ANALYSIS OF STRESS, ANXIETY, AND DEPRESSION AMONG STUDENTS USING REDDIT COMMUNITY DATA

Prahathish Kameswaran   **Roopak Narayanasamy**   **Soundarya Venkatesh**   Team: Soup Boys

INTRODUCTION

Education is an important milestone in our lives, but it can be very stressful. For some the impact might be more than others. We aim to provide insights into the stress patterns of students in universities. We hope this will help the universities understand the cause of their students' distress and come up with ways to support them.

BACKGROUND

Studies show that 8/10 students reported symptoms of anxiety during their time at the university [1].

SDG 3:
Good Mental Health and Well-being

Stressors:
- Examinations
- Competition
- Finance
- Personal

Effects:
- Cognitive deficit
- Insomnia
- Mood swings
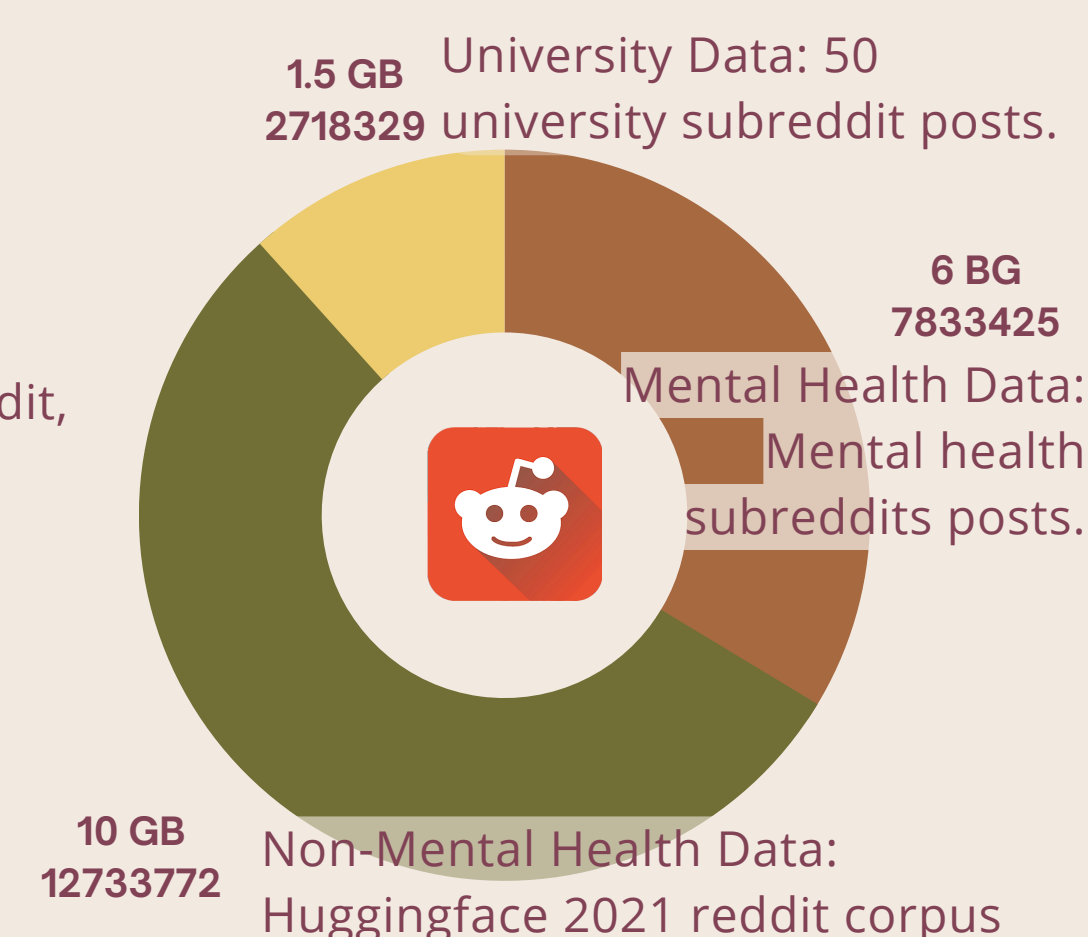- Physical illness

