

**LEONARDO JÚNIO ALVES DOS SANTOS**

**EXTRAINDO PERFIS DE USUÁRIOS MÓVEIS COM BASE NA LISTA DE  
APLICATIVOS E NA CIDADE**

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

Orientador: Fabrício Aguiar Silva

**VIÇOSA - MINAS GERAIS  
2021**

**Ficha catalográfica elaborada pela Biblioteca Central da Universidade  
Federal de Viçosa - Campus Viçosa**

T

Santos, Leonardo Júnio Alves dos, 1996-  
S237e            Extraíndo perfis de usuários móveis com base na lista de  
2021            aplicativos e na cidade / Leonardo Júnio Alves dos Santos. –  
Viçosa, MG, 2021.  
105 f. : il. (algumas color.) ; 29 cm.

Inclui anexo.

Orientador: Fabrício Aguiar Silva.

Dissertação (mestrado) - Universidade Federal de Viçosa.

Referências bibliográficas: f. 86-89.

1. Mineração de dados (Computação). 2. Computação móvel. 3. Aplicativos móveis. 4. Sistemas de comunicação móvel. I. Universidade Federal de Viçosa. Departamento de Informática. Programa de Pós-Graduação em Ciência da Computação. II. Título.

CDD 22. ed. 005.74

Bibliotecário(a) responsável: Renata de Fatima Alves CRB6/2578

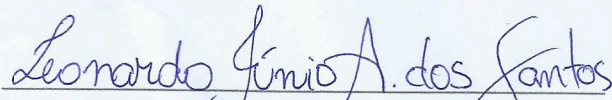
LEONARDO JÚNIO ALVES DOS SANTOS

EXTRAINDO PERFIS DE USUÁRIOS MÓVEIS COM BASE NA LISTA DE  
APLICATIVOS E NA CIDADE

Dissertação apresentada à Universidade Federal de Viçosa, como parte das exigências do Programa de Pós-Graduação em Ciência da Computação, para obtenção do título de *Magister Scientiae*.

APROVADA: 09 de abril de 2021.

Assentimento:



Leonardo Júnio Alves dos Santos  
Autor



Fabrício Aguiar Silva  
Orientador

*Dedico este trabalho ao meu pai, irmãos, avós e minha noiva Raissa.*

## AGRADECIMENTOS

Agradeço primeiramente a Deus por colocar as pessoas certas em meu caminho.

Agradeço meu pai, João, e meus irmãos, Amanda e Filipi, por me incentivarem desde o início dos estudos. A minhas avós, Alice e Rita, e tia Iris por acreditarem em mim.

Agradeço minha noiva/esposa Raissa pela paciência e toda ajuda durante esta etapa. Aos meus sogros, Rose e Rogério, que também nos apoiaram nesta fase da vida.

Agradeço ao meu professor e orientador Fabrício Aguiar Silva por todos os conhecimentos, profissionais e pessoais, que foram fundamentais durante o Mestrado e serão necessários para a vida.

Agradeço ao NESPeD (Núcleo de Estudos em Sistemas Pervasivos e Distribuídos) por toda parceria e contribuições imprescindíveis para o desenvolvimento deste trabalho. Em especial ao então aluno de graduação Germano Barcelos dos Santos por toda ajuda.

Agradeço à Universidade Federal de Viçosa por todas as oportunidades. Especialmente, o *Campus Florestal* ao qual tive o prazer de estudar desde o ensino técnico, durante a graduação e boa parte do mestrado.

Agradeço à Coordenação de Aperfeiçoamento de Pessoal de Nível Superior (CAPES) por todo apoio financeiro.

Por fim, agradeço a todos que tiveram alguma participação, mesmo que com palavras de apoio, durante esta etapa da minha vida.

*Sem sacrifício não há vitória.*

*Autor Desconhecido*

## RESUMO

SANTOS, Leonardo Júnio Alves dos, M.Sc., Universidade Federal de Viçosa, abril de 2021. **Extraíndo perfis de usuários móveis com base na lista de aplicativos e na cidade.** Orientador: Fabrício Aguiar Silva.

Nos últimos anos tem-se percebido um crescimento considerável na coleta e disponibilidade dos dados de usuários móveis. Esse tipo de dado é importante para ajudar nas tomadas de decisões em diferentes cenários, como cidades inteligentes, redes sociais, comércio eletrônico e físico, além de permitir que o comportamento de usuários móveis seja conhecido e que seus interesses sejam utilizados em seu benefício. Porém, gerar um perfil capaz de contribuir com informações relevantes utilizando somente dados brutos não é uma tarefa trivial. Além disso, os tipos de coletas (i.e., ativas ou passivas.) e de dados (e.g., *log* de ligações e localizações) utilizados impactam diretamente na generalização dos modelos preditivos e na privacidade do usuário. Outro ponto relevante é que os trabalhos encontrados na literatura não consideram a cidade do usuário. Dessa forma, usuários que residem em cidades completamente diferentes, uma metrópole e uma cidade do interior, por exemplo, não possuem distinção. Para mitigar estes problemas, este trabalho propõe uma investigação do perfil de usuários de dispositivos móveis utilizando somente a lista de aplicativos e cidade como entrada para os modelos. Para isso, são analisados os perfis dos usuários em cinco perspectivas: preço do dispositivo, tipos de estabelecimentos visitados, funcionalidades das áreas frequentadas, mobilidade e seus traços de personalidade. Afim de obter um perfil mais completo, é proposto o Perfil MAPED que engloba todos os perfis analisados. Para construir e avaliar todos estes perfis, uma base de dados real contendo informações de milhares de usuários móveis foi enriquecida com dados de diferentes fontes. Através de tal enriquecimento, foi possível criar modelos para segmentar usuários com base em seus aplicativos instalados e tipo de cidade em que residem. Ademais, para agilizar a etapa de enriquecimento semântico, foi desenvolvido um *framework* facilmente extensível e que pode ser utilizado para diversos problemas.

Palavras-chave: Mineração de Dados. Perfil de Usuários. Dados Móveis.

# ABSTRACT

SANTOS, Leonardo Júnio Alves dos, M.Sc., Universidade Federal de Viçosa, April, 2021. **Extracting profile of mobile users based on their list of applications and city.** Advisor: Fabrício Aguiar Silva.

In recent years we are facing a significant increase in the collection and availability of mobile user's data. This data type is relevant to help decision-making in different scenarios, such as smart cities, social networks, electronic and physical commerce. Besides that, this data allows the knowledge of mobile users' behavior, benefiting themselves by using their interests. However, generating a profile capable of contributing relevant information using only raw data is not a trivial task. Also, the types of collections (i.e., active or passive) and data (e.g., call log and locations) used directly impact the generalization of predictive models and user privacy. Besides, the works found in the literature do not consider the city in which the user resides. Thus, users who live in very different cities, such as a metropolis and a city in the countryside, have no distinction. This work uses only the users' list of applications and city to propose an investigation of the mobile devices users' profiles to mitigate these problems. For this purpose, we analyze the users' profiles from five perspectives: the device's price, types of establishments and functionalities of the areas visited, way of commuting, and their personality traits. We propose the MAPED Profile, which encompasses all the other profiles analyzed to obtain an integrated user profile. Also, we enriched a real database containing data from thousands of mobile users with data from different sources. In this way, it was possible to segment users based on their installed applications and the kind of city where they reside. Besides, to streamline the semantic enrichment stage, we developed an easily extensible framework that may resolve various problems.

Keywords: Data Mining. Users Profile. Mobile Data.

## LISTA DE FIGURAS

2.1	Exemplo de área considerando três funcionalidades. A área <i>A1</i> é 66,67% governamental e 33,33% comercial. Já a área <i>A2</i> é 100% comercial. A área <i>A3</i> é composta por 50% de locais comerciais, 25% educacional e 25% governamental. A área <i>A4</i> é 100% educacional. . . . .	20
2.2	Distâncias entre vizinhos (Baseada na Figura 6 de (Brodsky, 2019)). . .	22
2.3	Exemplo de indexação hierárquica (Baseada na Figura 8 de (Brodsky, 2019)). . . . .	22
2.4	Exemplo de <i>returner</i> e <i>explorer</i> para $k = 2$ ( $cm$ = centro de massa). . . . .	23
2.5	Ilustração da estrutura dos Perfis . . . . .	28
3.1	Visão geral dos componentes do Framework . . . . .	34
3.2	Implementação do módulo Enriquecedor . . . . .	35
3.3	Implementação do módulo Construtor . . . . .	37
3.4	Exemplos das telas do <i>APPprofile</i> . . . . .	41
4.1	Distribuição de usuários por quantidade de PoIs. A distribuição é de cauda longa onde há usuários com cerca de 100 pontos de interesse. . .	43
4.2	Proporção de usuários que possui instalado pelo menos um aplicativo da categoria. Algumas categorias, como, por exemplo, redes sociais, possuem aplicativos que podem vir de fábrica (ex. <i>Youtube</i> ) ocasionando altos valores mesmo que o usuário não os utilizem. . . . .	45
4.3	Quantidade de usuários por classe de população. . . . .	46
4.4	Quantidade de usuários por classe de PIB <i>per capita</i> . . . . .	46
4.5	Fluxo da criação dos modelos. . . . .	47
5.1	Os 30 modelos de aparelhos móveis mais utilizados. . . . .	52
5.2	Quantidade de usuários por classe de preço. . . . .	53
5.3	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Entrada</i> (sup=Suporte) . . .	53
5.4	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Intermediário</i> (sup=Suporte) . . .	54
5.5	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Mid-high</i> (sup=Suporte) . . .	55
5.6	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>High-end</i> (sup=Suporte) . . .	55
5.7	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Premium</i> (sup=Suporte) . . .	56
5.8	Resultados das métricas Precisão, Revocação e F-Score para as 5 classes de preços. . . . .	56
5.9	Quantidade de locais coletados para cada tipo do OSM. . . . .	57
5.10	Quantidade de usuários que possuem pelo menos um ponto de interesse em cada um dos locais do OSM. Os valores apresentados são exclusivamente para cada tipo, logo podem haver usuários em comum. . . . .	58
5.11	Os 5 maiores e 5 menores <i>Lifts</i> para <i>Aeroporto</i> (sup=Suporte) . . . . .	59
5.12	Os 5 maiores e 5 menores <i>Lifts</i> para <i>Hospital</i> (sup=Suporte) . . . . .	60
5.13	Os 5 maiores e 5 menores <i>Lifts</i> para <i>Shopping</i> (sup=Suporte) . . . . .	60
5.14	Os 5 maiores e 5 menores <i>Lifts</i> para <i>Universidade</i> (sup=Suporte) . . . . .	61
5.15	Resultados das métricas Precisão, Revocação e F-Score para os 4 tipos de locais. . . . .	62

5.16	Quantidade de usuários que possuem interesse na funcionalidade. Os valores apresentados são exclusivamente para cada funcionalidade, logo podem haver usuários em comum. . . . .	64
5.17	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Alimentação</i> (sup=Suporte)	65
5.18	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Companhia</i> (sup=Suporte) .	65
5.19	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Educação</i> (sup=Suporte) . .	66
5.20	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Entretenimento</i> (sup=Suporte)	66
5.21	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Esportes</i> (sup=Suporte) . . .	67
5.22	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Governo</i> (sup=Suporte) . . .	67
5.23	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Shopping</i> (sup=Suporte) . .	68
5.24	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Turismo</i> (sup=Suporte) . . .	68
5.25	Resultados das métricas Precisão, Revocação e F-Score para os 8 classes de funcionalidade. . . . .	69
5.26	Quantidade de usuários por classe de mobilidade. . . . .	70
5.27	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Returner</i> (sup=Suporte) . .	71
5.28	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Explorer</i> (sup=Suporte) . . .	72
5.29	Resultados das métricas Precisão, Revocação e F-Score para as classe <i>Returner</i> e <i>Explorer</i> considerando o k igual a 4. . . . .	72
5.30	Percentagens de usuários por categoria de aplicativo da <i>Play Store</i> . . . .	75
5.31	A distribuição dos 5 traços de personalidade. . . . .	76
5.32	Quantidade de usuários por Traço de Personalidade. . . . .	76
5.33	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Agradabilidade</i> (sup=Suporte)	77
5.34	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Aberto a Novas Experiências</i> (sup=Suporte) . . . . .	77
5.35	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Conscienciosidade</i> (sup=Suporte)	78
5.36	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Estabilidade Emocional</i> (sup=Suporte) . . . . .	78
5.37	Os 5 maiores e 5 menores <i>Lifts</i> para a classe <i>Extroversão</i> (sup=Suporte) .	79
5.38	Resultados das métricas Precisão, Revocação e F-Score para os 5 traços de personalidade. . . . .	80
5.39	Ilustração da estrutura do Perfil MAPED. . . . .	81
5.40	Resultado da média das execuções com intervalo de confiança de 95% do <i>Lift</i> para o Perfil MAPED. . . . .	83

## LISTA DE TABELAS

2.1	Comparação entre APIs de coleta de dados geolocalizados . . . . .	19
2.2	Questionário TIPI (Adaptado de <a href="#">Gosling et al. (2003)</a> ) . . . . .	25
2.3	Trabalhos Relacionados a Perfil de Usuário . . . . .	27
3.1	Trabalhos Relacionados ao <i>Framework</i> . . . . .	32
4.1	Exemplificação dos dados para três usuários fictícios . . . . .	43
5.1	Exemplificação dos dados finais para três usuários fictícios . . . . .	75
5.2	Resultado dos modelos de regressão para os traços de personalidade. . . . .	80
5.3	Exemplo de saída do Perfil MAPED. . . . .	82
A.1	Aplicativos e suas categorias . . . . .	90

# SUMÁRIO

1	INTRODUÇÃO	12
1.1	O problema e sua importância	14
1.2	Objetivos	14
1.3	Contribuições	15
1.4	Estrutura do Texto	15
2	REFERENCIAL TEÓRICO	17
2.1	Dados de Sistemas Móveis	17
2.2	Pontos de Interesse	18
2.3	Dados Demográficos	19
2.4	Dados de Estabelecimentos	19
2.5	Área Funcional	20
2.6	Mobilidade	22
2.7	Traços de Personalidade	24
2.8	Extração de Perfil de Usuário	27
3	FERRAMENTAS	30
3.1	<i>MoreData</i> : Um <i>Framework</i> para Enriquecimento de Dados	30
3.1.1	Trabalhos Relacionados	31
3.1.2	Arquitetura	33
	Enriquecedor	34
	Conector Elasticsearch	35
	Conector API	35
	Conector OSM	36
	Conector Banco Relacional	36
	Construtor	36
	Conversor	37
3.1.3	Exemplo	37
3.2	Aplicativo <i>APPprofile</i>	41
4	MATERIAIS E MÉTODOS	42
4.1	Dados Base	42
4.1.1	Base de Usuários	42
4.1.2	Base de Categorias de Aplicativos	44
4.1.3	Base de Cidades	44
4.2	Modelos Preditivos	46
4.2.1	<i>Resampling</i>	47
4.2.2	<i>Split</i>	48
4.2.3	<i>PCA</i>	48
4.2.4	<i>Grid Search CV</i>	48
4.2.5	<i>Predição</i>	48
4.3	Métricas	49
5	RESULTADOS E DISCUSSÕES	51
5.1	Perfil de Dispositivo	51

5.1.1	Descrição dos Dados . . . . .	51
5.1.2	Análise . . . . .	53
5.1.3	Modelos Preditivos . . . . .	55
5.2	Perfil de Estabelecimento . . . . .	57
5.2.1	Descrição dos Dados . . . . .	57
5.2.2	Análise . . . . .	58
5.2.3	Modelos Preditivos . . . . .	61
5.3	Perfil de Área Funcional . . . . .	63
5.3.1	Descrição dos Dados . . . . .	63
5.3.2	Análise . . . . .	64
5.3.3	Modelos Preditivos . . . . .	68
5.4	Perfil de Mobilidade . . . . .	70
5.4.1	Descrição dos Dados . . . . .	70
5.4.2	Análise . . . . .	71
5.4.3	Modelos Preditivos . . . . .	72
5.5	Perfil de Personalidade . . . . .	74
5.5.1	Descrição dos Dados . . . . .	74
5.5.2	Análise . . . . .	76
5.5.3	Modelos Preditivos . . . . .	78
	Modelo de Regressão . . . . .	79
5.6	Perfil Completo MAPED . . . . .	81
5.6.1	Análises e Resultados . . . . .	82
6	CONCLUSÃO . . . . .	84
6.1	Publicações . . . . .	85
	REFERÊNCIAS BIBLIOGRÁFICAS . . . . .	86
	APÊNDICE A APLICATIVOS E SUAS CATEGORIAS . . . . .	90

# Capítulo 1

## Introdução

Atualmente, os usuários de sistemas móveis geram uma grande variedade de dados como localizações de acessos, modelo do aparelho, nível de bateria, tipo de rede, lista de aplicativos, entre outros. A fim de conhecer melhor seus usuários, as empresas e a comunidade científica podem utilizar estes dados para a criação de perfis dos usuários. Contudo, utilizar somente os dados brutos gerados pelos usuários não é suficiente para a criação de um perfil com informações relevantes, tornando esta uma tarefa dispendiosa.

A complexidade de criar um perfil que possa contribuir com informações pertinentes está diretamente relacionada com os dados utilizados. Afirma-se isso pois, utilizar somente os dados brutos gerados pelos usuários pode produzir um perfil com baixa generalização. Isso visto que, estes tipos de dados são, em sua maioria, bastante específicos, o que dificulta o agrupamento de usuários para encontrar padrões capazes de auxiliar em algum problema; em outras palavras, formar comunidades de usuários com informações relevantes. Assim sendo, para desenvolver perfis com informações expressivas é preciso realizar manipulações e enriquecimentos nos dados brutos capazes de contribuir para compreender melhor o usuário, e assim impactar positivamente o perfil em questão.

Há diferentes possibilidades de fontes e tipos de dados para enriquecer uma base (e.g., dados climáticos, preço do aparelho, tipo de local visitado, categorias dos aplicativos instalados). No âmbito de dispositivos móveis, os avanços tecnológicos permitem a fácil coleta da localização em diferentes contextos (Naboulsi et al., 2016), tais como: *check-ins*, trajetórias, ocorrência de crimes, eventos climáticos, dentre outros. Dessa forma, é possível agregar dados geoespaciais coletados de diferentes fontes e em diversos contextos, fazendo com que haja várias maneiras de enriquecer uma única base de dados. Por exemplo, podemos enriquecer uma base de localizações de acessos de usuários de um aplicativo com informações dos tipos de locais que o usuário visitou (e.g., se é um aeroporto, uma universidade, ou algum outro tipo), dados demográficos da região em que o dado foi gerado (e.g., com base em dados de censo), dados climáticos, tempo gasto em transporte (Santin et al., 2020), entre outras formas. Com isso, o acréscimo deste tipo de informação contribui para que as empresas e a comunidade científica possam realizar análises mais detalhadas e específicas sobre os

clientes e seus lugares de preferência, melhorando seus serviços.

Outra maneira de enriquecer os dados semanticamente é utilizar dados comuns (não georreferenciados), muitos destes disponíveis na *Web*. Como as lojas *Google Play Store* e *Apple Store* que forcem diferentes dados sobre seus aplicativos, como a categoria ao qual o aplicativo pertence, o nome de seu pacote e o seu nome usual. Outro exemplo disso é a utilização do dado bruto referente ao modelo do aparelho. Com este dado, é possível utilizar informações de *e-commerce* para agregação de dados como preço do aparelho, capacidade de armazenamento e processamento, dentre outras informações técnicas. Assim, pode-se gerar uma base de dados mais completa, levando a mais possibilidades de analisar os perfis dos usuários em diferentes perspectivas.

O enriquecimento semântico pode ajudar a solucionar diversos tipos de problemas, como os dois exemplos descritos a seguir. [Sardianos et al. \(2018\)](#) realiza o enriquecimento de arquivos de *log* do Google Maps com informações do *OpenStreetMap* para resolver o problema de melhoramento de rota e obtenção de pontos de interesse do usuário. Já o trabalho de [Maynard and Bontcheva \(2015\)](#) utiliza dados de uma ontologia para enriquecer *tweets* e assim entender as mudanças climáticas que estão ocorrendo no mundo. Portanto, o enriquecimento semântico é um recurso que pode contribuir para o melhoramento dos mais diversos tipos de problemas, sendo a criação de perfis um deles.

Levando estes fatos em consideração, analisar o perfil do usuário sob diferentes perspectivas pode ajudar na solução de diversos problemas. Assim, nesta dissertação são apresentados modelos para extrair o perfil do usuário em cinco diferentes perspectivas utilizando como base somente a lista de aplicativos e a sua cidade. Para isso, vários enriquecimentos semânticos de diferentes fontes são realizados para complementar a base de dados inicial.

Os trabalhos encontrados na literatura abordam diferentes tipos de perfis, como psicológico ([Alvarez-Lozano et al., 2014](#)), social ([Li et al., 2008](#)) e de mobilidade ([Varlamis et al., 2020](#)). Muitas destas pesquisas necessitam da participação ativa do usuário, isto é, solicitação recorrente de dados (e.g., respostas diretas a questionários) ou validação das previsões realizadas (e.g., se o usuário esteve em determinado lugar em determinada hora). Em virtude disso, estas pesquisas geralmente possuem um número reduzido de voluntários, algumas somente 1 usuário, o que provoca a baixa generalização das análises e modelos gerados. Ademais, nenhum dos trabalhos encontrados considera diretamente o impacto do tipo de cidade em que o usuário reside em suas análises. O problema disto é que a forma como um habitante de uma metrópole se comporta é diferente se comparado com um residente de cidade pequena, por exemplo. Visando sanar estas dificuldades, os perfis propostos utilizam dados ativos somente para treinamento dos modelos preditivos. Também, os dados referentes à

cidade do usuário são utilizados sempre que forem pertinentes nos modelos.

## 1.1 O problema e sua importância

Este trabalho é motivado pela crescente necessidade das empresas e pesquisadores conhecerem o comportamento de usuários móveis. Em relação às empresas, atualmente, há um grande número de prestadores de serviços do mesmo segmento o que gera concorrência. Com o conhecimento sobre o perfil de seus clientes será possível resolver diferentes problemas como, por exemplo: recomendações mais assertivas de serviços móveis, identificação dos laços sociais (familiar, amizade, etc.), identificação de rotinas para detecção de fraude, entre outros. Com isso, as empresas podem prestar serviços cada vez mais personalizados para seus clientes e assim os mesmos poderão ter serviços com mais qualidade e que atendam melhor suas necessidades.

Já em relação à comunidade científica, os pesquisadores têm investigado como os usuários móveis se comportam para avançar o conhecimento sobre psicologia e comportamento humano. Além disso, o conhecimento sobre o perfil dos usuários móveis pode auxiliar nos direcionamentos de novas pesquisas e no desenvolvimento de produtos inovadores e de sistemas de transportes inteligentes das cidades do futuro.

## 1.2 Objetivos

Considerando todas as informações citadas anteriormente, é possível perceber que pode-se utilizar diferentes dados para a criação de um perfil. Estes dados podem invadir a privacidade do usuário (e.g., *log* de ligações) ou necessitar da participação ativa do mesmo. Além disso, o perfil criado pode não considerar o tipo de cidade em que o usuário reside, logo usuários de contextos completamente diferentes são tratados da mesma forma. Nesse contexto, uma pergunta que surge é se os aplicativos instalados de um usuário, e também as características de sua cidade, são fontes de informações relevantes que permitem traçar o perfil do usuário em diferentes perspectivas.

O objetivo geral desta dissertação é validar a hipótese de que os aplicativos instalados e dados demográficos da cidade do usuário favorecem a extração de perfis de usuários móveis. Para alcançar esse objetivo, são considerados os seguintes objetivos específicos:

- Investigar se os aplicativos instalados e dados sobre a população e a renda/capita da cidade do usuário favorecem a extração de perfil em termos do preço do dispositivo do usuário;

- Investigar se os aplicativos instalados e dados sobre a população e a renda/capita da cidade do usuário favorecem a extração de perfil em termos do interesse do usuário em visitar locais como Shoppings, Aeroportos, Universidades e Hospitais;
- Investigar se os aplicativos instalados favorecem a extração de perfil em termos dos traços de personalidade do usuário;
- Investigar se os aplicativos instalados e dados sobre a população e a renda/capita da cidade do usuário favorecem a extração de perfil em termos de características de mobilidade relacionadas ao raio de giro do usuário;
- Investigar se os aplicativos instalados e dados sobre a população e a renda/capita da cidade do usuário favorecem a extração de perfil em termos das características funcionais da regiões visitadas pelo usuário.

### 1.3 Contribuições

Dentre as contribuições deste trabalho, pode-se citar as seguintes:

- Cinco estudos diferentes do perfil do usuário, sendo eles:
  - Dispositivo: análise da faixa de preço do dispositivo do usuário;
  - Estabelecimento: verificação dos tipos de locais mais visitados pelo usuário;
  - Área Funcional: análise do tipo de região em que o usuário visita;
  - Mobilidade: estudo da maneira em que o usuário se locomove no dia a dia;
  - Personalidade: análise dos traços de personalidade do usuário.
- Perfil MAPED, perfil completo contendo todos as cinco perspectivas;
- *MoreData*, um framework facilmente extensível e que pode ser utilizado em diferentes problemas para o enriquecimento semântico de dados;
- *APPProfile*, um aplicativo *Android* para coletar dados dos usuários, como as listas de aplicativos e as respostas sobre traços de personalidades.

### 1.4 Estrutura do Texto

Esta dissertação está estruturada da seguinte forma:

No Capítulo 2 são apresentados o referencial teórico referente ao tema. No Capítulo 3 são descritas as ferramentas desenvolvidas, o *framework* e o aplicativo, bem

como suas componentes. Também são apresentados exemplos de aplicações do *framework* para ilustrar sua simplicidade de uso. No Capítulo 4 são descritos todos os dados e métricas utilizados. O Capítulo 5 apresenta os perfis analisados para formarem o perfil completo. Ademais, apresenta os modelos gerados e discutir seus resultados. No Capítulo 6 são discutidas as conclusões finais e descritos possíveis trabalhos futuros. Por fim, o Apêndice A mostra os aplicativos utilizados e suas respectivas categorias.

# Capítulo 2

## Referencial Teórico

Os dados de sistemas móveis podem ser utilizados para traçar os perfis dos usuários em diferentes perspectivas. Com base nisso, esse tipo de dado será utilizado como base para diversos enriquecimentos semânticos visando a extração dos perfis. Logo, faz-se necessário explicar os diferentes contextos explorados nesta dissertação. Levando isso em consideração, neste capítulo são apresentados os conceitos fundamentais e os trabalhos relevantes para o entendimento deste trabalho.

### 2.1 Dados de Sistemas Móveis

Os avanços tecnológicos facilitaram a coleta das mais diversas fontes de dados. No âmbito de sistemas móveis, é possível coletar dados como: *log* de acesso, localização, aplicativos instalados, uso de aplicativos, modelo do dispositivo, nível de bateria, dados da rede móvel, *check-ins*, entre outros. Com isso, é possível realizar estudos mais detalhados e específicos sobre o comportamento humano.

No que concerne aos tipos de coleta, duas delas se destacam: as coletas ativa e passiva. A coleta ativa é aquela em que faz-se necessária a participação direta do usuário, através de perguntas referentes ao nome, *e-mail*, preferência de alguma configuração, entre outras formas. Este tipo de coleta é mais vantajosa quando se deseja obter dados pessoais mais precisos (e.g., a idade exata de um usuário). Entretanto, a coleta ativa pode depender diretamente do número de voluntários, o que pode levar a um menor volume de dados. Além disso, os dados coletados podem estar diretamente relacionados à percepção e vontade do usuário, o que pode levar à má classificação desses.

Já a coleta passiva de dados não necessita da participação direta do usuário. Nela são coletados, de forma não perceptível ao usuário, dados como o nome do navegador usado, lista de aplicativos, modelo do aparelho, localização, dentre outros. Por estes motivos, a coleta passiva pode obter um volume de dados maior e com menos ruídos que a ativa por não depender diretamente do usuário.

Considerando estes fatores, somente uma base de dados utilizada nesta dissertação foi coletada de forma ativa, respostas referentes aos traços de personalidade, pois são dados da autoavaliação do usuário. Mesmo assim, esta base é utilizada somente

para a etapa de treinamento dos modelos preditivos do Perfil de Personalidade. Independente da forma de coleta utilizada, é possível perceber a importância do uso de dados de sistemas móveis na construção de perfis em diferentes perspectivas. Pois, é possível utilizar desta fonte de dados como base para vários enriquecimentos, a fim de conhecer melhor o usuário.

## 2.2 Pontos de Interesse

Um dos dados provenientes de sistemas móveis são as localizações brutas que podem ser utilizadas para diversas finalidades; dentre elas destacam-se a identificação e classificação de Pontos de Interesse (em inglês PoIs, *Points of Interest*). Estes pontos representam locais que um usuário visita com certa frequência, podendo corresponder a locais como residência, trabalho, atividades de lazer e educação, entre outros. Neste âmbito, a importância da identificação de PoIs se justifica pela filtragem de localizações pouco visitadas e na sintetização de localizações muito próximas que podem representar um mesmo local. Com isso, após identificado, é possível classificar este ponto (i.e., residência, trabalho, entre outros) para auxiliar a compreensão dos padrões de deslocamento do usuário. Logo, é possível perceber que PoIs são fontes importantes nas elaborações de perfis dos usuários.

Os trabalhos na literatura que abordam este tema podem ser divididos de acordo com o tipo dos dados que se baseiam para identificação dos pontos de interesse. Os trabalhos que utilizam dados densos, dados de GPS coletados com uma alta frequência (i.e., na ordem de poucos segundos), podem extrair informações mais detalhadas dos usuários, por exemplo, tempo de permanência em um determinado local e trajeto de deslocamento. Contudo, estes trabalhos tornam-se custosos computacionalmente quando o número de usuários é muito grande, pois o volume de dados coletados necessita de uma alta capacidade de armazenamento e de processamento no servidor. Em contrapartida, os trabalhos que utilizam dados esparsos, ou seja, que foram coletados em intervalos longos de tempo, possuem coletas mais simples e menos custosas. Porém, esse tipo de dado leva a uma dificuldade maior para a identificação de PoIs pois se tem menos detalhes das visitas.

Levando todas essas informações em consideração, esta dissertação baseia-se no trabalho de [Capanema et al. \(2019\)](#) para identificar e classificar pontos de interesse. Esta escolha deve-se pelo fato dos autores apresentarem uma abordagem baseada em dados esparsos e personalizada para cada usuário, ou seja, não é fixado nenhum período do dia para a classificação do PoI. Com isso, nesta dissertação as localizações brutas do usuário foram transformadas nas seguintes classes de ponto de interesse: *Principal*, *Casa*, *Trabalho* e *Outro*. Sendo o PoI *Principal* o mais visitado pelo usuário, ou seja, ele pode ser qualquer uma das outras classes, e o PoI *Outro* aquele que não é

identificado como *Casa* ou *Trabalho*.

## 2.3 Dados Demográficos

Dados demográficos são fontes importantes para compreender e analisar os diferentes aspectos de uma população. Estes dados são, geralmente, referentes a uma amostra representativa da população e possuem dados como distribuição por raça, distribuição por nível de escolaridade, distribuição por renda, dentre outros. Com isso, dados demográficos podem ser utilizados para compreender os diferentes perfis dos usuários.

O IBGE (Instituto Brasileiro de Geografia e Estatística) é um instituto público que produz e analisa dados estatísticos referentes ao Brasil. Uma das fontes de dados demográficos disponibilizadas por este instituto é o IBGE Cidades (IBGE, 2020). Esta base descreve características de 5.570 cidades, de acordo com uma pesquisa realizada pelo IBGE em 2018, nela encontra-se a geometria (no formato de polígono) que representa a cidade, além de fornecer 27 indicadores socioeconômicos como PIB/capita, tamanho da população, atividade com maior valor adicionado bruto, se pertence ao Semiárido, dentre outros. O IBGE disponibiliza estes dados através de uma API onde é possível realizar requisições para coletar estes dados.

