Depression Detection On Social Media With NLP

University: Duke Kunshan University
Course: Stats 402
Instructor: Sun Peng

Group 3: Lin He (lh317), Jingcheng Wu (gw93)

Date: Dec. 15, 2022

Agenda

1. Background
2. Objectives and Database
3. Methodology
4. Results and Evaluation
5. Conclusion

Background

Background: Depression

- **Symptoms of depression**

  A mental disorder that could disturb the psychosocial functioning and severely diminish the quality of daily life; symptoms include loss of sleep and appetite, restlessness and disinterest in daily activity and job, etc. Estimated population who suffer from depression **takes up 5% globally**.

- **Significance of depression**

  The sixth cause of disability and years of productive life lost,the top contributor to global disability, highly prevalent (**over 5% population**)

- **A problem that affects early intervention**

  A web-based survey by Yoshikawa et al., around **55%** people reported unwilling to seek help for depression

Detection Based on Social Media

- **Data available on social media:**
  - Content that show subtle cues of their symptoms
  - Activities on social media (number of friends, number of interaction, when do they post on social media, etc.)
- **The development of using social media as detection source:**
  - Social media is becoming a rich data source for AI techniques' application in depression diagnosis and detection.
  - The user base of social media is increasingly wide, which allows tremendous real-world impact for researches.

Objectives & Database

Main goal: Test how effective it is to <u>detect depressed users</u> based on their <u>social accounts activities</u>.

Objectives:

1. Test if we can use **NLP** method to classify depressed users and normal users solely based on the written content they post.
2. Check if the **NLP** method captures any language pattern of depressed patients in general.
3. Test if adding more **social account activities features** will improve the detection accuracy.

Database

- Curated by Shen et al.
- Separated JSON format files
- Data set has binary labels on users' ids
  - User diagnosed with depression
  - Normal user
- Each user account has over 100 posts
- Each user account contains several profile features
  - Friend count, follower count, following count...
  - Profile photo color
  - ...

- G. Shen, J. Jia, L. Nie, F. Feng, C. Zhang, T. Hu, T.-S. Chua, and W. Zhu, "Depression detection via harvesting social media: A multimodal dictionary learning solution," Proceedings of the Twenty-Sixth International Joint Conference on Artificial Intelligence , 08 2017.

Methodology

1. NLP + Classification

- **Encode the posts**
  - For each user, randomly selected the user's posts and combined the posts. Use the user's label as the merged post's label.
  - Two encode methods:
    - DistilBERT-based pretrained uncase model
    - Linguistic Inquiry and Word Count (LIWC)
- **Classify the posts**
  - Apply Logistic Regression, SVM, and Multi Layer Perceptron (MLP) on the encoded posts.

2. Analyze the BERT encoding

Test BERT on single post

- Among the False Negative group, break down the merged posts into single texts and run Bert + LR model on the single texts

LIWC analysis

- **Preprocessing**
  - Separate the merged posts of users that are classified by Bert + Logistic Regression into four categories: True Positive, True Negative, False Positive, False Negative
- **LIWC**
  - Run the merged texts on LIWC
- **Group analysis**
  - Analyze the mean, std, higher quartile and lower quartile of the four groups

3. User Feature Extraction

- **Profile Background Color**

  HEX-encoded color -> HSL-encoded color-> Extract multiplication of Saturation and L -> A feature that measures the “colorfulness”

- **User Active Time**

  Divided a day into morning, afternoon, evening and late night, then calculate the portions of the posts sent in four periods and assign scores to the periods.

