



**INSTITUTO FEDERAL DE EDUCAÇÃO, CIÊNCIA E TECNOLOGIA DO CEARÁ
IFCE CAMPUS ARACATI
COORDENADORIA DE CIÊNCIA DA COMPUTAÇÃO
BACHARELADO EM CIÊNCIA DA COMPUTAÇÃO**

Vinícius Lima Costa

**SiCliSIG - Identificando Similaridades entre Perfis de Ciclistas
Através de Sistemas de Informações Geográficas e Técnicas de
Ciência de Dados**

**ARACATI-CE
2018**

Vinícius Lima Costa

SICLISIG - IDENTIFICANDO SIMILARIDADES ENTRE PERFIS DE CICLISTAS ATRAVÉS DE SISTEMAS DE INFORMAÇÕES GEOGRÁFICAS E TÉCNICAS DE CIÊNCIA DE DADOS

Trabalho de Conclusão de Curso (TCC) apresentado ao curso de Bacharelado em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE - Campus Aracati, como requisito parcial para obtenção do Título de Bacharel em Ciência da Computação.

Orientador: Prof. Dr. Reinaldo Bezerra Braga (orientador)

Co-Orientador: Prof. Dr. Tales Paiva Nogueira

Aracati-CE
2018

Dados Internacionais de Catalogação na Publicação
Instituto Federal do Ceará - IFCE
Sistema de Bibliotecas - SIBI
Ficha catalográfica elaborada pelo SIBI/IFCE, com os dados fornecidos pelo(a) autor(a)

- C837s Costa, Vinícius Lima.
SiCliSIG - Identificando Similaridades entre Perfis de Ciclistas Através de Sistemas de Informações Geográficas e Técnicas de Ciência de Dados / Vinícius Lima Costa. - 2018.
64 f. : il. color.
- Trabalho de Conclusão de Curso (graduação) - Instituto Federal do Ceará, Bacharelado em Ciência da Computação, Campus Aracati, 2018.
Orientação: Prof. Dr. Reinaldo Bezerra Braga.
Coorientação: Prof. Dr. Tales Paiva Nogueira.
1. Cidades Inteligentes. 2. Sistemas de Informações Geográficas. 3. Sistemas de Recomendações. 4. Ciclismo. 5. Internet das Coisas. I. Título.
-

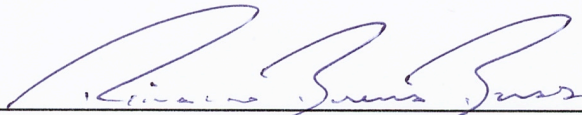
Vinícius Lima Costa

SICLISIG - IDENTIFICANDO SIMILARIDADES ENTRE PERFIS DE CICLISTAS ATRAVÉS DE SISTEMAS DE INFORMAÇÕES GEOGRÁFICAS E TÉCNICAS DE CIÊNCIA DE DADOS

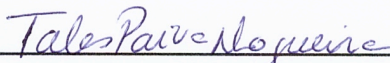
Trabalho de Conclusão de Curso (TCC) apresentado ao curso de Bacharelado em Ciência da Computação do Instituto Federal de Educação, Ciência e Tecnologia do Ceará - IFCE - Campus Aracati, como requisito parcial para obtenção do Título de Bacharel em Ciência da Computação.

Aprovada em 16/10/2018

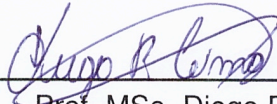
BANCA EXAMINADORA



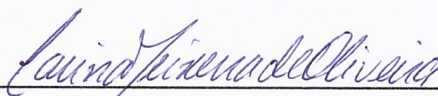
Prof. Dr. Reinaldo Bezerra Braga (Orientador)
Instituto Federal de Educação, Ciência e Tecnologia do Ceará



Prof. Dr. Tales Paiva Nogueira (Co-Orientador)
Universidade Federal do Ceará



Prof. MSc. Diego Rocha Lima
Instituto Federal de Educação, Ciência e Tecnologia do Ceará



Profa. Drá. Carina Teixeira de Oliveira
Instituto Federal de Educação, Ciência e Tecnologia do Ceará

DEDICATÓRIA

Dedico este trabalho primeiramente aos meus pais Ernesto e Lucilene, meu irmão Filipe, minha vó Lúcia e minha namorada Arielma que estiveram sempre ao meu lado me dando forças. Dedico também a todos aqueles que de alguma forma estiveram e estão próximos de mim, fazendo esta vida valer cada vez mais a pena. E por último dedico a todos meus professores, os quais sou muito grato e que contribuíram amplamente no meu aprendizado e formação.

AGRADECIMENTOS

Gostaria de agradecer a Deus, à toda minha família, amigos, namorada e pessoas que ajudaram na realização desse trabalho. Gostaria de agradecer também aos meus orientadores Dr. Reinaldo Bezerra Braga e Dr. Tales Paiva Nogueira que dedicaram seu tempo na orientação desse trabalho, me transmitindo vários conhecimentos. Agradeço também às pessoas que se voluntariaram a utilizar o aplicativo, ajudando nos resultados desse trabalho. Agradeço também aos professores MSc. Diego Rocha Lima e Dra. Carina Teixeira de Oliveira por terem aceitado participar da banca. E finalmente, gostaria de agradecer ao Instituto Federal de Educação, Ciência e Tecnologia do Ceará (IFCE) e a todos os professores dessa instituição pelos ensinamentos e apoio que me foi dado.

Sou imensamente grato pela paciência e incentivo de todos!!!

RESUMO

Com os avanços tecnológicos que vêm ocorrendo, juntamente com a integração de diversos serviços à Internet, ocorreu um crescimento na quantidade de dados gerados, começando a buscar soluções através de dados existentes. Além desse contexto, as pessoas vêm se preocupando mais com a saúde, buscando meios para melhorá-la como a prática de atividades físicas, aumentando a utilização de aplicativos nesse segmento e abrindo a possibilidade de dados gerados nesse eixo que podem ser dos diversos tipos. Neste trabalho é abordado a ideia de cidades inteligentes juntamente com dados do sistema global de posicionamento (GPS), para isso ele é dividido em algumas etapas. Primeiramente o foco está na captura e modelagem dos dados de GPS, buscando obter uma captura otimizada das rotas economizando espaço dos dispositivos. Após a captura e modelagem busca-se dispor os dados de forma que eles fiquem mais acessíveis e possam ser melhores analisados pelo SiCliSIG. Então os dados foram analisados, tendo como objetivo promover o crescimento do número de ciclistas, melhorar a segurança deles e incitar o pedal em grupo. Para tanto foram traçados os perfis dos ciclistas de acordo com as rotas que eles percorreram, com o objetivo de realizar a recomendação entre eles. Outra abordagem foi aplicada ao planejamento urbano, no qual foi utilizado o conceito de mapas de calor a fim de detectar as rotas mais percorridas, auxiliando na tomada de decisões das autoridades competentes a respeito de obras favoráveis ao ciclismo, como ciclovias e sistemas de compartilhamento de bicicleta. Unindo essas abordagens propostas, pode-se atingir os ideais de incitar o crescimento do número de ciclistas, melhorar a segurança deles e promover o pedal em grupo.

Palavras-chaves: Cidades Inteligentes, Sistemas de Informações Geográficas, Sistemas de Recomendações, Ciclismo, Sistema de Posicionamento Global, Dados Geográficos, Internet das Coisas

ABSTRACT

With the technological advances that have been occurring, along with the integration of several Internet services, there has been a growth in the amount of data generated, beginning to seek solutions through existing data. In addition to this context, people are becoming more concerned about health, seeking ways to improve it, such as practicing physical activities, increasing the use of applications in this segment and opening the possibility of data generated in this axis that can be of various types. In this work the idea of smart cities together with data of the global positioning system (GPS) is approached, for this it is divided in some steps. Firstly, the focus is on capturing and modeling GPS data, seeking to optimize route capture by saving space on the devices. After the capture and modeling, the data will be arranged so that they are more accessible and can be better analyzed by SiCliSIG. Then the data were analyzed, aiming to promote the growth of the number of cyclists, to improve the safety of them and to incite the pedal in group. In order to do so, the profiles of the cyclists were drawn according to the routes that they traveled, with the purpose of making the recommendation among them. Another approach was applied to urban planning, in which the concept of heat maps was used to detect the most traveled routes, assisting in the decision-making of the competent authorities regarding works favorable to cycling, such as bike paths and bicycle-sharing systems. Combining these proposed approaches, one can achieve the ideals of encouraging the growth of the number of cyclists, improving their safety and promoting the cycling in groups.

Keywords: Smart Cities, Geographic Information Systems, Recommendations Systems, Cycling, Global Positioning System, Geographic data, Internet of Things.

LISTA DE ILUSTRAÇÕES

Figura 1 – Gráfico sobre a porcentagem da população americana que pratica atividades físicas regularmente	15
Figura 2 – Gráfico sobre o mercado esportivo global, mostrando total entre os anos de 2005 a 2017 (em bilhões de dólares)	16
Figura 3 – Arquitetura proposta por Tyagi e Kumar (2017)	24
Figura 4 – Estrutura de um SIG proposta por Camara e Ortiz (1998)	27
Figura 5 – Triangulação realizada pelos satélites	29
Figura 6 – Funcionamento do algoritmo <i>K-means</i>	35
Figura 7 – Modelo em camadas para análise de dados	40
Figura 8 – Arquivo gerado pelo aplicativo móvel com os dados coletados	43
Figura 9 – Estrutura do Banco de dados responsável pelo armazenamento das informações	45
Figura 10 – Tela principal do aplicativo para coletar dados.	49
Figura 11 – Mapa com todas as rotas coletadas	54
Figura 12 – Mapa de calor com as alterações feitas nos parâmetros e baseado em todas as rotas percorridas pelos usuários	55
Figura 13 – Mapa de calor padrão, sem as alterações feitas nos parâmetros, e baseado em todas as rotas percorridas pelos usuários	56
Figura 14 – Mapas com todas as rotas percorridas pelo usuário 1 e pelo usuário 2	57

LISTA DE TABELAS

Tabela 1 – Tabela da análise temporal	57
-------------------------------------------------	----

LISTA DE ABREVIATURAS E SIGLAS

IFCE	Instituto Federal de Educação, Ciência e Tecnologia do Ceará
TCC	Trabalho de Conclusão de Curso
API	<i>Application Programming Interface</i>
XML	<i>Extensible Markup Language</i>
IoT	<i>Internet of Things</i>
GPS	<i>Global Positioning System</i>
SQL	<i>Structured Query Language</i>
MTB	<i>Mountain Bike</i>
RFID	<i>Radio-Frequency Identification</i>
P&G	<i>Procter & Gamble</i>
LoRa	<i>Long Range</i>
LPWAN	<i>Low Power Wide Area Network</i>
SIG	Sistema de Informações Geográficas
AVC	Acidente Vascular Cerebral
DoD	<i>Department of Defence</i>
EUA	Estados Unidos da América
NAVSTAR	<i>Navigation Satellite with Time and Ranging</i>
SA	<i>Selective Availability</i>
FCC	<i>Federal Communications Commission</i>
GNSS	<i>Global Navigation Satellite Systems</i>
EV	<i>Electric Vehicle</i>
HV	<i>Hybrid Vehicles</i>
TIC	Tecnologia de Informação e Comunicação
RS	<i>Recommender System</i>

CF *Collaborative Filtering*

k-NN *k-Nearest Neighbors*

SUMÁRIO

1	INTRODUÇÃO	14
1.1	Objetivos	18
1.2	Organização do Trabalho	19
2	FUNDAMENTAÇÃO TEÓRICA	20
2.1	Cidades Inteligentes	20
2.1.1	Internet das Coisas	22
2.2	Sistemas de Informações Geográficas	25
2.2.1	Sistema de Posicionamento Global	28
2.3	Sistemas de Recomendação	30
2.3.1	<i>K-means</i>	33
3	TRABALHOS RELACIONADOS	36
4	Proposta	40
4.1	Camada de Captura de Dados	40
4.1.1	Sensores	41
4.2	Camada de Interoperabilidade	41
4.2.1	Modelo de Dados	42
4.2.2	Extração e Transformação dos Dados	43
4.3	Camada de Armazenamento	45
4.4	Camada de Análise dos Dados	46
4.5	Camada de Aplicação	48
4.5.1	Cidades Inteligentes	50
4.5.1.1	Planejamento Urbano	51
4.5.1.2	Recomendação de Ciclistas	51
5	RESULTADOS	53
5.1	Planejamento Urbano	53
5.2	Recomendação de Ciclistas	56
6	CONCLUSÃO	58
	REFERÊNCIAS	59

1 INTRODUÇÃO

É perceptível o aumento na quantidade de dados gerados na Internet devido à ampla adoção de redes sociais e aplicativos. Nesse contexto, foi detectada uma oportunidade de lidar com tais dados e gerar informações para serem usadas no cotidiano das pessoas por meio da utilização da ciência de dados (*data science*).

A ciência de dados é um ramo da computação que tem como principal objetivo extrair e interpretar os dados de forma eficaz e apresentá-los em uma linguagem simples e não técnica para os usuários finais (DATASCIENCEACADEMY, 2018). Esse campo diz respeito às técnicas para extrair conhecimento de dados diversos, com um foco particular em dados “grandes” exibindo atributos “V”, como volume, velocidade, variedade, valor e veracidade (MANETH; POULOVASSILIS, 2017). Esse ramo vem se popularizando e engloba algumas outras áreas para o seu uso, como estatística, mineração de dados, aprendizado de máquina, entre outras.

A ciência de dados é geralmente suscetível à integração com outros campos, procurando potencializar as informações dessas áreas integradas. Uma das informações que podem ser exploradas pela ciência de dados é a localização fornecida pelo Sistema de Posicionamento Global (*Global Positioning System* - GPS), podendo obter informações diferentes de acordo com o local onde os dados foram gerados, conseguindo em alguns casos categorizar melhor os dados.

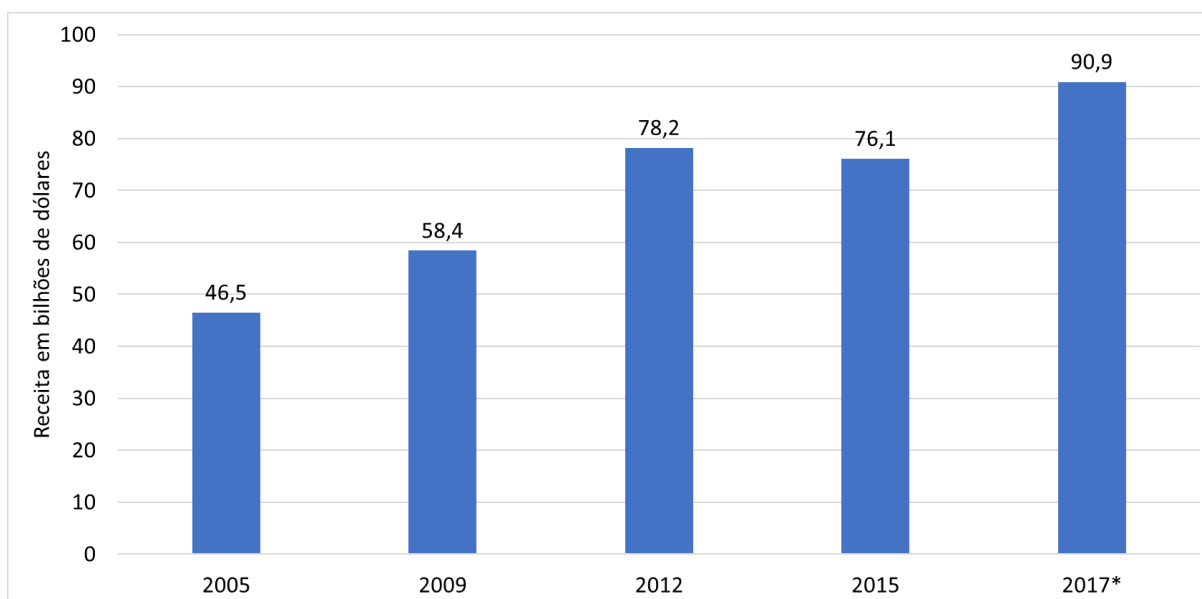
De acordo com o livro *Engineering Satellite-Based Navigation and Timing* (BETZ, 2016), o programa de GPS foi estabelecido no início dos anos 70 pelos Estados Unidos, baseado em trabalhos anteriores em navegação de rádio. Sendo que o primeiro satélite GPS operacional foi lançado em 1978. Já em 1983, após um crescimento contínuo e maturação do sistema, os Estados Unidos anunciaram formalmente o fornecimento de GPS para uso civil, seguido de um anúncio em 1991 de que o GPS civil seria fornecido gratuitamente à comunidade internacional. O tamanho da constelação então especificado para tal serviço é de 24 satélites operacionais, sendo que tal número foi alcançado em 1994.

O sistema de posicionamento global é incorporado em alguns dispositivos móveis como é o caso dos *smartphones*, conseguindo auxiliar em várias atividades como: localização automotiva, uso em redes sociais, práticas esportivas, etc. Esta última atividade vem se popularizando, pois uma parcela maior de pessoas está se preocupando mais com a saúde e bem-estar, surgindo a oportunidade para aplicativos que estejam relacionados a esses temas.

O crescimento da prática esportiva é ilustrado na Figura 1 e Figura 2. A Fi-

gura 1 é um gráfico extraído do site *Bureau of Labor Statistics* (WOODS, 2017), que mostra a porcentagem de americanos praticante de atividade física variando de 2003 a 2015 e a Figura 2 é um gráfico extraído do site *Statista* (KEARNEY, 2018), que mostra a receita do mercado esportivo global nos anos de 2005, 2009, 2012, 2013 e estima quanto foi a receita em 2017.

