Utilizing Machine Learning in Early Detection of Mental Health Disorders

Aryan Khanna, SCAI
Advisor: Yiran Luo, Assistant Teaching Professor
Arizona State University

Motivation

Traditional diagnostic methods for mental health disorders face hurdles in access--ibility and subjective bias, which often cause delays in timely treatment.

Task Setting

Explore how different **Natural Language Processing (NLP)** methods in Machine Learning can improve early detection on the open source de-identified **Mental Health Dataset (MHD)^**.

References

^ Jikadara, B, (2024). *Mental Health Dataset*. Kaggle. https://www.kaggle.com/datasets/bhavikjikadara/mental-health-dataset

Babu, M. S., & Kanaga, E. (2022). *Supervised machine learning models for depression sentiment analysis.* Frontiers in Psychology.

Eichstaedt, J. C., et al. (2018). *Facebook language predicts depression in medical records.* PNAS.

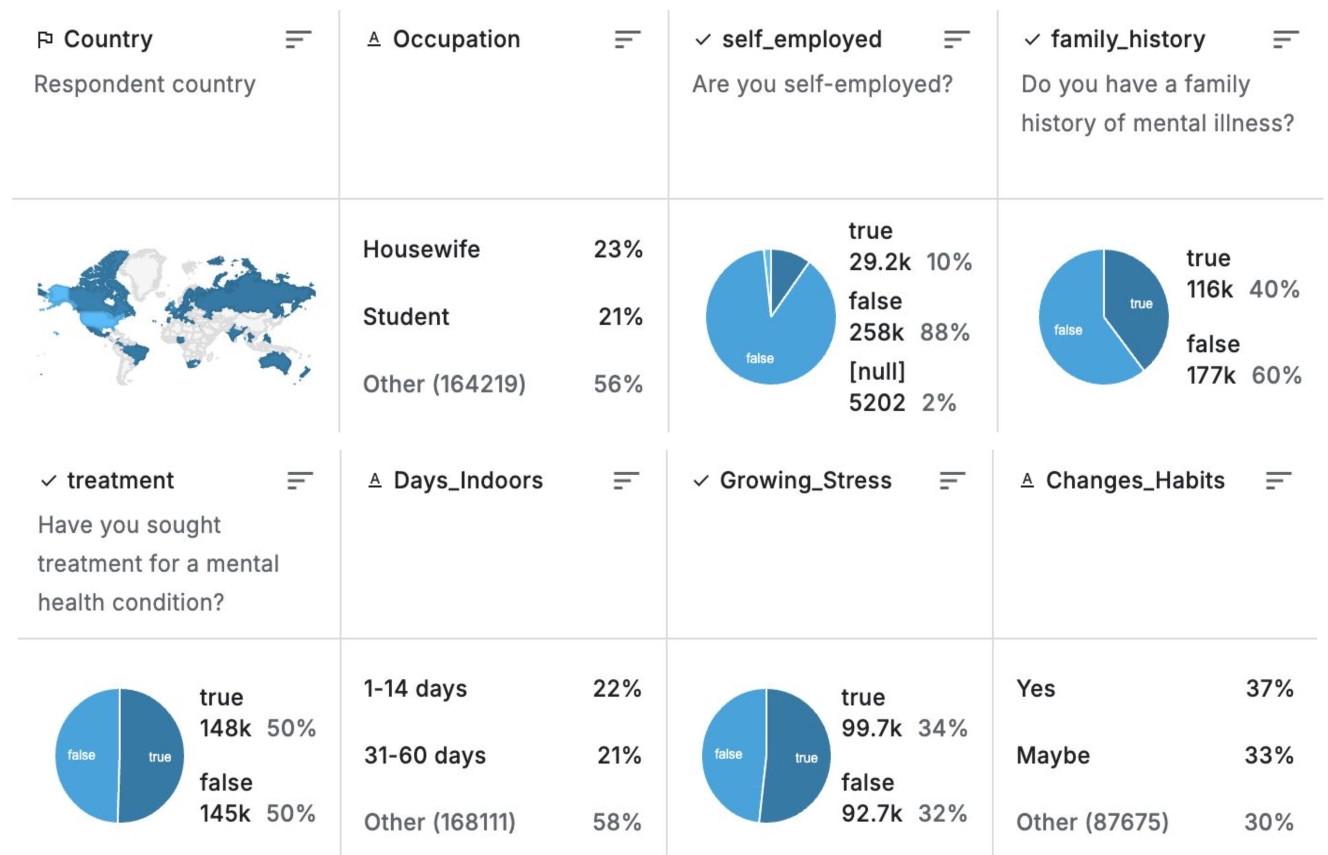
Nature Medicine (2020). *Machine learning model to predict mental health crises from electronic health records.*

Nemesure, M.D., Heinz, M.V., Huang, R. et al. *Predictive modeling of depression and anxiety using electronic health records and a novel machine learning approach with artificial intelligence.* Sci Rep 11, 1980 (2021).

Experiments and Results

Data Analyses

MHD includes 3 mood classes (low, mid, high) and contains various patient attributes/features such as occupation, habit, gender, and etc. We need to figure out which attributes are actually impactful for classification.



