

## **DECLARATION**

I hereby declare that I carried out the work reported in this thesis in the Department of Information Science under the Faculty of Information and Communication Technology, University of Technology (Yatanarpon Cyber City), under the supervision of Dr Yi Yi Hlaing. I solemnly declare that to the best of my knowledges, no part of this thesis has been submitted here or elsewhere in a previous application for award of a degree. All sources of knowledge used have been duly acknowledged.

.....

29<sup>th</sup> October 2019

ZWE HTET PAING

6IST-70

## APPROVAL

This is to certify that the B.E thesis titled “**PERSONALITY PREDICTION USING SUPPORT VECTOR MACHINE**” carried out by **ZWE HTET PAING, 6IST-70** has been read and approved for meeting part of the requirements and regulations governing the award of the degree of Bachelor of Engineering (Information Science and Technology), Department of Information Science under the Faculty of Information and Communication Technology, University of Technology (Yatanarpon Cyber City), Myanmar.

1. Dr. Reenu

Pro-Rector (Academic)

University of Technology (Yatanarpon Cyber City)

.....

(Chairman)

2. Dr. Hnin Aye Thant

Professor and Head of Department of Information Science

Faculty of Information and Communication Technology

.....

(Co-Chairman)

3. Dr. Htet Ne Oo

Lecturer and Course Coordinator

Department of Information Science

Faculty of Information and Communication Technology

.....

(Course Coordinator)

4. Dr. Yi Yi Hlaing

Associate Professor

Department of Information Science

Faculty of Information and Communication Technology

.....

(Supervisor)

5. Daw Phyu Hnin Lwin

Lecturer

Department of Information Science

Faculty of Information and Communication Technology

.....

(Member)

## **ACKNOWLEDGEMENTS**

First of all, I would like to express my special thanks to Dr. Aung Win, Rector, University of Technology (Yatanarpon Cyber City), for initiating the Bachelor programme at the University of Technology (Yatanarpon Cyber City).

I would like to express my respectful gratitude to Dr. Reenu, Pro-rector, University of Technology (Yatanarpon Cyber City), for her kind permission to complete this thesis. I am thankful to her for giving enough consideration to her ideas and views.

I would like to express grateful thank to Dr. Hnin Aye Thant, Professor and Head of Department of Information Science, Faculty of Information and Communication Technology, University of Technology (Yatanarpon Cyber City), for her kind guidance, encouragement and supervision and patience in making my thesis to complete successfully.

I am very grateful to Dr. Htet Ne Oo, Lecturer and Course Coordinator of Final Year (Information Science and Technology), Department of Information Science, Faculty of Information and Communication Technology, University of Technology (Yatanarpon Cyber City), for her kind guidance and encouragement and supervision and patience in making my thesis to complete successfully.

I am very grateful to my supervisor, Dr. Yi Yi Hlaing, Associate Professor, Department of Information Science, Faculty of Information and Communication Technology, University of Technology (Yatanarpon Cyber City), for her kind guidance and encouragement. She has been very supportive in this thesis, and also guides a lot, particularly, at the level of quality of presentation.

Very Special thanks to Daw....., Lecturer, English Department, University of Technology (Yatanarpon Cyber City) for her valuable supports from the language point of view in my thesis work.

I would like to thank a lot to all my teachers for their mentoring, encouragement, and recommending this dissertation.

Finally, I am grateful to my parents and friends who specially offered strong moral and physical support, care and kindness, during the year of my thesis study.

## **ABSTRACT**

Use of social networking has increased tremendously in recent times. It has become popular method for information distribution and social interaction. Personality has been considered as the most difficult human attribute to understand. It is very important as it can be used to define the uniqueness of a person. Personality detection from text means to extract the behavior characteristics of authors written the text. Personality detection models could be very useful in various domains like e-learning, information filtering, collaboration and e-commerce by a user interface that adapts the interaction according to user's personality. This thesis presents a method by which a user's personality can be accurately predicted through the publicly available information on their Twitter profile using Support Vector Machine (SVM). Twitter is a popular social media platform with millions of users. The tweets shared by these users have recently attracted the attention of researchers from diverse fields. This thesis focuses primarily on predicting user's personality from the analysis of tweets shared by the user.