Figura 1 – Gráfico sobre a porcentagem da população americana que pratica atividades físicas regularmente



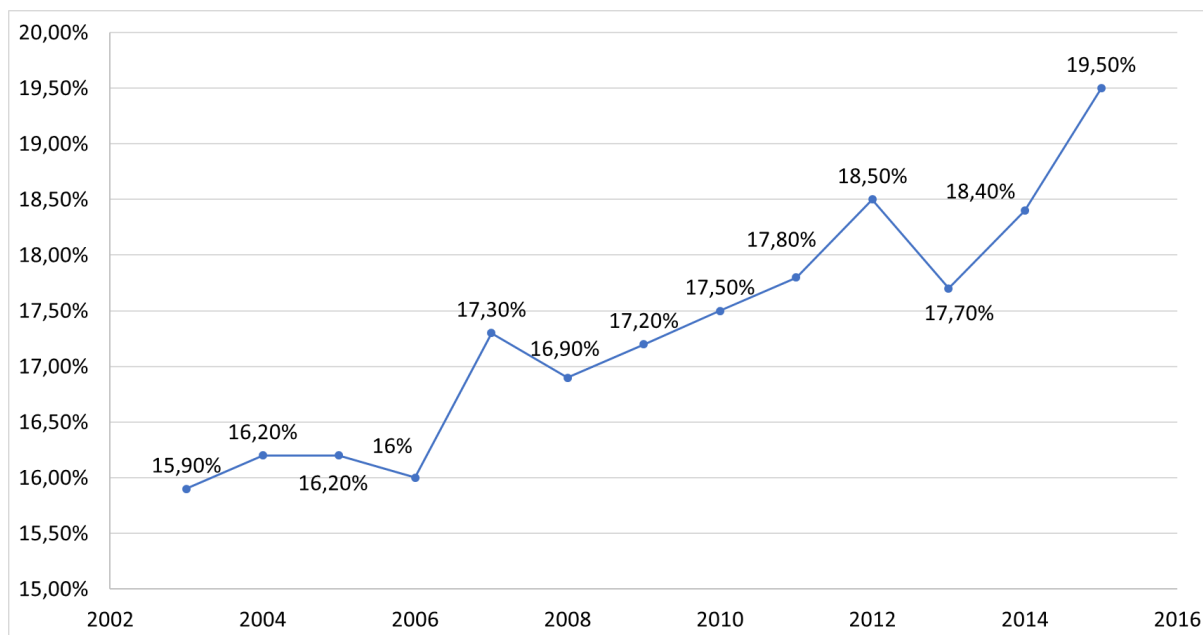
Fonte: Adaptado de Woods (2017)

No gráfico apresentado na Figura 1, é exposto o crescimento da quantidade de esportistas americanos: em 2003 a parcela de população americana que praticava esportes era de 13% e em 2015 esse número subiu para quase 20%, mostrando que houve um aumento na quantidade de esportistas americanos. Enquanto que no gráfico mostrado na Figura 2 é expresso o crescimento global do mercado esportivo. Em 2005, esse mercado teve uma receita de 46,5 bilhões de dólares e em 2017 foi estimada uma receita de 90,9 bilhões de dólares, evidenciando que o esporte também cresceu no mundo, não só nos Estados Unidos.

O crescimento da prática esportiva atinge também o Brasil, como mostra o jornal *DiarioDoNordeste* (2017). Nesse site, são mostrados dados divulgados pelo Ministério da Saúde, no qual, entre outros pontos, é falado que a prática de atividade física no tempo livre cresceu, passando de 30,3% em 2009 para 37,6% em 2016. Inserido nesse contexto das atividades físicas, o ciclismo vem se popularizando.

A bicicleta sempre foi bastante utilizada com o propósito de mobilidade e locomoção, porém ultimamente vem sendo agregada uma nova aplicação para ela, que é o seu uso como atividade física. Essa nova forma de utilização da bicicleta vem

Figura 2 – Gráfico sobre o mercado esportivo global, mostrando total entre os anos de 2005 a 2017 (em bilhões de dólares)



Fonte: Adaptado de [Kearney \(2018\)](#)

se popularizando, sendo esse crescimento notório no estado do Ceará, onde muitos ciclistas se tornaram amadores e competidores de eventos locais. Isso é evidente quando se observa o constante aumento de inscritos em eventos de competição, que vão desde níveis iniciais até níveis profissionais.

Em uma entrevista realizada na Bike Sport¹, o proprietário destacou as mudanças ocorridas no município de Aracati, quando o assunto é ciclismo. É notável o aumento de inscritos no Aracati Adventure², competição de *Mountain Bike* (MTB) realizada em Aracati-CE, cuja primeira edição ocorreu em 2011, que na época era chamada de MTB Racing. Na primeira edição, o evento contou com a presença de 50 ciclistas, enquanto que na sua última edição, que ocorreu em 2017, contou com 600 ciclistas participantes. Esse crescimento no número de ciclistas esportistas não se deu apenas pelo incentivo à prática de competições, com um número crescente de competições e campeonatos na região, mas também ao surgimento de mercado especializado no assunto. Ou seja, o surgimento e o crescimento de lojas especializadas (*bike shop*) também são responsáveis pela proliferação do ciclismo como uma forma de prática esportiva, disponibilizando informações a respeito dessa atividade. Isso está associado à facilidade no fornecimento de materiais próprios para a prática esportiva, fazendo com que esse esporte fique mais próximo das pessoas.

Em paralelo às práticas esportivas, o uso da bicicleta também tem crescido

¹ <https://www.facebook.com/BikeSportAracati/>

² <https://www.facebook.com/Aracati-Adventure-557297660995229/>

para fins de locomoção, substituindo o uso de outros meios de transportes, como motos ou carros. Esse crescimento foi ilustrado pelo jornal [Estadão \(2014\)](#), que reportou um aumento de 50% na quantidade de pessoas que usam bicicletas todos os dias como meio de transporte em São Paulo. Esse reuso da bicicleta como meio de transporte se deu por causa de alguns pontos como: fugir do trânsito, economizar dinheiro, melhorar a saúde, etc. Além desses pontos, pesa também o incentivo de algumas cidades que constroem pontos de bicicletas compartilhadas e ciclovias, tornando ainda mais agradável e barato o uso da bicicleta para locomoção. Segundo a revista [Exame \(2016\)](#), algumas das cidades brasileiras que estão se adaptando às bicicletas são: Fortaleza-CE, São Paulo-SP, Rio de Janeiro-RJ, Brasília-DF, Florianópolis-SC, Recife-PE, Belo Horizonte-MG, Curitiba-PR, Aracaju-SE e Manaus-AM.

O uso da bicicleta como meio de transporte, substituindo o carro ou moto, acarreta uma redução da quantidade de veículos automotivos que circulam nos grandes centros urbanos. Dessa forma, o uso da *bike* para locomoção agrega mais algumas vantagens. Inicialmente essa substituição do meio de transporte utilizado acarreta em uma redução nas emissões de gases poluentes na atmosfera ([BIKERADAR, 2018](#)), que segundo a revista [Veja \(2018\)](#), a poluição mata mais de 50 mil pessoas ao ano somente no Brasil. Além disso, ao utilizar a bicicleta para a locomoção, seu usuário tende a fugir do estresse que o trânsito pode proporcionar e também devido a essa fuga do trânsito, o ciclista está propício a levar menos tempo no seu trajeto já que as bicicletas não enfrentam os congestionamentos que os veículos automotivos enfrentam, especialmente os carros.

Esses novos entusiastas conseguem usufruir de algumas vantagens com o uso da *bike*. Uma delas é que seus adeptos se tornarão mais saudáveis, pois a atividade física regular previne doenças cardíacas, Acidentes Vasculares Cerebrais (AVCs), hipertensão, ajuda a prevenir e a controlar o diabetes, aumenta a resistência aeróbica, reduz a obesidade, ativa a musculatura de todo o corpo, diminui a ocorrência de doenças crônicas, faz bem para a saúde do idoso e aumenta o tempo de vida ([VADEBIKE, 2016](#)). Outra vantagem está relacionada à melhoria da coordenação motora, pois andar de bicicleta envolve todo o corpo, portanto, braços, pernas e mãos, além do mais, favorece o reflexo reduzindo o tempo de reação em alguma adversidade ([MUNDOBOAFORMA, 2015](#)). Finalmente, andar de *bike* acarreta também melhorias físicas, ocorrendo aumento de massa muscular, queima de calorias e melhoria da capacidade respiratória ([VADEBIKE, 2016](#)).

Porém, esse crescimento da utilização de bicicletas como prática esportiva e meio de transporte, despertou o interesse não somente de pessoas que almejam atividades físicas, mas também de pessoas mal-intencionadas com o intuito de roubar bicicletas e se aproveitar do crescente mercado de bicicleta para vender *bikes* rouba-

das. De acordo com a revista [EpocaNegocios \(2017\)](#), só em São Paulo houve 1351 furtos e roubos de bicicletas entre janeiro e setembro de 2016, logo, foram roubadas em média 46 bicicletas por dia.

Uma das formas de prevenir os roubos de bicicletas é pedalar em grupo. Além disso, um grupo de pedal garante maior visibilidade para os ciclistas no trânsito, reduzindo assim o risco de acidentes com outros meios de transportes, garantindo uma maior segurança para os ciclistas. Porém, pedalar em grupos pode ser benéfico também para servir de estímulo para os ciclistas, pois pedalar em grupo e ter uma interação com outras pessoas pode ser mais prazeroso do que pedalar sozinho.

Portanto, o incentivo ao pedal (andar de bicicleta) em grupo pode abrir portas para novos ciclistas, pois com o pedal em conjunto existe uma tendência de aumento na divulgação desse meio de transporte que pode ser utilizado como esporte. Essa divulgação pode ser através de conversas com colegas no cotidiano ou até mesmo por postagens em redes sociais providas pelos próprios ciclistas, podendo despertar o interesse de pessoas a se tornarem adeptos à bicicleta.

No contexto da realização da prática do ciclismo em grupos, existem algumas maneiras que podem incitar essa prática do ciclismo em grupos, uma delas é que pode se utilizar de ferramentas computacionais que identifiquem os perfis dos ciclistas, de maneira que possa sugerir ciclistas compatíveis. Desse modo, é benéfico ter formas de utilizar os sensores já existentes nos *smartphones*, a fim de identificar ciclistas com perfis semelhantes com o intuito de contribuir com o pedal em grupos, aumentando a segurança dos ciclistas e incitando a prática do ciclismo. Enquanto que no conceito de planejamento urbano, pode-se utilizar também de ferramentas computacionais, a fim de facilitar na detecção de locais mais frequentados por ciclistas, locais esses que podem vir a ser alvo de construções em prol do ciclismo como ciclovias.

1.1 Objetivos

Com o propósito de motivar o pedal em grupo e incitar o crescimento do ciclismo, o objetivo geral desse trabalho é propor uma solução que consiga, a partir de dados coletados pelo percurso de um ciclista traçar perfis, sugerir possíveis companheiros de pedal e auxiliar no planejamento urbano.

Os objetivos específicos desse trabalho são:

- Coletar dados através de sensores usando uma aplicação móvel;
- Interoperabilidade entre *smartphone* e o servidor;
- Armazenar dados em um servidor;

- Propor um algoritmo de análise de perfis para identificar similaridades entre ciclistas;
- Utilização das informações para auxiliar no planejamento urbano;
- Recomendação de ciclistas com perfis compatíveis.

1.2 Organização do Trabalho

O restante desse trabalho está organizado conforme exposto a seguir. O Capítulo 2 apresenta a fundamentação teórica, que consiste em um detalhamento de conceitos importantes que serão explorados no decorrer do trabalho. No Capítulo 3 são mostrados os trabalhos relacionados. O Capítulo 4 detalha toda a proposta desse trabalho, expondo um modelo em camadas na qual o trabalho se insere e a implementação do mesmo. Já no Capítulo 5 são mostrados os resultados obtidos. Por fim, no Capítulo 6 são mostradas as conclusões e alguns encaminhamentos para trabalhos futuros.

2 FUNDAMENTAÇÃO TEÓRICA

As Tecnologias de Informação e Comunicação (TICs) são bastante utilizadas com o intuito de auxiliar em atividades humanas a fim de otimizar a rotina da população. Com o decorrer dos anos, as TICs foram se desenvolvendo e passaram a se tornar indispensáveis em algumas atividades. Alguns dessas formas de tecnologia serão explanadas nesse capítulo, como as cidades inteligentes, que serão mostradas na Seção 2.1, Internet das Coisas (Seção 2.1.1), Sistemas de Informações Geográficas (Seção 2.2) e Sistema de Posicionamento Global (Seção 2.2.1).

2.1 Cidades Inteligentes

O conceito de cidades inteligentes (*Smart Cities*) se originou na busca por uma definição ideal de cidade com avanços tecnológicos. Segundo Mardacany (2014), esse conceito foi concebido em projetos pós Segunda Guerra Mundial, que abrangiam toda a ideia de cidade industrializada como “uma fábrica para trabalhar” com a construção de “máquinas para viver”, com princípios como futuro, inteligência, digital, avançado, verde, habitável e assim por diante.

Na virada do século, a noção de cidade ideal e tecnológica foi conceituada como fenômeno técnico, social e urbano. Com isso, os atuais modelos conceituais de uma cidade inteligente são dominados pelo modelo de negócios apoiado por instituições financeiras que buscam retorno sobre o investimento em meio ambiente construído (MARDACANY, 2014). Desta forma, as cidades inteligentes surgiram como uma maneira para melhorar uma cidade normal, a fim de que ela se torne uma cidade ideal ou próxima disso, por meio das tecnologias.

De acordo com Mohanty, Choppali e Kougianos (2016), as cidades inteligentes possuem dois conceitos mais aceitos. Uma das definições é que a *smart city* busca conectar a infraestrutura física, a infraestrutura de tecnologia da informação, a infraestrutura social e a infraestrutura de negócios para alavancar a inteligência coletiva da cidade. A outra definição formal e abrangente é que uma cidade inteligente e sustentável é uma cidade inovadora, que usa tecnologias de informação e comunicação e outros meios para melhorar a qualidade de vida da população. Essa qualidade de vida relaciona-se à eficiência das operações e serviços urbanos e competitividade, que atende às necessidades das gerações presentes e futuras em relação aos aspectos econômicos, sociais e ambientais.

Dessa forma, é importante utilizar sensores para captar informações relevan-

tes para serem agregados, tornando uma cidade inteligente de fato, como foi exposto por Afaneh e Shahrour (2017). Nesse trabalho é exposto que o conceito de Cidades Inteligentes visa usar a tecnologia digital, bem como a inovação social, para o monitoramento e gestão de infraestruturas e serviços urbanos. Também é retratado que o monitoramento inclui sensores inteligentes conectados por meio de redes de comunicação com e sem fio, que permitem uma supervisão e controle em tempo real das infraestruturas urbanas. Dessa forma, eles dizem que as leituras de sensores e os dados dos usuários são armazenados em grandes conjuntos de dados, juntamente com informações sobre o recurso de infraestrutura e outros dados úteis, como informações sobre o clima, tráfego, perfis e consumos dos usuários.

Unindo os conceitos expostos pelos dois autores, pode-se concluir que a cidade inteligente consta da utilização do sensoriamento disponível para uma melhor relação das pessoas com a cidade, conseguindo melhorar as necessidades da população.

As cidades inteligentes podem se aprimorar de várias formas, se utilizando dos sensores disponíveis nos diversos tipos de dispositivos existentes. De acordo com Malik e Shah (2017), as cidades inteligentes podem ser utilizadas em diferentes áreas de atuação. Algumas dessas áreas são:

- **Energia Inteligente:** Com a utilização de sensores, é possível verificar as fontes de energia utilizadas nas cidades, analisando a eficiência energética, bem como a poluição gerada. Neste contexto, algumas propostas apontam para a identificação das melhores fontes de energia para cada tipo de região (país, estado, cidade).
- **Transporte Inteligente:** Pode-se utilizar diferentes fontes de dados e sensores acoplados em transportes urbanos, a fim de conseguir um sistema de tráfego controlado, melhorando a situação da mobilidade urbana, conseguindo implantar soluções como estacionamento inteligente, padrões de mobilidade, etc.
- **Saúde Inteligente:** Ao implantar sensores e utilizá-los através de diferentes aplicativos, os pacientes podem se tornar mais ativos e conscientes sobre sua saúde. Um desses serviços pode ser um sistema de resposta imediata a chamadas de emergências.
- **Casas Inteligentes:** Equipada com sensores e outros dispositivos inteligentes, garantindo o uso eficiente de serviços, resultando em redução do consumo de energia nas moradias.
- **Iluminação Inteligente:** Pode ser implantado um sistema de iluminação baseado em sensores que detecta a presença de um objeto, ligando-se ou desligando-

se, podendo também ajustar sua intensidade de acordo com a necessidade. Dessa forma, é possível proporcionar uma iluminação otimizada.

- **Industria Inteligente:** Utilizando a tecnologia da informação e comunicação é possível se alcançar um aumento de produtividade nas indústrias.
- **Cidadãos Inteligentes:** Se os cidadãos não se engajarem no conceito das cidades inteligente, ela se torna inútil, pois não haverá quem utilize seus serviços ou forneça informações. Além disso, esses serviços podem ser utilizados inclusive para engajamento social.
- **Governança Inteligente:** Tendo em vista os problemas da cidade, os planos e políticas para a cidade são projetados e implementados a partir das informações coletadas pelos sensores nesse contexto de cidades inteligentes.

Como mostrado, as cidades inteligentes possuem diversas áreas de atuação, conseguindo aprimorá-las utilizando-se de sensores para tal. Porém, esse uso de sensoriamento não se restringe a cidades inteligentes, mas também podem ser utilizados em outros conceitos que é o caso da Internet das Coisas.

2.1.1 Internet das Coisas

A Internet das Coisas é um conceito que descreve um futuro onde os objetos físicos do dia-a-dia serão conectados à Internet e serão capazes de identificar-se para outros dispositivos. IoT é uma rede de dispositivos que se comunicam usando conectividade IP sem interferência humana. O ecossistema da Internet das Coisas consiste em objetos inteligentes, dispositivos inteligentes, smartphones e *tablets* etc. Ele pode utilizar de identificação por radiofrequência, resposta rápida, códigos, sensor ou tecnologia sem fio para permitir intercomunicação entre dispositivos (SINGH; SINGH, 2015).

O termo IoT surgiu ainda no século passado. Segundo Mancini (2017), em setembro de 1999, Kevin Ashton, cofundador e diretor executivo do Auto-ID Center, proferiu uma palestra para a Procter & Gamble (P&G), e apresentou uma nova ideia do sistema de identificação por rádio frequência (*Radio-Frequency Identification* - RFID) para a rastreabilidade do produto na cadeia de suprimentos. Para chamar a atenção dos executivos, ele colocou no título da apresentação a expressão *Internet of Things*.

Para o criador do termo IoT, K.Ashton (2009), os computadores e, portanto, a Internet, são quase totalmente dependentes dos seres humanos para obter informações. Porém as pessoas têm tempo, atenção e precisão limitados, o que significa que eles não são muito bons em capturar dados sobre as coisas no mundo real. Dessa

forma o ideal seria que os computadores possam coletar as informações através de seus próprios meios, conseguindo criar um mundo mais inteligente.

Portanto o IoT se utiliza de sensores, coletando suas informações, as transportando por meio de conectividade IP, seguindo um rumo um pouco parecido com o de cidades inteligente, sendo que esses dois conceitos podem ser integrados, como foi explorado por [Dlodlo, Gcaba e Smith \(2016\)](#).

Atualmente a Internet das Coisas já é explorada economicamente por algumas empresas, as principais são LoRa Alliance¹ e Sigfox². Cada empresa possui uma forma distinta de lidar com o IoT, tanto em proposta de trabalho como em arquitetura.

