Project Journal

Mental Health Analysis using Machine Learning

Aryan Ramani
23235438

This document contains the project journal that is maintained over course of the project. This journal will be divided into 2 parts: The tasks carried out by me and the timeline of the tasks.

Tasks

1. **Selecting Text Dataset related to Mental Health**: A dataset available on Kaggle was selected for this tasks. The aim was to select a dataset containing text that is suitable for the entire project.
2. **Conducting Data Analysis**: This part consisted of conducting data analytics for the text dataset. It consisted of data preprocessing, cleaning, stop word removal, and lemmatisation. Since, matplotlib or seaborn libraries does not consists of word clouds a separate library was used to create word cloud for data visualisation.
3. **Uploading Data to Database**: We used MongoDB as a database. I configured a cluster on cloud to get a connection string so that we can upload our datasets to the remote database. Both cleaned and uncleaned dataset were uploaded on the MongoDB cluster.
4. **Training of Machine Learning Models**: I used a couple of machine learning algorithms that could be suitable for the task in hand. The models used were logistic regression and Naive Bayes as they are popular in text classification.
5. **Training of Deep Learning Models**: I conducted deep learning experiments for text dataset using BERT which is a very popular model for sequence classification. I conducted an experiment where we use 2 custom iterations of BERT model, one with dropout and one without dropout. These experiments were conducted on Google Colab but a download copy is also available on the GitHub repository.
6. **Report Compilation**: I documented my experiments and incorporated it with the project report for the entire group.
7. **Video presentation**: I presented the second part of the video which was related to the text dataset.
8. **README file**: Developed the README file for the project.

Timeline

Each task was performed on a separate date. Here is a timeline for each task performed.

15 December, 2024: Task 1
20 December, 2024: Task 2
22 December, 2024: Task 4
26 December, 2024: Task 3
29 December, 2024: Task 5
4 January, 2025: Tasks 6, 7 and 8.

- Each of the first 5 tasks were done in 3-4 hours session.
- Tasks 6 required conducting research for references, documenting the work which took around 6-8 hours of work.
- Tasks 7 and 8 took around 10 minutes and 30 minutes each.