CPS803 Group 26: Project Proposal

Adriano Mariani
Ryerson University
Toronto, Canada
adriano.mariani@ryerson.ca

Arsalan Khuwaja
Ryerson University
Toronto, Canada
akhuwaja@ryerson.ca

Alize De Matas
Ryerson University
Toronto, Canada
adematas@ryerson.ca

Jasmine Joy
Ryerson University
Toronto, Canada
jasmine.joy@ryerson.ca

Our project aims to exploit machine learning techniques in detecting a probable suicide message based on social media posts. Our training data was taken from *r/SuicideWatch* and *r/Depression* Reddit communities. For this purpose, we will train and test classifiers such as Naïve Bayes, Support Vector Model and Logistic Regression to distinguish Reddit posts that indicate *suicide* and *non-suicide*. The word associations derived from each method could be used to identify posts with suicidal tendencies. The least biased method of identifying intent of suicide in text will be used to evaluate social media posts accumulate before and during the time of COVID-19.

Index Terms – Machine Learning, Supervised Learning, Natural Language Processing, Naïve Bayes, Binary Classification, Logistic Regression, Suicide.

I. Introduction

As a result of technology, we can live in a virtual world where people can create their own digital personas. Social media platforms like Reddit are spaces where people can escape the realities of the real world to openly vent their feelings and emotions online. Sharing memes, getting insight from others' and being a part of communities are all helpful coping mechanisms. However, many people in the state of depression share sensitive information about wanting to commit suicide, which should not go unnoticed. Suicide behaviours such as suicide ideation and attempts are regarded as a major predictor of death by suicide. Even individuals who experience persistent and severe suicide ideation, but do not all subsequently commit suicide, are at an increased risk of attempting suicide. Therefore, predicting the individuals who engage in suicide ideation by screening messages posted on Reddit would be effective in preventing suicide and helping clinical psychiatrists gain more insight on social media engagements of high-risk patients.

Every day, an average of more than 10 Canadians die by suicide. Groups in higher risk of suicide include men and boys, people serving federal sentences, survivors of suicide loss and survivors of a suicide attempt, First Nation and Metis communities and LGBTQ youth.[1] There are several known socio-demographic, physical and psychological factors which influence suicide mortality. However, using machine learning, a subset of artificial intelligence in which the computer generates predictive rules based on raw data can efficiently predict suicide risk in the virtual world and aid in the creation of preventative strategies.[2] [3] [4]

II. Method

A. Datasets

Three datasets will be used for this project.

1. *"Suicide and Depression Detection"* contains 232,074 unique values and classifies a user as suicidal or non-suicidal based on the text of their reddit post.
2. *"Depression_suicide"* contains 20,364 unique values and contains posts from *r/depression* and *r/SuicideWatch*. For the purposes of our project, we will consider *r/SuicideWatch* posts to be "*suicide*", and *r/depression* posts to be "*non-suicide*".
3. "*Suicide_notes*" contains 464 unique values. The notes were written by users who were confirmed with suicidal tendencies.

TABLE 1: DATASETS

Datasets	Intended Use	Rows	Description
suicide_detection.csv	Training	232,074	Data from Kaggle[5]. Data contains reddit posts that have been labelled as suicide and non-suicide.
depression_suicide.csv	Training & Test	20,364	Data from Kaggle[6]. Data contains reddit posts from *r/depression* and *r/suicidewatch*.
suicide_notes.csv	Test	464	Data from Kaggle[7]. Notes written by users who were confirmed with suicidal tendencies.

B. Text Processong and Conversion

1. Preprocessing our input data would include removing whitespace, non-alphanumeric characters, and Unicode characters.
2. Parsing large input data with more than 150 characters into smaller sentences.
3. Using the NLP library BERT, to give us the word embeddings of the input sentences in vector form. These vectors will be used to represent the features of the inputs used in the classification and training algorithms.

III. Classification Algorithms

The classification and training algorithms considered for this project are: Naive Bayes, Support Vector Model, and Logistic Regression. *Suicide_Detection.csv* will be used as the training dataset. The algorithms will classify inputs as suicidal or non-suicidal. The results from each algorithm will then be compared.