A LoRa Alliance é uma aliança de tecnologia que conta com mais de 500 membros estando em operação desde o final de março de 2015. O principal objetivo da Aliança LoRa é padronizar a rede de área ampla de baixa potência (*Low Power Wide Area Network* - LPWAN) e, por meio da padronização, permitir implantações de grande volume de IoT em larga escala ([LoRa Alliance, 2018a](#)).

Nessa padronização da LPWAN para a implementação de IoT, a LoRa Alliance criou um protocolo próprio derivado da LPWAN, que foi o LoRaWAN. De acordo com o site oficial da [LoRa Alliance \(2018b\)](#), o LoRaWAN é um protocolo de rede projetado para conectar sem fio 'coisas' operadas por bateria à Internet em redes regionais, nacionais ou globais e tem como alvo os principais requisitos da Internet das Coisas.

Além do protocolo LoRaWAN, a LoRa Alliance criou também um hardware para realizar a comunicação entre os dispositivos, que foi o longo alcance (Long Range - LoRa). LoRa é uma tecnologia de rádio frequência que permite comunicação a longas distâncias com consumo mínimo de energia ([BRAGA, 2018](#)). Com isso a LoRa Alliance possui um protocolo próprio e aberto para implementação da internet das coisas e um hardware para a troca de informações entre os dispositivos.

A Sigfox é uma empresa francesa fundada em 2009 com foco na construção de redes sem fio para conectar objetos de baixo consumo de energia, como dispositivos de IoT ([FAGUNDES, 2018](#)). De acordo com o site oficial da [Sigfox \(2018\)](#), a tecnologia Sigfox oferece uma solução de comunicação baseada em software onde toda a complexidade da rede e da computação é gerenciada em nuvem e não nos dispositivos. Isso permite que a Sigfox melhore constantemente seus recursos de rede e os disponibilize através de atualizações de software.

Além disso, a empresa [Sigfox \(2018\)](#) oferece também outros serviços como o Sigfox Atlas, que é um serviço de localização de baixo custo e baixo consumo de energia para todos os dispositivos equipados com o módulo Sigfox mais simples, sem

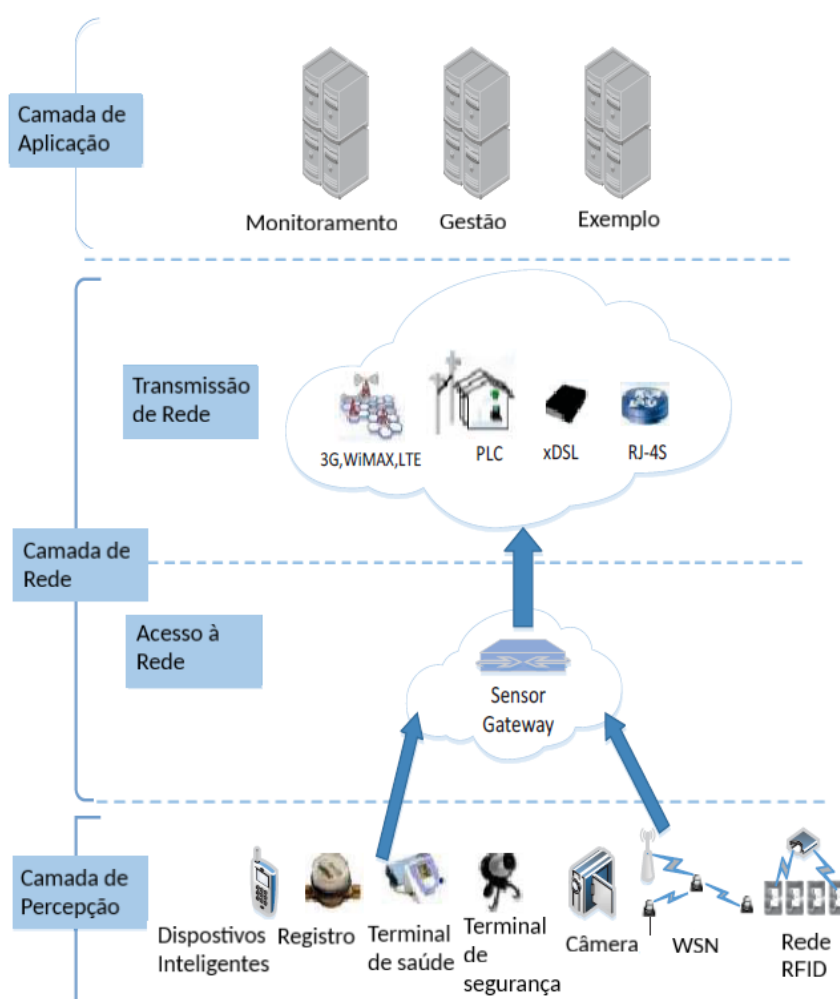
¹ <https://lora-alliance.org/>

² <https://www.sigfox.com/en>

o Sistema de Posicionamento Global (*Global Positioning System* - GPS). Outro serviço é o Sigfox Monarch, que fornece um serviço exclusivo de reconhecimento de rádio que permite que os dispositivos gerenciem as alterações de radiofrequência sem qualquer hardware adicional como GPS ou chipset Wi-Fi.

A empresa Sigfox também se utiliza de um protocolo próprio para a implementação do IoT, porém, diferentemente do LoRaWAN pertencente a LoRa Alliance, a arquitetura da Sigfox não está aberta, sendo seu uso restrito a empresa Sigfox e parceiros.

Figura 3 – Arquitetura proposta por Tyagi e Kumar (2017)



Fonte: Adaptado de <http://www.decom.ufop.br/imobilis/industria-4-0-e-iot/>

Mesmo sendo um tema bastante difundido, a Internet das Coisas ainda não possui uma arquitetura padrão definida, ainda que a LoRa Alliance esteja tentando definir a LoRaWAN como padrão. Dessa forma, a estrutura varia de acordo com quem implementará o conceito de IoT. Porém, visto essa diversidade de arquitetura, Tyagi e Kumar (2017) propuseram uma arquitetura simples e unificada, baseando-se em diferentes arquiteturas já existentes para Internet das Coisas. Essa arquitetura é dividida

em três camadas e é ilustrada pela Figura 3.

Na camada de aplicação são fornecidos os serviços solicitados pelos clientes. Essa camada aceita a solicitação do cliente para fornecer os serviços e contém os métodos de intercomunicação com aplicativos de uso pessoal. Já na camada de rede é responsável por adquirir dados dos sensores e dispositivos. A finalidade dos dispositivos e sensores é conectar-se por meio da camada de rede fornecida. Dessa forma a camada de redes é responsável por coletar e transmitir as informações. Enquanto que a camada de percepção lida principalmente com dados coletados, tags RFID e sensores, também se inclui leitores de códigos de barras, câmeras, GPSs, etc. Além disso, essa camada também é responsável por detectar e controlar o mundo físico (TYAGI; KUMAR, 2017).

Alguns dos conceitos das camadas expostos na Figura 3 se assemelham aos conceitos utilizados no modelo em camadas proposto na seção 4. Além dos conceitos das camadas, existem outros conceitos relacionados à Cidades Inteligentes e Internet das Coisas que se relacionam com tópicos desse trabalho.

2.2 Sistemas de Informações Geográficas

Um Sistema de Informação Geográfica (SIG) é formado por um conjunto de ferramentas relacionadas a informações espaciais e que são especializadas em adquirir, armazenar, recuperar, transformar e emitir tais informações. Esses dados geográficos descrevem objetos do mundo real em termos de posicionamento, seus atributos não aparentes e das relações topológicas existentes (P.A.BURROUGH, 1986 apud CAMARA; ORTIZ, 1998).

O SIG fornece um ambiente integrado para realizar análises de engenharia, segurança e ambientais. Ele permite que engenheiros e gerentes transformem dados em informações operacionais úteis para o gerenciamento ideal e seguro de infraestruturas durante seu ciclo de vida. Ele também oferece recursos avançados para visualização interativa 2D e 3D de dados dinâmicos, coletados por sensores inteligentes ou de qualquer outra forma (AFANEH; SHAHROUR, 2017).

Os sistemas de informações geográficas buscam facilitar o acesso e interpretação dos dados geográficos que são expressos normalmente em forma de números. Dessa forma, o SIG busca transformar esses números em informações perceptíveis pelos usuários, melhorando a análise das informações e propiciando melhores conclusões na manipulação de dados geográficos.

De acordo com Ferreira (2006), as primeiras tentativas de automatizar parte do processamento de dados com características espaciais, que até então eram feitos

somente de forma manual através de mapas, aconteceram na Inglaterra e nos Estados Unidos, em meados de 1950, com o objetivo principal de reduzir os custos de produção e manutenção de mapas.

Porém, os primeiros sistemas de informações geográficas só surgiram na década de 1960, no Canadá, como parte de um esforço governamental para criar um inventário de recursos naturais. Estes sistemas, no entanto, eram muito difíceis de usar. Não existiam monitores gráficos de alta resolução, os computadores necessários eram excessivamente onerosos e a mão de obra tinha que ser altamente especializada (FERREIRA, 2006).

Nesse contexto histórico, os SIGs surgiram como uma alternativa prática ao uso de mapas físicos, mesmo que no início devido as limitações tecnológicas existentes, ainda não se mostrassem tão vantajosos. Porém, os SIGs já mostravam uma vantagem em relação aos mapas físicos: um menor custo na sua manutenção. Isso serviu de justificativa e motivação para que outros países continuassem no seu estudo e desenvolvimento.

No final da década de 90 e início do século XXI, os SIGs começam a se tornar corporativos e orientados à sociedade, com a utilização da Internet, de bancos de dados geográficos distribuídos e com os esforços realizados em relação a interoperabilidade dos sistemas (FERREIRA, 2006).

Com essas evoluções, os SIGs passaram a ser utilizados por várias áreas e com diversas finalidades, sendo seu uso agregado também a outros conceitos, como cidades inteligentes e Internet das Coisas.

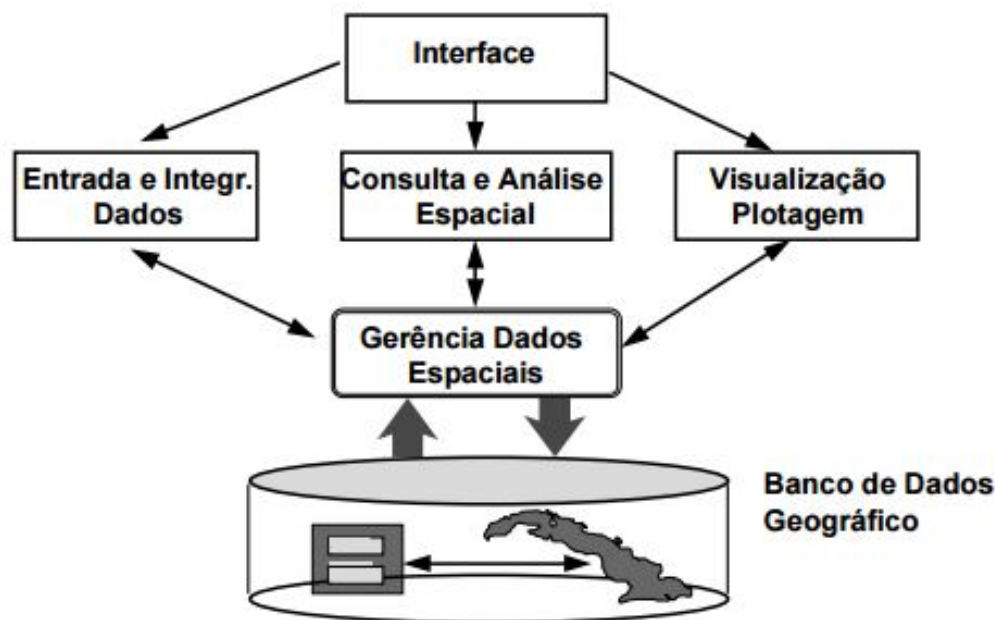
No contexto de cidades inteligentes, pode utilizar-se do SIG para realizar o planejamento urbano, auxiliando na instalação de, por exemplo: paradas de ônibus, estações de metrô, semáforos, rotatórias, etc. Podendo ser utilizado também para monitorar a cidade, juntamente com outros sensores. No cenário de Internet das Coisas, o SIG pode ser usado no contexto de mapear dispositivos, conseguindo catalogar as informações dos sensores e analisá-las da melhor forma, levando em consideração sua posição geográfica. A utilização do SIG pode unir esses dois conceitos. E esse contexto foi explorado por Afaneh e Shahrour (2017), onde os autores utilizaram alguns sensores em conjunto com as informações geográficas para conseguir monitorar alguns serviços de uma cidade, como abastecimento de água e de eletricidade.

A estrutura de Um SIG possui componentes que se relacionam de forma hierárquica, sendo essa estruturação abordada por Camara e Ortiz (1998) como mostrado na Figura 4.

Conforme a Figura 4, no nível mais próximo ao usuário, a interface homem-máquina define como o sistema é operado e controlado, enquanto que no nível in-

intermediário, um SIG deve ter mecanismos de entrada, processamento, visualização e saída de dados espaciais. No nível mais interno do sistema, um banco de dados geográficos lida com os dados espaciais e seus atributos (CAMARA; ORTIZ, 1998).

Figura 4 – Estrutura de um SIG proposta por Camara e Ortiz (1998)



Fonte: Camara e Ortiz (1998)

Verificando a estruturação de um SIG, pode-se analisar que a interface em uma aplicação pode servir apenas para a captura de dados, ou para visualizar os resultados obtidos através dos dados ou consultar e realizar a análise dos dados a fim de se obter resultados. Pode ainda ter esses serviços conjugados, conseguindo, assim, realizar a coleta e análise, ou coleta e visualização, entre outras formas de relação desses serviços. Entretanto, eles estão sempre interligados à gerência dos dados espaciais e ao banco de dados geográfico, que são a essência de um SIG. Essa variação de serviços que a interface pode oferecer vai depender de qual a proposta da aplicação e de que serviços essa proposta precisa para ser executada. Vale ressaltar que podem existir aplicações heterogêneas que unidas se completam, tornando-se um SIG. Ou então pode existir apenas uma aplicação com todos os serviços que só ela aliada ao banco de dados já é o SIG.

Nesse contexto, para realizar a captura dos dados geográficos, é necessário de um sistema de navegação por satélite, atualmente existem algumas opções desses sistemas, alguns existentes são:

- **GPS:** O Sistema de Posicionamento Global é um Sistema Global de Navegação por Satélite (*Global Navigation Satellite Systems - GNSS*) civil e militar, alimentado por informações de um sistema de 24 satélites (INFOESCOLA, 2018). Esse

sistema foi o primeiro sistema de navegação por satélites a ser desenvolvido e foi criado pelos Estados Unidos da América (EUA).

- **GLONASS:** É um GNSS que vem sendo desenvolvido pela Rússia desde 1976 e já conta com 24 satélites em órbita, responsáveis pelo o fornecimento de dados de posicionamento. Em português a sigla GLONASS significa Sistema de Navegação Global por Satélite ([OFICINADANET, 2016b](#)).
- **BeiDou:** É um sistema de satélites de posicionamento e navegação chinês. O sistema está atualmente em sua terceira geração e afirma atingir uma precisão de nível milimétrico que supera outros sistemas. No entanto, com apenas 22 satélites operacionais, o BeiDou apresenta uma pequena desvantagem em termos de precisão quando comparado ao GPS e ao GLONASS. É esperado mais satélites e maior precisão até 2020 ([TELIT, 2018](#)).
- **Galileo:** Foi Desenvolvido pela União Europeia em 2011, ele possui atualmente 14 satélites e seu objetivo é fornecer um serviço de posicionamento mais preciso em latitudes mais altas do que outros GNSS. Espera-se que o Galileo concorra com o alcance global do GPS até 2020 com o uso de 24 satélites em seis planos orbitais ([TELIT, 2018](#)).

Dentre essas tecnologias de posicionamento global que podem ser utilizadas para a captura dos dados geográficos, a mais popular é o GPS que é empregado em vários dispositivos, propiciando seu uso. Devido a esse maior uso, o GPS será explorado na subseção posterior.

2.2.1 Sistema de Posicionamento Global

O GPS é um sistema eletrônico de navegação civil e militar que emite coordenadas em tempo real e é alimentado por informações de um sistema de 24 satélites chamados de Satélite de Navegação com Tempo e Variação (*Navigation Satellite with Time and Ranging* - NAVSTAR) e controlado pelo Departamento de Defesa (*Department of Defence* - DoD) dos Estados Unidos da América ([INFOESCOLA, 2018](#)).

Estes satélites enviam sinais para o receptor (o aparelho de GPS), e então, o aparelho de GPS interpreta esses sinais dizendo exatamente a localização do aparelho. Para realizar tal processo, o GPS usa um sistema chamado de triangulação, que é ilustrado na Figura 5. Nesse processo, três satélites enviam o sinal para o receptor, que calcula quanto tempo cada sinal demorou a chegar nele ([OFICINADANET, 2016a](#)). Com isso, o GPS consegue fornecer a latitude e longitude do aparelho que requisitou a localização.

Figura 5 – Triangulação realizada pelos satélites

Fonte: <https://www.oficinadanet.com.br/post/12406-como-funciona-o-gps>

Além da sua localização terrestre obtida pela triangulação, o receptor GPS também consegue saber a sua altura em relação ao nível do mar (altitude), porém, para tal informação, é necessário um quarto satélite no processo de triangulação (OFICINADANET, 2016a).

O programa de GPS foi estabelecido no início dos anos 70, baseado em trabalhos anteriores em navegação de rádio. Tendo seu primeiro satélite operacional lançado em 1978. Em 1983, foi anunciado formalmente pelos Estados Unidos o fornecimento de GPS para uso civil, seguido de um anúncio em 1991 de que o GPS civil seria fornecido gratuitamente à comunidade internacional. Para realizar esse serviço, foi necessária uma constelação de 24 satélites, que foi chamada de NAVSTAR. Essa constelação só atingiu seu tamanho necessário de 24 componentes no ano de 1994.

Naquela época, o GPS estava sendo reconhecido e aprovado por diversas usuários que estavam explorando seus benefícios, como topógrafos, comunidades de aviação e organizações militares. Porém, mesmo com esse grau de sucesso, existiam limitações nos dispositivos GPS civis, chamado de disponibilidade seletiva (*selective availability* - SA), uma redução intencional da precisão do sinal civil destinada a garantir que os Estados Unidos e seus usuários militares aliados tivessem acesso exclusivo

à precisão total do GPS enquanto que os demais usuários tinha uma precisão limitada (BETZ, 2016).

Como exposto nos últimos parágrafos, o GPS surgiu para fins militares no período da Guerra Fria (1947 – 1991) e foi aberto ao público logo após esse conflito ideológico se encerrar. Porém, os Estados Unidos ainda não confiavam nas outras nações, mantendo o SA para que somente os militares americanos usufruíssem da precisão total do GPS.

