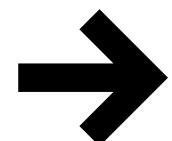
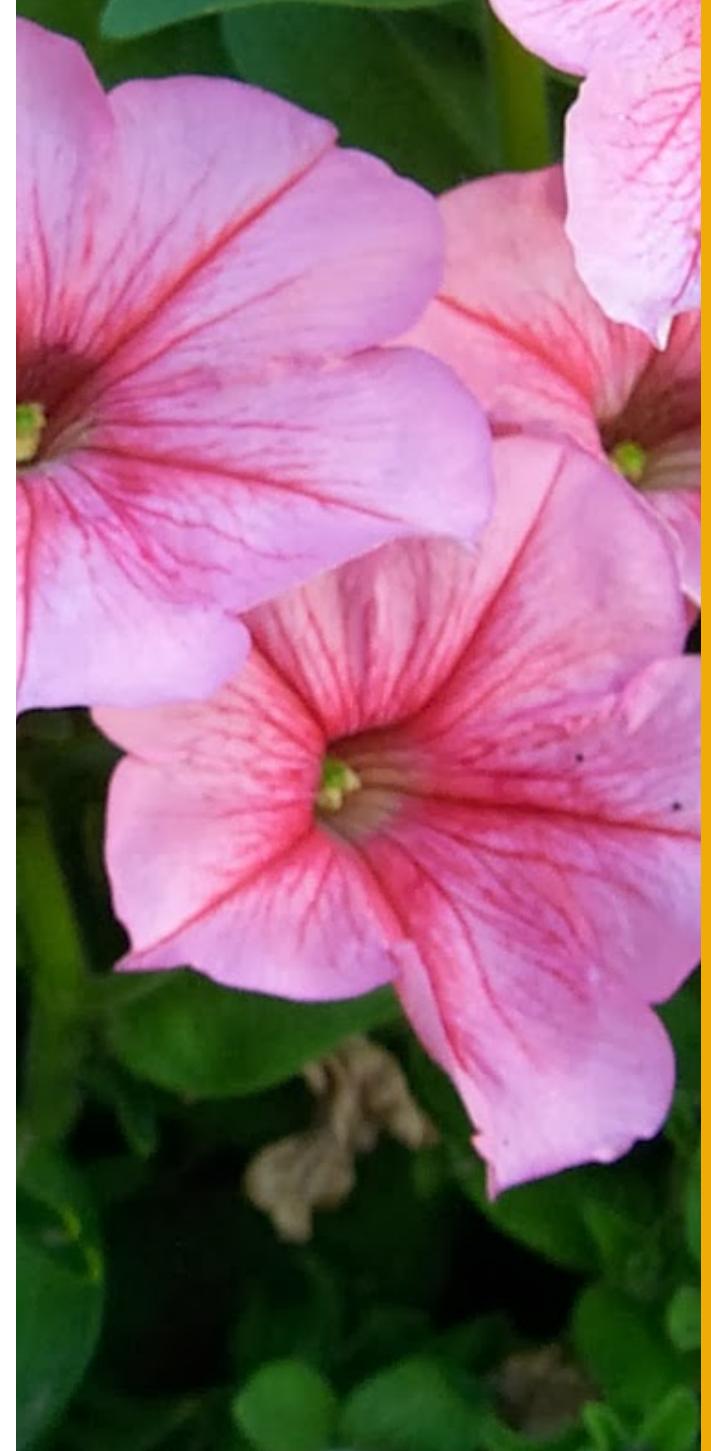


Predicting Anger Proneness Using Deep Learning Techniques

Adhinegoro, A.Y.; Tionando, V.C.; Adiputra, I.B.G.P.M.



Introduction



- Anger proneness is the likelihood of experiencing anger in response to certain stimuli or situations.
- Anger proneness varies widely among individuals.
- A person with higher anger proneness tends to face more problems, either in relationship, health, or legal troubles.

