

Depression Detection

- Week 1

Name: Sinchana Kumbale

University: Manipal Institute of Technology

Internship Period: Jan – July 2024

Undergraduate 4th year

Duration of the presentation: ~7 minutes

Agenda

1. Overview of the Research Landscape

1. Overview of existing literature
2. Highlights from key papers

2. Key Findings and Insights

1. Insights from prior work
2. Ideas for the project

3. Potential areas of exploration

4. Tentative plan for the week

Overview of Research Landscape

1. Built and tested several basic ML models (SVM, DT etc) with their best performing LR model trained on Twitter data and tested on Reddit data and a diary. [2]
2. Used a hierarchical network attention model compared against pretrained transformers. Use LIWC to analyse content and style [3]

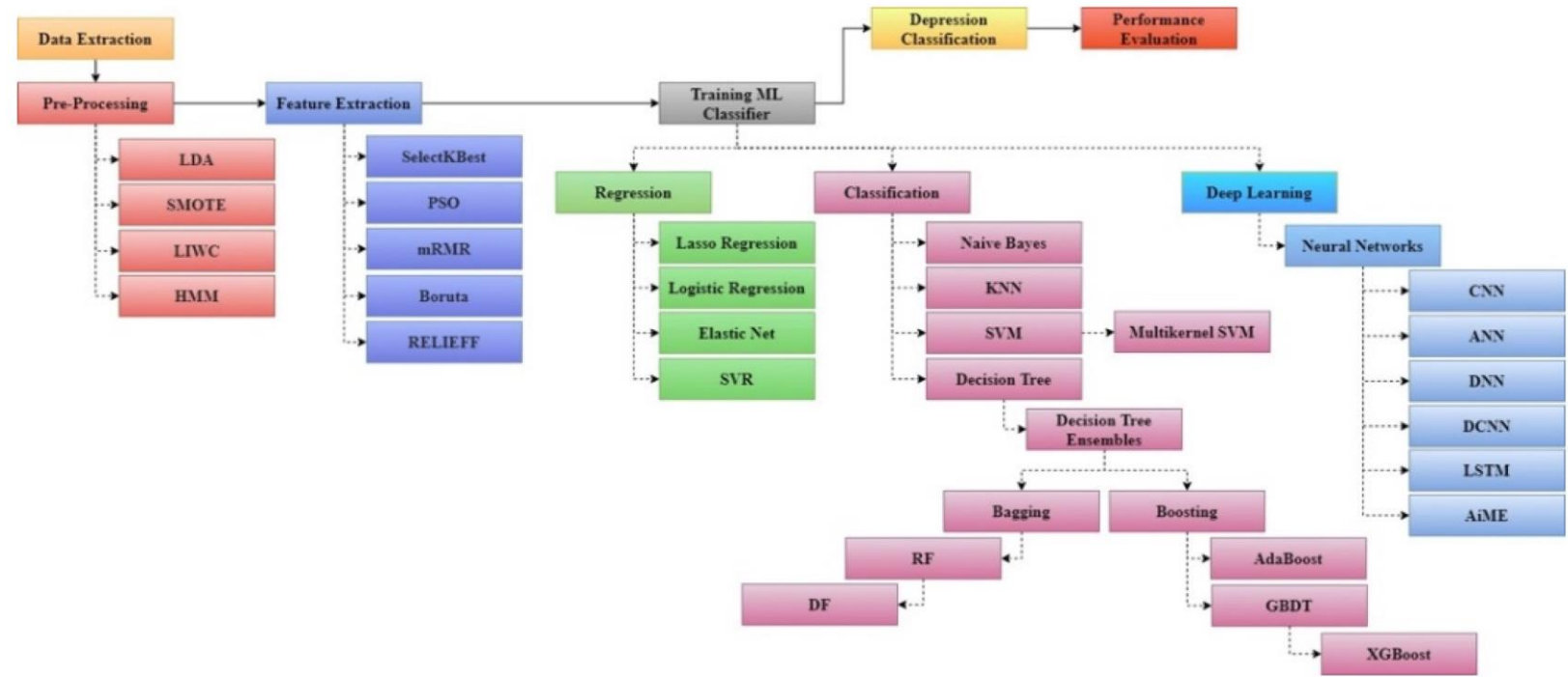


Fig 1. General model reflecting potential algorithms [1]