Previous work:

Mental health has received a lot of attention in the past, this is mainly because we have access to huge volumes of people's thoughts over the internet and big data frameworks now make it possible to analyze them. We particularly took inspiration from the below work:

- A Social Media Based Index of Mental Well-Being in College Campuses [2]
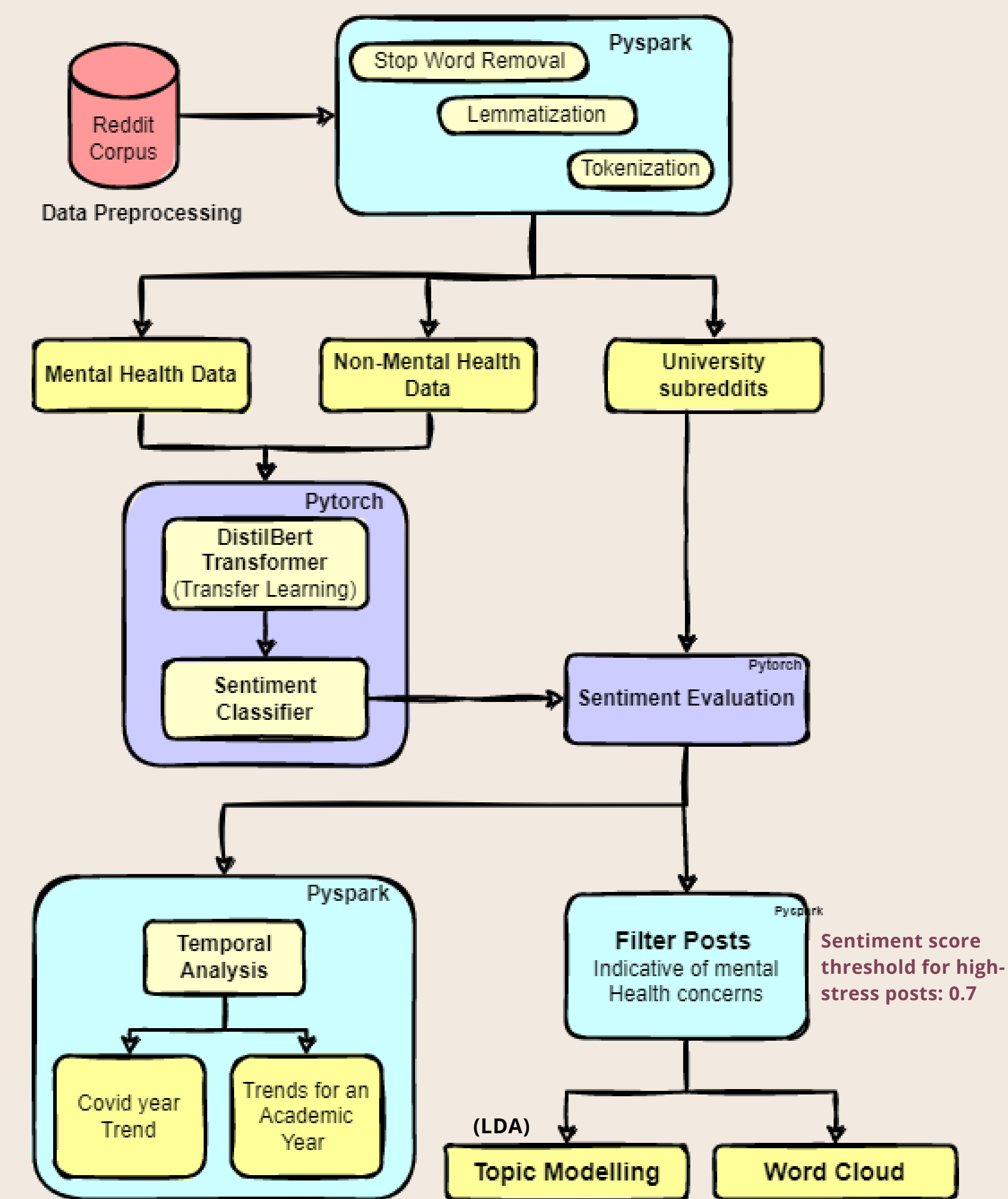- Modeling Stress with Social Media Around Incidents of Gun Violence on College Campuses. [3]