## 2.4 Dados de Estabelecimentos

Outra fonte importante de dados são as bases de dados geoespaciais provenientes de mapas. Este tipo de base possui diversos dados sobre uma região geográfica, podendo haver dados do mundo inteiro. Pode-se citar três APIs que facilitam a coleta de dados geoespaciais, e elas são: *Overpass* (OpenStreetMap, 2021) que recupera dados do *OpenStreetMap*, *Google Places* (Google, 2020) que coleta dados das bases do *Google* e *Places Graph* (Facebook, 2021) do *Facebook*. Uma comparação entre as APIs é apresentada na Tabela 2.1.

Tabela 2.1: Comparação entre APIs de coleta de dados geolocalizados

	<i>Overpass</i>	<i>Google Places</i>	<i>Places Graph</i>
<b>Limitação de Coleta</b>	10.000 solicitações diárias ou 1 GB diário	Ilimitado	500 solicitações diárias
<b>Custo</b>	Grátis	Grátis somente para desenvolvimento	Grátis

Como é possível ver com a Tabela 2.1 as diferentes APIs podem ser utilizadas em diferentes contextos. Caso seja necessária realizar muitas solicitações diárias e seja

possível pagar pela coleta, o *Google Places* é o mais indicado. Já o *Overpass* possui um limite diário de coleta consideravelmente bom além de ser gratuito e *software* livre.

Levando tudo isso em consideração, o *Overpass* foi escolhido por atender as necessidades para a realização deste trabalho. Dessa forma, os dados geoespaciais tiveram importante participação para a criação do perfil de Estabelecimento por permitir caracterizar os pontos de interesse do usuário com os tipos de estabelecimentos correspondentes.

## 2.5 Área Funcional

Outras associações que podem ser feitas utilizando as localizações de acesso do usuário são as referentes às Áreas Funcionais. Uma Área Funcional pode ser explicada como uma região que é caracterizada por diferentes tipos de locais, por exemplo a porcentagem de locais dos tipos educacional, governamental e comercial (Yu et al., 2018). Uma exemplificação da área funcional considerando esses tipos é apresentada na Figura 2.1. Com isso, é possível utilizar as localizações do usuário para analisar os tipos de áreas que ele frequenta para serem correlacionados com seus aplicativos. Assim, será possível avaliar, por exemplo, se um usuário que visita uma área que possui em sua maioria locais do tipo educação (escolas, universidades, entre outros) possui muitos aplicativos educacionais instalados. A seguir será formalizado como calcular o *score* a partir da porcentagem dos tipos de áreas que o usuário visita.

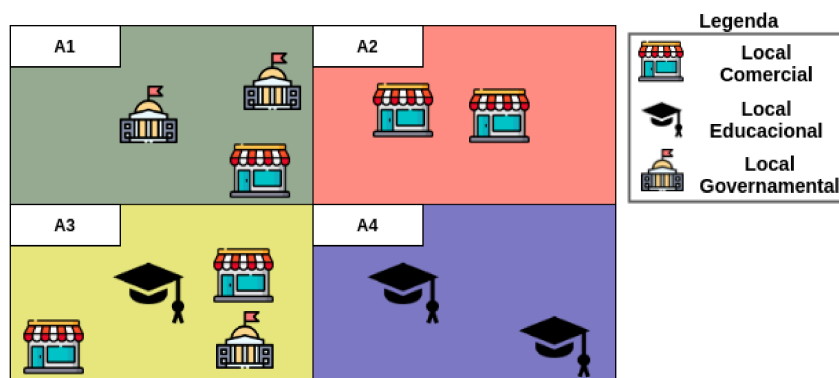


Figura 2.1: Exemplo de área considerando três funcionalidades. A área  $A1$  é 66,67% governamental e 33,33% comercial. Já a área  $A2$  é 100% comercial. A área  $A3$  é composta por 50% de locais comerciais, 25% educacional e 25% governamental. A área  $A4$  é 100% educacional.

Considere  $\mathbf{D}$  um conjunto de dados que contenha as localizações de  $m$  usuários e  $n$  áreas que podem possuir  $q$  diferentes funcionalidades. Inicialmente, é necessário contabilizar  $freq_{ij}$ , número de visitas do usuário  $u_i$  que estão contidas na área  $a_j$ , sendo  $\{i, j \in \mathbb{N} | 1 \leq i \leq m; 1 \leq j \leq n\}$ . Assim, o objeto  $\vec{u}_i = (freq_{i1}, freq_{i2}, \dots, freq_{in})$  representa a frequência de visitas do usuário  $u_i$  nas  $n$  áreas. Com isso, cada linha da

matriz  $\mathbf{F}$  representa um vetor de frequências de visita de cada um dos  $m$  usuários nas  $n$  áreas.

$$\mathbf{F}_{m \times n} = \begin{matrix} & a_1 & a_2 & \dots & a_n \\ \vec{u}_1 & \left( \begin{array}{cccc} freq_{11} & freq_{12} & \dots & freq_{1n} \\ \vec{u}_2 & freq_{21} & freq_{22} & \dots & freq_{2n} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \vec{u}_m & freq_{m1} & freq_{m2} & \dots & freq_{mn} \end{array} \right) \end{matrix} \quad (2.1)$$

Já a matriz  $\mathbf{P}$  denota os pesos de cada uma das  $q$  funcionalidades nas  $n$  áreas. Para calcular o peso  $p_{jk}$  de uma funcionalidade  $func_k$ , sendo  $\{k \in \mathbb{N} | 1 \leq k \leq q\}$ , em uma área  $a_j$  é necessário computar a quantidade de locais de cada uma das funcionalidades naquela área, para então calcular a percentagem de locais da funcionalidade que a área possui.

$$\mathbf{P}_{n \times q} = \begin{matrix} & func_1 & func_2 & \dots & func_q \\ a_1 & \left( \begin{array}{cccc} p_{11} & p_{12} & \dots & p_{1q} \\ a_2 & p_{21} & p_{22} & \dots & p_{2q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ a_n & p_{n1} & p_{n2} & \dots & p_{nq} \end{array} \right) \end{matrix} \quad (2.2)$$

Finalmente, para obter os *scores*, matriz  $\mathbf{S}$ , dos usuários é necessário realizar um produto entre as matrizes  $\mathbf{F}$  e  $\mathbf{P}$ .

$$\mathbf{S}_{m \times q} = \mathbf{F}_{m \times n} \times \mathbf{P}_{n \times q} = \begin{matrix} & func_1 & func_2 & \dots & func_q \\ u_1 & \left( \begin{array}{cccc} s_{11} & s_{12} & \dots & s_{1q} \\ u_2 & s_{21} & s_{22} & \dots & s_{2q} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ u_m & s_{m1} & s_{m2} & \dots & s_{mq} \end{array} \right) \end{matrix} \quad (2.3)$$

Logo, se analisarmos a matriz resultante  $\mathbf{S}$  por linha é possível verificar quais tipos funcionais o usuário possui maior contato. Já com a análise por coluna é possível verificar quais usuários mais visitam áreas de uma determinada funcionalidade. Dessa forma, é possível perceber que há várias possibilidades de correlacionar as áreas funcionais com algum outro dado do usuário para a criação de um perfil.

Há várias maneiras de expressar uma área geometricamente, geralmente, para facilitar a realização de operações espaciais (interseção, cruzamento, entre outras) são utilizadas formas geométricas. Nesta dissertação as áreas são formadas a partir do índice H3 (*A Hexagonal Hierarchical Geospatial Indexing System*) (Brodsky, 2019). Uma das motivações para se utilizar o H3 é o fato que ele utiliza hexágonos para representar as áreas. Isto diminui os erros de mapeamento e também faz com que as distâncias

entre todos os seus vizinhos sejam praticamente iguais. Desse modo, impactando positivamente em algumas operações geoespaciais, ao contrário da utilização de outras formas geométricas como pode ser visto na Figura 2.2. Já outra motivação é a implementação de um sistema hierárquico de indexação o que agiliza a busca de uma área específica, Figura 2.3.

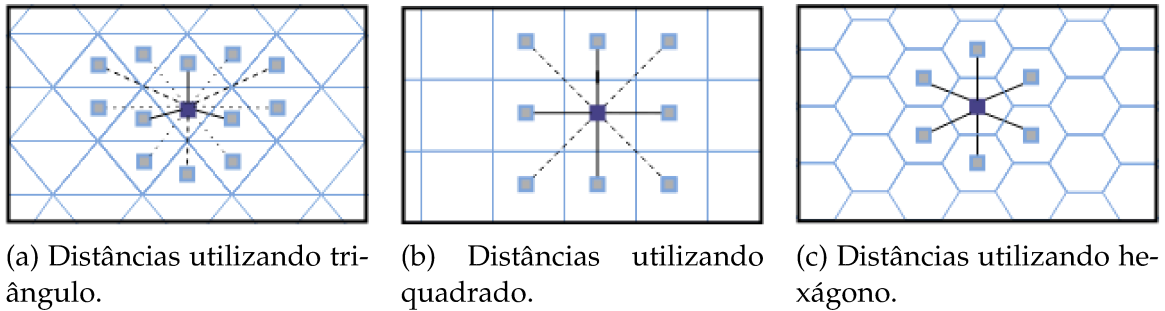


Figura 2.2: Distâncias entre vizinhos (Baseada na Figura 6 de (Brodsky, 2019)).

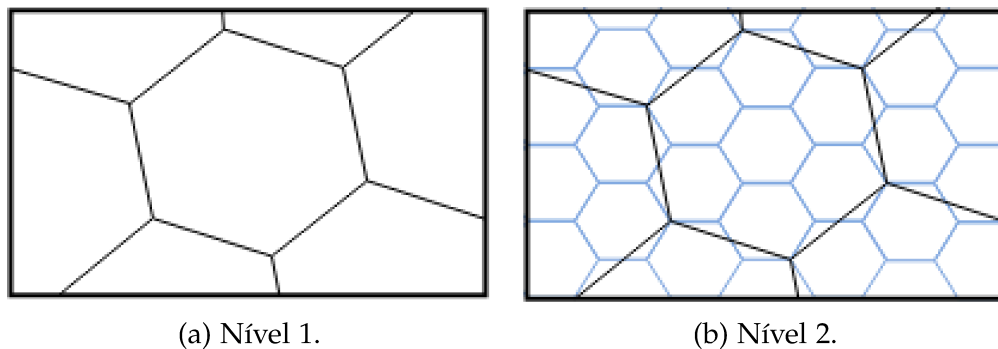


Figura 2.3: Exemplo de indexação hierárquica (Baseada na Figura 8 de (Brodsky, 2019)).

## 2.6 Mobilidade

A forma como as pessoas se deslocam pode também dizer muito sobre elas. Pappalardo et al. (2015) apresentam um estudo do perfil de deslocamento social que é dividido em duas classes, *returners* e *explorers*. São denominados *returners* aqueles usuários que geralmente visitam um número menor de locais de forma mais regular. Já os *explorers* são aqueles usuários que visitam um grande número de locais diferentes. Para classificar os usuários em uma das duas classes é necessário selecionar suas localizações mais frequentes (pontos de interesse) e avaliar seu raio de giro geral e relativo. Raio de giro pode ser definido como a distância média entre uma localização de referência e as outras localizações visitadas. Com isso, para avaliar o impacto das  $k$  localizações mais frequentadas pelo usuário, são realizadas comparações entre o

raio de giro de todas as localizações mais frequentadas e o raio de giro considerando somente as  $k$  localizações mais frequentes.

Levanto tudo isso em consideração, considere classificar um usuário que possua um conjunto  $L$  de localizações frequentes, onde a localização mais frequente  $L_1$  é o local com maior probabilidade de visita do usuário, geralmente sua casa. Assim, a localização  $L_k$  é o  $k$ -ésimo local mais frequentado pelo usuário. Sendo uma localização  $L_i$  representada por um vetor bidimensional  $\vec{r}_i$  e  $n_i$  é a frequência de visitas ou tempo total gasto do usuário na localização  $L_i$ .  $N = \sum_{i \in L} n_i$  é o número total de visitas ou tempo gasto pelo usuário em todas as localizações e  $r_{cm}$  é o centro de massa. Dessa forma, o raio de giro  $r_g$  do usuário é dado por:

$$r_g = \sqrt{\frac{1}{N} \sum_{i \in L} n_i (\vec{r}_i - r_{cm})^2} \quad (2.4)$$

Para compreender a importância da  $k$ -ésima localização mais frequente do usuário para definir a classe do comportamento de mobilidade,  $k$ -returners ou  $k$ -explorers, é necessário calcular o  $k$ -raio de giro  $r_g^{(k)}$ . O  $k$ -raio de giro é simplesmente o raio de giro considerando as localizações até o  $k$ -ésimo termo, ou seja, as localizações  $L_1, \dots, L_k$ . Assim,  $N_k$  e  $r_{cm}^{(k)}$  são, respectivamente, a soma dos pesos e o centro de massa para as  $k$  localizações mais frequentes. Logo,

$$r_g^{(k)} = \sqrt{\frac{1}{N_k} \sum_{i=1}^k n_i (\vec{r}_i - r_{cm}^{(k)})^2} \quad (2.5)$$

Com os valores dos raios de giro  $r_g$  e  $r_g^{(k)}$  calculados, é possível separar a população em uma das duas classes. Para isso, os autores especificam o uso de três diferentes métodos: SVM (*Support Vector Machine*), clusterização *expectation-maximization* e a classificação simples. Nesta dissertação foi selecionada a classificação simples, pois, de acordo com os autores, conseguiu alcançar resultados similares aos demais. Dessa forma, a classificação simples retorna que o usuário é  $k$ -returner se  $r_g^{(k)} > \frac{r_g}{2}$  e  $k$ -explorer, caso contrário.

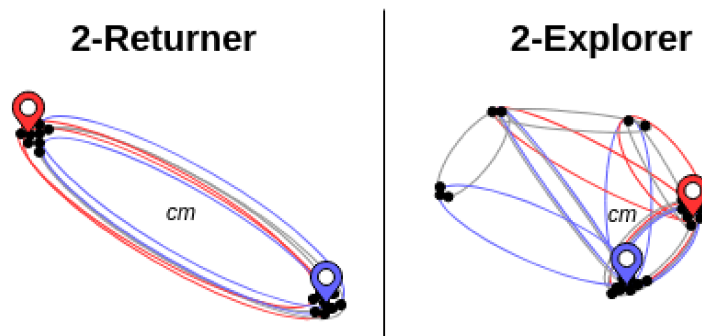


Figura 2.4: Exemplo de *returner* e *explorer* para  $k = 2$  ( $cm$  = centro de massa).

## 2.7 Traços de Personalidade

Um item relevante para a construção do perfil do cliente é a personalidade, que pode ser dividida em cinco traços distintos, conhecidos como *Big Five* (McCrae and Jr., 1999). Conhecer os traços de personalidade de um usuário é interessante pois permite resumir, prever e explicar a conduta de um indivíduo (Silva and Nakano, 2011) e com isso é possível correlacioná-los com produtos e serviços. Os cinco traços são descritos abaixo:

- *Agradabilidade* (AGR): Pessoas agradáveis, são gentis, confiáveis, tolerantes e mais propensas a ajudar os outros;
- *Aberto a Novas Experiências* (ANE): Pessoas abertas a novas experiências, são pessoas criativas, mente aberta, independentes e sempre estão procurando por diferentes experiências;
- *Conscienciosidade* (CON): Indivíduos deste traço são bem organizados, possuem autocontrole, são cuidadosos, persistentes e confiáveis;
- *Estabilidade Emocional* (EME): Pessoas emocionalmente estáveis são bem humoradas, não costumam sentir ansiedade, não são nervosas, possuem autoconfiança e tendem a não ter uma reação negativa diante diferentes situações da vida;
- *Extroversão* (EXT): Pessoas extrovertidas, são comumente pessoas que gostam de se comunicar, ativas socialmente.

Uma das formas de estimar os traços de personalidade é utilizando o questionário TIPI (*Ten-Item Personality Inventory*) (Gosling et al., 2003). Para calcular quantitativamente os traços de personalidade é necessário responder a dez perguntas do questionário TIPI de acordo com cada conceito e depois fazer os cálculos correspondentes para cada traço de personalidade. Na Tabela 2.2 é mostrado uma adaptação do questionário TIPI.

Para calcular a nota do usuário em determinado traço de personalidade, é necessário somar a resposta de uma pergunta com o reverso de outra resposta e calcular a média. Para isso, há um mapeamento entre os valores, caso o usuário tenha respondido com o valor 7 em uma pergunta, este valor será mapeado para o valor 1. Caso tenha respondido com o valor 6, será mapeado para o valor 2 e assim sucessivamente. Por exemplo, para calcularmos o resultado do usuário no traço Extroversão, utilizamos a resposta dele no primeiro item (Extrovertido, entusiasta) e o reverso da resposta no sexto item (Reservado, quieto). As equações para cada um dos traços de personalidade estão descritas abaixo, onde  $R$  é a função que calcula o reverso do valor do conceito dado pelo usuário.

Tabela 2.2: Questionário TIPI (Adaptado de Gosling et al. (2003))

(a) Valores dos Conceitos

Conceito	Valor
Discordo Totalmente	1
Discordo Moderadamente	2
Discordo um Pouco	3
Não discordo nem concordo	4
Concordo um pouco	5
Concordo Moderadamente	6
Concordo Fortemente	7

(b) Itens

Itens
$I_1$ . ___ Extrovertido, entusiasta.
$I_2$ . ___ Crítico, conflituoso.
$I_3$ . ___ Confiável, autodisciplinado.
$I_4$ . ___ Ansioso, facilmente perturbável.
$I_5$ . ___ Aberto à novas experiências, complexo.
$I_6$ . ___ Reservado, quieto.
$I_7$ . ___ Simpático, caloroso.
$I_8$ . ___ Desorganizado, sem cuidado.
$I_9$ . ___ Calmo, emocionalmente estável.
$I_{10}$ . ___ Convencional, sem criatividade.

- Agradabilidade:

$$AGR = \frac{R(I_2) + I_7}{2} \quad (2.6)$$

- Aberto a Novas Experiências:

$$ANE = \frac{I_5 + R(I_{10})}{2} \quad (2.7)$$

- Conscienciosidade:

$$CON = \frac{I_3 + R(I_8)}{2} \quad (2.8)$$

- Estabilidade Emocional:

$$EME = \frac{R(I_4) + I_9}{2} \quad (2.9)$$

- Extroversão:

$$EXT = \frac{I_1 + R(I_6)}{2} \quad (2.10)$$

Porém, extrair os traços de personalidade de um indivíduo é uma tarefa complexa. Até o momento, vários trabalhos já vem buscando resolver esse problema de diversas formas, sendo que muitas vezes usam de mecanismos invasivos que demandam infor-

mações sensíveis dos usuários. Em relação à extração dos traços de personalidade de usuários de serviços móveis, algumas soluções requerem uma coleta ativa, exigindo um comprometimento do usuário.

Alguns trabalhos encontrados na literatura utilizam a lista de aplicativos ou suas respectivas categorias, e seus respectivos usos, para prever os traços de personalidade do usuário. [Runhua Xu and Ilic \(2016\)](#) utilizaram somente as categorias dos aplicativos mais frequentes entre os usuários para correlacionar com os traços de personalidade através de algoritmos de classificação. Já em [Runhua Xu and Ilic \(2015\)](#) somente 13 aplicativos das categorias de redes sociais, comunicação, viagens, leitores de textos e compras foram utilizados. O trabalho de [Raihana Ferdous and Mayora \(2015\)](#) considera somente 5 categorias diferentes para aplicativos e com o objetivo de prever o nível de estresse no ambiente de trabalho. O presente estudo considera apenas as categorias dos aplicativos instalados pelo usuário, sem a necessidade de mais dados, para prever os traços de personalidade.

Outros trabalhos utilizam informações de contexto para estimar os traços do usuário, tais como: nível de bateria do celular, quando o dispositivo está carregando, quando o usuário está conectado em uma rede *WiFi*, quantidade de pessoas com bluetooth ativado perto do usuário, dentre outros. O trabalho de [Viana et al. \(2018\)](#) utiliza essas informações para correlacionar com os traços de personalidade. [Gokul Chittaranjan and Gatica-Perez \(2011\)](#) utilizaram dados dos logs de *Bluetooth*, SMS e ligações juntamente com as classificações de aplicativos como entrada para os classificadores SVM (*Support Vector Machine*) e Árvores de Decisão (C4.5) para prever os traços de personalidade dos usuários. Em [Yves-Alexandre de Montjoye and Pentland \(2013\)](#), os autores utilizam a quantidade de vezes que o usuário mandou mensagens, realizou ligações e ambos juntos para correlacionar com os traços de personalidade.

Outra base para prever os traços de personalidade é o comportamento do usuário nas redes sociais. [Ryan and Xenos \(2011\)](#) e [Jianqiang Shen and Liu \(2014\)](#) baseiam-se no uso do *Facebook*, como o número de álbuns publicados, as quantidades de compartilhamentos e atualizações do *status*, para correlacionar com os traços de personalidade. Contudo, para realização destes estudos são necessárias as coletas de informações invasivas dos usuários. Nesta dissertação foram utilizados somente as respostas dos usuários e sua lista de aplicativos.

Todos estes trabalhos conseguiram estabelecer uma conexão entre os traços de um indivíduo (personalidade, felicidade, comportamento) e alguma informação contextual de perfil do usuário (social ([Doyle and Youn, 2000](#)), mobilidade ([Viana et al., 2018](#)), uso de dispositivos móveis e aplicativos ([Yusong Gao and Zhu, 2014](#))). Esta dissertação tem como objetivo prever os traços de personalidade do indivíduo levando em consideração apenas as categorias de aplicativos que o mesmo tem instalado em seu *smartphone*. Diferentemente dos outros trabalhos, a informação sobre as catego-

rias dos aplicativos é mais fácil de obter e menos invasiva em termos de privacidade do usuário. Ou seja, utilizar somente as categorias de aplicativos diminui a complexidade do modelo gerado e mostra que é possível inferir os traços de personalidade de um usuário com precisão considerável.

## 2.8 Extração de Perfil de Usuário

Com todas as informações citadas anteriormente é possível perceber as diferentes abordagens possíveis para extrair informações a respeito do usuário. Com isso, é possível compreender os padrões sociais e preferências do usuário para construção de perfis que auxiliem na solução de diferentes tipos de problemas. Assim, as empresas e a comunidade científica podem criar soluções cada vez mais personalizadas para compreender melhor o comportamento do usuário. A Figura 2.5 apresenta as camadas de uma estrutura para extração do perfil do usuário.

Tabela 2.3: Trabalhos Relacionados a Perfil de Usuário

Trabalho	Tipo de Perfil	Fonte	Volume
Li et al. (2008)	Nível de Relacionamento	Log do GPS	65 usuários
Do and Gatica-Perez (2010)	Uso de Aplicativos	Log de Aplicativos	111 usuários
Abel et al. (2011)	Perfil Social	Tweets	Aproximadamente 45.000 usuários
Alvarez-Lozano et al. (2014)	Distúrbios de Bipolaridade	Log de Aplicativos	18 usuários
Seneviratne et al. (2015)	Gênero	Atributos Demográficos e Lista de Aplicativos	218 usuários
Malmi and Weber (2016)	Dados Demográficos	Atributos Demográficos e Lista de Aplicativos	3760 usuários
Sardianos et al. (2018)	Tipos de Locais Visitados e Rotas	Log do Google Maps e OSM	1 usuário
Varlamis et al. (2020)	Tipos de Locais Visitados e Rotas	Log do Google Maps e OSM	1 usuário

Na Tabela 2.3 são apresentados os trabalhos encontrados na literatura que abordam a criação de algum tipo de perfil. Seneviratne et al. (2015) realizam um estudo utilizando os dados respondidos em um questionário demográfico e a lista de aplicativos de 218 usuários para correlacionar e prever o gênero do usuário. Para isso, eles utilizaram SVM (*Support Vector Machine*) e *Naive Bayes* com diferentes combinações

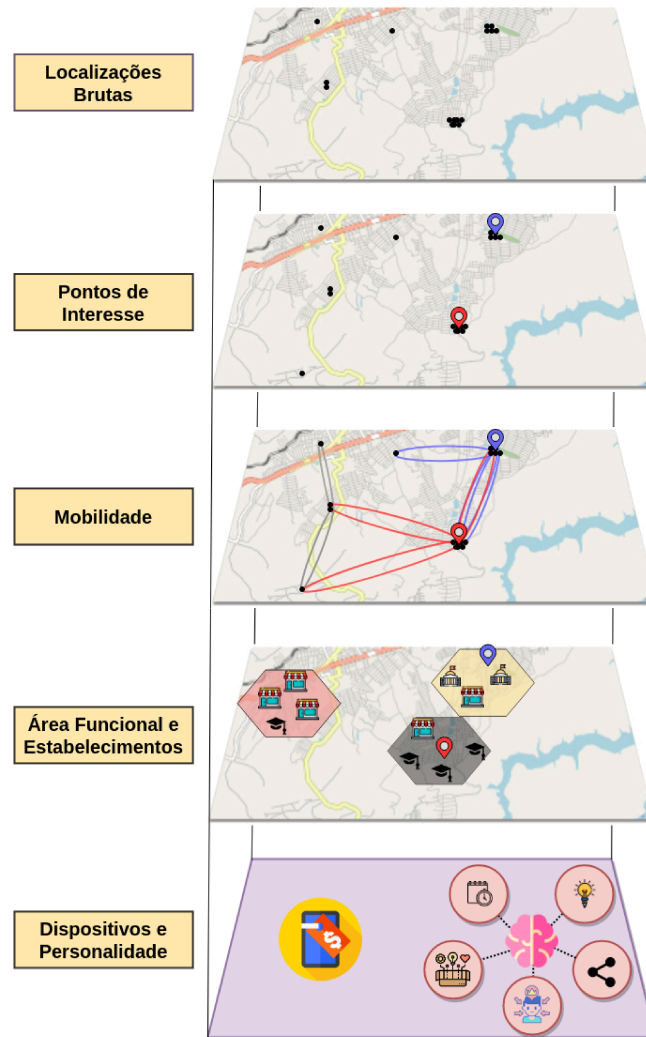


Figura 2.5: Ilustração da estrutura dos Perfis

de tipos de entrada (numérica, categórica, entre outros) na previsão. No trabalho de [Malmi and Weber \(2016\)](#) também utilizam atributos demográficos e lista de aplicativos, porém para criar um perfil do usuário considerando, além do gênero, atributos como faixa de idade e raça. Dessa forma, utilizaram estes dados como entrada para diferentes algoritmos como SVM, Floresta Aleatória e Regressão Logística para obter os melhores resultados. Além disso, enriqueceram os aplicativos com suas respectivas categorias da *Play Store* para realizar diferentes testes, mas não obtiveram melhora em relação ao uso direto da lista de aplicativos.

[Varlamis et al. \(2020\)](#) e [Sardianos et al. \(2018\)](#) apresentam trabalhos visando resolver o problema de melhoramento de rota e detecção de pontos de interesse. Para isso, em ambos são feitas as análises dos locais onde há uma concentração de acessos do usuário, sendo que para coletar as informações dos locais visitados realizam uma engenharia reversa com o *Open Street Map* (utilizam os dados do usuário para enri-

quecer com o tipo de local em que ele se encontra). Dessa maneira, eles conseguem criar um perfil dos tipos de locais em que o usuário visita e em quais rotas ele trafega.

Do and Gatica-Perez (2010) coletam dados de 111 voluntários para identificar e prever padrões do uso de cinco tipos de aplicativos (chamada de voz, envio de mensagem, internet, câmera e galeria). Já Alvarez-Lozano et al. (2014) trabalham com as correlações entre quatro categorias de aplicativos (entretenimento, social, estilo de vida e ferramenta) e distúrbios de bipolaridade de 18 pacientes de uma clínica. Com estes trabalhos é possível verificar a variedade de perfis que podem ser criados para um melhor entendimento do comportamento do usuário.

Há na literatura diversos trabalhos que analisam algum tipo de perfil do usuário, como, por exemplo, análise de laços sociais Li et al. (2008) e uso de redes sociais Abel et al. (2011). Dessa maneira, é possível analisar o comportamento do usuário em diferentes perspectivas para uma melhor tomada de decisão. Assim, é possível que empresas melhorem seus serviços e os clientes sejam beneficiados com produtos mais personalizados. Já a comunidade científica poderá avançar o estado da arte na área de comportamento humano. Por isso, esta dissertação apresenta a criação de um perfil que considera cinco perspectivas diferentes para melhor compreensão dos usuários.

# Capítulo 3

## Ferramentas

Neste capítulo são apresentadas as ferramentas desenvolvidas para auxiliar este trabalho.

### 3.1 *MoreData*: Um *Framework* para Enriquecimento de Dados

Nos últimos anos, tem-se percebido um crescimento considerável na coleta e disponibilidade de dados geospaciais (Naboulsi et al., 2016). Empresas e a comunidade científica estão percebendo a importância desse tipo de dado, e criaram iniciativas para que esses dados cheguem ao alcance de quem possa utilizá-los para ajudar nas tomadas de decisão.

Dados geospaciais são úteis em diferentes cenários (Barbosa et al., 2018). No contexto das cidades inteligentes, é possível entender a mobilidade urbana para ajudar na elaboração de políticas públicas. No contexto de redes sociais, padrões de comportamento geolocalizado podem ser explorados. No comércio eletrônico, as empresas podem compreender melhor o interesse de seus clientes de acordo com as características das regiões em que moram. Em linha geral, com dados geospaciais é possível extrair conhecimento para entender melhor os interesses de usuários móveis de diferentes seguimentos de mercado.

No entanto, os dados brutos originais, na maioria das vezes em forma de uma tupla <latitude, longitude, *timestamp*>, podem não agregar valor ao problema de negócio para o qual foram coletados. Por isso, é importante que os dados brutos sejam enriquecidos com fontes externas, muitas vezes de diferentes padrões e formatos, para que a informação semântica de uma localização seja conhecida. Por exemplo, é muito mais útil saber que um usuário móvel visita frequentemente um Shopping, do que apenas agrupar localizações sem o conhecimento sobre a semântica das mesmas. Porém, o enriquecimento de dados geospaciais é uma tarefa custosa, muitas vezes manual, que requer conhecimento específico, e dependente do formato das fontes externas.

Para sanar essa dificuldade, foi desenvolvido um *framework* flexível para que seja

possível realizar o enriquecimento semântico de diferentes fontes, além de ser facilmente expansível. O *framework*, chamado *MoreData*<sup>1</sup>, permite que diferentes formatos e diferentes fontes de dados sejam facilmente integrados a dados brutos originais, enriquecendo-os com novas informações. Com isso, a partir de dados de localização representados por uma tupla <latitude, longitude> ou por uma chave de identificação qualquer (e.g., CEP), o framework atua no casamento dos dados de entrada com dados de outras fontes, como, por exemplo dados do IBGE ou de alguma API disponível, de acordo com o interesse do usuário.

O framework já tem disponibilizado conectores para dados oriundos de APIs, banco de dados relacional, banco de dados de consulta rápida (i.e., *Elasticsearch*<sup>2</sup>) e para a base colaborativa do *OpenStreetMap*.<sup>3</sup> Além disso, o mesmo é facilmente expansível para que novos conectores sejam integrados pelo usuário de acordo com a demanda.

O *MoreData* foi projetado inicialmente para o enriquecimento de dados geoespaciais pela complexidade da operação. Contudo, também é possível utilizá-lo para enriquecer outros tipos de dados (e.g., listas de aplicativos). Dessa forma, o framework ajudou a reduzir o tempo de uma das tarefas mais custosas na área de ciência dos dados, que é a preparação dos dados.

