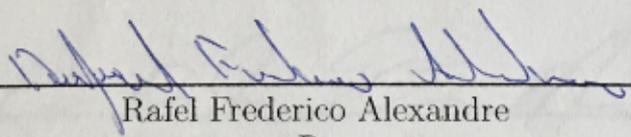
Sistema para classificação de malha viária baseado em Smartphones por meio de aprendizado supervisionado

Davidson Esteves Nunes

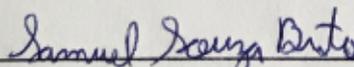
Monografia apresentada ao Instituto de Ciências Exatas e Aplicadas da Universidade Federal de Ouro Preto como requisito parcial da disciplina CSI499 – Trabalho de Conclusão de Curso II do curso de Bacharelado em Sistemas de Informação e aprovada pela Banca Examinadora abaixo assinada:



Talles de Medeiros
MSc
DECSI – UFOP



Rafael Frederico Alexandre
Dr
Examinador
DECSI – UFOP



Samuel Brito
MSc
Examinador
DECSI – UFOP

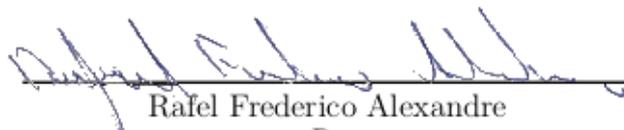
ATA DE DEFESA

No dia 08 do mês de Fevereiro de 2018, às 11:00 horas, na sala C304 do Instituto de Ciências Exatas e Aplicadas, foi realizada a defesa de Monografia pelo(a) aluno(a) **Davidson Esteves Nunes**, sendo a Comissão Examinadora constituída pelos professores: Talles de Medeiros, Rafael Frederico Alexandre, Samuel Brito. O(a) candidato(a) apresentou a monografia intitulada: "**Sistema para classificação de malha viária baseado em Smartphones por meio de aprendizado supervisionado**". A comissão examinadora deliberou, por unanimidade, pela aprovação do candidato, com nota 10,0 (DEZ), concedendo-lhe o prazo de 15 dias para incorporação das alterações sugeridas ao texto final. Na forma regulamentar, foi lavrada a presente ata que é assinada pelos membros da Comissão Examinadora e pelo(a) graduando(a).



Talles de Medeiros

MSc
DECSI – UFOP



Rafael Frederico Alexandre

Dr
Examinador(a)
DECSI – UFOP



Samuel Brito

MSc
Examinador(a)
DECSI – UFOP



Davidson Esteves Nunes

Este trabalho é dedicado a todo aquele que acredita que seu limite sempre está além do próprio horizonte.

Agradecimentos

Agradeço a minha família, especialmente minha mãe Eva que sempre acreditou que de alguma maneira podemos sempre melhorar. Minha irmã Amanda que pôs na minha cabeça a ideia de fazer Enem um dia desses. Meu irmão Diego que sempre foi um incentivo a buscar ser alguém melhor.

Agradeço à turma da salinha, pois teria sido tudo muito mais difícil sem essa turma que estuda muito.

Agradeço à família BlackJack, pelos quatro anos de convivência e amizade.

Agradeço aos meus professores, que do primeiro ao oitavo período, me estimularam a expandir minha mente. À professora Anliy, que embora nunca tenha me dado uma aula, ensinou muito sobre dedicação ao que se faz.

Agradeço a Deus por ter colocado todas essas boas pessoas em meu caminho.

“O correr da vida embrulha tudo. A vida é assim: esquenta e esfria, aperta e daí afrouxa, sossega e depois desinquieta. O que ela quer da gente é coragem.”

— Guimarães Rosa (1908 – 1967),
in: Grande Sertão: Veredas.

Resumo

O transporte rodoviário é o mais utilizado no país e ainda assim, a estrutura rodoviária é um dos grandes problemas enfrentados pelos brasileiros. Segundo a Confederação Nacional dos Transportes (NOTÍCIAS, 2017) 61,8% da extensão das rodovias pesquisadas foram consideradas ruins ou péssimas. Tal fator influi diretamente nos custos logísticos associados, uma vez que a má qualidade pode incorrer em um gasto de até 50% a mais de combustível (RODRIGUES, 2002), sem contar ainda o risco de acidentes. O conforto ou desconforto percebido no uso da via também são fatores relevantes uma vez que, segundo a norma ISO 2631 (STANDARDIZATION, 1978), existem limites toleráveis para a exposição humana à vibrações. Saber de antemão a qualidade do trecho, levando em conta o referencial de conforto do usuário, pode influenciar na escolha entre uma rota e outra.

Existem diversas iniciativas para mensurar a qualidade do pavimento com uso de smartphones, contudo, as mesmas levam em consideração apenas os dados objetivos mensurados a partir de leituras feitas com o dispositivo. Um dos objetivos deste trabalho é considerar a opinião dos usuários da via como fator qualificante. Isso traz alguns desafios para um cenário que já era difícil, uma vez que sistemas dessa natureza estão sujeitos a uma grande variedade de ruídos na informação. Este trabalho propõe um sistema de classificação de vias que mensura a qualidade do asfalto baseado no sensoriamento participativo coletando dados como: geolocalização, acelerômetro e opinião dos usuários sobre a qualidade da via.

O sistema é composto por uma aplicativo *Android* para a atividade de sensoriamento, um servidor em nuvem para recepcionar e classificar os dados coletados por diversos usuários e um painel de visualização web para consulta dos trechos classificados pelos usuários. Algoritmos de aprendizado supervisionado utilizam a opinião dada pelos usuários como argumento para a formação de clusters que serão utilizados no processo de classificação da massa de dados.

Neste trabalho foram percorridos 1.218,34 km entre ruas e estradas nas cidades de João Monlevade, Belo Horizonte, Santa Luzia, Ipatinga, Caratinga e Ouro Preto. Os dados foram coletados pelo autor e usuários voluntários que utilizaram o aplicativo, sendo 1.036,64 km em carros e 181,70 km em ônibus. Deste volume, considerando apenas os dados classificados 63,72% foram de trechos considerados bons e 36,28% ruins.

Palavras-chaves: qualidade viária. acelerômetro. crowdsensing. machine learning.

Abstract

Road transport is the most used in the country and yet, the road structure is one of the great problems faced by Brazilians. According the Confederação Nacional de Transportes (NOTÍCIAS, 2017) 64.8% of the surveyed road extension was considered bad or terrible. This factor has a direct bearing on the associated logistical costs, since poor quality can cost up to 50% more fuel (RODRIGUES, 2002), and increase the risk of accidents. The comfort or discomfort perceived in the use of the roadway are also relevant factors since, according to ISO 2631 (STANDARDIZATION, 1978) there are tolerable limits for human exposure to vibrations. Knowing in advance the quality of the pavement, taking into account the user comfort reference, can influence the choice between one route and another.

There are several initiatives to measure the quality of the pavement with the use of smart phones, however they take into account only the objective data measured from readings made with the device. One of the objectives of this work is to consider the users' opinions as a qualifying factor. This brings some challenges to a scenario that was already difficult, since systems of this nature are subject to a wide variety of information noises. This work proposes a road classification system that measures the quality of asphalt based on participatory sensing, collecting data such as: geolocation, accelerometer and opinion of the users on the quality of the road.

The system consists of a textitAndroid application for the sensing activity, a cloud server to receive and classify the data collected by several users and a web visualization panel to query the pavement classified by the users. Supervised learning algorithms use the opinion given by users in the formation of clusters that will be used in the classification process of the data mass.

In this work 1,218.34 km were traveled among streets and roads in the cities of João Monlevade, Belo Horizonte, Santa Luzia, Ipatinga, Caratinga and Ouro Preto. Data was collected by the author and volunteer users who used the application, where 1,036.64 km in cars and 181.70 km in buses. Of this volume, considering only the data classified 63.72 % were of good and 36.28 % bad segments.

Key-words: road quality. accelerometer. crowdsensing. machine learning.

Lista de ilustrações

Figura 1 – Veiculo de Diagnóstico de Rodovias	22
Figura 2 – Eixos calculados pelo acelerômetro	22
Figura 3 – Exemplo de estrutura de dados de um SSP.	23
Figura 4 – Waze	24
Figura 5 – Aprendizagem de máquina	25
Figura 6 – Propriedades do canopy	27
Figura 7 – Clusterização com canopies	27
Figura 8 – Elementos no cálculo de silhoueta	28
Figura 9 – Curva de $s(i)$ plotada, 2 e 3 clusters	29
Figura 10 – Curva de $s(i)$ plotada, 4 e 5 clusters	29
Figura 11 – Matriz de confusão	30
Figura 12 – Algoritmo kNN	31
Figura 13 – Algoritmo J48 - Árvore de decisão	32
Figura 14 – Algoritmo Random Forest	33
Figura 15 – SVM - Seleção de hiperplano	33
Figura 16 – RoadScan	36
Figura 17 – Total Pave	37
Figura 18 – Interface do Weka	39
Figura 19 – Primeiro protótipo do sistema	40
Figura 20 – Protótipo no painel do veículo	40
Figura 21 – StreetCheck 1.0, versão inicial	41
Figura 22 – Diagrama da arquitetura do sistema. Arquitetura do aplicativo cliente-servidor.	42
Figura 23 – StreetCheck - versão atualizada	43
Figura 24 – Influência da velocidade sobre suavidade	46
Figura 25 – Clusterização usando algoritmo Kanopy	47
Figura 26 – Silhoueta com dois clusters	47
Figura 27 – Silhoueta com três clusters	48
Figura 28 – Eixos com referencial no veículo	49
Figura 29 – Soma das leituras em relação à velocidade	50
Figura 30 – Cálculo vetorial do vetor resultante	50
Figura 31 – Eixos x, y e z comparados ao resultante	51
Figura 32 – Comparação entre um trecho BOM e um trecho RUIM	51
Figura 33 – Desvio padrão, comparação entre um trecho BOM e um trecho RUIM	52
Figura 34 – Angulo entre superfícies triangulares	53
Figura 35 – Valor limiar para análise - Trecho BOM	54

Figura 36 – Valor limiar para análise - Trecho RUIM	54
Figura 37 – Análise de picos/s	55
Figura 38 – Segmento de trecho em estrada	56
Figura 39 – Variação da suavidade conforme posição	57
Figura 40 – Mapa de captura em João Monlevade	59
Figura 41 – Mapa referente ao usuário 1	61
Figura 42 – Mapa referente gerado com dados da comunidade	61
Figura 43 – Acurácia vs Features	62
Figura 44 – Diagrama do banco de dados	72
Figura 45 – Mapa de Belo Horizonte	73
Figura 46 – Mapa de Santa Luzia	74
Figura 47 – Mapa de Ouro Preto	75
Figura 48 – Mapa de Confins	75
Figura 49 – Mapa de Ipatinga	76
Figura 50 – Mapa de Caratinga	77

Lista de tabelas

Tabela 1 – Comparação de suavidade entre carros e ônibus	44
Tabela 2 – Fator de silhoueta para clusters	48
Tabela 3 – Atributos relevantes	55
Tabela 4 – Especificações do ambiente de sistema	57
Tabela 5 – Amostras de treino dos usuários	59
Tabela 6 – Distâncias percorridas e posicionamento do smartphone	60
Tabela 7 – Distâncias percorridas e posicionamento do smartphone em dados clas- sificados	60
Tabela 8 – Remoção de features e acurácia	63

Lista de abreviaturas e siglas

API	Application Programming Interface
DECSI	Departamento de Computação e Sistemas
FTP	File Transfer Protocol, protocolo de transferência de arquivos
IDE	Integrated Development Interface
ICEA	Instituto de Ciências Exatas e Aplicadas
IPEA	Instituto de Pesquisa Econômica Aplicada
JM	João Monlevade
SI	Sistemas de Informação
SSP	Sistema de Sensoriamento Participativo
UFOP	Universidade Federal de Ouro Preto
WEKA	Waikato Environment for Knowledge Analysis

Sumário

1	INTRODUÇÃO	16
1.1	Objetivo	18
1.2	Organização do trabalho	18
2	REVISÃO BIBLIOGRÁFICA	20
2.1	Vibrações e conforto do usuário	20
2.2	Qualidade do pavimento	21
2.3	Acelerômetro	22
2.4	Aplicações Crowdsensing	23
2.5	Aprendizagem de máquina	25
2.5.1	Aprendizagem não supervisionada	26
2.5.1.1	Algoritmo de clusterização Canopy	26
2.5.1.2	Validação de clusters com Silhouette	28
2.5.2	Aprendizagem supervisionada	29
2.5.3	Algoritmos de classificação	31
2.5.3.1	kNN	31
2.5.3.2	J48	32
2.5.3.3	Random Forest	32
2.5.3.4	SVM	33
2.5.4	Filtragem de features	34
3	TRABALHOS RELACIONADOS	35
3.0.1	Pothole Patrol	35
3.0.2	Bag of words	35
3.0.3	RoadScan	36
3.0.4	Total Pave	37
4	DESENVOLVIMENTO	38
4.1	Software Utilizado	38
4.1.1	Netbeans	38
4.1.2	Android Studio	38
4.1.3	R-Studio	38
4.1.4	WEKA	39
4.1.5	Framework Vaadin	39
4.2	Protótipo inicial	40
4.3	Arquitetura do sistema	41

4.4	Interface do aplicativo	42
4.5	Metodologia	44
4.5.1	Veículos	44
4.5.2	Velocidade de leitura	45
4.5.3	Classes de dados	46
4.5.4	Features utilizadas	48
4.5.4.1	Eixos X, Y e Z	49
4.5.4.2	Soma	49
4.5.4.3	Vetor Resultante	50
4.5.4.4	Desvio Padrão	51
4.5.4.5	Suavidade	52
4.5.5	Features experimentais	53
4.5.5.1	Distorção Média - DISTM	53
4.5.5.2	Limites superiores e inferiores - QX1, QY1, QZ1, QX3, QY3, QZ3	53
4.5.5.3	Pico por segundo - PEAKS	54
4.5.6	Filtragem de features	55
4.5.7	Dados de geolocalização	56
4.5.8	Posicionamento do dispositivo	57
4.5.9	Servidor	57
4.5.10	Classificadores	58
5	RESULTADOS	59
5.1	Distância percorrida	59
5.2	Análise e discussão dos dados	60
5.3	Classificação e visualização	61
5.4	Comparação dos algoritmos	62
6	CONCLUSÃO	64
6.1	Trabalhos futuros	65
	REFERÊNCIAS	66
	APÊNDICES	70
	APÊNDICE A – MATERIAIS ELABORADOS PELO AUTOR	72
	ANEXOS	78
	ANEXO A – OUTROS MATERIAIS	79

1 Introdução

Dentre os meios de transporte utilizados no Brasil, o rodoviário é o mais amplamente utilizado dada a sua capilaridade e a extensão do território nacional. Considerando o grande número de municípios as estradas são, em grande parte, o principal meio de acesso (62,8%), seguido pelo transporte ferroviário (21,0%) e fluvial (12,6%) (ILOS, 2017). Dessa maneira, tem-se nas estradas o principal meio de transporte de cargas fazendo com que esse modal seja o mais representativo nos custos logísticos do setor. Além da questão econômica, o fator social que permite acesso de pessoas entre diferentes municípios em um tempo contabilizado em horas e não dias, permite a integração, a cidadania e avanço social de modo geral. Tendo isso em perspectiva, há de se considerar a grande importância deste tipo de acesso para o desenvolvimento da sociedade como um todo, de modo que o estado de conservação da malha viária é de fundamental importância não sendo apenas uma questão de conforto. É de conhecimento comum que o estado de conservação da malha viária brasileira oferece riscos a seus usuários, fator este apontado pelo CNT (TRANSPORTE, 2017) como um dos principais responsáveis pelo elevado índice de acidentes e incidentes.

Em termos econômicos, considerando que a qualidade da pista pode aumentar entre 50% e 58% o consumo de combustível (RODRIGUES, 2002), saber de antemão a qualidade da mesma pode gerar economia ao se definir uma rota que considere este fator. Caminhos alternativos que ofereçam melhor condição de trânsito permitem que a viagem siga em um ritmo mais uniforme (especialmente em cidades, onde existem várias possibilidades para se chegar ao mesmo destino) gerando economia tanto de tempo quanto de combustível .

Em uma pesquisa feita pela Confederação Nacional de Transportes cerca de 61.8% das vias brasileiras estão em condição ruim ou péssima (NOTÍCIAS, 2017). Dependendo do caso, a qualidade da via pode ser decisiva na tomada de decisão de uma rota uma vez que percorrer um trecho de péssima qualidade irá gerar desgastes no veículo, aumentar o consumo de combustível e reduzir o conforto do passageiro. Sendo assim, obter conhecimento prévio sobre a situação do trânsito e qualidade da via pode ser fundamental para escolha eficiente de rotas.

É de se destacar também que a exposição à vibrações causadas pela qualidade do pavimento pode gerar riscos à saúde e, conseqüentemente, ser fator de causas trabalhistas. O Tribunal Regional do Trabalho de Santa Catarina (TRT Santa Catarina, 2013) considerou relevante a norma ISO2631 em um acórdão dispendo sobre a questão de insalubridade no trabalho de dirigir ônibus, sendo que tal decisão pode se estender à trabalhadores da área de transportes de modo geral. A proposta deste projeto, denominado StreetCheck, é relevante nesse sentido uma vez que, dependendo da atividade exercida, o usuário poderá

planejar rotas que ofereçam menor risco à saúde.

Uma das formas de obter informações sobre situação do trânsito e qualidade da via em tempo real é por meio do sensoriamento participativo, também conhecido como *Crowdsensing*. O sensoriamento participativo é caracterizado como sensoriamento e coleta de dados de forma coletiva que permite gerar informação agregada útil aos participantes (BURKE et al., 2006). Aplicativos como *Waze* e *MapLink* coletam dados de seus usuários para mensurar o tráfego, reportar incidentes nas vias e com isso, sugerirem melhores rotas. Obter resultados, neste caso, depende do envolvimento ativo dos usuários para produzir a informação. O usuário precisa interagir com o dispositivo informando as condições da via e do trânsito em si, além de que o aplicativo também poderá coletar dados dos sensores do telefone que incluem acelerômetros, giroscópios e sistemas de GPS.

É interessante, todavia, automatizar o processo da coleta de dados afim de obter as informações de via sem depender exclusivamente do usuário. Nesse sentido diversos trabalhos vem sendo desenvolvidos em que o smartphone é utilizado como dispositivo de processamento e classificação do estado da via ((ERIKSSON et al., 2008), (LIMA, 2016)). Contudo, tais vias podem conter buracos, lombadas e principalmente, asfalto tão irregular que torna a tarefa de classificação automática desafiadora. Alguns trabalhos tem o foco em mapear irregularidades na via, outros procuram medir a qualidade de forma geral, tendo como referência o parâmetro *IRI* (*International Roughness Index*) (SAYERS; KARAMIHAS, 1998). Os dados de acelerômetro são combinados com as leituras de *GPS* para obter um mapeamento qualitativo da via. Contudo, na realidade brasileira, as vias podem ser consideradas boas ou ruins de acordo com a percepção de seus usuários uma vez que, dependendo da região um trecho pode ser considerado bom ao passo que em outra seria considerado ruim. Este trabalho leva em consideração essa percepção do usuário que pode variar em função da noção de conforto de cada um.