1. Naïve Bayes is a supervised classification algorithm where the variables are independent of each other. It is simple, fast, and scalable. We will use the *sklearn* and scikit-learn libraries available in Python to access the *StandardScaler, LabelEncoder* and *GaussianNB* models.
2. Support Vector Model is a supervised classification algorithm that finds the hyperplane that optimally separates a dataset into two distinct classes. We will use the *sklearn* library available in Python to access the *svm* models.
3. Logistic Regression is a classification algorithm used to predict the probability that the outcome is equal to 1 (*suicide*). It models predictions through the use of a sigmoid function, and classifies the posts are *suicidal* or *non-suicidal*. We will use the *sklearn* library available in Python to access the *LogisticRegression* models.[8]

IV. Intended Experiments

For the scope of our project, posts are only targeted for indication of suicide. The outputs generated will be labelled as 'suicide;' or 'non-suicide'.

To identify biases and to ensure that classification models are accurately predicting the outcome in a suitable manner, we have designed two experiments.

A. Experiment 1

Dataset 'Suicide_notes.csv' contains notes that were written by users who were confirmed with suicidal tendencies.

Anaysis of Experiment 1: For Suicide_notes.csv, it is expected that a large percentage of the posts will be labelled as 'suicide'. The notes in suicide_notes.csv will be classified independently by the three classification models. The model that labels the largest number of posts as 'suicide' would be considered the most accurate model of this experiment.

B. Experiment 2

Dataset '*depression_suicide.csv*' contains reddit posts from *r/depression* and *r/SuicideWatch.*

Anaysis of Experiment 1: It is assumed that the posts from *r/Suicidewatch* would be classified as sucicide. The posts in reddit_depression_suicidewatch.csv will be classified independently by the three classification models. The model that labels most of the *r/suicidewatch* posts as '*suicide*', and *r/depression* posts as '*non-suicide*' will be considered the most accurate model of this experiment. The average of the posts classified as '*suicide*' by each method is taken, and the model with the value closest to the average would be considered the more accurate algorithm.

C. Project Analysis

The words that are used to classify a post as 'suicide' are identified. The correlation between the frequency of words used in classification and prediction are compared. The least biased method that uses the most accurate definition is then identified.

D. Additional Analysis

Social media posts from 2020 to 2021 will be gathered. The least biased classification method identifying intent of suicide will be used to evaluate the posts.

V. Planning and Milestones

Our milestones will be broken down as follows:

A. Preprocessing

Milestone 1 (2 weeks): Preprocessing input data, parsing data with more than 150 characters into smaller sentences.
Milestone 2 (2 weeks): Using the BERT for feature extractions and to develop word embeddings of the input sentences in vector form.

B. Training

Milestone 3 (3 weeks): Training and modelling classification algorithms

- Training Naive Bayes pipeline and model
- Training Support Vector Model pipeline and model
- Training Logistic Regression pipeline and model

C. Evaluation

Milestone 4 (3 weeks): Evaluating classification algorithms

- Evaluate Naive Bayes pipeline and model
- Evaluate Support Vector Model pipeline and model
- Evaluate Logistic Regression pipeline and model

D. Wrap-up

Milestone 5 (2 weeks): Failure analysis and refinement
Milestone 6 (1 week): Report and video

References

[1] Namratha, P., Kishor, M., Sathyanarayana Rao, T. S., & Raman, R. (2015). Mysore study: A study of suicide notes. Indian journal of psychiatry, 57(4), 379–382. https://doi.org/10.4103/0019-5545.171831

[2] Ryu, S., Lee, H., Lee, D. K., Kim, S. W., & Kim, C. E. (2019). Detection of Suicide Attempters among Suicide Ideators Using Machine Learning. *Psychiatry investigation*, *16*(8), 588–593. https://doi.org/10.30773/pi.2019.06.19

[3] Canada, P. H. A. of. (2021, September 17). *Government of Canada*. Canada.ca. Retrieved October 5, 2021, from https://www.canada.ca/en/public-health/services/suicide-prevention/suicide-canada.html.

[4] McMullen, L., Parghi, N., Rogers, M. L., & Yao, H. (10/01/2021). *The role of suicide ideation in assessing near-term suicide risk: A machine learning approach* Elsevier. doi:10.1016/j.psychres.2021.114118

[5] Komati, N. (2021, 05 19). Suicide and Depression Detection. Retrieved from Kaggle: https://www.kaggle.com/nikhileswarkomati/suicide-watch

[6] Mashaly, M. (2020, 07 04). Suicide Notes. Retrieved from Kaggle: https://www.kaggle.com/mohanedmashaly/suicide-notes

[7] Rigoulet, X. (2021, 08 21). Reddit dataset: r/depression and r/SuicideWatch. Retrieved from Kaggle: https://www.kaggle.com/xavrig/reddit-dataset-rdepression-and-rsuicidewatch

[8] Reference: M. P. LaValley, "Logistic Regression," Circulation, vol. 117, no. 18, pp. 2395–2399, 2008.