A disponibilidade seletiva só veio a ser desligada em maio de 2000, atraindo ainda mais o uso do GPS. Porém, um evento independente do desligamento da SA acelerou indiretamente o uso generalizado do GPS, esse evento foi o requerimento da Comissão Federal de Comunicação (*Federal Communications Commission - FCC*), nesse requerimento os prestadores de serviços determinaram que o GPS era a maneira mais econômica para fazer com que os telefones celulares fossem “capazes de localização” ao fazer chamadas de emergência. Isso resultou em um investimento significativo em *chipsets* GPS pequenos, baratos e de baixo consumo que aumentaram drasticamente a utilidade e a utilização do GPS (BETZ, 2016).

Com o fim do SA, o GPS passou a ser explorado e empregado em diversas problemáticas, se tornando mais popular e ganhando mais investimento. A popularização se deu por causa da FCC e seu requerimento, que incitou a implantação dos sensores GPS nos celulares a fim de identificar a localização do aparelho nas chamadas de emergência, mas que posteriormente seria mais impactante com o surgimento dos smartphones, agregando várias funcionalidades ao GPS.

Atualmente o GPS é amplamente utilizado para diversas funcionalidades. Para atender a essa demanda mais de 30 satélites operam simultaneamente, com satélites adicionais em órbita em estado de espera, prontos para serem ativados caso um satélite em operação for retirado de serviço (BETZ, 2016).

2.3 Sistemas de Recomendação

Com o desenvolvimento e crescimento da internet e aplicativos afins, ocorreu uma grande quantidade de informações geradas e que poderiam ser melhor tratadas. Para atender essa demanda surgiram os Sistemas de Recomendação (*Recommender Systems - RSs*), que buscam acessar a informações relevantes de acordo com o perfil do usuário que as buscam, sem a necessidade de uma pesquisa web manual (KURMASHOV; LATUTA; NUSSIPBEKOV, 2015).

Nos dias atuais, esses sistemas de recomendação são utilizados por diversos sites para promover a venda de seus produtos. Um dos principais atrativos é evitar que

o consumidor precise demandar de muito esforço para encontrar o que procura, dessa forma os RSs são aplicados para resolver questões como essas: “Outros produtos que você pode gostar”, “clientes que compraram este item também compraram”, “você pode gostar disso” (SHAH et al., 2017).

Numa visão mais específica, os RSs são ferramentas e técnicas de software para recuperação e filtragem de informações que visam fornecer recomendações de itens significativas e eficazes para o usuário ativo (SHAH et al., 2017 apud GATZIOURA; SÀNCHEZ-MARRÈ, 2015).

O principal objetivo do RS, como seu nome já evidencia, é gerar recomendações significativas de itens ou produtos para um conjunto de usuários. Alguns exemplos de seu uso na indústria são: sugestões de livros na Amazon³, ou filmes no Netflix⁴ (MELVILLE; SINDHWANI, 2010).

O formato dos mecanismos de recomendação varia de acordo com o domínio e as características particulares dos dados disponíveis. Com isso, os sistemas de recomendação diferem na maneira como eles analisam as fontes de dados para desenvolver noções de diversidade entre usuários e itens e com essas noções conseguir detectar pares para recomendação (MELVILLE; SINDHWANI, 2010). Devido a essa especificidade requerida, onde o RS varia de acordo com a base de dados utilizada e em qual contexto é empregado, existem algumas categorias para os sistemas de recomendação, eles se classificam em: filtragem colaborativa, filtragem baseados em conteúdo (cognitiva) e híbrido (SHAH et al., 2017).

No método de Filtragem Colaborativa (Collaborative Filtering - CF), as recomendações para cada usuário são geradas fazendo comparações com o gosto de uma alternativa em relação a outro usuário que qualificou o produto de maneira semelhante ao usuário ativo. A CF baseia-se na ideia de que as pessoas que concordam com a avaliação de itens no passado provavelmente concordarão novamente no futuro. Os métodos de CF são agrupados em dois métodos gerais baseados em vizinhança e baseados em modelos (SHAH et al., 2017).

De acordo com Melville e Sindhvani (2010), na filtragem colaborativa baseada em vizinhança um subconjunto de usuários é escolhido com base em sua similaridade com o usuário ativo, e uma combinação ponderada de suas classificações é usada para produzir previsões para esse usuário. A maioria dessas abordagens pode ser generalizada pelo algoritmo resumido nas seguintes etapas:

- Atribuir um peso a todos os usuários com relação à semelhança com o usuário ativo;

³ <https://www.amazon.com.br/>

⁴ <https://www.netflix.com/>

- Selecione k usuários que tenham a maior semelhança com o usuário ativo (comumente chamado de vizinhança);
- Calcule uma previsão a partir de uma combinação ponderada das classificações dos vizinhos selecionados.

Já as técnicas baseadas no modelo de filtragem colaborativa baseada em modelos assumem que a similaridade entre usuários e itens é simultaneamente induzida por alguma estrutura oculta de dimensão inferior nos dados. Por exemplo, a classificação atribuída a um usuário a um filme pode depender de alguns fatores implícitos, como o gosto do usuário em vários gêneros de filme (MELVILLE; SINDHWANI, 2010).

Visto o conceito de filtragem colaborativa baseada em vizinhança e baseado em métodos, pode-se perceber que o método baseado em vizinhança se utiliza da estatística entre valores explícitos relacionados aos usuários para poder realizar a recomendação, ou seja, através dos dados já coletados, é realizada a predição. Enquanto que o conceito baseado em métodos se utiliza de valores explícitos para adquirir conhecimentos e realizar as recomendações de novos itens, ou seja, através dos dados já coletados é possível verificar se um novo item é compatível ou não.

Além do método de filtragem colaborativa, existe também o método de filtragem baseado em conteúdo que pode ser chamado também de filtragem cognitiva. Nesse método, ocorre a recomendação dos itens com base em uma comparação entre o conteúdo dos itens e os dados de um perfil de usuário. O conteúdo de cada item é representado como um conjunto de descritores, ou vetor de recursos. O filtro baseado em conteúdo analisa o conjunto de descritores que são alimentados como entrada para um item específico que é avaliado anteriormente pelo usuário. Este filtro, então, constrói um modelo de interesse dos usuários que gera recomendações (SHAH et al., 2017). Um exemplo para tal método é a recomendação de livros, onde se um usuário ler livros do gênero romance, como: “Para Todos os Garotos que Já Amei”, “Simplesmente Acontece” e “P.S: Eu Te Amo”, serão recomendados mais livros do mesmo gênero, como: “A Seleção”, “Um Dia”, etc.

E por último existe também a abordagem híbrida que consiste na união entre o método de filtragem colaborativa e filtragem baseado em conteúdo a fim de alavancar os pontos fortes existentes nesses dois métodos (MELVILLE; SINDHWANI, 2010). Existem diversas formas de integração entre esses dois métodos e de acordo com Shah et al. (2017), o método de recomendação híbrida pode ser colocado em prática das seguintes maneiras:

- **Ponderado:** Nesse tipo de recomendação, o peso de um item recomendado é calculado a partir dos resultados de todas as técnicas de recomendação dispo-

níveis implementadas no sistema.

- **Comutação:** Nesta abordagem, o sistema escolhe um componente em particular e aplica a técnica de recomendação que melhor atende a sua finalidade. Por exemplo, pode-se utilizar uma técnica para recomendar livros e outra para recomendar usuários.
- **Mista:** Diferentes abordagens de recomendação fornecem recomendações diferentes que serão introduzidas juntas. Este sistema híbrido é baseado na fusão e apresentação de várias listas classificadas em uma única lista classificada.
- **Combinação de Recursos:** O sistema híbrido é dividido ou separado em duas partes: contribuição e recomendação real. Esses dois coexistem no sistema. O usuário que recomenda depende dos resultados ou da saída de dados da recomendação de contribuição. A técnica híbrida de combinação de recursos permite que o sistema considere os dados de saída de recomendação colaborativa sem depender dele exclusivamente, o que diminui a sensibilidade do sistema.
- **Aumento de Recursos:** O híbrido de aprimoramento de recursos é semelhante aos híbridos de combinação de recursos, mas a diferença é que o colaborador fornece uma característica interessante. O híbrido de aprimoramento de recursos é mais elástico do que o método de combinação de recursos.

Esses tipos de sistemas de recomendação podem ser utilizados em conjunto com algoritmos de aprendizagem de máquina (*machine learning*). Segundo o site [SAS \(2018\)](#), “O aprendizado de máquina é um método de análise de dados que automatiza a construção de modelos analíticos. É um ramo da inteligência artificial baseado na ideia de que sistemas podem aprender com dados, identificar padrões e tomar decisões com o mínimo de intervenção humana”. Alguns exemplos de algoritmos de *machine learn* são o *k*-vizinhos mais próximos (*k-Nearest Neighbors* - *k*-NN), árvore de decisão, *K-means*, etc.

Para a realização da análise dos dados nesse trabalho foi optado pelo uso do *K-means*. Sua escolha é justificada pelo seu método de implementação menos complexo, e pelo conceito do centroide que possibilita uma comparação de dados mais simples, já que o objetivo do *k-means* nesse trabalho não é realizar a recomendação propriamente dita, mas sim de auxiliar na mesma reduzindo os cálculos.

2.3.1 *K-means*

Antes de iniciar o detalhamento do algoritmo *K-means* em si, é necessário entender primeiramente que os algoritmos de aprendizagem de máquina se di-

videm em duas categorias, a aprendizagem supervisionada e aprendizagem não-supervisionada.

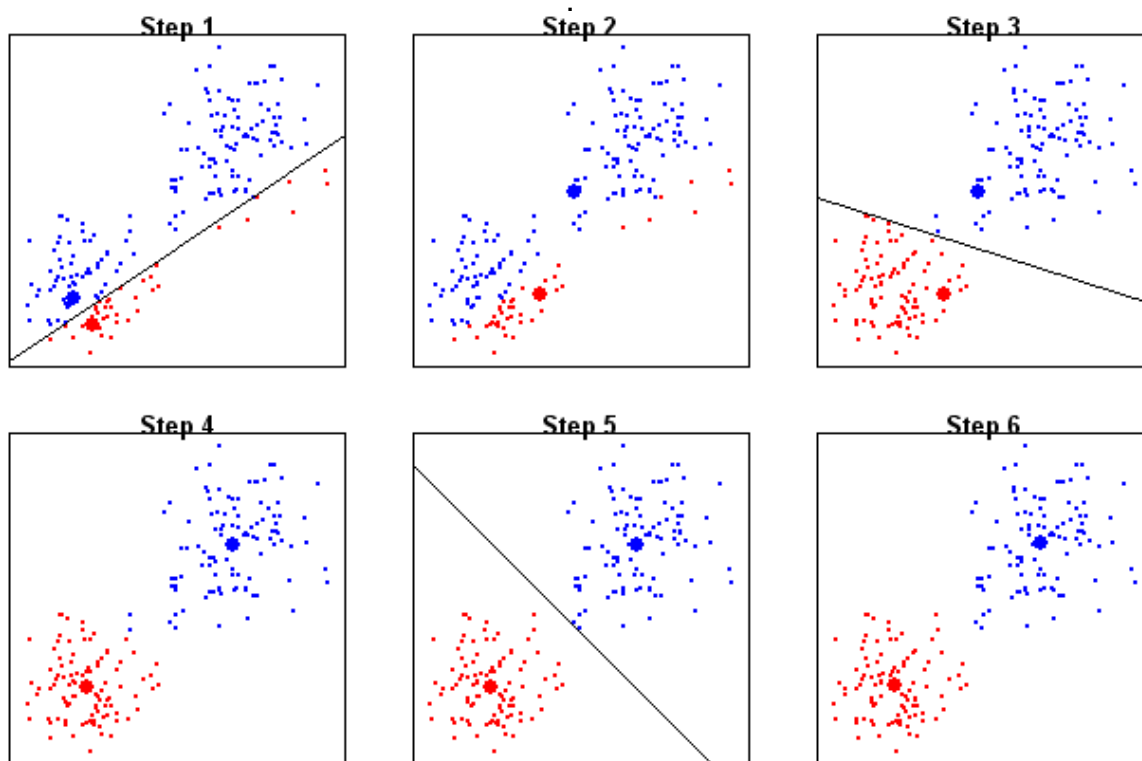
A aprendizagem supervisionada é o termo usado sempre que o programa é “treinado” sobre um conjunto de dados pré-definido. Baseado no treinamento com os dados pré-definidos, o programa pode tomar decisões precisas quando recebe novos dados (CIENCIAEDADOS, 2017). Com isso o algoritmo supervisionado precisa de dados para treinamento antes de realmente poder tomar decisões.

Já a aprendizagem não supervisionada ou *clusterização* é um método de segmentação de dados que partilham tendências e padrões semelhantes. Este tipo de estudo não tem uma variável dependente, não precisando ser treinado para tomar decisões. As técnicas de clusterização procuram semelhanças e diferenças num conjunto de dados e agrupam os registos semelhantes em segmentos ou *clusters*, de uma forma automática, de acordo com algum critério ou métrica. Não é necessário definir os grupos nem os atributos que devem ser utilizados para segmentar o conjunto de dados. (UNIVERSIDADEDEPORTO, 2018).

Dentro desse contexto de aprendizagem de máquina o *K-means* se insere nos algoritmos de aprendizagem não-supervisionada, ele foi proposto por JBMacQueen em 1967 e é um algoritmo clássico de agrupamento (YANG; DENG, 2010).

Primeiramente, o *K-means* seleciona aleatoriamente k pontos de todo o conjunto de dados que representam os centroides iniciais (ou sementes). Cada ponto remanescente no conjunto de dados é atribuído a um *cluster* cujo centroide está mais próximo desse ponto. Após isso, as coordenadas do centroide são recalculadas atualizando o seu valor que corresponde à média de todos os pontos atribuídos ao seu *cluster*. Esse processo itera até que uma função de custo seja atendida, ou após ocorrer os cálculos de atualização do valor do centroide e esse valor não for modificado, significando que os *clusters* já estão definidos. A precisão das análises de *cluster* de grandes conjuntos de dados usando *K-means* depende da precisão dos métodos de inicialização do centroide que são adaptados para conjuntos de dados distribuídos em várias máquinas (Qi et al., 2016).

Na Figura 6, o funcionamento do *K-means* é ilustrado. A imagem mostra seis passos (*Steps*) e a categorização dos dados em dois *clusters*, possuindo, consequentemente, dois centroides. No primeiro passo, são atribuídos valores aleatórios aos centroides e então os dados são divididos em dois grupos de acordo com esses centroides. No passo 2, os valores dos dois centroides são recalculados a partir dos dados que estavam em seus respectivos *clusters*, possuindo novos valores. No terceiro passo os dados são realocados nos dois grupos com base nos novos valores dos centroides. No quarto passo os valores dos dois centroides são recalculados a partir dos dados que estavam em seus respectivos *clusters*, assim como no passo 2. Já no

Figura 6 – Funcionamento do algoritmo *K-means*

Fonte: <https://medium.com/@dilekamadushan/introduction-to-k-means-clustering-7c0ebc997e00>

quinto passo os dados são realocados nos dois *clusters* com base nos novos valores dos centroides, assim como ocorreu no passo 3. E no sexto passo os valores dos dois centroides são recalculados novamente, porém como não houveram alterações nesses valores o algoritmo foi encerrado.

Como visto o *K-means* basicamente alterna entre duas ações, atribuir valor aos centroides e organizar os dados em *clusters* de acordo com sua distância para os centroides existentes, sendo que os dados são alocados aos grupos em que os centroides estão mais próximos aos dados. Vale ressaltar que, para a atribuição de valores, no primeiro caso, os centroides recebem valores aleatórios de dados existentes no conjunto e que, após isso, os valores atribuídos aos centroides têm como base os dados existentes dentro do seu *cluster*.

3 TRABALHOS RELACIONADOS

A ciência de dados vem sendo explorada por vários pesquisadores com o intuito de oferecer soluções que podem ser aproveitadas no cotidiano, às vezes se utilizando de conceitos como Cidades Inteligentes, Internet das Coisas e Sistemas de Informações Geográficas. Para oferecer essas soluções, os pesquisadores podem utilizar dados já existentes gerados pelas pessoas através de, por exemplo, redes sociais, ou então pela criação de aplicações próprias para auxiliarem na coleta de dados. Essas informações podem ser utilizadas para diversas finalidades e podem ser de diversos tipos, inclusive de localização.

Uma dessas funcionalidades foi explorada por [Gao et al. \(2013\)](#), a ideia proposta é de utilizar a localização juntamente com o tempo, ele se baseia em *check-ins* dependentes do tempo, utilizando um *dataset* da rede social Fousquare¹, a qual ele analisa os *check-ins* dos usuários baseando-se no tempo do mesmo. Após essa análise, busca-se sugerir locais para o usuário tomando como base a localização, horário atual do mesmo e *check-ins* passados.

Além dessa abordagem de localização pura, pode ser utilizada a ideia de trajetória, segundo [Nogueira et al. \(2018\)](#), “quando os dados de localização são constantemente coletados e associados a um instante no tempo, temos uma trajetória”. A ideia de trajetória foi abordada por [Zheng, Xie e Ma \(2010\)](#), que em seu trabalho extraíram as informações das redes sociais e para tratar os dados foi utilizado a ideia de árvore de decisão, com o objetivo de sugerir localizações, viagens, recomendações de atividades ou recomendações de amizade, baseado em trajetórias feitas pelos usuários e na localização deles. Já [Dodge, Weibel e Frootan \(2009\)](#) fazem a análise de trajetórias com a finalidade de identificar o modo de transporte de trajetórias. Com isso, além da trajetória, eles se baseiam em outras variáveis como distância, velocidade, aceleração, etc. Com essas informações, é possível determinar se a rota foi realizada por pedestre, bicicleta, moto ou carro.

Esse contexto de trajetórias foi abordado por [Damerow, Klingelschmitt e Egger \(2016\)](#). Eles analisam os padrões de movimentações de veículos a fim de verificar a probabilidade de não interação e com isso fazer uma predição precoce de acidentes para tentar evitá-los. A interação é uma alteração no padrão, como por exemplo: desacelerar ou mudar de faixa. Foi proposto que, se houver chance de ocorrer interseção (colisão) entre dois veículos, é analisada a probabilidade de não interação, ou seja, a probabilidade de os motoristas não fazerem nada para evitar a colisão. Caso essa probabilidade seja maior que 70%, significa que há chance de haver colisão. Com

¹ <https://pt.foursquare.com/>

isso, o sistema alerta o motorista para ele interagir e evitar o acidente.

As trajetórias podem ser exploradas também mais voltadas às atividades físicas, como abordado por [Evans et al. \(2013\)](#). Nesse trabalho, foi utilizada a ideia de “rotas quentes”, usando a localização e analisando rotas. Essas podem ser de corredores ou ciclistas. Com a análise, são buscadas as mais frequentes e podem ser utilizadas a fim de oferecer indicação de rotas para corredores ou ciclistas próximos a elas.