Como o processo de coleta dos dados da via gera uma grande quantidade de dados, se faz necessário o uso de técnicas para busca por padrões nestes dados. Aprendizagem de Máquina consiste em uma área de estudos em Inteligência Artificial que busca, por meio de algoritmos especializados, encontrar padrões em grandes volumes de dados. Através destes padrões, amostras de dados desconhecidos podem ser categorizadas (BISHOP, 2013). Por meio de modelos de aprendizagem supervisionada, os dados coletados no StreetCheck são classificados e a partir daí trazem a informação das vias para seus usuários. O sistema busca comparar trechos desconhecidos com os modelos de "BOM" e "RUIM" fornecidos pelos usuários e assim categorizar itens com características similares. Diversos contribuições tem sido publicadas no intuito de monitorar a qualidade do pavimento dividindo-se em abordagens direcionadas à imagens do solo (através de câmeras e sensores a laser) (GARBOWSKI; GAJEWSKI, 2017), e informações de acelerômetro e GPS (ERIKSSON et al., 2008) e (WANG et al., 2015).

O presente trabalho propõe o software chamado *StreetCheck*, um sistema composto de um aplicativo para dispositivos Android e também um *backend* na nuvem, cujo objetivo é fazer a classificação das vias com base na percepção dos usuários. Para isso, é realizado um procedimento que é dividido em duas etapas: A primeira etapa consiste na coleta de dados a partir das medições do acelerômetro do *smartphone* no qual o aplicativo está instalado. Nesse momento é essencial que o usuário classifique alguns trechos percorridos para que sejam armazenados dados de vias consideradas boas e ruins. Na segunda etapa essas informações são enviadas para um servidor que as utiliza em algoritmos de *Machine Learning* para definir um modelo de classificação que corresponda à percepção do usuário.

1.1 Objetivo

O objetivo deste trabalho é desenvolver um sistema de coleta de dados colaborativa e classificação de malha viária baseado em algoritmos de aprendizagem de máquina. O sistema deverá "aprender" o que o usuário entende por bom e ruim e considerar este fator no modelo a ser usado na classificação. Este projeto continua o trabalho iniciado em (REIS; NUNES; MOTTA, 2017) abrangendo as etapas de captura, pré-processamento, envio, classificação e exibição dos resultados. Nesse sentido, foram definidos os seguintes objetivos específicos para este trabalho.

- Desenvolver uma interface da aplicação Android que permita ao usuário expressar sua opinião.
- Coletar, pré-processar e enviar para o servidor os dados no smartphone.
- Criar uma base de dados colaborativa onde dados possam ser utilizados por diferentes usuários.
- Classificar os dados coletados considerando o perfil do usuário utilizando algoritmos de aprendizagem de máquina.
- Exibir um mapa em que seja possível, para cada usuário, consultar a qualidade das vias conforme sua percepção.

1.2 Organização do trabalho

O texto é organizado em seis capítulos, seguindo a seguinte ordem: no Capítulo 2 - Revisão Bibliográfica onde são introduzidos conceitos, técnicas e equipamentos existentes para tratar do problema em questão. No Capítulo 3 - Trabalhos relacionados, contém referência a trabalhos que de alguma maneira contribuíram para este projeto. No Capítulo 4 - Desenvolvimento, são tratados aspectos relevantes sobre a construção do sistema bem

como da metodologia. No Capítulo 5 - Resultados são discutidos os resultados alcançados e também as dificuldades encontradas. No Capítulo 6 é feita a conclusão do projeto e sua aplicabilidade juntamente com os trabalhos futuros.

2 Revisão bibliográfica

Para desenvolver o software StreetCheck foi necessário buscar na literatura diversos conceitos e metodologias para contemplar cada etapa. Foram encontrados diversos materiais entre artigos, livros e seminários de forma que foi necessário selecionar os que contribuíram de forma mais relevante.

2.1 Vibrações e conforto do usuário

Estudos sobre a sensação de conforto do usuário e qualidade do pavimento apontam haver uma relação entre a quantidade de vibração gerada pela via com o conforto do usuário. Dada a complexidade da questão, que também envolve fatores subjetivos, índices como o IRI (*International Roughness Index*) podem não ser suficientes para expressar o grau de desconforto do usuário (CANTISANI; LOPRENCIPE, 2010). A norma ISO 2631-1 (STANDARDIZATION, 1978) estabelece valores que relacionam o tempo de exposição e a magnitude dos valores de aceleração vertical e transversal com a fadiga, ou nível de eficiência reduzida do indivíduo exposto. Em seu trabalho Balbinot (BALBINOT, 2001) afirma que motoristas de ônibus ou caminhões estão expostos a riscos de saúde ocupacional devido à vibrações transmitidas pela via, podendo desenvolver problemas como hérnias e dores lombares. Essas vibrações geradas pelo próprio funcionamento de veículos pesados podem ser intensificadas em pavimentos ruins.

É importante considerar que no transporte rodoviário nem sempre o passageiro estará na posição sentada em uma poltrona. Diferentes veículos permitem ajustes de reclinção ou ainda é possível que se viaje em pé. Dessa maneira os vetores de aceleração a que estarão sujeitos os usuários não serão sentidos da mesma maneira. Essa questão será discutida na parte de Metodologia (4.5) em que trata da representação dos eixos de captura do acelerômetro.

O StreetCheck considera a percepção do usuário sobre o conforto e como as vibrações provocadas por inconsistências na via interferem nessa percepção. Alguns usuários podem considerar um trecho ruim ao passo que outros considerem aceitável, tendo em vista as diferentes percepções. Usuários do setor de transportes podem selecionar vias de acordo com um certo nível de conforto ao transportar cargas delicadas ou animais. O trânsito de pacientes poderia se beneficiar ao se escolher rotas que causem o menor estresse possível em seus usuários.

2.2 Qualidade do pavimento

O pavimento rodoviário mais comumente encontrado pode ser classificado em duas categorias básicas: o pavimento composto por concreto de cimento e o pavimento asfáltico. Os pavimentos de concreto são aqueles em que o revestimento é uma placa de concreto *Portland* onde a espessura é definida em função da resistência à flexão das placas de concreto e camadas subjacentes. Já o pavimento asfáltico é formado por compostos diversos dispostos em camadas sobre a superfície de forma a criar um revestimento resistente às ações do tráfego (BERNUCCI L.B, 2010). Considerando o veículo que trafega na via, defeitos no pavimento irão causar um maior impacto na estrutura o que conseqüentemente irá afetar a noção de conforto do usuário. Tais impactos podem ainda causar danos estruturais e ainda aumentar custos operacionais em até 58% no consumo de combustível e aumento de 40% em custos operacionais (BERNUCCI, 2008).

Os defeitos no pavimento surgem devido ao uso natural da via, em que o material deteriora com tempo, ou devido a fatores externos como calor, frio ou chuvas que, entre outras coisas, causam deformidades, fendas, trincas, buracos, corrugações, panelas, e deslocamentos no substrato que sustenta a via. A consequência é a piora nas condições de uso e necessidade de reparos. Considera-se ainda como defeito os remendos feitos na via, uma vez que estes além de diminuir a uniformidade evidenciam pontos de fragilidade (BERNUCCI, 2008). A ideia de sistemas, como o proposto neste trabalho, que permitem o acompanhamento em tempo real da condição da via, também proporcionam um monitoramento da degradação, tendo em vista que a medida que os usuários classificam e reclassificam um trecho são gerados dados históricos dos trechos revelando sua evolução.

Para quantificar a qualidade da via são utilizados índices como o *IRI* que determina em m/km os desvios encontrados na superfície em relação ao projeto. Ao somatório desses desvios é dado o nome de irregularidade longitudinal. Em termos de equipamento a medição da qualidade de vias é feita por perfilômetros que podem ser mecânicos, a laser ou ultrassom. Atualmente o sistema mais sofisticado em uso no Brasil é utilizado pelo DNIT (Departamento Nacional de Infraestrutura de Transportes) (TRANSPORTES, 2017) que consiste em um veículo onde são instalados diversos componentes de medição tais como, perfilômetro a laser, odômetro de alta precisão e câmeras para foto e filmagem da rodovia.

Figura 1 – Veículo de Diagnóstico de Rodovias



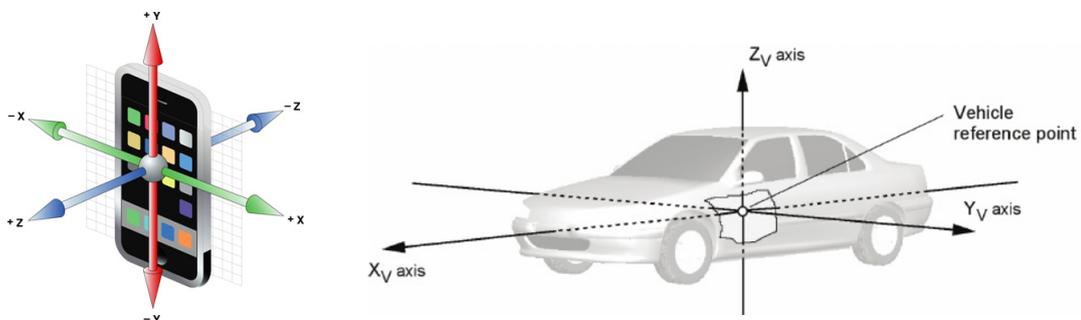
Fonte – (TRANSPORTES, 2017)

O StreetCheck não calcula o IRI, mas considera uma métrica própria de suavidade para quantificar o nível de qualidade dos trechos percorridos.

2.3 Acelerômetro

Segundo (CALACHE, 2013) um acelerômetro é um instrumento eletrônico capaz de medir a aceleração da gravidade em m/s^2 . Os dados lidos são uma tripla formada pelos eixos X, Y e Z. O eixo X (longitudinal) corresponde à aceleração veículo para frente ou para trás. O eixo Y corresponde ao movimento lateral ocorrido à direita ou a esquerda. O eixo Z corresponde à aceleração vertical do veículo de modo que quando este se encontra em repouso o valor apurado é aproximadamente $9,8 m/s^2$ correspondentes à aceleração de gravidade da terra. Devido ao seu baixo custo este se tornou parte integrante de quase todo smartphone a venda atualmente.

Figura 2 – Eixos calculados pelo acelerômetro



Fonte – (TERMINOLOGY, 2008, p. 8)

Com o smartphone estando embarcado no interior do veículo, a aceleração sofrida pelo mesmo em qualquer dos seus eixos automaticamente irá refletir no dispositivo que irá

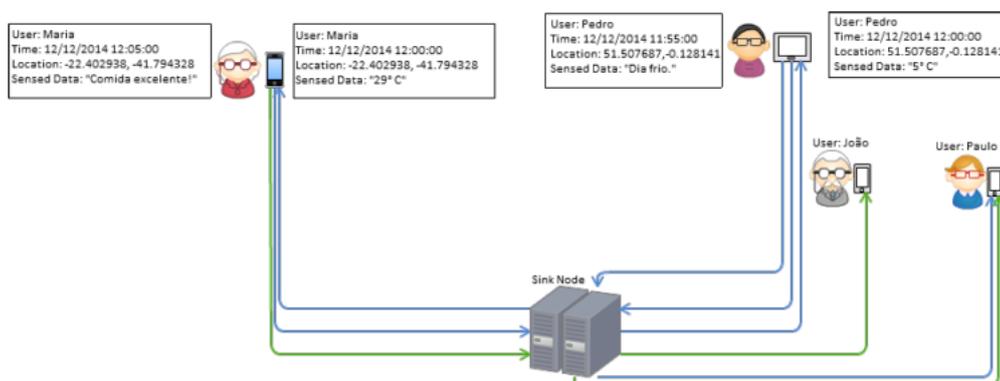
coletar e armazenar os valores.

Neste trabalho, tanto quanto em outros relacionados, o acelerômetro é o dispositivo mais importante por ser a principal fonte de dados. Entretanto, é importante notar que conforme pode ser visto em (BISCONSINI et al., 2016) e também em (BISCONSINI; JÚNIOR, 2015) a qualidade dos dados é afetada pela imprecisão desses acelerômetros. Tal imprecisão precisa ser levada em conta na implementação de projetos que fazem uso desse recurso.

2.4 Aplicações Crowdsensing

Aplicações *Crowdsensing* são baseadas no conceito chamado sensoriamento participativo, que segundo (FREITAS, 2015) trata-se do sensoriamento realizado através de dispositivos móveis portados por pessoas, também chamados Sistemas de SSP (Sistema de Sensoriamento Participativo). Entretanto, os dados obtidos por este sistemas não se restringem apenas aos obtidos por meio de sensores, mas também por informações derivadas de seres humanos (visão, audição, olfato, tato, paladar). No caso do StreetCheck, além de informações derivadas dos sensores, o usuário fornece um *input* com sua opinião sobre o trecho. Em (GANTI; YE; HEI, 2011) define-se aplicações *Crowdsensing* como sistemas que fazem uso de dados coletados por grupos de indivíduos afim de gerar algum resultado de interesse coletivo. O interesse, neste caso, pode ser segmentado ou não. Uma única aplicação SSP pode coletar dados que atendem à diversos escopos e interesses, sejam dos próprios usuários, de empresas ou do poder público que munido da informação de interesse coletivo pode agir nesse sentido.

Figura 3 – Exemplo de estrutura de dados de um SSP.



Fonte – (FREITAS, 2015)

Na figura anterior pode-se notar entidades geradoras e ou consumidoras da informa-

ção. Os usuários João e Paulo consomem a informação gerada pela comunidade, cada um em seu escopo. Pedro e Maria geram e consomem a informação, geralmente num contexto mais restrito. No StreetCheck os atores consumidores poderiam ser tanto órgãos públicos interessados no estado de conservação da malha viária, quanto entidades particulares que tem interesses específicos em rotas de melhor qualidade.

Um exemplo típico desse tipo de aplicação é o *Waze* que além de mostrar mapas rodoviários também permite que os usuários compartilhem informações uns com os outros.

Figura 4 – Waze



Fonte – Disponível em: <www.waze.com>

Na prática, os usuários do Waze fazem uso do serviço de interesse comum (os mapas) e compartilham informações úteis para a comunidade (informações de rotas, acidentes, congestionamento, etc). O modelo de negócio usa um mecanismo de gamificação onde os usuários são premiados de acordo com seu nível de colaboração estimulando uma participação mais ativa da comunidade.

Como plataforma colaborativa, o StreetCheck se encaixa nesse quesito uma vez que a informação de interesse comum, no caso o mapeamento viário, é utilizada no interesse de todos.

Comparadas com as metodologias de sensoriamento tradicionais, as aplicações crowdsensing permitem uma leitura ampla e dinâmica de ambientes que podem sofrer mudanças com o decorrer do tempo. A exemplo do que acontece com pavimentação

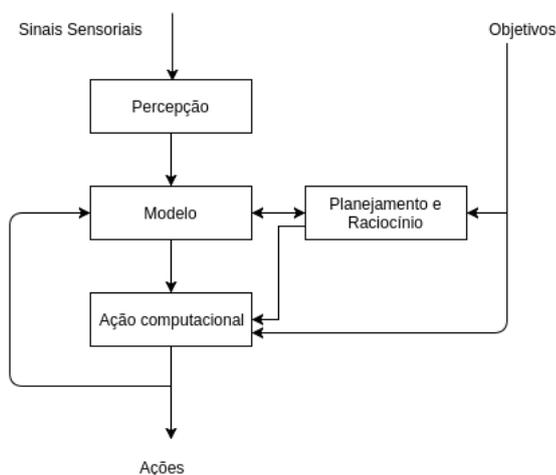
viária, onde com a ação do do tempo ocorre deterioração da superfície, o monitoramento participativo pode revelar a evolução da degradação uma vez que com o uso constante da aplicação os usuários estarão sempre mantendo atualizadas as bases de dados do sistema.

Tendo a informação em mãos, os órgãos responsáveis pela manutenção podem fazer reparos preventivos, o que geralmente é menos custoso que a recuperação total de trechos. Em um contexto de cidades inteligentes seria possível fazer planejamento de rotas com intuito de reduzir o desgaste em áreas fragilizadas.

2.5 Aprendizagem de máquina

Aprendizagem de máquina ou *Machine Learning* é uma área de Inteligencia Artificial em que, segundo (NILSSON, 2005) consiste em um conjunto de técnicas derivadas de campos como estatística, modelagem neural, teoria de controle adaptativo, modelos psicológicos e modelos evolucionários. O objetivo é obter comportamentos em que um software reaja com base em padrões encontrados na massa de dados analisada.

Figura 5 – Aprendizagem de máquina



Fonte – (NILSSON, 2005)

A partir de sinais oriundos de fontes de dados (sensores, bancos de dados, arquivos, etc) um processo de percepção mapeia objetivos com auxílio de um modelo. A máquina pode executar ações a partir das decisões tomadas, sendo que paralelamente o modelo é retroalimentado. O processo de aprendizagem possui duas vertentes, uma baseada em modelos pré-definidos onde o programa segue exemplos previamente registrados, chamado aprendizagem supervisionada. A outra, não supervisionada, é baseada na busca por padrões, sem que haja um conhecimento prévio do objeto de estudo.

2.5.1 Aprendizagem não supervisionada

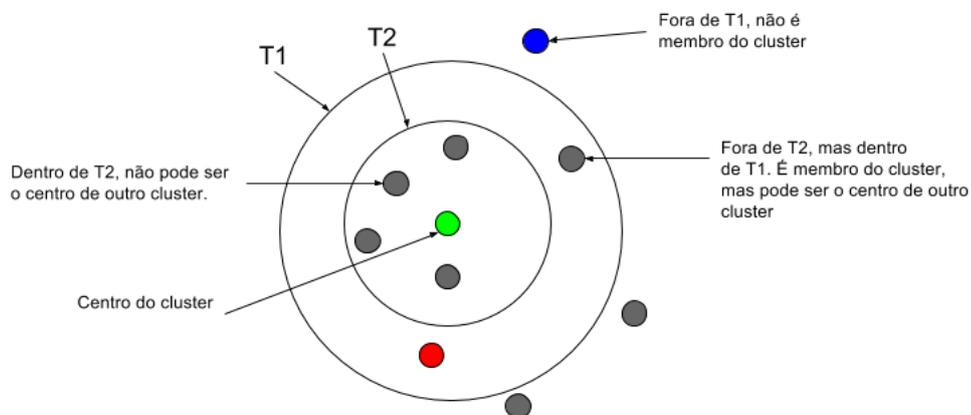
O aprendizado não supervisionado é um método que busca separar conjuntos de dados tendo como base a similaridade entre seus atributos (NILSSON, 2005). Amostras similares formam agrupamentos chamados *clusters* que se destacam pela proximidade em que seus atributos se encontram. Para quantificar a proximidade são usadas funções de distância, sendo a mais comumente utilizada a distância euclidiana.

