

SCOrE: um serviço de classificação de usuários móveis com base em seus aplicativos e suas cidades

Leonardo J. A. S. Figueiredo¹, Germano B. dos Santos², Raissa P. P. M. Souza¹
Fabrício A. Silva², Thais R. M. B. Silva², Antonio A. F. Loureiro¹

¹Universidade Federal de Minas Gerais, Belo Horizonte/MG

²Universidade Federal de Viçosa, Florestal/MG

{leonardo.alves, raissa.papini, loureiro}@dcc.ufmg.br

{germano.santos, fabricio.asilva, thais.braga}@ufv.br

Resumo. *Tem-se percebido um aumento significativo na coleta e disponibilidade de dados oriundos de dispositivos móveis. Esses dados, em sua maioria geolocalizados, podem ser usados para se extrair o perfil de comportamento e interesse de usuários. Com isso, empresas podem oferecer serviços melhores e mais personalizados, e a comunidade científica pode avançar no entendimento sobre o comportamento humano. Neste contexto, este trabalho investiga se é possível extrair com qualidade o perfil de usuários móveis considerando apenas os aplicativos instalados, a população e o PIB/capita da cidade do usuário. Foi criado o serviço SCOrE que classifica o usuário em três perspectivas: interesses em locais específicos, visitas a regiões com características funcionais específicas e mobilidade. Com isso, foi possível entender alguns interesses de usuários móveis com base em dados menos invasivos dos mesmos, avançando assim o conhecimento sobre a área.*

Abstract. *There has been a significant increase in the collection and availability of mobile devices data, making possible its use to extract the profile of users' behavior and interests. With this information, companies can offer better and more personalized services, and the scientific community can advance the understanding of human behavior. In this context, this work investigates whether it is possible to extract mobile users' profiles with assertiveness. For this purpose, we considered only the installed applications, the population, and the income/capita of the user's city. So, we created the SCOrE, a service that classifies users from three perspectives: interests in specific locations, visits to regions with particular functional characteristics, and mobility patterns. So, it was possible to understand some mobile users' interests based on less invasive data, advancing then the knowledge about the area.*

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 [Intelligence 2019]. A fim de conhecer melhor os usuários, esses dados podem ser usados para a extração de perfis de comportamento dos mesmos. Esses perfis podem indicar características relevantes dos usuários em diferentes perspectivas, que podem beneficiar tanto as empresas quanto a comunidade científica.

Em relação às empresas, atualmente há uma grande concorrência na prestação de serviços. Com o conhecimento sobre o perfil de seus clientes, será possível tomar melhores decisões, como: recomendações mais assertivas de serviços móveis, ofertas direcionadas, identificação de rotinas para detecção de fraude, entre outros. Já em relação à comunidade científica, os pesquisadores têm investigado características dos usuários móveis para avançar o conhecimento sobre o 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.

Para extrair perfis de comportamento de usuários móveis, diferentes fontes de dados podem ser utilizadas. Isso pode levar a problemas como invasão de privacidade (e.g., registros de ligações ou mensagens) ou necessitar da participação ativa dos usuários (e.g., questionários). Nesse contexto, uma pergunta que surge é se apenas o conhecimento sobre os aplicativos instalados de um usuário, e também algumas características de sua cidade, são fontes de informações relevantes que permitem traçar um perfil consistente do usuário em diferentes perspectivas. Essa pergunta é relevante pois essas informações sobre os aplicativos e a cidade são fáceis de se obter e menos invasivas.

