

**Faculty of Technology**

**Department of Computer Science**

**PLANNING REVIEW**

|  |
| --- |
| PROM02  Academic Year: 2021/22 |

# Terms of Reference (50%)

## Project Title

*A Comparison of Machine Learning Techniques for Twitter Cyberbullying Detection*

## Project Overview

*Guidance: provide the aim, objectives, research question and practical outcomes of your project:*

***Aim****: This study aims to evaluate the performance of feature extraction techniques and machine learning classifiers employed in cyberbullying detection using F1 score, precision, recall, and accuracy as performance indicators.*

***Objectives****:*

* *Conduct a thorough literature review to discover the machine learning classifiers and feature extraction techniques extensively utilized for the identification of cyberbullying on social media platforms.*
* *Compare the performance of three different feature extraction techniques on three popular machine learning classifiers for cyberbullying detection.*
* *Develop an automated detection model with improved classifier efficiency by incorporating the best of three tested feature extraction techniques*
* *Present a real-time cyberbullying detection platform capable of detecting and reporting cyberbullying tweets for flagging.*

***Research Question****: What sorts of machine learning techniques and methodologies are widely utilized in twitter cyberbullying identification? which feature extraction techniques can be used to improve identification? Can a model for the automated identification of cyberbullying be built in real time with accuracy and F1score over 90% and less processing time than 2mins?*

***Practical outcomes****:*

* *A comparison of previous various cyberbullying detection techniques*
* *A model which presents comparison of various techniques to detect cyberbullying on twitter. This is further divided into data acquisition, data pre-processing, feature extraction, classification and evaluation*
* *A prototype data driven web app for detecting and reporting cyberbullying tweets*

## Underpinning research with Literature Review

|  |  |  |  |
| --- | --- | --- | --- |
| *Citation* | *Brief summary of paper* | *Relevance to your research question* | *Relevance to practical outcome of project* |
| Ates, E.C., Bostanci, E. and Guzel, M.S. (2021) ‘Comparative Performance of Machine Learning Algorithms in Cyberbullying Detection: Using Turkish Language Preprocessing Techniques’. arXiv. Available at: <https://doi.org/10.48550/arXiv.2101.12718>. | *This paper compared the performance of various machine learning algorithms for identifying cyberbullying in Turkish communications.*  *Precision, recall, accuracy and F1 score values were used to evaluate the performance of classifiers* | *Some of the machine learning algorithms used in the study will be compared alongside different feature extraction techniques* | *Facilitates comparative study of the machine learning algorithms* |
| Arif, M. (2021) ‘A Systematic Review of Machine Learning Algorithms in Cyberbullying Detection: Future Directions and Challenges’, *Journal of Information Security and Cybercrimes Research*, 4(1), pp. 01–26. Available at: <https://doi.org/10.26735/GBTV9013>. | *Evaluated the present state of the art in cyberbullying research.*  *Proposed a paradigm that considers all potential cyberbullying actors* | *Comprehensive information on the state-of-the-art classifiers and feature extraction techniques used in cyberbullying* | *Gives Basis for comparative studies on feature extraction techniques* |
| Gholizadeh, S. (2022) *Top Popular Python Libraries in Research*. Available at: <https://doi.org/10.22541/au.164580055.55493761/v1>. | *This paper discusses the most common python packages/libraries used in research* | *Important packages for developing the model on python programming language* | *Useful for model development and web app deployment* |
| Gupta, P. *et al.* (2018) ‘A Proposed Framework to Analyze Abusive Tweets on the Social Networks’, *International Journal of Modern Education and Computer Science*, 10, pp. 46–56. Available at: <https://doi.org/10.5815/ijmecs.2018.01.05>. | *A comparative analysis of multiple classification techniques that facilitates scalability.*  *Presented a strategy based on adjectives that surpasses the existing approach based on cosine similarity in terms of recall and precision.* | *Compared some machine learning algorithms and presented a novel feature extraction technique* | *Facilitates comparative study and the introduction of new models.* |
| Hani, J., Nashaat, M., Ahmed, M., Emad, Z., Amer, E. and Mohammed, A., 2019. Social media cyberbullying detection using machine learning. *Int. J. Adv. Comput. Sci. Appl*, *10*(5), pp.703-707. | *Using various classifiers, this paper proposed a supervised machine learning approach for detecting and preventing cyberbullying.*  *The proposed approach was evaluated based on the accuracy of the classifiers, with Neural Network outperforming others such as Support Vector Machine.* | *It shows some of the machine learning methods used in the detection and prevention of twitter cyberbullying.* | *One of the related studies cyberbullying techniques to be compared* |
| Hee, C.V, Jacobs G, Emmery C, Desmet B, Lefever E. *et al.* (2018) ‘Automatic detection of cyberbullying in social media text’, *PLOS ONE*, 13(10), p. e0203794. Available at: <https://doi.org/10.1371/journal.pone.0203794>. | * *described the collection and fine-grained annotation of a cyberbullying corpus for English and Dutch* * *Made use of LSVM classifier* * *Evaluated model performance with f1score* | *Shows the nlp and machine learning techniques used in cyberbullying detection on social media.* | *One of the cyberbullying detection techniques to be compared.* |
| Liu, Q. *et al.* (2018) ‘Text Features Extraction based on TF-IDF Associating Semantic’, in *2018 IEEE 4th International Conference on Computer and Communications (ICCC)*. Chengdu, China: IEEE, pp. 2338–2343. Available at: <https://doi.org/10.1109/CompComm.2018.8780663>. | *After excluding words with low TD-IDF value this paper used the word2vec model to train the word vector in the corpus to obtain its semantic features* | *Feature extraction techniques to improve text classification* | *Gives insight to TD-IDF, one of the feature extraction techniques that will be compared and analyed in this work* |
| Saha, T., Saha, S. and Bhattacharyya, P. (2019) ‘Tweet Act Classification : A Deep Learning based Classifier for Recognizing Speech Acts in Twitter’, in *2019 International Joint Conference on Neural Networks (IJCNN)*, pp. 1–8. Available at: <https://doi.org/10.1109/IJCNN.2019.8851805>. | *This work provides a deep learning-based tweet act classifier for comprehending the content and meaning of tweets and identifying the most useful conversations among the tweeters.* | *Novel technique for extracting features* | *Helps in understanding hand picked feature extraction techniques* |