## CONTENTS

	Page
DECLARATION	i
APPROVAL	ii
ACKNOWLEDGEMENTS	iii
ABSTRACT	iv
CONTENTS	v
LIST OF FIGURES	vii
LIST OF EQUATIONS	viii
<b>CHAPTER 1 INTRODUCTION</b>	
1.1 Objectives	2
1.2 Field Background	2
1.2.1 Big Five Mode	2
1.3 Overview of the System	3
1.4 Organization of the Thesis	4
<b>CHAPTER 2 THEORY BACKGROUND</b>	
2.1 Data Managing	5
2.1.1 Numpy	5
2.1.1 Pandas	6
2.2 Natural Language Processing	6
2.2.1 NLTK	7
2.2.2 TextBlob	7
2.3 Twitter API	8
2.4 Tweepy	8
2.5 Machine Learning	9
2.5.1 Scikit-learn	9
2.5.2 TF-IDF	10
2.6 Support Vector Machine	12
2.6.1 Pros and Cons associated with SVM	18
2.6.2 Support Vector Regression (SVR)	19
2.6.3 Different between SVR and simple regression model	21
2.7 The Big Five personality traits	23
2.7.1 Descriptions of the particular personality traits	24

2.8 Chapter Summary	28
<b>CHAPTER 3 SYSTEM DESIGN AND IMPLEMENTAION</b>	
3.1 System Design	29
3.2 Methodology	29
3.2.1 Data Collection	29
3.2.2 Preprocessing Text	30
3.2.3 Feature Extraction	30
3.2.4 Classification and Regression using SVM	31
3.2.4.1 Classification	32
3.2.4.2 Regression	33
3.3 Evaluation Approaches	34
3.4 The Basic SVM Classification Calculation Example	35
3.5 Experiment and Results	38
3.6 System Implementation	40
3.7 Chapter Summary	42
<b>CHAPTER 4 CONCLUSION</b>	
4.1 Conclusion	43
4.2 Limitation of the system	43
4.3 Future extension	43
<b>REFERENCES</b>	

## LIST OF FIGURES

<b>Figure</b>	<b>Page</b>
2.1 Data Splitting and Output Prediction	10
2.2 Support Vectors	13
2.3 Kernel Trick	16
2.4 Maximal Margin SVM	17
2.5 Kernel SVR	21
2.6 Blue line: Hyper Plane; Red Line: Boundary Line	22
2.7 Data Points	22
3.1 System Design	29
3.2 Support Vector Machine	37
3.3 Main Page	40
3.4 Login to Twitter Page	40
3.5 Result Page	41
3.6 Timeline Page	41
3.7 Post Page	42

## LIST OF EQUATIONS

Equation	Page
2.1 Support Vector Machine	14
2.2 Support Vector Machine	14
2.3 Support Vector Machine	14
2.4 Kernel SVM	16
2.5 Kernel SVM	16
2.6 Support Vector Machine	17
2.7 SVM decision function	18
2.8 Linear SVR	20
2.9 Non-linear SVR	20
2.10 Non-linear SVR	20
2.11 Kernel SVR	21
2.12 Kernel SVR	21
2.13 Hyper plane equation	22
2.14 Boundary line equation	23
2.15 Boundary line equation	23
2.16 Support Vector Regression	23
3.1 TF-IDF	31
3.2 TF-IDF	31
3.3 TF-IDF	31
3.4 Accuracy	34
3.5 Precision	34
3.6 Recall	34
3.7 F1 Score	34
3.8 RSME	35
3.9 $R^2$	35