DATA

Source: PushShift API, HuggingFace dataset

Labels: created_utc, subreddit, title, selftext, score, over_18, upvote_ratio and is_video

1.5 GB — University Data: 50
2718329 university subreddit posts.

6 BG
7833425
Mental Health Data: Mental health subreddits posts.

10 GB
12733772
Non-Mental Health Data: Huggingface 2021 reddit corpus

METHOD

Data Preprocessing — Reddit Corpus → Pyspark (Stop Word Removal, Lemmatization, Tokenization)

Mental Health Data | Non-Mental Health Data | University subreddits

Pytorch — DistilBert Transformer (Transfer Learning) → Sentiment Classifier

Pytorch — Sentiment Evaluation

Pyspark — Temporal Analysis (Covid year Trend, Trends for an Academic Year)

Pyspark — Filter Posts: Indicative of mental Health concerns — Sentiment score threshold for high-stress posts: 0.7

(LDA) → Topic Modelling | Word Cloud

Hyperparameters for the Sentiment Classifier:
- batch_size = 64,
- max_length = 512,
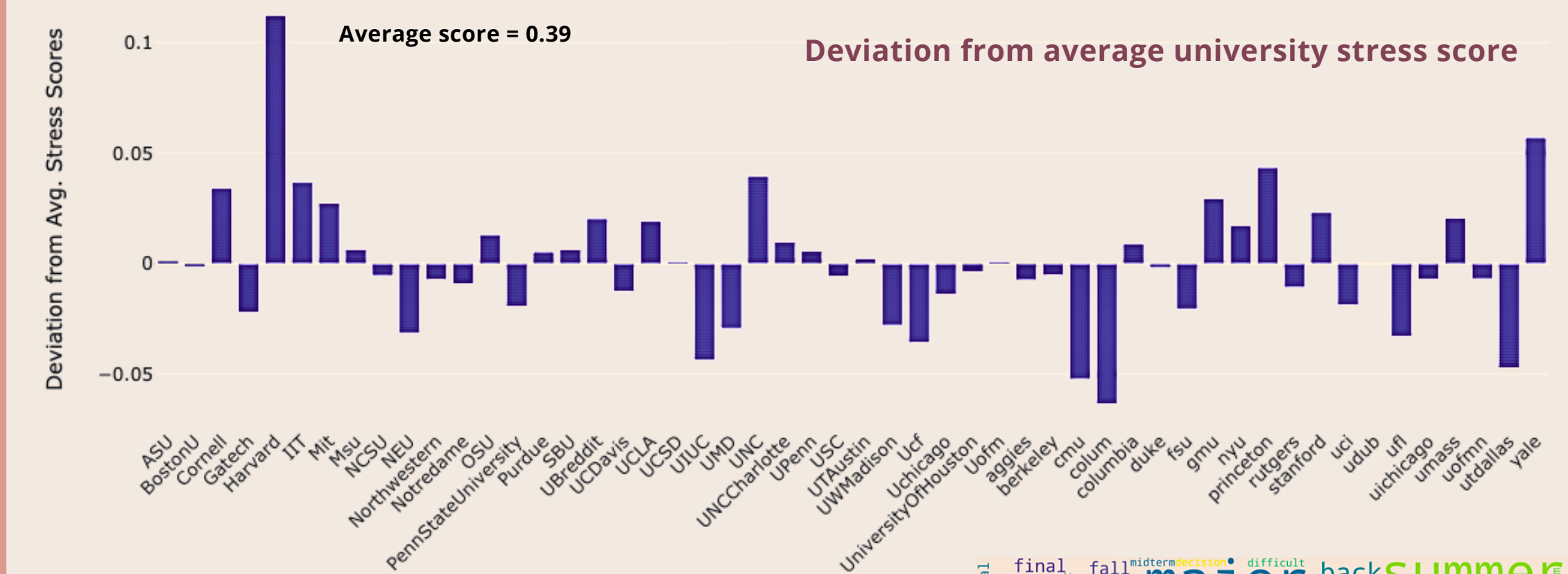- learning_rate = 16e-4,
- num_epochs = 1,
- dropout = 0.1

