

Team 2

Depressive text classification.

Why we chose this project?

- Depression is “a common and serious medical illness that negatively affects how you feel, the way you think and how you act.”
- The symptoms of depression ranges from changes in appetite, loss of interest or pleasure in activities once enjoyed, thoughts of death or suicide, and etc.
- During the COVID-19 Pandemic there has been an elevated amount of adverse mental health conditions with depression being the lead condition.

How this project will help?

- A text classification model that can score how depressed a piece of text could be used for monitoring service.
- The project can help provide infographics that medical organizations can use to educate the public.
- Provide the ability to analyze any given text to determine the “rate of depression” for that particular text
- “Globally, more than 264 million people of all ages suffer from depression.”³ So, with researching this topic we’d help understand an issue that impacts a significant amount of people globally.

Subreddit Summations (Bryan and Ashish)

- Counted all the posts coming from r/depression and r/SuicideWatch
- Conclusion
 - Dataset contains equal amount of threads for each class (subreddit)
 - No subreddit holds a bigger influence in data

```
+-----+-----+
|      class| count|
+-----+-----+
| depression|304886|
|SuicideWatch|304886|
+-----+-----+
```

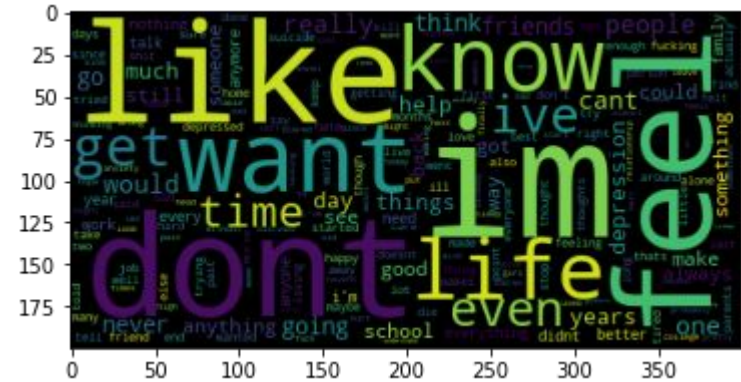
Post Length Statistics (Davith)

- Basic Stats done on the lengths of all posts.
- No cool graphics here :(
- Conclusion:
 - Thought we may be able to see something important here. Other than the weird outlier min value nothing to useful here.

```
(count: 609772, mean: 1176.795928642231, stdev: 1422.9257077693858, max: 40297.0, min: 2.0)
```

WordCount (Ashish and Davith)

- Analyzed word frequencies.
 - Pre Processing
 - Removed Punctuation
 - Lowercased
 - Removed Stop Words (the, I, am)
 - Created Word Cloud Visualization
- Conclusion
 - With the word frequencies we can use this information in accounts to our labeling process.



WordCount (Ashish and Davith)

- The following is a snippet of our results. ->

```
+-----+-----+
|      _1|      _2|
+-----+-----+
|      im|1202526|
|     dont| 833336|
|     like| 810509|
|     feel| 704298|
|     want| 627313|
|     know| 609032|
|     life| 547364|
|      get| 517916|
|      ive| 438814|
|     even| 432674|
|     time| 426501|
|   really| 416692|
|   people| 410211|
|     cant| 402223|
|    would| 378815|
|      one| 373920|
|    think| 319247|
|    going| 316714|
|    never| 314062|
|      go| 309134|
+-----+-----+
only showing top 20 rows
```

TF-IDF (Bryan)