2.5.1.1 Algoritmo de clusterização Canopy

Neste trabalho foi utilizado o algoritmo Canopy para encontrar clusters nos dados coletados pelos usuários. O algoritmo é descrito em (WHIZBANG; NIGAM; UNGAR, 2000) e definido como um método rápido para detecção de clusters em grandes volumes de dados envolvendo uma técnica barata para medição de distâncias. O objetivo aqui era verificar se as categorias "BOM" e "RUIM" se ajustavam adequadamente aos dados ou se era justificável alguma outra categoria como "REGULAR". Esta metodologia é utilizada quando não se conhece a existência de padrões nos dados observados. A implementação do algoritmo que foi usada encontra-se na biblioteca do WEKA. Este algoritmo foi escolhido especificamente por não precisar de nenhum parâmetro e revelar, de forma automática, os clusters existentes.

O processo de clusterização é feito representando todos os objetos como um ponto no espaço multidimensional de suas características. São definidos dois limiares de distância T_1 e T_2 , sendo $T_1 > T_2$. O algoritmo funciona em dois estágios, no primeiro é selecionado um ponto aleatório e criado um canopy. Os pontos próximos deste ponto central, que estão no raio de T_2 , não podem ser centros de outros canopies. Uma função menos custosa é usada para calcular essas distâncias.

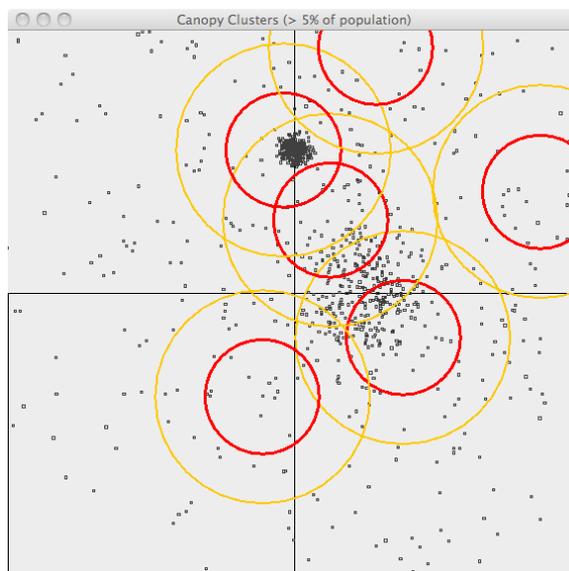
Figura 6 – Propriedades do canopy



Fonte – Próprio autor

Outros canopies são criados, tentando manter uma distância dos centros uns dos outros, contudo, naturalmente haverá sobreposições. A primeira etapa termina quando não existem mais canopies a serem criados, conforme visto na figura a seguir ¹.

Figura 7 – Clusterização com canopies



Fonte – (Apache Software Foundation, 2005)

O próximo passo consiste em usar uma função de distância mais robusta para resolver os pontos que ficaram dentro de dois ou mais canopies que se sobrepuseram. O

¹ Disponível em: <https://docs.opencv.org/>

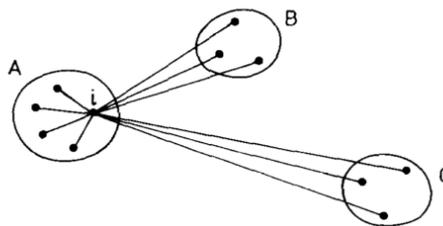
ponto pertence ao canopy de centro mais próximo. O passo seguinte para a validação do conjunto de clusters é a verificação de adequação dos mesmos aos dados, o que neste trabalho foi feito utilizando o algoritmo Silhouette apresentado a seguir.

2.5.1.2 Validação de clusters com Silhouette

Conquanto o algoritmo de clusterização Canopy seja uma ferramenta para encontrar rapidamente clusters em grandes conjuntos de dados, o número de clusters encontrados não é necessariamente o ideal. Uma metodologia para avaliar se o número de *clusters* encontrados se adequam da melhor forma aos dados é encontrada em (ROUSSEEUW, 1987). O processo consiste em computar um vetor $s(i)$ para a coleção de elementos i pertencentes a um cluster, utilizando uma lista de dissimilaridades $d(i, C)$ afim de calcular um vetor $b(i)$ correspondente à dissimilaridade mínima de i .

- $a(i)$ = corresponde à dissimilaridade média entre o elemento i com todos os pontos do cluster A.
- $d(i, C)$ = dissimilaridade média do elemento i com os elementos de C.
- $b(i)$ = valor mínimo de $d(i, C)$

Figura 8 – Elementos no cálculo de silhoueta



Fonte – (ROUSSEEUW, 1987, p. 55)

Em um dado vetor $a(i)$ é colocada a média de dissimilaridade entre cada elemento i de A com os outros elementos do mesmo conjunto. O vetor $b(i)$ representa o vizinho mais próximo do elemento i , para o caso o mesmo não ser considerado como parte integrante de A. Esse fato implica que para o algoritmo funcionar o mesmo precisa de pelo menos dois clusters. Dessa maneira $s(i)$ é dado pela fórmula:

$$s(i) = \frac{b(i) - a(i)}{\max(a(i), b(i))} \quad (2.1)$$

O resultado de $s(i)$ é então plotado em um gráfico que mostra a curva de adequação.

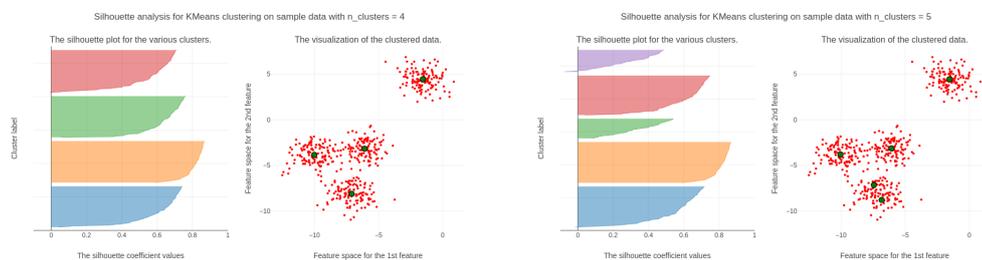
Figura 9 – Curva de $s(i)$ plotada, 2 e 3 clusters



Fonte – Disponível em <<https://plot.ly/scikit-learn/plot-kmeans-silhouette-analysis>>

O mesmo conjunto de dados organizado em dois clusters e três clusters tem respectivamente 0.70 e 0.58 pontos de silhoueta.

Figura 10 – Curva de $s(i)$ plotada, 4 e 5 clusters



Fonte – Disponível em <<https://plot.ly/scikit-learn/plot-kmeans-silhouette-analysis>>

Para quatro e cinco clusters o score vai para 0,65 e 0,56 respectivamente. Como será visto em 4.5.3 o número de clusters adequados ao StreetCheck é dois, correspondendo às opções de "BOM" e "RUIM" para qualificar o pavimento.

2.5.2 Aprendizagem supervisionada

Com a utilização de sensores e múltiplos agentes de coleta de dados, o volume de informação cresce de maneira rápida. Analisar grandes conjuntos de dados afim de se extrair conhecimento sobre a qualidade do pavimento é uma tarefa humanamente extenuante e improdutiva. Neste ponto, a capacidade de processamento de informação dos computadores pode ser usada para extração de padrões indicativos nos dados.

Machine Learning é a metodologia que vem sendo empregada em diversos trabalhos onde o foco é a extração de padrões e classificação de informações. Trata-se de uma técnica de busca de padrões em grandes conjuntos de dados de modo que o objetivo é encontrar características semelhantes entre conjuntos de amostras e assim agrupa-las.

O método de aprendizagem supervisionada é um técnica de classificação que consiste em definir um modelo a partir de exemplos. O usuário fornece um conjunto de amostras que consistem em dados X associados a um resultado Y . Por X entende-se como um grupo de uma ou mais características observados em um objeto ou evento qualquer. Y corresponde à saída dada para tais características em forma de classe (para problemas de classificação) ou valores (para problemas de regressão). Um algoritmo é executado sobre um conjunto expressivo de amostras e um modelo de classificação é definido de tal maneira que seja possível classificar dados de entrada desconhecidos a partir desse modelo (BROWLEE, 2016).

Para a análise ser bem sucedida são definidas as características do objeto de estudo que no caso são chamadas de *features*. Cada *feature* mostra uma característica objetiva dos dados de forma que, por meio dela, seja possível distinguir uma amostra de outra. Um algoritmo de classificação irá utilizar essas características para separar em classes a massa de dados.

O grau de precisão do algoritmo depende da quantidade de amostras ser suficientemente expressiva e da qualidade dos *features* existentes na amostra. Um número insuficiente de amostras ou *features* de pouca qualidade podem levar a erros de classificação uma vez que a técnica procura agrupar elementos por sua semelhança. O desempenho do algoritmo de classificação é medido através de sua matriz de confusão que é gerada por meio de testes nas amostras de treinamento.

Figura 11 – Matriz de confusão

		Valor Previsto	
		Positivo	Negativo
Valor Verdadeiro	Negativo	Verdadeiros Positivos	Falsos Negativos
	Positivo	Falsos Positivos	Verdadeiros Negativos

Fonte – (ANDREONI, 2014, p. 28)

Neste trabalho o usuário pode, por meio do aplicativo, classificar o trecho em duas categorias como "BOM" ou "RUIM", o que se encaixa na matriz de confusão como classe positiva e classe negativa. O desempenho do algoritmo é medido considerando como base os dados de trecho classificados pelo usuário e essa base é chamada de dados de treino. Contudo, pode ocorrer do usuário classificar um trecho em uma categoria e em outro momento reclassificar o mesmo trecho em outra categoria. Tal feito introduz um dilema de classificação que pode diminuir a acurácia. No presente projeto este fato é tratado inserindo novas instâncias da amostra na base de dados. As variações vistas nestes dados poderão revelar uma deterioração em progresso na via.

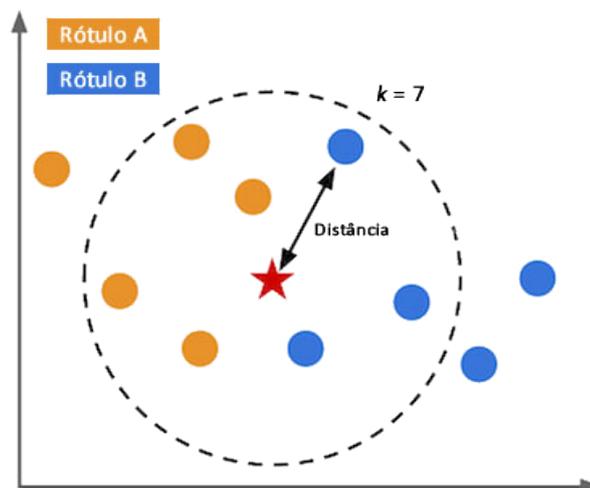
2.5.3 Algoritmos de classificação

O objetivo básico de um algoritmo de classificação é criar uma linha divisória entre instâncias. Tendo sido definidas classes de objetos o objetivo é inferir, por meio das características de cada amostra, a qual classe a mesma pertence. Neste trabalho foram testados alguns algoritmos para classificar os dados coletados:

2.5.3.1 kNN

O kNN classifica uma amostra desconhecida relacionando-a com amostras conhecidas mais próximas da mesma. Tem esse nome por considerar os k vizinhos mais próximos, ou k -nearest neighbors, e o método mais comum para avaliar a proximidade é a distância euclidiana. Diversas características podem funcionar como preditores de modo que o espaço amostral terá tantas dimensões quanto se queira avaliar.

Figura 12 – Algoritmo kNN



Fonte – Disponível em <https://goo.gl/4mksrk>

Para cada amostra analisada é criado um vetor de distâncias contendo as distâncias de todos os vizinhos. A classe dos "k" vizinhos conhecidos mais próximos determina a classe da amostra.

2.5.3.2 J48

O J48 trabalha com o conceito de árvore de decisão, considerando os dados de treinamento como diferentes ramificações. Cada folha é o resultado de uma combinação dos diferentes valores para os atributos dados em cada amostra do conjunto de testes.

Figura 13 – Algoritmo J48 - Árvore de decisão

```

J48 pruned tree
-----
vm <= 57.480551
|
| peaks <= 11.738592
| |
| | distm <= 3.20907
| | |
| | | dp <= 0.701277
| | | |
| | | | distm <= 0.084359
| | | | |
| | | | | vm <= 14.868346: RUIW (21.0/2.0)
| | | | | vm > 14.868346
| | | | | | qx1 <= 0.666509: BOM (19.0)
| | | | | | qx1 > 0.666509
| | | | | | | qz1 <= 10.207955
| | | | | | | | distm <= 0.061538: BOM (4.0)
| | | | | | | | distm > 0.061538: RUIW (2.0)
| | | | | | | | qz1 > 10.207955: RUIW (4.0)
| | | | | distm > 0.084359
| | | | | | dp <= 0.557654
| | | | | | | distm <= 1.59674: BOM (751.0/19.0)
| | | | | | | distm > 1.59674
| | | | | | | | suavidade <= 0.085167: BOM (62.0/3.0)
| | | | | | | | suavidade > 0.085167
| | | | | | | | | qy1 <= 8.476183: BOM (6.0/1.0)
| | | | | | | | | qy1 > 8.476183: RUIW (4.0)

```

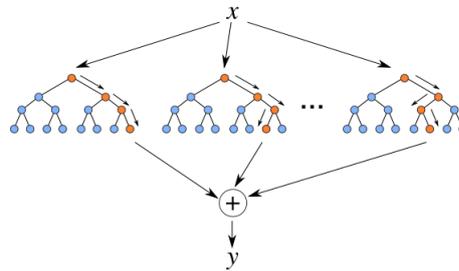
Fonte – Próprio autor

A árvore é construída após sucessivas divisões dos seus exemplos de acordo com seus valores preditivos (Von Zuben; ATTUX, 2007). No exemplo, com dados do StreetCheck é possível observar que a amostra é da classe "BOM" se $qz1 \leq 10$, $qx1 > 0.66$, $distm \leq 10.20$, $vm > 14.86$ até chegar no nó inicial.

2.5.3.3 Random Forest

Segundo (BREIMAN, 2001) Random Forests são combinações de árvores de decisão cada qual com pelo menos três atributos preditores. A estratégia do Random Forest é utilizar um conjunto de árvores de decisão, de forma que cada uma delas analisando um conjunto de características da amostra e chegue a uma conclusão. O algoritmo reúne então esse conjunto de "votos" e a classe mais votada é considerada a classe da amostra analisada.

Figura 14 – Algoritmo Random Forest

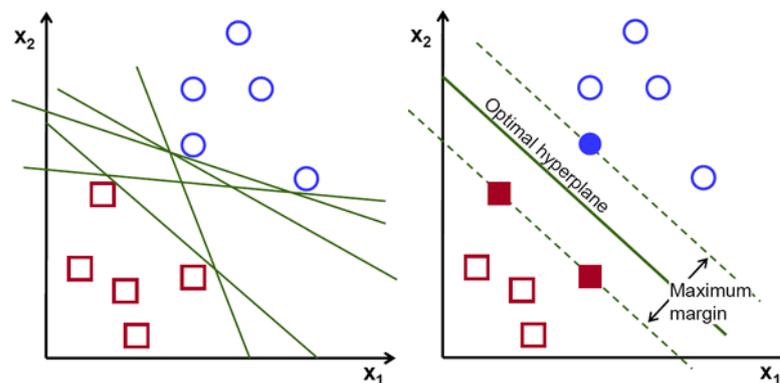


Fonte – (BREIMAN, 2001)

2.5.3.4 SVM

O SVM (Support Vector Machine) procura separar as amostras em um espaço dimensional linearmente separável por meio de uma função de transformação denominada Kernel. O objetivo dessa função é definir um hiperplano que separe os dados. A fase de modelagem com os dados de treinamento serve para encontrar uma função separadora. O Algoritmo SVM também encontra um hiperplano que maximiza a margem de separação das classes (LORENA, 2007)

Figura 15 – SVM - Seleção de hiperplano

Fonte – Disponível em: <<https://docs.opencv.org/>>

Essa abordagem tem encontrado muitos bons resultados na literatura, contudo, o desafio é encontrar os melhores parâmetros para o ajustamento do modelo de forma que não ocorra nem *overfitting* (ajuste excessivo ao modelo, não reconhecendo novas amostras) e nem o *underfitting* (errar mesmo nos dados de treinamento). Esse processo de ajuste é um problema de otimização. Para o cálculo do hiperplano separador utiliza-se os vetores de suporte que são compostos pelas amostras que são as mais informativas no conjunto de

treinamento. Uma vez que tais amostras estão na fronteira entre os grupos, o algoritmo um plano de separação e pode, a partir daí, classificar as novas amostras.

2.5.4 Filtragem de features

Um tópico importante a ser considerado é que para qualquer algoritmo de classificação os atributos utilizados precisam, cada um a sua maneira, descrever o objeto observado. Em (HOUEI, 2005) é discutido o algoritmo *ReliefF*, uma abordagem para a seleção de features relevantes. A partir de uma amostra, o algoritmo compara os atributos dos vizinhos de mesma classe mais próximos e também os vizinhos mais próximos de classe oposta. A ideia é verificar o quanto os atributos aproximam amostras de mesma classe ou afastam amostras de classes diferentes.

Neste trabalho, essa metodologia foi utilizada para considerar quais os atributos deveriam ser utilizados pelos algoritmos classificadores na etapa de classificação das amostras.

3 Trabalhos relacionados

A utilização de smartphones para mapeamento da qualidade tem crescido tendo em vista se tratar de um hardware relativamente barato que conta com capacidade de sensoriamento e processamento embutido. Concomitantemente a grande popularização destes dispositivos possibilita a obtenção em larga escala de dados em diferentes regiões. A maior parte dos trabalhos tem o foco em classificar numericamente a via ou detectar anomalias como buracos. Este trabalho busca qualificar trechos inteiros de via.

Todavia é importante observar que os dados de aceleração coletados por smartphones estão sujeitos a ruídos e imprecisões que podem variar de modelo para modelo. Em (BISCONSINI et al., 2016) o autor recomenda que estes dados sejam utilizados com cautela, uma vez que existe uma tendência do *IRI* calculado por smartphones serem de menor magnitude do que o calculado por sistemas tradicionais de aferição da qualidade.