Nessa ideia existe também o FrameSTEP, proposto [Nogueira et al. \(2018\)](#). O FrameSTEP tem como objetivo gerar episódios utilizando a ontologia STEP, esse conceito de episódio é explicado por [Mountain e Raper \(2001\)](#), “um episódio representa um período de tempo discreto para o qual o comportamento espaço temporal do usuário era relativamente homogêneo”. Esses episódios são gerados a partir de dados da localização de corredores, como latitude, longitude e *timestamp*. Então é possível formular consultas que identifiquem padrões nos episódios e encontrem, por exemplo, usuários que compartilhem o mesmo comportamento.

O contexto de trajetórias está presente no SiCliSIG, uma vez que os dados utilizados para análise são rotas de ciclistas, rotas essas que estão enquadradas em trajetórias. Como diferencial, o trabalho proposto busca, através dessas informações, realizar a recomendação de ciclistas.

Como visto, a análise de dados recolhidos através de sensoriamento pode ser explorada para diversas finalidades, entrando no contexto de internet das coisas, cidades inteligentes e sistemas de informações geográficas. Dessa forma, a análise de trajetórias de informações coletadas pode ser utilizada também para o planejamento urbano. Ele se insere no contexto das cidades inteligentes, esse tema foi abordado por alguns autores buscando utilizar de dados gerados por ciclistas, a fim de auxiliar nesse planejamento.

Nesse contexto de analisar as informações, [Chen e Jakubowicz \(2015\)](#). buscaram identificar padrões de mobilidade urbana de bicicletas através de dados coletados pelo sistema de bicicletas compartilhadas. Esse estudo de padrões de viagem pode ser utilizado em outros meios de transportes, auxiliando na detecção da mobilidade urbana e pode ser utilizado no planejamento urbano, a fim de detectar quais pontos de bicicletas compartilhadas possui maior demanda ou carência de bicicletas.

O estudo de bicicletas com planejamento urbano foi abordado por outros autores no contexto de *bikes* compartilhadas, onde são estudadas maneiras de melhorar a cidade através de dados coletados pelas estações ou detectar formas de aprimorar a distribuição das bicicletas através da demanda de empréstimos e devoluções. Esses tópicos foram explorados por [Wang e Wang \(2013\)](#) e [Yang, Lin e Chang \(2010\)](#).

Nesse cenário de planejamento urbano, o SiCliSIG busca, através de rotas percorridas por ciclistas, fornecer maneiras de auxiliar na tomada de decisão referente a obras de infraestrutura que beneficiem o ciclismo, sendo uma abordagem de planejamento urbano diferente das citadas anteriormente.

Além do planejamento urbano, existe a vertente de utilizar informações geográficas para a sugestão de rotas, que é explorado por [Torres et al. \(2015\)](#). Nesse trabalho, é utilizado um aplicativo para coletar as informações dos ciclistas e então oferecer alternativas de trajetos mais seguros para os mesmos. Esse aplicativo funciona como um planejador de rotas, onde o ciclista coloca seu destino e a aplicação calcula o melhor caminho para o ciclista, levando em consideração a distância, o tipo de rua (avenida, ciclovia, etc.) e a frequência em que os ciclistas trafegam nessas ruas. Com isso, o ciclista pode realizar seu trajeto com mais segurança.

No contexto de sugestão de rotas, foi estudado também por [Hrnčíř et al. \(2017\)](#), no trabalho em questão foram explorados alguns critérios que podem auxiliar no planejamento de rotas, melhorando a experiência do ciclista na sua viagem. Dessa forma, foi proposto um algoritmo de roteamento multicritério que sugere caminhos para o destino pedido pelo usuário se baseando em critérios como por exemplo: menor tempo, conforto e ganho de elevação.

A ideia do ciclismo como meio de transporte foi explorada por [Herrmann et al. \(2018\)](#), que buscaram agregar bem-estar ao percurso dos ciclistas. Nesse artigo, foi falado a respeito da emissão de poluentes que carros com motor à combustão podem proporcionar. Além disso, foi mostrado como solução o uso de Veículos Híbridos (*Hybrid Vehicles* - HV) e Veículos Elétricos (*Electric Vehicle* - EV), para reduzir a emissão de poluentes. Além do mais, o trabalho proposto por [Herrmann et al. \(2018\)](#) tem como objetivo incitar a prática de ciclismo, buscando detectar os ciclistas próximos aos veículos, para que esses veículos parem ou reduzam sua emissão de poluentes por um tempo a fim de proporcionar um ar mais limpo aos ciclistas. Para tanto, se o veículo possuir motor à combustão, ele mudará sua programação para uma forma mais econômica a fim de reduzir emissão de poluentes. Porém, se o veículo for do tipo HV, ele passará a utilizar o motor elétrico para cessar a poluição e se for do tipo EV, nada precisará ser feito, pois ele não emite poluentes. Além disso, quando um carro detecta um ciclista, ele emite um aviso para os carros próximos avisando sobre a presença do mesmo, para que os outros veículos também entrem no modo menos poluente.

O trabalho proposto compactua de visões dos trabalhos relacionados, como o uso de tecnologia da informação e comunicação aliadas a bicicletas. Nessa visão, este trabalho utiliza de ciência de dados para analisar informações e conseguir resultados através de rotas geradas por ciclistas. O diferencial desse trabalho, está

primeiramente em classificação de ciclistas com perfis parecidos através de análise espacial e temporal. Além disso, também está o diferencial de utilizar das informações disponíveis para montar um mapa de calor a fim de auxiliar no planejamento urbano. Neste trabalho a cidade em questão é Aracati-CE no Brasil.

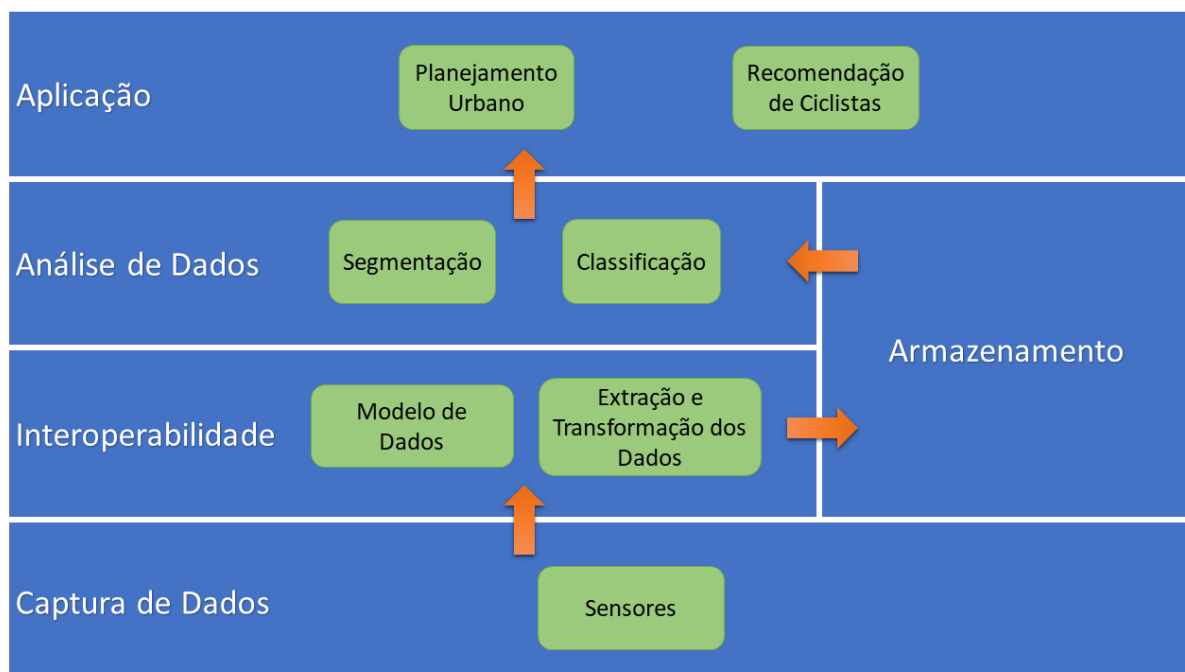
4 Proposta

O smartphone é um poderoso aliado na geração de dados devido aos seus diversos sensores embarcados. Por meio da utilização destes sensores, propõe-se a captura e análise das trajetórias de ciclistas.

Esta proposta possui uma arquitetura em camadas, sendo cada uma responsável por uma atividade dentro do contexto da contribuição deste trabalho. O SiCliSIG é dividido em duas etapas, uma delas é a coleta de dados através do smartphone, a outra consiste em uma aplicação *desktop* para a análise das informações coletadas, buscando identificar similaridades entre os ciclistas e auxiliar no planejamento urbano.

Nesse capítulo, é exposta cada uma das camadas existentes, mostrando-as e explicando suas principais atribuições. Como mostra a Figura 7, a proposta desse trabalho está dividida em cinco camadas: Captura de Dados, Interoperabilidade, Armazenamento, Análise de Dados, e Aplicação.

Figura 7 – Modelo em camadas para análise de dados



Fonte: Elaborado pelo autor.

4.1 Camada de Captura de Dados

A primeira camada consiste na captura de dados. Esta camada é responsável por interagir com o ambiente sensoreado e coletar os dados gerados por esse

ambiente. Para que isso ocorra, é necessário utilizar sensores que representam o comportamento do usuário naquele ambiente, capturando as informações.

4.1.1 Sensores

Os sensores, como mostrado no Capítulo 2, são bastante utilizados nos conceitos de cidade inteligentes, internet das coisas e sistemas de informações geográficas. Nesse contexto, essa etapa de captura dos dados se utiliza do sensoriamento, que é fundamental por se tratar de dados geográficos. Ao identificar a necessidade de alguns sensores, foi visto que o *smartphone* poderia suprir essa exigência, pois a grande maioria vem equipado com diversos sensores, os quais são capazes de coletar e armazenar essas informações geográficas.

Os *smartphones* possuem alguns sensores e entre eles os mais comuns são: acelerômetro, giroscópio, localização, etc. Além desses sensores, o *smartphone* é compatível com alguns tipos de sistemas como o WiFi e o Bluetooth. A parte de sensoriamento de *smartphones* já foi explorada e mostrada como uma boa aliada à internet das coisas, oferecendo uma diversidade de funções (LANGMANN, 2018).

Dentre os sensores existentes nos celulares, foram utilizados o sensor GPS para coletar os dados geográficos, trabalhando em conjunto com a internet, a fim de melhorar a precisão dos dados. A internet é fornecida pelo sistema WiFi ou então pelos dados móveis providos pelo *chip* de telefonia do celular.

4.2 Camada de Interoperabilidade

Após a realização da coleta de dados na primeira camada, tem-se a necessidade da camada de interoperabilidade. De acordo com o site [Governo Digital \(2018\)](#):

A interoperabilidade pode ser entendida como uma característica que se refere à capacidade de diversos sistemas e organizações trabalharem em conjunto (interoperarem) de modo a garantir que pessoas, organizações e sistemas computacionais interajam para trocar informações de maneira eficaz e eficiente.

Visto esse conceito, a camada de interoperabilidade tem o objetivo de armazenar essas informações obtidas pelos sensores nos *smartphones*, garantindo uma comunicação entre sistemas heterogêneos.

Essa camada se divide em duas etapas, a primeira determina a modelagem dos dados armazenados em arquivos das informações coletadas pelos sensores e a segunda descreve a extração e a transformação desses dados para serem armazenados em um servidor a fim de melhor manipulá-los.

4.2.1 Modelo de Dados

Os dados, ao serem coletados, são armazenados em um arquivo de Linguagem de Marcação Extensiva (*Extensible Markup Language* - XML) cujo objetivo é organizar os dados, facilitando a manipulação dos mesmos. Esse arquivo de texto é criado após selecionar a opção iniciar do aplicativo, ele recebe como nome a data em que o percurso está sendo traçado, juntamente com o nome do percurso inserido pelo usuário, podendo ser o nome do destino ou algum outra nomenclatura de preferência do usuário. Dessa forma, caso o usuário faça um percurso dia 22/03/2018 e o nome do percurso seja “canoa”, o nome do arquivo será: *22-03-18_canoa.txt*.

A estrutura do arquivo é composta por *tags* e, internamente a essas *tags*, existem *sub-tags* como ilustrado na Figura 8. Nessa estrutura, a cada captura de dados realizada pelo aplicativo móvel, é gerada uma *tag* chamada *captura*. Dentro dessa *tag* existem *sub-tags*, que são preenchidas com os dados coletados pela aplicação móvel. Logo, a cada captura de dados, é adicionado ao arquivo uma *tag captura*, juntamente com suas *sub-tags* com os valores já preenchidos.

Na *sub-tag* tempo fica armazenado o tempo da captura dos dados no formato *timestamp* que segundo o site *Unix TimeStamp*, “o registro *timestamp* do *Unix* é uma maneira de rastrear o tempo como um total de segundos em execução. Essa contagem começa na *Unix Epoch* em 1º de janeiro de 1970, na UTC. Portanto, o registro de data e hora do *Unix* é meramente o número de segundos entre uma data específica e a *Unix Epoch*” ([UNIXTIMESTAMP, 2018](#)). Esse formato de tempo é uma alternativa ao formato *datetime* que é comumente utilizado, além disso o *timestamp* consegue ser compatível com o formato *geography* do banco de dados, facilitando sua armazenagem juntamente com os dados geográficos capturados pelo aplicativo.

Além da *sub-tag* tempo, existe também as *sub-tags* latitude e longitude. Latitude é a distância, em graus, de qualquer ponto da superfície terrestre em relação à Linha do Equador, já a longitude é a distância também em graus de qualquer ponto da superfície terrestre em relação ao Meridiano de Greenwich ([ALVES., 2018](#)).

Por último, existe a *sub-tag* altitude, que representa a distância vertical (em metros) de determinado ponto em relação ao nível do mar ([TODAMATERIA, 2018](#)). Essas três últimas *sub-tags* têm o objetivo de determinar o ponto exato da coleta dos dados. Com a combinação desses dados, pode-se ter informações a respeito do percurso, como a velocidade, distância, tempo para percorrer o percurso, altimetria etc.

Após mostrar e especificar as *tags* e *sub-tags* existentes no arquivo, vale ressaltar que cada captura de dados ocorre de 10 em 10 segundos, caso o objeto esteja em movimento. Nessa captura, é escrito no arquivo uma *tag* captura com as *sub-tags*

tempo, latitude, longitude e altitude preenchidos. A Figura 8 mostra como os dados ficam distribuídos no arquivo, observando-se a existência de duas capturas.

Figura 8 – Arquivo gerado pelo aplicativo móvel com os dados coletados

```
<captura>
  <tempo>1521640504</tempo>
  <latitude>-4.5736984</latitude>
  <longitude>-37.790986</longitude>
  <altitude>0.0</altitude>
</captura>
<captura>
  <tempo>1521640510</tempo>
  <latitude>-4.5736386</latitude>
  <longitude>-37.7909949</longitude>
  <altitude>0.0</altitude>
</captura>
```

Fonte: Elaborado pelo autor.

4.2.2 Extração e Transformação dos Dados

Após a coleta de dados realizadas pelos sensores, as informações das trajetórias se encontram armazenadas em diversos arquivos espalhados nos celulares dos usuários que as coletaram. Isso se torna um dos desafios para acessar essas informações, por estarem em dispositivos diferentes e dificultando a manipulação de dados por estarem armazenados em arquivos, mesmo que sejam arquivos estruturados.

Com o intuito de facilitar o acesso aos dados, é necessário agrupar os arquivos existentes nos diversos dispositivos dos usuários que coletaram os dados para um local centralizado, melhorando o acesso às informações dos arquivos. Já para facilitar a manipulação dos dados, as informações contidas nos arquivos devem ser extraídas e armazenadas em um banco de dados. Finalmente, para viabilizar esses processos, é necessária a utilização de um *scanner* e de um *parser*.

Para extrair as informações dos arquivos, foi necessário realizar a criação de um *scanner*, que em português é conhecido como analisador léxico. O *scanner* é a primeira etapa ocorrida em compiladores e sua tarefa é transformar um fluxo de caracteres em um fluxo de palavras na linguagem de entrada. Cada palavra precisa ser classificada em uma categoria sintática, ou “classe gramatical” (COOPER; TORCZON,

2014a). Com o analisador léxico, é possível identificar as *tags* e *sub-tags* contidas no arquivo, extraindo as informações para serem armazenadas no banco de dados.

Após a extração das informações do arquivo, primeiramente são calculados alguns valores a partir das informações da rota, como o tempo e distância do percurso. Para realizar o cálculo do tempo, foi feita a diferença do tempo inicial (horário da primeira captura) e do tempo final (horário da última captura), obtendo-se assim o tempo do percurso.

Já para calcular a distância do percurso, foi utilizada a fórmula de *Haversine* mostrada na Equação 4.1, que leva em consideração a curvatura da terra, possibilitando resultados mais precisos.

A fórmula de *Haversine* se baseia na latitude (ϕ) e longitude (λ) de dois pontos para determinar a distância entre eles. Os valores de latitude e longitude são convertidos em radianos para serem inseridos na fórmula e a constante "r" representa o raio da terra que pode possuir o valor de 6371 km quando a medida da distância buscada for em quilômetro ou possuir o valor de 3956 mi para a medida da distância em milhas (CAMPOS, 2017). A distância total da rota, representada por "d" é obtida através da soma das distâncias entre cada ponto que integra a rota.

$$d = 2r * \arcsin\left(\sqrt{\sin^2\left(\frac{\phi_2 - \phi_1}{2}\right) + \cos(\phi_1) * \cos(\phi_2) * \sin^2\left(\frac{\lambda_2 - \lambda_1}{2}\right)}\right) \quad (4.1)$$

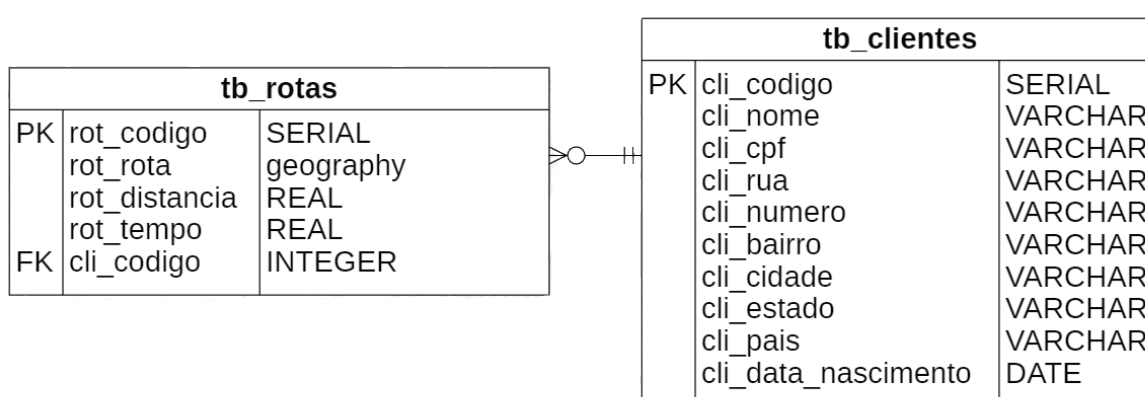
Porém, após extrair os dados contidos no arquivo e calcular as informações necessárias a partir deles, é necessário armazená-los em um banco de dados. Para isso, é utilizado um *parser*.