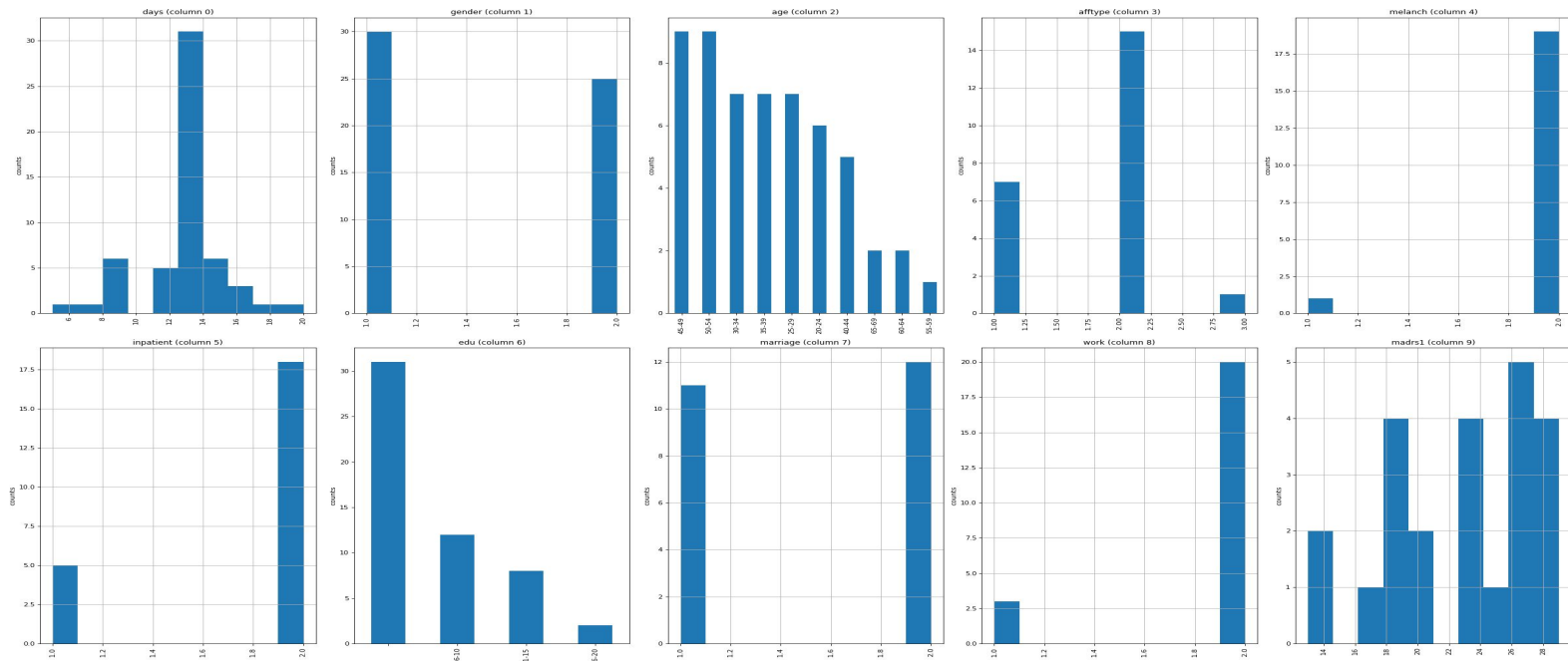
- Cleaning process
 - Removed punctuation
 - Lowercase all words using Tokenizer
 - Split the text using white spaces
- Analysis
 - Calculated the term frequency and hashed it using a hash function
 - Calculated the inverse document frequency
- Conclusion
 - With the tf and idf, we will be able to determine words that holds the least importance and disregard it.

```
+-----+-----+-----+-----+
|      class|      text|      words|      tf|      idf|
+-----+-----+-----+-----+
|SuicideWatch|Feeling a bit dep...|[feeling, a, bit,...|(65536,[1714,1903...|(65536,[1714,1903...|
|SuicideWatch|Was going to hang...|[was, going, to, ...|(65536,[181,600,1...|(65536,[181,600,1...|
| depression|Have you ever mai...|[have, you, ever,...|(65536,[353,1546,...|(65536,[353,1546,...|
| depression|I haven't felt po...|[i, haven't, felt...|(65536,[32,835,11...|(65536,[32,835,11...|
| depression|Partners of those...|[partners, of, th...|(65536,[12,351,83...|(65536,[12,351,83...|
+-----+-----+-----+-----+
only showing top 5 rows
```


Exploratory Analysis (Saurav)

- Used pandas library to visualize the relation between the patients and their contributing factors for depression that are given in the datasets. Such as Age, marital status, work, and so on.

Graphs (Saurav)



Bigrams (Ami)

- Pairs of words that were likely to come up together
- We filtered results that had “depression” to get only the words that were associated with “depression”
- They are sorted by likelihood ratio (descending)
- Order matters - for example “depression and anxiety” is different from “anxiety and depression”

```
((('depression', 'anxiety'), 131.70544652948615)
(('anxiety', 'depression'), 72.22559030989585)
(('severe', 'depression'), 45.797027239998144)
(('suffering', 'depression'), 38.95904130927086)
(('major', 'depression'), 30.13600332875144)
(('suffer', 'depression'), 27.362614897322388)
(('struggling', 'depression'), 26.779261833819106)
(('chronic', 'depression'), 22.07980264585341)
(('diagnosed', 'depression'), 21.879939849418278)
(('noticed', 'depression'), 21.32728644706868)
(('dealing', 'depression'), 20.763148455855518)
(('ptsd', 'depression'), 20.66112977381583)
(('depression', 'suicidal'), 18.334598430478266)
(('depression', 'excuse'), 18.147180861872506)
(('struggled', 'depression'), 17.75320355097563)
(('history', 'depression'), 17.038040829154518)
(('depression', 'comes'), 12.459294837020794)
(('depression', 'meds'), 11.763935979998784)
(('understand', 'depression'), 11.436903905808226)
(('due', 'depression'), 10.702910067258937)
(('depression', 'since'), 9.537978108930261)
(('back', 'depression'), 6.552105196987837)
(('living', 'depression'), 5.9822099935789845)
(('depression', 'doesnt'), 4.6062403734209925)
(('depression', 'started'), 4.440231845630621)
```