Esta seção está organizada da seguinte forma. Na Subseção 3.1.1 são apresentados os trabalhos relacionados. O framework proposto é melhor descrito na Subseção 3.1.2. Por fim, na Subseção 3.1.3 são apresentados dois exemplos utilizando o *MoreData*.

### 3.1.1 Trabalhos Relacionados

Os avanços tecnológicos permitiram uma coleta mais fácil das mais diversas fontes de dados, fazendo com que novos conhecimentos fossem extraídos. Até o momento, não foram encontrados trabalhos que auxiliem no enriquecimento de dados geoespaciais oriundos de diferentes fontes. Portanto, os trabalhos descritos a seguir, listados na Tabela 3.1 se referem a projetos que precisaram de fazer algum tipo de enriquecimento, e que os responsáveis tiveram que investir um tempo significativo para isso. Ou seja, esses trabalhos e outros existentes se beneficiariam com o *MoreData*.

Li et al. (2008) apresentaram um *framework* para extrair informações dos logs de aparelhos GPS para identificar o nível de relacionamento (familiar, amizade, estranho na mesma cidade, estranho em outra cidade) entre os usuários de acordo com as informações espaço-temporais. Hung et al. (2009) também apresentam um *framework*, mas utilizando dados de uma aplicação Web (CarWeb) para descobrir comunidades dentre uma população de usuários, onde cada comunidade é a representação de vários usuá-

---

<sup>1</sup><https://github.com/gegen07/more-data>

<sup>2</sup><http://elastic.co/>

<sup>3</sup><https://www.openstreetmap.org/>

Tabela 3.1: Trabalhos Relacionados ao *Framework*

Artigo	Problema	Fonte	Formato	Volume
Li et al. (2008)	Laços Sociais e Pontos de Interesse	Log de GPS	-	65 usuários
Hung et al. (2009)	Identificação de Trajetória e Laços Sociais	CarWeb	-	-
Abel et al. (2011)	Traçar Perfil	Twitter API	REST (XML, JSON, RSS ou ATOM)	Aproximadamente 45.000 usuários
Farrahi and Gatica-Perez (2011)	Pontos de Interesse	Smartphone	-	97 usuários
McInerney et al. (2013)	Identificação de Trajetória e Pontos de Interesse	Nokia Lausanne	-	38 usuários
Maynard and Bontcheva (2015)	Mudanças Climáticas	Redes Sociais (Twitter) e Reegle	XML e RDF	-
Wu et al. (2015)	Pontos de Interesse	Twitter API	REST (XML, JSON, RSS ou ATOM)	-
Sardianos et al. (2018)	Melhoramento de Trajetória e Pontos de Interesse	Log Google Maps e OpenStreet-Map	KML e XML	1 usuário
Varlamis et al. (2020)	Melhoramento de Trajetória e Pontos de Interesse	Log Google Maps e OpenStreet-Map	KML e XML	1 usuário

rios que possuem rotas em comum, frequentando lugares semelhantes. Nestes dois estudos, os *frameworks* foram desenvolvidos para auxiliar em problemas específicos, ou seja, foram despendidos tempo e esforço nas tarefas de extração e enriquecimento dos dados. Já com a utilização do *MoreData*, os envolvidos poderiam se beneficiar na agilização destas etapas.

Os trabalhos Varlamis et al. (2020) e Sardianos et al. (2018) abordam os problemas de melhoramento de rota e detecção de pontos de interesse. Nestes dois trabalhos, foram utilizados os formatos KML *Keyhole Markup Language* (KML) e *eXtensible Markup Language* (XML) referentes aos dados de localizações de acessos dos usuários e dos locais do *OpenStreetMap*. No trabalho de Maynard and Bontcheva (2015) é utilizada parte dos dados em XML, mas também em *Resource Description Framework* (RDF), pro-

venientes de uma ontologia, para desenvolverem um *toolkit* que busca extrair, agregar e visualizar conhecimentos baseados em uma análise automática de dados de mídias sociais para extraírem os termos chave e indicadores das mudanças climáticas. Com isso, é possível perceber que os dados são disponibilizados em diferentes fontes e formatos, assim as etapas de coleta e agregação dos dados podem ser diretamente impactadas caso haja uma mudança na fonte dos dados, causando retrabalho. Isto poderia ser evitado com a utilização do conversor e dos conectores do *MoreData*.

Wu et al. (2015) utilizam quase 38 milhões de *tweets* para relacionar o local naquela hora a uma série de comentários que também estão relacionados àquele mesmo local e hora, sendo possível minerar termos importantes de todos os documentos e anotar o propósito daquela visita. Com isso, além de conhecerem o local de interesse do usuário, descobrem o motivo da visita. Abel et al. (2011) utilizam mais de 3 milhões de *tweets* para relacioná-los com artigos de notícias, e assim, correlacionar um usuário com tópicos que ele normalmente comenta, criando um perfil. Estes trabalhos não especificam as técnicas utilizadas para enriquecer, mas o grande volume de dados pode impactar diretamente no custo dessa tarefa. Dessa forma, o *MoreData* poderia agilizar este processo utilizando um dos seus conectores já desenvolvidos, como, por exemplo, o *Elasticsearch*.

Até o momento, muitos trabalhos da literatura fazem uso de dados de fontes externas, incluindo geoespaciais, para realizarem diferentes tipos de análises, como identificar trajetórias e laços sociais (McInerney et al., 2013) e detectar pontos de interesse (Farrahi and Gatica-Perez, 2011). Com isso, os envolvidos em cada um desses trabalhos precisaram tratar e enriquecer seus dados originais de forma manual e particular. Não foi encontrado nenhum trabalho que preencha essa lacuna entre os dados de fontes externas (com diferentes formatos e tipos de acesso) com os dados brutos originais. Por isso, este trabalho propõe um *framework* capaz de estabelecer conectores para padronizar a utilização de diferentes fontes e formatos de dados. Além disso, o *framework* foi desenvolvido para ser facilmente extensível, garantindo assim seu funcionamento para diversas condições.

### 3.1.2 Arquitetura

A ideia principal do *framework* proposto é unir dados, que podem ser georreferenciados, a partir da relação entre seus atributos. O *framework* é composto por dois módulos principais: o enriquecedor e o conversor. A Figura 3.1 apresenta uma visão geral dos seus componentes, descritos a seguir.

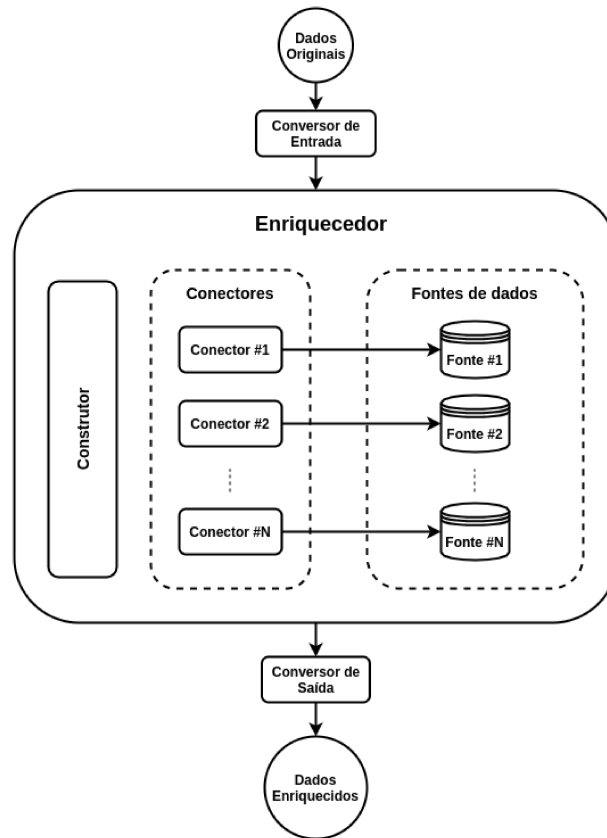


Figura 3.1: Visão geral dos componentes do Framework

### Enriquecedor

O enriquecedor é o principal ator do *framework*. Ele é composto por dois componentes principais: conector, responsável por fazer a conexão com uma fonte externa de dados; e construtor, responsável por gerar o enriquecimento. A versão atual já possui quatro conectores implementados: um para utilizar o *Elasticsearch*, ferramenta de busca rápida; outro conector que acessa uma fonte de dados por meio de uma *Application Programming Interface (API)*; um conector que utiliza dados do *OpenStreetMap (OSM)*; além desses, um conector para utilizar bancos de dados relacionais. Cada conector deve implementar a interface **IEnricherConnector**, como mostra a Figura 3.2.

Essa modelagem do Enriquecedor, utilizando o padrão de projeto *Strategy*, permite uma alta extensibilidade do *framework*, sendo possível criar outros conectores que implementem a interface e utilizá-los na classe **Enricher**, criando assim novas conexões com diversas fontes de dados de acordo com a necessidade do usuário.

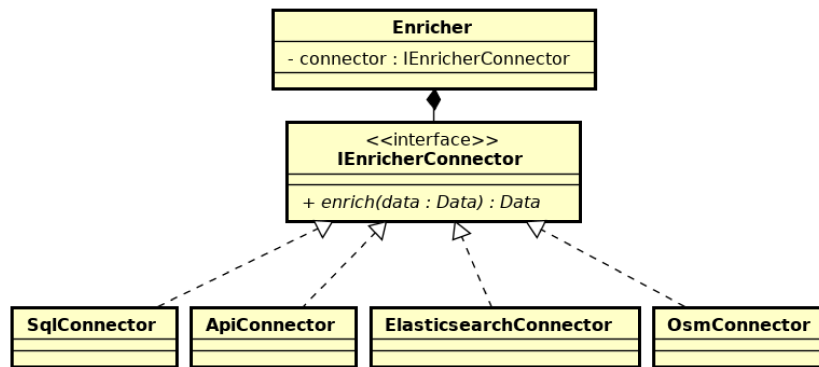


Figura 3.2: Implementação do módulo Enriquecedor

### Conector Elasticsearch

O *Elasticsearch*<sup>4</sup> é uma base de busca rápida que vem sendo muito utilizado nos últimos anos. O conector criado utiliza duas estruturas *JavaScript Object Notation* (JSON): *pipeline*, responsável por definir qual atributo do índice será enriquecido e qual será utilizado para fazer a relação; e *policy*, responsável por definir qual campo será utilizado para enriquecer um índice.

Durante a criação das estruturas *pipeline* e *policy*, é possível definir qual relacionamento entre os índices será utilizado para fazer o enriquecimento, sendo consulta geoespacial ou consulta simples. O primeiro caso oferece a possibilidade de utilizar 4 operações, oferecidas pelo *Elasticsearch*, entre as formas geométricas: interseção, disjunção, inclusão e pertinência; já o segundo caso permite uma consulta de correspondência entre chaves identificadoras de índices diferentes.

Após a definição de quais campos, relações e qual tipo de *query* serão utilizados para o enriquecimento, é executado o processo de reindexação utilizando o *pipeline* criado.

### Conector API

Esse conector executa requisições *HyperText Transfer Protocol* (HTTP) a partir de uma *Uniform Resource Identifier* (URI) definida pelo usuário a fim de coletar informações necessárias para a associação dos dados da API aos dados originais. Através de uma *cache* desenvolvida, o conector guarda valores que já foram requisitados, acelerando o processo de enriquecimento.

O processamento da informação retornada pela API deve ser implementado pelo usuário, possibilitando que diferentes tratamentos sejam realizados sobre o que enriquecer e como enriquecer, e quais campos serão criados para guardar o resultado.

<sup>4</sup><https://www.elastic.co>

## Conector OSM

O conector OSM utiliza dados provenientes da API do *OpenStreetMap*. Também há a opção de o usuário utilizar dados previamente já coletados, sem a necessidade da coleta dos dados na base OSM por parte do conector. Após o processo de obtenção dos dados, o conector indexa localmente os elementos retornados em uma *Árvore R* com a finalidade de melhorar o tempo da pesquisa georreferenciada.

Essa pesquisa pode ser feita de duas maneiras: a partir de um raio fornecido pelo usuário ou pela relação de pertinência entre uma localização e os locais do OSM representados como polígonos. No primeiro modo é feita uma mudança de projeção, do Sistema Geodésico Mundial para a Projeção Azimutal, para uma melhor representação do círculo resultante e utilização do metro como unidade de comprimento. Após a mudança da projeção e a transformação do ponto em um círculo, o conector enriquece a fonte original com o polígono intersectado pelo círculo e seus dados.

Já na segunda maneira, é realizada uma verificação da relação de pertinência entre o polígono coletado do OSM e uma localização. Após a verificação, o conector enriquece a fonte de dados original com o polígono correspondente e seus atributos.

## Conector Banco Relacional

Esse conector realiza uma conexão com bancos relacionais. O conector oferece algumas possibilidades de uso por meio da biblioteca *SQLAlchemy*<sup>5</sup>, permitindo o uso de diferentes sistemas como *MySQL*, *PostgreSQL*, *SQLite*, entre outros.

Após a conexão com o banco, é feita uma consulta simples entre uma tabela que possui os dados, que serão utilizados para o enriquecimento, e a fonte de dados original. Essa consulta é realizada relacionando uma coluna, que pertence à tabela, definida pelo usuário e um atributo da fonte original. Após a consulta, o conector gera um novo atributo, com um nome também definido pelo usuário, na fonte original, contendo os dados do banco de dados utilizado.

## Construtor

O construtor implementa o padrão de projeto *Builder* para utilizar os diversos conectores criados através da classe **Enricher** e compor o enriquecimento dos dados utilizando a classe **EnricherBuilder**, como mostra a Figura 3.3. Utilizando o padrão mostrado é possível enriquecer com diferentes conectores, possibilitando uma combinação na utilização dessas estruturas.

A classe **Data** desempenha o papel de modelar genericamente os dados originais que serão enriquecidos. Portanto, possui atributos que descrevem ao *framework* me-

---

<sup>5</sup><https://www.sqlalchemy.org/>

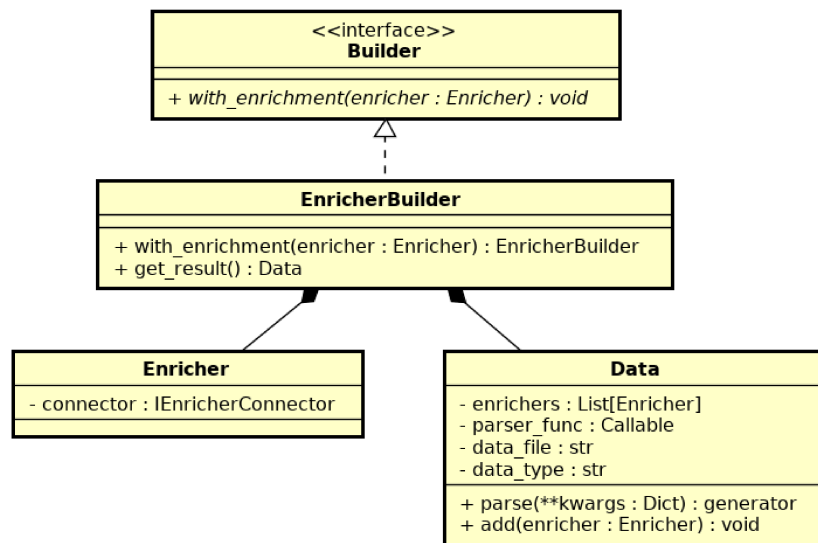


Figura 3.3: Implementação do módulo Construtor

tadados como o tipo de arquivo utilizado e onde se encontra o dado. Além disso, possui um atributo do tipo função, responsável por fazer o *parser* dos dados. O *MoreData* desempenha essa função genericamente. No entanto, esse atributo permite ao usuário criar sua própria função e analisar seus dados de uma forma específica antes de enriquecê-los.

### Conversor

O *MoreData* utiliza o formato JSON ao longo do enriquecimento, visionando a manutenibilidade. No entanto, a partir desse módulo é possível fornecer como entrada ao *framework* arquivos com outros formatos, como CSV e Parquet. Além disso, esse módulo oferece opções de saída podendo ser dos tipos CSV, JSON e Parquet. Vale destacar que o conversor também é facilmente extensível para novos formatos. Ademais, para abranger arquivos de tamanhos diversos, o conversor ainda possui a opção de separar a saída em múltiplos arquivos menores.

### 3.1.3 Exemplo

Para mostrar como é feito o enriquecimento utilizando o *MoreData* a seguir são mostrados dois exemplos. No primeiro deles, é utilizado o módulo *ApiConnector* para enriquecer uma base inicial com dados provenientes da API IBGE Cidades. Já o outro exemplo, enriquece os pontos de interesse dos usuários com dados de estabelecimentos do tipo *Fitness Centre* (e.g., academias) do OSM que estão em um raio de 100 metros. Assim, utilizando o módulo *OSMPlacesConnector*, é simplificado o enriquecimento semântico da base de usuários.

- Exemplo 1: Dados da API IBGE Cidades

```

import os
import sys
sys.path.insert(0, os.path.abspath('../..'))

from enrichment.enricher import EnricherBuilder, Enricher
from enrichment.enricher.api_connector import ApiConnector
from enrichment.models.data import Data
from enrichment.parser import parse_document
from enrichment.utils.util import read_json_from_file, Converter

DATASETS_DIR = "../.././././datasets/"

CIDADES_DIR_CSV = DATASETS_DIR + "cidades_info_polygon.csv"
CIDADES_DIR_JSON = "../../././data/output/conversion/cidades_info_polygon.
                    json"

URL_PATTERN = r'https://servicodados.ibge.gov.br/api/v1/pesquisas/-/
                indicadores/47001/resultados/{
                localidade}'

PARAMETERS = {
    "fields": [
        {
            "key": "localidade",
            "name": "Codigo do Municipio"
        }
    ]
}

if __name__ == "__main__":
    def response_parser(response):
        for res in response:
            for res in res["res"]:
                return {
                    "pib_per_capita": res["res"]["2017"]
                }

    # Converter.csv_to_json(CIDADES_DIR_CSV, CIDADES_DIR_JSON)
    cidades = Data(data_file=CIDADES_DIR_JSON, parser_func=parse_document,
                   data_type="csv")

    api_ibge_enricher = Enricher(connector=ApiConnector(response_parser=
                                                         response_parser, url_pattern=
                                                         URL_PATTERN, params=PARAMETERS))

    cidades_enriched = \

```

```

    EnricherBuilder(cidades) \
        .with_enrichment(api_ibge_enricher) \
        .get_result()

import enrichment.utils.util as util
util.write_json_generator_to_json("../data/output/cidades/json/
                                cidades-enriched", cidades_enriched,
                                1024)
util.Converter.json_enriched_to_csv("../data/output/cidades/json/*.
                                    json", "../data/output/cidades/
                                    csv/")

```

- Exemplo 2: Coleta de Estabelecimentos

```

import os
import sys
sys.path.insert(0, os.path.abspath('../..'))

from enrichment.enricher import EnricherBuilder, Enricher
from enrichment.enricher.osm.osm_places_connector import
    OSMPlacesConnector
from enrichment.models.data import Data
from enrichment.parser import parse_document
from enrichment.utils.util import read_json_from_file, Converter

DATASETS_DIR = "../../datasets/"
FITNESS_CENTRE = DATASETS_DIR + "Locais_OSM/csv/leisure-fitness_centre.
                                csv"

USER_DATA = "../../datasets/user_profile_17092019_preprocessed.json"

if __name__ == "__main__":

    user = Data(data_file=USER_DATA, parser_func=parse_document,
                data_type="json")

    osm_enricher = Enricher(connector=OSMPlacesConnector(key="leisure",
                                                            value="fitness_centre", file=
FITNESS_CENTRE, radius=100,
                                                            dict_keys=["points_of_interest"]
                                                            ))

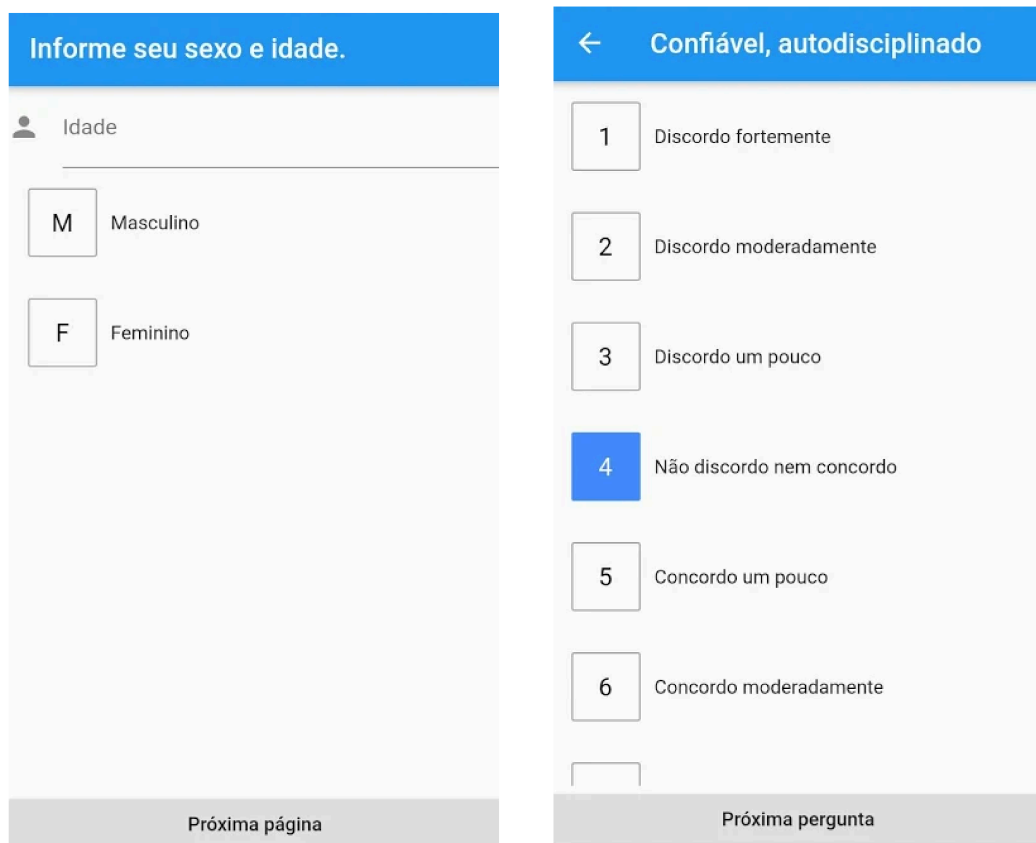
    user_enriched = \
        EnricherBuilder(user) \
            .with_enrichment(osm_enricher) \
            .get_result()

```



## 3.2 Aplicativo *APProfile*

Para obter a base de dados necessária para a análise dos traços de personalidade, foi desenvolvido um aplicativo, denominado *APProfile*, para dispositivos com sistema operacional *Android* usando as tecnologias *Flutter* e *Java*. Neste aplicativo, o usuário informa sua idade, sexo e responde aos dez itens do formulário TIPI (Seção 2.7), e ao final, suas respostas são enviadas para um banco de dados junto com a lista de aplicativos instalados no aparelho móvel do usuário. Por fim, os dados dos usuários são armazenados no *Firebase Realtime Database*, banco de dados hospedado na nuvem, onde é possível gerar um arquivo *JSON* contendo todos os dados coletados de cada usuário voluntário.



(a) Tela inicial para o preenchimento de Idade e Sexo (b) Tela do *Item<sub>3</sub>* do questionário TIPI

Figura 3.4: Exemplos das telas do *APProfile*

Algumas medidas foram tomadas para impedir ou eliminar possíveis ruídos nos dados. Para evitar que usuários respondessem mais de uma vez, é permitido o re-envio das respostas, porém elas são sobrescritas no banco de dados. Além disso, também foram coletadas a hora em que o usuário começa a responder o questionário e a hora em que envia os dados para verificar a necessidade de remoção de usuários que possam ter respondido aleatoriamente.

# Capítulo 4

## Materiais e Métodos

A seguir são apresentados os dados utilizados como base para a criação dos perfis, exceto o perfil de traços de personalidade que terá todos os seus dados apresentados em sua seção.

### 4.1 Dados Base

Os dados utilizados como base para a criação dos perfis são as listas de aplicativos e as localizações de acessos dos usuários. A exceção são os dados para o perfil dos traços de personalidade que serão apresentados na seção correspondente. Isso porque, os traços de personalidade são dependentes das respostas dos usuários no questionário TIPI, assim o número de usuários para este perfil é restrito.

Inicialmente, a lista de aplicativos do usuário é enriquecida com as respectivas categorias para contabilizar a quantidade de aplicativos de cada categoria que ele possui. As localizações dos usuários são utilizadas para determinar seus pontos de interesse e seus respectivos tipos. Por fim, o ponto de interesse do tipo *Principal* é utilizado para determinar as classes de população e PIB/capita da cidade do usuário.

Na Tabela 4.1 é apresentado um exemplo dos dados fictícios de três usuários após os enriquecimentos iniciais. Cada usuário possui a quantidade de aplicativos instalados de cada categoria, as categorias da cidade do usuário em termos do IBGE e a lista de PoIs. Além disso, na tabela também é possível ver que os usuários podem possuir diferentes quantidades de pontos de interesse.

Abaixo são especificados cada um dos tipos de dados utilizados como base.

#### 4.1.1 Base de Usuários

Foram utilizados como base para o enriquecimento, dados reais de 46.439 usuários móveis do sistema operacional *Android*, obtidos sob confidencialidade de uma empresa parceira da área de serviços móveis. Essa base foi coletada entre os dias 01/11/2019 e 31/01/2020. Dentre os dados coletados, os utilizados neste trabalho foram: lista de aplicativos instalados, as localizações geradas pelos acessos ao aplicativo da empresa parceira e o modelo do aparelho. A lista de aplicativos é, simplesmente,

Tabela 4.1: Exemplificação dos dados para três usuários fictícios

(a) Dados das quantidades de aplicativos de cada categoria e as categorias do IBGE

Identificador Usuário	Categoria de Aplicativo						Categoria do IBGE	
	C1	C2	C3	...	C37	População	PIB/capita	
0001	5	3	10	...	0	Pequena	Baixo	
0002	4	8	1	...	2	Grande	Médio	
0003	0	2	3	...	0	Média	Alto	

(b) Pontos de Interesses dos Usuários

Identificador do Usuário	Latitude	Longitude	Tipo
0001	-29.167028	-51.168350	<i>Principal</i>
0001	-05.907247	-35.264202	<i>Casa</i>
0002	-07.028732	-35.861667	<i>Principal</i>
0002	-10.190794	-48.310523	<i>Outro</i>
0002	-22.671720	-43.255123	<i>Casa</i>
0002	-22.832166	-47.163181	<i>Trabalho</i>
0002	-25.597251	-49.334327	<i>Outro</i>
0003	-15.603462	-56.111846	<i>Principal</i>

uma lista de identificadores dos aplicativos que estavam instalados no aparelho do usuário. Já as localizações foram utilizadas para aplicar o algoritmo proposto por [Capanema et al. \(2019\)](#) e encontrar os pontos de interesse (PoIs) do usuário (explicado mais detalhadamente na Seção 2.2), totalizando 178.877 PoIs distintos considerando todos os usuários. Após a aplicação do algoritmo de [Capanema et al. \(2019\)](#), foram removidos os usuários sem nenhum PoI, e o número total desses reduziu para 34.428. Na Figura 4.1 é mostrada a distribuição da porcentagem de usuários por quantidade de pontos de interesse.

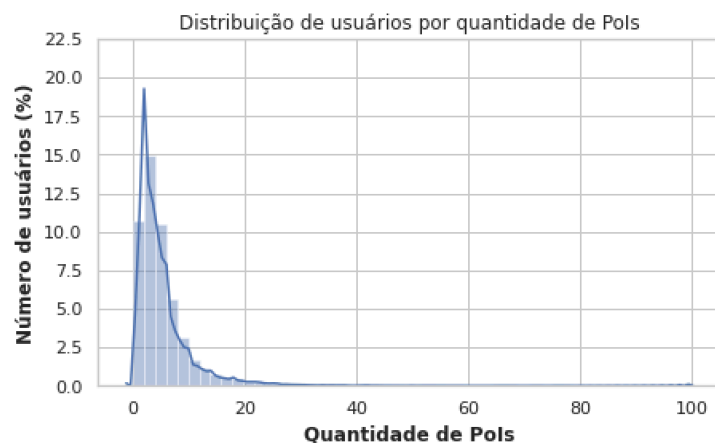


Figura 4.1: Distribuição de usuários por quantidade de PoIs. A distribuição é de cauda longa onde há usuários com cerca de 100 pontos de interesse.

### 4.1.2 Base de Categorias de Aplicativos

Esta base foi disponibilizada pela empresa parceira no qual desenvolveram um *Crawler* para coletar as informações dos aplicativos disponíveis na *Google Play Store* (loja de aplicativos *Android*). Nesta base, muitas informações dos aplicativos são fornecidas pelo próprio desenvolvedor, sendo a categoria uma delas. Com isso, muitas vezes, não há um padrão de categoria para aplicativos de mesma funcionalidade. Um exemplo disso é que alguns aplicativos de jogos esportivos estão categorizados como *Esporte* ao invés de *Jogos*. O problema disso é que não há como afirmar que um usuário tem interesse na prática de esportes, já que ele pode simplesmente ter um jogo deste tema. Além disso, algumas categorias são genéricas e não representam de fato as reais categorias dos aplicativos. Levando tudo isso em consideração, para a realização deste trabalho foram criadas 37 categorias específicas que agrupam de maneira mais consistente 533 dos principais aplicativos disponíveis (Apêndice A). Por fim, esta base foi armazenada em um banco relacional MySQL.

Após inserir os dados das categorias no MySQL, utilizando o conector *SQL* do *MoreData*, os enriquecimentos das listas de aplicativos dos usuários foram feitos a partir de uma consulta simples. Para isso, foram relacionadas a lista de identificadores de aplicativos que cada usuário possui ao identificador correspondente da base de aplicativo. Resultou-se, portanto, em uma lista que continha a categoria de cada aplicativo do usuário. Na Figura 4.2 é possível ver as proporções de usuários que possuía pelo menos um dos aplicativos de cada categoria.

### 4.1.3 Base de Cidades

Por fim, a base original foi enriquecida, utilizando o conector *API* do *MoreData*, com informações dos municípios do Brasil usando a consulta geoespacial. Para isso, foram relacionados os pontos de interesses do usuário e a geometria especificada na base do IBGE. Portanto, foi adicionado um campo contendo os atributos sobre a cidade resultante da consulta geoespacial feita pelo conector. Após este enriquecimento, como haviam PoIs que não estavam localizados no Brasil, o número final de usuários foi reduzido a 32.466.

Para este trabalho foram escolhidos somente o PIB/capita e o tamanho da população porque foram considerados os mais relevantes para se caracterizar o perfil de um usuário em termos de seus locais de interesse. Uma exemplificação disto é que, possivelmente, não faria sentido classificar um usuário que more em um município com baixo PIB/capita e pequena população como interessado em Shopping, já que nestas condições dificilmente há este tipo de estabelecimento no local. Em contrapartida, as outras informações dos municípios, como se é semiárido, participante do Amazônia Legal, dentre outras, não indicam as condições do local em que a pessoa frequenta.