- [1] Machine Learning Algorithms for Depression: Diagnosis, Insights, and Research Directions (Aleem et al, 2022) | [Paper](#)
- [2] A textual based feature approach for depression detection using machine learning classifiers and social media texts (Chiong et al, 2021) | [Paper](#)
- [3] An emotion and cognitive based analysis of mental health disorders from social media data (Uban et al, 2021) | [Paper](#)

Key Findings and Insights

1. Using data from social media for text-based detection of depression is supported by Fig 2
2. Removing words like "depression" "diagnosis" from dataset increases generalizability [2]
3. A potential point of bias results from the demographic of social media users whose age lie between 18-49 [4]

Some ideas:

1. Utilizing a dataset along the lines of [1] and [5] where users' tweets over a period of time is analysed
2. Potentially using a hierarchical architecture by varying the classification model trying out LLMs or transformers seems a possible avenue
3. Trying out unsupervised learning or using an imbalanced dataset

[4] Fair and Explainable depression detection in Social Media, (Adarsh et al, 2023) | [Paper](#)

[5] Monitoring Depression Trends on Twitter During the COVID-19 Pandemic: Observational Study (Zhang et al, 2021) | [Paper](#)

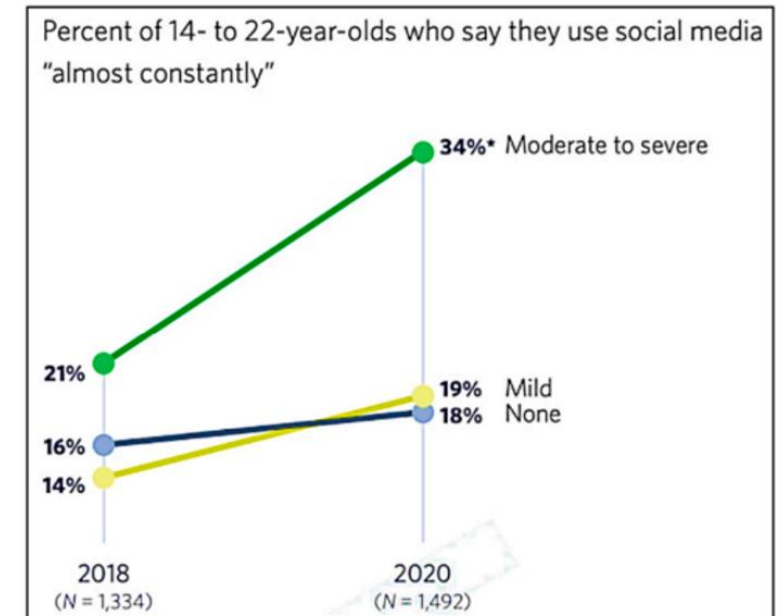


Fig 2. Frequency of usage of social media and depressive symptom level [2]

Tentative Plan

Plan for next week

1. Analyse papers along similar lines
2. Look at codes for relevant papers at paperswithcode
3. Try a preliminary data collection methodology
4. Get data from <https://erisk.irlab.org/>

Relevant Links

1. Overall project plan and timeline: [Link](#)
2. Analysis and notes from relevant papers: [Link](#)

Overall Timeline:

- Month 1:
 - Week 1 and 2 – Intense **literature review** on methodologies used for this task
 - Week 3 and 4 – Select weigh potential baseline models
- Month 2
 - Week 1 and 2 – **Implement the baseline systems** and debug potential issues
 - Week 3 and 4 – **Measure the performance** of systems and compare them
- Month 3
 - Week 1 and 2 – Conduct a gap analysis and start **developing potential solutions**
 - Week 3 and 4 – Conduct an initial test, refine, and optimise based on results you get
- Month 4
 - Week 1 and 2 – Draft a literature review based on what is conducted as of now
 - Week 3 and 4 – Iteratively **continue developing the solution** while outline what the methodology and results section might look like
- Month 5
 - Week 1 and 2 – **Evaluate** the developed model and collect metrics. Find potential cases of failure in your system and areas of future development.
 - Week 3 and 4 – **Complete writing** all remaining sections of the research
- Month 6
 - Week 1 and 2 – Make **final refinements** to the code and the written report
 - Week 3 and 4 – Ensure that all **documentation** is completed, proofread and finalise.