NLP PROJECT BASED ON SAUDI NATIONAL OPEN DATA

Business Name Classification using Arabic BERT

Prepared by : Sulaiman Alatwah

Why This Project Matters

Understanding the importance of categorizing business names effectively

Improved Searchability

Classifying names helps consumers find relevant businesses easily.

Enhanced Marketing Strategies

Categories allow businesses to tailor promotions to target audiences.

Data Analytics Insights

Classification provides valuable data for market analysis and trends.

Data Preprocessing Overview

Effective preparation is crucial for model accuracy and performance.
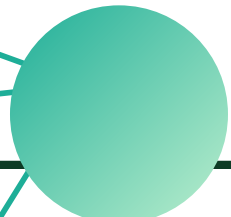
Manual Labeling

Ensures data quality by accurately categorizing names for training.

Text Cleaning

Removes noise, improving the model's ability to learn patterns.

Consistent Formatting

Standardizes input, making data processing streamlined and efficient.

Overview of Classification Techniques

Exploring various algorithms for business name classification

This section discusses the **key algorithms** utilized in classifying business names, including decision trees, support vector machines, and neural networks. Each technique offers unique advantages, enabling effective categorization and better understanding of business entities in the dataset.

Model Accuracy Comparison Over Time

High accuracy reflects model's performance and reliability

The line chart illustrates the **impressive growth** in model accuracy, peaking at 99.80%. This indicates a strong ability to classify business names effectively, showcasing the model's reliability and effectiveness in real-world applications.

Technologies and Tools Used in Project

01

Python for data analysis and model development, ensuring efficient processing.

02

Pandas library for data manipulation, making handling datasets easier and more effective.

03

Scikit-learn for implementing machine learning algorithms, providing robust functionalities for classification tasks.

04

Matplotlib for visualizing results, allowing clear representation of model accuracy and insights.