A seguir alguns trabalhos encontrados na literatura que, em algum aspecto, contribuíram para este projeto.

3.0.1 Pothole Patrol

Um dos primeiros trabalhos para a detecção de irregularidades em vias baseado em acelerômetro e posicionamento por satélite foi proposto em um sistema de (ERIKSSON et al., 2008) batizado *Pothole Patrol P²* para detectar irregularidades nas vias. Para isto, utilizaram um acelerômetro nos painéis de um conjunto de táxis em Los Angeles e um algoritmo *Naive* para classificar as irregularidades. Uma outra proposta baseada em heurísticas de limites de leitura dos sensores é apresentada em (MEDNIS et al., 2011). Os autores conseguem uma taxa de acerto de até 92% na detecção de anomalias.

Assim como os autores destes trabalhos propuseram, a abordagem inicial do projeto StreetCheck usou a metodologia de manter o hardware de captura instalado no interior do veículo. O objetivo neste caso residia em buscar coletar amostras com o máximo de fidelidade, sem a interferência causada por movimentações no smartphone. Contudo, neste trabalho a abordagem utilizada foi usar apenas o smartphone ao invés do hardware embarcado no veículo.

3.0.2 Bag of words

No trabalho encontrado em (GONZÁLEZ et al., 2017), os autores propõem o uso da abordagem *Bag of words* (metodologia usada em processamento de linguagem natural) para criar o vetor de características (*features*) como entrada para algoritmos de aprendizado

supervisionado. Embora os algoritmos apresentados tiveram resultados estatisticamente similares, o algoritmo de rede neural apresentou melhor desempenho. As medições foram feitas com o smartphone em diversas posições no veículo de forma a verificar a influência deste fator nos dados. No atual projeto esse fator também foi levado em conta, de forma que o usuário pode registrar o posicionamento do dispositivo durante o uso.

A maioria dos trabalhos propostos focam na detecção de anormalidades como buracos, lombadas ou junções de pistas. Devido a qualidade das vias brasileiras, essas podem conter irregularidades como ondulação, e ainda assim ser caracterizada como boa. Nesse contexto, este trabalho propõe uma classificação de vias urbanas brasileiras. Além disto, tem também como objetivo a caracterização de trechos completos e não apenas detectar anormalidades.

3.0.3 RoadScan

Vem sendo desenvolvido no laboratório iMóbilis do Instituto de Ciências Exatas e Aplicadas da Universidade Federal de Ouro Preto e é um aplicativo cuja proposta de sensoriamento participativo é bastante parecida com este projeto (LIMA, 2016). Ao contrário de aplicações que tentam encontrar valores equivalentes ao IRI, o RoadScan busca trabalhar diretamente com o desvio padrão e o algoritmo "z-peak" encontrado na literatura.

Figura 16 – RoadScan



Fonte – (LIMA, 2016)

Baseado em experimentos de coleta de dados os autores definiram limiares de referência chamados "LOWTHRESH" e "HIGHTHRESH". A partir destes referenciais as mudanças de gradação indicavam se o veículo estava em um pavimento bom ou ruim.

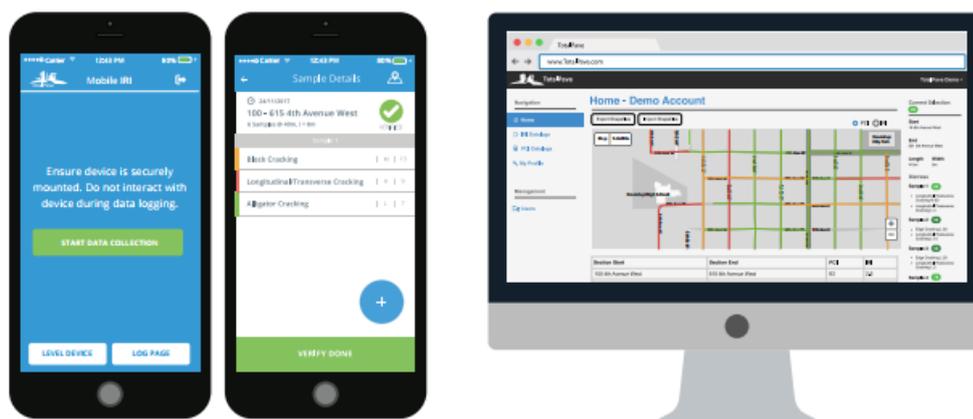
O StreetCheck usa a mesma metodologia para a coleta de dados, contudo a proposta difere no sentido de considerar a opinião do usuário e de explorar os algoritmos de *Machine Learning* para encontrar padrões nos dados coletados que representam segmentos de pista.

3.0.4 Total Pave

Aplicativo disponível para dispositivos Android e IOS ([TOTALPAVE](http://totalpave.com/),). A proposta da empresa desenvolvedora (que leva o mesmo nome do aplicativo) é fornecer uma aproximação do IRI utilizando leituras do acelerômetro do celular. A ideia é que os utilizadores, que no caso tem a incumbência de coletar ativamente os dados, percorram os trechos de via. A informação deve ser acessada por meio de um web-site que mostra os resultados apurados.

A semelhança com o StreetCheck reside no fato da informação ser coletada por um grupo de usuários, processada em um servidor e usar o mesmo para emitir resultados. Contudo, pelo fato deste sistema buscar o índice IRI não pode haver interação do usuário com o smartphone durante a medição. O celular neste caso é como um sensor instalado no veículo. Também é necessária uma calibração do dispositivo antes das medições.

Figura 17 – Total Pave



Fonte – Disponível em: <<http://totalpave.com/>>

O fato de haver empresas explorando a abordagem de uso de smartphones para esse tipo de trabalho demonstra a relevância da pesquisa. Existe um mercado consumidor e o interesse em informações de qualidade viária. Por ser um produto proprietário as informações técnicas sobre a abordagem utilizada não são divulgadas.

4 Desenvolvimento

O desenvolvimento do aplicativo teve como principal objetivo a praticidade de uso. Espera-se que uma aplicação *Crowdsensing* possa ser usada por um grande número de usuários em diversos contextos. Nesse sentido o sistema permite o uso, tanto em carros quanto ônibus, podendo variar a posição e local onde se encontre o smartphone. A abordagem de análise dos dados precisou levar em consideração questões de imprecisão, ruído e informações inconsistentes por parte do usuário.

4.1 Software Utilizado

Todo o software utilizado para o projeto é de código livre, usando basicamente a linguagem Java em todos os módulos.

4.1.1 Netbeans

Para o desenvolvimento de software complexo é fundamental o uso de uma IDE (Integrated Development Environment) que reúna todas as bibliotecas necessárias para a compilação e implantação do sistema. Foi escolhida a IDE NetBeans <<https://www.netbeans.org>> devido à sua maturidade e estabilidade, o que facilita o trabalho de programação. Neste ambiente foram desenvolvidos os módulos de servidor, sistemas de testes de algoritmos e a interface web.

4.1.2 Android Studio

Atualmente é a principal ferramenta de desenvolvimento para a plataforma *Android*. Sendo uma IDE possui em si todas as bibliotecas necessárias para a criação, teste, depuração e implantação do aplicativo. A linguagem utilizada é Java embora exista suporte para outras linguagens como Kotlin e C++. A ferramenta pode ser baixada gratuitamente em <<https://developer.android.com>>.

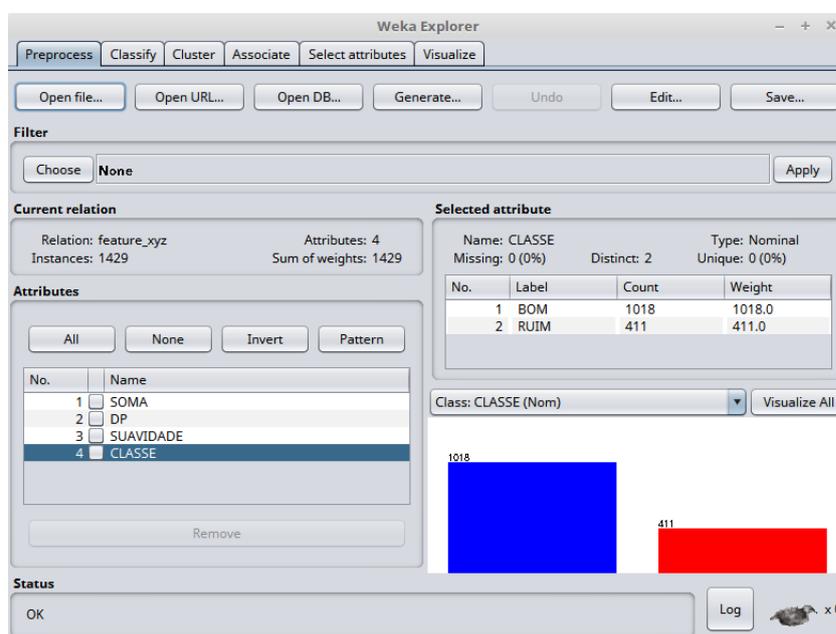
4.1.3 R-Studio

R é uma linguagem e também um ambiente de desenvolvimento voltado para análises estatísticas. Fortemente voltada para manipulação de conjuntos de dados, permite análises complexas e usa bibliotecas diversas para geração de gráficos. Neste projeto foi utilizada como ferramenta auxiliar para fazer cálculos estatísticos, testar features, gerar gráficos, validar clusters e manipular a massa de dados gerada no processo de coleta. Pode ser baixada tanto para ambiente Linux quanto Windows no website <<https://www.rstudio.com>>.

4.1.4 WEKA

WEKA (Waikato Environment for Knowledge Analysis) é um conjunto de bibliotecas e também de uma aplicação que permite execução de tarefas na área de *Machine Learning* e mineração de dados. Possui uma extensiva coleção de algoritmos implementados tornando mais fácil a tarefa de desenvolver projetos nessa área. Através da ferramenta Weka Explorer foi feita a análise de desempenho dos algoritmos de classificação tratada em Resultados.

Figura 18 – Interface do Weka



Fonte – Print do aplicativo

O WEKA pode ser baixado em <<https://www.cs.waikato.ac.nz/ml/weka/index.html>>.

4.1.5 Framework Vaadin

Para desenvolver em ambiente web o uso de frameworks aumenta a produtividade eliminando o retrabalho com a reutilização de componentes. O Vaadin é um conjunto de bibliotecas Java que abstrai toda a parte de programação Javascript e HTML em um único conjunto de componentes Java. O programador trabalha apenas com objetos Java tradicionais de maneira semelhante à interface gráfica Swing. São vendidos componentes avançados pela empresa que leva o mesmo nome do framework, existe também um portal de componentes gratuitos desenvolvidos pela comunidade. Neste projeto o uso do framework contribuiu para a construção do painel do usuário de maneira simples e direta.

4.2 Protótipo inicial

O primeiro protótipo deste projeto foi desenvolvido durante a disciplina Tópicos em Redes e Internet das Coisas no primeiro período de 2017 (REIS; NUNES; MOTTA, 2017). O sistema foi constituído em dois módulos distintos, sendo o primeiro por um hardware de captura utilizando uma placa *Raspberry Pi3*. Este hardware além de ter a capacidade de processamento adequada ao projeto, permitia a instalação de diversos tipos de sensores o que no caso em questão foram dois, um acelerômetro acoplado ao corpo da placa e um receptor de GPS instalado na parte externa da caixa de proteção.

Figura 19 – Primeiro protótipo do sistema



Fonte – Próprio autor

Este conjunto foi fixado ao painel do veículo usado durante os testes de modo que as vibrações sofridas durante o percurso pudessem ser capturadas pelo dispositivo. Embora a controladora permitisse a inclusão de um cartão de memória, optou-se por enviar os dados capturados via rede *wifi* ao módulo de processamento e armazenamento implementado em um aplicativo *Android*.

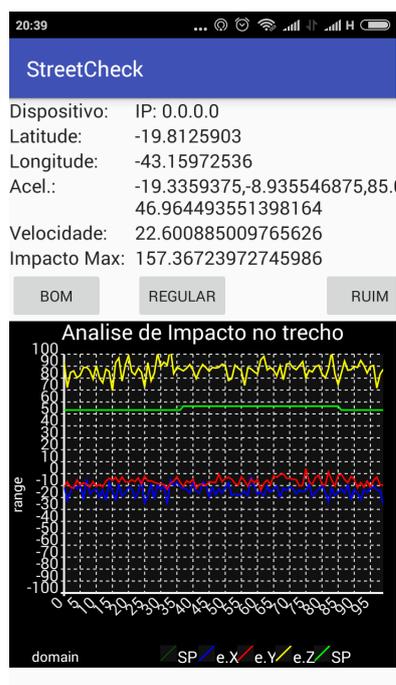
Figura 20 – Protótipo no painel do veículo



Fonte – Próprio autor

O aplicativo por sua vez recebia os pacotes de dados gerados pelo hardware de captura, fazia o armazenamento em arquivos de texto e atualizava um gráfico em tempo real com as informações obtidas. O usuário tinha a opção de classificar o trecho percorrido como "BOM", "REGULAR" ou "RUIM".

Figura 21 – StreetCheck 1.0, versão inicial



Fonte – Próprio autor

Os *features* armazenados nesta fase eram constituídos pelas leituras dos três eixos do acelerômetro e a velocidade momentânea calculada pelo GPS. As informações de latitude e longitude também eram armazenadas para futuro georreferenciamento dos dados.

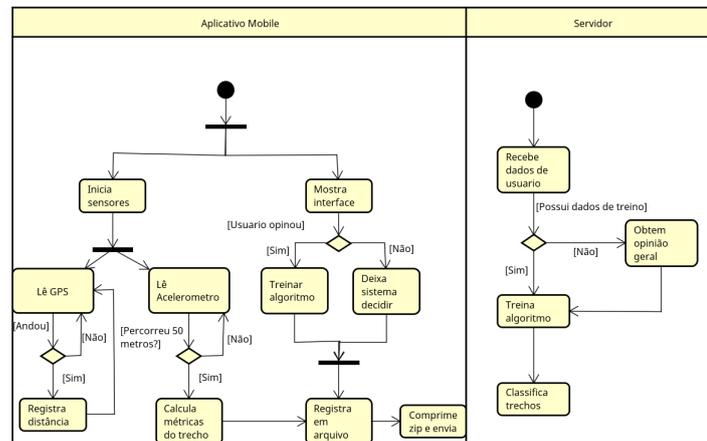
A continuidade desse projeto adaptou o modelo para funcionar exclusivamente em smartphone, uma vez que todos os sensores necessários, a infraestrutura de sistema bem como a capacidade de processamento e armazenamento já se encontram disponíveis no dispositivo. A opção de "REGULAR" foi removida por não ter uma correspondência relevante em relação aos dados coletados, tal questão será discutida mais adiante em 4.5.3.

4.3 Arquitetura do sistema

Do ponto de vista do usuário a execução do sistema se dá pela interface da aplicação. Contudo a interface é apenas o meio de interação do usuário com a aplicação, cabendo a uma camada de serviço em forma de *threads* a captura dos dados de acelerômetro e

localização. Sendo assim, o sistema executa de forma paralela diversas funções para a montagem do registro de dados.

Figura 22 – Diagrama da arquitetura do sistema. Arquitetura do aplicativo cliente-servidor.



Fonte – Próprio autor

Quando o aplicativo é aberto, são iniciadas as *threads* de monitoramento do GPS e obtenção de dados do acelerômetro. As informações são continuamente capturadas e armazenadas em arquivo e no momento que o usuário classifica algum trecho é criado um registro à parte que irá constituir os dados de treinamento do mesmo.

No ambiente do servidor é implementado um serviço de *FTP* responsável por receber os pacotes de dados provenientes dos clientes. Ao obter os dados, um algoritmo de classificação é treinado afim de se ter um modelo para classificação geral para o usuário em específico. Para melhorar a performance tanto do armazenamento quanto da transmissão de dados, todas as informações são gravadas diretamente em arquivos zip, por meio de uma biblioteca de compressão no próprio aplicativo.

4.4 Interface do aplicativo

Para facilitar a utilização do aplicativo por parte de usuários leigos optou-se por uma interface mais simples. Apenas as opções necessárias à coleta de dados são exibidas em forma de *radio buttons*.

Figura 23 – StreetCheck - versão atualizada



Fonte – Próprio autor

Os itens de interface utilizados foram os seguintes:

- Frase explicativa orientando o usuário a informar sua opinião sobre o trecho
- Duas opções em forma de ícone caracterizando "BOM" e "RUIM".
- Posicionamento do dispositivo. Afim de analisar possíveis alterações nos dados em função do local onde o dispositivo se encontra. Estes dados serão discutidos em Análise e discussão dos dados em 5.2.
- O tipo de veículo para fins de análise comparativa. Discutido mais adiante em Análise e discussão dos dados.
- Um componente do tipo *gauge* indicando a velocidade do veículo.
- Ícone de wifi indicando a presença de rede wireless.
- Checkbox para o usuário autorizar o sistema a enviar os dados. Dessa maneira o usuário pode escolher um momento adequado para o envio sem comprometer seu plano de dados.
- Identificador de usuário. Embora os dados sejam anônimos afim de evitar violação de privacidade, o usuário poderá consultar a classificação das vias com base em sua

percepção por meio deste ID. Não foi feita nenhuma implementação que associe o ID a um dispositivo em específico.

Um problema detectado durante a fase inicial de coleta de dados foi que o usuário ao marcar a opção "BOM" ou "RUIM" em um trecho, se esquecia de desmarcar a opção levando o sistema a marcar amostras subsequentes da mesma maneira, mesmo quando o perfil da via mudava. Esse problema foi resolvido introduzindo uma marcação de distância percorrida no sistema. Ou seja, no momento que o usuário seleciona uma opção, o sistema conta, a partir daquele ponto, uma distância de 50 metros e em seguida desmarca automaticamente a opção. Isso permitiu a coleta de amostras mais precisas.

4.5 Metodologia

Apenas smartphones *Android* foram contemplados no desenvolvimento deste projeto, tendo em vista o custo de desenvolvimento para outras plataformas como IOS. Em um primeiro momento foi desenvolvido o módulo de captura de dados, seguido da implementação do servidor. A parte final se deu na implementação dos algoritmos de classificação.

4.5.1 Veículos

Inicialmente o público-alvo eram os motoristas de carros particulares, contudo optou-se por permitir a classificação em ônibus tendo em vista que a grande maioria dos usuários de meios de transporte terrestres usam transporte público (65% segundo pesquisa feita pelo IPEA (BRASIL, 2011)). Assim o sistema considera dados capturados tanto em carros quanto em ônibus na base de dados. Fazendo uma análise dos conjuntos de dados ficou evidenciado que os dados ficam relativamente próximos nas amostras classificadas manualmente conforme a tabela:

Tabela 1 – Comparação de suavidade entre carros e ônibus.

