Beykoz University

FACULTY OF ENGINEERING AND ARCHITECTURE

Department of "Computer Engineering"

"CAPSTONE PROJECT - Literature Review"
60610CPEOZ-CME0168

Social Media Sentiment Analysis in Azerbaijani Language using NLP - Python

Istanbul, December - 2023

Adviser: Dr. Alp Sardağ

Leyla Abdullayeva - 1904010038

Capstone Project Design Document: Social Media Sentiment Analysis using NLP-Python

1. Introduction

This document outlines the design for a capstone project titled "Social Media Sentiment Analysis using NLP-Python." The project aims to leverage Natural Language Processing (NLP) techniques and Python libraries to analyze sentiment in social media content.

2. Project Motivation

Understanding public opinion and sentiment has become increasingly important in today's social media-driven world. This project seeks to develop a tool that can analyze sentiment in social media data, providing valuable insights for various applications like:

Market research: Analyzing customer sentiment towards brands and products
Political analysis: Understanding public opinion on political figures and events
Social change monitoring: Tracking sentiment around social issues and movements

3. Project Objectives

The project aims to achieve the following objectives:

Data Collection: Develop a mechanism to collect relevant social media data (e.g., tweets, comments) from chosen platforms based on specific keywords or topics.
Data Preprocessing: Clean and pre-process the collected data, including handling noise, tokenization, and normalization.
Sentiment Analysis: Implement NLP techniques, like sentiment lexicons and machine learning models, to analyze sentiment in the preprocessed data.
Visualization and Analysis: Develop visualizations and reports to present the extracted sentiment insights in a clear and effective manner.

4. Tools and Technologies

The project will utilize the following tools and technologies:

Programming Language: Python
NLP Libraries: NLTK, spaCy, TextBlob, VADER (Valence Aware Dictionary for sEntiment Reasoning)
Machine Learning Libraries: Scikit-learn, TensorFlow (optional)

Data Collection Tools: Twitter API, Social Media APIs (depending on chosen platforms)
Visualization Tools: Matplotlib, Seaborn, Tableau (optional)
Cloud Platform: Google Colab for efficient computing resources

5. References

- Bird, Steven, Edward Klein, and Ewan Klein. "Natural language processing and sentiment analysis: The birds of a feather sentiment lexicon." Journal of Cognitive Therapy and Research 53.4 (2009): 482-490.
- Pang, Bo, Lillian Lee, and Shivakumar Vaithyanathan. "Thumbs up!: Sentiment classification using machine learning techniques." Proceedings of the ACL on empirical methods in natural language processing. Association for Computational Linguistics, 2002.
- Sentiment Analysis and Opinion Mining by Bing Liu and Minqing Hu
- Natural Language Processing with Python by Bird, Klein and Loper

6. Project Design and Implementation

- Data Collection: Define the specific social media platforms and criteria for data collection (keywords, topics, timeframe). Explore APIs and scraping methods to gather relevant data.
- Data Preprocessing: Design a cleaning and preprocessing pipeline to handle tasks like text normalization, tokenization, removal of stop words, and handling noise like emojis and URLs.
- Sentiment Analysis: Choose appropriate NLP techniques for sentiment analysis. Implement sentiment lexicons like VADER or train a machine learning model on labeled sentiment data for more nuanced analysis.
- Visualization and Analysis: Develop visualizations (e.g., word clouds, sentiment distribution charts) and reports to effectively present the extracted sentiment insights and draw conclusions.

7. Evaluation and Testing

Evaluate the performance of the sentiment analysis model by comparing its predictions with human-labeled data. Analyze the accuracy, precision, and recall of the model and refine it based on the results.

8. Timeline and Deliverables

Define a project timeline with key milestones for each phase of development. Establish deliverables for each milestone, including code notebooks, reports, and presentations.

9. Conclusion

This capstone project aims to develop a practical tool for analyzing sentiment in social media data using NLP techniques and Python libraries. By providing valuable insights into public opinion and sentiment, this project has the potential to be beneficial for various applications across different domains.