



Simpósio de Integração Acadêmica

Inteligência Artificial: A Nova Fronteira da Ciência Brasileira
SIA UFV Virtual 2020



Uma Solução para Análise de Grandes Volumes de Dados Georreferenciados

UFV-Campus Florestal

Isabella Menezes Ramos, Fabrício A. Silva e Wandella Maia de Oliveira

dados georreferenciados, perfil demográfico, análise de regressão

Introdução

A disseminação de dispositivos móveis tem aumentado em uma proporção significativa entre a população brasileira. Isso acontece porque os smartphones servem de fonte de dados sobre os usuários e o ambiente onde estão inseridos. Sendo assim, as empresas têm buscado extrair conhecimento útil desses dados com o intuito de conhecer seus clientes e oferecer serviços personalizados.

Com isso, é esperado que os usuários de serviços móveis se sintam mais satisfeitos com a qualidade dos mesmos. Este trabalho pode ajudar às empresas a conhecerem melhor seus usuários e oferecerem melhores serviços. Ao obter informações sobre o usuário tem-se dados sobre sua preferência de dispositivos e com isso, obtendo a localização de sua residência, pode-se desvendar características específicas daquela região.

Objetivos

O objetivo deste trabalho é explorar dados de dispositivo móvel e setor censitário de milhares de usuários no Brasil para verificar se seu local de residência está correlacionado com as características de seus smartphones.

Material e Métodos

Para a execução do projeto, utilizamos a linguagem Python, IDE Jupyter Notebook e bibliotecas como Sklearn, Matplotlib e Seaborn para a fazer as análises do banco de dados, composto por mais de 60.000 usuários de várias partes do Brasil.

Resultados e Discussão

Foram feitas caracterizações, a fim de conhecer melhor os dados com o qual trabalhamos, em termos de população, região rural, região urbana, Amazônia-legal, região semiárida e atividades com maior valor adicionado bruto, com isso conseguimos observar características diferentes entre o perfil de usuários Android e Apple, mas não há grande variação de preço entre os dispositivos no geral.

Após a caracterização, foram feitas análises com modelos de regressão, com o intuito de encontrar correlações entre os dados, e de machine learning para tentar prever o preço dos dispositivos móveis. Verificamos que não existe uma correlação clara no modelo de regressão e não foi possível treinar um modelo eficiente para realizar a classificação de preços.

Conclusões

Ao se tentar várias combinações de variáveis explicativas, verificou-se que não existe uma correlação clara e, conseqüentemente, não foi possível estabelecer com precisão um modelo de regressão para prever o preço médio dos smartphones com base em características da região de residência. Apesar disso, o presente trabalho é importante pois revela conhecimentos sobre o perfil de usuários que pode ajudar as empresas a oferecerem melhores serviços.

Bibliografia

- Neves, E. C. (2019). Modelos de predição – ensemble learning. medium. Acesso em: 5 de Agosto de 2020.
- Rodrigues, V. (2018). Entenda o que é AUC e ROC nos modelos de machine learning. medium. Acesso em: 5 de Agosto de 2020

Apoio Financeiro



Agradecimentos