[7]

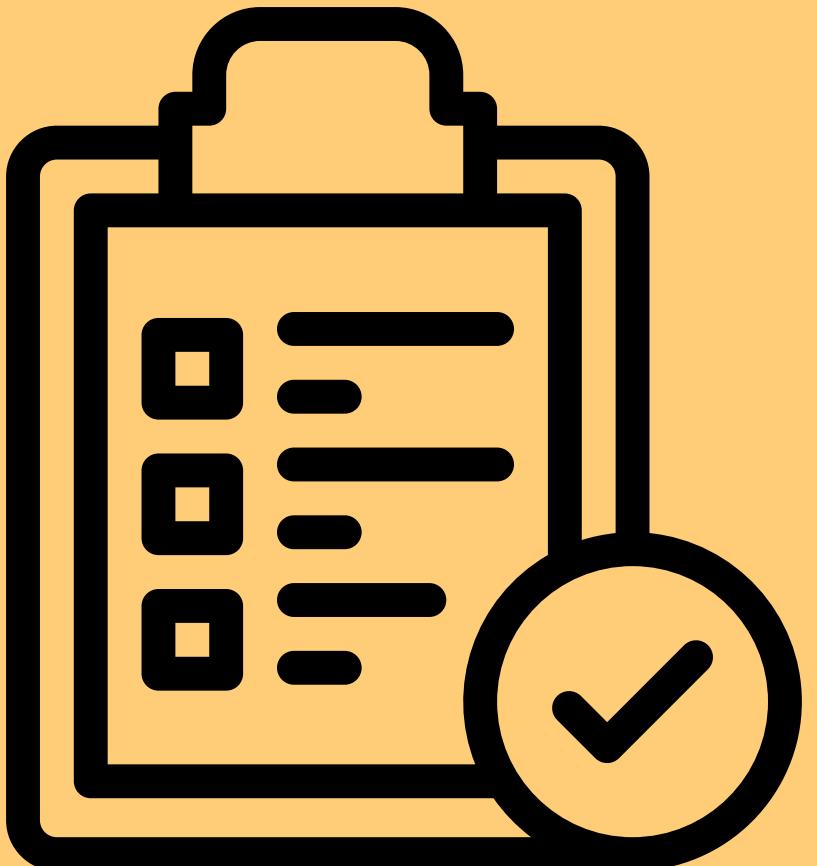
Problem Statement



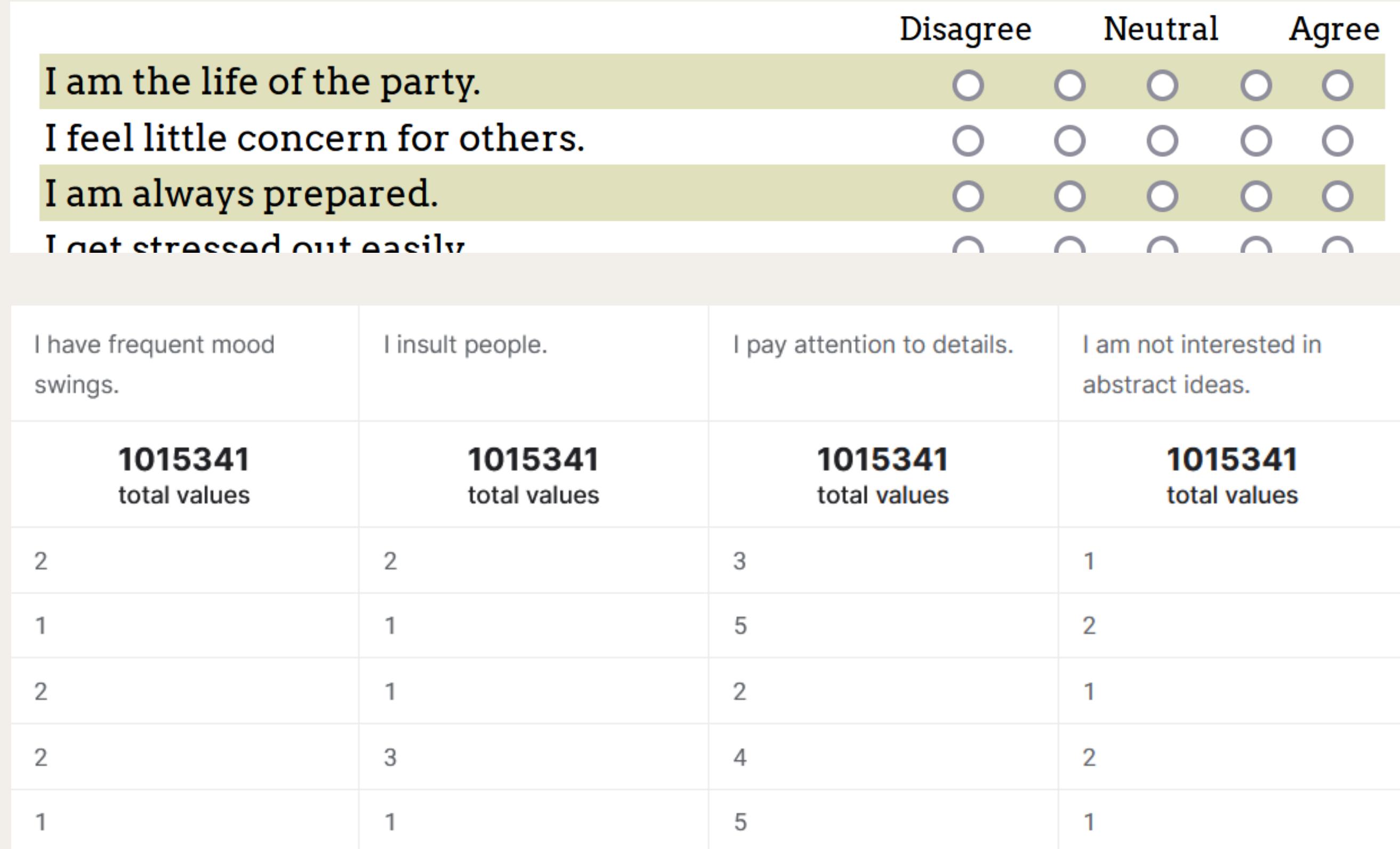
- The level of anger proneness of every individual is unpredictable and difficult to determine accurately in a short time.
- Determining the level of anger proneness through