O objetivo deste trabalho é investigar como esses dados (aplicativos instalados e cidade) podem ser utilizados para extrair o perfil de usuários móveis, criando um Serviço de Classificação de Perfil (SCOrE) baseado em três perspectivas: interesse em locais específicos (i.e., Shoppings, Aeroportos, Hospitais e Universidades), interesse em regiões com características funcionais específicas (i.e., Alimentação, Companhia, Educação, Entretenimento, Esportes, Governo, Shopping, Turismo), e características de mobilidade (i.e., *Returners* e *Explorers* [Pappalardo et al. 2015]). Para isso, foram utilizados dados reais de mais de 30.000 usuários enriquecidos com informações de outras fontes. Foram analisadas as associações entre características dos usuários e os seus aplicativos instalados, e criados modelos preditivos capazes de classificar o perfil de um usuário nas diferentes perspectivas. Os resultados mostraram que foi possível alcançar valores satisfatórios de precisão (até 90%) e revocação (até 100%).

O restante deste trabalho está organizado da seguinte forma. Na Seção 2 são discutidos os principais trabalhos encontrados que visam extrair perfil de usuários. Os dados e os métodos utilizados são descritos na Seção 3. O SCOrE e seus resultados são discutidos na Seção 4. Por fim, a Seção 5 apresenta as conclusões e trabalhos futuros.

2. Trabalhos Relacionados

Existem trabalhos na literatura que visam extrair perfis de usuários com base em diferentes dados. [Seneviratne et al. 2015] realizam um estudo utilizando as respostas de um questionário demográfico e a lista de aplicativos de 218 usuários para prever o gênero do usuário. [Malmi and Weber 2016] utilizam dados demográficos e lista de aplicativos para criar um perfil do usuário considerando, além do gênero, atributos como idade e raça.

[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, são feitas as análises dos locais onde há uma concentração de acessos do usuário, sendo feita uma engenharia reversa com o *OpenStreetMap* (utilizam os dados do usuário para enriquecer com o tipo de local em que ele se encontra).

[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). O trabalho de [Silva et al. 2018] considera o uso de 8 aplicativos durante um ano, identificando padrões diferentes de frequência de uso de acordo com períodos do dia e outros contextos. [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. Já [da Silva Ferreira et al. 2021] utilizam as categorias dos aplicativos para classificar o usuário de acordo com os traços de personalidade do *Big Five*.

Por fim, também foram encontrados trabalhos que investigam perfis sociais dos usuários, como [Li et al. 2008] e [Abel et al. 2011]. No primeiro são utilizados *logs* de GPS de 65 voluntários para criar grafos que são usados na identificação do laço social. O segundo também utiliza grafos, mas para correlacionar os comentários de mais de 45.000 usuários em tópicos no *Twitter*, e assim criar um perfil. [Sahoo and Gupta 2021] também utiliza laços sociais, mas para criar um perfil que facilite a detecção de notícias falsas.

O presente trabalho se difere dos demais por utilizar dados de fácil obtenção e menos invasivos. Isso porque, para inferir um novo perfil é necessário apenas a lista de aplicativos e uma única localização referente à cidade do usuário. Além disso, para a criação dos modelos preditivos foram utilizados dados reais de 32.466 usuários, criando perfis analisados em três perspectivas diferentes.

3. Dados e Métodos

3.1. Dados Base

Os dados utilizados neste trabalho para a criação dos modelos para os perfis são as listas de aplicativos instalados e as localizações de acessos dos usuários. Vale destacar que esses dados são utilizados apenas para a construção dos modelos preditivos de perfis. Para o uso do modelo, apenas os aplicativos instalados e a cidade do usuário, que são menos invasivos, são necessários. Além disso, todos os enriquecimentos necessários neste trabalho foram feitos utilizando o *MoreData* [Figueiredo et al. 2021].

Base de Usuários

Foram utilizados, inicialmente, dados reais de 46.439 usuários móveis do sistema *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. Esses dados incluem a lista de aplicativos instalados e diversas localizações geradas pelos acessos do usuário ao aplicativo da empresa. Tal lista é simplesmente um conjunto 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, totalizando 178.877 PoIs distintos considerando todos estes. Foram removidos os usuários sem nenhum PoI, reduzindo o número para 34.428.

