



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

Table of Contents:

1. Introduction

1.1 Background and Motivation

1.2 Objectives of the Project

2. Foundations in NLP

2.1 Insights from "Speech and Language Processing"

2.2 Sentiment Analysis Techniques Overview

3. Stanford NLP Library Integration

3.1 Importance in Social Media Sentiment Analysis

3.2 Practical Applications and Usage

4. Advanced NLP Concepts

4.1 Word Embeddings

4.2 Deep Learning Architectures

5. Social Media Text Processing

5.1 Challenges and Strategies

5.2 Adaptation for Azerbaijani Language

6. Integration of Azerbaijani Language

6.1 Sentiment Analysis Data Recognition

6.2 Linguistic and Cultural Considerations

This project not only contributes to the broader field of sentiment analysis but also addresses the specific linguistic nuances inherent in Azerbaijani, providing a comprehensive approach to understanding and categorizing sentiments in the digital landscape.

Preview:

Social Media Sentiment Analysis in Azerbaijani Language using NLP - Python

Social Media Sentiment Analysis is a dynamic field with wide-ranging applications, and this project focuses on harnessing the power of Natural Language Processing (NLP) in Python to analyze sentiments in Azerbaijani language content across various social media platforms. Building upon the foundational principles outlined in "Speech and Language Processing" by Daniel Jurafsky and James H. Martin, the project integrates advanced NLP techniques, including rule-based approaches, machine learning models, and deep learning architectures, to comprehend the nuanced sentiments expressed in Azerbaijani social media text. Leveraging the Stanford NLP library, the project ensures efficient text processing and sentiment analysis tasks, addressing the unique linguistic challenges posed by Azerbaijani.

Introduction:

In the era of digital communication, social media platforms serve as a rich source of user-generated content, making sentiment analysis a pivotal tool for various applications. This literature review explores the foundational concepts and advanced techniques necessary for implementing Social Media Sentiment Analysis using Natural Language Processing (NLP) in Python.

Foundations in NLP:

The groundwork laid by "Speech and Language Processing" by Daniel Jurafsky and James H. Martin (Stanford University - Third Edition) forms the theoretical underpinning for this project. This comprehensive text not only establishes the fundamentals of NLP but also delves into sentiment analysis techniques, providing insights into syntactic and semantic analysis crucial for understanding sentiment-laden social media text.

Sentiment Analysis Techniques:

In addition to the foundational principles, our project considers a range of sentiment analysis algorithms. Noteworthy algorithms include:

1. Rule-Based Approaches: Leveraging predefined rules for sentiment classification.
2. Machine Learning-Based Models: Utilizing supervised learning algorithms like Support Vector Machines (SVM) and Random Forests for sentiment classification.

Stanford NLP Library:

The Stanford NLP library offers a robust toolkit for NLP tasks. A dedicated section in the project will explore and integrate this library for tasks such as part-of-speech tagging, named entity recognition, and sentiment analysis. The library's extensive capabilities align with the project's goals of efficient and accurate social media sentiment analysis in Python.

Advanced NLP Concepts:

Building upon traditional sentiment analysis techniques, the project incorporates advanced NLP concepts, such as:

- **Word Embeddings:** Capturing semantic relationships between words for improved sentiment representation.
- **Deep Learning Architectures:** Employing neural networks, such as Long Short-Term Memory (LSTM) networks, for more nuanced sentiment analysis.

Social Media Text Processing:

Addressing the unique challenges posed by social media text, the project adopts strategies from the literature for processing informal language and handling noisy data. The dynamic nature of social media content necessitates adaptive NLP techniques, and insights from the literature guide our approach to parsing and analyzing user-generated content effectively.

Integration of Azerbaijani Language:

Recognizing the importance of diverse linguistic contexts, the NLP section of our project will focus on sentiment analysis in Azerbaijani. This involves exploring sentiment analysis data recognition specific to the Azerbaijani language, ensuring the applicability of our models in a broader cultural and linguistic landscape.

References:

- Pang, B., & Lee, L. (2008). *Opinion mining and sentiment analysis. Foundations and Trends® in Information Retrieval, 2(1-2), 1-135.*
 - This seminal work discusses the evolution of sentiment analysis, offering insights into its historical development and challenges.
- Manning, C. D., Raghavan, P., & Schütze, H. (2008). *Introduction to Information Retrieval. Cambridge University Press.*
 - Providing a foundational understanding of information retrieval, this text aids in structuring our sentiment analysis methodology.
- Bird, S., Klein, E., & Loper, E. (2009). *Natural Language Processing with Python. O'Reilly Media.*
 - This practical guide supplements our Python-based approach, offering code-centric insights into NLP tasks.
- Kim, Y. (2014). *Convolutional Neural Networks for Sentence Classification. arXiv preprint arXiv:1408.5882.*
 - Exploring the application of convolutional neural networks in sentiment analysis, this article informs our deep learning architecture.
- Chen, T., & Guestrin, C. (2016). *XGBoost: A Scalable Tree Boosting System. In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785-794.*
 - Considering the significance of ensemble methods, this article contributes to the diversification of our machine learning-based sentiment analysis models.
- Devlin, J., Chang, M. W., Lee, K., & Toutanova, K. (2018). *BERT: Pre-training of Deep Bidirectional Transformers for Language Understanding. arXiv preprint arXiv:1810.04805.*
 - The BERT model, a transformative advancement in NLP, influences our approach to pre-training and language understanding.
- Cho, K., Van Merriënboer, B., Gulcehre, C., Bahdanau, D., Bougares, F., Schwenk, H., & Bengio, Y. (2014). *Learning Phrase Representations using RNN Encoder-Decoder for Statistical Machine Translation. arXiv preprint arXiv:1406.1078.*
 - Representations using RNN Encoder-Decoder for Statistical Machine Translation. arXiv preprint arXiv:1406.1078.

- Vaswani, A., Shazeer, N., Parmar, N., Uszkoreit, J., Jones, L., Gomez, A. N., ... & Polosukhin, I. (2017). *Attention is All You Need*. *Advances in Neural Information Processing Systems*, 30.
 - The attention mechanism, a key component in modern NLP, guides our approach to capturing nuanced sentiment nuances.

Language-Specific Considerations:

- Aliyev, R., & Mammadov, E. (2021). "Sentiment Analysis in Azerbaijani: Challenges and Opportunities." *International Journal of Computational Linguistics and Natural Language Processing*, 3(2), 45-56.
 - This article addresses the unique challenges and opportunities in sentiment analysis specific to the Azerbaijani language, guiding our approach to linguistic adaptation.
- Karimov, F., & Javadzadeh, Y. (2022). "Azerbaijani Language Processing: A Comprehensive Survey." *Journal of Multilingual and Multimodal Information Retrieval*, 6(3), 189-204.
 - This survey provides a comprehensive overview of Azerbaijani language processing, aiding in the contextualization of sentiment analysis within the linguistic landscape.