O *parser*, também conhecido como analisador sintático, remete a uma etapa ocorrida em compiladores chamada de análise sintática. De acordo com o livro *Construindo Compiladores* (COOPER; TORCZON, 2014b), o *parser* vê um fluxo de palavras, onde cada palavra está associada a uma categoria sintática. Nessa aplicação, o *parser* tem como objetivo armazenar as informações extraídas pelo analisador léxico em um banco de dados, analisando os comandos em linguagem de programação, juntamente com as informações e processando os dados no banco de dados. Dessa forma, o *scanner* é utilizado para extrair as informações dos arquivos, enquanto que a utilização do *parser* está relacionada ao armazenamento das informações extraídas do arquivo no banco de dados.

4.3 Camada de Armazenamento

Após a extração das informações contidas no arquivo de texto realizadas pelo *parser*, elas são armazenadas em um banco de dados, que é composto por duas tabelas relacionadas. Uma é responsável por armazenar as informações dos usuários e a outra armazena as rotas que cada utilizador percorreu durante o uso da aplicação. Nessa subseção, é exposto um pouco mais a respeito da disposição do banco de dados e suas tabelas.

Figura 9 – Estrutura do Banco de dados responsável pelo armazenamento das informações



Fonte: Elaborado pelo autor.

Como visto na Figura 9, a tabela `tb_clientes` é a tabela responsável por armazenar as informações dos usuários. Nela estão armazenados os dados das pessoas que utilizaram o aplicativo e tiveram suas rotas salvas no banco de dados.

As informações armazenadas nessa tabela têm objetivos distintos. Os campos que armazenam essas informações são divididos em duas classes: identificação e prescrição. Na primeira classe, os campos são responsáveis por garantir a identificação do usuário, distinguindo-o dos demais, esses campos são: `cli_codigo`, `cli_nome`, `cli_cpf`. Já na classe de prescrição, os campos fornecem informações extras que podem auxiliar na indicação de rotas aos usuários. Esses campos são: `cli_ rua`, `cli_número`, `cli_bairro`, `cli_cidade`, `cli_estado`, `cli_pais` e `cli_data_nascimento`. Com as informações dessa segunda classe é possível ampliar o leque de sugestões aos usuários, pois, além de se basear nas informações das rotas, pode-se basear na localização ou idade do ciclista. Assim, o usuário só recebe sugestões de rotas compatíveis com ele, caso essa tenha sido percorrida por outro usuário da mesma cidade, ou então de faixa etária parecida.

Em paralelo, as informações armazenadas na tabela `tb_rotas` são referen-

tes aos percursos traçados pelos usuários. As rotas são armazenadas na coluna `rot_rota` e o formato utilizado é o *geography*, que é um tipo de dado espacial usado para representar um recurso no sistema de coordenadas da Terra. Ele é derivado do PostGIS (POSTGIS, 2018), que é um extensão para o banco de dados PostgreSQL. Ele adiciona suporte a objetos geográficos, permitindo que as consultas de localização sejam executadas na Linguagem de Consulta Estruturada (*Structured Query Language* - SQL). Além de armazenar a rota em si, também são armazenadas informações derivadas dela a fim de auxiliar na montagem de perfil do usuário. Essas informações são: `rot_distancia` e `rot_tempo`. Elas possibilitam analisar a distância que o ciclista consegue percorrer e quanto tempo poderá levar, ajudando a traçar um perfil para o usuário.

Essa tabela possui não apenas os dados geográficos do arquivo e as informações provenientes dele, mas ela também possui a coluna `rot_codigo`, que identifica as rotas inseridas. Ela recebe também a coluna `cli_codigo`, que é o código do usuário que percorreu a rota e uma chave estrangeira, relacionando a rota com o usuário que a percorreu. Dessa forma, é possível classificar os usuários de acordo com as rotas que eles percorreram, verificando quais usuários se encaixam na classificação de outros usuários, podendo sugerir pessoas compatíveis para serem parceiros de pedal. Podendo utilizar também informações da tabela `tb_cliente` para auxiliar nas sugestões de companheiros aos usuários.

4.4 Camada de Análise dos Dados

Com as informações armazenadas no banco de dados, elas se tornam mais acessíveis e podem ser melhor manipuladas e categorizadas, propiciando a análise desses dados. Nessa etapa, buscou-se, através das informações de rotas dos usuários (ciclistas), chegar a uma recomendação de ciclistas com perfis semelhantes, além de ter sido buscado também uma forma de auxiliar no planejamento urbano.

Para realizar a sugestão de ciclistas, foi criado um sistema de recomendação baseado no método de filtragem cognitiva (baseado em conteúdo), abordado na Seção 2.3. Nesse tipo de filtragem, ocorre a recomendação com base em itens já utilizados pelo usuário. Por exemplo, se o usuário escuta músicas do gênero rock, esse método de filtragem identificará isso e recomendará outras músicas do gênero rock para o usuário.

Os tipos de dados coletados e a forma como eles estão dispostos contribuíram na escolha da filtragem cognitiva para a recomendação, pois dessa forma os usuários são classificados de acordo com as rotas que ele percorreu. Já os outros dois métodos (filtragem colaborativa e híbrida) foram preteridos devido a necessidade

de o usuário classificar algo, sendo um método que funciona melhor quando os dados estão agrupados em gêneros, assim como ocorre em filmes, livros, musicas, etc.

Em conjunto com o método de filtragem cognitiva, foi utilizado o algoritmo *K-means*, explicado na Seção 2.3.1. Ele é um algoritmo de aprendizagem de máquina que funciona de forma não-supervisionada, ou seja, ele consegue gerar resultados com os dados inseridos sem a necessidade de passar por um treinamento. O *K-means* classifica os dados com base em centroides. Por exemplo, dado um conjunto de dados e um número de grupos igual a dois ($k=2$), são definidos dois centroides e os dados vão para o grupo representado pelo centroide que possui a menor distância para o dado comparado. Dessa forma o valor do centroide representa o grupo que ele classifica.

Com base no funcionamento dos centroides, esse trabalho utiliza o algoritmo *K-means*, buscando aproveitar o centroide para representar os usuários. Com isso, cada ciclista passa pelo algoritmo *K-means*, sendo inseridos nesse algoritmo os dados das rotas que o usuário percorreu, tendo como retorno desse algoritmo o centroide que representa todas as rotas percorridas pelo ciclista.

O algoritmo *K-means* foi escolhido com o objetivo de auxiliar na análise das informações, pois resume os diversos valores das rotas percorridas por um usuário em apenas uma informação que é o valor do centroide. Reduzindo a complexidade dos cálculos para verificar os perfis dos ciclistas e também tornando os dados homogêneos, pois um usuário pode ter percorrido duas rotas e outro usuário percorrido cinco, com a utilização do *k-means*. Nesse caso, as duas rotas do primeiro usuário resultam em um centroide e as cinco rotas do segundo usuário se torna outro centroide, dessa forma, ao invés de comparar dois valores de um lado com cinco valores de outro, pode comparar apenas um valor de cada lado que são os valores dos centroides.

Para melhorar a eficiência do código, além da utilização do *K-means*, o SiCliSIG teve uma inspiração na técnica de programação dinâmica. De acordo com o livro *Problems on Algorithms*, escrito por Ian Parberry e William Gasarch, “Programação dinâmica é um nome sofisticado para dividir e conquistar com uma tabela. Em vez de resolver subproblemas recursivamente, resolva-os sequencialmente e armazene suas soluções em uma tabela” (PARBERRY; GASARCH, 2002, p. 87).

Nesse contexto de programação dinâmica, foi utilizada pelo SiCliSIG uma tabela local para armazenar os cálculos entre os usuários, a fim de evitar a repetição de operações, e assim caso precise calcular novamente, só é necessário acessar o valor e não efetuar o recálculo. Um exemplo prático, é quando o algoritmo compara o usuário 1 e o usuário 2 e obtêm a resposta, porém futuramente irá ocorrer a comparação inversa, que é do usuário 2 com o usuário 1, contudo o resultado será o mesmo que da operação inversa, dessa forma, pode-se armazenar o resultado da outra operação

a fim de evitar o recálculo.

Já para auxiliar no planejamento urbano, foi utilizado o conceito de mapas de calor. De acordo com o site ArcGIS, "um mapa de calor (*heat map*) representa a densidade geográfica de elementos de pontos num mapa, utilizando áreas coloridas para representar os pontos. As áreas serão maiores quando existe uma maior concentração de pontos" (ARCGIS, 2018). Com a representação de um mapa de calor é possível verificar as rotas mais percorridas pelos ciclistas e então realizar um planejamento urbano voltado a esses dados.

4.5 Camada de Aplicação

Nessa camada, são utilizadas as informações e processos das camadas anteriores para gerar serviços. Essa camada fica dividida em duas etapas: coleta de dados e análise.

Primeiramente, foi desenvolvida uma aplicação Android. Esse aplicativo é instalado no celular do ciclista e funciona da seguinte forma: antes de um ciclista começar um pedal ou treino, ele deve ativar o aplicativo, que começará a coletar dados em intervalos de tempo e a partir desse conjunto de dados coletados, conseguir ter várias informações como distância do percurso, velocidade, média de altimetria, etc.

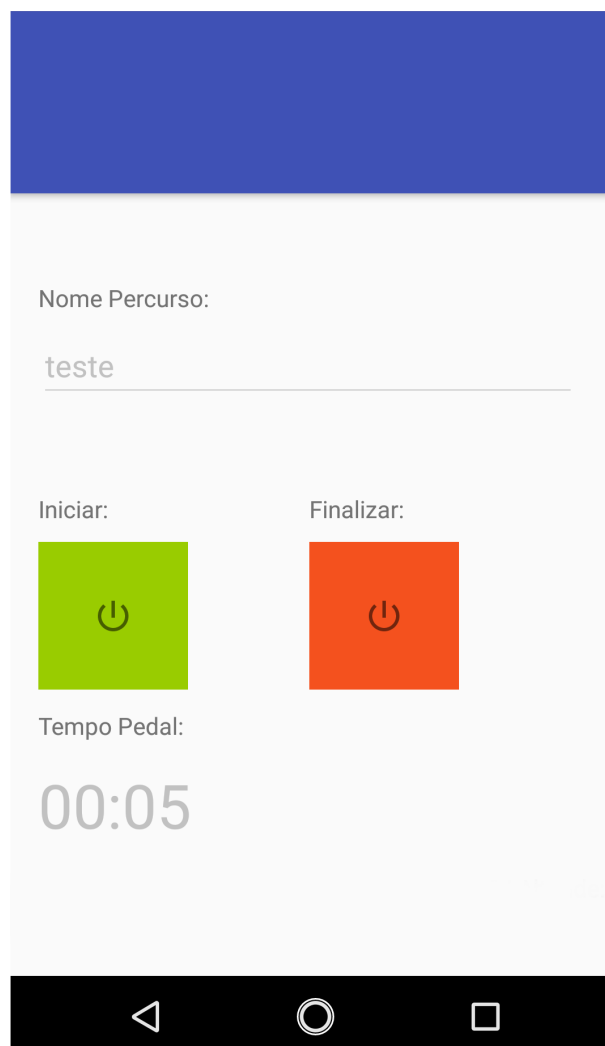
Para o funcionamento do aplicativo mostrado na Figura 10, deve ser inserido o nome do percurso a ser traçado (o nome que o usuário preferir para retratar o percurso) e então selecionar o botão iniciar, com isso o cronômetro será acionado começando a contagem e indicando ao usuário que o aplicativo já está em funcionamento. Feito isso, a aplicação passa a coletar os dados e inserí-los periodicamente em um arquivo. Quando o usuário terminar de traçar o percurso, o botão finalizar da aplicação deve ser pressionado, encerrando as tarefas (coleta de dados) e reiniciando o cronômetro, estando a aplicação apta para um novo percurso.

Na realização da coleta de dados foi utilizada uma Interface de Programação de Aplicações (*Application Programming Interface - API*) de localização para Android que se utiliza do sensor GPS disponível no celular e dos dados móveis ou WiFi a fim de capturar as informações. O uso dessa API foi aliado com o método serviço que segundo o site [AndroidDevelopers \(2018\)](#):

Um *Service* é um componente do aplicativo que pode realizar operações longas e não fornece uma interface do usuário. Outro componente do aplicativo pode iniciar um serviço e ele continuará em execução em segundo plano mesmo que o usuário alterne para outro aplicativo.

Dessa forma, os dados podem ser coletados mesmo com o celular bloqueado ou realizando outra ação, como por exemplo utilizar um aplicativo para ouvir música, ou monitorar a frequência cardíaca.

Figura 10 – Tela principal do aplicativo para coletar dados.



Fonte: Elaborado pelo autor.

O aplicativo proposto foi pensado para identificar a movimentação do ciclista que o utiliza, capturando as informações apenas quando ele estiver em movimento, evitando capturar os dados quando o ciclista estiver parado, buscando reduzir o consumo energético e gastos com dados móveis. Além disso, tem o objetivo de ocupar menos espaço no dispositivo, evitando capturas com informações repetidas, pois quando o ciclista estivesse parado, continuariam a ocorrer capturas das informações do mesmo de forma repetida.

Buscando atender esses requisitos, ao ser inicializado, o aplicativo coleta as informações em um intervalo de 10 segundos. Inicialmente foi fixado um intervalo de 5 segundos, porém após alguns testes foi visto que ao aumentar esse intervalo

para 10 segundos não acarretaria prejuízos para a análise e reduziria o consumo de memória, pois diminui pela metade o número de coletas. Para ocorrer cada coleta deve ter sido percorrida no mínimo uma distância de 20 metros nesse intervalo de tempo, mostrando que o ciclista deve manter no mínimo uma velocidade de 2m/s que equivale a 7,2 km/h.

Essa velocidade média mínima pode parecer um pouco baixa para a categoria de ciclistas que podem atingir velocidades como 30 km/h ou 40 km/h. Porém, essas médias podem variar de acordo com a bicicleta utilizada, com o terreno percorrido e também depende do preparo do ciclista em questão. Dessa forma, se o ciclista optar por fazer uma trilha (*off-road*), em alguns casos ele enfrentará terrenos mais severos, como subidas íngremes ou passagens de areia, onde reduzirá bastante a velocidade e, em alguns casos, impossibilitará o ciclista de transpor esse obstáculo pedalando, o forçando a vencê-lo empurrando a bicicleta.

Pensando nessa hipótese para coletar dados de rotas do tipo trilha, a velocidade média mínima definida foi uma velocidade que poderia ser alcançada por um ciclista empurrando sua bicicleta em terrenos mais severos, não perdendo esses pontos na coleta de dados. Dessa forma, consegue-se coletar os dados mesmo em condições mais adversas de rotas, como acontece nas trilhas, ou em escaladas de serras e montanhas. Além disso, evita a coleta de pontos repetidos que ocorre quando o ciclista está parado.

A partir desses dados coletados na aplicação móvel, é possível conseguir efetuar alguns serviços como: Planejamento Urbano e Sugestão de Ciclistas. Esses serviços se enquadram no contexto de cidades inteligentes e serão explorados nesse trabalho.

4.5.1 Cidades Inteligentes

Ao unir a Tecnologia da Informação e Comunicação (TIC), que é proposto na aplicação móvel, com o planejamento urbano, esses temas passam a estar integrados no contexto de cidades inteligente (*smart cities*). Nas cidades inteligentes tendem a ocorrer a integração entre o mundo físico e os sensores, buscando melhorar as necessidades da população como mostrado na Seção 2.1.

Com as informações coletadas pela aplicação móvel é possível conseguir resultados que afetem algumas áreas das cidades inteligentes, que foram citadas na Seção 2.1. Essas áreas afetadas são governança inteligente e cidadãos inteligentes, o primeiro é ilustrado pelo planejamento urbano e o segundo pela recomendação de ciclistas e serão abordados a seguir.

4.5.1.1 Planejamento Urbano

O planejamento urbano está inserido no conceito de governança inteligente, visando melhorar a qualidade de vida da população. Para essas melhorias ocorrerem, são tomadas decisões e efetuadas modificações que possam ser vantajosas para a comunidade, como por exemplo: criar praças, asfaltar ruas, alterar sentidos de ruas, etc.

Uma vez que os dados coletados pela aplicação móvel são de cunho geográfico, elas podem indicar através de uma análise informações relevantes a respeito de locais da cidade, como por exemplo: os locais de destaque da cidade em que os ciclistas passam, quais as rotas mais percorridas, quais bairros são mais percorridos, etc. Essas informações que podem ser extraídas dos dados coletados conseguem auxiliar no planejamento urbano, podendo gerar como respostas: locais propícios para construção de ciclovias ou instalação de sistemas de bicicletas compartilhadas.

Como foi citado na seção “Análise de Dados”, para a realização do planejamento urbano, é utilizado o conceito mapa de calor (*heat map*), que através dos dados existentes consegue colocar em evidência os dados que possuem maior incidência. No caso dessa aplicação, o mapa de calor funciona com o objetivo de evidenciar as rotas mais trafegadas pelos ciclistas, onde as rotas menos percorridas possuem uma coloração mais clara (transparente) e as mais percorridas possuem uma coloração mais forte (nítida). Com esse tipo de tratamento dos dados, é propiciada uma análise melhor, conseguindo então auxiliar no planejamento de construções voltadas aos ciclistas, como é o caso de ciclovias e pontos de bicicletas compartilhadas.

4.5.1.2 Recomendação de Ciclistas

Além do planejamento urbano, os dados obtidos podem auxiliar na recomendação de ciclistas, que se insere no conceito de cidadãos inteligentes. Como falado no Capítulo 2.1, esse conceito prega que as pessoas podem se utilizar das informações obtidas para melhorar sua interação com outros cidadãos.

Para realizar a recomendação de ciclistas é preciso inicialmente identificar o perfil de cada ciclista, para tanto, as rotas que o ciclista percorreu vão servir de parâmetro para essa identificação. E dessas rotas serão extraídos os dados espaciais e temporais que são, respectivamente a distância e o tempo de cada trajeto.

Após isso, esses dados passam pelo algoritmo *K-means* que retornará um centroide representando as distâncias percorridas pelo ciclista e um outro centroide representando o tempo que o ciclista levou para percorrer as rotas. Com essas informações ocorre o cálculo da diferença entre os perfis, procurando encontrar perfis com

a menor diferença possível, diferença essa que pode variar entre 0% e 100%.