	Carro	Onibus
Trecho bom	0.4811	0.5724
Trecho ruim	0.3536	0.3540

Fonte – Próprio autor

Pode-se observar em trechos de melhor qualidade a média de leituras no ônibus é um pouco maior que no carro, provavelmente em função da estrutura de suspensão do veículo. Já em trechos de qualidade inferior as médias ficaram praticamente iguais. Os veículos usados neste trabalho foram:

- Chevrolet Corsa Hatch

- Honda Civic
- Fiat Uno
- Mercedes Sprinter
- Scânia onibus intermunicipal (Gontijo)
- Mercedes onibus urbano (Enskon)

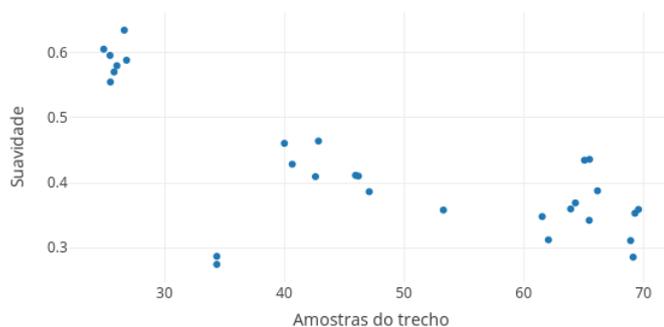
A título de curiosidade foi feito um teste no metrô de Belo Horizonte, imaginando que um veículo sobre trilhos poderia ter um índice de suavidade perto do máximo. O que foi observado, no entanto, é que o índice não passou de 0.82 pontos. O fator de suavidade é discutido mais adiante.

4.5.2 Velocidade de leitura

Considerando que o uso da aplicação será feito por um usuário comum, e que o mesmo nem sempre terá o interesse ou a possibilidade de manter uma velocidade constante durante o uso do aplicativo, optou-se por não fixar uma velocidade específica ao fazer a captura dos dados. Alguns trabalhos exigem que a coleta seja feita a uma velocidade fixa, ou que amostras capturadas a uma certa velocidade sejam comparadas apenas a outras coletadas na mesma velocidade (LIMA, 2016). Para superar este problema foi feita uma abordagem que levasse em consideração velocidades variáveis entre as amostras. Ao processar um segmento de pista é calculada a velocidade média entre todas as amostras do conjunto e a partir dessa velocidade é calculado o fator de suavidade.

Em (ALESSANDRONI et al., 2017) é feito um estudo sobre a influência da velocidade sobre o fator de imperfeição percebido ao trafegar na pista. Para cada tipo de terreno, com características distintas de asfalto existe um fator denominado *gamma law* que, combinados com características da suspensão do veículo, afetam a percepção das irregularidades da via. A relação pode ser observada em um experimento feito percorrendo um mesmo trecho em velocidades diferentes.

Figura 24 – Influência da velocidade sobre suavidade



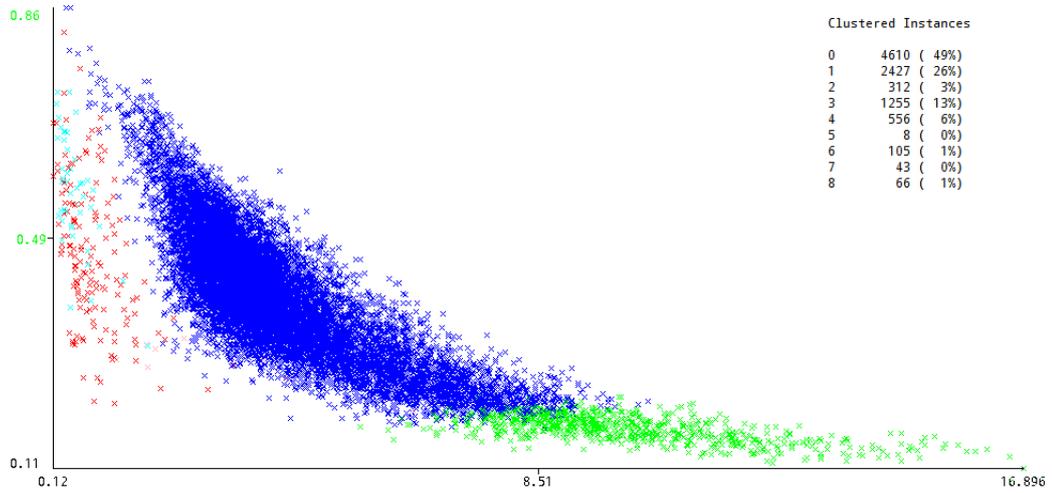
Fonte – Próprio autor

Como a velocidade acentua o impacto sentido as amostras apresentam suavidade decrescente. Os autores do estudo propuseram uma metodologia matemática para obter o perfil de cada trecho e assim determinar um fator de correção, o que pode ser visto no artigo citado. Uma desafio, porém, é replicar o mesmo estudo em vias brasileiras, que, diferentemente do contexto italiano (onde foi feito o estudo), na prática não exibem padrões de qualidade de primeiro mundo. Para procurar contornar este problema foram testadas features que diferenciassem as amostras quando apenas o fator de suavidade não fosse suficiente.

4.5.3 Classes de dados

Diferentemente da primeira versão do aplicativo, implementada em (REIS; NUNES; MOTTA, 2017) onde se podia escolher a opção "REGULAR", neste projeto só é possível escolher de forma binária entre "BOM" e "RUIM". Tal decisão foi tomada após ser feita uma análise dos clusters gerados por meio da técnica de curva de silhoueta (ROUSSEEUW, 1987). O método para esta análise foi primeiramente aplicar o algoritmo Kanopy (WHIZBANG; NIGAM; UNGAR, 2000) implementado no WEKA para estimar de maneira rápida a quantidade de clusters encontrados nos dados não classificados.

Figura 25 – Clusterização usando algoritmo Kanopy

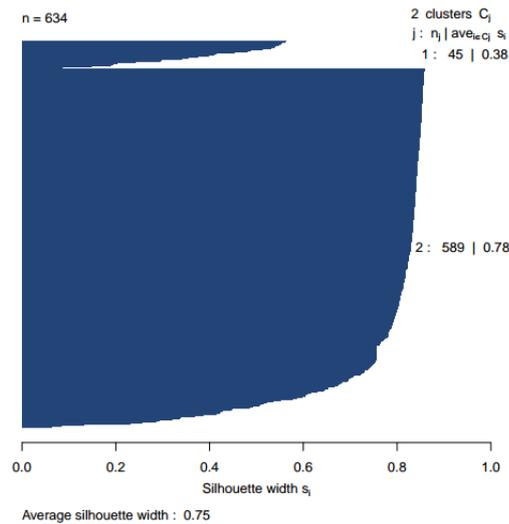


X=Velocidade em m/s, Y=Suavidade

Fonte – Próprio autor

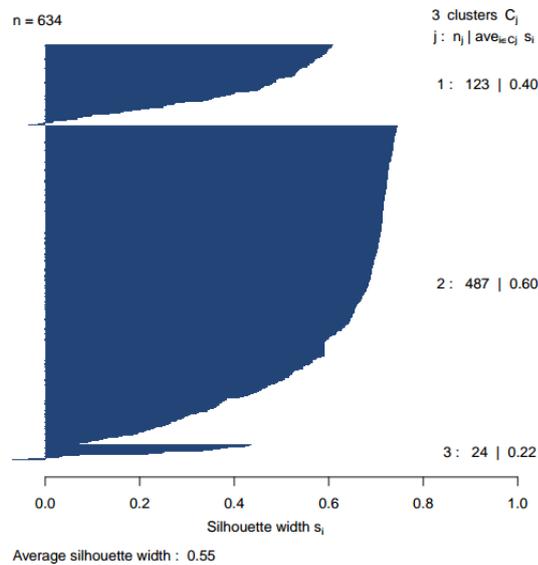
Como a quantidade de clusters encontrados pelo Kanopy não necessariamente expressa a melhor adequação de conjuntos (vide 2.5.1.2) para os dados, por meio do algoritmo *Silhouette* implementado no R-Studio foi feita uma classificação para os clusters.

Figura 26 – Silhoueta com dois clusters



Fonte – Próprio autor

Figura 27 – Silhoueta com três clusters



Fonte – Próprio autor

Observando as curvas o que pôde ser notado é que a medida que o número de clusters ia aumentando a qualidade da curva de silhoueta ia caindo conforme tabela abaixo.

Tabela 2 – Fator de silhoueta para clusters.

Clusters	Silhoueta
2	0.75
3	0.55
4	0.48
5	0.44
6	0.37
9	0.36
10	0.36
15	0.33
20	0.32

Fonte – Próprio autor

Por meio dos resultados justificou-se a utilização de apenas duas classes para o trabalho, ou seja "BOM" e "RUIM".

4.5.4 Features utilizadas

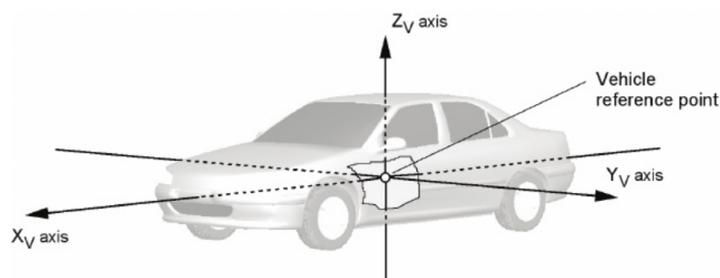
As *features* caracterizam um objeto de estudo em *Machine Learning*. Neste trabalho temos *features* geradas de forma instantânea, no momento da leitura do acelerômetro na forma de valores dos eixos X, Y e Z. Temos também *features* geradas no momento em que um trecho é fechado e processado, e por último *features* geradas no servidor, durante o processamento. Tais *features* caracterizam o conjunto de dados capturados em um intervalo

de espaço percorrido e seu objetivo é adicionar dimensões em que os dados possam ser diferenciados uns dos outros.

4.5.4.1 Eixos X, Y e Z

Considerando que o smartphone esteja no interior do veículo, as vibrações sofridas por este refletirão naquele. Contudo é necessário observar que neste trabalho não se especifica uma posição obrigatória para o smartphone. Em outras palavras não haverá uma correspondência direta entre os eixos relativos ao veículo e os eixos de captura dos sensores. Em outros trabalhos, como em (MENEGUZZI; TREIS; CENDRON, 2017) é utilizada a matriz de rotação para recalcular a posição relativa dos eixos e assim fazer o alinhamento que corresponda à posição do smartphone.

Figura 28 – Eixos com referencial no veículo



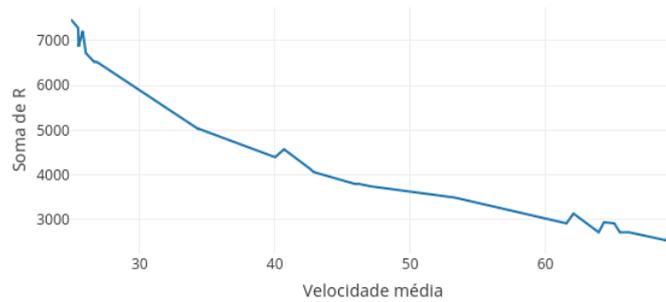
Fonte – (TERMINOLOGY, 2008, p. 8)

Contudo, neste trabalho, consideramos que a posição física do usuário (pode estar sentado, de pé, reclinado, etc) interfere na classificação de um trecho, e que não há como ter um controle dessa posição. Optou-se então por combinar as leituras coletadas em um único vetor chamado Resultante que é obtido através da soma vetorial dos três eixos a um custo computacional menor.

4.5.4.2 Soma

A cada trecho percorrido o sistema armazena a soma das leituras do vetor resultante. O objetivo é distinguir amostras de trechos coletadas em diferentes velocidades, uma vez que existe uma relação entre a velocidade e a quantidade de amostras lidas no trecho.

Figura 29 – Soma das leituras em relação à velocidade



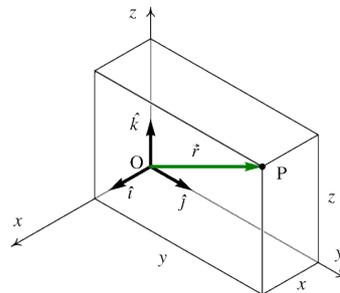
Fonte – Fonte: próprio autor

Estes dados foram coletados no mesmo trecho de via em velocidades diferentes. Observa-se, entretanto, que não se trata de uma relação perfeitamente linear pois a quantidade de leituras feitas por segundo depende do desempenho do dispositivo. Tal desempenho pode ser afetado por diversos fatores como bateria, temperatura, aplicativos sendo executados entre outros.

4.5.4.3 Vetor Resultante

O valor de resultante é uma *feature* que representa de forma vetorial a aceleração sofrida em um determinado ponto independente da posição do dispositivo. O valor é calculado fazendo a soma vetorial dos três eixos x,y e z de forma que o resultado é representado em módulo.

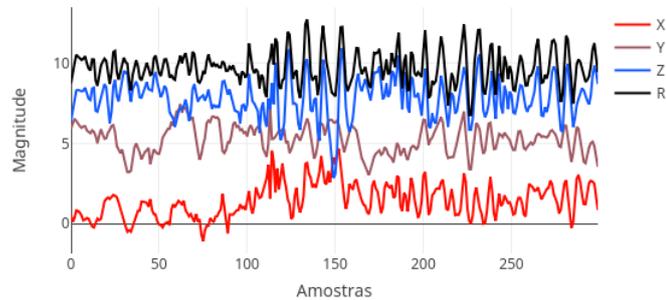
Figura 30 – Cálculo vetorial do vetor resultante



Fonte – Disponível em: <<https://goo.gl/jhCcMv>>

No gráfico que segue podemos observar uma comparação entre os três eixos juntamente com o resultante. O eixo R representa a soma vetorial de X, Y e Z em cada ponto do gráfico.

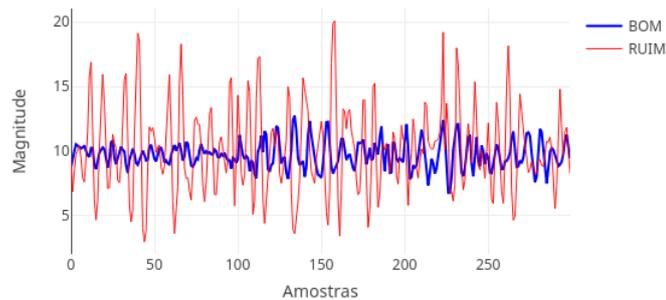
Figura 31 – Eixos x, y e z comparados ao resultante



Fonte – Próprio autor

Dessa maneira, pôde-se observar que este dado mostrou-se adequado para representar a quantidade total de aceleração sofrida em cada leitura executada. Como o valor estará sempre em módulo, não haverão valores abaixo de zero, o que facilita a leitura do gráfico. Cabe lembrar que neste trabalho não se leva em consideração os movimentos laterais e longitudinais sofridos em X e Y de forma separada, mas sim toda a mudança de aceleração da gravidade ocasionada por deformidades da pista.

Figura 32 – Comparação entre um trecho BOM e um trecho RUIM



Fonte – Próprio autor

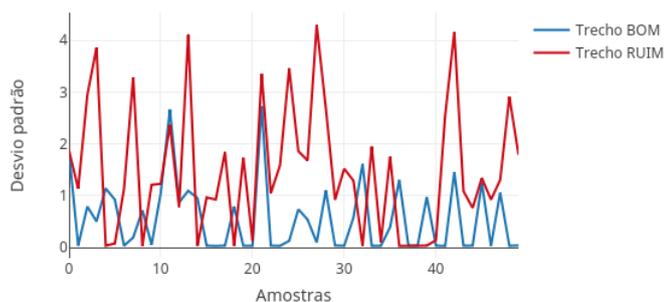
Conforme pode ser visto no gráfico, a diferença entre trechos bons e ruins é perceptível ao se comparar a magnitude das leituras. Sendo assim, neste projeto foi usado o valor do vetor resultante para os cálculos referentes à análise do pavimento viário.

4.5.4.4 Desvio Padrão

Ao percorrer um trecho de 50 metros o sistema calcula a média e o desvio padrão de todas as leituras. Indica o quão espalhadas estão em relação à média estão as amostras,

podendo indicar o grau de irregularidade da via. Em (LIMA, 2016) é utilizado como fator para classificação da via associado a valores limítrofes. Neste trabalho é utilizado como suporte para outros cálculos.

Figura 33 – Desvio padrão, comparação entre um trecho BOM e um trecho RUIM



Fonte – Próprio autor

Assim como na análise estatística leituras extremas afetam o desvio padrão, buracos e irregularidades acentuadas geram valores extremos no acelerômetro. Uma leitura pontual pode indicar a presença de um buraco de forma que na média geral do trecho o mesmo pode ser considerado bom. Contudo, uma sequência de irregularidades tornará os valores lidos espalhados em torno da média de maneira aleatória fazendo com o desvio padrão suba.

4.5.4.5 Suavidade

Suavidade foi definida como uma maneira de normalizar os dados coletados de forma independente da velocidade. Parte do princípio de que quanto mais regular e plana é a superfície, menor será o desvio padrão da amostra. Quanto mais próximo de 0 for o desvio padrão, menos o mesmo irá dividir a velocidade. Sendo assim, a suavidade quanto mais próxima de 1, mais plano e regular é o trecho percorrido e quanto mais próximo de 0 maior o número de irregularidades. Suavidade é dada pela fórmula:

$$S = \frac{vm}{1+dp} \quad (4.1)$$

Sendo vm a velocidade média no trecho e dp o desvio padrão calculado. Cabe lembrar que mesmo com a normalização a suavidade ainda é afetada pela velocidade, conforme discutido em (ALESSANDRONI et al., 2017). Essa correção, porém, não entrou no escopo deste trabalho por depender de outras pesquisas relacionadas à malha rodoviária brasileira.

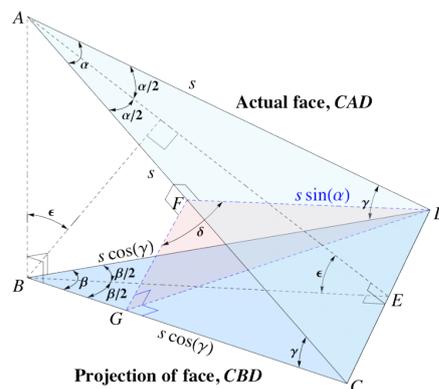
4.5.5 Features experimentais

Este conjunto de features foi elaborado para auxiliar o classificador a diferenciar amostras quando o fator de suavidade não for suficiente. Foram considerados aspectos estatísticos e geométricos.

4.5.5.1 Distorção Média - DISTM

Distorção Média (DISTM) Considera cada triângulo formado pelos eixos X,Y e Z como um plano e a partir disso calcula o ângulo entre planos subsequentes. A soma dos ângulos de todos os planos encontrados na via é dividida por 360 afim de se avaliar em termos de voltas completas o índice de distorção do trecho. Quanto menor este valor, mais plana é a via, com menos irregularidades.