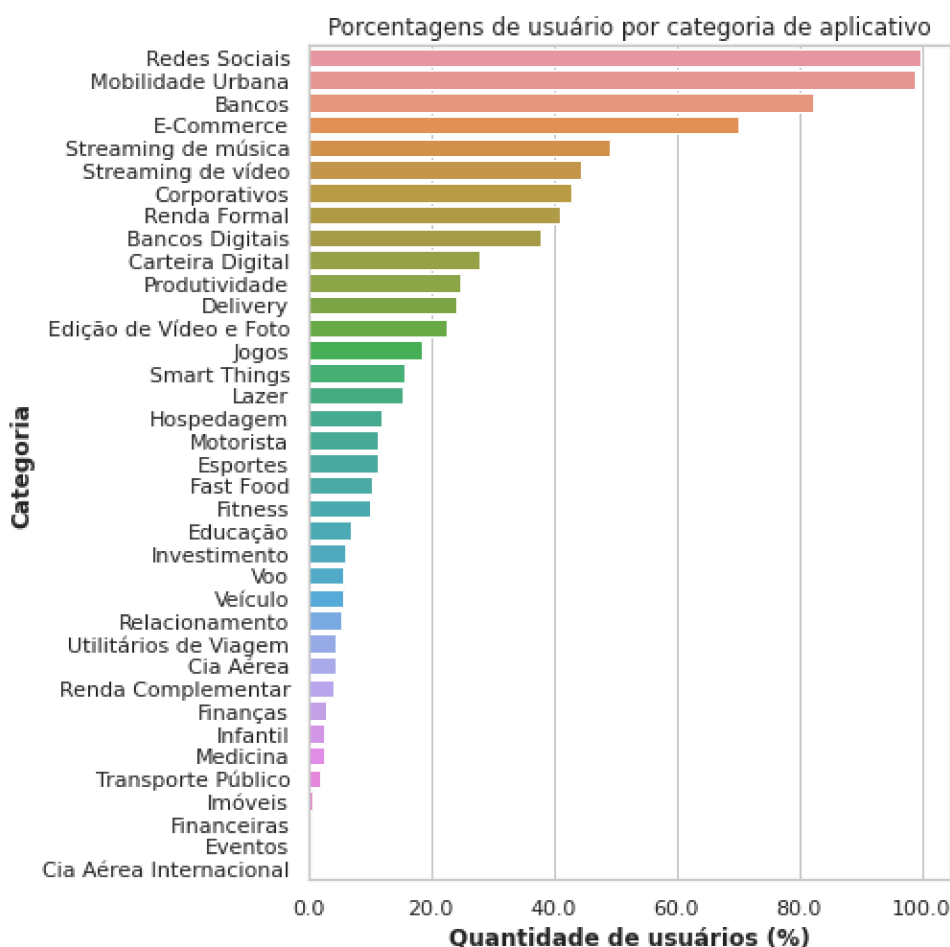


Figura 4.2: Proporção de usuários que possui instalado pelo menos um aplicativo da categoria. Algumas categorias, como, por exemplo, redes sociais, possuem aplicativos que podem vir de fábrica (ex. *Youtube*) ocasionando altos valores mesmo que o usuário não os utilizem.

Como os usuários possuem diferentes tipos de PoIs, foram considerados os valores do IBGE do seu PoI principal (que o usuário mais frequenta). Com estes valores, cada um dos dois atributos foram divididos em três classes. A população foi dividida de acordo com uma adaptação dos valores fornecidos pelo Ipea (2020), dessa maneira as classes de população estão distribuídas da seguinte forma:

- Classe Pequena: População  $\leq 100.000$
- Classe Média:  $100.000 < \text{População} \leq 500.000$
- Classe Grande: População  $> 500.000$

Na Figura 4.3 é possível ver que as classes estão bem distribuídas, mas que o número de usuários de cidade pequena é o maior. Já as classes para o PIB/capita, Figura 4.4, foram calculadas considerando os valores dos *quartis* e limite superior para os municípios. Os valores dos quartis Q1, Q2 e Q3 são, respectivamente R\$

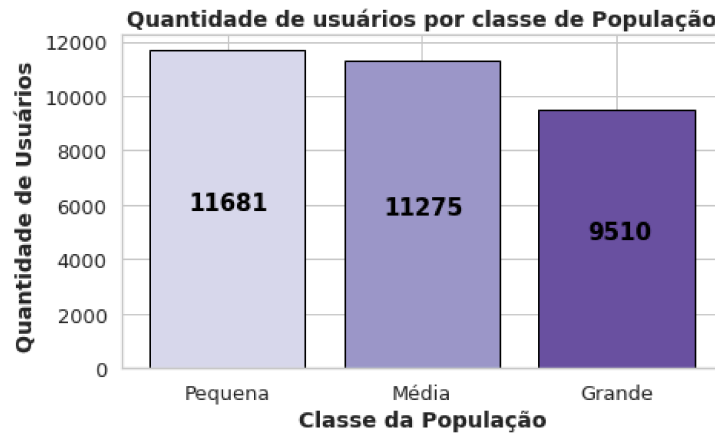


Figura 4.3: Quantidade de usuários por classe de população.

13.315,23, R\$ 21.839,10 e R\$ 34.232,97. Assim, o limite superior  $LS$  foi calculado da seguinte forma:

$$LS = Q3 + 1,5 * (Q3 - Q1) = R\$65.609,58 \quad (4.1)$$

Com isso, as classes do PIB/capita possuem a seguinte configuração:

- Classe Baixo: PIB *per capita* < R\$ 21.839,10
- Classe Médio: R\$ 21.839,10 ≤ PIB *per capita* ≤ R\$ 65.609,58
- Classe Alto: PIB *per capita* > R\$ 65.609,58

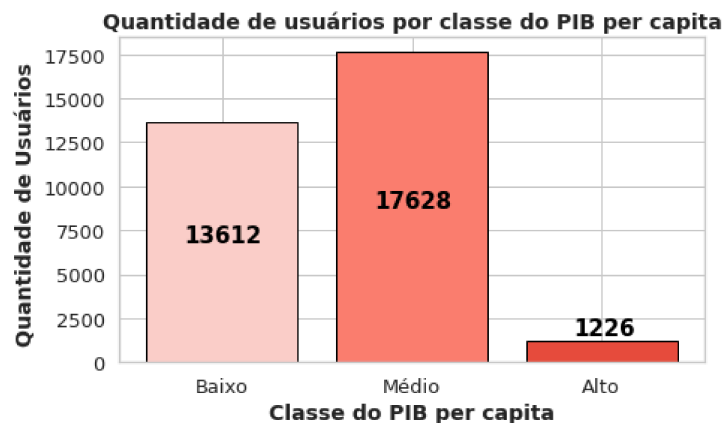


Figura 4.4: Quantidade de usuários por classe de PIB *per capita*

## 4.2 Modelos Preditivos

Foram criados modelos preditivos para inferir o perfil de um usuário em cada uma das cinco perspectivas. Dessa maneira, para cada classe do perfil foram criados mo-

delos binários diferentes, além de um modelo multi-classe que considera todas as classes do perfil. Ainda, foram geradas entradas para os modelos considerando uma variável *dummy* para as categorias de aplicativos, sendo 0 caso o usuário não possua nenhum aplicativo daquela categoria, e 1 caso contrário.

Afim de encontrar os melhores resultados, foram testados 4 diferentes algoritmos de classificação: Regressão Logística, Floresta Aleatória, *Naive Bayes* e *K-Nearest Neighbors*. Cada classificador foi testado com os dados em duas formas diferentes: Simples, onde nenhuma alteração é feita; e com aplicação de PCA (Análise de Componentes Principais), em que os componentes principais são usados como atributos. Além disto, os modelos foram testados com e sem os dados do IBGE, exceto para o Perfil de Personalidade, pois este não utiliza dados referentes à cidade do usuário.

Para corrigir o desbalanceamento das classes de entrada, foram testadas a aplicação ou não de técnicas de *Resampling*. Para evitar que todo o conjunto de técnicas aplicadas afete os resultados, o processo foi repetido 35 vezes para cada uma das combinações de entrada, sendo que a amostragem é diferente em cada uma delas. A seguir, o fluxo da criação de modelos é ilustrado na Figura 4.5 e cada etapa melhor descrita.

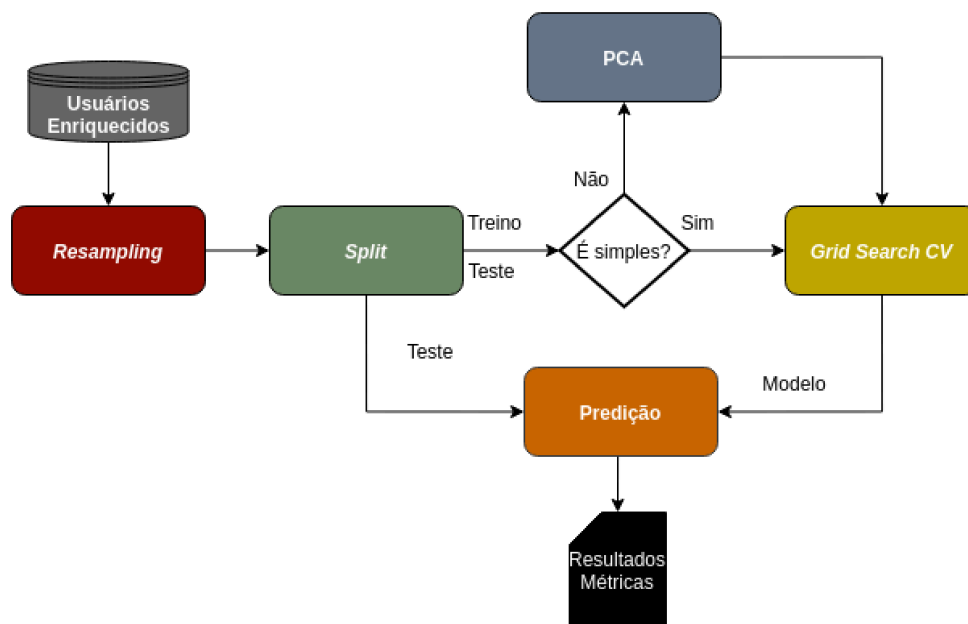


Figura 4.5: Fluxo da criação dos modelos.

### 4.2.1 *Resampling*

Na etapa de *Resampling* é realizado o balanceamento da entrada de dados. Foram consideradas três formas de *sampling*: *Undersampling*, *Oversampling* e o *No-sampling*. A primeira delas, o *Undersampling*, balanceia a entrada de dados de forma que remove itens da classe de maior quantidade. Já *Oversampling* adiciona itens na classe com

menor quantidade. Por fim, a *No-sampling* não faz qualquer alteração na entrada de dados, deixando-a desbalanceada.

Uma observação é quanto a comparação dos resultados entre as técnicas de *resampling*. A técnica de *Oversampling* obteve, na maioria das vezes, os melhores resultados se comparada com as demais. Contudo, ao utilizar desta técnica ocorreu *overfitting* dos modelos, ou seja, os modelos ficaram ruins para prever novos resultados. Isso porque, são feitas cópias de alguns itens para balancear a classe de menor quantidade. Por este motivo, os modelos gerados utilizando *oversampling* foram descartados.

### 4.2.2 *Split*

Esta etapa divide os dados em treino e teste, sendo correspondentes a 80% e 20%, respectivamente. Além disso, ela salva em arquivos temporários os conjuntos de treino e testes para que todos os classificadores executem com a mesma amostragem. Caso for selecionado o modo simples, o fluxo dos dados é direcionado diretamente para a etapa de *Grid Search CV*.

### 4.2.3 *PCA*

Após receber os dados de treino e teste da etapa *Split*, o PCA (Análise de Componentes Principais) é aplicado considerando o número de colunas da entrada. Dessa forma, aplicando o classificador nas componentes do PCA é selecionado o melhor número de componentes daquela iteração baseado na métrica *F-Score*. Assim, os conjuntos de treino e teste do melhor número de componente são enviados para a próxima etapa.

### 4.2.4 *Grid Search CV*

Após receber o conjunto de treino da etapa de *Split*, caso for simples, ou da etapa de PCA, o modelo final da iteração é gerado. Para verificar a generalização do modelo é aplicada a técnica de validação cruzada com 10 partições e também é feita a mixagem dos parâmetros do classificador. Após este processo, o melhor modelo daquela iteração é salvo.

### 4.2.5 *Predição*

Por fim, a etapa de predição recebe os dados de treino e teste, além do melhor modelo da etapa *Grid Search CV*. Caso seja aplicado PCA, o dado de teste é transformado de acordo com o número de componentes. Assim, as métricas de precisão, revocação e *F-Score* são aplicadas e seus resultados salvos.

### 4.3 Métricas

Como dito anteriormente, foram criados alguns modelos preditivos para cada um dos perfis, assim faz-se necessária a utilização de métricas para avaliação da qualidade destes modelos. Seja um perfil  $X$  qualquer que possui  $m$  classes, foram aplicadas algumas métricas que fazem referência aos seguintes conjuntos: sendo  $U$  o conjunto de todos os usuários, então  $R_i \subset U$  é um subconjunto de  $U$  com todos os usuários que realmente possuem relação com a classe  $i$  e  $P_i \subset U$  o subconjunto de  $U$  com todos os usuários que o modelo previu que possuíssem relação com a classe  $i$ , sendo  $\{i \in \mathbb{N} | 1 \leq i \leq m\}$ .

- Precisão: porcentagem do total de acertos em relação ao total de previsões da classe  $i$ .

$$Precisao_i = \frac{|R_i \cap P_i|}{|P_i|}$$

- Revocação: do total de usuários que pertencem à classe  $i$ , quantos foram previstos corretamente.

$$Revocacao_i = \frac{|R_i \cap P_i|}{|R_i|}$$

- F-Score: é o balanceamento através da média harmônica entre precisão e revocação da característica  $i$ .

$$F\_Score_i = 2 * \frac{Precisao_i * Revocacao_i}{Precisao_i + Revocacao_i}$$

Considerando também que são analisadas  $n$  categorias de aplicativos, logo  $A_j$  é o subconjunto de todos os usuários que possuem pelo menos 1 aplicativo da categoria  $j$  instalado, sendo  $\{j \in \mathbb{N} | 1 \leq j \leq n\}$ . Dessa forma, a intercessão  $R_i \cap A_j$  resulta no subconjunto contendo todos os usuários que possuem alguma relação com a classe  $i$  e também possui instalado pelo menos 1 aplicativo da categoria  $j$ .

- Suporte: indica a taxa de usuários relacionados com a classe  $i$  que possuem algum aplicativo da categoria  $j$ .

$$Sup_{ij} = \frac{|R_i \cap A_j|}{|R_i|}$$

- Lift (taxa de aumento): é o cálculo do ganho (ou perda, caso o valor for negativo) que se pode ter ao comparar amostras dos usuários que pertencem e não pertencem à classe  $i$ , e que possuem aplicativos da categoria  $j$ .

$$Lift_{ij} = \frac{\frac{|R_i \cap A_j|}{|R_i|} - \frac{|\overline{R_i} \cap A_j|}{|\overline{R_i}|}}{\frac{|\overline{R_i} \cap A_j|}{|\overline{R_i}|}}$$

# Capítulo 5

## Resultados e Discussões

Como apresentado no Capítulo 2, é possível explorar o perfil do usuário em diferentes perspectivas. Dessa forma, neste capítulo serão apresentados cinco perfis considerando as seguintes perspectivas: dispositivo, estabelecimento, área funcional, mobilidade e personalidade. Para cada um destes perfis foram utilizados dados reais de usuários que foram enriquecidos de diferentes formas utilizando o *MoreData* (descrito na Seção 3.1). Por fim, será apresentado um perfil completo, nomeado Perfil MAPED (Mobilidade, Área funcional, Personalidade, Estabelecimento e Dispositivo), que agrupa todos os cinco perfis.

### 5.1 Perfil de Dispositivo

Há atualmente diversos modelos de dispositivos móveis disponíveis no mercado. Logo, estes dispositivos apresentam diferentes capacidades de armazenamento e processamento o que impacta diretamente nos preços. Com isso, este perfil correlaciona a faixa de preço do dispositivo com as categorias de aplicativos e a cidade do usuário. Sendo assim, será possível avaliar se usuários com aparelhos mais caros possuem mais aplicativos de alguma categoria específica, por exemplo.

#### 5.1.1 Descrição dos Dados

Para este perfil, os usuários foram enriquecidos com uma base de dados que continha as informações dos modelos dos aparelhos móveis, totalizando 654 modelos distintos. Para a coleta desta base foi utilizado o *CrawlMobi* apresentado por [Maia et al. \(2020\)](#) que enriquece o nome técnico do modelo com 22 características (e.g., marca, capacidade de memória, tamanho da tela, preço). No escopo desta dissertação, foram utilizados somente os nomes usuais e preços dos modelos, sendo o primeiro usado unicamente para a caracterização dos dados. Com os dados já coletados, foi utilizado o conector *Elasticsearch* do *MoreData* para enriquecer os modelos dos usuários com os dados coletados. Após o enriquecimento, o número total de usuários reduziu para 32.451, pois alguns modelos não estavam contidos na base. A Figura 5.1 apresenta as percentagens dos 30 modelos mais utilizados pelos usuários.

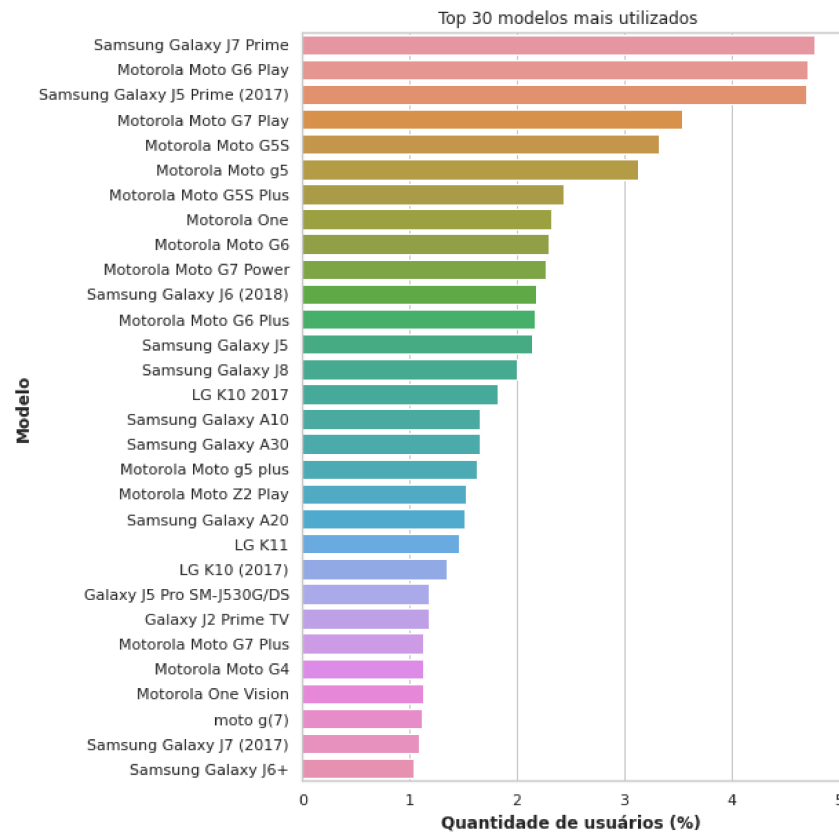


Figura 5.1: Os 30 modelos de aparelhos móveis mais utilizados.

De acordo com a adaptação das classes dadas por [Medeiros \(2019\)](#), os preços dos aparelhos são divididos em cinco classes distribuídas da seguinte forma:

- Classe *Entrada*: Preço < R\$ 700,00
- Classe *Intermediário*: R\$ 700,00 ≤ Preço ≤ R\$ 1.000,00
- Classe *Mid-high*: R\$ 1.000,00 < Preço ≤ R\$ 2.000,00
- Classe *High-end*: R\$ 2.000,00 < Preço ≤ R\$ 3.000,00
- Classe *Premium*: Preço > R\$ 3.000,00

A Figura 5.2 mostra a distribuição dos usuários entre as classes de preço. As classes *Intermediário* e *Premium* são as que apresentam as maiores e menores quantidades de usuários, respectivamente.

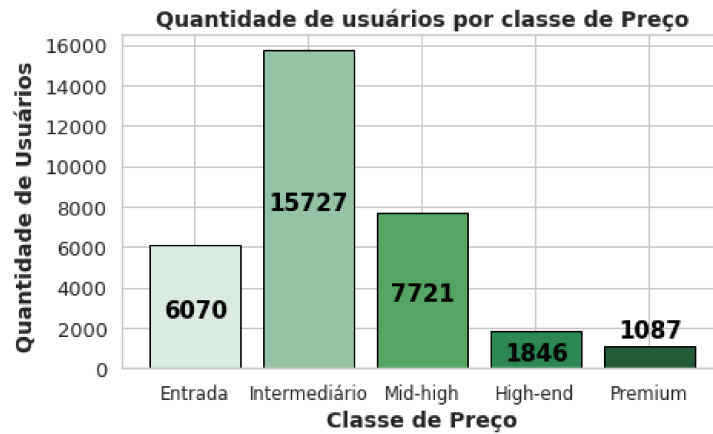


Figura 5.2: Quantidade de usuários por classe de preço.

### 5.1.2 Análise

Para verificar a relação entre as classes de preços e as categorias de aplicativos que o usuário possui instalado foi aplicada a métrica de *Lift*. As Figuras 5.3 a 5.7 apresentam os 5 maiores e menores *Lifts* para as classes de preço. Dessa maneira, serão comparados os usuários que possuem dispositivos de uma classe de preço com os usuários restantes.

Com a Figura 5.3 é possível perceber que 56,94% dos usuários que possuem aparelhos móveis baratos (classe *Entrada*) têm instalados aplicativos corporativos. Além disso, é possível perceber que a maioria dos *Lifts* são negativos, ou seja, usuários com aparelhos baratos possuem menos, ou não possuem, aplicativos da maioria das categorias. Logo, é possível perceber que há relação entre os aplicativos instalados e o aparelho ser mais barato. Além disso, é possível ver que as categorias *Eventos* e *Cia Aérea Internacional* possuem suportes e *Lifts* muito baixos indicando que a maioria dos usuários com aparelhos baratos não instalam aplicativos destas categorias.

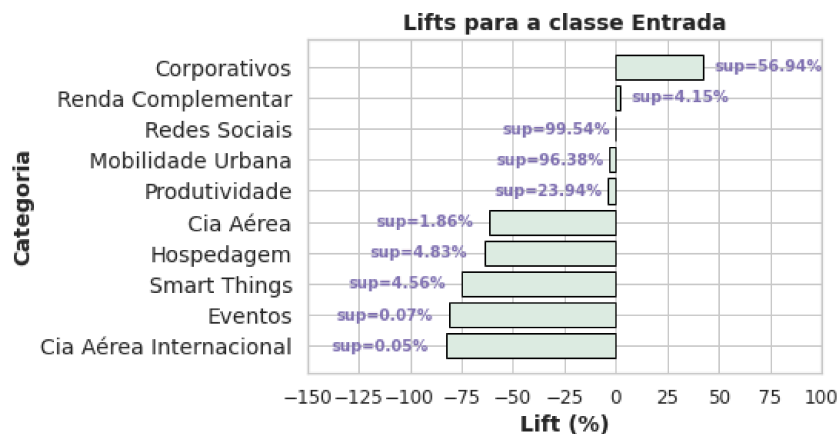


Figura 5.3: Os 5 maiores e 5 menores *Lifts* para a classe *Entrada* (sup=Suporte)

Analisando a Figura 5.4 dos *Lifts* para as classe *Intermediário* é possível verificar que os usuários que possuem aparelhos dessa classe possuem baixa relação com aplicativos relacionados a viagens (Utilitários de Viagem, Cia Aérea e Cia Aérea Internacional). Isso porque, além dos *Lifts* negativos para essas classes, os valores de suporte baixos mostram que poucos usuários com aparelhos intermediários possuem aplicativos relacionados a viagens. Ainda, é possível perceber também que, como a classe *Barato*, a maioria dos *Lifts* são negativos o que pode ser justificado pela baixa capacidade de memória e processamento de celulares destas categorias.

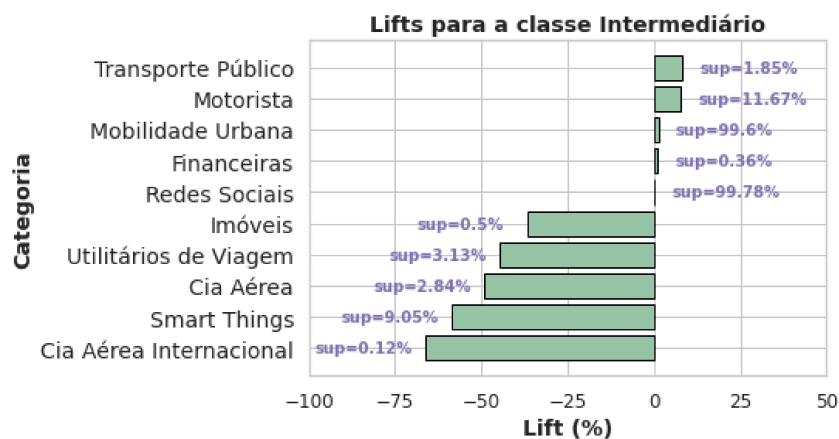


Figura 5.4: Os 5 maiores e 5 menores *Lifts* para a classe *Intermediário* (sup=Suporte)

A partir da Figura 5.5 vemos que aparelhos da classe *Mid-high* possuem *Lifts* não negativos para a maioria das categorias de aplicativos. Isso pode ser explicado pelo fato de que a partir desta faixa de preço os aparelhos possuem mais capacidade de armazenamento e processamento, logo os usuários podem instalar mais aplicativos. Ainda, a categoria *Smart Things* que nas classes anteriores está entre os 5 piores *Lifts*, para a classe *Mid-high* ocupa a quinta colocação, o que pode estar relacionado com as melhores capacidades dos aparelhos, pois geralmente *Smart Things* necessitam de uma maior capacidade energética e de processamento.

Analisando a Figura 5.6, é possível ver que dentre o 5 maiores *Lifts* da classe *High-end*, 4 são relacionados com viagens. Esta relação pode ser explicada pelas melhores condições financeiras dos usuários que possuem aparelhos desta faixa de preço se comparados com as faixas anteriores. O que favorece esta hipótese é o *Lift* de -01,85% com suporte 4,01% para Renda Complementar, mostrando que a maioria dos usuários com aparelhos *High-end* não possuem aplicativos desta categoria.

Como é mostrado na Figura 5.7, há várias categorias de aplicativos em comum com a Figura 5.6. Além dos usuários com aparelhos da faixa *Premium* possuírem relações com aplicativos de viagem, ainda possuem poucos usuários com Renda Complementar. Ademais, usuários com aparelhos *Premium* possuem *Lift* positivo (162,62%) para aplicativos de Finanças.

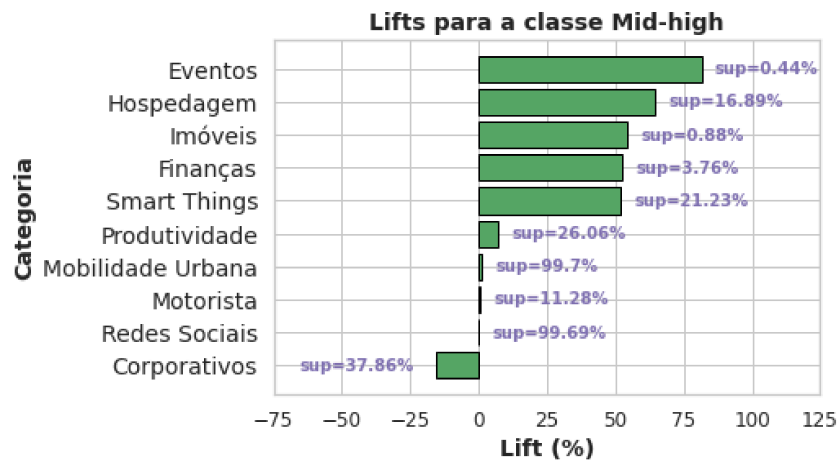


Figura 5.5: Os 5 maiores e 5 menores *Lifts* para a classe *Mid-high* (sup=Suporte)

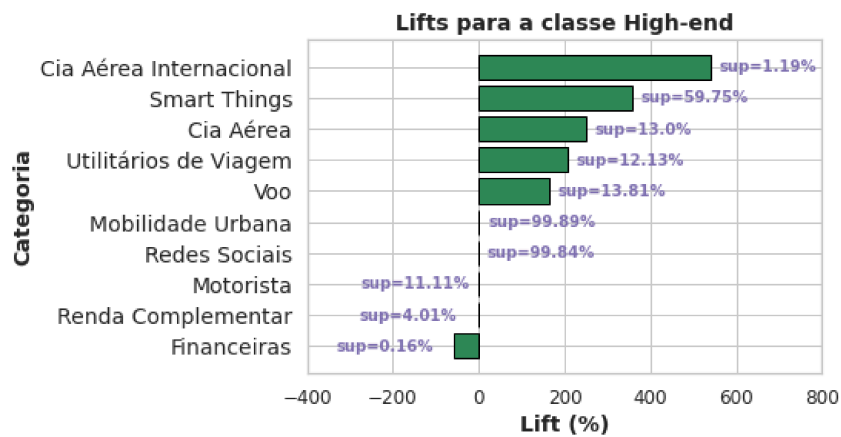


Figura 5.6: Os 5 maiores e 5 menores *Lifts* para a classe *High-end* (sup=Suporte)

### 5.1.3 Modelos Preditivos

As análises feitas na seção anterior mostram que é possível diferenciar os aplicativos instalados com as faixas de preços dos aparelhos móveis. Por este motivo, nesta seção são testados modelos preditivos para prever a faixa de preço do aparelhos do usuário utilizando como entrada seus aplicativos e sua cidade.

Para simplificar a discussão dos resultados, os modelos binários, com os componentes do PCA, adicionados dos atributos do IBGE e aplicando Floresta Aleatória foi a combinação que alcançou, na média das 35 iterações, os melhores resultados das métricas precisão, revocação e F-Score.

A Figura 5.8 apresenta o resultado das métricas para as 35 execuções da melhor combinação. As classes *High-end* e *Premium* obtiveram, no geral, os melhores resultados das métricas, algo que pode ser explicado pelo fato das duas classes possuírem os maiores números de categorias com *Lifts* positivos. Ao mesmo tempo, as classes com poucos *Lifts* positivos, *Entrada* e *Intermediário*, são também as que possuem os piores resultados.

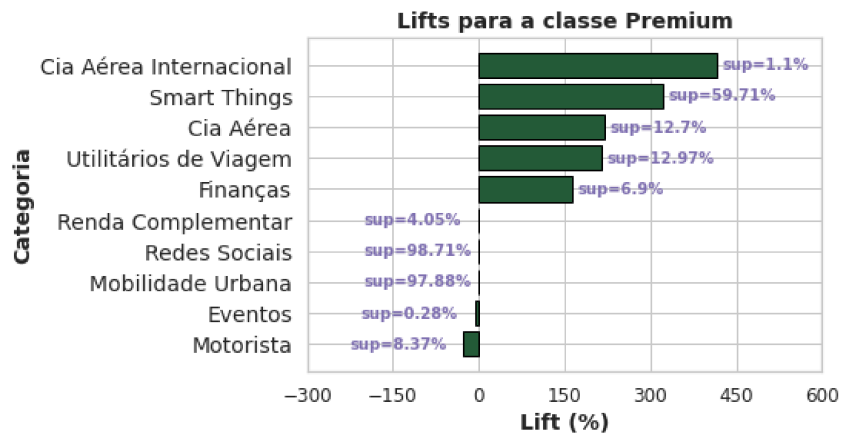


Figura 5.7: Os 5 maiores e 5 menores *Lifts* para a classe *Premium* (sup=Suporte)

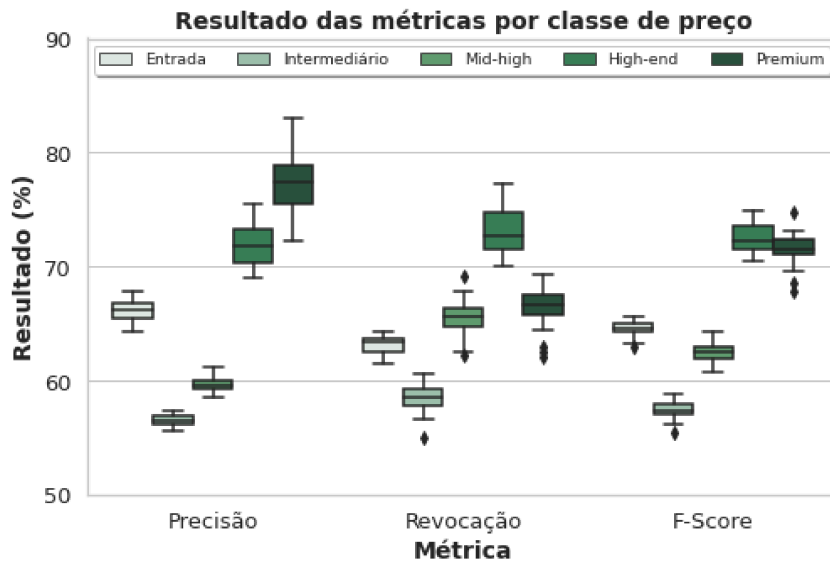


Figura 5.8: Resultados das métricas Precisão, Revocação e F-Score para as 5 classes de preços.

