Analyzing Persian Twitter Sentiments on the Arbaeen Walk: A Comparative Study of LDA and BERTopic with the Arbaeen Tweets Dataset

Zeynab Didehkhani*

Abstract

As the world becomes increasingly interconnected, social media platforms like Twitter play a pivotal role in shaping public discourse. One significant topic that has emerged in recent years is the Arbaeen walk, a religious and political movement that has attracted global attention. This paper has two main objectives: first, to construct the "Arbaeen Tweets dataset," curated specifically for sentiment analysis; and second, to analyze Persian tweets related to the Arbaeen walk from 2021-2022. By employing natural language processing (NLP) techniques—namely Latent Dirichlet Allocation (LDA) and BERTopic—we aim to uncover prevalent themes and sentiments, referred to as "topics" in topic modeling. The "Arbaeen Tweets dataset" consists of 2,622 colloquial Persian texts from Twitter, labeled for sentiment analysis. Our findings indicate that while LDA demonstrates slightly superior quantitative performance, BERTopic offers greater coherence in topic interpretation.

Keywords: Text Mining, Topic Modeling, LDA, BERTopic, Dataset construction, Sentiment Analysis, Arbaeen, Foot-pilgrimage, Twitter

---

*Independent Researcher, `zkhani5319@yahoo.com`