Figura 34 – Angulo entre superfícies triangulares

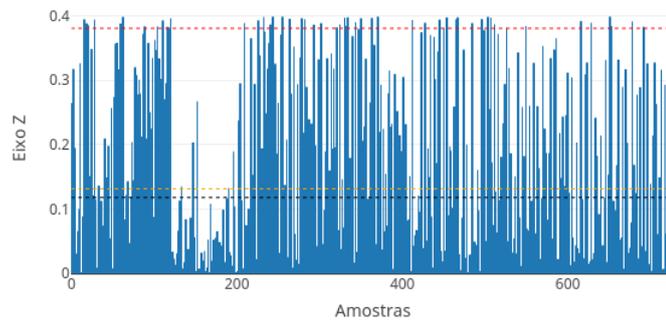


Fonte – Disponível em: <<http://www.gregegan.net/SCIENCE/Dihedral/Dihedral.html>>

4.5.5.2 Limites superiores e inferiores - QX1, QY1, QZ1, QX3, QY3, QZ3

Tendo por base a ideia de análise de quartis, onde as amostras são separadas em segmentos, o objetivo neste caso é observar o comportamento de um único eixo do acelerômetro dentro de limiares. Usa como referência a média e o desvio padrão como parâmetros para definir um limiar mínimo ou máximo.

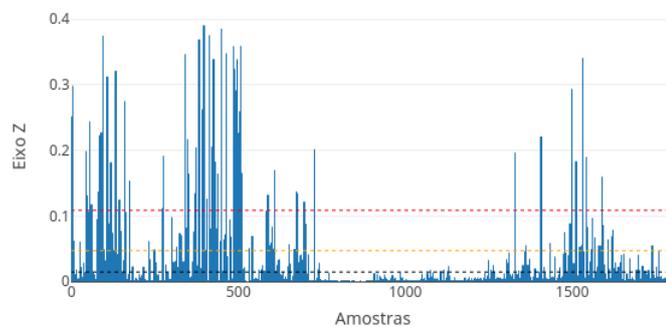
Figura 35 – Valor limiar para análise - Trecho BOM



Fonte – Próprio autor

O limiar é definido como a média mais duas vezes o desvio padrão. Para avaliar a relevância do atributo, foram executadas algumas rodadas com alguns algoritmos de classificação mais adiante em 5.4.

Figura 36 – Valor limiar para análise - Trecho RUIM



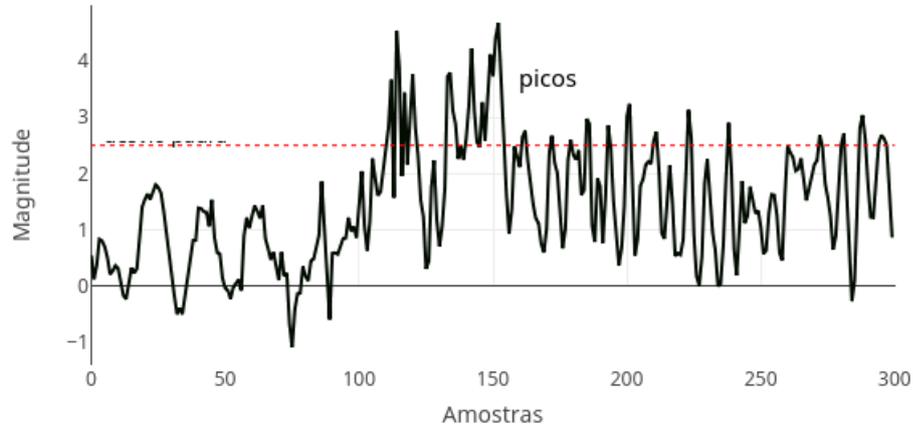
Fonte – Próprio autor

O objetivo é avaliar o volume dos dados encontrados nos primeiros e terceiros quartis partindo da hipótese que tais valores terão magnitudes diferenciadas em trechos ruins e bons.

4.5.5.3 Pico por segundo - PEAKS

O objetivo deste é diferenciar amostras em função da quantidade de picos lidos em um trecho em um intervalo de tempo. Foi definido um limiar assim como no item anterior e por esse limiar se definiu o que constituía um pico.

Figura 37 – Análise de picos/s



Fonte – Próprio autor

Contudo, como demonstrado em resultados, esse *feature* não possui boa representação contribuindo inclusive como fator de incremento no nível de ruído. Dessa maneira o mesmo foi descartado para uso neste projeto.

4.5.6 Filtragem de features

Para determinar as *features* mais relevantes, foi utilizado o algoritmo *ReliefF* que faz um ranking de acordo com o grau de diferenciação que determinado atributo produz nos dados. Após ter um volume de dados suficiente para testes foi feita a filtragem de atributos pelo WEKA. Segue o resultado:

Tabela 3 – Lista de atributos e relevância.

Atributo	Score
vm	0.03079
suavidade	0.01328
qx3	0.01228
qx1	0.01202
dp	0.01032
qz1	0.00949
qy1	0.00916
distm	0.00882
qy3	0.00835
qz3	0.00825
media	0.00795
peaks	0.00683

Fonte – Produzido pelo autor

O ranking depende da qualidade dos dados inseridos e também de fatores decorrentes da natureza do problema como, por exemplo, a questão do usuário inserir classificações conflitantes entre trechos muito parecidos. Contudo, como o objetivo de cada *feature* é mostrar um aspecto distinto de cada amostra, o conjunto destes aspectos é o que determina ao final a classe da amostra. Isso será discutido mais adiante em Resultados 5.

4.5.7 Dados de geolocalização

Utilizando o sensor de GPS do smartphone o sistema avalia a distância percorrida para fins de marcação de início e fim de um trecho de pista. O sistema monitora constantemente os dados de GPS e quando ocorre uma movimentação é calculada a velocidade. Quando, entre as leituras, é avaliado que se percorreu 50 metros o sistema armazena os pontos de início e fim do movimento definindo um segmento de pista.

Figura 38 – Segmento de trecho em estrada



Fonte – Próprio autor

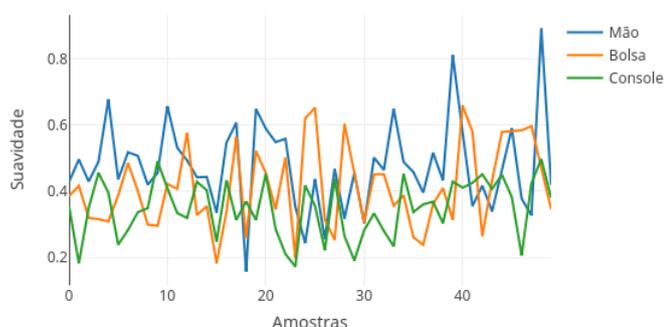
Pode-se observar na imagem contudo, que os segmentos de pista traçados são de tamanho irregular. Isso ocorre em função da falta de precisão do módulo de GPS dos smartphones, especialmente em áreas rurais. Um outro desafio dessa abordagem de mapeamento é o fato de gerar inúmeros trechos sobrepostos uma vez que não se tem o controle de em que ponto o usuário irá acionar o aplicativo. Sendo assim o mesmo trecho poderia ser registrado no banco de dados inúmeras vezes variando pequenas unidades na latitude e longitude. O método utilizado para resolver este problema foi diminuir o número de casas decimais de 8 para 4. Com 4 casas decimais a precisão do GPS vai para 11 metros, o que já é suficiente para identificar uma via por meio de duas coordenadas.

Para casos de pistas de mão dupla o sentido do veículo no momento da coleta também conta. Este problema foi resolvido ao se armazenar as coordenadas em forma de duas tuplas (LAT,LON) contendo a latitude e longitude dos pontos iniciais e finais da coleta. Considerando um ponto A e B da trecho, em um sentido serão armazenadas as coordenadas AB e, no sentidos oposto, BA.

4.5.8 Posicionamento do dispositivo

Na interface do aplicativo o usuário pode definir onde se encontra o smartphone no momento da coleta dos dados. Para efeito de comparação de dados foram dadas as opções de MÃO, CONSOLE, BOLSA e BOLSO (este último não entrou na análise por não haver amostras relevantes). Conforme pode ser visto no gráfico, não há muita diferença se o dispositivo estando na bolsa ou no console tendo uma média entre 0.35 e 0.36 da média de suavidade respectivamente. Quando se encontra na mão pôde-se observar que a média subiu para 0.48 pontos, o que poderia ser explicado pela atenuação causada pelo próprio usuário ao manipular o dispositivo.

Figura 39 – Variação da suavidade conforme posição



Fonte – Próprio autor

4.5.9 Servidor

Para este projeto foi contratado um serviço de computação em nuvem na forma de um servidor virtual. Serviços de virtualização de servidor permitem que se tenha uma máquina disponível, com sistema operacional próprio totalmente dedicada ao cliente. É uma modalidade mais barata uma vez que o hardware sendo virtual diversos clientes dividem os recursos de um servidor físico de grande porte de maneira completamente isolada. São essas a especificação do servidor:

Tabela 4 – Especificações do ambiente de sistema.

Endereço da aplicação web	http://191.252.184.157:8080/StreetCheck
Sistema operacional	Ubuntu Linux Server - 16.04
Memoria RAM	1G
Tamanho do HD	40G
Endereço IP	191.252.184.157
Servidor de Banco de dados	Mysql
Servidor de aplicações	JBoss - Wildfly-11

Fonte – Próprio autor

Nessa configuração o servidor não é capaz de suportar um grande número de usuários, porém os serviços de virtualização contratados permitem uma rápida expansão de recursos caso seja necessário.

4.5.10 Classificadores

Durante a utilização do aplicativo o sistema divide o conjunto de dados capturados em duas partes. A primeira é composta pelas informações de trecho que o usuário fornece ao marcar a opção "BOM" ou "RUIM", neste caso são os dados de treinamento. O segundo conjunto de dados, normalmente maior que primeiro, é composto por todos os registros que o usuário não classifica, obtidos automaticamente pelo sistema, e ocorrem normalmente quando o usuário ativa o aplicativo e guarda o celular. Estes dados, não classificados, formam a base pública geral. Pode ocorrer de um usuário não fornecer dados de treinamento e nesse caso o sistema irá utilizar um conjunto aleatório de amostras de outros usuários.

Tendo esses dois conjuntos de dados o objetivo dos classificadores é, por meio dos dados de treinamento, classificar os dados gerais de acordo com o conceito de cada usuário. A saída deste processo é dada em forma de mapa utilizando a API do *Google Maps*.

O classificador utilizado pelo StreetCheck foi o *Random Forest* implementado na biblioteca do WEKA. Durante os testes de desempenho os resultados variaram pouco entre os algoritmos de classificação, fato que pode ser explicado devido à natureza dos dados e do volume de classificações distintas entre amostras próximas.

Tabela 6 – Distâncias percorridas e posicionamento do smartphone

Tipo de veículo	Posição do smartphone	Km
CARRO	BOLSA	76,30
CARRO	BOLSO	25,78
CARRO	CONSOLE	664,24
CARRO	MAO	269,82
ONIBUS	BOLSA	26,04
ONIBUS	CONSOLE	56,22
ONIBUS	MAO	99,38

Fonte – Próprio autor

Obs. Neste trabalho quando se registrou "CONSOLE" no ônibus referiu-se a um espaço na janela onde foi possível fixar o smartphone. Em carros refere-se ao painel dianteiro do mesmo.

5.2 Análise e discussão dos dados

Os trechos que os usuários classificaram, onde cada amostra corresponde a 50 metros de via, foram de 187,82 Km entre as cidades já mencionadas.

O maior desafio encontrado foi encontrar um conjunto de *features* que criasse uma boa linha de separação entre as amostras, especialmente quando em velocidades maiores. Conforme visto nos testes e também confirmado em (ALESSANDRONI et al., 2017) a velocidade cria um fator de ruído intensificando a percepção de impacto.

Do total de trechos percorridos e classificados, os que se tornam dados de treinamento, até o momento 63,72% foram de trechos considerados bons e 36,28% ruins. Alguns poucos trechos (0,6%) foram registrados mais de uma vez na base de treinamento com reclassificação, ou seja, quando a opinião muda de bom para ruim ou vice-versa. Espera-se, contudo, que este volume cresça a medida que mais usuários utilizarem o sistema, de forma que com efeito trechos cuja opinião forem mudando de bom para ruim poderão ser indicativos de deterioração da via.

Tabela 7 – Distâncias percorridas e posicionamento do smartphone em dados classificados

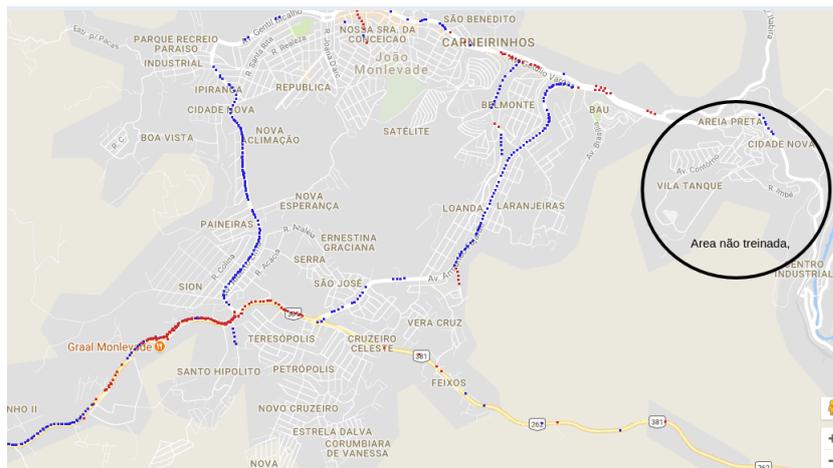
Tipo de veículo	Posição do smartphone	Km
CARRO	BOLSA	18,76
CARRO	BOLSO	0,30
CARRO	CONSOLE	89,30
CARRO	MÃO	80,30
CARRO	PAINEL	49,58
ÔNIBUS	CONSOLE	2,16
ÔNIBUS	MÃO	5,44

Fonte – Próprio autor

5.3 Classificação e visualização

Na prática, o funcionamento da aplicação como resposta ao usuário pode ser visto nos mapas de classificação. É possível observar que o usuário classificou apenas alguns trechos no mapa. Pontos em azul representam trechos bons e em vermelho ruins. Supondo que um usuário A classificou os trechos destacados no mapa.

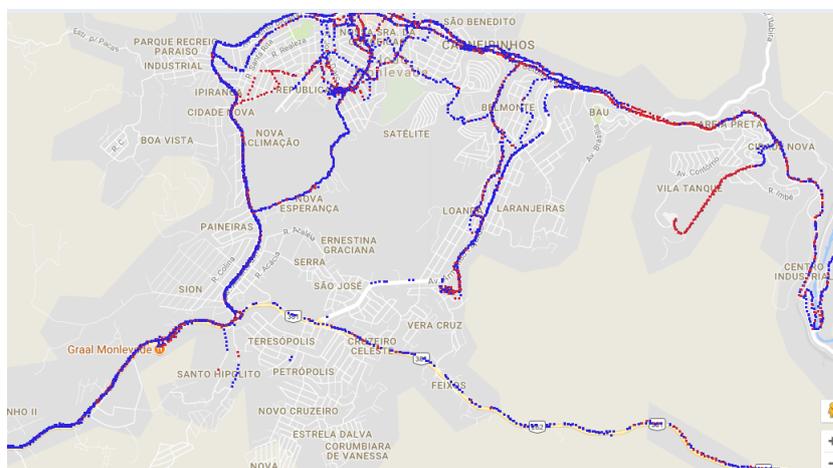
Figura 41 – Mapa referente ao usuário 1



Fonte – Próprio autor

Com os dados coletados pela comunidade de usuários foi possível classificar trechos que o usuário ainda não tinha percorrido, de modo que ao consultar o mapa ele toma conhecimento do estado de áreas que não visitou.

Figura 42 – Mapa referente gerado com dados da comunidade



Fonte – Próprio autor

Como sistema colaborativo o StreetCheck permite que uma comunidade de usuários alimente uma base de dados global acessível a todos.

5.4 Comparação dos algoritmos

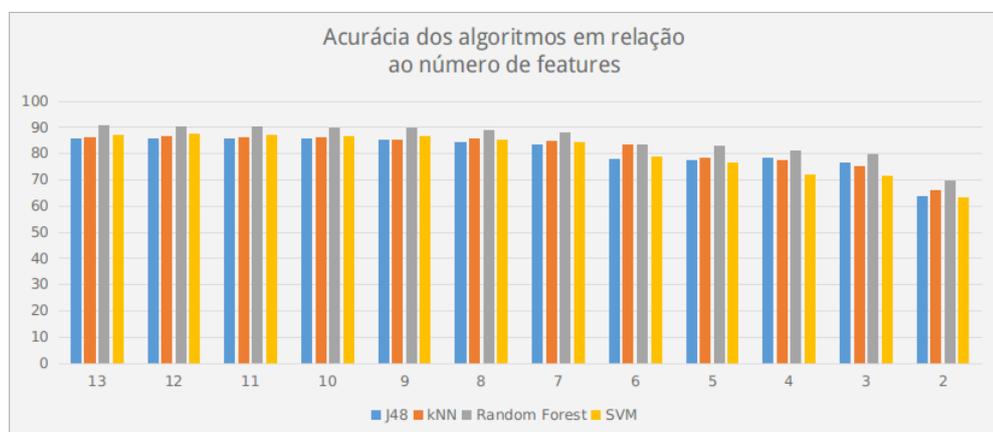
Em um primeiro teste de classificação, foram utilizados o KNN, o Random Forest, J48 e o SVM. Foram utilizadas 9.382 amostras classificadas do banco de dados afim de testar o desempenho de cada algoritmo.

Algoritmo	Acurácia
Random Forest	90,64%
SVM	87,02%
Knn 5 vizinhos	86,05%
J48	85,71%

Fonte – Próprio autor

Como se pode observar o classificador com o melhor desempenho foi o Random Forest. Todos os classificadores utilizados neste sistema fazem parte da biblioteca do WEKA. O desempenho de cada classificador depende, contudo, do conjunto de *features* utilizado. Individualmente os features podem não dizer muito sobre uma amostra, mas a combinação destes contribui para o desempenho da classificação. Para demonstrar este fato, no conjunto utilizado foram feitos repetidos testes sendo que a cada rodada foi removido um atributo considerando o grau de relevância do atributo conforme mostrado na tabela de relevância (3) vista anteriormente. Foi gerada então a seguinte tabela indicando o efeito da remoção de features.

