DRUG ADVERSE EVENTS CLASSIFICATION USING SOCIAL MEDIA CONTENT

Ranith Sachintha Ranawaka

(179346C)

Degree of Master Of Science in Computer Science

Department of Computer Science and Engineering

University of Moratuwa

Sri Lanka

May 2021

Declaration

I declare that this is my own work and this dissertation does not incorporate without acknowledgement any material previously submitted for a Degree or Diploma in any other University or institute of higher learning and to the best of my knowledge and belief it does not contain any material previously published or written by another person except where the acknowledgement is made in the text. Also, I hereby grant to University of Moratuwa the non-exclusive right to reproduce and distribute my thesis, in whole or in part in print, electronic, or another medium. I retain the right to use this content in whole or part in future works.

....	....
Ranith Sachintha Ranawaka	Date

The above candidate has carried out research for the Master's thesis/ Dissertation under my supervision.

....	....
Dr. Surangika Ranathunga	Date

Acknowledgements

I would like to express profound gratitude to my advisor, Dr. Surangika Ranathunga, for her invaluable support by providing relevant knowledge, materials, advice, supervision, and useful suggestions throughout this research work. Her expertise and continuous guidance enabled me to complete my work successfully. Further, I would like to thank Mr. Sanmugan Aravinthan, for providing valuable resources for this project.

I am grateful for the support and advice given by Dr. Uthayasanker Thayasivam. Further, I would like to thank all my colleagues for their help in finding relevant research material, sharing knowledge and experience, and for their encouragement.

I am as ever, especially indebted to my parents and family for their love and support throughout my life. Finally, I wish to express my gratitude to all my colleagues at IQVIA Sri Lanka, for the support given to me to manage my MSc research work.

Abstract

On-time detection of possible adverse events a drug may have has been a critical issue for the pharmaceutical industry, although it undergoes rigorous clinical trials there still can be adverse effects once it reaches the market, this is known as post-market drug safety surveillance. The ordinary way to collect these was through physicians who prescribe the drug reporting back to the pharmaceutical company. But this process consumes time and has the risk of missing important drug adverse reactions.

The recent popularity of social media has led people to communicate extensively about their aspects in day-to-day life, this includes the communications of the experience regarding the drugs and their adverse events. This makes social media a rich resource for monitoring drugs after they reach the market.

In this research, we experiment with machine learning models including deep learning models using social media contents manually verified by health care professionals for the presence of drug adverse events. The Social media data has been acquired through popular health care social media channels from their respective APIs.

Well-known Text classification algorithms such as SVM and Logistic Regression provide the best accuracy for ADR mining, CNN's which has recently shown high accuracy levels for text classification also shows high levels of accuracy for ADR classification tasks.

Table of Contents

List of Figures **Page**

List of Tables

List of Equations

List of Abbreviations

Abbreviation	Description
ADR	Adverse Drug Reaction
OTC	Over the Counter
FDA	Food and Drug Administration
DoTs	Dose Time and Susceptibility
PV, PHV	Pharmacovigilance
CRM	Customer Relationship Management
URL	Universal Resource Locator
POS	Part of Speech
UMLS	Unified Medical Language System
SVM	Support Vector Machines
ME	Maximum Entropy
MNB	Multinomial Naïve Bayes
CTakes	Clinical Text Analysis and Knowledge Extraction System
jSRE	Java simple relation extraction
NLP	Natural Language Processing
BOW	Bag of Words
TF	Term Frequency
BOAW	Bag of Audio Words
TF-IDF	Term Frequency Inverse Document Frequency
SNOMED CT	Systematized Nomenclature of Medicine - Clinical Terms
ICH	International Council for Harmonization of Technical Requirements for Pharmaceuticals for Human
CRF	Conditional Random Fields
RNN	Recurrent Neural Network

CNN	Convolutional Neural Network
LSTM	Long Short-Term Memory