Base de Categorias de Aplicativos

Dentre os milhares de aplicativos existentes, foram selecionados os 533 mais frequentes (instalados por pelo menos 3 usuários), para evitar que aplicativos muito raros acabem interferindo nos modelos. Esses aplicativos foram agrupados em 37 categorias específicas, que foram mapeadas nos aplicativos de cada usuário.

Base de Cidades

Visando adicionar informações das cidades dos usuários, suas casas foram mapeadas em suas respectivas cidades a partir das geometrias do IBGE Cidades. Também foram removidos os usuários com casas fora do Brasil, reduzindo a base para 32.466 usuários.

Neste trabalho, foram escolhidos o PIB/capita e o tamanho da população da cidade por serem considerados os mais relevantes para caracterizar o perfil de um usuário. 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 dificilmente há este tipo de estabelecimento nestas condições. Em contrapartida, as outras informações dos municípios, como participante do Amazônia Legal não indicam as condições do local em que a pessoa frequenta.

Cada um desses dois novos atributos foram divididos em três classes. A população foi dividida de acordo com uma adaptação dos valores fornecidos pelo Ipea¹: pequena (menos que 100.000 habitantes) média (entre 100.000 e 500.000 habitantes) e grande (mais que 500.000 habitantes).

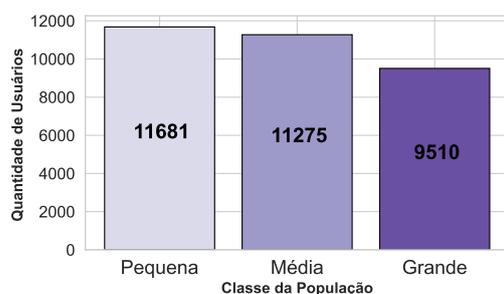


Figura 1. Quantidade de usuários por classe de população.

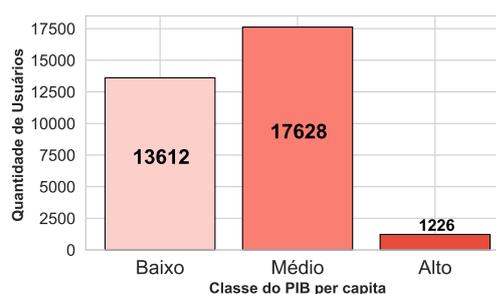


Figura 2. Quantidade de usuários por classe de PIB per capita

Na Figura 1, é possível ver que as classes estão bem distribuídas, mas que o número de usuários de cidades pequenas é o maior. Já as classes para o PIB/capita, Figura 2, foram calculadas considerando os valores dos *quartis* e limite superior para os municípios. Os valores dos quartis Q1, Q2 e Q3 são, respectivamente R\$ 13.315,23, R\$ 21.839,10 e R\$ 34.232,97, com um limite superior *LS* de R\$ 65.609,58. Com isso, as classes do PIB/capita ficaram assim: baixo (PIB/capita < R\$ 21.839,10), médio (PIB/capita entre R\$ 21.839,10 e R\$ 65.609,58) e alto (PIB/capita > R\$ 65.609,58).

3.2. Métodos

Foram criados modelos preditivos para inferir o perfil de um usuário em cada uma das três perspectivas (Interesses em Locais, Área Funcional e Mobilidade), utilizando como entrada apenas os aplicativos instalados e as categorias da população e do PIB/capita da cidade do usuário. Para cada classe do perfil, foram criados modelos binários diferentes. 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. O mesmo foi feito para as categorias de população e renda.

¹https://web.archive.org/web/20090819081149/http://www.ipea.gov.br/003/00301009.jsp?ttCD_CHAVE=5499

A fim de encontrar os melhores resultados, foram testados 4 diferentes algoritmos de classificação: *Logistic Regression*, *Random Forest*, *Naive Bayes* e *K-Nearest Neighbors*. Foi utilizado o PCA (Análise de Componentes Principais) para que os componentes principais sejam usados como atributos. Quando necessário, para corrigir o desbalanceamento das classes de entrada, foi utilizada a técnica de *Undersampling*, que remove dados da classe com maior número de entradas para que seja balanceada com a de menor número. O processo foi repetido 35 vezes para cada uma das combinações de entrada, já que o balanceamento gera amostras diferentes em cada rodada. A técnica de *Grid Search* foi utilizada para se encontrar as melhores configurações para os algoritmos.