Considerando os resultados obtidos pelos modelos, pode-se verificar que é possível inferir o preço do aparelho do usuário utilizando como entrada sua lista de aplicativos e os dados sobre sua cidade, com qualidade satisfatória. Assim, é possível indicar serviços e produtos condizentes com a capacidade do aparelho do usuário. Logo, empresas e pesquisadores podem realizar recomendações mais assertivas.

## 5.2 Perfil de Estabelecimento

Este perfil visa compreender o comportamento do usuário no que tange os tipos de locais que ele visita. O objetivo principal é analisar a relação entre os aplicativos e a cidade do usuário com os tipos de locais que possam interessá-lo. Estas informações podem ser úteis, por exemplo, para uma empresa conhecer melhor seus clientes ou para uma campanha governamental. Para isso, a base de dados dos usuários foi enriquecida com quatro tipos de locais do *OpenStreetMap* para realização do estudo. A seguir serão apresentados os dados utilizados e como eles foram enriquecidos, além disso a análise realizada para validar a construção de modelos preditivos e os resultados parciais obtidos.

### 5.2.1 Descrição dos Dados

Foi utilizado o conector OSM do *MoreData*, no mês de Janeiro de 2020, para a coleta de quatro diferentes tipos de locais do OSM: *Aeroporto*, *Hospital*, *Shopping* e *Universidade*. A escolha desses tipos de locais foi feita considerando a precisão de coleta do GPS dos *smartphones*. Isso porque, durante a coleta da localização pode ocorrer um erro e com isso estabelecimentos que, em sua maioria, são geograficamente pequenos, como academias e padarias, por exemplo, poderiam perder vários usuários ou conter muitos usuários que na realidade visitaram estabelecimentos vizinhos e de tipos completamente diferentes. Além disso, muitos locais são mapeados como pontos, ou seja, seria necessário estabelecer um raio para verificar se um ponto de interesse do usuário está contido ou não. O problema disto é que não há um valor de raio genérico que funcionaria para todos os estabelecimentos. Por estes motivos, somente os locais geograficamente significativos e mapeados como polígonos foram considerados. Dessa maneira, foram coletados 7.540 locais pertencentes ao território brasileiro, distribuídos conforme mostra a Figura 5.9.

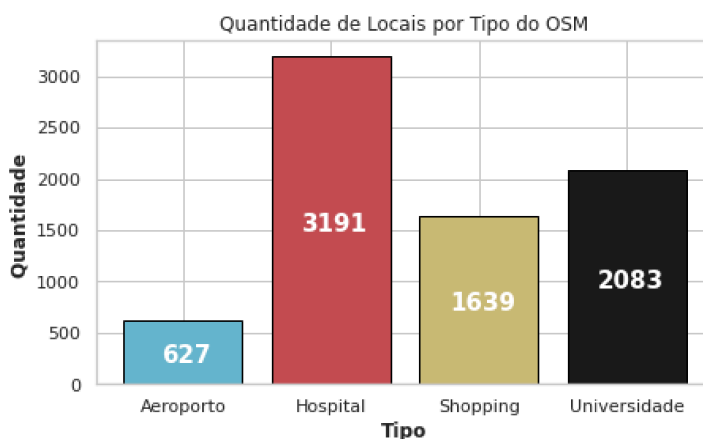


Figura 5.9: Quantidade de locais coletados para cada tipo do OSM.

O enriquecimento dos locais por meio do conector OSM foi realizado a partir de uma consulta geoespacial com uma relação de inclusão entre a geometria específica de um local e o ponto de interesse do usuário, resultando em um objeto JSON contendo o tipo do local que contém o ponto de interesse. A Figura 5.10 mostra a quantidade de usuários que possuem pelo menos um ponto de interesse em algum tipo de local do OSM.

Como o objetivo deste estudo é a relação entre os aplicativos e a cidade de um usuário, e um local que possa interessá-lo, cada tipo de local foi individualmente analisado. Além disso, foi considerado que se pelo menos um dos POIs do usuário esteja contido em um local, então o usuário é interessado no tipo referente. Levando essas questões em consideração, as quantidades de usuários que são apresentadas na Figura 5.10 podem possuir interseções, ou seja, um usuário pode possuir interesse em mais de um tipo de local.

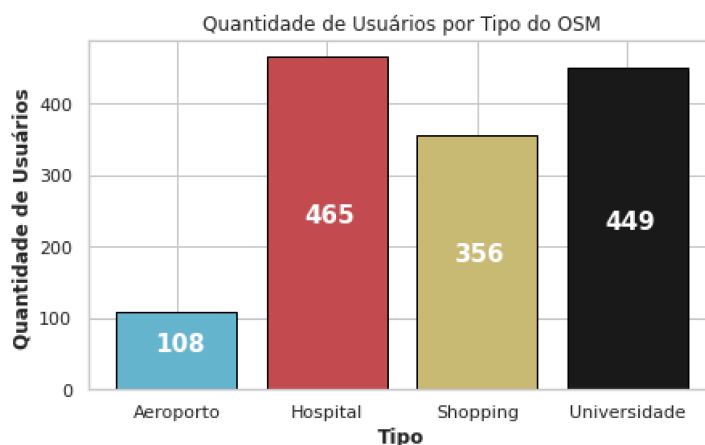


Figura 5.10: Quantidade de usuários que possuem pelo menos um ponto de interesse em cada um dos locais do OSM. Os valores apresentados são exclusivamente para cada tipo, logo podem haver usuários em comum.

## 5.2.2 Análise

Nesta seção, será investigado se existe correlação entre algumas categorias de aplicativos e o interesse de usuários em algum dos 4 locais considerados neste trabalho. Para isto, é calculado o *Lift* de cada categoria de aplicativo considerando cada tipo de local. Ou seja, considerando uma amostra que possua somente usuários com interesse em um tipo de local e outra de usuários que não possua interesse neste tipo, é comparada a porcentagem de usuários que possui algum aplicativo de uma categoria. Com isso, é possível avaliar quantas vezes há um ganho (ou perda, caso o valor seja negativo) na relevância de uma categoria de aplicativo ao saber o interesse em um tipo de local. As Figuras 5.11 a 5.14 mostram os 5 maiores e menores valores de *Lift* para as 4 categorias de locais estudados.

Como mostra a Figura 5.11, é possível observar que 15,74% dos usuários que possuem um ponto de interesse em Aeroportos possuem aplicativos relacionados a *Voo*, o que influencia na taxa dos usuários que frequentam esse tipo de local em comparação aos usuários que não possuem interesse, demonstrado pelo *Lift* de 173,68%. O gráfico mostra, ainda, que os *Lifts* de aplicativos pertencentes a categoria de *Cia Aérea Internacional* e *Cia Aérea* são positivos, sendo explicado pela procura da melhor companhia aérea e pelo planejamento dos voos feito por usuários que possuem interesse nesse tipo de local. Portanto, há uma relação entre aplicativos dessas categorias e o local Aeroporto. Além disso, o gráfico mostra duas categorias (*Eventos* e *Financeiras*) com suporte 0% seguido de *Lift* -100%, indicando que poucos usuários instalaram aplicativos relacionados a essas categorias.

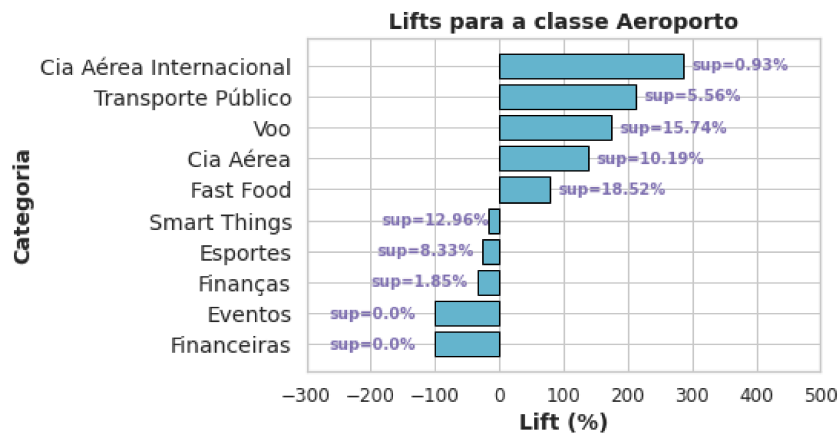


Figura 5.11: Os 5 maiores e 5 menores *Lifts* para *Aeroporto* (sup=Suporte)

Analisando a Figura 5.12, é possível perceber que a taxa dos usuários que possuem algum ponto de interesse em hospitais e têm aplicativos instalados sobre *Medicina* é 64,36% maior em comparação aos usuários que não possuem interesse nesse tipo de local. Portanto, é possível afirmar que existe uma associação entre essa categoria e Hospitais. O gráfico também mostra que a categoria *Transporte Público* possui *Lift* positivo, sendo explicado pelo fato de usuários utilizarem esse tipo de transporte para irem aos Hospitais. Assim como usuários que possuem PoI em aeroportos, é possível observar que existe um *Lift* alto para *Companhia Aérea Internacional*; porém o suporte é menor se comparado à figura 5.11. Além disso, a categoria de aplicativos de *Utilitários de Viagem* possui um *Lift* negativo, não sendo possível afirmar que usuários que possuem ponto de interesse em hospital tem relação a aplicativos sobre voos.

A Figura 5.13 mostra que o *Lift* de aplicativos da categoria de *Investimento* é positivo e o de aplicativos de *Renda Complementar* é negativo, para usuários que possuem algum PoI em *Shopping*. É possível afirmar que essa é uma característica da base utilizada nesse trabalho, já que existem usuários de diversas fontes de renda que frequentam shoppings. Portanto, não é possível generalizar e relacionar o uso de

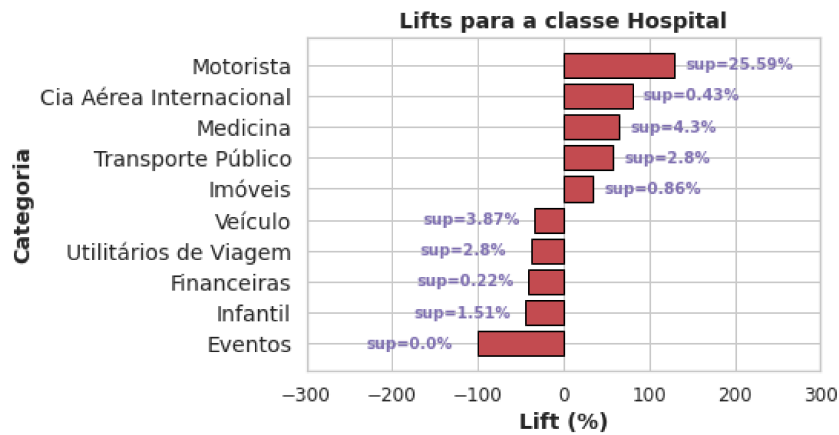


Figura 5.12: Os 5 maiores e 5 menores *Lifts* para *Hospital* (sup=Suporte)

aplicativos de renda e esse local de interesse. No entanto, é possível observar que existe uma relação entre os usuários que possuem PoI em *Shopping* e categoria de aplicativos de *Transporte Público* pelo *Lift* de 90,60%, visto que é um tipo de transporte utilizado pelos usuários para visitar shoppings.

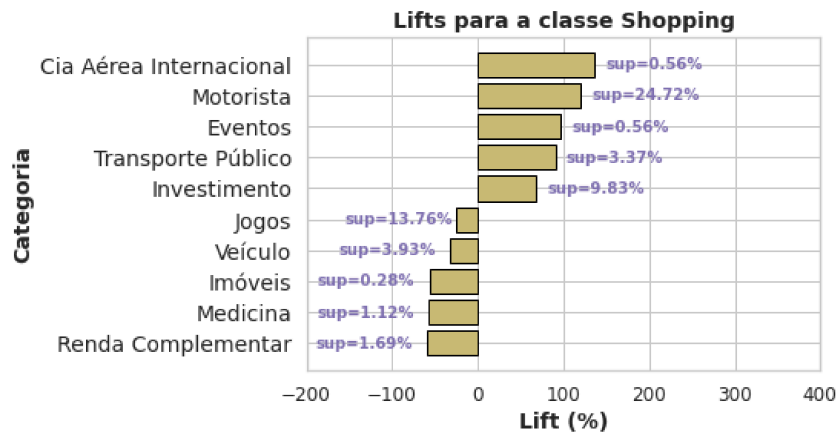


Figura 5.13: Os 5 maiores e 5 menores *Lifts* para *Shopping* (sup=Suporte)

A Figura 5.14 mostra que usuários que possuem ponto de interesse em universidades possuem um alto número de aplicativos sobre *Educação* (14,48%) e um *Lift* de 111,64%. Ademais, o gráfico demonstra que a categoria de aplicativos *Infantil* possui *Lift* negativo, o que condiz com a faixa etária dos usuários que frequentam universidades, na maioria estudantes. Esse gráfico ainda mostra uma grande quantidade de usuários que têm aplicativos da categoria de *Motorista*, o que pode ser explicado pelos usuários que são motoristas que esperam nesses locais. Existe, ainda um *Lift* alto para aplicativos de *Transporte Público* (89,38%), sendo explicado pelo fato de que muitos usuários que frequentam universidades utilizam transporte público para a locomoção.

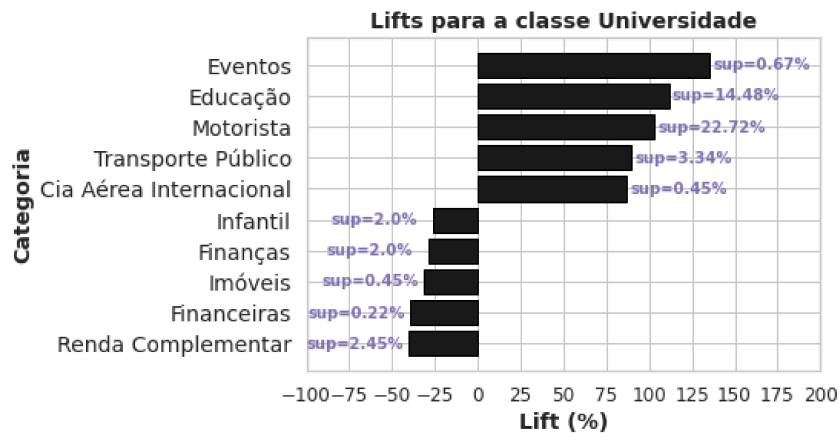


Figura 5.14: Os 5 maiores e 5 menores *Lifts* para *Universidade* (sup=Suporte)

### 5.2.3 Modelos Preditivos

Como apresentado anteriormente, é possível perceber algumas diferenças em relação aos aplicativos instalados pelos usuários e os seus interesses em algum dos locais estudados (*Aeroporto*, *Hospital*, *Shopping* e *Universidade*). Portanto, nessa seção será verificado se é possível criar modelos preditivos para classificar um usuário interessado em algum desses locais, considerando como entrada os seus aplicativos e a sua cidade.

Com os dados enriquecidos dos 31.457 usuários, foram geradas as entradas para os modelos considerando uma variável *dummy* para as categorias de aplicativos, sendo 0 caso o usuário não possua nenhum aplicativo da categoria e 1 caso contrário, além das respectivas classes da população e PIB/capita. Foram criados 4 modelos binários diferentes, um para cada tipo de local separadamente, além de um modelo multi-classe que considera todos os 4 tipos simultaneamente. Dessa maneira, é previsto a possibilidade do usuário possuir interesse ou não em algum dos 4 tipos de locais.

Para simplificar a discussão dos resultados, os modelos binários utilizando Floresta Aleatória, os componentes do PCA como atributos de entrada e com adição dos atributos do IBGE foi a combinação que conseguiu, em geral, os melhores resultados de precisão, revocação e F-Score. Por este motivo, a seguir são discutidos os resultados dos modelos binários para esta configuração.

Uma questão importante que deve ser considerada para a avaliação dos resultados é uma limitação quanto ao real interesse do usuário no local. Neste trabalho, está sendo assumido que um usuário tem interesse em algum tipo de local (e.g., Shopping) se ele possui algum PoI em algum estabelecimento desse tipo. Porém, não estão sendo tratados os diferentes tipos de interesse, como funcionário ou visitante, por exemplo.

A Figura 5.15 apresenta os resultados para as métricas precisão, revocação e F-Score para as 35 execuções da melhor combinação. Para a precisão é possível ver que

Aeroporto se destaca em relação aos outros, o que pode ser explicado pelo fato de que ele possui mais categorias diretamente relacionadas como é possível visualizar os valores do *Lift* na Figura 5.11. Também considerando os valores do *Lift* é possível supor o motivo de Universidade possuir os melhores resultados de revocação, já que este tipo obteve mais categorias com *Lift* altos. Finalmente, com o F-Score é possível ver que, no geral, os 4 tipos de lugares obtiveram resultados parecidos com uma pequena vantagem para Universidade que possui bons resultados para precisão e revocação.

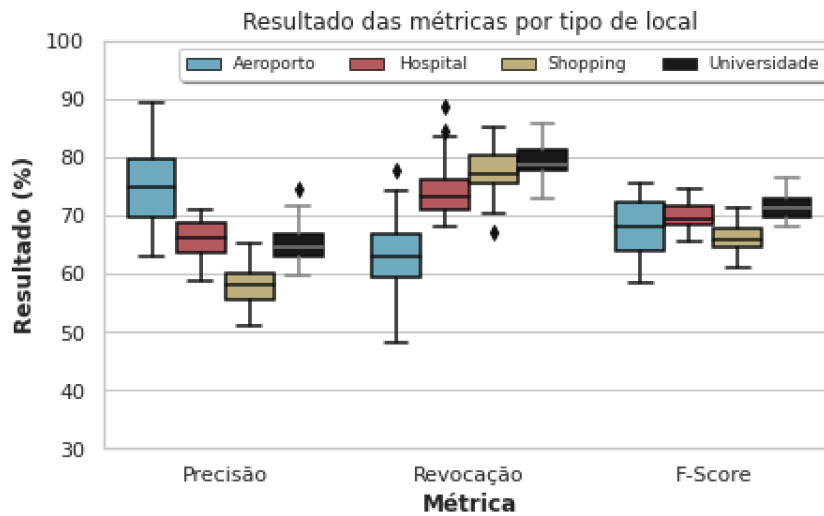


Figura 5.15: Resultados das métricas Precisão, Revocação e F-Score para os 4 tipos de locais.

Os resultados obtidos mostram que para a base de dados utilizada, o Perfil de Estabelecimento é satisfatório. Dessa maneira, é possível utilizar somente a lista de aplicativos do usuário e sua cidade para inferir seu interesse em *Aeroporto*, *Hospital*, *Shopping* e *Universidade*. Com isso, a comunidade científica e o mercado podem compreender melhor as preferências do usuário. Assim, soluções podem ser desenvolvidas considerando estas informações para maior assertividade.

## 5.3 Perfil de Área Funcional

Diferente da análise considerando exclusivamente o tipo do local, somente o fato do usuário possuir algum ponto de interesse contido na região já é suficiente para avaliação da área funcional. Em contra partida, no perfil dos tipos de locais, apresentados anteriormente, o usuário necessita ter algum PoI contido em um local de um dos tipos analisados para conseguir realizar o estudo. Logo, o perfil de área funcional possibilita uma análise mais ampla das regiões que o usuário visita. Dessa maneira, serão relacionados os aplicativos e cidade do usuário com os tipos de região que ele visita.

### 5.3.1 Descrição dos Dados

Para construir as áreas foram utilizadas as localizações brutas dos 31.457 usuários base, totalizando 4.966.909 localizações. Utilizando o índice H3, foram formadas áreas com aproximadamente 500 metros de raio, equivalente a 0,7373 km<sup>2</sup> (resolução igual a 8, segundo a configuração do H3<sup>1</sup>). Com estas configurações, foram geradas 217.262 áreas, mas foram removidas as áreas com menos de 6 visitas. Isso porque, estas áreas fazem com que suas funcionalidades tenham altos pesos nos cálculos. Logo, foram consideradas 106.051 áreas.

A Wiki<sup>2</sup> do *OpenStreetMap* fornece as relações entre as *tags* que podem ser coletadas. Dessa maneira, foi coletado através de uma requisição todas as combinações possíveis de locais para servirem como entrada para o conector *OSM* do *MoreData*. Baseando-se no trabalho de Yu et al. (2018), esta dissertação considera oito tipos de funcionalidades, são elas: *Companhia, Shopping, Turismo, Entretenimento, Esportes, Educação, Alimentação e Governo*. Assim, os tipos de locais foram mapeados em uma das oito funcionalidades.

Após as etapas de coletas dos locais e definição das áreas, cada área foi enriquecida com dados dos locais que possui. Com isso, os locais foram agrupados de acordo com a funcionalidade para calcular o peso (no caso a percentagem) de cada funcionalidade em cada uma das áreas. Com as áreas funcionais já definidas, a frequência de visitas de cada usuário em cada área foi calculada. Por fim, foram computadas as notas de cada usuário em relação as oito funcionalidades, com base nos detalhes descritos na Seção 2.5.

As notas foram utilizadas para separar os usuários em interessados ou não em cada funcionalidade. Esta classificação foi feita utilizando o limite superior de cada classe de funcionalidade. Logo, os usuários com notas maiores que o limite superior possuem interesse, e sem interesse caso contrário. Os limites de cada classe ficaram

<sup>1</sup><https://github.com/uber/h3/blob/master/docs/core-library/restable.md>

<sup>2</sup>[https://wiki.openstreetmap.org/wiki/Map\\_features](https://wiki.openstreetmap.org/wiki/Map_features)

da seguinte forma:

- *Alimentação*: 23,0680
- *Companhia*: 1,0316
- *Educação*: 32,5000
- *Entretenimento*: 45,9156
- *Esportes*: 10,2961
- *Governo*: 0,9036
- *Shopping*: 25,6207
- *Turismo*: 16,6667

A Figura 5.16 mostra a quantidade de usuários que cada classe de funcionalidade possui de acordo com a classificação.

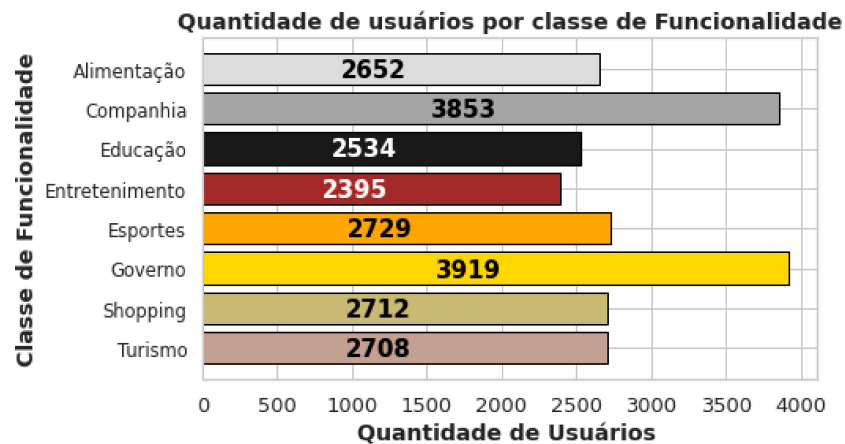


Figura 5.16: Quantidade de usuários que possuem interesse na funcionalidade. Os valores apresentados são exclusivamente para cada funcionalidade, logo podem haver usuários em comum.

### 5.3.2 Análise

Para verificar a relação entre as funcionalidades e os aplicativos instalados pelos usuários foi aplicada a métrica *Lift*. As Figuras 5.17 a 5.24 apresentam os *Lifts* das oito classes de funcionalidade.

A Figura 5.17 mostra que áreas relacionadas com alimentação possuem *Lift* positivo para *Transporte Público*. Isso pode estar relacionado com a escolha deste tipo de transporte por pessoas que visitam este tipo de região. Além disso, a falta das

categorias de *Delivery* e *Fast Food* na lista dos cinco maiores *Lifts* pode estar relacionada ao fácil acesso dos usuários a restaurantes, bares, entre outros. Ademais, com o *Lift* negativo da Categoria de *Infantil* não é possível sugerir relações entre aplicativos desta categoria e visitas em área de alimentação.

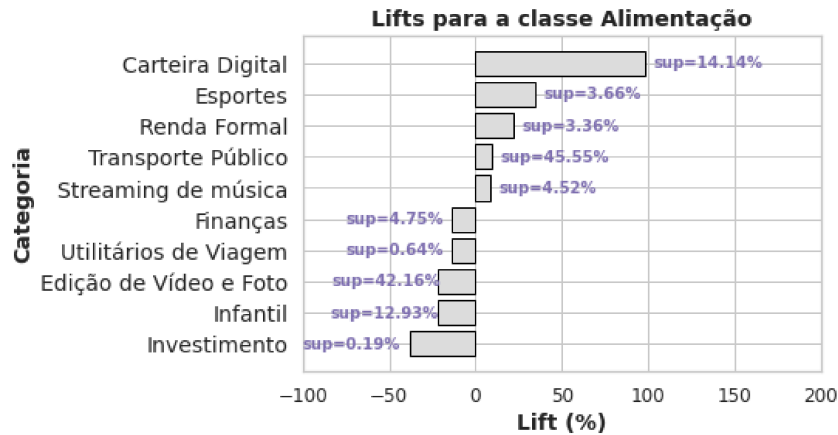


Figura 5.17: Os 5 maiores e 5 menores *Lifts* para a classe *Alimentação* (sup=Suporte)

Com a Figura 5.18 é possível ver o *Lift* positivo e suporte alto para *Corporativos*, sendo explicado pela necessidade dos usuários que trabalham em companhias de acompanhar tarefas e fazer reuniões. Também, o gráfico exhibe *Lift* negativo para a categoria *Fast Food*, mas com suporte baixo, logo não se pode tirar conclusões dessa relação.

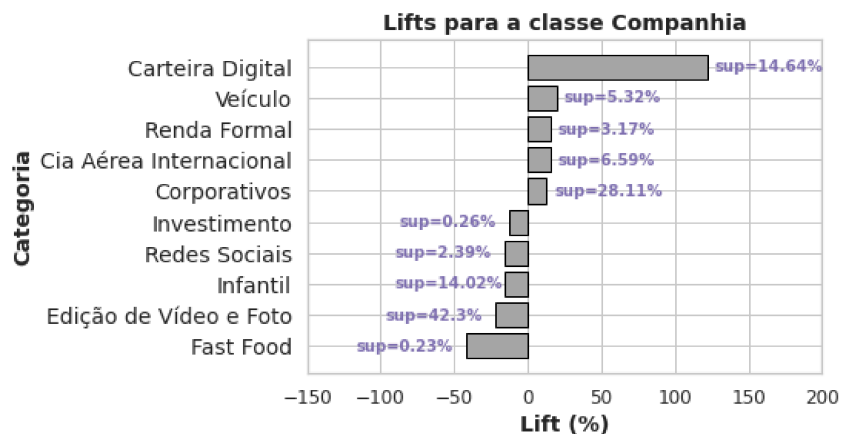


Figura 5.18: Os 5 maiores e 5 menores *Lifts* para a classe *Companhia* (sup=Suporte)

A Figura 5.19 mostra que o *Lift* para categoria *Carteira Digital* é positivo para usuários que visitam com frequência áreas educacionais. Já o *Lift* negativo para *Finanças* também pode estar diretamente relacionado ao público que frequenta áreas educacionais, em sua grande maioria, estudantes.

A Figura 5.20 mostra que usuários que frequentam áreas de entretenimento possuem *Lifts* positivos para aplicativos de *Cia Aérea Internacional*. Algo que pode indicar

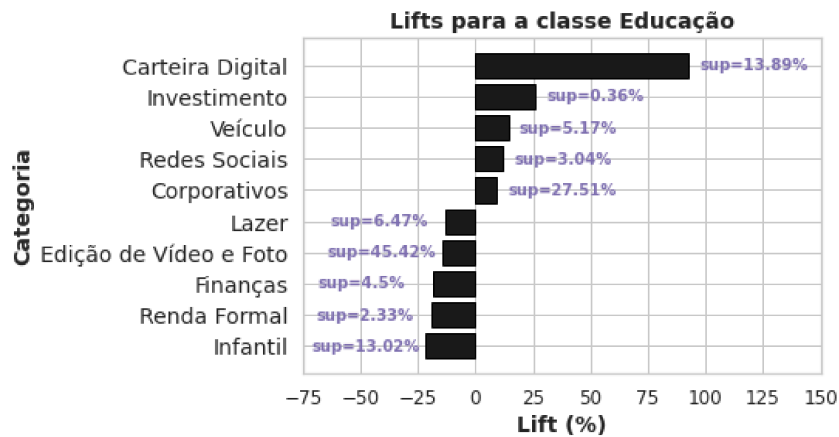


Figura 5.19: Os 5 maiores e 5 menores *Lifts* para a classe *Educação* (sup=Suporte)

que pessoas com estes aplicativos viajam mais, e por isso estão mais em contato com este tipo de área. Outra categoria com *Lift* positivo é *Esportes*, que está diretamente relacionado com formas de entretenimento. Ademais, o *Lift* negativo e o baixo suporte para a categoria *Fast Food*, não permite afirmar que usuários que visitam áreas de entretenimento tem relação a aplicativos desta categoria.

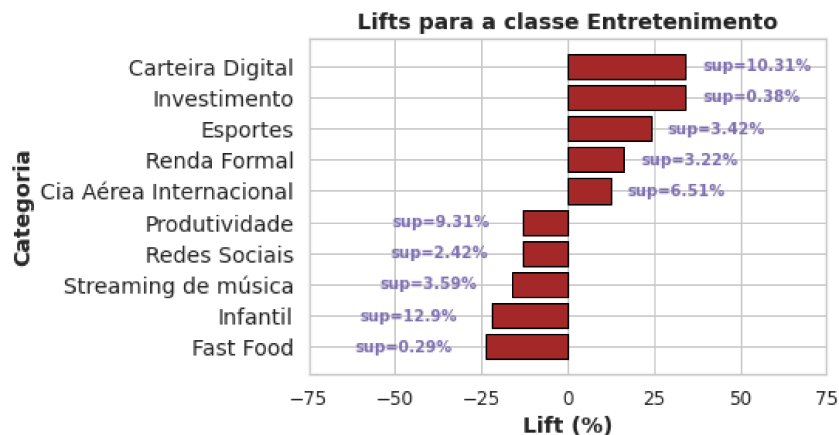


Figura 5.20: Os 5 maiores e 5 menores *Lifts* para a classe *Entretenimento* (sup=Suporte)

A Figura 5.21 mostra os *Lifts* positivos para aplicativos da categoria *Esportes* para usuários que visitam áreas esportivas. Logo, é possível supor que pessoas que frequentam este tipo de região também acompanham informações relacionadas a esportes. O *Lift* negativo para *Infantil* pode estar relacionado com a faixa etária do público que geralmente frequenta áreas de Esportes, que podem ser jovens ou adultos ainda sem filhos.