a personality test can yield inaccurate results.



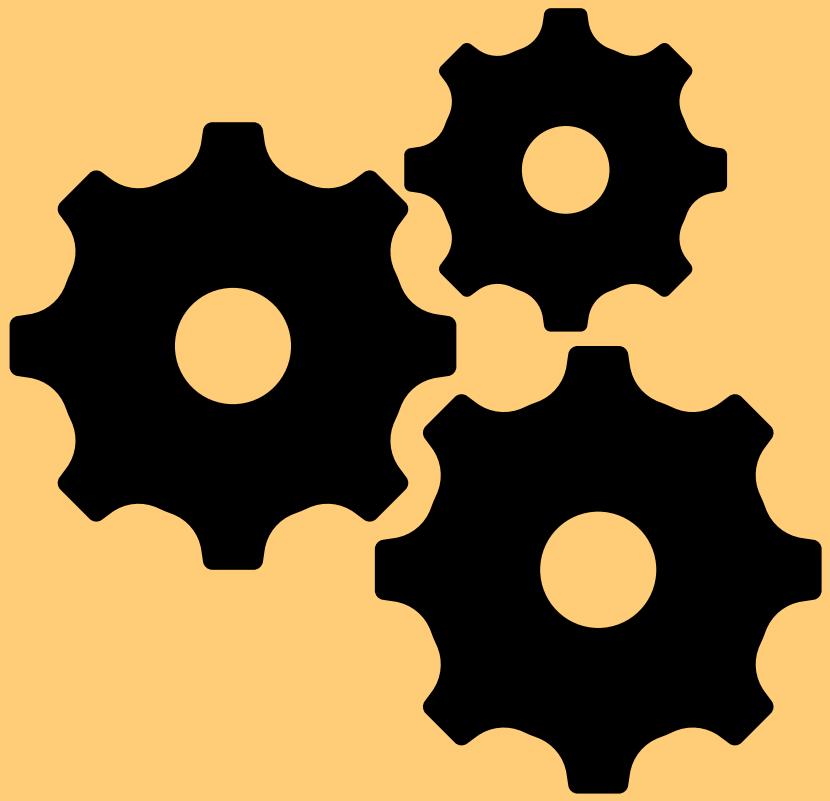
Introduction: Dataset



The dataset consists responses to 50 questions of a personality questionnaire done by 1,015,342 online participants through openpsychometrics.org website.