Foram utilizadas as métricas *precisão*, *revocação* e *f-score* para avaliar a qualidade dos modelos preditivos. Além disso, para uma análise preliminar sobre a associação entre os aplicativos dos usuários e seus respectivos interesses, foi utilizada a métrica de *lift*. Essa métrica indica o ganho (ou perda, caso o valor for negativo) que se pode ter ao comparar amostras dos usuários que pertencem e não pertencem à classe de interesse, e que possuem aplicativos de alguma categoria.

4. SCOrE

Com base nas discussões anteriores, foi criado o serviço SCOrE, recebendo como entrada a lista de aplicativos instalados e uma única localização. Sua primeira etapa é enriquecer a entrada com as categorias dos aplicativos, e a localização com os dados de população e PIB/capita. Além disso, foram criados modelos para cada uma das três perspectivas e estes foram utilizados para calcular as probabilidades do usuário pertencer a cada uma das classes dos três perfis. A seguir são apresentados cada uma das perspectivas separadamente e os resultados obtidos para o SCOrE.

4.1. Perfil de Interesses em Locais

Este perfil visa compreender o comportamento dos usuários (31.457) 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, foi verificado se cada PoI de um usuário está localizado em algum tipo específico de local conforme os dados do *OpenStreetMap (OSM)*².

Foram coletados quatro diferentes tipos de locais do OSM: *Aeroporto*, *Hospital*, *Shopping* e *Universidade*. A escolha desses tipos foi feita pois se tratam de locais com maior área geográfica, facilitando o mapeamento de uma localização de um usuário a um desses locais. Foram encontrados 7.540 locais pertencentes ao território brasileiro, distribuídos em: 627 aeroportos, 3.191 hospitais, 1.639 Shoppings e 2.083 universidades.

Dentre os resultados obtidos dos *Lifts*, um destaque vai para usuários que possuem um ponto de interesse em Aeroportos e os valores positivos para aplicativos das categorias *Cia Aérea Internacional* (257,31%), *Voo* (178,31%) e *Cia Aérea* (146,68%). Também foi possível observar que a taxa de usuários que possuem algum ponto de interesse em hospitais e têm aplicativos instalados sobre *Medicina* é 64,36% maior em comparação com os usuários que não possuem interesse nesse tipo de local. O *Lift* de aplicativos da categoria

²<http://openstreetmap.org/>

de *Investimento* é positivo (68,90%) e o de aplicativos de *Renda Complementar* é negativo (-59,27%), para usuários que possuem algum PoI em *Shopping*. Por fim, 14,48% dos usuários que possuem PoI em universidades possuem aplicativos sobre *Educação*, com *Lift* de 111,64%. Existe ainda um *Lift* alto para aplicativos de *Transporte Público* (89,38%) e *Eventos* (132,05%), explicado pelo perfil de usuários de universidades.

A Figura 3 apresenta os resultados dos modelos preditivos para as métricas precisão, revocação e F-Score para as 35 execuções³ para o *Random Forest*, que foi o algoritmo com melhores resultados. É possível prever mais precisamente se um usuário tem interesse em *Aeroporto* pois possui mais categorias de aplicativos diretamente relacionadas. 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.

