INFORMATICS INSTITUTE OF TECHNOLOGY

INFORMATICS INSTITUTE OF TECHNOLOGY

In Collaboration with

UNIVERSITY OF WESTMINSTER

Suicide Detection using Sentimental analysis

A Project Proposal by

Mr. S.C Fernando

Supervised by

Mr Ishara Erandi

This Project Proposal is submitted in partial fulfilment of the requirements for the BSc (Hons) Software engineering degree at the University of Westminster

October 2022

Contents

1. Project Proposal

1.1 Introduction

In the age where internet is the leading source of information sharing almost 63% of the whole global population has access to internet and social media, as more people spend time in social media everyday it eventually becomes a major part of our day to day life. Every year an average of 700,000 people are recorded to die from suicide mostly in the age gaps of 15-29 years, as most of the people spend their time on social media suicidal or not gives us a glimpse of their habits and emotions through their digital footprint left on social media made by sharing or posting certain content, From Platforms such as twitter, Facebook and even WhatsApp data of the content people share can be used as a tool to understand human emotion which will enable us to prematurely detect suicidal intents in people.

1.2 Problem domain

1.2.1 NLP and Sentimental analysis

Natural language processing also known as NLP are one of the core branch in machine learning that uses models to understand and analyze human language in depth, This technology is widely used in text to speech, google assistant, text analysis to name a few. Sentiment analysis is one of the sub branches of NLP where use cases for such a technology can be seen in Amazon or eBay product recommendation system where it processes our likes and dislikes of certain products and recommends items we are more likely to be interested in.

This technology is further used in analyzing online reviews on certain products and predicting the customers emotion towards that certain product.

1.2.2 Suicide detection with sentiment analysis

When talking about online suicide detection the users behavior can be tracked by the things she/him share ,texts or posts this is where sentiment analysis comes into play, training the model with full sentences and categorizing the sentence with keywords related to suicide and depressing such as “bullying”, stress etc.…shows one way of approaching this problem(Jung, Park and Song, 2017).

1.3 Problem Definition

With the growth of the internet in the past few decades cyberbullying has become a prominent problem in the internet. Sentimental analysis is mostly used in market research to understand what the customer likes and dislikes and stay ahead of the competition. Currently the research done on online suicide detection systems with sentiment analysis is very little and most of them are not explored in depth, work like (Theng et al., 2021)shows how sentimental analysis can be used in detecting cyberbullying with a trained model with twitter datasets. The same concept could be implemented in our suicide research using sentimental analysis.

1.3.1 Problem Statement

Sentimental analysis has been used in cyberbullying research but much research or implementations haven't been done on suicide detection with sentimental analysis.

1.4 Research Motivation

In the modern era where technology and social media is a part of the day to day life of a person it is a common occurrence of cyberbullying, this is a root cause of suicide and depression. Many in depth research has been done on cyberbullying detection, but very minimal research or implementation have been done for suicide detection using sentimental analysis.

1.5 Existing work

Citation	Brief Description	Limitations	Contribution
(Jung et al., 2021)	Classification using text features and metadata, training model with random forest and gradient boosting machines.	Need for very large datasets for accurate results.	Utilizing metadata features to increase effectiveness and accuracy.

(Castillo-Sánchez et al., 2020)	Sentiment analysis, deep learning, gradient boosting machines	Need for large data sets.	Identifying the machine learning algorithms for suicide risk assesment.

1.6 Research Gap

The research gap that you will be addressing in your research.

Expectation versus actual

There are different type research gaps

1. Theoretical gap
2. **Performance gap**
3. Empirical gap

1.7 Contribution to the Body of Knowledge

By addressing the above gap what is the contribution you are going to make

1. Technological contribution
2. Domain contribution

1.8 Research Challenge

Evidence for complexity and challenge to achieve, you need to write such that it gives reason why it could lead to a publication.

Publishable doesn't mean it is publishable in a conference but publishable in a https://mjl.clarivate.com/search-results journal

Further evidence to show that this can be further extended to PhD research

1.9 Research question/s

https://www.scribbr.com/research-process/research-questions/

1.10 Research Aim

One sentence

Further elaborate on the aim

1.11 Research Objective

Elaborate the steps of atomic activities that you need to carryout to achieve the aim

Research Objectives	Explanation	Learning Outcome
Problem Identification		LO1
Literature Review	RO1 RO2 RO3	LO1
Data Gathering and Analysis		LO2, LO3
Research Design		
Implementation		
Testing and Evaluation		

1.12 Project Scope

1. **In-scope**
2. **Out-scope**
3. **Diagram showing prototype feature**

2.1 Methodology

1. **Research methodology**

Research Philosophy	The author of the research has selected the positivism as the research philosophy
Research Approach	Deductive or inductive why?
Research Strategy	Experiment, survey => questionnaire (can be quantitative or qualitative) or interview (can be quantitative or qualitative),
Research Choice	Mono method => only one method can quantitative (Positivist) or qualitative (interpretivist), Multi method (More than one method but all belong to same paradigm (positivist or interpretivist)) or Mixed method (only pragmatist can mix the method => mixing the method from positivism and interpretivism)
Time zone	Cross-sectional or longitudinal

2. **Development methodology**
    a. **What is the life cycle model and why?**

b. **Design methodology => SSADM or OOAD or Anything else?**

c. **Evaluation methodology => Evaluation metrics and/or benchmarking**

3. **Project management methodology**

    a. **Schedule using the Gantt Chart after doing a WBS (Do not have to provide the WBS)**

    b. **Deliverables, milestones and dates of deliverables**

    c. **Resource requirements**

        i. **Hardware requirements**

        ii. **Software requirements**

        iii. **Skills requirements**

        iv. **Data Requirements**

    d. **Risk Management**

Risk Item	**Severity**	**Frequency**	**Mitigation Plan**
	5	**5**	
	5	**4**	
	5	**1**	

References

Castillo-Sánchez, G. et al. (2020). Suicide Risk Assessment Using Machine Learning and Social Networks: a Scoping Review. *Journal of Medical Systems*, 44 (12). Available from https://doi.org/10.1007/s10916-020-01669-5.

Jung, H., Park, H.A. and Song, T.M. (2017). Ontology-based approach to social data sentiment analysis: Detection of adolescent depression signals. *Journal of Medical Internet Research*, 19 (7). Available from https://doi.org/10.2196/jmir.7452.

Jung, W. et al. (2021). Suicidality Detection on Social Media Using Metadata and Text Feature Extraction and Machine Learning. *Archives of Suicide Research*. Available from https://doi.org/10.1080/13811118.2021.1955783.

Theng, C.P. et al. (2021). Cyberbullying Detection in Twitter Using Sentiment Analysis. *International Journal of Computer Science and Network Security*, 21 (11).