

Nanyang Technological University

**AI6102:
Machine Learning Methodologies & Applications
Term Paper**

Kaggle Competition (Knowledge)
Real or Not? NLP with Disaster Tweets

Ron Kow Kheng Hui
Brandon Chua Shao Jie
School of Computer Science and Engineering

Abstract

Table of Contents

| | |
|---|------------|
| Abstract | i |
| List of Tables | iii |
| List of Figures | iv |
| List of Abbreviations | v |
| 1 Introduction | 1 |
| 1.1 Problem Statement | 1 |
| 1.2 XX | 1 |
| 2 Proposed Solution | 3 |
| 2.1 Challenges | 3 |
| 2.2 Raw Data | 3 |
| 2.3 Bag of Words Model | 3 |
| 2.4 Performance Metrics | 3 |
| 3 Model Development | 4 |
| 3.1 Modelling Process | 4 |
| 3.2 Feature Cleaning | 4 |
| 3.3 Feature Engineering | 4 |
| 3.4 Algorithms | 4 |
| 3.5 Experiments | 4 |
| 3.5.1 Classification Algorithms | 4 |
| 3.5.2 Cross-validation Results | 4 |
| 3.6 Leaderboard Results | 4 |

| | |
|--|----------|
| 4 Conclusion | 5 |
| 4.1 Summary | 5 |
| 4.2 Lessons Learned | 5 |
| 4.3 Kaggle Competition Page and Source Files | 5 |
| 4.4 Allocation of Work | 5 |
| Bibliography | 5 |
| Appendices | 7 |
| A xxx | 7 |

List of Tables

| | | |
|-----|---------------|---|
| 1.1 | XXX | 2 |
|-----|---------------|---|

List of Figures

| | | |
|-----|---------------|---|
| 1.1 | XXX | 1 |
|-----|---------------|---|

List of Abbreviations

Chapter 1

Introduction

1.1 Problem Statement

In this project we will:

-

1.2 XX

Figure 1.1 illustrates this process.

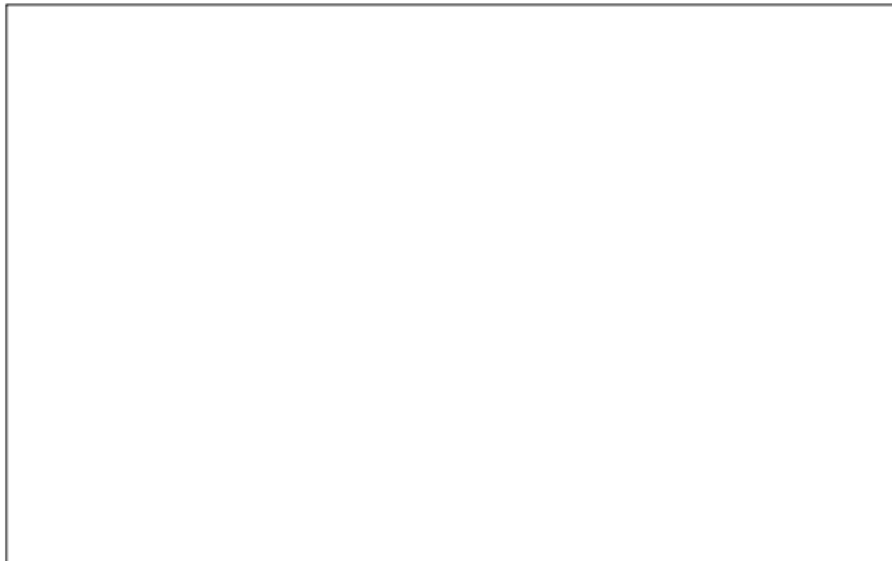


Figure 1.1: XXX

$$xxx \tag{1.1}$$

Table [1.1](#) shows XXX.

Table 1.1: XXX

| Notation | Description |
|----------|-------------|
|----------|-------------|

xxx

Chapter 2

Proposed Solution

2.1 Challenges

2.2 Raw Data

2.3 Bag of Words Model

2.4 Performance Metrics

Chapter 3

Model Development

3.1 Modelling Process

3.2 Feature Cleaning

3.3 Feature Engineering

3.4 Algorithms

3.5 Experiments

3.5.1 Classification Algorithms

3.5.2 Cross-validation Results

3.6 Leaderboard Results

Chapter 4

Conclusion

4.1 Summary

4.2 Lessons Learned

4.3 Kaggle Competition Page and Source Files

4.4 Allocation of Work

The source files, including the data, models, and results, are accessible at:

<https://github.com/ronkow/kaggle-disaster-tweets>

Bibliography

Appendices

A xxx