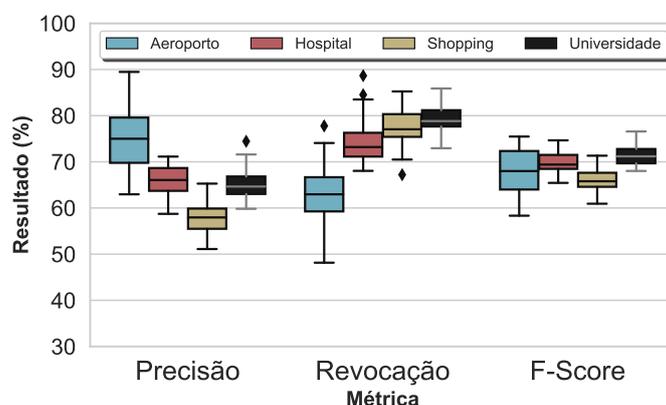


Figura 3. Resultados de Precisão, Revocação e F-Score para os 4 tipos de locais.

4.2. Perfil de Área Funcional

Outro perfil estudado é em relação às áreas funcionais visitadas pelos usuários. Uma Área Funcional pode ser entendida 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]. Com isso, é possível analisar os tipos de áreas que ele frequenta e estas são 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. Neste trabalho, foi utilizada a técnica de fatoração de matrizes para calcular o *score* a partir da porcentagem dos tipos de áreas visitadas, conforme descrito a seguir.

Para construir as matrizes foram utilizadas 4.966.909 localizações brutas dos 22.747 usuários que possuíam localização próxima de algum estabelecimento dos tipos analisados. Utilizando o índice H3⁴, foram formadas regiões com aproximadamente 500 metros de raio, equivalente a 0,73 km² (resolução igual a 8, segundo a configuração do

³foram feitas várias execuções devido ao efeito aleatório do *undersampling*

⁴<https://eng.uber.com/h3>

H3⁵). Foram extraídas 106.051 regiões com pelo menos 5 visitas cada uma, para evitar o uso de áreas pouco relevantes. Os *scores* foram utilizados para separar os usuários em interessados ou não em cada funcionalidade, através do limite superior de cada uma das classes de funcionalidade. Logo, é assumido que os usuários com *score* maior que o limite superior possuem interesse, e os demais não. Dessa forma, somente usuários com *score* alto são considerados como interessados, diminuindo o risco de falsos-positivos.

A análise dos *Lifts* mostrou que existem algumas associações interessantes, o que indica que os aplicativos dos usuários podem estar relacionados com seus interesses em regiões com determinadas características funcionais. Por exemplo, usuários com alto *score* em *Alimentação* não possuem aplicativos das categorias *Delivery* (-3,77%) e *Fast Food* (1,03%) com frequência. Além disso, também foi observado que usuários com *score* alto para *Companhia* possuem valor alto de *Lift* para aplicativos *Corporativos* (128,02%).

A Figura 4 apresenta os resultados dos modelos preditivos para o algoritmo *Logistic Regression*, que foi o melhor para esse cenário. É possível observar que os resultados de precisão foram similares para todas as classes de funcionalidades, com somente a classe *Esportes* com resultados inferiores a 50% para esta métrica. Já para revocação, os resultados foram significativamente distintos, com algumas classes alcançando 100% nesta métrica. Este é um resultado relevante para campanhas de *marketing*, por exemplo, pois indica que todos os usuários que possuem algum interesse na funcionalidade foram previstos corretamente. Dessa maneira, ao invés de sugerir um produto para um 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, com destaque para *Educação* e *Esportes*. Estes resultados mostram que a maioria dos usuários com interesse nessas funcionalidades foram classificados corretamente pelos modelos. Dessa forma, há uma alta assertividade de usuários que possuem alguma relação com determinada funcionalidade. Considerando esses 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, propondo assim soluções mais personalizadas para os usuários.

4.3. 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*, quando regularmente visitam poucos locais diferentes, ou *explorers*, caso contrário. Dessa maneira, foram utilizadas a lista de aplicativos e a cidade do usuário como entrada para o estudo de [Pappalardo et al. 2015], classificando-as como *returners* ou *explorers*.

Conforme o trabalho original [Pappalardo et al. 2015], foram testados valores de $k \in \{2, 4, 8\}$. 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*. Neste trabalho, foi selecionado o valor de $k = 4$. Isso porque, considerar $k = 2$ limita a somente *Casa* e *Trabalho*, para a maioria dos usuários. Este comportamento

⁵<https://github.com/uber/h3/blob/master/docs/core-library/restable.md>

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.