Introduction: Dataset

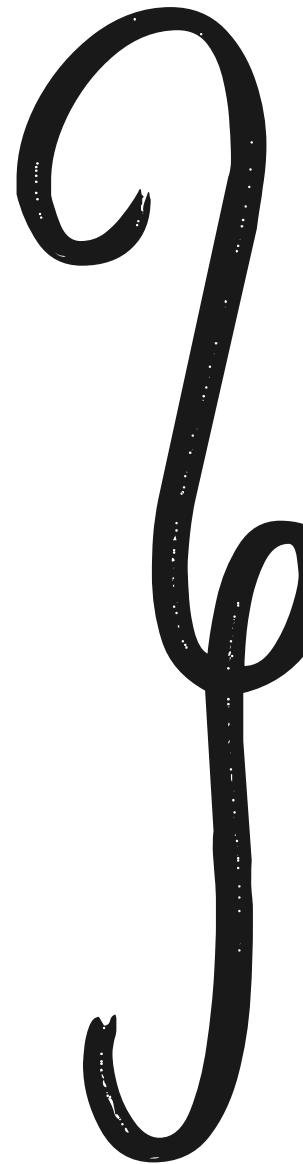


Features

I am relaxed most of the time.
I worry about things.
I seldom feel blue.
I am easily disturbed.
I get upset easily.

...

I change my mood a lot.
I am interested in people.
I insult people.
I sympathize with others' feelings.



