School of Electronic Engineering and Computer Science
Queen Mary University of London

MSc PROJECT DEFINITION
2020-21

This project definition must be undertaken in consultation with your supervisor. The feasibility of the project should have been assessed and the project aims should be clearly defined.
Submission of this document implies that you have discussed the specification with your supervisor.

Project Title: Identifying Hate Speech Categories In Social Media

Supervisor: Dr Arkaitz Zubiaga

Student name: Jack Ryan Cracknell

Student email: EC20046@QMUL.ac.uk

Student phone number: 07754568563

PROJECT AIMS:

The aim of this project is to develop a model which is able to detect hate speech in social media posts alongside the category that the hate speech belongs to. Example categories that will be assigned to hate speech messages include Homophobia, Racism, Sexism, Transphobia and Ableism. While hate speech detection is a popular topic for research, the idea of adapting the problem by implementing subclasses adds a unique twist and by creating such a model, it should improve the ease at which social media sites can identify what type of hate speech is most rampant and aid in the solution of stemming it.

There are a few notable challenges that this project must be able to overcome. As this project won't be creating a dataset from scratch but rather combining a number of pre-existing datasets used for other research papers, there may be some difficulty when merging multiple sources together. There is no standard way to create datasets for this task, therefore some datasets will need to be adapted.
Furthermore, one technical challenge will be the definition of what is regarded as hate speech. A formal definition will be decided on within the project.

PROJECT OBJECTIVES:
List a series of objectives you need to achieve in order to fulfil the aims of your project.

An investigation into the current state-of-the-art techniques for hate speech detection is a key objective that will fuel the rest of the project. The information gathered here will lead to an informed decision on how to progress with selecting datasets and building models.
Merge together datasets from various sources, overcoming any compatibility issues. A plan for a standardised data format may need to be developed if the composition of the datasets are vastly different.
Develop, test and evaluate models against the current state-of-the-art methods that were investigated in the early stages of the project.

METHODOLOGY:

1) Research the topic area, what are the state of the art methodologies? Is there much research into hate speech subclasses? What do they do?

2) Decide upon some datasets from different hate speech detection areas. If the datasets are different in format, choose a standard and reformat the data.

3) Combine datasets so we have a mix of different hate speech types

4) Develop a model which can classify hate speech / non hate speech

5) Develop a model which takes the identified hate speech messages and labels them with a subclass depending on the type of hate used

6) Create visualisations and other evaluation metrics to report on the successfulness of the project

PROJECT MILESTONES
Indicate what measurable/tangible components you will produce as part of this project. This may take the form of deliverable document(s) or developmental milestones such as a working piece of software/hardware.

REQUIRED KNOWLEDGE/ SKILLS/TOOLS/RESOURCES:

Indicate as far as possible the skills that are required for you to undertake this project. Also include any software, hardware or other tools or resources that you believe you will need.

TIMEPLAN

This can be a GANTT chart submitted with this document or a list of tasks, milestones and deliverables with timings.