MSc Project Report

Topic Recommendation System For

Digital Marketing In Sports

Using Natural Language Processing Techniques

Jayanga Anushan Jayathilake

2023

A report submitted as part of the requirements for the degree of

MSc Big Data Analytics at Robert Gordon University, Aberdeen, Scotland

1814812 – JA Jayathilake

Abstract

Using social media data, this research project intended to create a machine-learning model to anticipate current sports themes while appealing to specific user categories. A user-friendly interface was also developed to visualize the model's output. The research problems addressed in the study were successful sports tweet classification, data sources, appropriate machine learning and NLP technologies, use cases, and output visualization. The LDA model, which was chosen after reviewing literature, efficiently classified tweets based on keywords, yielding trending sports topics.

The research accomplished its first goal by developing the LDA model for topic prediction. It addressed the second goal with an interactive dashboard that provided easy access to recommended subjects and country-specific tweet distribution. Nonetheless, there were constraints. The laborious data integration procedure hampered the discovery of real-time trends, which is crucial for dynamic marketing campaigns. Unstructured data storage was also underutilized, limiting data processing and decision-making. The lack of a persistence layer for storing LDA models made retraining and real-time deployment difficult.

Future research recommendations included incorporating real-time data via systems such as Apache Kafka and NoSQL databases. Using Azure Databricks and growing datasets might improve model output and relevance. Data from platforms such as Facebook and Instagram, as well as an investigation of social media networks other than Twitter, were proposed. Finally, the machine learning technique used in this work improved the digital marketing environment by forecasting sport-related trends, demonstrating its potential for larger applications and its ability for continuous development.

Keywords: Machine learning, topic modeling, recommendation systems, digital marketing, social media, LDA model, real-time data integration, user-friendly visualization, natural language processing.