Para realizar o cálculo da diferença, são comparados os valores dos centroides dos usuários, para determinar o maior e o menor desses valores, que são representados na Equação 4.2 por “b” e “a”. Após essa classificação, é realizada a subtração entre eles e então esse valor resultante é dividido pelo maior dos dois centroides e multiplicado por 100, com isso é conseguida a porcentagem de diferença entre os dois usuários, representado pela variável “x”. Se esse valor resultante for menor ou igual a 10% significa que os usuários são compatíveis e podem pedalar juntos. Esse processo é repetido para todas as combinações de usuário.

$$x = \left(\frac{b - a}{b}\right) * 100 \quad (4.2)$$

O valor 10% foi escolhido como parâmetro para determinar a compatibilidade por ser um valor que no pior caso de detecção, caso este que apresentará maior disparidade entre os ciclistas, o usuário menos preparado consiga acompanhar o usuário mais preparado e ao pedalarem juntos o ciclista mais qualificado não se sinta entediado. Proporcionando assim uma pedalada mais prazerosa para ambos usuários.

5 RESULTADOS

Neste capítulo, são mostrados os resultados obtidos através da implementação da proposta, abordada no capítulo anterior. Para chegar a esses resultados, foram utilizados dados de 5 ciclistas da cidade de Aracati, Ceará. Para a coleta de dados, foi desenvolvido um aplicativo móvel que era iniciado no princípio das atividades e encerrado ao concluir a pedalada. O aplicativo foi instalado em aparelhos de 10 ciclistas, no entanto, alguns usuários esqueceram de utilizar o aplicativo e outros tiveram problemas em seu uso. Dessa forma, impossibilitando a coleta de informações de todos os ciclistas, sendo coletado com sucesso os dados de 5 ciclistas, totalizando 58 rotas.

Após a coleta, os dados passaram por uma etapa de extração e transformação (Seção 4.2.2), que consistiu em uma aplicação escrita na linguagem Python que extrai as informações da rota (tempo e distância) e as armazena em um banco de dados no servidor juntamente com a própria rota, facilitando a manipulação desses dados posteriormente.

Para realizar o armazenamento das informações no servidor, foi utilizado o banco de dados PostgreSQL¹ juntamente com a extensão PostGIS² para auxiliar na manipulação dos dados geográficos.

A análise de dados, como falado no capítulo anterior, teve dois objetivos. Primeiro, procurou-se reconhecer ciclistas com perfis semelhantes que poderiam pedalar em conjunto. Segundo, buscou-se auxiliar no planejamento urbano. Para auxiliar no entendimento dos dados, a Figura 11 mostra um mapa com todas as rotas percorridas pelos ciclistas que foram coletadas pela aplicação desenvolvida e utilizadas nesse trabalho.

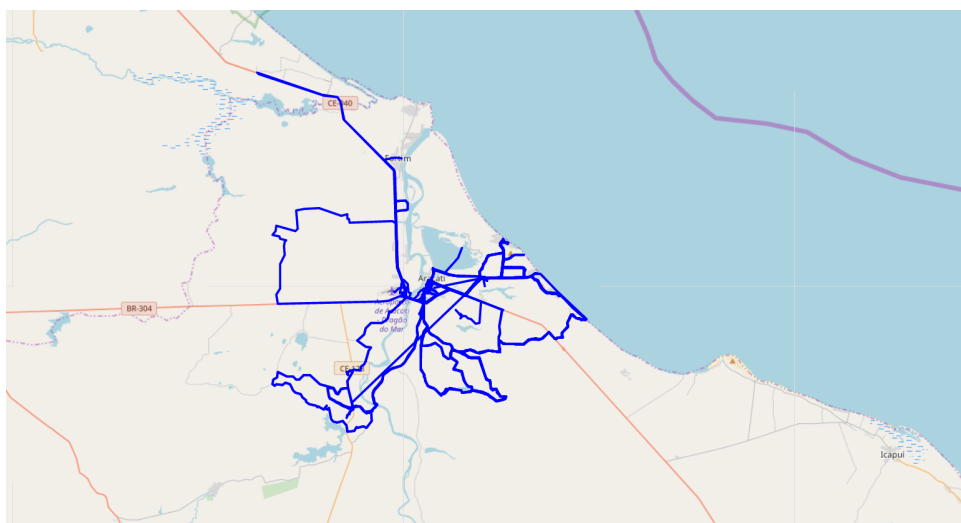
5.1 Planejamento Urbano

Para auxiliar no planejamento urbano, foi utilizado o conceito de mapas de calor, abordado na Seção 4.5.1.1. Como mencionado anteriormente, mapas de calor tem como objetivo evidenciar pontos mais frequentados. Porém, neste trabalho, o objetivo não consiste em evidenciar apenas um ponto mais acessado, mas sim destacar rotas mais frequentadas, rotas essas que consistem em conjuntos de pontos.

O padrão do mapa de calor é mostrar pontos maiores e com destaque para uma cor quando são estes muito frequentados, ou seja, quando existem muitos pontos

¹ <https://www.postgresql.org/>

² <https://postgis.net/>

Figura 11 – Mapa com todas as rotas coletadas

Fonte: Elaborado pelo autor

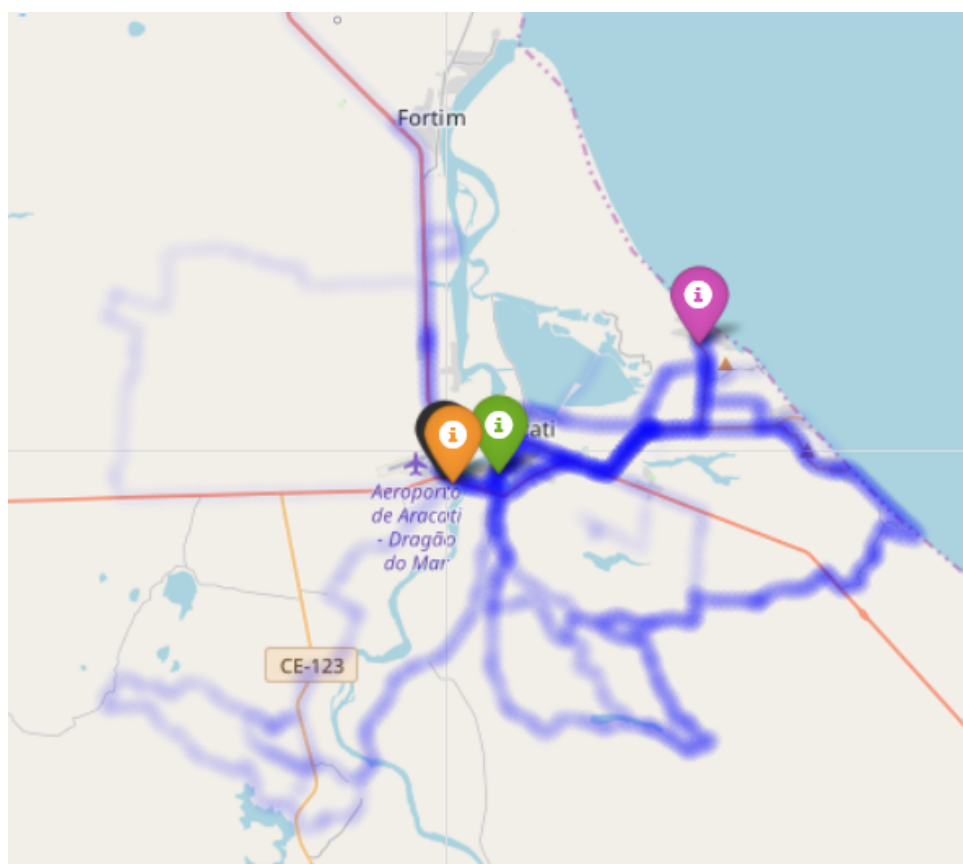
próximos a um mesmo local. Já quando um local possui poucos pontos à sua volta, a coloração tende a uma outra cor. No Folium, biblioteca do Python utilizada, quanto mais frequentado o ponto, ele tende a vermelho e quando menos frequentado ele tende a azul ou verde.

O mapa de calor gerado nessa aplicação retrata as rotas sempre na coloração azul, porém, para se obter uma melhor diferenciação visual entre as rotas mais e menos percorridas, foi alterada a transparência das rotas. Dessa forma, quanto mais transparente um trecho, significa que o mesmo foi percorrido menos vezes e quanto mais nítido, mais transitado ele foi. A transparência aplicada na geração do mapa de calor foi 0.5, baseado na quantidade de rotas analisada.

Além da transparência dos pontos, foram alterados também seus tamanhos, pois no mapa de calor padrão da biblioteca Folium os pontos possuem tamanhos maiores a fim de dar ênfase às suas posições. Porém, para a visualização das rotas utilizadas neste trabalho, o uso dos valores padrão de tamanho tornava a interpretação do mapa mais difícil por impedir uma visão detalhada dos trechos percorridos. Além disso, ao ser traçada, uma rota poderia englobar uma rua que não tinha sido percorrida devido ao tamanho dos pontos que, dependendo da concentração, poderiam tomar mais de uma rua. Dessa forma, foi reduzido o tamanho dos pontos, para que o mapa de calor pudesse representar mais fielmente os trechos percorridos mesmo com uma grande aglomeração de pontos.

A Figura 12 mostra o mapa de calor obtido através das rotas coletadas pelos ciclistas. Analisando a Figura, podemos perceber que existem locais que possuem tráfego maior que outros. Entre esses locais, existem dois pontos de destaque. O primeiro é a rota da entre a sede de Aracati (ponto verde) e os bairros Pedregal (ponto

Figura 12 – Mapa de calor com as alterações feitas nos parâmetros e baseado em todas as rotas percorridas pelos usuários

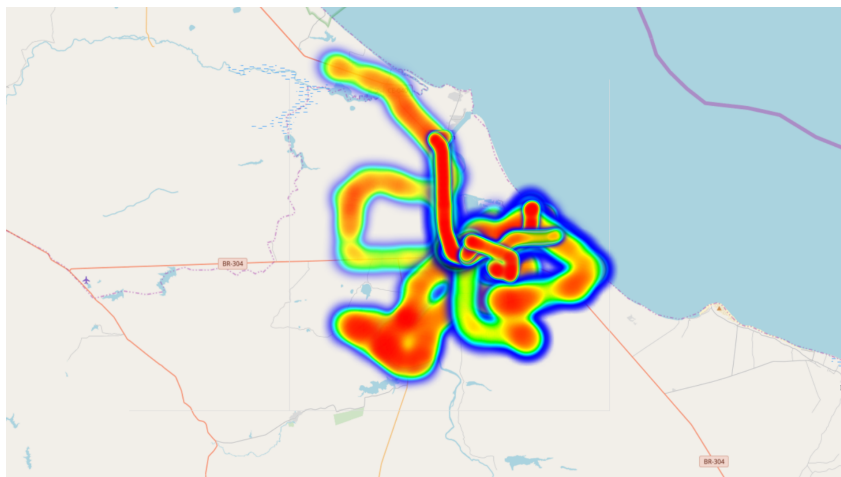


Fonte: Elaborado pelo autor

preto) e Vila Grega (ponto laranja). Uma possível interpretação para a popularidade desses trechos pode ser explicada por uma grande afluência de ciclistas que se locomovem entre suas casas e seus locais de trabalhos, o que é característico dessa região. Dessa forma, pode-se concluir que pontos dessa rota podem ser propícios para a instalação de estações de bicicletas compartilhadas. Outro percurso bastante percorrido é o trecho entre a cidade de Aracati (Ponto verde) e a praia de Canoa Quebrada (ponto lilás), onde pode ser um local oportuno para a instalação de uma ciclovia, já que, em muitas partes desse trajeto, não existe acostamento.

Caso não tivessem sido alterados os parâmetros de transparência e tamanho dos pontos, já mencionados neste capítulo, o mapa de calor obtido com os mesmos dados usados na Figura 12, seria o representado pela Figura 13, onde fica difícil distinguir quais as rotas existentes e quais são as mais percorridas.

Figura 13 – Mapa de calor padrão, sem as alterações feitas nos parâmetros, e baseado em todas as rotas percorridas pelos usuários



Fonte: Elaborado pelo autor

5.2 Recomendação de Ciclistas

As informações capturadas pelos ciclistas podem ser utilizadas também para a detecção dos perfis dos mesmos. A análise realizada através dos dados pode ser espacial ou temporal. A análise espacial leva em consideração a distância das rotas percorridas pelos ciclistas e a análise temporal se baseia no tempo que foi demandado para concluir as rotas.

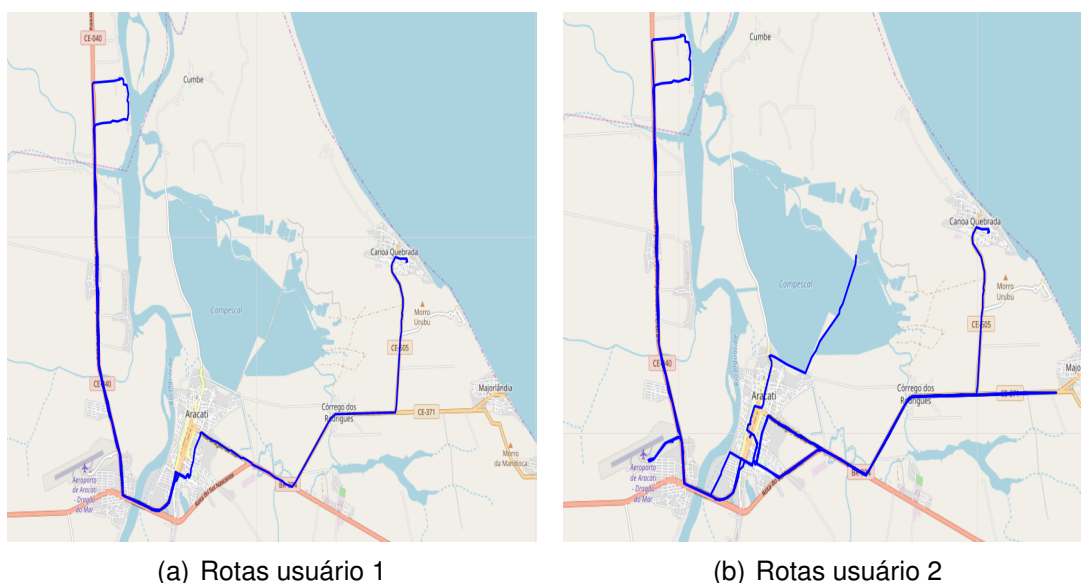
Unindo os dois conceitos anteriores surge a análise espaço-temporal, que se baseia em quanto tempo um objeto levou para fazer determinado trajeto, ou seja, ela se sustenta na velocidade desenvolvida ao longo do trajeto. Porém o cenário correto para se basear na velocidade é quando o tempo utilizado corresponde apenas ao tempo de locomoção do objeto, não levando em consideração o tempo em que o objeto ficou parado. Esse cenário é diferente do armazenado no banco de dados que computa também o tempo em que o usuário ficou parado, dessa forma ao medir a velocidade ela pode se tornar inferior ao que realmente foi devido o tempo que o usuário ficou parado, provocando mudanças na análise.

Na análise espacial, as distâncias das rotas percorridas pelos usuários foram inseridas no algoritmo *K-means*, que, por sua vez, retornou o centroide correspondente a essas distâncias. Então foram comparados os centroides dos usuários a fim de encontrar uma diferença menor que 10% entre eles. Após essa análise espacial, a qual todos os usuários foram comparados entre si, foi detectado um relacionamento entre o usuário 1 e o usuário 2, que possuem suas rotas representadas na Figura 14. O valor do centroide do usuário 1 é 18.28, enquanto que o valor do centroide correspondente ao usuário 2 é 17.02. Dessa forma, após a aplicação do cálculo, foi

mostrada uma diferença menor que 10% entre eles, caracterizando então uma similaridade significativa.

Analisando a Figura 14, é possível observar uma semelhança nas trajetórias percorridas pelos dois usuários. Dessa forma, evidenciando que a compatibilidade entre os dois ciclistas, detectada pelo algoritmo, é existente.

Figura 14 – Mapas com todas as rotas percorridas pelo usuário 1 e pelo usuário 2



Na análise temporal realizada, os dados inseridos no algoritmo *K-means* foram referentes ao tempo de duração das rotas em minutos. O resultado retornado pelo algoritmo consistiu no centroide correspondente a esse tempo. Com isso, os centroides foram comparados, com o objetivo de encontrar uma diferença menor do que 10%. Após a análise temporal, foram encontradas similaridades entre o usuário 2 e o usuário 7 e entre o usuário 1 e o usuário 5, que são exemplificados pela Tabela 1.

Tabela 1 – Tabela da análise temporal

A - B	Centroide usuário A	Centroide usuário B	Porcentagem
2 - 7	88.313	93.033	5.072%
1 - 5	74.4	67.016	9.923%

Com a realização da análise dos dados, foram identificados três relacionamentos entre ciclistas, sendo dois desses relacionamentos através da análise temporal e um através da análise espacial. Mediante esses relacionamentos, é possível realizar a recomendação dos ciclistas compatíveis, incitando o pedal em grupo, que acarreta uma maior segurança para os ciclistas e os estimulam a perseverarem na atividade, pois é mais prazeroso pedalar em grupos que sozinho.

6 CONCLUSÃO

Neste trabalho foram abordadas formas de utilizar dados geográficos provenientes de rotas de ciclistas. Essas rotas foram obtidas através de uma aplicação móvel instalada nos *smartphones* dos usuários. Após isso, as rotas dos usuários e dados provenientes delas foram armazenados em um banco de dados em uma máquina servidora com o intuito de serem manuseados pelo SiCliSIG. Posteriormente, esses dados foram manipulados e geraram informações úteis para ciclistas e autoridades públicas. Com os dados, foi possível verificar o perfil dos ciclistas e verificar se eles são compatíveis a fim de fazer uma recomendação de parceiros de pedalada. Foi possível também elaborar um mapa de calor, que pode auxiliar na tomada de decisão das autoridades referente a obras de infraestrutura, pois dessa forma, facilita a detecção de áreas mais trafegadas por ciclistas, auxiliando na implantação de ciclovias ou sistemas de bicicletas compartilhadas em locais adequados.

Com as informações abordadas neste trabalho, com relação ao referencial teórico, pode-se observar que o contexto de cidades inteligentes é algo que está em evidência atualmente, sendo útil para resolver diversas problemáticas e que pode ser utilizado em conjunto com outras tecnologias e contextos. Dessa forma, esse trabalho exemplifica essa conjugação de tecnologias, onde o contexto de cidades inteligentes foi abordado unido à sistemas de informações geográficas e sistemas de recomendação. Sendo um reflexo de diversos trabalhos e aplicações que unem diversas tecnologias a fim de melhorar a qualidade de vida e a mobilidade urbana, oferecendo formas acessíveis para o usuário.

