



# Simpósio de Integração Acadêmica

## “Ciências Básicas para o Desenvolvimento Sustentável”

SIA UFV 2023



## Classificação Semântica de Pontos de Interesse

Pedro Augusto Maia Silva ([pedro.maia@ufv.br](mailto:pedro.maia@ufv.br)), Fabrício A. Silva ([fabricaoasilva@ufv.br](mailto:fabricaoasilva@ufv.br))  
Laboratório de Inteligência em Sistemas Pervasivos e Distribuídos (NESPED-Lab)  
IEF - Universidade Federal de Viçosa - Campus Florestal

**Palavras chave - Dados geoespaciais, Pontos de Interesse, Algoritmo de classificação**

**Área temática: Ciência da computação**  
**Grande área: Ciências Exatas e Tecnológicas**  
**Categoria: Pesquisa**

### Introdução

O planejamento das cidades de forma metódica atualmente tende a utilizar de novas tecnologias e ferramentas de análise semântica de pontos de interesse para esclarecer a divisão da estrutura espacial e fortalecer o planejamento racional das áreas funcionais urbanas. Atualmente, com o uso de redes sociais baseadas em localização, é possível extrair os check-ins dos usuários. Uma forma comum de se obter dados é por meio de posts compartilhados no Twitter. O algoritmo de K vizinhos mais próximos (K-NN) foi utilizado para a análise dos dados, sendo enriquecido com os padrões de mobilidade.

### Objetivos

O objetivo geral do trabalho desenvolvido ao longo dos últimos 12 meses foi um estudo e análise de possibilidades de otimização do algoritmo de classificação semântica de pontos de interesse apresentado utilizando o K-FN[1] e integrá-lo na biblioteca SENDAS. O algoritmo usa padrões de mobilidade e o K-NN modificado para identificar o tipo de local de um check-in não rotulado. Isso inclui também a capacitação do bolsista nas áreas de ciência dos dados, incluindo análise estatística, mineração de dados e aprendizagem de máquina, além de ferramentas úteis para a coleta, processamento e análise de grandes quantidades de dados.

### Material e Método

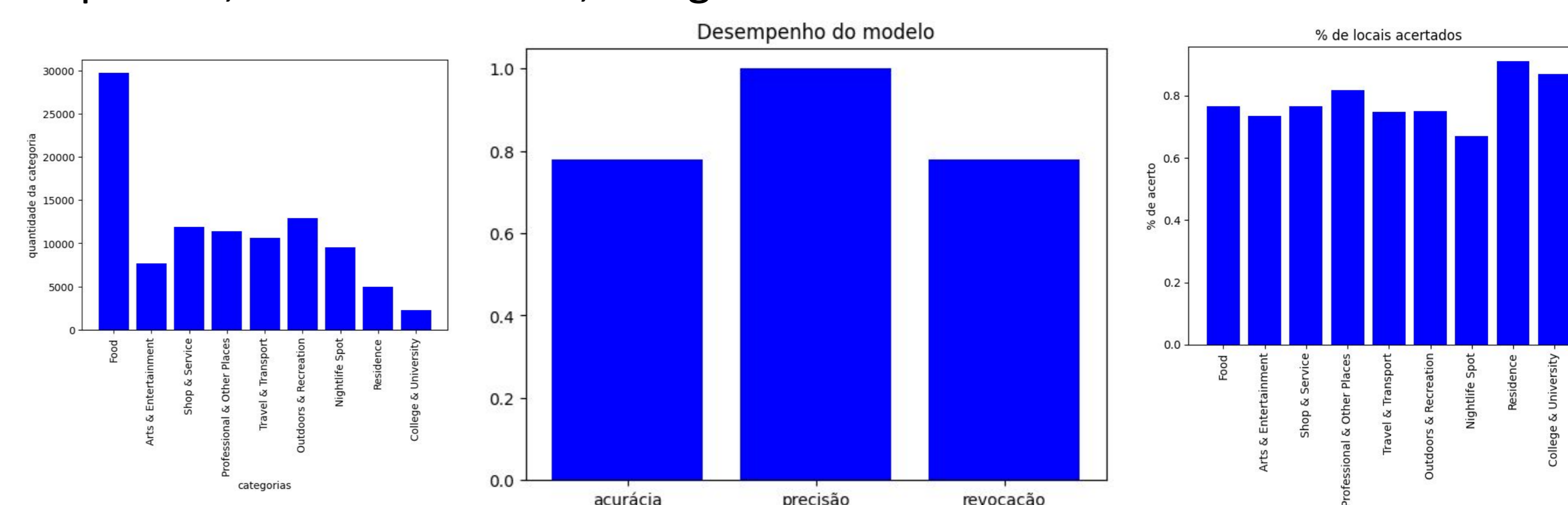
A tecnologia Spark foi utilizada, visto que na linguagem podem ser aplicados pipelines para etapas complexas. As bibliotecas utilizadas foram: SQL (pertencente ao pacote apache do Spark) e geom (pertencente ao pacote jts do locationtech). Foram usados dados de mobilidade urbana e os dados oriundos das LBSN's, as quais forneciam os check-ins dos usuários com as informações necessárias para a análise dos rótulos. Dessa forma, foi possível alterar e melhorar a classificação no algoritmo K-FN.

### Apoio financeiro

Bolsa de iniciação científica PIBIC/CNPq.

### Resultados e Discussão

Partindo do K-FN, construímos a solução realizando as modificações para se alinhar com a biblioteca do SENDAS, aproveitando a estrutura do TrajDataFrame contido na biblioteca, o modelo teve um bom desempenho na identificação dos locais visitados pelos usuários, especialmente os de maior frequência, como residência, colégio e universidade.



### Conclusões

A solução implementada teve como objetivo trabalhar com os dados de maneira paralela e distribuída. Após feita a implementação existem possíveis melhorias para o modelo, como incluir a análise do horário dos check-ins. Apesar disso, as aplicações práticas da proposta em análise de mobilidade e recomendação de locais se mostraram interessantes e precisas.

### Bibliografia

- [1] Helen C. M. Senefonte, Thiago H. Silva, Ricardo Luders, Myriam R. B. S. Delgado. "Classifying Venue Categories of Unlabeled Check-ins Using Mobility Patterns" p.29-31, 2019.
- [2] Gao, Song, Krzysztof Janowicz, and Helen Couclelis. "Extracting functional urban regions from points of interest and human activities in location-based social networks." Transações em GIS 21.3 (2017): 446-467.

### Agradecimentos

É importante destacar a participação do meu orientador Fabrício por todo o apoio e pela oportunidade. Vale destacar também a participação do Guilherme Sérgio com o auxílio e ambientação do projeto. E também ao CNPq pelo apoio financeiro.