A Figura 5.22 mostra que o maior *Lift* é para a categoria de *Carteira Digital* (105,19%) para usuários que visitam áreas governamentais. Já o *Lift* negativo para *Infantil* com suporte alto pode estar relacionado com a faixa etária dos frequentadores deste tipo de local.

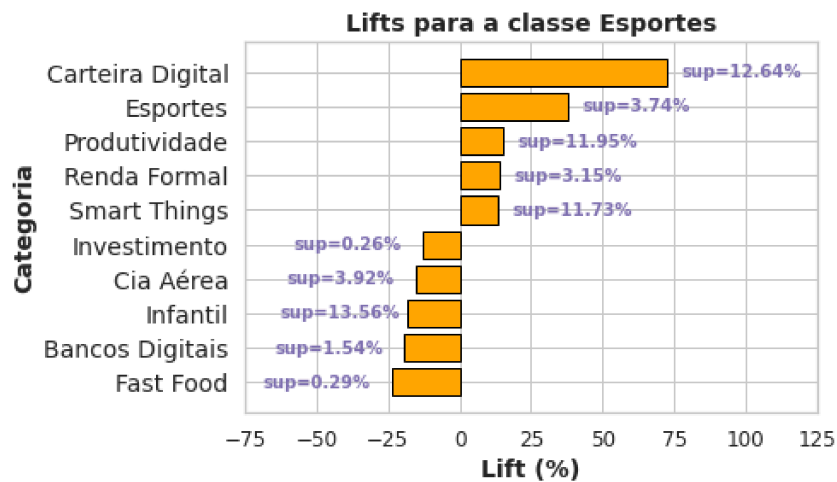


Figura 5.21: Os 5 maiores e 5 menores *Lifts* para a classe *Esportes* (sup=Suporte)

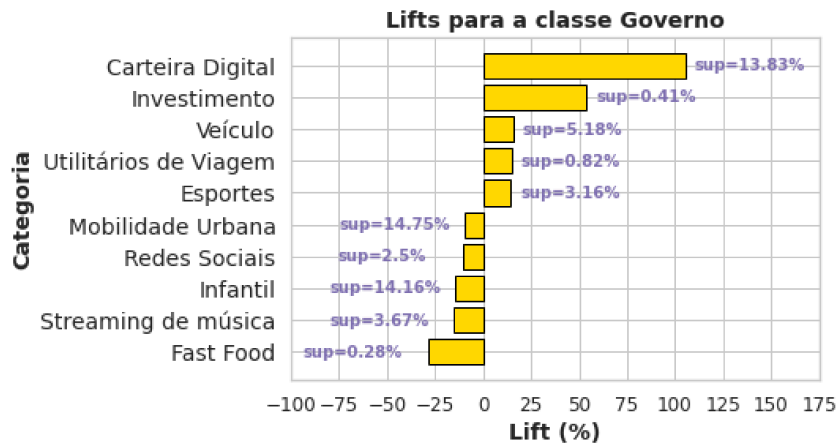


Figura 5.22: Os 5 maiores e 5 menores *Lifts* para a classe *Governo* (sup=Suporte)

A Figura 5.23 mostra que o *Lift* para a categoria *Transporte Público* é positivo e o suporte alto para usuários que frequentam áreas de *Shopping*. Isso pode ser explicado por ser um forma de deslocamento bastante utilizada por clientes e funcionários deste tipo de estabelecimento. Já com os *Lifts* negativos para *Investimento* e *Infantil* não é possível realizar nenhuma inferência, pois diferentes perfis de pessoas frequentam este tipo de região.

O *Lift* para aplicativos relacionados a viagem (*Utilitários de Viagem* e *Cia Aérea Internacional*) é positivo para áreas turísticas, mostrado na Figura 5.24. Isto pode ser explicado pela necessidade de turistas verificarem a situação de voos e de outros suportes que estes tipos de aplicativos fornecem. Esse gráfico ainda mostra o *Lift* negativo para a categoria *Investimento*, porém não é possível realizar inferências entre a relação desta categoria e a funcionalidade *Turismo*.

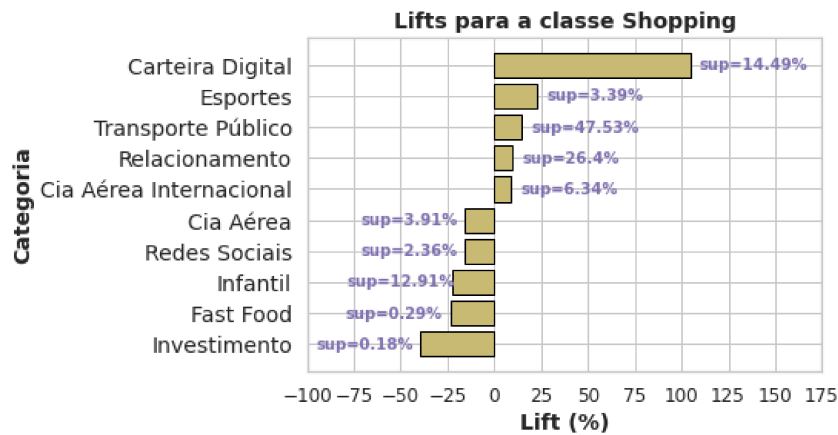


Figura 5.23: Os 5 maiores e 5 menores *Lifts* para a classe *Shopping* (sup=Suporte)

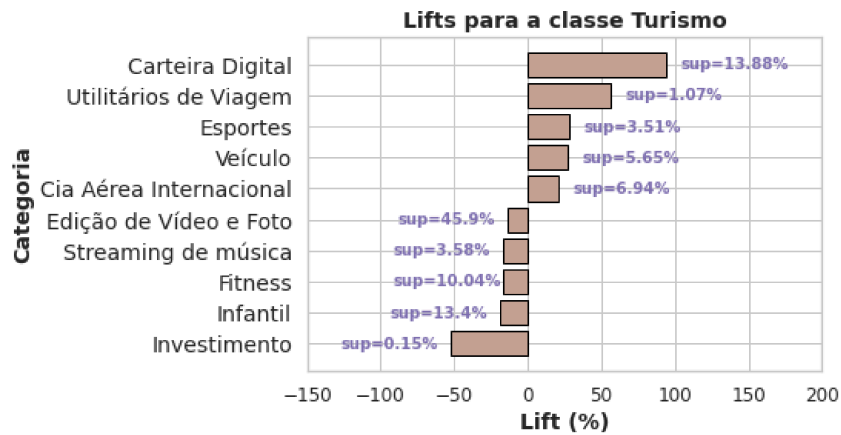


Figura 5.24: Os 5 maiores e 5 menores *Lifts* para a classe *Turismo* (sup=Suporte)

### 5.3.3 Modelos Preditivos

Com a análise apresentada, é possível perceber que há alguma correlação entre as categorias de aplicativos e as funcionalidades das áreas visitadas por eles. Afim de prever o interesse do usuário por determinada funcionalidade foram criados modelos que utilizam como base a lista de aplicativos e a cidade do usuário.

Uma limitação que deve ser ressaltada é o real motivo da visita do usuário em determinada área. Uma pessoa pode frequentar uma região para diferentes finalidades (e.g., trabalhar, estudar, comprar), e esta visita não precisa estar diretamente relacionada com a funcionalidade da região. Por exemplo, um atendente de um banco situado em uma área turística, pode visitar a região somente para trabalhar. Logo, esta pessoa é classificada como interessada neste tipo de região.

Para simplificar as discussões, os modelos binários, com os componentes do PCA adicionados aos atributos do IBGE e aplicando Regressão Logística foi a combinação que alcançou, na média das 35 iterações, os melhores resultados nas métricas precisão, revocação e *F-Score*. A Figura 5.25 apresenta os resultados referentes as 35 execuções.

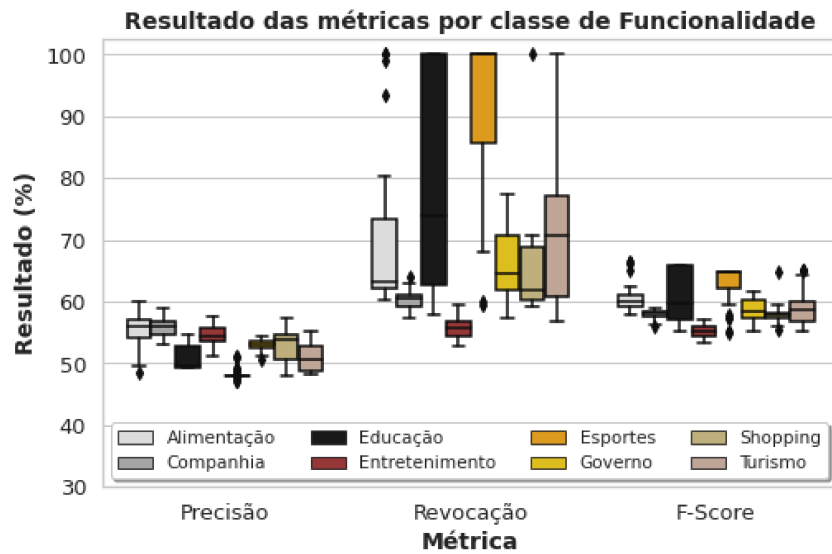


Figura 5.25: Resultados das métricas Precisão, Revocação e F-Score para os 8 classes de funcionalidade.

Na figura, é possível ver que os resultados de precisão foram similares para todas as classes de funcionalidades. Somente a classe *Esportes* obteve, na média das 35 execuções, resultados inferiores a 50% para esta métrica. Já para a métrica revocação, os resultados foram bastantes distintos. Várias classes conseguiram alcançar 100% de revocação, isso é muito relevante para campanhas de *marketing*, por exemplo. Isso pois, valores altos de revocação indicam que todos os usuários que possuem algum interesse na funcionalidade foram previstos corretamente. Dessa maneira, ao invés de sugerir um produto para qualquer usuário escolhido aleatoriamente, o direcionamento da campanha seria mais assertivo.

No geral, as classes de funcionalidade não obtiveram resultados altos para precisão, limitando-se à 60%. Contudo, é possível verificar que muitas classes obtiveram resultados satisfatórios para revocação, destaque de *Educação* e *Esportes*. Estes resultados mostram que a maioria das previsões dos modelos estão corretas. Dessa forma, há uma alta assertividade de usuários que possuem alguma relação com determinada funcionalidade. Considerando estes resultados, utilizando somente a lista de aplicativos e conhecendo a cidade do usuário é possível conhecer os tipos de áreas mais visitadas por ele. Assim, é possível utilizar do Perfil de Área Funcional para compreender padrões de mobilidade social. Logo, propor soluções mais personalizadas para os usuários.

## 5.4 Perfil de Mobilidade

Outra perspectiva analisada é a mobilidade do usuário, baseando-se no trabalho [Pappalardo et al. \(2015\)](#), em que os usuários são classificados como *returners* ou *explorers*. Dessa maneira, com base na lista de aplicativos e a cidade do usuário ele é classificado como *returner* ou *explorer*. Assim, será possível analisar quais categorias de aplicativos cada classe de mobilidade possui.

### 5.4.1 Descrição dos Dados

O perfil de mobilidade utiliza as localizações brutas dos usuários como base para as classificações. Como localizações que se referem a mesma área não são totalmente iguais (i.e., diferem em casas decimais), foi necessário truncar as localizações para serem agrupadas, pois é preciso conhecer as localizações mais frequentadas pelo usuário. As localizações agrupadas serviram como entrada para a biblioteca *scikit-mobility*<sup>3</sup> calcular o raio de giro e o *k\_raio* de giro do usuário.

Como dito anteriormente, este perfil se baseia no trabalho de ([Pappalardo et al., 2015](#)) para a classificação em *returner* ou *explorer*. Igualmente apresentado pelos autores, foram testados valores de *k* iguais a 2, 4 e 8. Para simplificar, o valor de *k* representa a quantidade de localizações mais visitadas pelo usuário que serão consideradas para a classificação em *returner* ou *explorer*. Como nem todos os usuários da base original possuíam pelos menos 8 localizações, restaram somente 25.744. A Figura 5.26 apresenta a distribuição do número de usuário entre as classes.

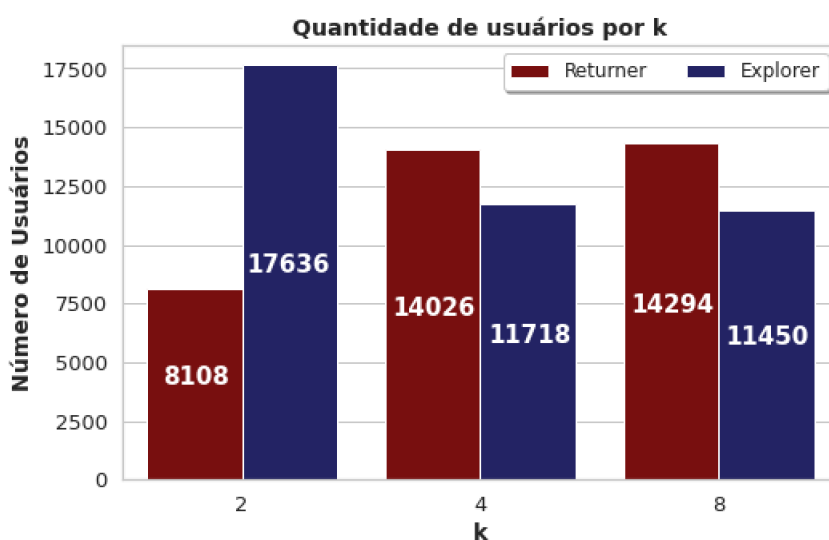


Figura 5.26: Quantidade de usuários por classe de mobilidade.

Para simplificar as discussões, foi selecionado o conjunto do *k* igual a 4. Isso

<sup>3</sup><https://github.com/scikit-mobility/scikit-mobility/>

porque, considerar somente o  $k$  igual a 2 fica limitado para somente *Casa* e *Trabalho*, para a maioria dos usuários. Este comportamento, não condiz com o perfil de muitos usuários que além de casa e trabalho possuem outros pontos de interesse relevantes, como local de lazer e educação. Além disso, a diferença de distribuição entre  $k$  igual 4 e 8 foi pequena.

## 5.4.2 Análise

Nesta seção, serão verificadas as correlações entre as categorias de aplicativos instalados e as classes de mobilidade. Dessa forma, foi utilizada a métrica *Lift* para cada categoria de aplicativo em relação às 2 classes. Por exemplo, considerando um subconjunto de todos os usuários que são *returners* e outro que possua somente *explorers*, são feitas comparações entre as porcentagens de usuários que possuem instalado algum aplicativo da categoria. Com isso, é possível calcular a relevância em saber o tipo de usuário (*returner* ou *explorer*). As Figuras 5.27 e 5.28 apresentam os 5 maiores e menores valores de *Lift* para as classes *returner* e *explorer*, respectivamente.

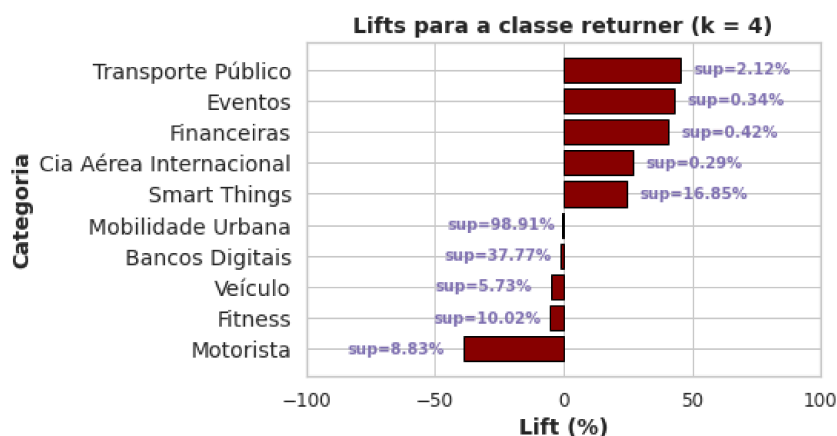


Figura 5.27: Os 5 maiores e 5 menores *Lifts* para a classe *Returner* (sup=Suporte)

Como o perfil de mobilidade considera duas classes, os gráficos de *Lift* apresentam as mesmas categorias, mas com as posições invertidas. A Figura 5.27 mostra que o *Lift* da categoria *Transporte Público* é positivo (24,37%) para usuários classificados como *returner*. Isso pode ser explicado pelo fato de que pessoas que não visitam locais muito distantes e variados não necessitam de veículo próprio, logo utilizam os transportes públicos. Já na Figura 5.28, o *Lift* para *Motorista* é positivo. Algo que pode estar relacionado diretamente com a profissão de motorista de aplicativo. Pois, motoristas visitam vários lugares diferentes durante o dia.

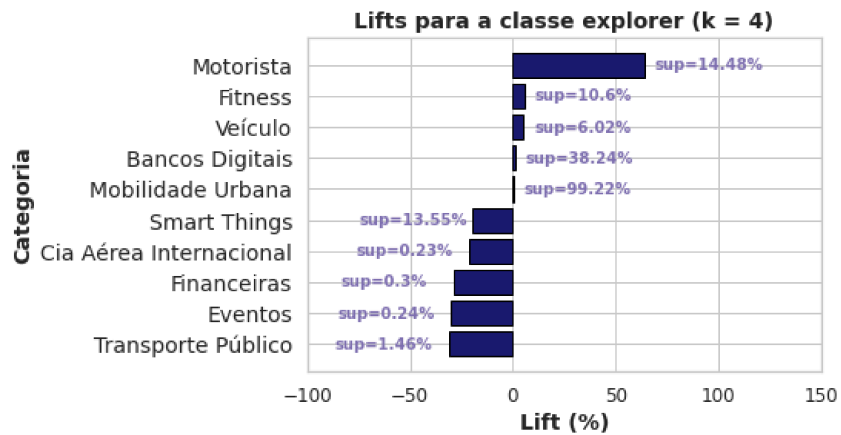


Figura 5.28: Os 5 maiores e 5 menores *Lifts* para a classe *Explorer* (sup=Suporte)

### 5.4.3 Modelos Preditivos

Com as análises é possível ver que há relação entre os aplicativos instalados e a classe de mobilidade. Por isso, foram criados e avaliados modelos preditivos que possuem como entrada as categorias de aplicativos instalados e a cidade do usuário. Afim de simplificar as discussões, são discutidos somente os resultados para o algoritmo *Naive Bayes*, utilizando as classes do IBGE, as componentes do PCA e aplicando *Undersampling*. Pois, foi a combinação que obteve os melhores resultados de precisão, revocação e *F-Score*, apresentados na Figura 5.29.

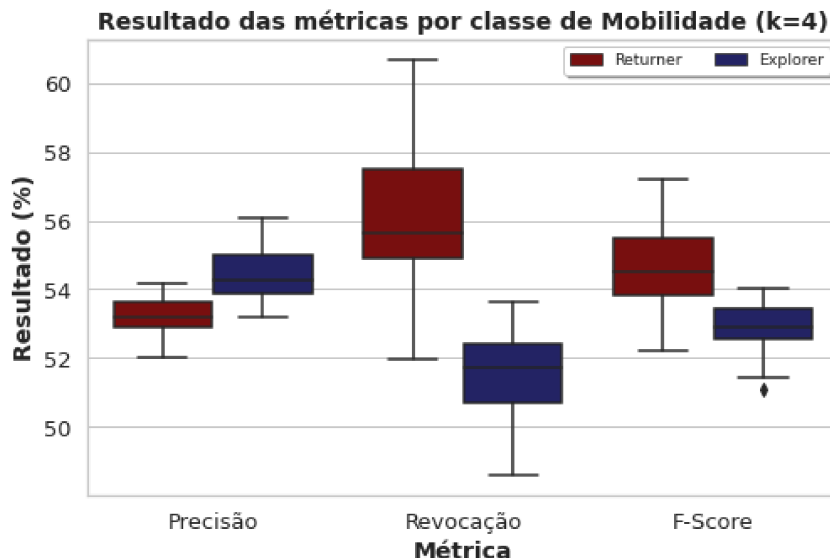


Figura 5.29: Resultados das métricas Precisão, Revocação e *F-Score* para as classe *Returner* e *Explorer* considerando o *k* igual a 4.

Os Gráficos mostram uma pequena vantagem do *Explorer* somente em relação a precisão. Já os modelos para a classe *Returner* obtiveram melhores resultados de revocação e *F-Score*. Isso pode ser explicado pelo maior número de categorias de

aplicativos com *Lift* positivo para *Returners*.

O Perfil de Mobilidade não obteve altos resultados para os modelos gerados. Isso pode estar relacionado aos diferentes perfis de pessoas que pertencem ao conjunto de *Returners* ou *Explorers*. Há pessoas de diferentes idades, preferências e classe social que podem ser classificadas nestas classes. Ainda assim, este perfil pode ser útil para selecionar usuários em comparação com amostras aleatórias. Dessa maneira, é possível melhorar assertividade da escolha do público alvo em campanhas, por exemplo.

## 5.5 Perfil de Personalidade

Diferente dos outros perfis, o Perfil de Personalidade não utiliza dados da cidade do usuário, pois esta informação não interfere no traço do mesmo. Logo, este perfil tem como objetivo prever os traços de personalidade do usuário de acordo com o *Big Five* (McCrae and Jr., 1999) utilizando como base somente sua lista de aplicativos. Assim, será possível verificar se usuários que possuem o traço de extroversão mais acentuado também possuem mais aplicativos de redes sociais, por exemplo.

### 5.5.1 Descrição dos Dados

A base de usuários utilizada é diferente da base dos demais perfis, já que fez-se necessária as respostas no questionário *TUPI* para construção dos modelos preditivos. Por isso, foi desenvolvido o aplicativo *APPProfile*, descrito na Seção 3.2, para coletar a lista de aplicativos e as repostas referentes ao questionário. O *APPProfile* foi divulgado em algumas redes sociais (*Facebook* e *Instagram*) e por *e-mail*, o tempo da coleta dos dados foi de aproximadamente um mês (entre Maio de 2020 e Junho de 2020), alcançando um total de 98 usuários. Destes usuários, 57 são do sexo masculino e 41 do sexo feminino. A média de idade foi de 22,6 anos, com o desvio padrão igual à 7,15, máximo igual à 63 e mínimo igual à 14. Para remover possíveis usuários que tenham respondido sem o propósito de ajudar o trabalho, foram considerados somente os usuários que levaram ao menos 45 segundos para responder ao questionário (no qual foi desconsiderado somente um usuário).

Considerando os aplicativos dos 97 usuários, foram coletados dados de 1.423 aplicativos distintos instalados. Contudo, destes aplicativos, somente 225 estavam contidos no mapeamento utilizado nos outros perfis. Por este motivo, foi necessário utilizar as categorias fornecidas pela *Play Store*, que permitiu que fossem mapeados 874 aplicativos em 45 categorias distintas. Porém, restaram 549 aplicativos sem categorização, sendo que alguns deles foram instalados por muitos usuários. Para solucionar parcialmente este problema, foi feita manualmente a categorização dos aplicativos que possuíssem pelo menos quatro usuários, totalizando um total de 41 aplicativos. Após a categorização manual, a base continha um total de 915 aplicativos categorizados e 508 aplicativos sem categorização, sendo que esses últimos foram desconsiderados por estarem presentes na lista de poucos usuários. Finalmente, foram removidas as categorias que continham menos de 4 usuários, totalizando 38 categorias. A Figura 5.30 apresenta a porcentagem de usuários por categoria de aplicativo.

Após a coleta dos dados e o enriquecimento dos aplicativos com suas respectivas categorias, as notas dos traços de personalidades dos usuários também foram agregadas. A Figura 5.31 mostra a distribuição das notas dos usuários em cada traço de

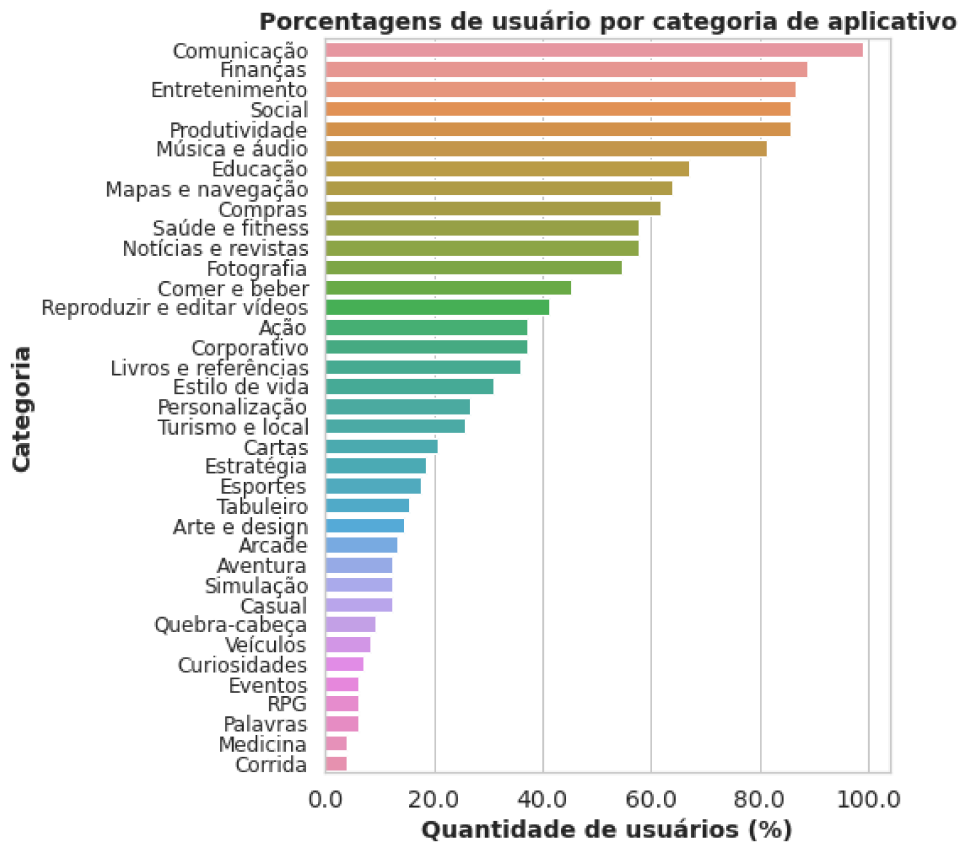


Figura 5.30: Percentagens de usuários por categoria de aplicativo da *Play Store*.

personalidade. A Tabela 5.1 ilustra a matriz completa dos dados após todas as agregações e agrupamentos, sendo que os valores das categorias indicam quantos aplicativos daquela categoria um determinado usuário possui, e o valor de cada traço é a nota do usuário para aquele traço específico, que varia de 1 a 7.

Tabela 5.1: Exemplificação dos dados finais para três usuários fictícios

Identificador Usuário	Categoria de Aplicativo					Traço de Personalidade				
	C1	C2	C3	...	C45	AGR	ANE	CON	EME	EXT
0001	5	3	10	...	0	1	5.5	2.5	6	4.5
0002	4	8	1	...	2	4.5	5	4	6.5	1.5
0003	0	2	3	...	0	7	3.5	1	2	3

Uma consideração importante é que, diferente dos outros perfis, o perfil de personalidade, além de considerar modelos de classificação, também aborda modelos de regressão. Isso é motivado pela variação do score (1 a 7). Logo, além de prever se o usuário possui o traço, é feita a previsão da nota do usuário. Por este motivo, para a criação dos modelos de regressão são utilizados os dados das notas dos usuários. Já para a criação dos modelos de classificação, foi realizado um mapeamento para os traços de personalidade, sendo 1 caso o usuário possua nota maior ou igual a 5.0, e 0 caso contrário, as quantidades de usuários por traço é mostrada na Figura 5.32.

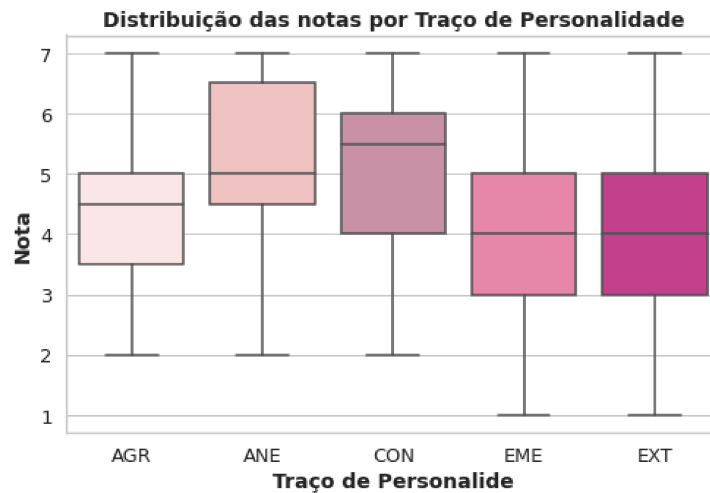


Figura 5.31: A distribuição dos 5 traços de personalidade.

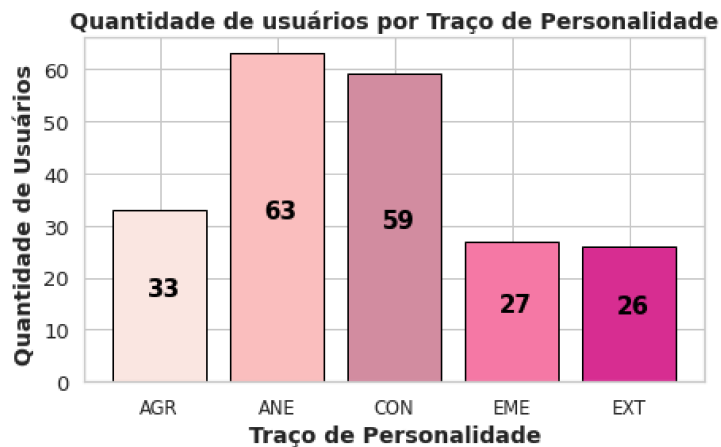


Figura 5.32: Quantidade de usuários por Traço de Personalidade.

## 5.5.2 Análise

Para avaliar a relação entre as categorias de aplicativos e os traços de personalidade, foram gerados os gráficos de *Lift*. Dessa maneira, é possível medir o ganho (ou perda, caso negativo) em saber o traço de personalidade do usuário. As Figuras de 5.33 a 5.37 apresentam os 5 maiores e menores *Lifts* para cada traço de personalidade. Entretanto, como foi necessária a utilização das categorias fornecidas pela *Play Store* não é possível realizar a maioria das inferências. Isso porque, como dito no início deste capítulo, as categorias são preenchidas pelos próprios desenvolvedores. Assim, várias das categorias não expressam exatamente o que o aplicativo faz.

A Figura 5.33 mostra os maiores *Lifts* são das categorias *Aventura* (38,52%) e *Casual* (29,29%) para usuários que possuem o traço *Agradabilidade*. Já as categorias com *Lifts* -100% e com suporte 0.0% indicam que poucos usuários com este traço instalaram aplicativos destas categorias.

A Figura 5.34 apresenta *Lifts* positivos para as categorias *Curiosidade* e *Simulação*

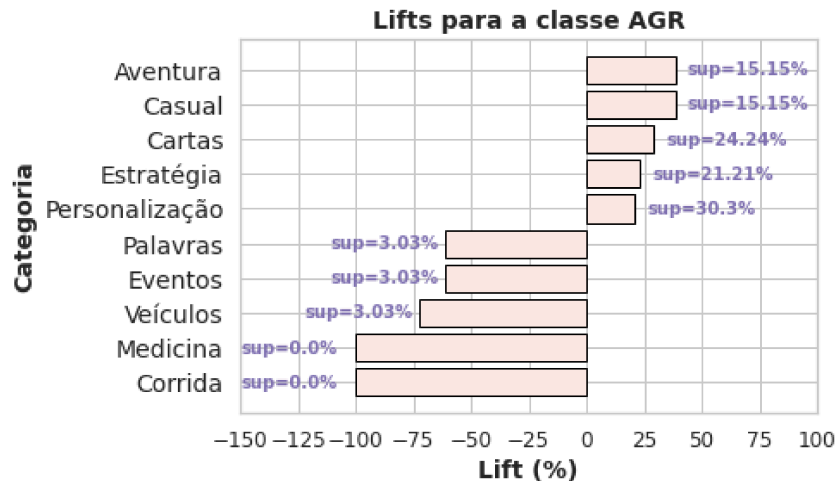


Figura 5.33: Os 5 maiores e 5 menores *Lifts* para a classe *Agradabilidade* (sup=Suporte)

para usuários que possuem o traço *Aberto a Novas Experiências*. Estes valores podem estar relacionados ao fato de que pessoas com este traço estão sempre procurando por diferentes experiências. Já com os valores negativos para *RPG* (-73,02%) e *Casual* (-46,03%) não se pode elaborar nenhuma hipótese.