Para o experimento, foram coletados dados de 5 ciclistas da cidade de Aracati, Ceará. A fim de realizar a análise dessas informações, foi utilizado o algoritmo *K-means*, juntamente com uma equação para descobrir o percentual de diferença entre os ciclistas. Foi realizada uma análise temporal que detectou duas similaridades entre os ciclistas, e ocorreu também uma análise espacial a qual detectou uma similaridade.

Como trabalhos futuros, um dos desafios é a detecção dos perfis das rotas coletadas. Isso pode ensejar a aplicação de outras técnicas de análise de dados resultando em outras aplicações como, por exemplo, a sugestão de rotas, ou então o aprimoramento da sugestão de ciclistas. Além disso, todas as etapas podem ser integradas originando uma aplicação móvel que consiga lidar com todos esses passos. A análise de momentos de parada constitui-se como um importante trabalho futuro por permitir identificar diferentes segmentos de rotas individuais e possibilitar a realização de uma análise espaço-temporal sem prejuízos. Outra abordagem seria rodar a proposta com mais usuários, a fim de conseguir um maior número de relacionamentos.

REFERÊNCIAS

- AFANEH, A.; SHAHROUR, I. Use of gis for sunrise smart city project, large scale demonstrator of the smart city. In: *2017 Sensors Networks Smart and Emerging Technologies (SENSET)*. [S.l.: s.n.], 2017. p. 1–4. Citado 3 vezes nas páginas 21, 25 e 26.
- ALVES., R. F. *Latitudes e Longitudes*. 2018. Disponível em: <<https://brasilecola.uol.com.br/geografia/latitudes-longitudes.htm>>. Acesso em: 16.06.2018. Citado na página 42.
- ANDROIDDEVELOPERS. *Serviços*. 2018. Disponível em: <<https://developer.android.com/guide/components/services.html?hl=pt-br>>. Acesso em: 14.06.2018. Citado na página 48.
- ARCGIS. *Adicionar um Mapa de Calor (Heat Map)*. 2018. Disponível em: <<https://doc.arcgis.com/pt-pt/maps-for-office/3.1/add-a-heat-map.htm>>. Acesso em: 06.10.2018. Citado na página 48.
- BETZ, J. W. Navstar global positioning system. In: _____. *Engineering Satellite-Based Navigation and Timing: Global Navigation Satellite Systems, Signals, and Receivers*. Wiley-IEEE Press, 2016. p. 672–. ISBN 9781119141167. Disponível em: <<https://ieeexplore.ieee.org/xpl/articleDetails.jsp?arnumber=7394726>>. Citado 2 vezes nas páginas 14 e 30.
- BIKERADAR. *30 great benefits of cycling*. 2018. Disponível em: <<https://www.bikeradar.com/beginners/gear/article/30-reasons-to-take-up-cycling-23965/>>. Acesso em: 13.08.2018. Citado na página 17.
- BRAGA, N. C. *Conheça a tecnologia LoRa® e o protocolo LoRaWAN™ (LOR001)*. 2018. Disponível em: <<http://newtoncbraga.com.br/index.php/eletronica/52-artigos-diversos/11992-conheca-a-tecnologia-lora-e-o-protocolo-lorawan-lor001>>. Acesso em: 08.08.2018. Citado na página 23.
- CAMARA, G.; ORTIZ, M. J. Sistemas de informação geográfica para aplicações ambientais e cadastrais: uma visão geral. *CONGRESSO BRASILEIRO DE ENGENHARIA AGRÍCOLA*, v. 27, p. 59–82, 1998. Citado 4 vezes nas páginas 9, 25, 26 e 27.
- CAMPOS, D. M. R. Biblioteca de criação de rotas entre coordenada em ambientes marítimos no sistema bio data. 2017. Citado na página 44.
- CHEN, L.; JAKUBOWICZ, J. Inferring bike trip patterns from bike sharing system open data. In: *2015 IEEE International Conference on Big Data (Big Data)*. [S.l.: s.n.], 2015. p. 2898–2900. Citado na página 37.
- CIENCIAEDADOS. *Conceitos Fundamentais de Machine Learning*. 2017. Disponível em: <<http://www.cienciaedados.com/conceitos-fundamentais-de-machine-learning>>. Acesso em: 23.07.2018. Citado na página 34.

- COOPER, K. D.; TORCZON, L. Capítulo 2 - scanners. In: COOPER, K. D.; TORCZON, L. (Ed.). *Construindo Compiladores*. Rio de Janeiro: Elsevier Editora Ltda., 2014. p. 19 – 67. ISBN 978-85-352-5564-5. Disponível em: <<http://www.sciencedirect.com/science/article/pii/B978853525564500024>>. Citado na página 44.
- COOPER, K. D.; TORCZON, L. Capítulo 3 - analisadores sintáticos (parsers). In: COOPER, K. D.; TORCZON, L. (Ed.). *Construindo Compiladores (Segunda edição)*. Segunda edição. Rio de Janeiro: Elsevier Editora Ltda., 2014. p. 69 – 140. ISBN 978-85-352-5564-5. Disponível em: <<https://www.sciencedirect.com/science/article/pii/B978853525564500036>>. Citado na página 44.
- DAMEROW, F.; KLINGELSCMITT, S.; EGGERT, J. Spatio-temporal trajectory similarity and its application to predicting lack of interaction in traffic situations. In: IEEE. *Intelligent Transportation Systems (ITSC), 2016 IEEE 19th International Conference on*. [S.l.], 2016. p. 2512–2519. Citado na página 36.
- DATASCIENCEACADEMY. *Curso, Introducao à Ciências de Dados 2.0*. 2018. Disponível em: <<https://www.datascienceacademy.com.br/path-player?courseid=introducao-de-dadosunit=5a93333b5e4cdedcff8b4571Unit>>. Acesso em: 13.08.2018. Citado na página 14.
- DIARIODONORDESTE. *Prática de atividade física vem crescendo entre a população*. 2017. Disponível em: <<http://diariodonordeste.verdesmares.com.br/cadernos/cidade/pratica-de-atividade-fisica-vem-crescendo-entre-a-populacao-1.1746220>>. Acesso em: 01.08.2018. Citado na página 15.
- DLODLO, N.; GCABA, O.; SMITH, A. Internet of things technologies in smart cities. In: *2016 IST-Africa Week Conference*. [S.l.: s.n.], 2016. p. 1–7. Citado na página 23.
- DODGE, S.; WEIBEL, R.; FOROOTAN, E. Revealing the physics of movement: Comparing the similarity of movement characteristics of different types of moving objects. *Computers, Environment and Urban Systems*, v. 33, n. 6, p. 419 – 434, 2009. ISSN 0198-9715. Spatial Data Mining–Methods and Applications. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0198971509000556>>. Citado na página 36.
- EPOCANEGOCIOS. *Uma bike é roubada a cada meia hora*. 2017. Disponível em: <<https://epocanegocios.globo.com/Brasil/noticia/2017/02/epoca-negocios-uma-bike-e-roubada-a-cada-meia-hora.html>>. Acesso em: 03.08.2018. Citado na página 18.
- ESTADAO. *Em São Paulo, número de ciclistas cresce 50% em 1 ano*. 2014. Disponível em: <<https://sao-paulo.estadao.com.br/noticias/geral,em-sao-paulo-numero-de-ciclistas-cresce-50-em-1-ano,1562460>>. Acesso em: 01.08.2018. Citado na página 17.
- EVANS, M. R. et al. Fast and exact network trajectory similarity computation: A case-study on bicycle corridor planning. In: *Proceedings of the 2Nd ACM SIGKDD International Workshop on Urban Computing*. New York, NY, USA: ACM, 2013. (UrbComp '13), p. 9:1–9:8. ISBN 978-1-4503-2331-4. Disponível em: <<http://doi.acm.org/10.1145/2505821.2505835>>. Citado na página 37.

- EXAME. *Força no pedal! Como as bikes estão mudando 10 capitais*. 2016. Disponível em: <<https://exame.abril.com.br/brasil/os-avancos-e-tropecos-das-bicicletas-em-10-cidades-do-brasil/>>. Acesso em: 14.08.2018. Citado na página 17.
- FAGUNDES, E. *Sigfox uma plataforma para IoT*. 2018. Disponível em: <<http://efagundes.com/blog/sigfox-uma-plataforma-para-iot/>>. Acesso em: 08.08.2018. Citado na página 23.
- FERREIRA, N. C. Apostila de sistema de informações geográficas. 2006. Citado 2 vezes nas páginas 25 e 26.
- GAO, H. et al. Exploring temporal effects for location recommendation on location-based social networks. In: ACM. *Proceedings of the 7th ACM conference on Recommender systems*. [S.l.], 2013. p. 93–100. Citado na página 36.
- GATZIOURA, A.; SÀNCHEZ-MARRÈ, M. A case-based recommendation approach for market basket data. *IEEE Intelligent Systems*, v. 30, n. 1, p. 20–27, Jan 2015. ISSN 1541-1672. Citado na página 31.
- Governo Digital. *Interoperabilidade*. 2018. Disponível em: <<https://www.governodigital.gov.br/transformacao/orientacoes/interoperabilidade>>. Acesso em: 23.06.2018. Citado na página 41.
- HERRMANN, A. et al. A new take on protecting cyclists in smart cities. *IEEE Transactions on Intelligent Transportation Systems*, p. 1–8, 2018. ISSN 1524-9050. Citado na página 38.
- HRNČIŘ, J. et al. Practical multicriteria urban bicycle routing. *IEEE Transactions on Intelligent Transportation Systems*, v. 18, n. 3, p. 493–504, March 2017. ISSN 1524-9050. Citado na página 38.
- INFOESCOLA. *Sistema de Posicionamento Global*. 2018. Disponível em: <<https://www.infoescola.com/cartografia/gps-sistema-de-posicionamento-global/>>. Acesso em: 16.08.2018. Citado 2 vezes nas páginas 27 e 28.
- K.ASHTON. That 'internet of things' thing. In: . [S.l.: s.n.], 2009. Citado na página 22.
- KEARNEY, A. *Global sports market - total revenue from 2005 to 2017 (in billion U.S. dollars)*. 2018. Disponível em: <<https://www.statista.com/statistics/370560/worldwide-sports-market-revenue/>>. Acesso em: 05.04.2018. Citado 2 vezes nas páginas 15 e 16.
- KURMASHOV, N.; LATUTA, K.; NUSSIPBEKOV, A. Online book recommendation system. In: *2015 Twelve International Conference on Electronics Computer and Computation (ICECCO)*. [S.l.: s.n.], 2015. p. 1–4. Citado na página 30.
- LANGMANN, R. Didactical experiments with smartphone sensors in the internet of things. In: *2018 IEEE Global Engineering Education Conference (EDUCON)*. [S.l.: s.n.], 2018. p. 918–922. Citado na página 41.
- LoRa Alliance. *About LoRa Alliance*. 2018. Disponível em: <<https://lora-alliance.org/about-lora-alliance>>. Acesso em: 08.08.2018. Citado na página 23.

- LoRa Alliance. *About LoRaWAN Specification?* 2018. Disponível em: <<https://loro-alliance.org/about-lorawan>>. Acesso em: 08.08.2018. Citado na página 23.
- MALIK, F.; SHAH, M. A. Smart city: A roadmap towards implementation. In: *2017 23rd International Conference on Automation and Computing (ICAC)*. [S.l.: s.n.], 2017. p. 1–6. Citado na página 21.
- MANCINI, M. Internet das coisas: História, conceitos, aplicações e desafios. In: . [S.l.: s.n.], 2017. Citado na página 22.
- MANETH, S.; POULOVASSILIS, A. Data science. *The Computer Journal*, v. 60, n. 3, p. 285–286, March 2017. ISSN 0010-4620. Citado na página 14.
- MARDACANY, E. Smart cities characteristics: importance of built environments components. In: *IET Conference on Future Intelligent Cities*. [S.l.: s.n.], 2014. p. 1–6. Citado na página 20.
- MELVILLE, P.; SINDHWANI, V. Recommender systems. In: _____. *Encyclopedia of Machine Learning*. Boston, MA: Springer US, 2010. p. 829–838. ISBN 978-0-387-30164-8. Disponível em: <https://doi.org/10.1007/978-0-387-30164-8_705>. Citado 2 vezes nas páginas 31 e 32.
- MOHANTY, S. P.; CHOPPALI, U.; KOUGIANOS, E. Everything you wanted to know about smart cities: The internet of things is the backbone. *IEEE Consumer Electronics Magazine*, v. 5, n. 3, p. 60–70, July 2016. ISSN 2162-2248. Citado na página 20.
- MOUNTAIN, D.; RAPER, J. Modelling human spatio-temporal behaviour: a challenge for location-based services. In: *Proceedings of 6th International Conference on Geocomputation*. [S.l.: s.n.], 2001. Citado na página 37.
- MUNDOBOAFORMA. *10 Benefícios de Andar de Bicicleta Para Boa Forma e Saúde*. 2015. Acessado: 10.08.2018. Disponível em: <<https://www.mundoboaforma.com.br/10-beneficios-de-andar-de-bicicleta-para-boa-forma-e-saude/>>. Citado na página 17.
- NOGUEIRA, T. P. et al. Framestep: A framework for annotating semantic trajectories based on episodes. *Expert Systems with Applications*, v. 92, p. 533 – 545, 2018. ISSN 0957-4174. Disponível em: <<http://www.sciencedirect.com/science/article/pii/S0957417417306796>>. Citado 2 vezes nas páginas 36 e 37.
- OFICINADANET. *Sistema de Posicionamento Global*. 2016. Disponível em: <<https://www.oficinadanet.com.br/post/12406-como-funciona-o-gps>>. Acesso em: 16.08.2018. Citado 2 vezes nas páginas 28 e 29.
- OFICINADANET. *Você sabe o que é o GLONASS?* 2016. Disponível em: <<https://www.oficinadanet.com.br/post/10569-voce-sabe-o-que-e-o-qlonass>>. Acesso em: 19.08.2018. Citado na página 28.
- PARBERRY, I.; GASARCH, W. Problems on algorithms. 2002. Citado na página 47.
- POSTGIS. *About PostGIS*. 2018. Disponível em: <<https://postgis.net/>>. Acesso em: 03.04.2018. Citado na página 46.

- QI, J. et al. K*-means: An effective and efficient k-means clustering algorithm. In: *2016 IEEE International Conferences on Big Data and Cloud Computing (BDCloud), Social Computing and Networking (SocialCom), Sustainable Computing and Communications (SustainCom) (BDCloud-SocialCom-SustainCom)*. [S.l.: s.n.], 2016. p. 242–249. Citado na página 34.
- SAS. *Machine Learn*. 2018. Disponível em: <https://www.sas.com/pt_br/insights-/analytics/machine-learning.html>. Acesso em: 30.09.2018. Citado na página 33.
- SHAH, K. et al. Recommender systems: An overview of different approaches to recommendations. In: *2017 International Conference on Innovations in Information, Embedded and Communication Systems (ICIIECS)*. [S.l.: s.n.], 2017. p. 1–4. Citado 2 vezes nas páginas 31 e 32.
- SIGFOX. *Sigfox Services*. 2018. Disponível em: <<https://www.sigfox.com/en-/solutions/sigfox-services>>. Acesso em: 08.08.2018. Citado na página 23.
- SINGH, S.; SINGH, N. Internet of things (iot): Security challenges, business opportunities amp; reference architecture for e-commerce. In: *2015 International Conference on Green Computing and Internet of Things (ICGCloT)*. [S.l.: s.n.], 2015. p. 1577–1581. Citado na página 22.
- TELIT. *The Business Advantages of a Multi GNSS Set-Up*. 2018. Disponível em: <<https://www.telit.com/blog/multi-gnss-business-advantages/>>. Acesso em: 19.08.2018. Citado na página 28.
- TODAMATERIA. *Altitude*. 2018. Disponível em: <<https://www.todamateria.com.br/altitude>>. Acesso em: 16.06.2018. Citado na página 42.
- TORRES, S. et al. Becity: sensing and sensibility on urban cycling for smarter cities. In: *2015 34th International Conference of the Chilean Computer Science Society (SCCC)*. [S.l.: s.n.], 2015. p. 1–4. Citado na página 38.
- TYAGI, V.; KUMAR, A. Internet of things and social networks: A survey. In: *2017 International Conference on Computing, Communication and Automation (ICCCA)*. [S.l.: s.n.], 2017. p. 1268–1270. Citado 3 vezes nas páginas 9, 24 e 25.
- UNIVERSIDADEDEOPORTO. *Tipos de Data Mining*. 2018. Disponível em: <<https://paginas.fe.up.pt/~mgi99021/it/tipos.htm>>. Acesso em: 23.07.2018. Citado na página 34.
- UNIXTIMESTAMP. *Timestamp*. 2018. Disponível em: <<https://www.unixtimestamp.com/>>. Acesso em: 16.06.2018. Citado na página 42.
- VADEBIKE. *12 bons motivos para ir de bicicleta*. 2016. Disponível em: <<http://vadebike.org/2006/06/por-que-ir-de-bicicleta/>>. Acesso em: 10.08.2018. Citado na página 17.
- VEJA. *Poluição mata 50 mil pessoas no Brasil a cada ano, alerta OMS*. 2018. Disponível em: <<https://veja.abril.com.br/saude/poluicao-mata-50-mil-pessoas-no-brasil-a-cada-ano-alerta-oms/>>. Acesso em: 17.10.2018. Citado na página 17.

- WANG, I.; WANG, C. Analyzing bike repositioning strategies based on simulations for public bike sharing systems: Simulating bike repositioning strategies for bike sharing systems. In: *2013 Second IIAI International Conference on Advanced Applied Informatics*. [S.l.: s.n.], 2013. p. 306–311. Citado na página 37.
- WOODS, R. A. *Sports and Exercise*. 2017. Disponível em: <<https://www.bls.gov/spotlight/2017/sports-and-exercise/home.htm>>. Acesso em: 05.04.2018. Citado na página 15.
- YANG, L.; DENG, M. Based on k-means and fuzzy k-means algorithm classification of precipitation. In: *2010 International Symposium on Computational Intelligence and Design*. [S.l.: s.n.], 2010. v. 1, p. 218–221. Citado na página 34.
- YANG, T.; LIN, J.; CHANG, Y. Strategic design of public bicycle sharing systems incorporating with bicycle stocks considerations. In: *The 40th International Conference on Computers Industrial Engineering*. [S.l.: s.n.], 2010. p. 1–6. Citado na página 37.
- ZHENG, Y.; XIE, X.; MA, W.-Y. Geolife: A collaborative social networking service among user, location and trajectory. *IEEE Data Eng. Bull.*, Citeseer, v. 33, n. 2, p. 32–39, 2010. Citado na página 36.