Assim, o *Lift* da categoria *Transporte Público* é positivo (24,37%) para usuários classificados como *returners*. 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. Por outro lado, pessoas que necessitam de transporte público, não têm oportunidades de ficarem visitando muitos locais frequentemente. Já para a classe *explorer*, o *Lift* para *Motorista* é positivo (64,07%), o que pode estar relacionado com a profissão de motorista de aplicativo, que visitam vários lugares diferentes.

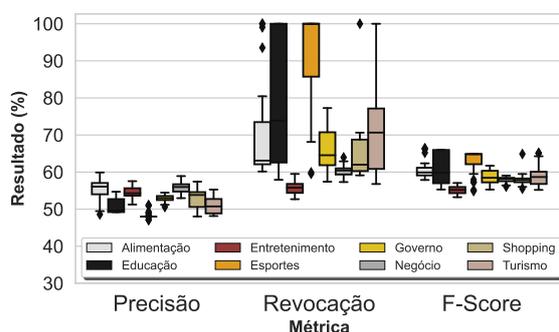


Figura 4. Resultado para Perfil de Área Funcional.

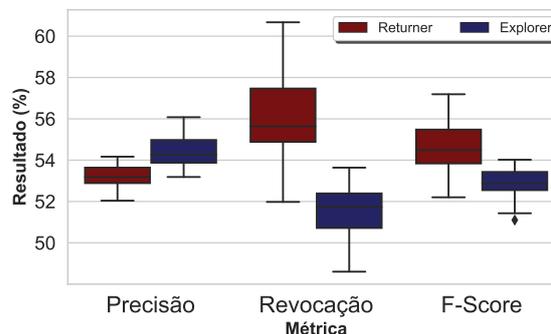


Figura 5. Resultado para Perfil de Mobilidade ($k = 4$).

Com as análises, é possível ver que há relação entre os aplicativos instalados e a classe de mobilidade. A Figura 5 apresenta os resultados dos modelos para o *Naive Bayes*, que foi o melhor algoritmo para esse cenário. 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 obteve resultados de precisão e revocação entre 50% e 60%. Isso pode estar relacionado aos diferentes perfis de pessoas que pertencem ao conjunto de *Returners* ou *Explorers*. Há pessoas de diferentes idades, preferências e classes sociais que podem pertencer a essas categorias. Ainda assim, este perfil pode ser útil para selecionar usuários em comparação com amostras aleatórias. Dessa maneira, é possível melhorar a escolha do público-alvo em campanhas de *marketing*, por exemplo.

4.4. Resultados

Como visto nas seções anteriores, foi possível criar modelos de aprendizado de máquina para extrair perfis de usuários móveis em diferentes perspectivas. Para avaliar a qualidade do serviço, foram selecionados os 780 usuários da interseção entre as três perspectivas, ou seja, aqueles que pertenciam a pelo menos uma das classes de cada perfil. Repetimos o processo de geração da amostra 35 vezes para cada classe dos perfis, e a Figura 6 mostra os valores de *lifts* obtidos. Todas as classes obtiveram *Lifts* positivos nas médias das execuções, sendo a *Companhia*, *Turismo* e *Aeroporto* as que apresentaram melhores resultados. Os valores baixos, porém positivos, das classes *Returner* e *Explorer* refletem aos resultados encontrados no Perfil de Mobilidade.