# Project Schedule (20%)

*Guidance: the project schedule should be provided as a series of tables as detailed below.*

## Table 1: Effort

*Guidance: Decompose each of the major tasks in your project into sub-tasks, identifying deliverables and deadlines*

|  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- |
| Task Id | Task Name | Start | Deadline | Hours | Deliverable |
| 1.0 Project conception, initiation and preliminary research | | | | | |
| 1.1 | Conceptualization and project proposal | 13/06/2022 | 30/06/2022 | 44 |  |
| 1.2 | Understanding the project requirements through supervisor’s help, project planning | 04/07/2022 | 13/08/2022 | 60 |  |
| 2.0 Literature Review | | | | | |
| 2.1 | Searching, reading and getting notes from quality papers | 15/08/2022 | 21/08/2022 | 24 |  |
| 2.2 | Literature review on cyberbullying domain | 22/08/2022 | 28/08/2022 | 20 |  |
| 2.3 | Literature review on NLP in detection of cyberbullying | 29/08/2022 | 04/09/2022 | 20 |  |
| 2.4 | Literature review on Machine learning in detection of cyberbullying | 05/09/2022 | 11/09/2022 | 20 |  |
| 2.5 | Literature review on Feature Extraction techniques used in cyberbullying detection (TD-IDF, Word2Vec, DistilBert, gloVe) | 12/09/2022 | 18/09/2022 | 20 |  |
| 2.6 | Classifiers used in cyberbullying detection (LR, SGD, RF) | 19/09/2022 | 25/09/2022 | 20 | D1 |
|  |  |  |  |  |  |
| 3.0 Methodology | | | | | |
| 3.1 | Methodology groundwork and writeup | 26/09/2022 | 30/09/2022 | 20 |  |
| 3.2 | Data Acquisition | 03/10/2022 | 12/10/2022 | 12 | D2 |
| 3.3 | Model Design (NLP Pipeline and model construction) | 13/10/2022 | 20/10/2022 | 24 |  |
| 3.4 | Model Development | 21/10/2022 | 30/10/2022 | 28 |  |
| 3.5 | Model Deployment | 31/10/2022 | 6/11/2022 | 20 | D3 |
| 3.6 | Model Testing and Management | 7/11/2022 | 13/11/2022 | 24 | D4 |
| 4.0 Result, Discussion, and Conclusion | | | | | |
| 4.1 | Evaluation Metrics | 14/11/2022 | 20/11/2022 | 20 |  |
| 4.2 | Performance Results of Feature Extraction Techniques | 21/11/2022 | 27/11/2022 | 20 |  |
| 4.3 | Performance Results of Classifiers | 28/11/2022 | 4/12/2022 | 20 | D5 |
| 4.4 | Conclusion | 5/12/2022 | 11/12/2022 | 20 |  |
| 4.5 | Correction, review, finalizing dissertation and viva preparation | 12/12/2022 | 25/12/2022 | 40 | D6 |

## Table 2: Deliverables

*Guidance: Add project deliverables in chronological. Deliverable numbers needed to be included.*

|  |  |  |
| --- | --- | --- |
| Del. No. | Name | Deadline |
| D1 | Table of comparison of related research, completion of literature review | 25/09/2022 |
| D2 | Acquiring an Open-Source Dataset | 12/10/2022 |
| D3 | Completed Jupiter notebook/python code | 06/11/2022 |
| D4 | Functional/Non functional Spec, A GUI/Data driven web App | 13/11/2022 |
| D5 | Result Analysis with tables and graphs | 04/12/2022 |
| D6 | Prepared viva presentation | 25/12/2022 |

## Table 3: Milestones

*Guidance: Add milestones in chronological order and how you will know these have been achieved.*

|  |  |  |  |
| --- | --- | --- | --- |
| Milestone | Name | Deadline | Evidence |
| M1 | Completed Literature review with table of comparison | 25/09/2022 | D1 |
| M2 | Cyberbullying Detection Model | 06/11/2022 | D2, D3 |
| M3 | Prototype Data Driven Web App | 13/11/2022 | D4 |
| M4 | Results and Conclusion | 25/12/2022 | D5, D6 |