Figura 43 – Acurácia vs Features



Fonte – Próprio autor

Através da análise, além da curva descendente, o que se pode perceber é a existência de features que, dependendo do algoritmo utilizado, reduzem a acurácia ao invés e aumentam-na como no caso da feature "PEAKS". A ordem de remoção dos atributos seguiu a recomendação feita pelo algoritmo *ReliefF*, sendo do menos importante para os mais importantes: VM, SUAVIDADE, QX3, QX1, DP, QZ1, QY1, DISTM, QY3, QZ3, MEDIA, PEAKS

Tabela 8 – Remoção de features e acurácia

	J48	kNN	R.F.	SVM
PEAKS	85,67*	86,04*	90,64	87,02
MEDIA	85,72	86,31	90,31	87,20*
QZ3	85,39*	85,85*	89,97	87,07
QY3	85,51	86,07	89,82	86,63
DISTM	85,35	85,35	89,60	86,43
QY1	84,29	85,55	88,64	85,17
QZ1	83,42	84,55	87,83	84,00
DP	77,77	83,30	83,11	78,93
QX1	77,56*	78,10	82,75	76,49
QX3	78,07	77,24	81,17	71,91
SUAVIDADE	76,47	75,22	79,64	71,30
VM	63,71	65,86	69,53	63,15

Fonte – Próprio autor

Os algoritmos J48 e kNN sofreram o efeito de um número elevado de atributos interferindo em sua acurácia. No caso do SVM, em um único caso a remoção do atributo QY1 aumentou a acurácia. O mesmo não aconteceu para o Random Forest em que se observou a redução da acurácia em todos os casos.

6 Conclusão

Este trabalho apresentou o StreetCheck como um sistema colaborativo que permite aos usuários classificarem vias de acordo com seus critérios de qualidade. Este conceito, diferentemente dos trabalhos citados, permite uma classificação subjetiva que leve em conta outros fatores além do IRI e correlatos. Partindo do princípio que diferentes usuários podem ter demandas diferenciadas no trânsito este sistema permite acrescentar este fator na equação. O desafio contudo não é simples, uma vez que sistemas dessa natureza dependem de muitos fatores e estão sujeitos a diversos tipos de ruído na informação. A própria subjetividade do usuário que pode classificar de maneira diversa coisas iguais pode representar o pior dos ruídos. Considerando esses fatores entendemos que o StreetCheck em sua forma é válido como sistema de auxílio na tomada de decisão.

A interface de uso no smartphone foi o mais simples possível, de forma que o usuário possa usar intuitivamente. Alguns detalhes estéticos podem ser melhorados, mas por se tratar de uma primeira versão foi suficiente para cumprir o objetivo do trabalho. O recurso de desligamento automático da amostra foi muito útil para diminuir o número de classificações incorretas durante a fase de teste, uma vez que o usuário ligava o *checkbox* e esquecia de desligar.

O processo de coleta de dados do acelerômetro e GPS se mostrou funcional. Entretanto, uma dificuldade encontrada foi que em alguns locais o sistema perdia o sinal de GPS e tal ocorrência não era notificada no sistema acarretando em perda da coleta em alguns trechos. A compressão dos dados se mostrou indispensável devido ao volume de dados gerado além de melhorar o desempenho da rede. Em relação ao pré-processamento no smartphone é recomendável que se faça apenas o básico para poupar recursos deixando o processamento de *features*, especialmente as experimentais, no servidor uma vez que estas demandam operações de ordenação, cálculos estatísticos e geométricos que teriam um custo computacional elevado para um smartphone.

Por seguir um modelo colaborativo o uso por parte da população de usuários pode, desde que seja contínuo, ir acertando distorções que eventualmente podem ocorrer na classificação dos dados. Mesmo usuários que não colaboram de forma direta fornecendo amostras de treinamento podem enriquecer a base de dados apenas mantendo o aplicativo em execução que enviaria para o servidor informações de trechos que posteriormente serão úteis para os demais usuários.

Os algoritmos de aprendizagem e classificação vistos são válidos para o uso da plataforma, o que se precisa melhorar é qualidade dos features para elevar o nível de classificação. A performance do StreetCheck nesse quesito foi de 90.64% de acertos, o que

pode ser considerado interessante levando em conta que as condições de uso do StreetCheck são mais amplas (posicionamento do dispositivo indiferente, usar tanto em carros quanto em ônibus), o que incrementa o nível de ruído na informação.

6.1 Trabalhos futuros

O StreetCheck, mesmo estando em uma fase inicial, apresenta resultados interessantes. Contudo algumas coisas precisam ser melhoradas como aplicar o fator de regularização para a suavidade em relação à velocidade. O efeito seria uma melhora no mecanismo de classificação que teria clusters mais afastados, facilitando bastante a tarefa do classificador.

Alguns aparelhos de celular apresentaram instabilidades devido à diferentes versões do Android e requisitos de segurança. Em relação à privacidade não existe nada que relacione o ID do usuário ao número do telefone ou a qualquer informação que o ligue ao dono do aparelho, contudo tendo em mãos o celular do usuário seria possível a alguém com acesso ao banco de dados rastrear os locais onde o mesmo classificou algum trecho. Isso pode ser resolvido mantendo os dados de treinamento unicamente no smartphone enviando os mesmos para o servidor apenas quando o usuário solicitasse a visualização do mapa. Após a renderização e visualização os dados seriam eliminados, o que não é possível no atual modelo que concentra todos os dados no servidor.

Melhorar a interface do aplicativo embutindo no mesmo o mapa. Dessa maneira toda a informação relevante para o usuário estaria em uma única interface, deixando o servidor apenas para a parte de processamento pesado.

Implementar o armazenamento de dados em um banco de dados com suporte a objetos geográficos. O PostgreSQL com a extensão Postgis é uma sugestão. Um banco de dados geográfico possibilita consultas georreferenciadas tornando mais eficiente a forma como os dados são filtrados.

Referências

- ALESSANDRONI, G. et al. A Study on the Influence of Speed on Road Roughness Sensing: The SmartRoadSense Case. *Sensors*, v. 17, n. 2, p. 305, feb 2017. ISSN 1424-8220. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/28178224><http://www.pubmedcentral.nih.gov/articlerender.fcgi?artid=PMC5336061><http://www.mdpi.com/1424-8220/17/2/305>>. Citado 3 vezes nas páginas 45, 52 e 60.
- ANDREONI, M. An intrusion detection and prevention architecture for software defined networking. 2014. Disponível em: <https://www.researchgate.net/publication/272220901_A_Proposal_of_Defense_Service_Against_Distributed_Denial_of_Service_Attacks>. Citado na página 30.
- Apache Software Foundation. *Canopy Clustering*. 2005. Disponível em: <<https://mahout.apache.org/users/clustering/canopy-clustering.html>>. Citado na página 27.
- BALBINOT, A. *Caracterização dos níveis de vibração em motoristas de ônibus : um enfoque no conforto e na saúde*. 311 p. Tese (Doutorado) — Universidade Federal do Rio Grande do Sul, 2001. Disponível em: <<http://www.lume.ufrgs.br/bitstream/handle/10183/2482/000320380.pdf?sequence=1>>. Citado na página 20.
- BERNUCCI, L. B. *Programa de Integração e Capacitação aos Engenheiros e Técnicos admitidos pelo DER/PR*. [S.l.], 2008. Citado na página 21.
- BERNUCCI L.B, d. M. M. L. C. J. S. J. *Pavimentação asfáltica - Formação Básica para Engenheiros*. [S.l.]: Petrobras Asfaltos, 2010. Citado na página 21.
- BISCONSINI, D. et al. *ANÁLISE ESPECTRAL DOS DADOS COLETADOS POR SMARTPHONES PARA A AVALIAÇÃO DA IRREGULARIDADE LONGITUDINAL DOS PAVIMENTOS*. 2016. Acesso em: 2018/01/05. Citado 2 vezes nas páginas 23 e 35.
- BISCONSINI, D. R.; JÚNIOR, J. L. F. O Uso De Smartphones Para a Avaliação Da Irregularidade Longitudinal De Pavimentos. n. December, p. 129–132, 2015. Citado na página 23.
- BISHOP, C. M. *Pattern Recognition and Machine Learning*. [S.l.: s.n.], 2013. v. 53. 1689–1699 p. ISSN 1098-6596. ISBN 978-0-387-31073-2. Citado na página 17.
- BRASIL, A. *Estudo do Ipea mostra que 65% da população usam transporte público nas capitais — Governo do Brasil*. 2011. Disponível em: <<http://www.brasil.gov.br/governo/2011/05/estudo-do-ipea-mostra-que-65-da-populacao-usam-transporte-publico-nas-capitais>>. Citado na página 44.
- BREIMAN, L. Random forests. *Machine Learning*, v. 45, n. 1, p. 5–32, 2001. ISSN 1573-0565. Disponível em: <<https://doi.org/10.1023/A:1010933404324>>. Citado 2 vezes nas páginas 32 e 33.
- BROWNLEE, J. *Master Machine Learning Algorithms: Discover how They Work and Implement Them from Scratch*. [s.n.], 2016. Disponível em: <<https://books.google.com.br/books?id=PCJnAQAAAJ>>. Citado na página 30.

- BURKE, J. et al. Participatory sensing. In: *In: Workshop on World-Sensor-Web (WSW'06): Mobile Device Centric Sensor Networks and Applications*. [S.l.: s.n.], 2006. p. 117–134. Citado na página 17.
- CALACHE, D. C. *Caracterização de um Acelerômetro Baseado em Sistemas Microeletromecânicos (MEMS)*. Tese (Doutorado), 2013. Disponível em: <<http://www.lee.eng.uerj.br/~jpaulo/PG/2013/PG-Acelerometro-MEMS-2013.pdf>>. Citado na página 22.
- CANTISANI, G.; LOPRENCIPE, G. Road Roughness and Whole Body Vibration: Evaluation Tools and Comfort Limits. *Journal of Transportation Engineering*, v. 136, n. September, p. 818–826, 2010. ISSN 0733-947X. Disponível em: <https://www.academia.edu/16793124/Road{_}Roughness{_}and{_}Whole{_}Body{_}Vibration{_}Evaluation{_}Too>. Citado na página 20.
- ERIKSSON, J. et al. The pothole patrol: using a mobile sensor network for road surface monitoring. In: ACM. *Proceedings of the 6th international conference on Mobile systems, applications, and services*. [S.l.], 2008. p. 29–39. Citado 2 vezes nas páginas 17 e 35.
- FREITAS, M. P. FACULDADE BATISTA DO ESTADO DO RIO DE JANEIRO PÓS-GRADUAÇÃO EM REDES DE COMPUTADORES. p. 21, 2015. Disponível em: <<http://www.faberj.edu.br/downloads/biblioteca/tcc/pos/MaykoPetersen-FABERJ-SensoriamentoParticipativo-ComFichaCatalografica.pdf>>. Citado na página 23.
- GANTI, R.; YE, F.; HEI, H. Mobile crowdsensing: Current state and future challenges. *IEEE Communications Magazine*, v. 49, n. 11, p. 32–39, 2011. ISSN 0163-6804. Disponível em: <<http://www.scopus.com/inward/record.url?eid=2-s2.0-81355138524{&}partnerID=40{&}md5=4c0b9f4b26ffd71cd8915cbea3>>. Citado na página 23.
- GARBOWSKI, T.; GAJEWSKI, T. Semi-automatic Inspection Tool of Pavement Condition from Three-dimensional Profile Scans. *Procedia Engineering*, v. 172, p. 310–318, 2017. Citado na página 17.
- GONZÁLEZ, L. C. et al. Learning roadway surface disruption patterns using the bag of words representation. *IEEE Transactions on Intelligent Transportation Systems*, PP, n. 99, p. 1–13, 2017. ISSN 1524-9050. Citado na página 35.
- HOUEI, D. L. *Uso do ReliefF para Seleção de Atributos em Dados Multirrotulo*. 182 p. Tese (Doutorado), 2005. Disponível em: <<http://sites.labic.icmc.usp.br/pub/mcmonard/SpolaorCLEI2011.pdf>>. Citado na página 34.
- ILOS. *Transporte de cargas e a encruzilhada do Brasil para o futuro*. 2017. Disponível em: <<http://www.ilos.com.br/web/cenario-da-infraestrutura-rodoviaria-no-brasil/>>. Citado na página 16.
- LIMA, L. C. de. *RoadScan: Sistema Crowdsourcing para Identificação de Qualidade de Pavimentos Asfálticos*. 2016. Monografia apresentada ao curso de Engenharia de Computação do Departamento de Computação e Sistemas da Universidade Federal de Ouro Preto como requisito parcial para obtenção do grau de Bacharel em Engenharia de Computação. Disponível em: <http://www.monografias.ufop.br/bitstream/35400000/226/1/MONOGRAFIA_RoadscanSistemaCrowdsourcing.pdf/>. Citado 4 vezes nas páginas 17, 36, 45 e 52.

- LORENA, A. C. *Revista de Informática Teórica e Aplicada*, 2007. ISSN 00978418. Citado na página 33.
- MEDNIS, A. et al. Real time pothole detection using Android smartphones with accelerometers. In: IEEE. *Distributed Computing in Sensor Systems and Workshops (DCOSS), 2011 International Conference on*. [S.l.], 2011. p. 1–6. Citado na página 35.
- MENEGUZZI, L.; TREIS, R. F.; CENDRON, M. M. Utilização de giroscópio e acelerômetro para identificação de movimentação em ambientes tridimensionais. 2017. Disponível em: <<http://videira.ifc.edu.br/fice/wp-content/uploads/sites/27/2016/09/10-Artigo-UTILIZAÇÃO-DE-GIRO>>. Citado na página 49.
- NILSSON, N. J. Introduction to Machine Learning. *Machine Learning*, v. 56, n. 2, p. 387–99, 2005. ISSN 10959572. Disponível em: <<http://www.ncbi.nlm.nih.gov/pubmed/21172442>>. Citado 2 vezes nas páginas 25 e 26.
- NOTÍCIAS, A. C. de. *Pesquisa CNT indica piora da qualidade das rodovias*. 2017. Disponível em: <<http://www.cnt.org.br/Imprensa/Noticia/pesquisa-cnt-indica-piora-qualidade-rodovias>>. Acesso em: 2017/12/26. Citado 3 vezes nas páginas 8, 9 e 16.
- REIS, D. M.; NUNES, D. E.; MOTTA, V. F. StreetCheck: Sensoriando e Inferindo a Qualidade de Vias Urbanas. In: *IV WORKSHOP DE INICIAÇÃO CIENTÍFICA EM SISTEMAS DE INFORMAÇÃO*. [s.n.], 2017. p. 21–24. Disponível em: <<http://sbsi2017.dcc.ufla.br/download/wicsi.pdf>>. Citado 3 vezes nas páginas 18, 40 e 46.
- RODRIGUES, A. *Introdução aos Sistemas de Transporte no Brasil e à Logística Intermunicipal*. [S.l.]: Aduaneiras, 2002. ISBN 9788571292390. Citado 3 vezes nas páginas 8, 9 e 16.
- ROUSSEEUW, P. J. Silhouettes: A graphical aid to the interpretation and validation of cluster analysis. *Journal of Computational and Applied Mathematics*, v. 20, n. C, p. 53–65, 1987. ISSN 03770427. Citado 2 vezes nas páginas 28 e 46.
- SAYERS, M. W.; KARAMIHAS, S. M. The little book of profiling. *Basic Information about Measuring and Interpreting Road Profiles*, n. September, p. 100, 1998. Disponível em: <<http://www.umtri.umich.edu/content/LittleBook98R.pdf>>. Citado na página 17.
- STANDARDIZATION, I. O. for. *Guia para avaliação da exposição humana à vibrações de corpo inteiro*. 1978. Disponível em: <<http://www.criffer.com.br/uploads/ISO-2631.pdf>>. Citado 3 vezes nas páginas 8, 9 e 20.
- TERMINOLOGY, V. D. Surface vehicle recommended practice. v. 4970, p. 724–776, 2008. Citado 2 vezes nas páginas 22 e 49.
- TOTALPAVE. *TotalPave - Pavement Management for the 21st Century*. Disponível em: <<http://totalpave.com/>>. Citado na página 37.
- TRANSPORTE, C. N. do. *Anuário CNT de transportes*. 2017. Disponível em: <<http://www.cnt.org.br>>. Citado na página 16.

TRANSPORTES, D. N. de Infraestrutura de. *Veículo de Diagnóstico de Rodovias*. 2017. 24 p. Disponível em: <<http://www.dnit.gov.br/download/veiculo-de-diagnostico-de-rodovias.pdf>>. Citado 2 vezes nas páginas 21 e 22.

TRT Santa Catarina. *ACÓRDÃO TRT 12ª / SECRETARIA DA 3ª TURMA / 2013-07-01 - 0007151-73.2011.5.12.0001,07151-2011-001-12-00-0,254915/13 :: Jurisprudência::Acórdão 0007151-73.2011.5.12.0001,07151-2011-001-12-00-0,254915/2013 (Federal::Judiciário::Tribunal Regional do Trabalho::3ª Turma - 12ª Região - Santa Catarina) ::*. 2013. Disponível em: <<http://www.lexml.gov.br/urn/urn:lex:br:justica.trabalho;regiao.12:tribunal.regional.trabalho;turma.3:acordao:2013-06-18;0007151-73.2011.5.12.0001,07151-2011-001-12-00-0,254915>>. Citado na página 16.

Von Zuben, F. J.; ATTUX, R. R. F. Árvores de Decisão. 2007. Disponível em: <ftp://ftp.dca.fee.unicamp.br/pub/docs/vonzuben/ia004{_}1s10/notas{_}de{_}aula/topico7{_}IA>. Citado na página 32.

WANG, H.-W. et al. A Real-Time Pothole Detection Approach for Intelligent Transportation System. *Mathematical Problems in Engineering*, Hindawi, v. 2015, aug 2015. ISSN 1024-123X. Disponível em: <<https://www.hindawi.com/journals/mpe/2015/869627/>>. Citado na página 17.

WHIZBANG, A. M.; NIGAM, K.; UNGAR, L. H. Efficient clustering of high-dimensional data sets with application to reference matching. 2000. Disponível em: <<http://www.kamalnigam.com/papers/canopy-kdd00.pdf>>. Citado 2 vezes nas páginas 26 e 46.

