Project Description:

This project aims to develop an intelligent system that analyzes and classifies Arabic tweets by topic using deep learning (LSTM) techniques.

The system was built through a series of steps, starting with text processing. Tweets are cleaned of symbols, links, and duplication, then divided into words and a stemming technique is applied to reduce the words to their basic form.

The data is then prepared using class balancing to overcome the topic imbalance problem. The texts are then converted into numerical representations that the neural network can handle, encoding the topics into numbers.

The model is built using a neural network consisting of an embedding layer, followed by an LSTM layer to process the text sequence, and finally a Dense layer with a Softmax feature to identify the topic. The model was trained on tweet data and evaluated using performance metrics: accuracy, specificity, recall, and F1-score.

In addition, the system can classify any person's tweets from an external CSV file and display the percentage of topics they write about statistically and graphically. The project thus provides an effective tool for analyzing Arabic textual content and understanding users' interests on Twitter.