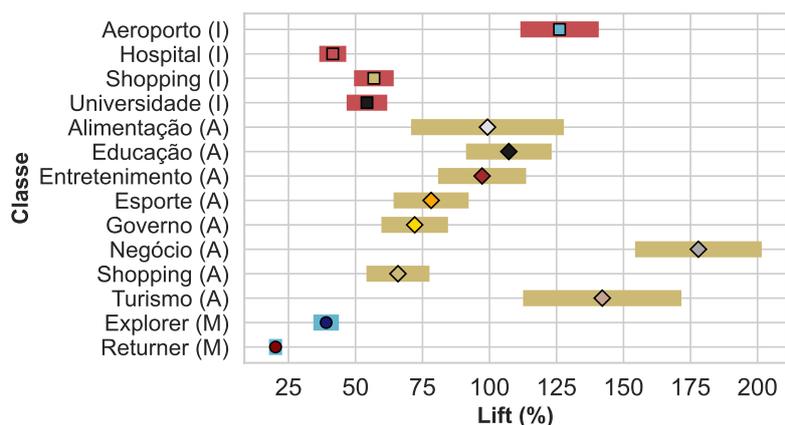


Figura 6. Lifts com intervalo de 95% de confiança

Em geral, esses resultados revelam que ao utilizar o serviço de extração de perfis proposto, pode-se obter uma lista de usuários classificados em tais classes em que se tenha interesse melhor do que com amostras aleatórias. Assim, empresas que não possuem dados geoespaciais detalhados, nem outras informações, ainda podem inferir perfis de mobilidade e interesse de seus usuários e utilizar esse conhecimento para oferecer-lhes serviços melhores e mais personalizados.

5. Conclusões e Trabalhos Futuros

Este trabalho apresentou o SCOrE, um serviço para que o perfil de um usuário seja extraído, recebendo como entrada apenas os aplicativos instalados e a cidade do mesmo. Nele, foi investigado o impacto de três diferentes perspectivas no perfil do usuário (interesses em locais específicos, visitas a regiões com características funcionais específicas e de mobilidade) com base apenas nos aplicativos instalados, e em informações de população e PIB/capita da cidade do usuário. Os resultados mostraram que existem associações relevantes e condizentes com as características do usuário. Além disso, os modelos alcançaram, no geral, resultados significativos em termos de precisão e revocação.

Os conhecimentos adquiridos com este trabalho são relevantes para a comunidade científica e para o mercado. Foi mostrado, com base em dados reais de dezenas de milhares de usuários, que é possível inferir o perfil de usuários com base em poucas informações, menos invasivas dos mesmos. Este é o primeiro trabalho que investiga, sob três diferentes perspectivas, a relevância dos aplicativos e das cidades para se extrair conhecimentos sobre os usuários. Como trabalhos futuros, pretende-se investigar outras perspectivas, como de laços sociais e propensão de adesão a serviços móveis. Além disso, aprimorar os perfis existentes, e analisar outros tipos de locais (e.g. estádios e parques).

Referências

- 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.
- 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.
- da Silva Ferreira, C., dos Santos Figueiredo, L. J. A., Silva, F. A., and Silva, T. R. B. (2021). Inferindo traços de personalidade com base em aplicativos móveis. In *Anais do XIII Simpósio Brasileiro de Computação Ubíqua e Pervasiva*, pages 162–171. 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.
- Figueiredo, L. J., dos Santos, G. B., Souza, R. P., Silva, F. A., and Silva, T. R. B. (2021). Moredata: A geospatial data enrichment framework. In *Proceedings of the 29th International Conference on Advances in Geographic Information Systems*, pages 419–422.
- Intelligence, G. (2019). The mobile economy 2019. online.
- 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.
- Malmi, E. and Weber, I. (2016). You are what apps you use: Demographic prediction based on user’s apps. *arXiv preprint arXiv:1603.00059*.
- 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.
- Sahoo, S. R. and Gupta, B. B. (2021). Multiple features based approach for automatic fake news detection on social networks using deep learning. *Applied Soft Computing*, 100:106983.
- 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, F. A., Domingues, A. C. S. A., and Silva, T. R. M. B. (2018). Discovering mobile application usage patterns from a large-scale dataset. *ACM Trans. Knowl. Discov. Data*, 12(5).
- Varlamis, I., Sardianos, 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.
- 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.