Correctness measures for sentiment classifier:
- Accuracy: 0.9101
- Precision: 0.8931
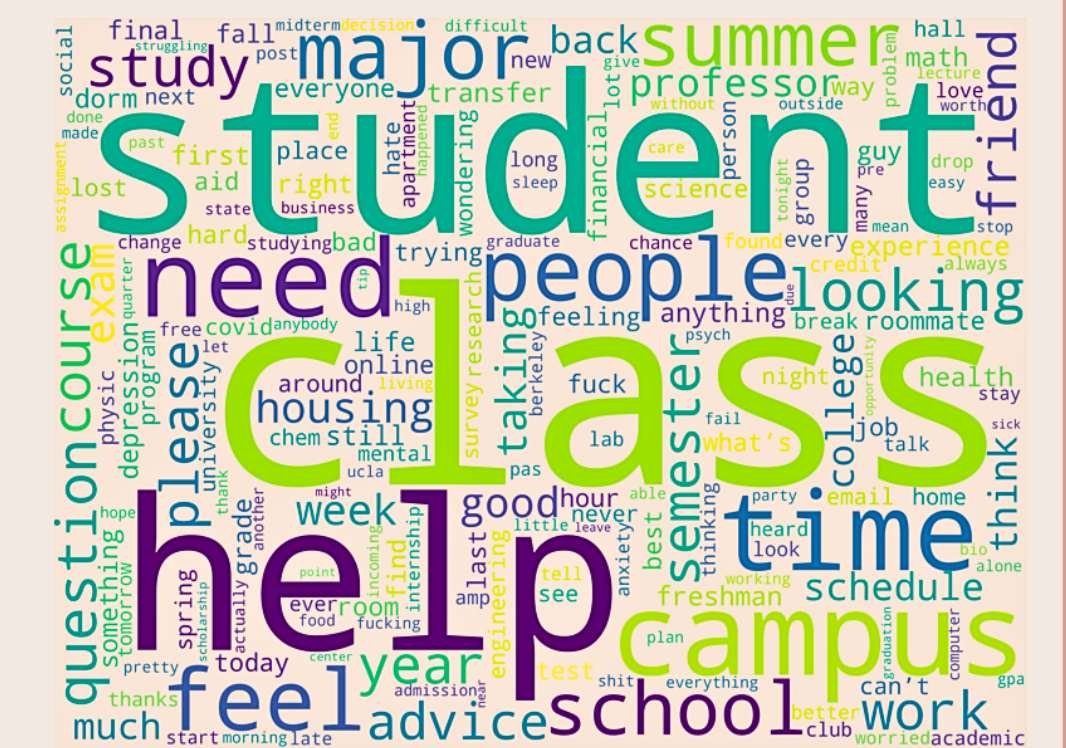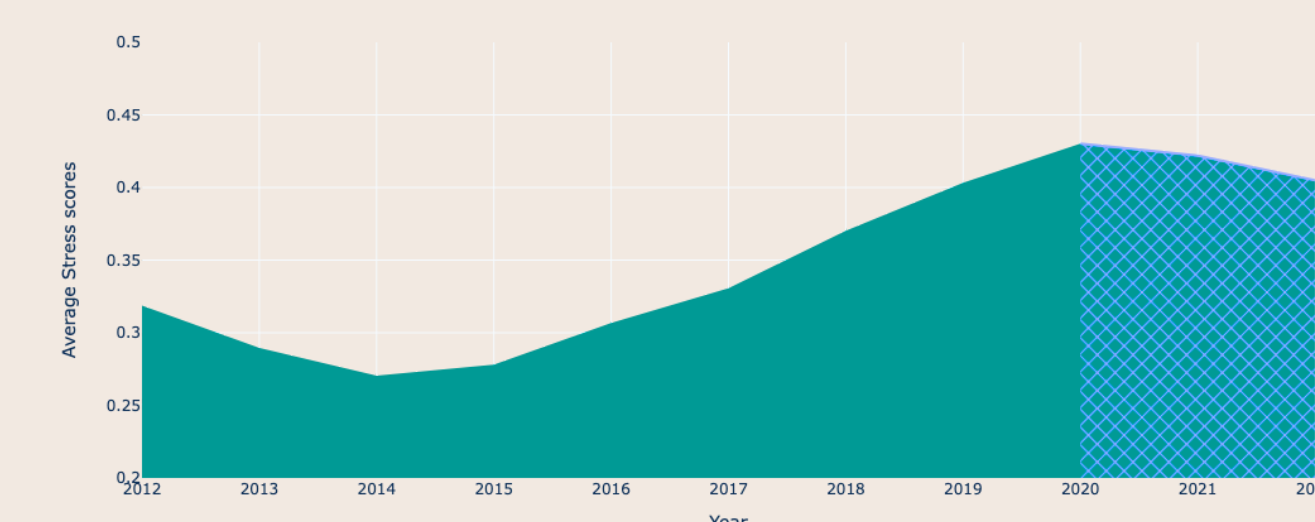- Recall: 0.9316
- F1-Score: 0.9120