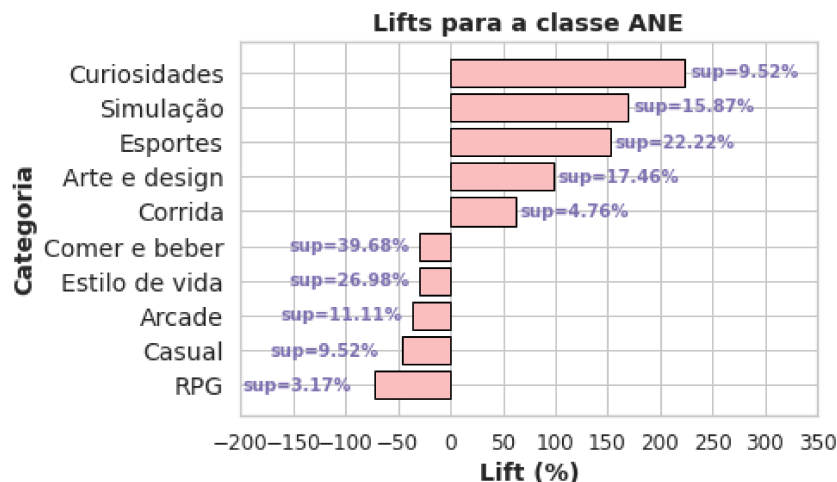


Figura 5.34: Os 5 maiores e 5 menores *Lifts* para a classe *Aberto a Novas Experiências* (sup=Suporte)

A Figura 5.35 mostra um *Lift* alto para aplicativos da categoria *Curiosidade* em relação a usuários com o traço *Conscienciosidade* (286,44%). Já o *Lift* mais baixo para esta classe de personalidade foi para a categoria *RPG* (-87,12%).

Os *Lifts* mais altos são dos aplicativos das categorias *Corrida* (159,26%) e *Veículos* (55,56%) para o traço *Estabilidade Emocional*, Figura 5.36. Já a categoria *Medicina* possui *Lift* -100% e suporte 0.0% indicando que poucos usuários que possuem este traço instalaram aplicativos destas categorias.

A Figura 5.37 mostra que as categorias que obtiveram os *Lifts* mais altos são de

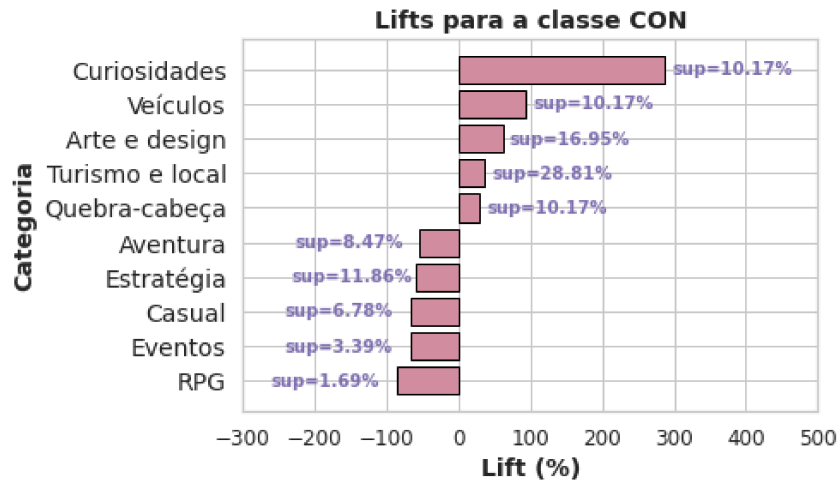


Figura 5.35: Os 5 maiores e 5 menores *Lifts* para a classe *Conscienciosidade* (sup=Suporte)

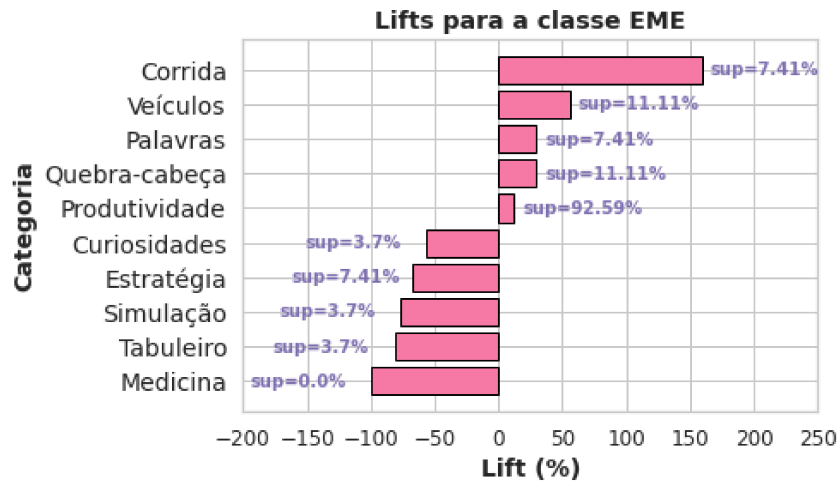


Figura 5.36: Os 5 maiores e 5 menores *Lifts* para a classe *Estabilidade Emocional* (sup=Suporte)

*Corrida* (173,08%) e *Esportes* (142,74%) para usuários que possuem o traço *Extroversão*. Os aplicativos das categorias *Eventos* e *Medicina* possuem *Lifts* -100% e suporte 0.0% mostrando que poucos usuários extrovertidos possuem aplicativos destas categorias.

### 5.5.3 Modelos Preditivos

Para validar a hipótese do trabalho, foram testadas duas abordagens: classificadores para prever se o usuário possui ou não o traço de personalidade e modelo de regressão para prever qual o nível daquele indivíduo em um determinado traço.

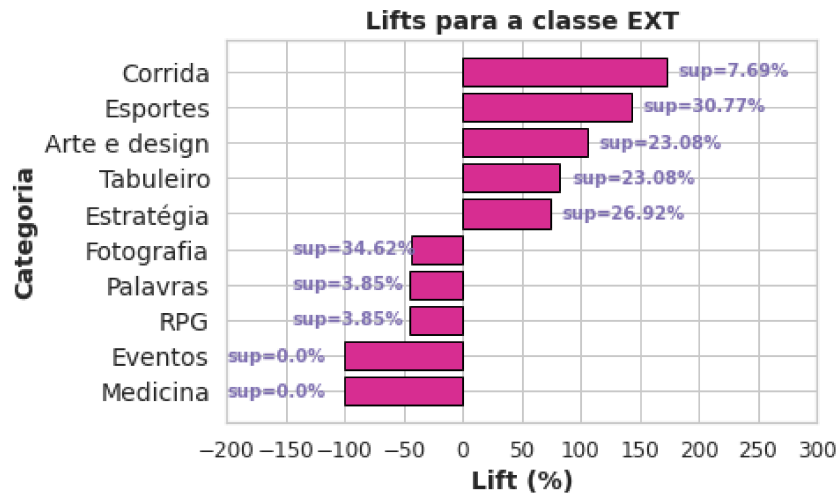


Figura 5.37: Os 5 maiores e 5 menores *Lifts* para a classe *Extroversão* (sup=Suporte)

### Modelos de Classificação

Para as entradas dos modelos de classificação, foi realizado um mapeamento para os traços de personalidade, sendo 1 caso o usuário possua nota maior ou igual a 5.0, e 0 caso contrário. Para corrigir o desbalanceamento da entrada, foi aplicada a técnica de *Undersampling*, que consiste em pegar amostras do maior grupo e deixá-lo com o mesmo tamanho do grupo menor. Para evitar que o uso desta técnica afete os resultados, todo o processo foi repetido 35 vezes para cada combinação de entrada. Além disso, a técnica PCA (Análise de Componentes Principais) foi aplicada para eliminar possíveis categorias que seriam prejudiciais à predição.

Afim de simplificar a discussão dos resultados, a combinação que utiliza o algoritmo *Naive Bayes*, componentes do PCA e aplicando *Undersampling*, foi escolhida por possuir, em geral, os melhores resultados de precisão, revocação e F-Score. A Figura 5.38 apresenta os resultados das métricas. A melhor precisão obtida foi no traço *Aberto a Novas Experiências*, alcançando 100%. Já para a métrica revocação o traço *Agradabilidade* foi o destaque. Ademais, os traços obtiveram, na média, resultados parecidos, com uma vantagem para o traço *Agradabilidade*.

### Modelo de Regressão

Na modelagem da regressão, o método escolhido foi o de Mínimos Quadrados Ordinários (OLS). Foi criado um modelo de regressão para cada traço de personalidade, em que a variável a ser predita é a nota (1 a 7) do usuário naquele traço. Para a entrada do modelo, foi considerado a quantidade de aplicativos instalados em cada uma das 38 categorias.

Alguns resultados sobre a regressão podem ser encontrados na Tabela 5.2. Pode-se perceber que o  $R^2$  está acima de 0,81 para todos os traços, o que indica que o modelo

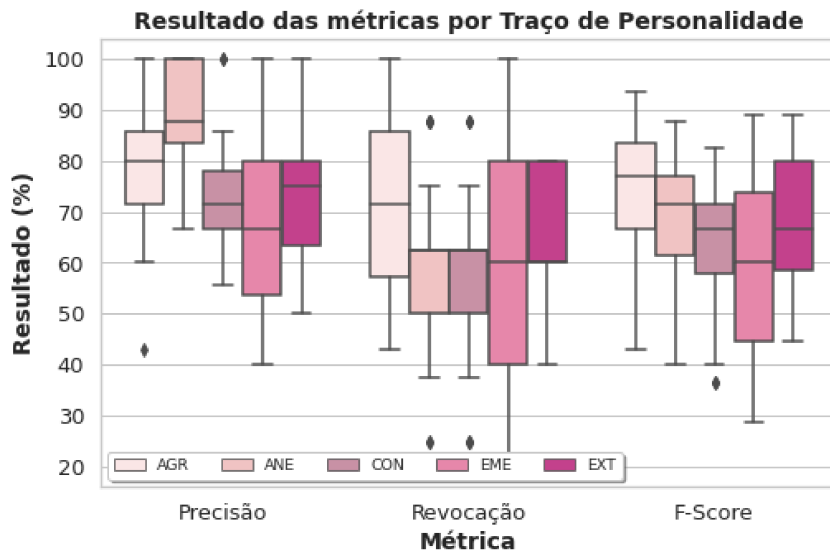


Figura 5.38: Resultados das métricas Precisão, Revocação e F-Score para os 5 traços de personalidade.

explica bem a variância da nota de personalidade se comparada com a média. Mesmo quando o  $R^2$  é ajustado de acordo com o número de variáveis, esse valor ainda é significativo, sendo acima de 0,68 para todos os traços. Por fim, pelos resultados do F-Test, a probabilidade abaixo de 0,05 indica que as variáveis predictoras utilizadas no modelo são relevantes para a estimativa da nota de personalidade com um nível de confiança de 95%.

Tabela 5.2: Resultado dos modelos de regressão para os traços de personalidade.

Traço	$R^2$	$R^2$ Ajus.	F-statistic	Prob (F-statistic)
AGR	0,833	0,726	7,748	$2,84 \times 10^{-12}$
ANE	0,857	0,764	9,281	$5,11 \times 10^{-14}$
CON	0,822	0,707	7,170	$1,50 \times 10^{-11}$
EME	0,811	0,689	6,664	$6,95 \times 10^{-11}$
EXT	0,824	0,711	7,266	$1,13 \times 10^{-11}$

Neste trabalho todos os traços de personalidade se mostraram positivamente associados a aplicativos da categoria Finanças (considerando um nível de confiança de 95%), isso indica que independentemente do traço, os usuários estão adotando a prática de *internet banking*, cartões virtuais e por pagamentos online. Todos os traços de personalidade se mostraram negativamente associados a aplicativos da categoria Estilo de Vida (considerando um nível de confiança de 95%), que incluem aplicativos de encontros (*Tinder*, *happn*, etc.), meditação, dieta, dentre outros. O possível motivo disto seria essa época de pandemia, onde as pessoas deixam de utilizar estes aplicativos de encontros e abandonam suas rotinas habituais devido a quarentena.

## 5.6 Perfil Completo MAPED

Como visto nas seções anteriores, é possível criar e analisar perfis de usuários de dispositivos móveis considerando diferentes perspectivas. Dessa maneira, foi elaborado um perfil completo que considera todos os perfis explorados neste trabalho, denominado Perfil MAPED (Mobilidade, Área funcional, Personalidade, Estabelecimento e Dispositivo). Assim, é possível aplicar esses modelos quando se conhece apenas a lista de aplicativos instalados e a uma localização dos usuários, para extrair informações capazes de auxiliar no entendimento do comportamento dos mesmos para resolução de diversos problemas.

O Perfil MAPED considera as cinco perspectivas, mesmo assim não necessita de muitos dados de entrada e da realização de vários enriquecimentos. Isso porque, são utilizados somente a lista de aplicativos e uma única localização referente a cidade do usuário. Assim, os aplicativos são enriquecidos com suas respectivas categorias (Personalizada ou fornecida pela *Play Store*) e a localização referente à cidade é enriquecida com os dados do IBGE. Desse modo, estes dados são utilizados como entradas para os melhores modelos apresentados em seus respectivos perfis. Assim, como saída o Perfil MAPED apresenta as probabilidades do usuário pertencer a cada uma das 24 classes dos perfis. Uma ilustração do Perfil MAPED é apresentada na Figura 5.39. Além disso, um exemplo de saída para três usuários é mostrado na Tabela 5.3.

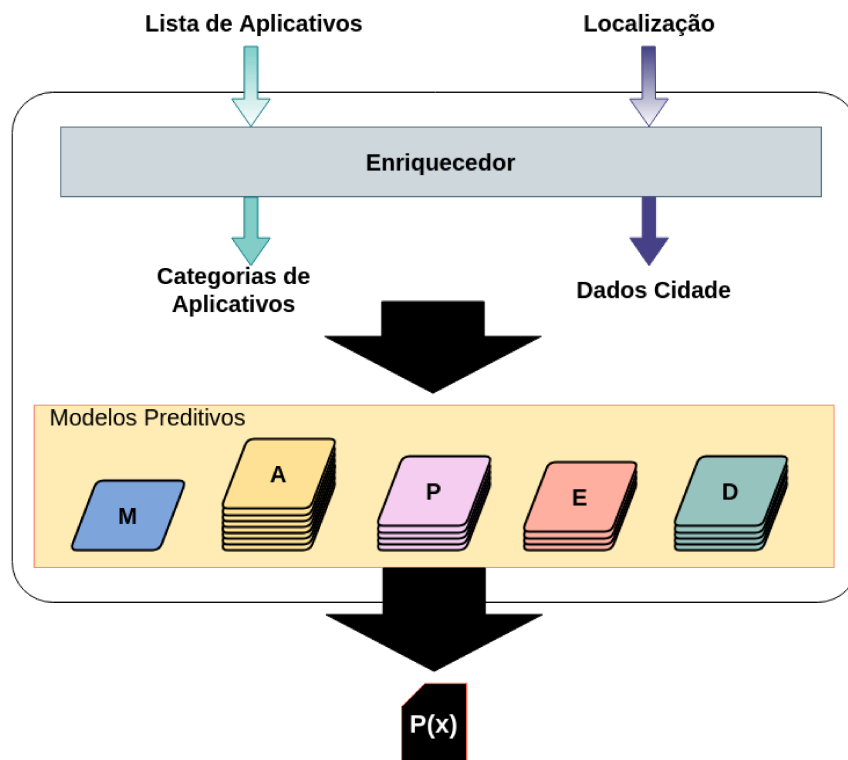


Figura 5.39: Ilustração da estrutura do Perfil MAPED.

Tabela 5.3: Exemplo de saída do Perfil MAPED.

Perfil	Classe	P(U1)	P(U2)	P(U3)
M	<i>Returner</i>	54,74	57,85	53,93
	<i>Explorer</i>	45,26	42,15	46,07
A	<i>Alimentação</i>	43,60	43,77	45,20
	<i>Companhia</i>	52,49	49,45	49,77
	<i>Educação</i>	52,03	47,80	46,25
	<i>Entretenimento</i>	61,84	53,40	48,65
	<i>Esportes</i>	50,49	50,38	50,04
	<i>Governo</i>	46,20	46,05	45,58
	<i>Shopping</i>	49,18	43,09	43,33
	<i>Turismo</i>	51,41	48,73	47,54
P	AGR	54,80	42,30	60,51
	ANE	51,58	45,45	57,30
	CON	39,09	65,80	38,65
	EME	32,52	50,34	47,24
	EXT	40,10	48,82	53,39
E	<i>Aeroporto</i>	59,91	57,43	37,21
	<i>Hospital</i>	57,05	49,12	37,64
	<i>Shopping</i>	55,99	52,10	45,55
	<i>Universidade</i>	51,56	64,25	41,53
D	<i>Entrada</i>	35,45	34,30	58,27
	<i>Intermediário</i>	42,80	43,65	54,94
	<i>Mid-high</i>	52,51	54,66	40,49
	<i>High-end</i>	60,05	62,05	30,43
	<i>Premium</i>	46,45	61,82	35,24

### 5.6.1 Análises e Resultados

Para análise do Perfil MAPED foram selecionados os usuários que pertenciam a 4 dos perfis estudados, totalizando 780. A exceção é somente o Perfil de Personalidade, pois somente ele utilizava um conjunto de usuários diferente, como explicado na Seção 5.5. Além disso, para as avaliações que serão realizadas o número de usuário deste perfil é restrito. Por estes motivos, não são analisados e discutidos os resultados considerando este perfil.

Seja  $R$  o conjunto dos 780 usuários,  $R_i$  é o subconjunto de todos os usuários que pertencem a classe  $i$ . Foi gerada aleatoriamente uma amostra  $P_i$  com 30 usuários que o respectivo modelo previu que pertenciam à classe  $i$ . Outra amostra aleatória, nomeada

$A_i$ , também com 30 usuários, foi retirada do conjunto  $R_i$  sem nenhum critério de filtragem. Assim, a métrica de *Lift* foi utilizada para avaliar a qualidade das previsões.

$$Lift_i = \frac{\frac{|P_i \cap R_i|}{|P_i|} - \frac{|A_i \cap R_i|}{|A_i|}}{\frac{|A_i \cap R_i|}{|A_i|}} = \frac{|P_i \cap R_i| - |A_i \cap R_i|}{|A_i \cap R_i|}$$

Sendo que,  $|P_i| = |A_i| = 30$ .

Todo o processo de geração das amostras foi repetido 35 vezes para cada classe  $i$ . A Figura 5.40 apresenta os valores obtidos para cada classe.

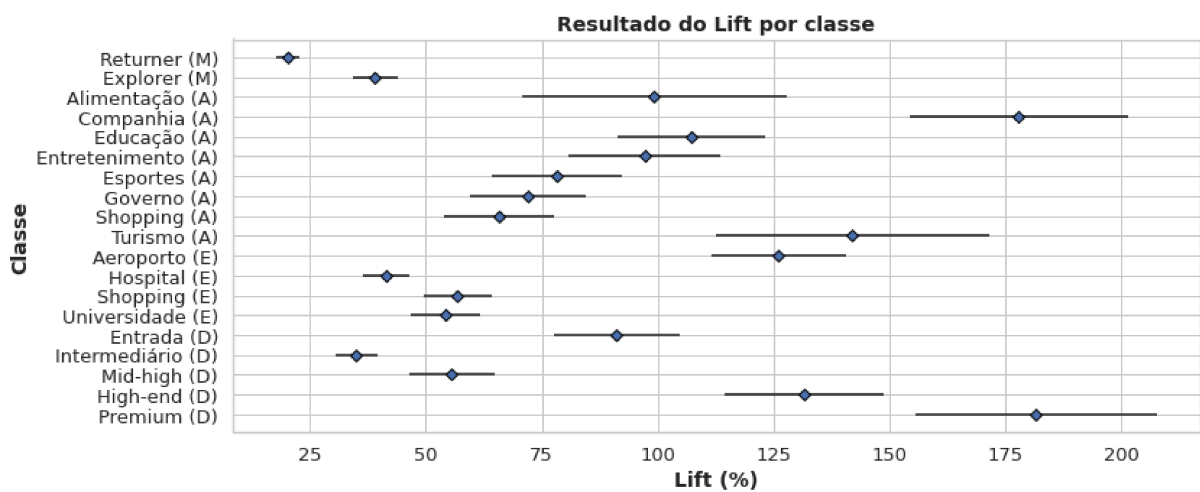


Figura 5.40: Resultado da média das execuções com intervalo de confiança de 95% do *Lift* para o Perfil MAPED.

A Figura 5.40 mostra que todas as classes obtiveram, na média das execuções, *Lifts* positivos. O melhor resultado foi para a classe *Premium* de preço do dispositivo. Este valor alto pode ser explicado pelo baixo número de usuários que realmente possuem aparelhos desta classe. Além disso, como pode ser visto na análise do lift, os valores foram positivos para a maioria das categorias de aplicativos. Já os valores baixos para as classes *returner* e *explorer* condizem com os resultados encontrados para o Perfil de Mobilidade. Além do mais, o número de usuários para estas classes estava bem distribuído, 441 para *returner* e 339 para *explorer*. Dessa maneira, não houve uma vantagem discrepante entre estas classes.

# Capítulo 6

## Conclusão

Esta dissertação investigou padrões do comportamento de usuários de dispositivos móveis em diferentes perspectivas. Os perfis apresentados neste trabalho necessitam somente da lista de aplicativos do usuário e uma única localização, quando necessária, para prever os padrões de interesse do usuário. Com as análises realizadas e resultados obtidos pelos modelos foi possível encontrar vários padrões.

Um destes padrões foi que usuários de dispositivos de entrada ou intermediário possuem significativamente menos aplicativos de categorias considerados mais restritos, como *Cia Aérea*, *Smart Things*, *Viagens* e *Eventos*. Por outro lado, o inverso é percebido para as classes *High-end* e *Premium*, em que um percentual muito maior dessas categorias (além de outras referentes a imóveis e finanças, por exemplo) é presente. Isso mostra que o tipo de dispositivo utilizado é um indicador relevante dos interesses e das classes sociais dos usuários.

Outra conclusão foi a de que usuários com PoIs em aeroportos possuem um aumento significativo na taxa de aplicativos das categorias relacionados a *Cias Aéreas* e *Voo*, o que indica usuários com um perfil viajante. O mesmo pode ser obtido com a categoria *Educação*, *Eventos* e *Transporte Público* e PoIs em Universidades. A mesma conclusão pode-se ter de *Medicina* e *Hospitais*.

Um número considerável de usuários que visitam áreas de companhias possuem aplicativos da categoria *Corporativos*, mostrando a necessidade destes usuários em realizar reuniões e acompanhar atividades. Já usuários com interesse em áreas de esportes possuem um aumento relevante em aplicativos de *Esportes*, mostrando a relação direta entre os mesmos. A mesma relação é encontrada entre áreas turísticas e aplicativos das categorias *Utilitários de Viagem* e *Cia Aérea Internacional*.

Usuários classificados como *returners* possuem uma taxa maior de aplicativos da categoria *Transporte Público*. Já os usuários *explorers*, possuem uma taxa maior de aplicativos de *Motorista*. Logo, é possível identificar a relação entre estes aplicativos e suas respectivas classes.

Já o Perfil de Personalidade ficou limitado pois foram necessárias as respostas dos usuários ao questionário *TUPI*. Dessa maneira, foram utilizados um número pequeno de usuários para que se possa generalizar estes resultados. Mesmo assim, foi possível ver que com base nos aplicativos instalados consegue-se prever os traços de

personalidade do usuário.

Por fim, o Perfil completo MAPED obteve resultados satisfatórios considerando as perspectivas analisadas. Com 95% de confiança é possível estipular o ganho em saber a classe do usuário em diferentes cenários, tendo alguns destes alcançado ganhos de mais de 200%. Com estes resultados, empresas e a comunidade científica podem melhorar o conhecimento à respeito do comportamento do usuário. Assim, é possível adaptar produtos e serviços à necessidade do usuário.

Outra contribuição deste trabalho foi o *MoreData*, um *framework* capaz de auxiliar no enriquecimento semântico de dados. O *MoreData* foi essencial para agilizar e simplificar o processo de enriquecimento de todos os perfis abordados. Além disso, também foi possível perceber com a aplicação, a fácil adaptação do *framework* para diferentes fontes e formatos de arquivos, independente de serem georreferenciados (e.g., dados de cidades e de estabelecimentos) ou não (e.g., categorias de aplicativos, preços dos aparelhos).

Baseado nos resultados alcançados, pretende-se utilizar um maior número de usuários para melhorar as análises. Outra melhoria seria quanto ao aumento do número de aplicativos categorizados. Também, explorar outras perspectivas que possam complementar o perfil MAPED. Ademais, testar o *framework* para outros casos de uso e, com isso, desenvolver mais conectores. Além disso, planeja-se paralelizar o processo de enriquecimento para torná-lo mais eficiente.

## 6.1 Publicações

Além do trabalho apresentado nesta dissertação, durante o mestrado foram desenvolvidos outros projetos em cooperação com outros pesquisadores. O primeiro deles foi [Souza et al. \(2020\)](#) onde são avaliados os impactos de dados demográficos na recomendação de aplicativos. Outra participação foi a escrita de um capítulo de livro sobre dados geoespaciais ([Domingues et al., 2020](#)). Este capítulo aborda vários aspectos importantes (e.g., tipos, coleta, armazenamento, visualização) considerando este tipo de dado. Por fim, foi publicado o trabalho [Souza et al. \(2021\)](#) sobre um modelo híbrido para o problema *indoor/outdoor detection*.

## Referências Bibliográficas

- Abel, F., Gao, Q., Houben, G.-J., and Tao, K. (2011). Semantic enrichment of twitter posts for user profile construction on the social web. In *Extended semantic web conference*, pages 375–389. Springer.
- Alvarez-Lozano, J., Osmani, V., Mayora, O., Frost, M., Bardram, J., Faurholt-Jepsen, M., and Kessing, L. V. (2014). Tell me your apps and i will tell you your mood: correlation of apps usage with bipolar disorder state. In *Proceedings of the 7th International Conference on Pervasive Technologies Related to Assistive Environments*, pages 1–7.
- Barbosa, H., Barthelemy, M., Ghoshal, G., James, C. R., Lenormand, M., Louail, T., Menezes, R., Ramasco, J. J., Simini, F., and Tomasini, M. (2018). Human mobility: Models and applications. *Physics Reports*, 734:1–74.
- Brodsky, I. (2019). H3: Uber’s Hexagonal Hierarchical Spatial Index. *Uber Engineering Blog*.
- Capanema, C., Silva, F. A., and Braga, T. M. (2019). Identificação e classificação de pontos de interesse individuais com base em dados esparsos. In *Anais Principais do XXXVII Simpósio Brasileiro de Redes de Computadores e Sistemas Distribuídos*, pages 15–28. SBC.
- Do, T.-M.-T. and Gatica-Perez, D. (2010). By their apps you shall understand them: mining large-scale patterns of mobile phone usage. In *Proceedings of the 9th international conference on mobile and ubiquitous multimedia*, pages 1–10.
- Domingues, A., Silva, F., Santos, L., Souza, R., Coimbra, G., and Loureiro, A. A. F. (2020). Dados geoespaciais: Conceitos e técnicas para coleta, armazenamento, tratamento e visualização. *Sociedade Brasileira de Computação*.
- Doyle, K. O. and Youn, S. (2000). Exploring the traits of happy people. In *Social Indicators Research*.
- Facebook (2021). Places Graph - Documentação - Facebook for Developers. [Online; accessed 10. Jan. 2021].

- Farrahi, K. and Gatica-Perez, D. (2011). Discovering routines from large-scale human locations using probabilistic topic models. *ACM Transactions on Intelligent Systems and Technology (TIST)*, 2(1):1–27.
- Gokul Chittaranjan, J. B. and Gatica-Perez, D. (2011). Who’s who with big-five: Analyzing and classifying personality traits with smartphones. In *15th Annual International Symposium on Wearable Computers*, pages 29–36.
- Google (2020). Overview | Places API | Google Developers. [Online; accessed 10. Jan. 2021].
- Gosling, S. D., Rentfrow, P. J., and Swann Jr, W. B. (2003). A very brief measure of the big-five personality domains. *Journal of Research in personality*, 37(6):504–528.
- Hung, C.-C., Chang, C.-W., and Peng, W.-C. (2009). Mining trajectory profiles for discovering user communities. In *Proceedings of the 2009 International Workshop on Location Based Social Networks*, pages 1–8.
- IBGE (2020). [Online; accessed 5. Nov. 2020].
- Ipea (2020). Ipea. [Online; accessed 3. Jul. 2020].
- Jianqiang Shen, O. B. and Liu, J. (2014). A study of facebook behavior: What does it tell about your neuroticism and extraversion? In *Computers in Human Behavior*.
- Li, Q., Zheng, Y., Xie, X., Chen, Y., Liu, W., and Ma, W.-Y. (2008). Mining user similarity based on location history. In *Proceedings of the 16th ACM SIGSPATIAL international conference on Advances in geographic information systems*, pages 1–10.
- Maia, W., Silva, F., and Silva, T. (2020). Um estudo sobre a relação entre smartphones e dados demográficos. In *Anais do IV Workshop de Computação Urbana*, pages 302–315, Porto Alegre, RS, Brasil. SBC.
- Malmi, E. and Weber, I. (2016). You are what apps you use: Demographic prediction based on user’s apps. *arXiv preprint arXiv:1603.00059*.
- Maynard, D. and Bontcheva, K. (2015). Understanding climate change tweets: an open source toolkit for social media analysis. In *EnviroInfo and ICT for Sustainability 2015*. Atlantis Press.
- McCrae, R. R. and Jr., P. T. C. (1999). A five-factor theory of personality. In *Handbook of personality: Theory and research*, pages 139–153.
- McInerney, J., Stein, S., Rogers, A., and Jennings, N. R. (2013). Breaking the habit: Measuring and predicting departures from routine in individual human mobility. *Pervasive and Mobile Computing*, 9(6):808–822.