  Higher score = on average more active in later times of a day

- **Antidepressant Count**

  Count how many times did the users mention the name of some popular antidepressants in their tweets

Results + Evaluation

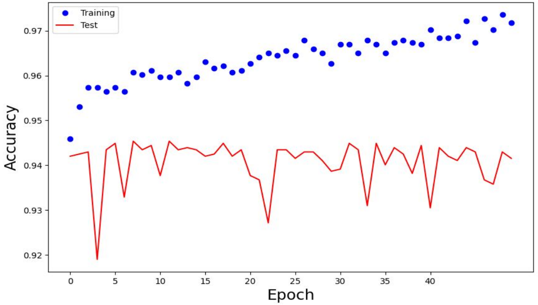
BERT + Classification

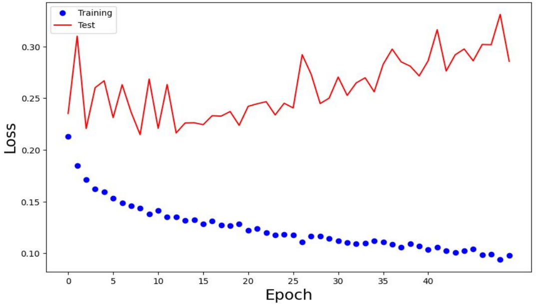
- The Bert + classification models perform better on balanced data
- Bert + LR performs better than Bert + SVM

	BERT + LR	BERT + SVM	Dummy Classifier
Text, imbalanced (847:153)	0.824	0.820	0.856
Text, balanced (500:500)	0.704	0.692	0.505
User, imbalanced (949:51)	0.960	0.960	0.945
User, balanced (200:200)	0.770	0.620	0.517

LIWC + MLP

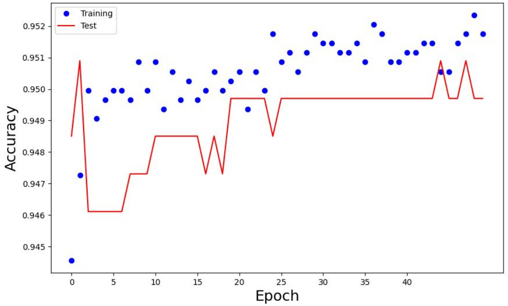
- LIWC gives 100+ features, we manually selected 33 related ones
- Run features and labels through Neural Network
- Use Scikit-learn feature selection to pick top 6 features, run through NN

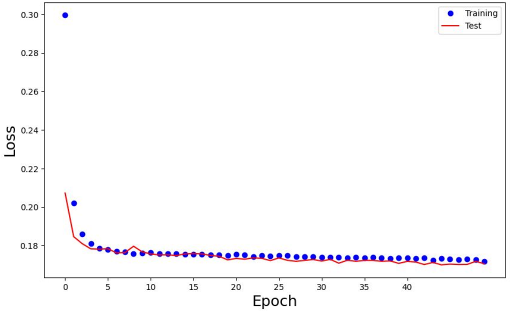




Accuracy & loss, 33 features, NN, imbalance dataset

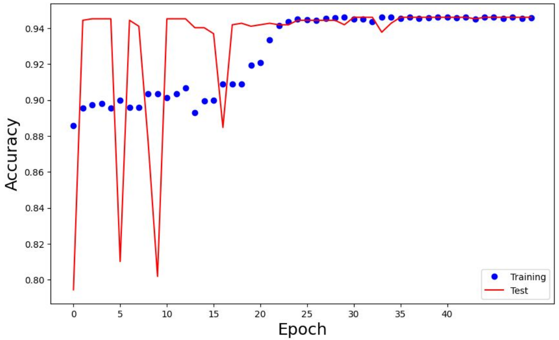
Accuracy & loss, 6 features, NN, imbalance dataset