RESULTS

From our analysis, we unearthed temporal trends over the academic year and through the pandemic. We were also able to find potential topics and words that are closely associated with high-stress posts.

Stress sentiment scores:

Average score = 0.39 — Deviation from average university stress score

Temporal analysis:

Analysis of stress scores of college reddit communities over the past 10 years

Word Cloud

Stress trends over an academic year

Topic modeling:

Some of the topics we found with the Latent Dirichlet Allocation algorithm:
- **Finance**: ['looking', 'housing', 'need', 'advice', 'roommate', 'aid', 'student', 'financial', 'freshman', 'fall']
- **Social**: ['depression', 'campus', 'people', 'night', 'today', 'tonight', 'food', 'around', 'party', 'day']
- **Study**: ['major', 'study', 'schedule', 'science', 'engineering', 'program', 'course', 'exam', 'research', 'grade']

CONCLUSION

We built a sentiment classifier using Reddit data to give stress scores to students' posts on University subreddits. We analyzed the university subreddit's stress levels over the years to find patterns and causes that might contribute to them. We found that students' stress levels peaked during the pandemic and dips during the breaks. We also found potential topics or words that are closely associated with the high-stress posts. Universities can understand these factors and take measures to support the students by raising awareness and providing support and resources to help.

REFERENCES

1. Asif S, Mudassar A, Shahzad TZ, Raouf M, Pervaiz T. Frequency of depression, anxiety and stress among university students. Pak J Med Sci. 2020 Jul-Aug;36(5):971-976. doi: 10.12669/pjms.36.5.1873. PMID: 32704273; PMCID: PMC7372668.
2. Bagroy, Shrey et al. "A Social Media Based Index of Mental Well-Being in College Campuses." Proceedings of the SIGCHI conference on human factors in computing systems. CHI Conference vol. 2017 (2017): 1634-1646. doi:10.1145/3025453.3025909.
3. Koustuv Saha and Munmun De Choudhury. 2017. Modeling Stress with Social Media Around Incidents of Gun Violence on College Campuses. Proc. ACM Hum.-Comput. Interact. 1, CSCW, Article 92 (November 2017), 27 pages. https://doi.org/10.1145/3134727