# Final Report

*June 22, 2020*

**Abstract**

Here is a summary

## Contents



# 1 Introduction

## 1.1 Objectives

## 1.2 Data

# 2 Data Science Methods

## 2.1 Information Extraction

## 2.2 Feature Engineering

## 2.3 Machine Learning Pipeline

The machine learning pipeline can be divided into 5 steps: data splitting, preprocessing, feature selection, hyperparameter tuning and model selection. The Feature experiment is conducted to evaluate the use of features using the finalist model. Baseline models are provided from two perspectives: dummy model is used as classifier baseline model while Bag-of-word (BOW) model and TF-IDF model are used as feature baseline models. Because of the limited dataset, cross-validation is applied through the whole pipeline in order to use our data in a more efficient way and make a more accurate out-of-sample estimate.

# 3 Data product and results

## 3.1 Data Product

## 3.2 Results