SWE557: Artificial Intelligence and Machine Learning Programming

Project Topic: 4G LTE Coverage Optimization: Identification of Heterogeneous Cellular Coverage Regions Using K-means Clustering on KPI Data.

Presenter: *Luis Serge Olivier Geralda Tutab*

ABSTRACT

In the modern landscape of telecommunication, optimizing the coverage and performance of 4G LTE networks is crucial for maintaining high-quality service and user satisfaction. This project, titled "4G LTE Coverage Optimization: Identification of Heterogeneous Cellular Coverage Regions Using K-means Clustering on KPI Data", explores the application of K-means clustering, an unsupervised machine learning algorithm, to classify LTE Key Performance Indicator (KPI) data points. The primary objective is to identify heterogeneous cellular coverage regions within a network site and optimize coverage by pinpointing potential coverage holes regions.

The project begins with data acquisition, where extensive LTE KPI data is collected from network monitoring tools and databases from Kaggle. Key metrics such as Reference Signal Received Power (RSRP), Signal-to-Interference-plus-Noise Ratio (SINR), and throughput are extracted, forming the basis of the analysis. The next step involves data cleaning and preprocessing, which includes handling missing values, normalizing the data, and ensuring consistency across the dataset.

Following data preparation, the K-means clustering algorithm is implemented. The algorithm segments the KPI data into distinct clusters, each representing different performance levels within the network. To determine the optimal number of clusters, the elbow method is employed, which helps in identifying the point where adding more clusters no longer significantly reduces the within-cluster sum of squares.

The dataset is then divided into training and testing subsets to evaluate the performance of the clustering model. The training set is used to fit the K-means model, while the testing set helps in validating the clusters. Post clustering, error evaluation is conducted to assess the accuracy and reliability of the model. Confusion matrix and Accuracy scores are calculated to measure the quality of the clustering results, ensuring that the clusters are well-defined and distinct.

The final stage of the project involves analyzing the clusters to identify heterogeneous coverage regions, specifically highlighting potential coverage holes. By mapping the clusters back to the geographic locations of the LTE network, areas with degraded performance are identified. This analysis provides actionable insights for network engineers, enabling them to pinpoint and address coverage deficiencies effectively.

The results should demonstrate that K-means clustering can provide valuable insights into 4G LTE network performance, contributing to more robust and reliable coverage. This approach not only aids in immediate troubleshooting but also supports long-term network optimization efforts, enhancing overall service quality and user experience.