Trigrams (Ami)

- Groups of 3 words next to each other in a sentence that were likely to show up together
- Again, results were filtered for “depression” and were sorted by likelihood ratio
- These can show words that are related to depression

```
((('feel', 'like', 'depression'), 9528.119876643184)
((('depression', 'feel', 'like'), 9524.726073164264)
((('dont', 'know', 'depression'), 7189.405642579027)
((('depression', 'dont', 'want'), 3945.9455799729335)
((('depression', 'suicidal', 'thoughts'), 1676.546969059987)
((('depression', 'feels', 'like'), 1621.1240780335927)
((('depression', 'long', 'time'), 1160.0917769141388)
((('depression', 'dont', 'think'), 1017.8117751787623)
((('social', 'anxiety', 'depression'), 923.6868735689229)
((('severe', 'depression', 'anxiety'), 739.8683300794883)
((('depression', 'social', 'anxiety'), 716.9740414879772)
((('diagnosed', 'depression', 'anxiety'), 699.5068469579847)
((('struggling', 'depression', 'anxiety'), 678.034050016812)
((('suffer', 'depression', 'anxiety'), 613.9660401327799)
((('due', 'depression', 'anxiety'), 557.0603968390519)
((('depression', 'anxiety', 'ptsd'), 552.3716514790202)
((('depression', 'anxiety', 'long'), 530.5193312587514)
((('pretty', 'sure', 'depression'), 467.60472509759904)
((('depression', 'bipolar', 'disorder'), 410.7831189223852)
((('worst', 'part', 'depression'), 406.3844335463595)
((('ive', 'struggling', 'depression'), 316.13682196086165)
((('diagnosed', 'clinical', 'depression'), 299.6137399335885)
((('depression', 'coming', 'back'), 268.52637204436166)
((('depression', 'long', 'remember'), 262.75191405049486)
((('struggling', 'depression', 'long'), 251.22579919545637)
((('depression', 'keeps', 'getting'), 235.9716322294812)
((('ive', 'suffered', 'depression'), 176.75338403513857)
((('severe', 'case', 'depression'), 118.57712762235414)
```

Other Filters for Trigrams

```
((('makes', 'feel', 'like'), 10293.597071115762))  
((('feel', 'like', 'im'), 10240.551236546688))  
((('make', 'feel', 'like'), 9838.977873964726))  
((('feel', 'like', 'shit'), 9801.689318345865))  
((('dont', 'feel', 'like'), 9787.57064497894))  
((('made', 'feel', 'like'), 9692.460861874346))  
((('still', 'feel', 'like'), 9640.49805014465))  
((('im', 'feel', 'like'), 9635.9875458193))  
((('feel', 'like', 'cant'), 9625.45291756278))  
((('sometimes', 'feel', 'like'), 9622.101446789893))  
((('feel', 'like', 'failure'), 9619.632002582111))  
((('honestly', 'feel', 'like'), 9606.239122530087))  
((('\n\ni', 'feel', 'like'), 9604.616125955748))  
((('making', 'feel', 'like'), 9600.948020161712))  
((('want', 'feel', 'like'), 9596.989974788255))  
((('else', 'feel', 'like'), 9595.660690160705))  
((('feel', 'like', 'burden'), 9590.524992547456))  
((('feel', 'like', 'theres'), 9589.093929021748))  
((('feel', 'like', 'ive'), 9588.20194676046))  
((('feel', 'like', 'complete'), 9584.568264908363))  
((('feel', 'like', 'crap'), 9576.534484925647))  
((('\ni', 'feel', 'like'), 9569.644244404575))  
((('get', 'feel', 'like'), 9567.684795254041))  
((('feel', 'like', 'idiot'), 9565.871007761132))  
((('doesnt', 'feel', 'like'), 9561.410372910446))  
((('always', 'feel', 'like'), 9560.297565405437))  
((('feel', 'like', 'nothing'), 9559.70138887553))
```

```
((('feel', 'like', 'lost'), 9528.730032416675))  
((('lost', 'feel', 'like'), 9527.328643443732))  
((('lost', 'best', 'friend'), 1712.8097130707906))  
((('like', 'ive', 'lost'), 290.3298650503469))  
((('ive', 'lost', 'everything'), 246.46665777638367))  
((('ive', 'lost', 'friends'), 245.98839845332736))  
((('ive', 'already', 'lost'), 165.78922948724522))  
((('im', 'completely', 'lost'), 129.51476728054848))  
((('lost', 'almost', 'friends'), 61.18199632962903))
```



Thanks

Questions?