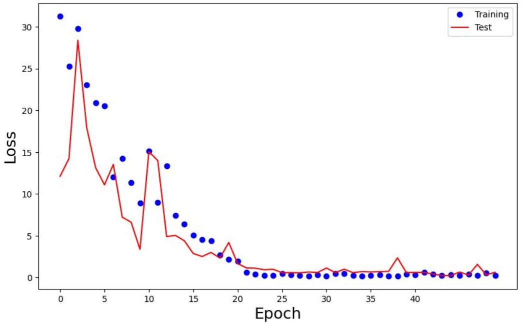




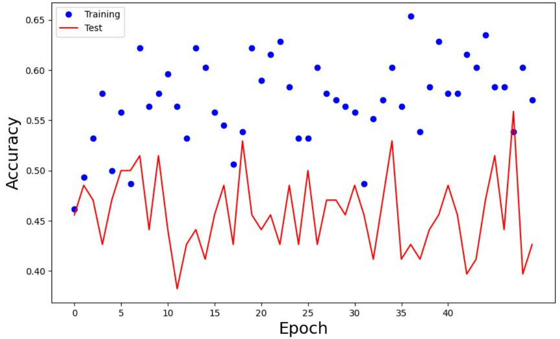
LIWC + Extended User Features + MLP

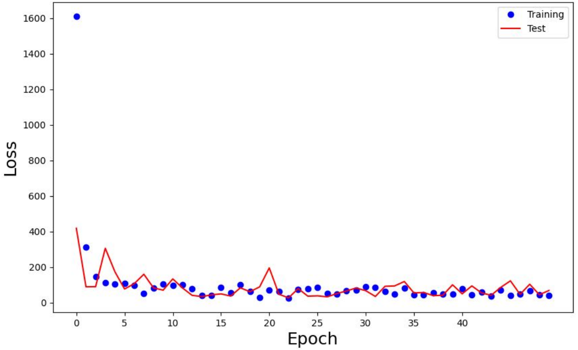
- 33 LIWC Features + 7 User Features, including time, color, user metrics
- Run features and labels through Neural Network
- Use Scikit-learn feature selection to pick top 6 features, run through NN





Accuracy & loss, 41 features, NN, imbalance data set

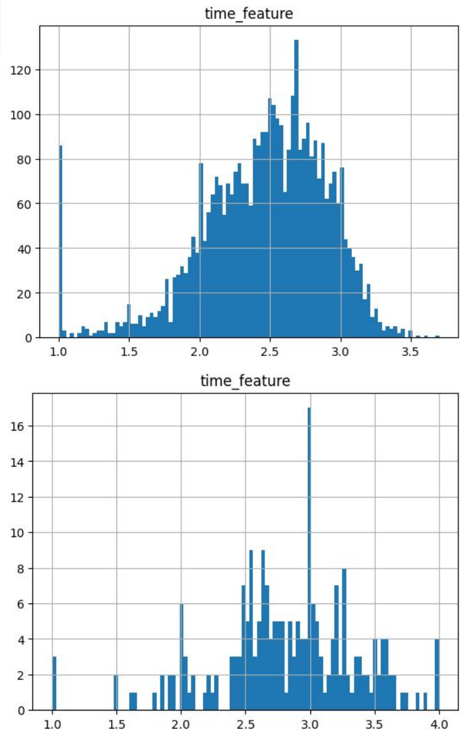




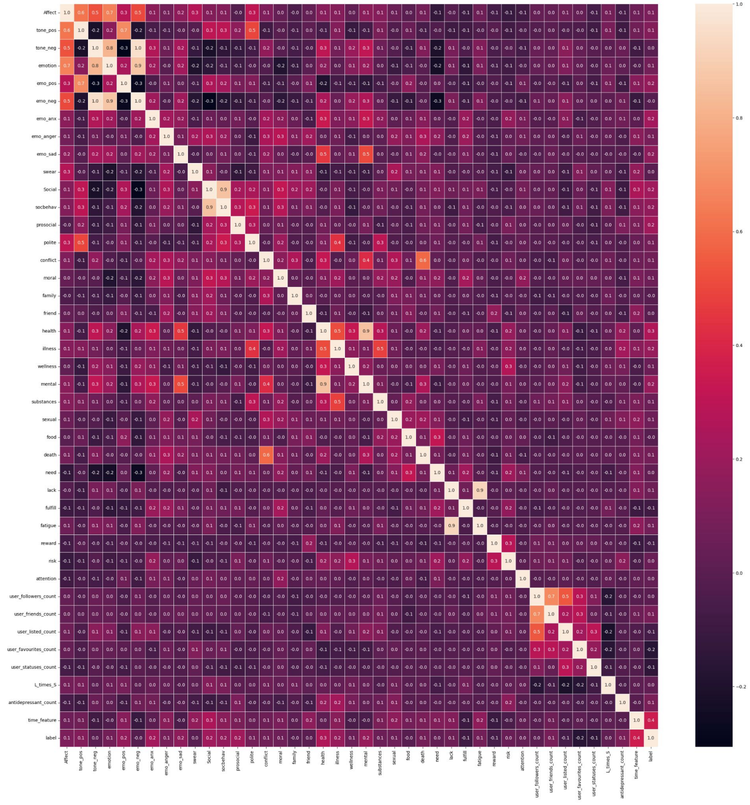
Accuracy & loss, 41 features, NN, Balance Dataset

