Exploring Big Data Analytics in Stroke Prediction and Comparative Study of Learning Algorithms

Presented by: Kritika and Rojina

Problem Statement

- Stroke leads to severe disabilities and death worldwide.
- Around 16 million strokes occur yearly; 5 million fatalities (WHO).
- Early recognition and prediction are crucial to minimize the devastating effects.
- Traditional diagnosis methods, like MRIs and CT scans, may not always provide timely results.

Research Question

- How can big data analytics improve the accuracy and efficiency of stroke prediction models?

Base Paper

- Title: "Exploring the Efficiency of Convolutional Neural Networks, Recurrent Neural Networks, and Deep Learning in Stroke Prediction Using Medical Image Data"
- Authors: Chidozie Shamrock Nwosu, S. D.
- Published in: PubMed Central (2019)
- Dataset Used: Medical images, likely including CT and MRIs for stroke prediction.

Dataset

Source: https://www.kaggle.com/datasets/fedesoriano/stroke-prediction-dataset

- Rows: 5110
- Columns: 12

Aim

To explore the potential of big data and machine learning in developing an effective stroke prediction system.

Objectives

1. Analyze large datasets of medical records to identify high-risk individuals.
2. Implement machine learning models for stroke prediction.
3. Compare the performance of models such as Logistic Regression, SVM, Decision Trees, and Random Forest.

Methodology

1. Data Collection
2. Data Cleaning
3. Feature Selection
4. Model Training

Machine Learning Models

1. Logistic Regression
2. Support Vector Machine (SVM)
3. Decision Tree
4. Random Forest

CONCLUSION

Our study demonstrates that using big data analytics and machine learning for stroke prediction can significantly improve early detection and prevention, ultimately saving lives.

We compared four machine learning models – Logistic Regression, SVM, Decision Tree, and Random Forest.

By analyzing large medical datasets, we can identify high-risk individuals, providing an opportunity for timely medical intervention.

Thank You