**Área temática:** Ciências Exatas e da Terra

**Desenvolvimento de um Serviço para Processamento de Dados de Mobilidade baseado em Mecanismos de Mineração de Dados**

Mikéias Gabriel Maia Azevedo, Daniel Faustino Lacerda de Souza, Angélica Félix de Castro

Com os avanços tecnológicos das últimas décadas, dispositivos cada vez menores alcançaram altos níveis de processamento, contando com cada vez mais funcionalidades. Uma das funções mais comuns é a de georreferenciamento, isto é, a capacidade de obter a localização do dispositivo em tempo real no globo. Com isso, por meio de redes passivas de captura, é possível ter acesso à esses dados, anonimizá-los e, ainda assim, extrair informações relevantes através do seu rastro no espaço tempo. A partir dessas informações surgem diversas aplicações relativas às variáveis espaço/tempo do dispositivo (anonimizado), propiciadas pela capacidade de aplicação de modelos de predição e agregação. Dito isto, no contexto dessa pesquisa foi concebida uma arquitetura de software, com seu conjunto de mecanismos e processos, objetivando viabilizar a construção de uma plataforma de serviços de análise de dados baseados na coleta passiva de informações sobre a mobilidade de dispositivos eletrônicos rastreáveis. Além disso, provas de conceitos foram desenvolvidas e testadas a fim de validar a solução idealizada.

A metodologia adotada foi dividida em três frentes realizadas paralelamente, consistindo na pesquisa e levantamento bibliográfica dos assuntos adjacentes relativos à, tanto as áreas de técnicas de mineração de banco de dados, quanto no contexto de anonimização de dados e redes de sensores em cidades inteligentes. Como uma segunda frente, foi feito o estudo das técnicas e tecnologias relevantes para o desenvolvimento da arquitetura pretendida. Por fim, como terceira frente, realizou-se o desenvolvimento da solução.

O resultado final alcançado foi uma arquitetura baseada em camadas que realiza o processamento de dados utilizando técnicas de clusterização em nuvem aplicando o framework Apache Spark. As camadas criadas, denominadas de (a) ACM (Mecanismo de Classificação e Agregação) e (b) MPM (Mecanismo de Predição e Mineração), são, respectivamente, responsáveis pelo (a) processo de agrupamento sistemático com base na morfologia dos dados (agrupando os dados de localização e retornando-os em um formato definido de pontos, áreas ou linhas); e (b) pela aplicação de modelos estatísticos com vistas a classificação de dados e inferência de informação. Como prova de conceito, no contexto do ACM foram desenvolvidos alguns filtros de classificação morfológica baseadas em operações de interseção, união e segmentação de áreas quadrangulares e circulares. No contexto do MPM, algoritmos para geração de mapas de calor, classificação por k-means e determinação de intervalos de confiança para classificação de padrões de mobilidade, foram incorporados. É importante considerar que a arquitetura desenvolvida foi pensada de forma a ser flexível, facilitando a incorporação de novos métodos para ambas as camadas de processamento de dados.

**Palavras-chave:** *smartcities*, geoprocessamento, processamento de dados, big data, mineração de dados.   
**Agência financiadora:** UFERSA.