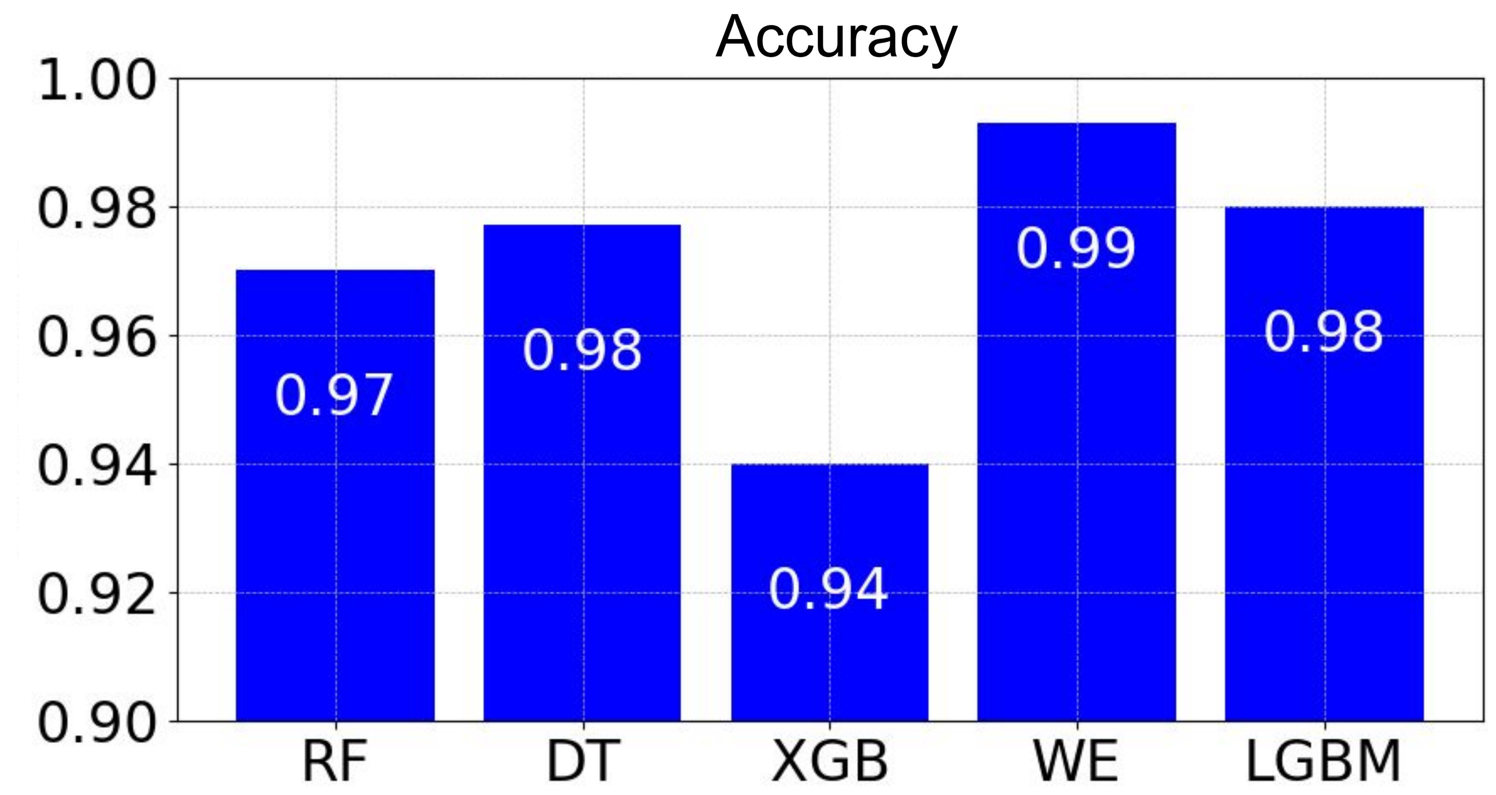
Distributions of example attributes/features in **MHD**

Weight	Feature	Weight	Feature
0.5391 ± 0.0073	occupation	0 ± 0.0000	care_options
0.4130 ± 0.0039	changes_habits	0 ± 0.0000	mental_health_interview
0.3879 ± 0.0030	days_indoors	0 ± 0.0000	treatment
0.3073 ± 0.0022	work_interest	0 ± 0.0000	family_history
0.3033 ± 0.0028	growing_stress	0 ± 0.0000	self_employed
0.2807 ± 0.0028	social_weakness		
0.2549 ± 0.0019	mental_health_history		
0.1728 ± 0.0010	coping_struggles		
0.1087 ± 0.0009	gender		

Using the regressive **XGBoost as a baseline,** we find the attributes of **Occupation** and **Days_indoors** have the most significant impact over mood swings, while **treatment** and **family_history** have the minimal.

Mood Classification w/ Machine Learning Models

We evaluate 5 models on **MHD** - Random Forest **(RF)**, Decision Tree **(DT)**, XGBoost **(XGB)**, AWS SageMaker WeightedEmsemble-L3 **(WE)**, and LightGBM-XT **(LGBM)** - over mood classification.



All 5 models achieve 94%+ overall prediction accuracy. However, we also observe several attributes (> 25% of all misclassified cases) that all models commonly intend to misclassify:

- **Occupation:** Housewife, Student
- **Day_Indoors:** 31-60 days, 15-30 days.

Moreover, the biases in **MHD** cause cases with **gender=Male** and **self_employed=true** more likely to be misclassified as well.

Future Directions

Experiments on Fairer Data: Use debiased and diverse mental health data and anonymized textual therapy transcripts for better model generalization.

Multi-Modal Analyses: Combine text with voice, facial expressions, and physiological data for comprehensive indicators.

FURI

Ira A. Fulton Schools of Engineering
Arizona State University