SENTIMENT ANALYSIS OF HAJJ-RELATED CONTENT ON X

MOHAMED TAREK ELSAYED MOHAMED TORKY

UNIVERSITI TEKNOLOGI MALAYSIA

NOTES : If the thesis is CONFIDENTIAL or RESTRICTED, please attach with the letter from the organization with period and reasons for confidentiality or restriction

SENTIMENT ANALYSIS OF HAJJ-RELATED CONTENT ON X

MOHAMED TAREK ELSAYED MOHAMED TORKY

A thesis submitted in fulfilment of the
requirements for the award of the degree of
Master of Data Science

School of Computing
Faculty of Computing
Universiti Teknologi Malaysia

CHAPTER 1

INTRODUCTION

1.1 Introduction

The rise of social media has changed the way people share experiences, express their opinions, and discuss world events. Hajj, the yearly Islamic pilgrimage to Mecca, is one of the most major religious events to get worldwide attention online. Every year, millions of Muslims perform the Hajj, and many of them express their thoughts and experiences on X (previously Twitter), giving it a valuable source of real-time emotional and sociological information.

Understanding what people think around Hajj-related experiences on X gives useful insights for religious groups, scholars, and the government. Sentiment analysis, a Natural Language Processing (NLP) approach, may assist categorize and comprehend this massive amount of unstructured text data. This chapter describes the study's history, research challenge, objectives, and scope, with the goal of analyzing sentiment in Hajj-related tweets using simple machine learning algorithms.

1.2 Problem Background

People now interact globally via X (previously Twitter) by sharing their opinions and debating global concerns, including significant religious pilgrimage like Hajj. People share their feelings, experiences, and opinions on Hajj, ranging from spiritual gratitude to practical problems, creating a public narrative around this holy pilgrimage. These expressions are frequently unstructured and uploaded in enormous quantities, rendering human analysis impossible.

While sentiment analysis has been widely applied in commercial fields such as product and service reviews, limited research has focused on analyzing sentiments

related to Islamic practices, particularly Hajj. Given the international attention that Hajj receives, and the complex sentiments associated with it, the accurate and automated understanding of these public perceptions is crucial. This study proposes an NLP-based sentiment analysis system specifically focused on Hajj-related tweets to fill this gap.

1.3 Problem Statement

Despite the enormous number of Hajj-related tweets sent annually on X, there is an obvious lack of specific tools for assessing public opinion toward Hajj. Due to the number and speed with which these tweets are published, manual analysis is wasteful and unsustainable. Most existing sentiment analysis methods are general-purpose or designed for commercial domains, therefore they do not capture the religious and cultural subtleties connected with Hajj.

This limitation has an impact on a variety of stakeholders, including Islamic groups, researchers, and the government, all of whom require accurate methods to analyze public opinion, detect recurrent concerns, and identify disinformation. There is a need for a straightforward, purpose-built sentiment analysis tool that can automatically categorize and show sentiments in Hajj-related tweets.

1.4 Research Questions

This study aims to explore how the Hajj pilgrimage is perceived on X (formerly Twitter) by applying sentiment analysis to tweets. The research will be guided by the following questions:

(a) What is the overall sentiment of Hajj-related tweets on X — positive, negative, or neutral?

(b) Can a simple sentiment analysis model accurately classify Hajj-related tweets into sentiment categories?

(c) What are the most frequently used words and phrases in each sentiment category?

(d) What conclusions and insights can be drawn from users' sentiments about Hajj based on their tweets?

1.5 Research Objectives

(a) To collect and preprocess Hajj-related tweet data from X (formerly Twitter).

(b) To develop a sentiment analysis model using simple machine learning and NLP techniques to classify tweets into positive, negative, or neutral sentiments.

(c) To visualize the sentiment classification results using appropriate techniques such as charts and word clouds to represent sentiment trends.

1.6 Scope of Research

This research is limited to sentiment analysis of tweets specifically related to Hajj. The scope and constraints of this project are outlined as follows:

(a) The dataset will consist exclusively of Hajj-related tweets collected from X (formerly Twitter).
(b) Only English-language tweets will be included to maintain compatibility with the selected sentiment analysis tools and models.
(c) Sentiment will be categorized into three classes: positive, negative, and neutral.
(d) The implementation will be done using the Python programming language.
(e) Python libraries such as NLTK, VADER, and TextBlob will be used for natural language processing and sentiment classification.