Name of Candidate: Abdalla Mahmoud Mohamed Abdelhamid
Degree: Master
Title of Thesis: An Approach for Real-Time IoT Data Stream Processing and Analytics
Supervisors:
1. Prof. Shereen Aly Taie   **2.** Prof. Haytham Tawfeek Al-Feel

Approval: /   /   **Department:** Information System

An Approach for Real-Time IoT Data Stream Processing and

Analytics

A thesis submitted to the Faculty of Computers and Artificial Intelligence - Fayoum University in Partial fulfillment of the requirements for the M. Sc.

degree in Information Systems

Submitted by:

Abdalla Mahmoud Mohamed Abdelhamid
Teaching Assistant, Information System Department
Faculty of Computers and Artificial Intelligence, Fayoum University

Supervised by:

Prof. Shereen Aly Taie
Professor, Computer Science Department,
Vice dean for postgraduate studies and
research affairs of the Faculty of Computers

and Artificial Intelligence,
Fayoum University

Prof. Haytham Tawfeek Al-Feel
Professor, Information Systems Department,
Faculty of Computers and Artificial

Intelligence,
Fayoum University

2023

The summary not more than 500 words

Name of Candidate: Abdalla Mahmoud Mohamed Abdelhamid
Degree: Master
Title of Thesis: An Approach for Real-Time IoT Data Stream Processing and Analytics
Supervisors:
1. Prof. Shereen Aly Taie   **2.** Prof. Haytham Tawfeek Al-Feel

Approval: /   /   **Department:** Information System

ABSTRACT

Online social networks like Twitter and Facebook have become essential for connecting people, disseminating information, and sharing interests since the internet's advent into our daily lives. Twitter, in particular, has grown rapidly and plays a significant role in analyzing user-generated data to personalize experiences and enhance recommendation tasks. Twitter posts serve as an important source of information for identifying users' positive interests on various topics, including their food preferences. A model of Food Interests Analysis (FIA) is proposed that leverages real-time processing and analytics of Twitter data streams. The primary objective is to extract insights and valuable information about users' food preferences and interests, enhancing user experiences, advancing the capabilities of personalized food recommendations, and leveraging them for Internet of Things (IoT)-based personalized services and artificial intelligence (AI) applications. Furthermore, business intelligence (BI) has become an important analytical technique for market forecasting and assessing consumer satisfaction and market demand. Since business intelligence requires in-depth analysis, sentiment analysis is the process of using natural language processing (NLP) and machine learning (ML) techniques to identify the emotional tone and attitude in text, making it useful for analyzing Twitter posts and customer reviews to identify user preferences and market demand. As a result, it's critical to choose relevant advertisements for users at particular locations to capture their attention and generate profit. The proposed FIA Model combines both topic modeling and sentiment analysis techniques by employing Latent Dirichlet Allocation (LDA) with term frequency-inverse document frequency (TF-IDF) for topic modeling and lexicon-based approaches for sentiment analysis. The proposed model integrates topic modeling and sentiment analysis techniques to enable the identification of user-specific food preference categories and interests, serving as a foundation for generating personalized food recommendations aligned with individual tastes and preferences. In addition, the proposed FIA model applied and evaluated two lexicon-based approaches to sentiment analysis for tweets using a dataset of 20,000 publicly available tweets collected, which were tested with five algorithms: Naive Bayes (NB), Support Vector Machine (SVM), Decision Trees (DT), Random Forest (RF), and Logistic Regression (LR). The Valence Aware Dictionary for Sentiment Reasoner (VADER) lexicon-based approach achieved the highest accuracy of 75% using the Logistic Regression classifier, while TextBlob achieved 73.1% accuracy using the Decision Tree classifier.

The summary not more than 500 words