- Medeiros, H. (2019). Faturamento com smartphones cresce 6% no Brasil e alcança R\$ 58 bilhões em 2018 - Mobile Time. [Online; accessed 8. Jan. 2021].
- Naboulsi, D., Fiore, M., Ribot, S., and Stanica, R. (2016). Large-scale mobile traffic analysis: a survey. *IEEE Communications Surveys & Tutorials*, 18(1):124–161.
- OpenStreetMap (2021). Overpass API/Overpass QL - OpenStreetMap Wiki. [Online; accessed 10. Jan. 2021].
- Pappalardo, L., Simini, F., Rinzivillo, S., Pedreschi, D., Giannotti, F., and Barabási, A.-L. (2015). Returners and explorers dichotomy in human mobility. *Nature communications*, 6(1):1–8.
- Raihana Ferdous, V. O. and Mayora, O. (2015). Smartphone app usage as a predictor of perceived stress levels at workplace. In *International Conference on Pervasive Computing Technologies for Healthcare*.
- Runhua Xu, Remo M. Frey, D. V. and Ilic, A. (2015). Towards understanding the impact of personality traits on mobile app adoption - a scalable approach. In *European Conference on Information Systems*.
- Runhua Xu, Remo Manuel Frey, E. F. and Ilic, A. (2016). Understanding the impact of personality traits on mobile app adoption – insights from a large-scale field study. In *Computers in Human Behavior*, pages 244–256.
- Ryan, T. and Xenos, S. (2011). Who uses facebook? an investigation into the relationship between the big five, shyness, narcissism, loneliness, and facebook usage. In *Computers in Human Behavior*.
- Santin, P., Gubert, F. R., Fonseca, M., Munaretto, A., and Silva, T. H. (2020). Characterization of public transit mobility patterns of different economic classes. *Sustainability*, 12(22):9603.
- Sardianos, C., Varlamis, I., and Bouras, G. (2018). Extracting user habits from google maps history logs. In *2018 IEEE/ACM International Conference on Advances in Social Networks Analysis and Mining (ASONAM)*, pages 690–697. IEEE.
- Seneviratne, S., Seneviratne, A., Mohapatra, P., and Mahanti, A. (2015). Your installed apps reveal your gender and more! *ACM SIGMOBILE Mobile Computing and Communications Review*, 18(3):55–61.
- Silva, I. B. and Nakano, T. C. (2011). Modelo dos cinco grandes fatores da personalidade: Análise de pesquisas. In *Avaliação Psicológica*, pages 51–62.

- Souza, R., Santos, L., Silva, M., Silva, F., and Silva, T. (2020). Impacto do uso de informações demográficas para a recomendação de aplicativos móveis. In *Anais do XII Simpósio Brasileiro de Computação Ubíqua e Pervasiva*, pages 111–120. SBC.
- Souza, R. P., dos Santos, L. J., Coimbra, G. T., Silva, F. A., and Silva, T. R. (2021). A big data-driven hybrid solution to the indoor-outdoor detection problem. *Big Data Research*, 24:100194.
- Varlamis, I., Sardanios, C., and Bouras, G. (2020). Mining habitual user choices from google maps history logs. In *Putting Social Media and Networking Data in Practice for Education, Planning, Prediction and Recommendation*, pages 151–175. Springer.
- Viana, A. C., Luzio, A. D., Jaffrès-Runser, K., Stefa, A. M., and Julinda (2018). Accurately inferring personality traits from the use of mobile technology. In *HAL Archives Ouvertes*.
- Wu, F., Li, Z., Lee, W.-C., Wang, H., and Huang, Z. (2015). Semantic annotation of mobility data using social media. In *Proceedings of the 24th International Conference on World Wide Web*, pages 1253–1263.
- Yu, D., Li, Y., Xu, F., Zhang, P., and Kostakos, V. (2018). Smartphone app usage prediction using points of interest. *Proceedings of the ACM on Interactive, Mobile, Wearable and Ubiquitous Technologies*, 1(4):1–21.
- Yusong Gao, H. L. and Zhu, T. (2014). Predicting subjective well-being by smartphone usage behaviors. In *International Conference on Health Informatics*.
- Yves-Alexandre de Montjoye, Jordi Quoidbach, F. R. and Pentland, A. S. (2013). Predicting personality using novel mobile phone-based metrics. In *SBP 2013*, pages 48–55.

# Apêndice A

## Aplicativos e suas Categorias

Tabela A.1: Aplicativos e suas categorias

	<b>Categoria</b>	<b>Aplicativo</b>
1	Bancos	br.com.banestes.rmb.mobile
2	Bancos	br.com.banrisul
3	Bancos	br.com.bb.android
4	Bancos	br.com.bb.oewallet
5	Bancos	br.com.gabba.Caixa
6	Bancos	br.com.santander.way
7	Bancos	br.com.sicoobnet
8	Bancos	br.gov.caixa.cartoes
9	Bancos	com.bradesco
10	Bancos	com.bradesco.exclusive
11	Bancos	com.bradesco.prime
12	Bancos	com.credicard.app
13	Bancos	com.itaou
14	Bancos	com.itaou.aco
15	Bancos	com.itaou.applight
16	Bancos	com.itaou.empresas
17	Bancos	com.itaou.pers
18	Bancos	com.itaucard.activity
19	Bancos	com.santander.app
20	Bancos Digitais	br.com.agipag.app
21	Bancos Digitais	br.com.bradesco.next
22	Bancos Digitais	br.com.digio
23	Bancos Digitais	br.com.intermedium
24	Bancos Digitais	br.com.neon
25	Bancos Digitais	br.com.original.bank
26	Bancos Digitais	br.com.original.reip
27	Bancos Digitais	br.com.uol.ps.myaccount

**Tabela A.1 Aplicativos e suas Categorias (continuação)**

	<b>Categoria</b>	<b>Aplicativo</b>
28	Bancos Digitais	com.coinbase.android
29	Bancos Digitais	com.nu.production
30	Carteira Digital	com.b2winc.amedigitalbr.com.meliuz
31	Carteira Digital	com.google.android.apps.walletnfcrel
32	Carteira Digital	com.mercadopago.wallet
33	Carteira Digital	com.paypal.android.p2pmobile
34	Carteira Digital	com.picpay
35	Carteira Digital	com.samsung.android.spay
36	Cia Aérea	br.com.avianca
37	Cia Aérea	br.com.edeploy.gol.checkin.activities
38	Cia Aérea	br.com.voeazul
39	Cia Aérea	mobile.latam.com.latamapp
40	Cia Aérea Internacional	com.aa.android
41	Cia Aérea Internacional	com.aeromexico.app
42	Cia Aérea Internacional	com.afklm.mobile.android.gomobile.klm
43	Cia Aérea Internacional	com.aircanada
44	Cia Aérea Internacional	com.airfrance.android.dinamoprd
45	Cia Aérea Internacional	com.alitalia.mobile
46	Cia Aérea Internacional	com.ba.mobile
47	Cia Aérea Internacional	com.delta.mobile.android
48	Cia Aérea Internacional	com.emirates.ek.android
49	Cia Aérea Internacional	com.ethiopianairlines.ethiopianairlines
50	Cia Aérea Internacional	com.flysaa.mobile.android
51	Cia Aérea Internacional	com.lufthansa.android.lufthansa
52	Cia Aérea Internacional	com.m.qr
53	Cia Aérea Internacional	com.megasis.android
54	Cia Aérea Internacional	com.mttnow.android.copa.production
55	Cia Aérea Internacional	com.mttnow.android.etihad
56	Cia Aérea Internacional	com.royalairmaroc.app
57	Cia Aérea Internacional	com.rytong.airchina
58	Cia Aérea Internacional	com.united.mobile.android
59	Cia Aérea Internacional	com.wow.taagcargo
60	Cia Aérea Internacional	Kal.FlightInfo
61	Cia Aérea Internacional	net.aper.ARMobile
62	Corporativos	appear.in.app
63	Corporativos	br.com.zeropaper.app

Tabela A.1 Aplicativos e suas Categorias (continuação)

	<b>Categoria</b>	<b>Aplicativo</b>
64	Corporativos	com.asana.app
65	Corporativos	com.dropbox.android
66	Corporativos	com.facebook.work
67	Corporativos	com.fiverr.fiverr
68	Corporativos	com.google.android.apps.giant
69	Corporativos	com.google.android.apps.meetings
70	Corporativos	com.google.android.apps.vega
71	Corporativos	com.google.android.keep
72	Corporativos	com.google.android.talk
73	Corporativos	com.indeed.android.jobsearch
74	Corporativos	com.infojobs.phone
75	Corporativos	com.kaching.merchant
76	Corporativos	com.linkedin.android.jobs.jobseeker
77	Corporativos	com.microsoft.office.lync15
78	Corporativos	com.microsoft.teams
79	Corporativos	com.pipedrive
80	Corporativos	com.salesforce.chatter
81	Corporativos	com.skype.m2
82	Corporativos	com.skype.raider
83	Corporativos	com.trovit.android.apps.jobs
84	Corporativos	com.tysdoapp.beerorcoffee
85	Corporativos	net.infojobs.mobile.androidcom.glassdoor.app
86	Corporativos	net.slideshare.mobile
87	Corporativos	org.wordpress.android
88	Corporativos	us.zoom.videomeetings
89	Delivery	br.com.brainweb.ifood
90	Delivery	br.com.jamesdelivery.android.james
91	Delivery	com.cerveceriamodelo.modelonow
92	Delivery	com.glovo
93	Delivery	com.grability.rappi
94	Delivery	com.ubercab.eats
95	E-Commerce	br.com.buscapes.MainPack
96	E-Commerce	br.com.cea.appb2c
97	E-Commerce	br.com.clubeextra
98	E-Commerce	br.com.dafiti
99	E-Commerce	br.com.kanui

Tabela A.1 Aplicativos e suas Categorias (continuação)

	<b>Categoria</b>	<b>Aplicativo</b>
100	E-Commerce	br.com.lojasrenner
101	E-Commerce	br.com.mobits.peixurbano
102	E-Commerce	br.com.netshoes.app
103	E-Commerce	br.com.pernambucanas.varejo.android
104	E-Commerce	br.com.riachuelo.app
105	E-Commerce	br.com.sbf.centauro
106	E-Commerce	br.com.viewit.xcommerce.ricardoeletro
107	E-Commerce	br.com.walmart.app
108	E-Commerce	br.com.zattini
109	E-Commerce	com.alibaba.aliexpresshd
110	E-Commerce	com.amazon.mShop.android.shopping
111	E-Commerce	com.amazon.windowshop
112	E-Commerce	com.artech.eaplic.havan
113	E-Commerce	com.b2w.americanas
114	E-Commerce	com.b2w.shoptime
115	E-Commerce	com.b2w.submarino
116	E-Commerce	com.boticario.mobile
117	E-Commerce	com.contextlogic.geek
118	E-Commerce	com.contextlogic.wish
119	E-Commerce	com.ebay.mobilecom.alibaba.intl.android.apps.poseidon
120	E-Commerce	com.epro.dx
121	E-Commerce	com.g2a.marketplace
122	E-Commerce	com.inditex.zara
123	E-Commerce	com.luizalabs.mlapp
124	E-Commerce	com.mercadolibre
125	E-Commerce	com.novapontocom.casasbahia
126	E-Commerce	com.schibsted.bomnegocio.androidApp
127	E-Commerce	com.sonyericsson.pontofrio
128	E-Commerce	com.sonyericsson.pontofriobr.com.kanui
129	E-Commerce	com.zoom.zoomandroid
130	Edição de Vídeo e Foto	com.adobe.lrmobile
131	Edição de Vídeo e Foto	com.adobe.psmobile
132	Edição de Vídeo e Foto	com.duapps.recorder
133	Edição de Vídeo e Foto	com.jsdev.instasize
134	Edição de Vídeo e Foto	com.kwai.video

Tabela A.1 Aplicativos e suas Categorias (continuação)

	<b>Categoria</b>	<b>Aplicativo</b>
135	Edição de Vídeo e Foto	com.linecorp.b612.android
136	Edição de Vídeo e Foto	com.lomotif.android
137	Edição de Vídeo e Foto	com.magisto
138	Edição de Vídeo e Foto	com.nexstreaming.app.kinemasterfree
139	Edição de Vídeo e Foto	com.niksoftware.snapseed
140	Edição de Vídeo e Foto	com.picsart.studio
141	Edição de Vídeo e Foto	com.pixlr.express
142	Edição de Vídeo e Foto	com.quvideo.xiaoying
143	Edição de Vídeo e Foto	com.roidapp.photogrid
144	Edição de Vídeo e Foto	com.stupeflix.replay
145	Edição de Vídeo e Foto	com.venticake.retrica
146	Edição de Vídeo e Foto	com.vSCO.cam
147	Edição de Vídeo e Foto	com.wondershare.filmorago
148	Edição de Vídeo e Foto	video.like
149	Educação	br.com.alura.lingua
150	Educação	br.com.alura.mobi
151	Educação	br.com.alura.startco.unlockyourbrain
152	Educação	br.com.descomplica.vod
153	Educação	br.com.passeidireto
154	Educação	com.babbel.mobile.android.en
155	Educação	com.bagatrix.mathway.android
156	Educação	com.br.twelvemin
157	Educação	com.busuu.android.enc
158	Educação	com.cambly.cambly
159	Educação	com.classdojo.android
160	Educação	com.duolingo
161	Educação	com.fusionprojects.edmodo
162	Educação	com.google.android.apps.classroom
163	Educação	com.google.android.apps.forscience.whistlepunk
164	Educação	com.google.android.apps.primer
165	Educação	com.google.android.spotlightstories
166	Educação	com.lumoslabs.lumosity
167	Educação	com.malmath.apps.mm
168	Educação	com.memrise.android.memrisecompanion
169	Educação	com.microblink.photomath
170	Educação	com.newsoftware.gabaritar

Tabela A.1 Aplicativos e suas Categorias (continuação)

	<b>Categoria</b>	<b>Aplicativo</b>
171	Educação	com.overdrive.mobile.android.libby
172	Educação	com.quizlet.quizletandroid
173	Educação	com.remind101
174	Educação	com.udemy.android
175	Educação	com.wlingua.curso
176	Educação	com.yousician.yousician
177	Educação	daldev.android.gradehelper
178	Educação	org.brilliant.android
179	Educação	org.coursera.android
180	Educação	org.edx.mobile
181	Educação	org.geogebra.android
182	Educação	org.khanacademy.android
183	Educação	org.socratic.android
184	Educação	us.mathlab.android
185	Esportes	atpwta.live
186	Esportes	br.com.golmobile.placaruo1
187	Esportes	br.com.mobits.cartolafc
188	Esportes	br.tv.horizonte.android.premierefc
189	Esportes	com.chwimaconcept.betpredict
190	Esportes	com.dazn
191	Esportes	com.dfstream.nfl
192	Esportes	com.dorna.officialmotogp
193	Esportes	com.fic.foxsports
194	Esportes	com.fourtaps.brpro
195	Esportes	com.fourtaps.libpro
196	Esportes	com.gcteam.toptvhd
197	Esportes	com.globo.ge.app
198	Esportes	com.gotv.nflgamecenter.us.lite
199	Esportes	com.nbaimd.gametime.nba2011
200	Esportes	com.nousguide.android.rbtv
201	Esportes	com.softpauer.f1timingapp2014.basic
202	Esportes	com.uefa.ucl
203	Esportes	com.uefa.uel
204	Esportes	com.wsl.android
205	Esportes	net.feelgoodapps.betpredictor
206	Eventos	com.evenyx.ingressoprime

Tabela A.1 Aplicativos e suas Categorias (continuação)

	<b>Categoria</b>	<b>Aplicativo</b>
207	Eventos	com.goodbarber.brasilgameshow
208	Eventos	com.greetingsisland.sam
209	Eventos	com.ingresse.ticketboothcom.sympla.tickets
210	Eventos	com.ingressolive
211	Eventos	com.nproducoes
212	Eventos	com.sympla.organizer
213	Eventos	outgo.android
214	Fast Food	burgerking.com.br.appandroid
215	Fast Food	com.mcdo.mcdonalds
216	Finanças	br.com.gerenciadorfinanceiro.controller
217	Finanças	br.com.guiabolso
218	Finanças	br.com.meubolsoemdia.Jimbo
219	Finanças	com.bookmark.money
220	Finanças	com.disrapp.coinkeeper.material
221	Finanças	com.ecda.wisecash
222	Finanças	com.minhaseconomias
223	Finanças	com.organizze.android
224	Finanças	com.Splitwise.SplitwiseMobile
225	Finanças	mic.app.gastosdiarios_licencia
226	Financeiras	br.com.creditas.mobile
227	Financeiras	br.com.crefisa.crefisa
228	Financeiras	br.com.noverde.noverde
229	Financeiras	com.enova.simplic
230	Financeiras	com.mutual.app
231	Fitness	br.com.livetouch.nutrabem
232	Fitness	br.com.tadashi.appmeditacao
233	Fitness	com.azumio.android.sleeptime
234	Fitness	com.brainbow.rise.app
235	Fitness	com.buscaalimento.android
236	Fitness	com.clue.android
237	Fitness	com.dailyyoga.inc
238	Fitness	com.endomondo.android
239	Fitness	com.eokoe.smartfitcoach
240	Fitness	com.fatsecret.android
241	Fitness	com.fitbit.FitbitMobile
242	Fitness	com.fitnesskeeper.runkeeper.pro

Tabela A.1 Aplicativos e suas Categorias (continuação)

	<b>Categoria</b>	<b>Aplicativo</b>
243	Fitness	com.garmin.android.apps.connectmobile
244	Fitness	com.google.android.apps.fitness
245	Fitness	com.mapmyrun.android2com.stt.android
246	Fitness	com.myfitnesspal.android
247	Fitness	com.nike.ntc
248	Fitness	com.nike.plusgps
249	Fitness	com.northcube.sleepcycle
250	Fitness	com.runtastic.android
251	Fitness	com.runtastic.android.pro2
252	Fitness	com.sec.android.app.shealthcom.northpark.drink-water
253	Fitness	com.sportractive
254	Fitness	com.strava
255	Fitness	com.weightwatchers.mobile
256	Fitness	homeworkout.homeworkouts.noequipment
257	Fitness	net.aspbrasil.keer.core.receitaslight
258	Hospedagem	br.com.hotelurbano
259	Hospedagem	br.com.mobilesaude.hotelaria
260	Hospedagem	com.accor.appli.hybrid
261	Hospedagem	com.airbnb.android
262	Hospedagem	com.booking
263	Hospedagem	com.hcom.android
264	Hospedagem	com.trivago
265	Imóveis	br.com.lopes.moving
266	Imóveis	br.com.quintoandar.inquilinos
267	Imóveis	br.com.zap.imoveis
268	Imóveis	com.minasapp.minhacasaminhaveda
269	Imóveis	com.mitula.homes
270	Imóveis	com.project.vivareal
271	Imóveis	com.wave.mcmv
272	Imóveis	es.roid.and.trovit
273	Imóveis	maya.im.imovelweb
274	Infantil	br.com.zeroum.patatipatata
275	Infantil	br.com.zeroum.turmadagalinha
276	Infantil	com.babysparks.babysparks
277	Infantil	com.bible.kids

Tabela A.1 Aplicativos e suas Categorias (continuação)

	<b>Categoria</b>	<b>Aplicativo</b>
278	Infantil	com.google.android.apps.youtube.kids
279	Infantil	family.tracker.my
280	Infantil	kr.co.smartstudy.babyshark_android_googlemarket
281	Infantil	pink.baby.phone.kid.music.songs.animal.princess .numbersorg.findmykids.app
282	Investimento	br.com.bradesco.trading
283	Investimento	br.com.clear.droid
284	Investimento	br.com.easynvest.rendafixa
285	Investimento	br.com.itaucorretorapro
286	Investimento	br.com.rico.mobile
287	Investimento	br.com.santander.mobilebroker
288	Investimento	br.com.xp.carteira
289	Investimento	br.gov.fazenda.tesouro.td
290	Investimento	com.binance.dev
291	Investimento	com.bloomberg.android.plus
292	Investimento	com.fusionmedia.investing
293	Investimento	com.iqoption
294	Investimento	com.italy.broker
295	Investimento	com.teletrader.android
296	Investimento	com.ticno.olymptrade
297	Investimento	net.metaquotes.metatrader4
298	Investimento	net.metaquotes.metatrader5
299	Jogos	com.activision.callofduty.shooter
300	Jogos	com.blizzard.wtcg.hearthstone
301	Jogos	com.dts.freefireth
302	Jogos	com.ea.game.pvzfree_row
303	Jogos	com.ea.game.simpsons4_row
304	Jogos	com.fingersoft.benjibananas
305	Jogos	com.fingersoft.hillclimb
306	Jogos	com.gameloft.android.ANMP.GloftA8HM
307	Jogos	com.gameloft.android.ANMP.GloftDMHM
308	Jogos	com.halfbrick.fruitninja
309	Jogos	com.kiloo.subwaysurf
310	Jogos	com.ludo.king
311	Jogos	com.mattel.jv.uno
312	Jogos	com.miniclip.plagueinc

Tabela A.1 Aplicativos e suas Categorias (continuação)

	<b>Categoria</b>	<b>Aplicativo</b>
313	Jogos	com.netmarble.mherosgb
314	Jogos	com.nianticlabs.pokemongo
315	Jogos	com.outfit7.mytalkingangelafree
316	Jogos	com.outfit7.mytalkingtomfree
317	Jogos	com.roblox.client
318	Jogos	com.robtopx.geometryjumplite
319	Jogos	com.rovio.angrybirds
320	Jogos	com.scimob.ninetyfour.percent
321	Jogos	com.skgames.trafficrider
322	Jogos	com.supercell.clashofclans
323	Jogos	com.supercell.hayday
324	Jogos	com.zeptolab.ctr.ads
325	Jogos	com.zeptolab.ctr2.f2p.googlejp.konami.duellinks
326	Jogos	logos.quiz.companies.game
327	Jogos	me.pou.app
328	Jogos	net.mobigame.zombietsunami
329	Jogos	uk.co.aifactory.chessfree
330	Lazer	br.com.sbt.app
331	Lazer	br.com.sky.selfcare
332	Lazer	com.amazon.kindle
333	Lazer	com.amazon.klite
334	Lazer	com.discord
335	Lazer	com.globo.app
336	Lazer	com.globo.globotv
337	Lazer	com.imdb.mobile
338	Lazer	com.imdb.pro.mobile.android
339	Lazer	com.imgur.mobile
340	Lazer	com.moviecity.app
341	Lazer	com.nickappintl.android.nickelodeon
342	Lazer	com.ninegag.android.app
343	Lazer	com.playstation.mobilemessenger
344	Lazer	com.reddit.frontpage
345	Lazer	com.scee.psxandroid
346	Lazer	com.ted.android
347	Lazer	com.valvesoftware.android.steam.community
348	Medicina	br.gov.datasus.cnsdigital

Tabela A.1 Aplicativos e suas Categorias (continuação)

	<b>Categoria</b>	<b>Aplicativo</b>
349	Medicina	com.baviux.pillreminder
350	Medicina	com.despdev.weight_loss_calculator
351	Medicina	com.healint.migraineapp
352	Medicina	com.herzberg.easyquitsmoking
353	Medicina	com.ladytimer.ovulationcalendar
354	Medicina	com.lbrc.PeriodCalendar
355	Medicina	com.macropinch.hydra.android
356	Medicina	com.medscape.android
357	Medicina	com.mysugr.android.companion
358	Medicina	com.smsrobot.period
359	Medicina	com.ufo.disease
360	Medicina	eu.smartpatient.mytherapy
361	Medicina	org.iggymedia.periodtracker
362	Medicina	ru.mobiledimension.kbr
363	Mobilidade Urbana	br.com.easytaxi
364	Mobilidade Urbana	com.cabify.rider
365	Mobilidade Urbana	com.goeuro.rosie
366	Mobilidade Urbana	com.google.android.apps.maps
367	Mobilidade Urbana	com.ridewith
368	Mobilidade Urbana	com.taxis99
369	Mobilidade Urbana	com.ubercab
370	Mobilidade Urbana	com.ubercab.uberlite
371	Mobilidade Urbana	com.waze
372	Mobilidade Urbana	me.lyft.android
373	Mobilidade Urbana	sinet.startup.inDriver
374	Motorista	br.com.easytaxista
375	Motorista	com.app99.driver
376	Motorista	com.cabify.driver
377	Motorista	com.ubercab.driver
378	Produtividade	com.adobe.reader
379	Produtividade	com.adobe.scan.android
380	Produtividade	com.anydo
381	Produtividade	com.appxy.tinyscanner
382	Produtividade	com.autodesk.autocadws
383	Produtividade	com.box.android
384	Produtividade	com.cloudmagic.mail

Tabela A.1 Aplicativos e suas Categorias (continuação)

	<b>Categoria</b>	<b>Aplicativo</b>
385	Produtividade	com.evernote
386	Produtividade	com.ifttt.ifttt
387	Produtividade	com.intsig.camscanner
388	Produtividade	com.lastpass.lpandroid
389	Produtividade	com.pushbullet.android
390	Produtividade	com.Slack
391	Produtividade	com.teamviewer.quicksupport.market
392	Produtividade	com.teamviewer.teamviewer.market.mobile
393	Produtividade	com.todoist
394	Produtividade	com.trello
395	Produtividade	mega.privacy.android.app
396	Redes Sociais	com.bitstrips.imoji
397	Redes Sociais	com.facebook.katana
398	Redes Sociais	com.facebook.lite
399	Redes Sociais	com.facebook.mlite
400	Redes Sociais	com.facebook.orca
401	Redes Sociais	com.google.android.apps.tachyon
402	Redes Sociais	com.instagram.android
403	Redes Sociais	com.linkedin.android
404	Redes Sociais	com.mobilemotion.dubsmash
405	Redes Sociais	com.pinterest
406	Redes Sociais	com.snapchat.android
407	Redes Sociais	com.tencent.mm
408	Redes Sociais	com.tumblr
409	Redes Sociais	com.twitter.android
410	Redes Sociais	com.twitter.android.lite
411	Redes Sociais	com.viber.voip
412	Redes Sociais	com.vimeo.android.videoapp
413	Redes Sociais	com.whatsapp
414	Redes Sociais	com.whatsapp.w4b
415	Redes Sociais	com.zhiliaoapp.musically
416	Redes Sociais	com.zhiliaoapp.musically.go
417	Redes Sociais	kik.android
418	Redes Sociais	org.telegram.messenger
419	Relacionamento	com.badoo.mobile
420	Relacionamento	com.bumble.app

Tabela A.1 Aplicativos e suas Categorias (continuação)

	<b>Categoria</b>	<b>Aplicativo</b>
421	Relacionamento	com.ftw_and_co.happn
422	Relacionamento	com.grindrapp.android
423	Relacionamento	com.hotornot.app
424	Relacionamento	com.jaumo.casual
425	Relacionamento	com.okcupid.okcupid
426	Relacionamento	com.tinder
427	Relacionamento	net.lovoe.android
428	Renda Complementar	air.com.avon.mobile.AvonMobile
429	Renda Complementar	br.com.app.gpu1964804.gpue731837757c8d4116c1a-439f4335cd57
430	Renda Complementar	br.com.avon.mobile_commerce
431	Renda Complementar	br.com.grupoboticario.boticariorepresentante
432	Renda Complementar	br.com.jequiti.zapsac3
433	Renda Complementar	br.com.kplay.evgrupohinode
434	Renda Complementar	br.com.marykay.myassistant
435	Renda Complementar	br.com.mobicare.cielo
436	Renda Complementar	br.com.userede
437	Renda Complementar	br.gov.caixa.bolsafamilia
438	Renda Complementar	com.monetizzeapp
439	Renda Complementar	com.superrevendedores.app
440	Renda Complementar	net.natura.cn
441	Renda Formal	air.br.com.alelo.mobile.android
442	Renda Formal	app.desenvolvimento.ECXSaldo
443	Renda Formal	br.com.amil.beneficiarios
444	Renda Formal	br.com.bradseg.segurobradescosaude
445	Renda Formal	br.com.mobile.ticket
446	Renda Formal	br.com.portoseguro.saudeodonto
447	Renda Formal	br.com.serasaexperian.consumidor
448	Renda Formal	br.com.softwaremanager.aleloar
449	Renda Formal	br.com.sulamerica.sam.saude
450	Renda Formal	br.gov.caixa.fgts.trabalhador
451	Renda Formal	br.gov.caixa.trabalhador
452	Renda Formal	br.gov.fazenda.receita.irpf
453	Renda Formal	br.gov.fazenda.receita.pessoafisica
454	Renda Formal	com.primety.sodexomobile
455	Renda Formal	com.unimed.guiamdico

Tabela A.1 Aplicativos e suas Categorias (continuação)

	<b>Categoria</b>	<b>Aplicativo</b>
456	Renda Formal	ws.hanzo.Vrrh
457	Smart Things	com.amazon.dee.app
458	Smart Things	com.fossil.wearables.fossil
459	Smart Things	com.google.android.apps.chromecast.app
460	Smart Things	com.google.android.wearable.app
461	Smart Things	com.nest.android
462	Smart Things	com.samsung.android.app.watchmanager
463	Smart Things	com.xiaomi.smarthome
464	Smart Things	wp.wattpad
465	Streaming de música	com.apple.android.music
466	Streaming de música	com.aspiro.tidal
467	Streaming de música	com.claro.claromusica.br
468	Streaming de música	com.google.android.apps.youtube.music
469	Streaming de música	com.musixmatch.android.lyrify
470	Streaming de música	com.rhapsody.napster
471	Streaming de música	com.rhapsody.vivomusica
472	Streaming de música	com.shazam.android
473	Streaming de música	com.soundcloud.android
474	Streaming de música	com.spotify.lite
475	Streaming de música	com.spotify.music
476	Streaming de música	com.studiosol.palcomp3
477	Streaming de música	deezer.android.app
478	Streaming de música	tunein.player
479	Streaming de vídeo	br.com.telecineplay.android
480	Streaming de vídeo	com.amazon.avod.thirdpartyclient
481	Streaming de vídeo	com.dla.android
482	Streaming de vídeo	com.justwatch.justwatch
483	Streaming de vídeo	com.netflix.mediaclient
484	Streaming de vídeo	com.tid.enjoy
485	Streaming de vídeo	com.tubitv
486	Streaming de vídeo	com.tuxera.streambels
487	Streaming de vídeo	org.xbmc.kodi
488	Streaming de vídeo	tv.twitch.android.app
489	Streaming de vídeo	xmw.app.playerxtreme
490	Transporte Público	br.com.embryo.rpc.android
491	Transporte Público	br.com.metrosp.appmetro

Tabela A.1 Aplicativos e suas Categorias (continuação)

	<b>Categoria</b>	<b>Aplicativo</b>
492	Transporte Público	br.gov.sp.cptm.mobile
493	Transporte Público	br.nanoit.viewbus
494	Transporte Público	com.citymapper.app.release
495	Transporte Público	com.marcobiscaro.busfinder.activities
496	Transporte Público	com.riocard.riocardmais
497	Transporte Público	com.thera.metrofacil
498	Transporte Público	com.tranzmate
499	Transporte Público	solucoes.zeno.sinetram
500	Utilitários de Viagem	br.com.livelo.app
501	Utilitários de Viagem	br.com.multipus.participantemultipus
502	Utilitários de Viagem	com.easy.currency.extra.androary
503	Utilitários de Viagem	com.expedia.bookings
504	Utilitários de Viagem	com.google.android.apps.travel.onthego
505	Utilitários de Viagem	com.makemytrip
506	Utilitários de Viagem	com.mapswithme.maps.pro
507	Utilitários de Viagem	com.tripadvisor.tripadvisor
508	Utilitários de Viagem	com.tweakersoft.aroundme
509	Utilitários de Viagem	com.xe.currency
510	Utilitários de Viagem	com.YRH.PackPoint
511	Utilitários de Viagem	net.xelnaga.exchanger
512	Utilitários de Viagem	taxi.android.client
513	Utilitários de Viagem	toilet.samruston.com.toilet
514	Veículo	br.com.carango
515	Veículo	br.com.ctncardoso.ctncar
516	Veículo	br.com.icarros.androidapp
517	Veículo	br.com.mobills.consultaplaca
518	Veículo	br.org.fipe.app
519	Veículo	com.devplank.masterfipe
520	Veículo	com.diariosassociados.novovrum
521	Veículo	com.dsdevelop.infocar
522	Veículo	com.kajda.fuelio
523	Veículo	com.multas.app
524	Veículo	com.nezdroid.cardashdroid
525	Veículo	hands.android.webmotors
526	Voo	br.com.passagensimperdiveis
527	Voo	com.amadeus.cmt.client.android

**Tabela A.1 Aplicativos e suas Categorias (continuação)**

	<b>Categoria</b>	<b>Aplicativo</b>
528	Voo	com.flightaware.android.liveFlightTracker
529	Voo	com.flightradar24free
530	Voo	com.gm.decolar
531	Voo	com.kayak.android
532	Voo	com.seatguru
533	Voo	net.skyscanner.android.main