

TABLE OF CONTENT

| | |
|--|------|
| TABLE OF CONTENT | ix |
| LIST OF TABLES | xi |
| LIST OF FIGURES | xii |
| ABSTRACT..... | xiii |
| Chapter 1 INTRODUCTION | 1 |
| 1.1 Background..... | 1 |
| 1.2 Problem | 2 |
| 1.3 Problem Limitation | 2 |
| 1.4 Objective..... | 3 |
| 1.5 Thesis Organization | 3 |
| Chapter 2 LITERATURE REVIEW | 4 |
| Chapter 3 BASIC THEORY | 7 |
| 3.1 Data Classification | 7 |
| 3.1.1 Multinomial Naïve Bayes | 8 |
| 3.1.2 Support Vector Machine | 9 |
| 3.1.3 Decision Tree | 10 |
| 3.1.4 Performance Measure..... | 12 |
| 3.2 Twitter User Historical Data | 13 |
| 3.2.1 Tweet..... | 14 |
| 3.2.2 Tweet Location | 14 |
| 3.2.3 Tweet Date | 15 |
| 3.2.4 User Hometown..... | 16 |
| 3.3 Geo-mapping Visualization..... | 16 |
| 3.3.1 CartoDB Geomap Visualization | 16 |
| 3.4 Python and Machine-Learning library | 17 |
| Chapter 4 ANALYSIS AND MODEL DESIGN..... | 18 |
| 4.1 Model Analysis..... | 18 |
| 4.1.1 Overview Model | 18 |
| 4.1.2 Generating insight and deciding features..... | 20 |
| 4.2 Model Design..... | 22 |
| 4.2.1 Pre-processing and Feature Extraction | 23 |
| 4.2.2 Labelling Data Training..... | 24 |
| 4.2.3 Classification..... | 25 |
| 4.2.4 Performance Measurement..... | 26 |
| Chapter 5 IMPLEMENTATION | 27 |
| 5.1 Pre-processing and Feature Extraction..... | 27 |
| 5.2 Labelling on Data Training and Predicted Data Testing..... | 33 |
| 5.3 Classification..... | 33 |
| 5.3.1 Multinomial Naïve Bayes Step | 33 |

| | | |
|---|--|----|
| 5.3.2 | Support Vector Machine Step | 35 |
| 5.3.3 | Decision Tree Step..... | 38 |
| 5.4 | Performance Measure..... | 40 |
| Chapter 6 EXPERIMENT AND DISCUSSION..... | | 41 |
| 6.1 | Experiment and Results | 41 |
| 6.1.1 | Classifier Prediction: | 41 |
| 6.1.2 | Accuracy on Data Training..... | 42 |
| 6.1.3 | Accuracy on Predicted Data Testing | 42 |
| 6.1.4 | GeoMap Visualization..... | 43 |
| 6.2 | Discussion | 44 |
| Chapter 7 CONCLUSION AND FUTURE WORK..... | | 45 |
| 7.1 | Conclusion | 45 |
| 7.2 | Future work..... | 45 |
| BIBLIOGRAPHY | | 47 |