Apêndices

APÊNDICE A – Materiais elaborados pelo autor

Figura 44 – Diagrama do banco de dados

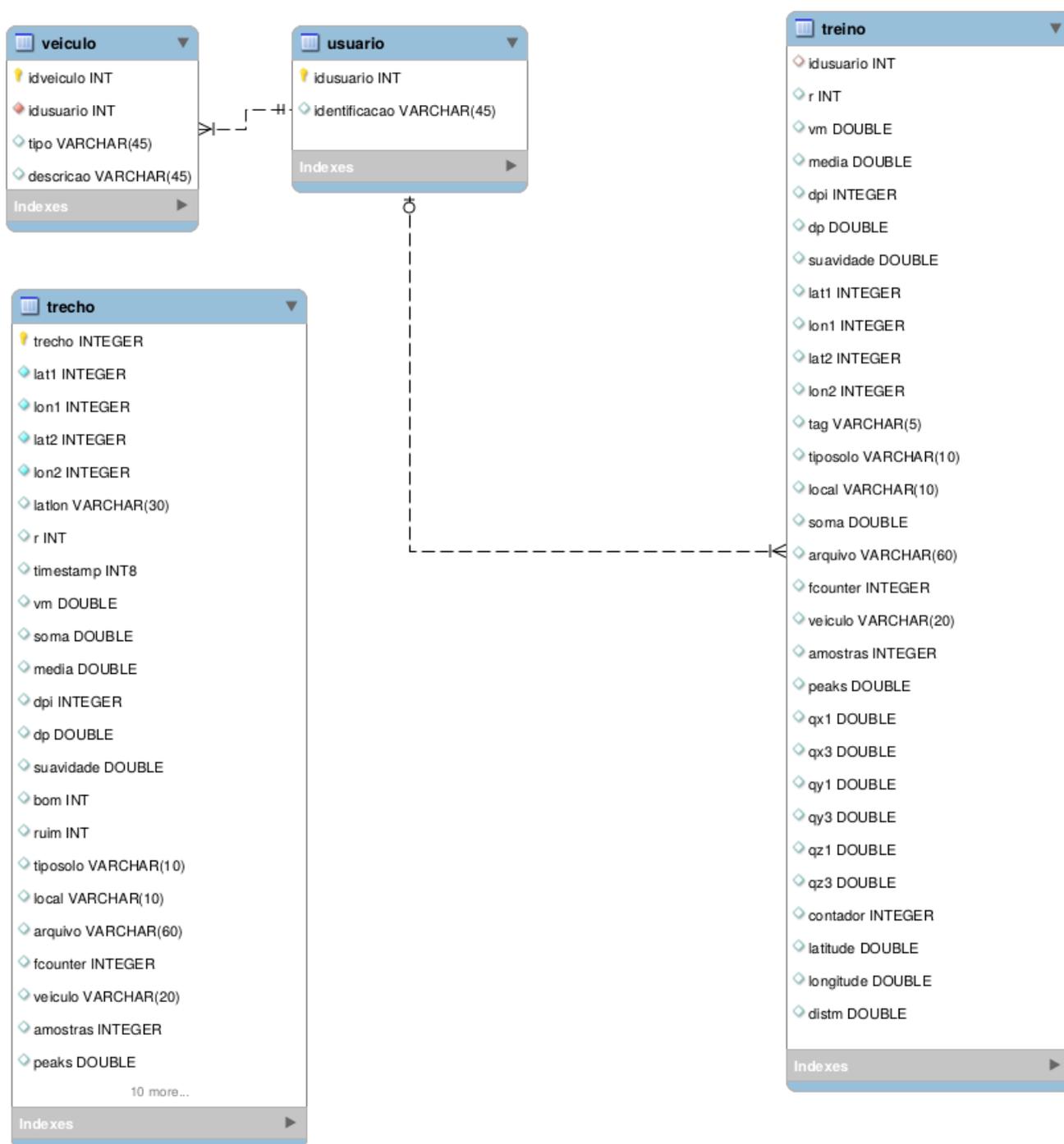
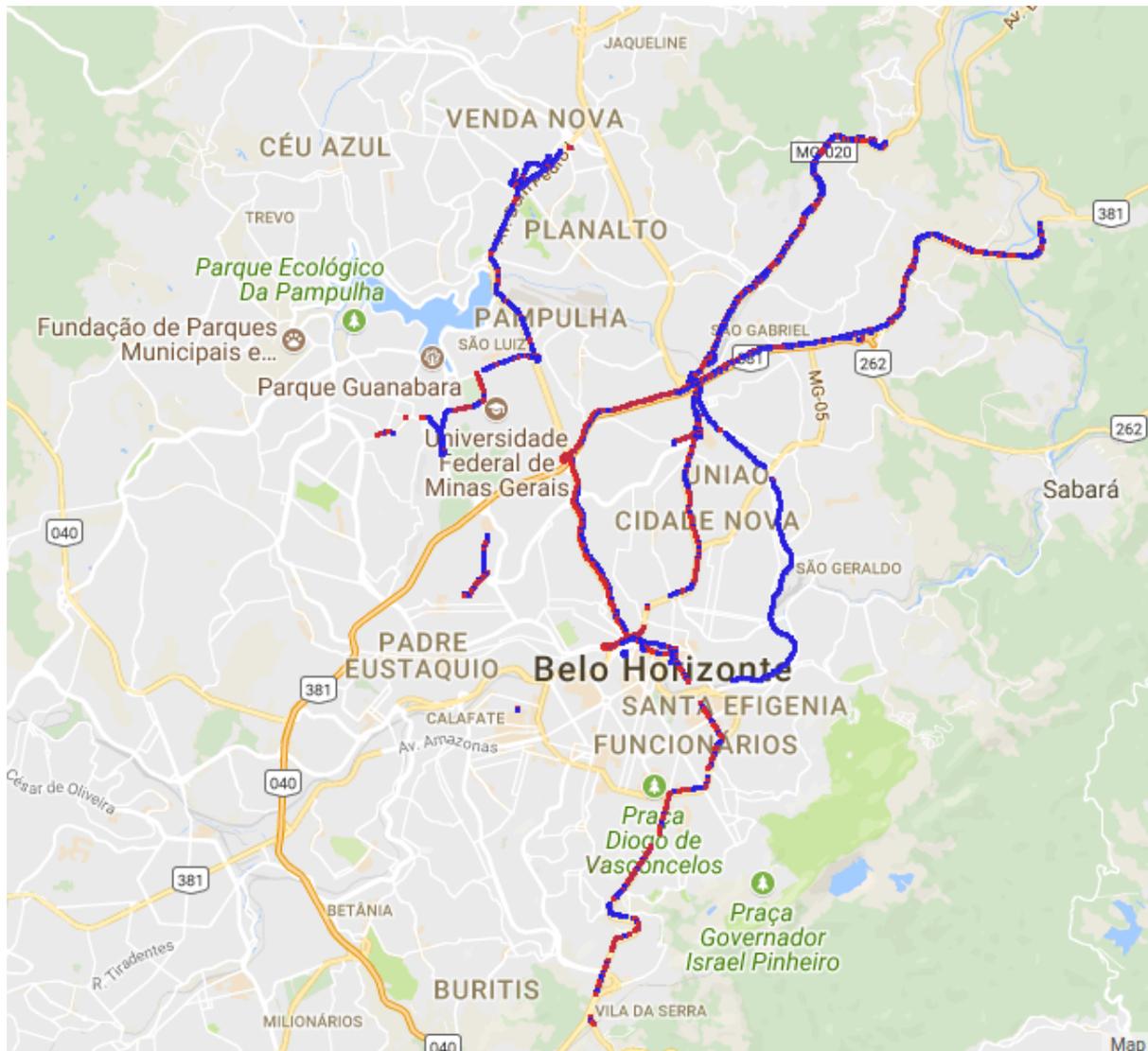
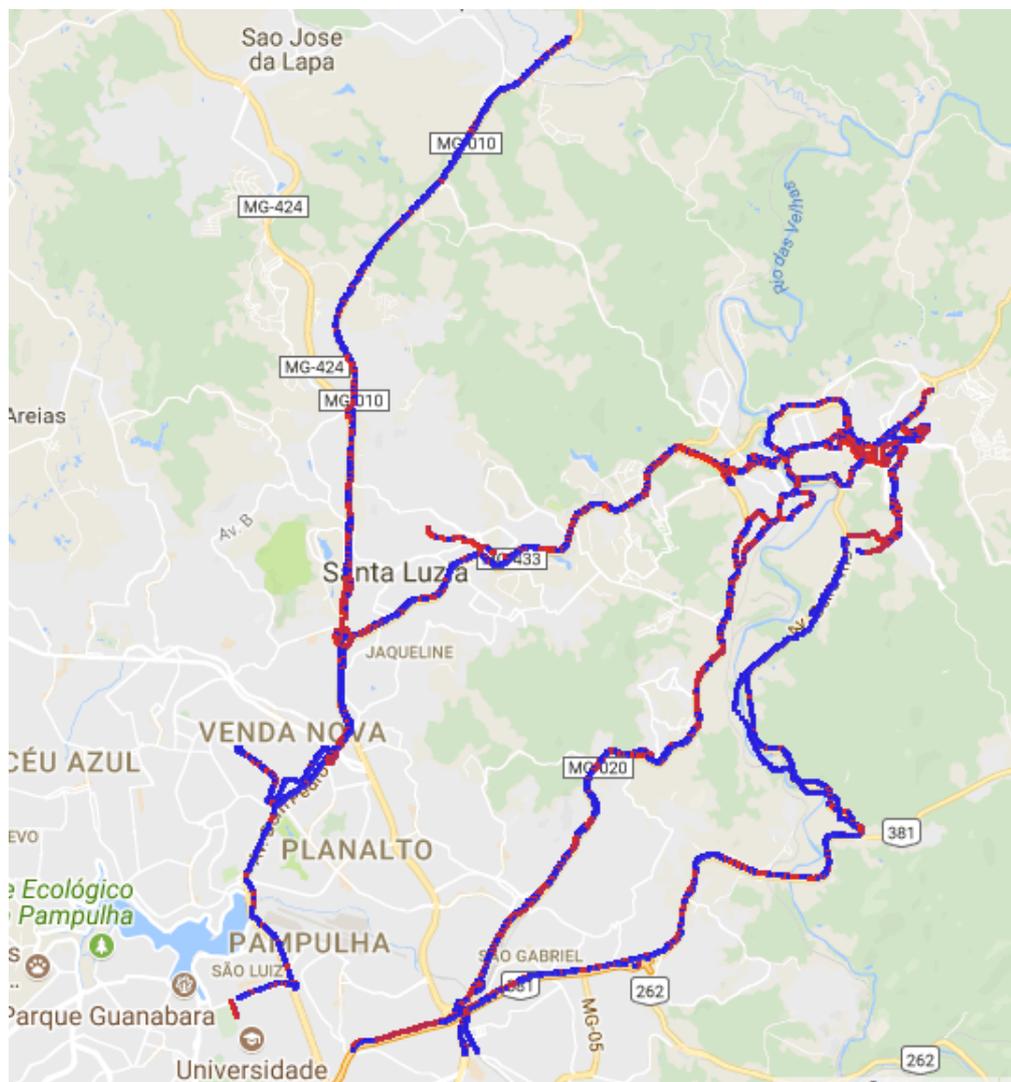


Figura 45 – Mapa de Belo Horizonte



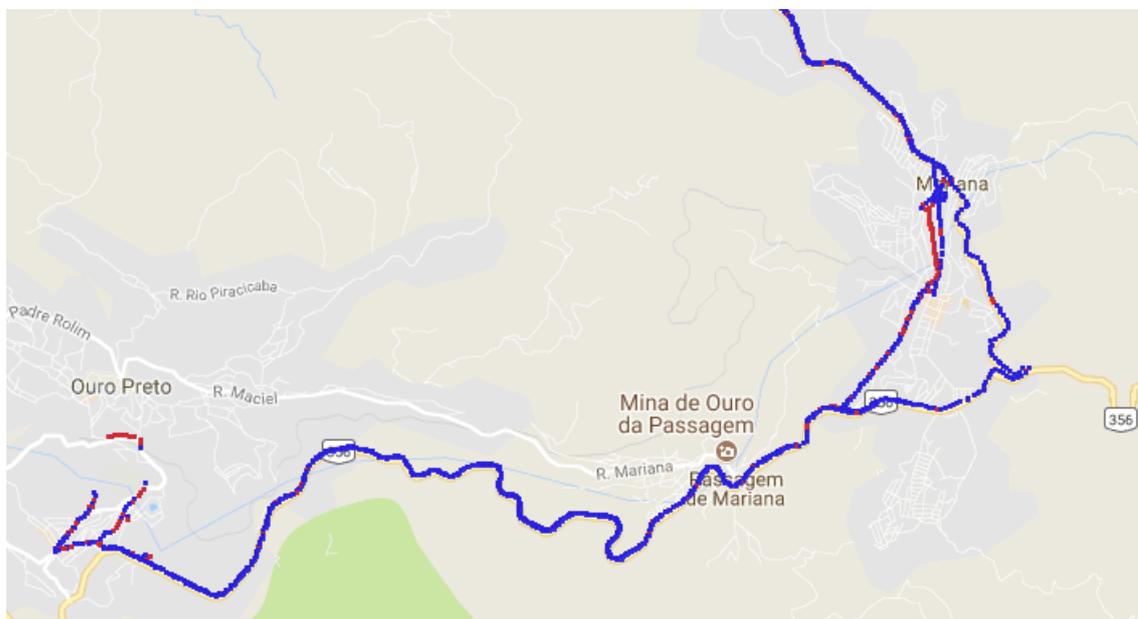
Fonte – Próprio autor

Figura 46 – Mapa de Santa Luzia



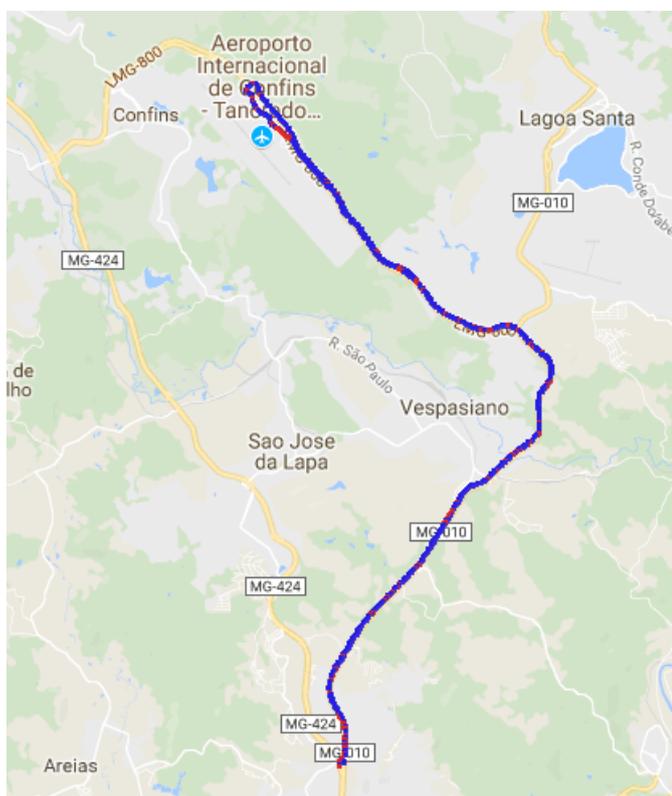
Fonte – Próprio autor

Figura 47 – Mapa de Ouro Preto



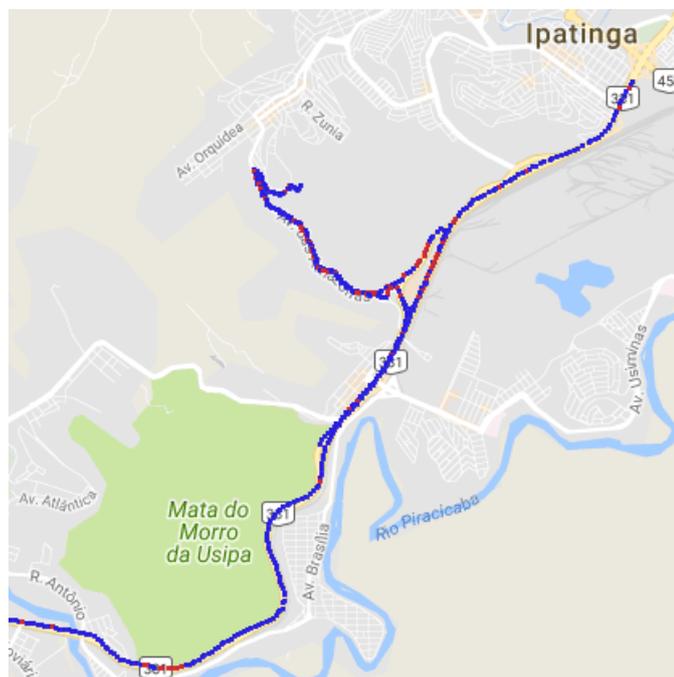
Fonte – Próprio autor

Figura 48 – Mapa de Confins



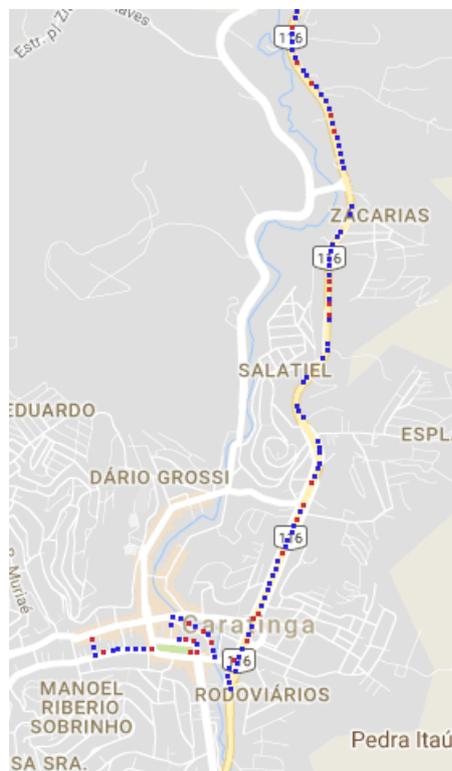
Fonte – Próprio autor

Figura 49 – Mapa de Ipatinga



Fonte – Próprio autor

Figura 50 – Mapa de Caratinga



Fonte – Próprio autor

Anexos

ANEXO A – Outros materiais

Anexos são materiais não elaborados pelo autor, que servem de fundamentação, comprovação e ilustração (Biblioteca JM).

TERMO DE RESPONSABILIDADE

Eu, **Davidson Esteves Nunes** declaro que o texto do trabalho de conclusão de curso intitulado “*Sistema para classificação de malha viária baseado em Smartphones por meio de aprendizado supervisionado*” é de minha inteira responsabilidade e que não há utilização de texto, material fotográfico, código fonte de programa ou qualquer outro material pertencente a terceiros sem as devidas referências ou consentimento dos respectivos autores.

João Monlevade, 08 de fevereiro de 2018



Davidson Esteves Nunes

Declaração

Certifico que o aluno DAVIDSON ESTEVES NUNES, autor do trabalho de conclusão de curso intitulado "Sistema para classificação de malha viária baseado em Smartphones por meio de aprendizado supervisionado", efetuou as correções sugeridas pela banca examinadora e que estou de acordo com a versão final do trabalho.



Talles de Medeiros - Orientador

Ouro Preto, 01 de agosto de 2018