LIWC + Extended User Features Analysis

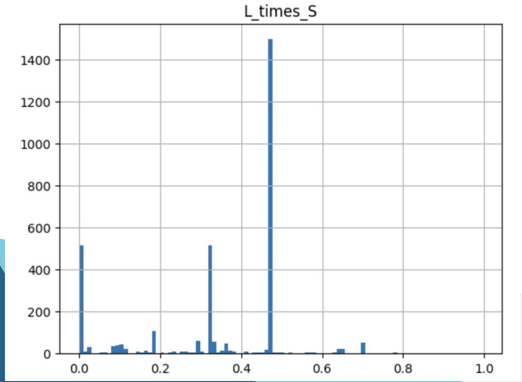
- Time feature has correlation of 0.4 with label
- Profile color feature is interfered by default choices



Distribution of time feature score of all users and depressed users

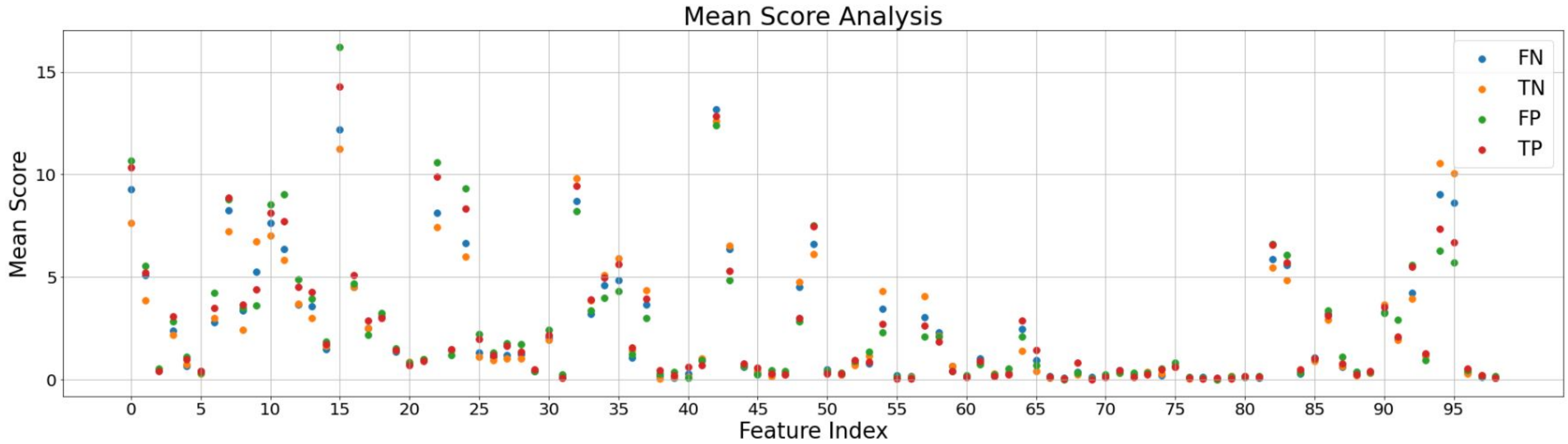


Correlation Heatmap of 33 LIWC features and 7 extended user features



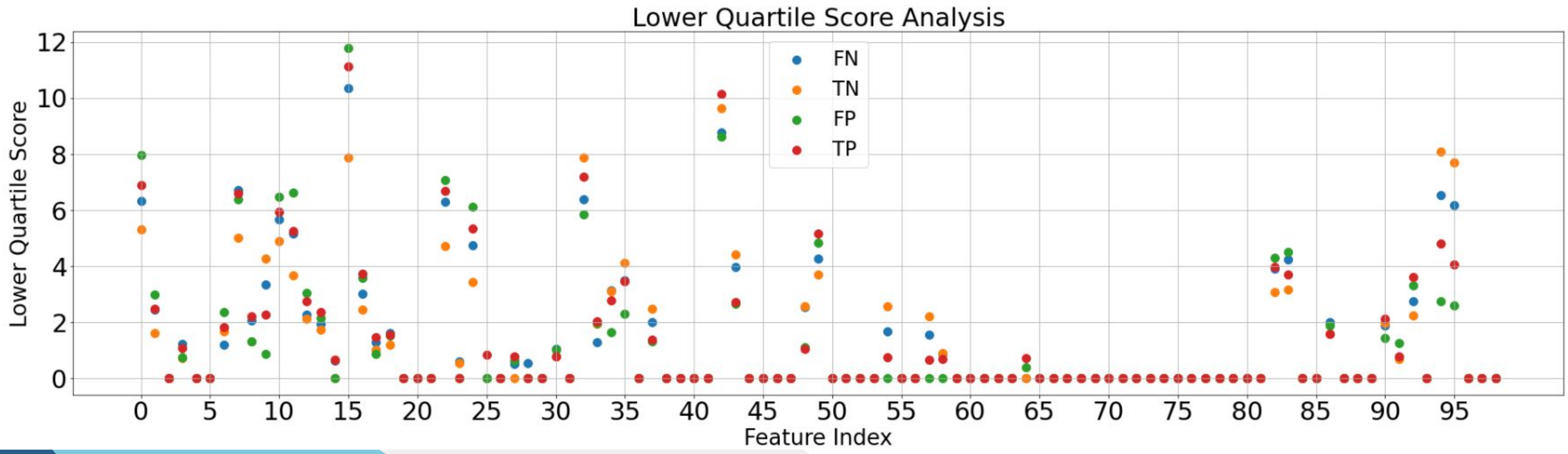
BERT Encoding Analysis (Mean Score)

- Among the 99 features, there are only 8 features in which the four classification groups have evidently different mean values.
- 8 features are personal pronoun use, verb use, cognition, cognition process, conversation, and internet slang use



BERT Encoding Analysis (Lower Quartile)

- ○ Many feature scores are 0 within the True Positive group.
- ○ It might a piece of evidence that BERT-logistic model does not make classification decisions based on the psychological nuances in language, or only a small fraction of features are useful.



BERT + LR test false negative group

- No apparent connections between the mislabeled separate text and the merged text

Total number of depressed users	200
Total number of posts by depressed users	2000
Mislabeled depressed users	83
Number of mislabeled users who have 0-3 posts mislabeled	36
Number of mislabeled users who have 4-6 posts mislabeled	23
Number of mislabeled users who have 7-10 posts mislabeled	24

Conclusion

1. BERT model is not capturing the language feature of the texts. Possible reasons:
   a. with **diluted** data: only ~5% of all users are depressed, and not all their tweets contain depression related expressions
   b. BERT does not capture many features that LIWC can pick up
   c. Probably the depressed users do not have a common, distinct language pattern from normal users

2. User post time could be derived as a useful feature
3. Future work:
   a. Use user features to screen out certain tweets, then feed into BERT for further analysis?
   b. Compare the texts that are labeled positive by BERT+LR model to the symptoms of depression and check for similarity

Thanks