## Table 4: Outline Schedule / Gantt chart

*Guidance: Identify the timeline for each major Task / Objective (as in the effort table above) with related deliverables / milestones – as in the example below.*

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | June | | | | | July | | | | | August | | | | | September | | |
|  | 1 | 2 | 3 | 4 | 5 | | 6 | 7 | 8 | 9 | | 10 | 11 | 12 | 13 | | 14 | 15 |
| Project Conception |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |
| Literature Review |  |  |  |  |  | |  |  |  |  | |  |  |  |  | | M1  D1 |  |
| Methodology |  |  |  |  |  | |  |  |  |  | |  |  |  |  | |  |  |

|  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
|  | October | | | | | November | | | | | December | | | | | January | | |
|  | 16 | 17 | 18 | 19 | 20 | | 21 | 22 | 23 | 24 | | 25 | 26 | 27 | 28 | | 29 | 30 |
| Methology |  | D2 |  |  | M2  D3 | | D4 |  |  |  | |  |  |  |  | |  |  |
| Result &Concl |  |  |  |  |  | |  |  |  |  | | D5 |  | M4  D6 |  | |  |  |

# Evaluation Plan (10%)

*Guidance: Complete the following table - one page maximum.*

|  |  |  |
| --- | --- | --- |
| Objective | Evaluation Approach | Evidence |
| Conduct a thorough literature review to discover the machine learning classifiers and feature extraction techniques extensively utilized for the identification of cyberbullying on social media platforms.  Compare the performance of popular machine learning classifiers for cyberbullying detection.  Develop an automated detection model with improved classifier efficiency by incorporating and testing various feature extraction techniques.  Present a real-time cyberbullying detection platform capable of detecting and reporting cyberbullying tweets for flagging. | Search online for recent related quality papers to study how the issue cyberbullying identification is addressed  Classifiers, feature extraction methods, datasets used and results from the related quality papers studied will be compared  Acquire a suitable dataset, pre-process/clean the data, then 3 or 4 feature extraction techniques will be used on the clean data and their effect on 3 of the best classifiers found in related studies will be compared. The feature extraction technique and the classifier pair to produce the best result when evaluated with accuracy, f1score, precision and recall will become the model.  The model will be pickled and turned to a light web app/graphical user interface which can detect a cyberbullying tweet and subsequently report it for it to be brought down. | Documentation of my literature review.  A Table of comparison between related studies showing feature extraction techniques, classifiers used and results from previous studies.  A written code in python programming language with clear and well organised package structures  A web app with Easy-to-use interface that can detect cyberbullying whose success will be demonstrated by its ability correctly classify a tweet as bullying or not and report it for flagging if it is |

# Social, Ethical, Legal and Professional issues (20%)

## Social, Ethical, Legal and Professional Issues Table

*Guidance: reflect on your project considering the context, sector, users, potential stakeholders, etc. Consider how your project sits with legislation, such as GDPR, are your users under 18 or vulnerable? How does it relate to current government strategy and focus? Why is it needed – is your project responding to a social or professional need? What impact might it have on that profession? Are there ethical issues for your project – is the data sensitive, could your results have implications for work or life? Not every project will have issues across all 4 categories but do critically consider this so you can be certain your project does not. The whole table should be no longer than a page.*

|  |  |
| --- | --- |
| Social issues | *The effects of cyberbullying can be devastating and corrosive for victims, particularly young people. These victims may develop socio-psychological issues such as depression, loneliness, low self-esteem, school phobias, suicide thoughts, and social anxiety as a result of being bullied online.*  *The aforementioned problems can be remedied by the identification and prevention of instances of cyberbullying, the basis of this project.* |
| Ethical issues | *Tweets are typically gathered for research and analysis via the platform's API. Users need not be informed that their tweets are being collected in any way. Users of Twitter are not contacted directly to obtain their informed consent. This is due to the fact that it has been granted when the user accepted the terms of service for Twitter. Notwithstanding, data collected are for solely for research and analysis in accordance with twitters term of service. There is no need to put in for approval to the university ethics committee because direct participants are not used* |
| Professional issues | *My professional approach to the project includes conducting literature research, attending drop-in sessions on dissertation writing, and obtaining the skills, information, and behaviors required to successfully navigate the data science environment.*  *Furthermore, regular consultation sessions with my supervisor are always eye-opening, keeping me on track and assisting me in developing competence.* |
| Legal Issues | *The entirety of legitimate access to Twitter’s data is based on legal contracts, and as Twitter's conditions encompass many aspects of good ethical practice, compliance with these terms may help simplify the legal rationale required for this research.*  *This study will make a legitimate use of the data by working with the data, or accessing the data, in a way that does not violate the terms of the contract that Twitter has with its users..* |

## Ethics Approval

*Proposal doesn’t require ethics approval as direct participant won’t be used in the acquisition of the dataset*

# Appendices

*For example…*

*Ethics Approval Application*