

# Relatório Técnico: Agrupamento e Reconhecimento de Atividades Humanas Utilizando o Algoritmo K-means

Nome do Residente: Leandro de Oliveira e Sumaia Suzart Argôlo Nunesmaia

Data de Entrega: 27/11/2024

RESTIC36-CEPEDI

## Resumo

Este projeto explora o uso do algoritmo K-means para o reconhecimento de atividades humanas com base em dados obtidos de sensores em smartphones. As etapas incluem análise exploratória, normalização dos dados, redução de dimensionalidade com PCA e a aplicação do K-means para identificar clusters representando diferentes atividades. Os resultados mostram que a abordagem adotada oferece boa separação e coesão dos grupos, embora haja limitações em relação à interpretação direta dos clusters formados.

## Introdução

Com o crescimento das tecnologias de dispositivos móveis, o monitoramento de atividades humanas usando sensores se tornou uma área de grande interesse. Este trabalho utiliza dados de sensores para agrupar atividades humanas com base em medições como aceleração e giroscópio. O algoritmo K-means foi escolhido devido à sua simplicidade, escalabilidade e eficácia em tarefas de agrupamento. A escolha do número de clusters e a avaliação do modelo foram feitas com base em métricas como inércia e silhouette score.

## Metodologia

A abordagem utilizada no projeto iniciou com uma análise exploratória dos dados, permitindo identificar padrões e relações entre as variáveis obtidas dos sensores de smartphones. As variáveis foram normalizadas usando a técnica de padronização (StandardScaler) para garantir que todas contribuíssem de forma equilibrada ao modelo. Em seguida, aplicou-se a técnica de redução de dimensionalidade com PCA (Análise de Componentes Principais), preservando 90% da variância explicada, o que reduziu a dimensionalidade do dataset sem perda significativa de informação. O algoritmo K-means foi implementado utilizando a inicialização **k-means++**, que assegura uma escolha inicial eficaz dos centróides, otimizando o tempo de convergência. Para determinar o número ideal de clusters (KKK), foram aplicados o método do cotovelo (elbow method), analisando a inércia, e o silhouette score, que avaliou a coesão interna e a separação dos clusters. A qualidade do agrupamento foi avaliada por meio de métricas, e os resultados foram visualizados em 2D e 3D após a aplicação do PCA.

## Resultados

Os resultados mostraram que o número ideal de clusters foi  $K=5$ , com base nos valores do silhouette score e no ponto de inflexão do gráfico do método do cotovelo. As visualizações em 2D e 3D confirmaram que os clusters estavam bem separados, indicando uma boa coesão e representatividade dos grupos. As métricas de avaliação reforçaram a eficácia do modelo, com um silhouette score indicando uma separação clara entre os clusters. A análise das variáveis mais significativas apontou que medições específicas dos sensores contribuíram de forma predominante para a formação dos clusters, sugerindo que o modelo conseguiu identificar padrões distintos de atividades humanas.

## Discussão

Os resultados obtidos mostraram que variáveis como o número de seguidores e postagens têm um impacto significativo na taxa de engajamento. No entanto, limitações foram identificadas, incluindo a possibilidade de que outras variáveis não consideradas possam influenciar os resultados. A escolha de variáveis independentes e o modelo de regressão impactaram diretamente o desempenho do modelo, corroborando a importância da seleção de recursos (Hastie et al., 2009).

## Conclusão e Trabalhos Futuros

Este trabalho demonstrou a aplicabilidade e eficácia do algoritmo K-means para agrupar atividades humanas a partir de dados de sensores de smartphones. A normalização e a redução de dimensionalidade desempenharam papéis fundamentais na melhoria do desempenho do algoritmo, enquanto o uso de métricas e visualizações permitiu avaliar e interpretar os resultados com clareza. Apesar das limitações relacionadas à ausência de labels para validação direta, os clusters formados indicaram padrões consistentes e significativos nos dados.

## Trabalhos Futuros

Para trabalhos futuros, recomenda-se a inclusão de variáveis contextuais, como horário das atividades e informações adicionais sobre os participantes, para enriquecer os dados e melhorar a precisão dos agrupamentos. Além disso, a aplicação de técnicas híbridas que combinem clustering com métodos supervisionados pode ajudar a validar os clusters formados e explorar relações mais complexas. Outro avanço seria incorporar informações temporais, utilizando séries temporais para capturar dependências dinâmicas entre as variáveis ao longo do tempo.

## Referências

1. Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The Elements of Statistical Learning: Data Mining, Inference, and Prediction*. Springer.
- James, G., Witten, D., Hastie, T., & Tibshirani, R. (2013). *An Introduction to Statistical Learning: with Applications in R*. Springer.
- Tukey, J. W. (1977). *Exploratory Data Analysis*. Addison-Wesley.

- Choudhury, M. D., & Sundararajan, A. (2016). *The Influence of Social Media on Engagement. International Journal of Research in Marketing.*
- Dados: *UCI HAR Dataset - Human Activity Recognition Using Smartphones.*