Target

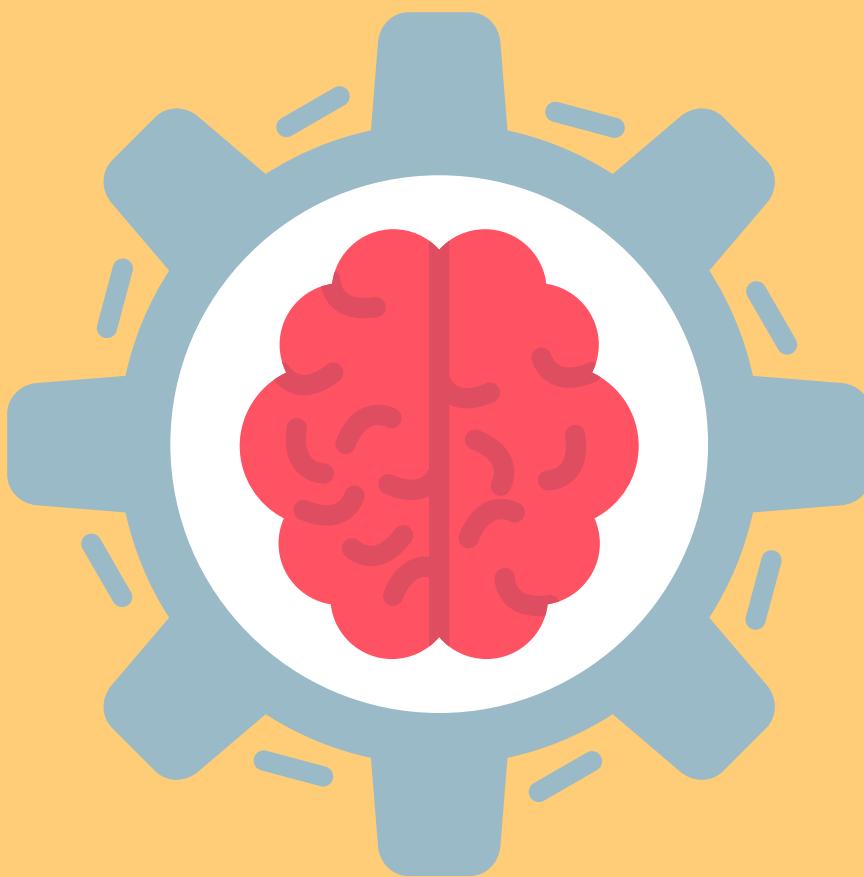
**I get
irritated
easily.**

disagree/slightly disagree/neutral/
slightly agree/agree

Literature Review

No	Research	Title	Result
[1]	Borisovi et al. (2022)	Deep Neural Networks and Tabular Data: A Survey	SAINT closely matches XGBoost in multiclassification task, with over 96% in both Accuracy and AUC.
[2]	Somepalli et al. (2021)	SAINT: Improved Neural Networks for Tabular Data via Row Attention and Contrastive Pre-Training	SAINT algorithms outperform various Machine Learning and Deep Learning classifiers in binary and multiclass classification tasks.
[3]	Khan et al. (2020)	Personality classification from online text using Machine Learning Approach	XGBoost without random oversampling : 86% Accuracy XGBoost with random oversampling : 99% Accuracy
[4]	Demirkaya et al. (2020)	Exploring the Role of Loss Functions in Multiclass Classification	Results in cross-entropy loss function outperforming quadratic loss.
[5]	Roma et al. (2018)	Could Time Detect a Faking-Good Attitude? A Study With the MMPI-2-RF	Suggests response time is a useful variable in detection of faking.

Methodology



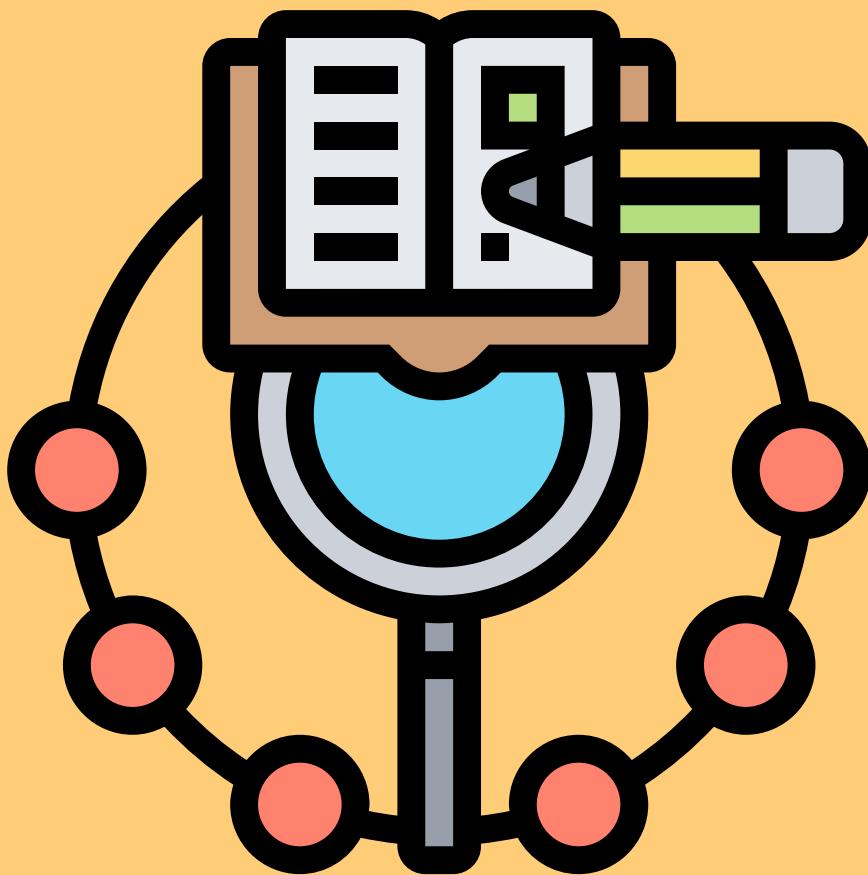
Data Preprocessing :

- Resampling with Random Oversampling
- Resampling with Random Undersampling
- Feature Selection using XGBoost

Model (for comparison):

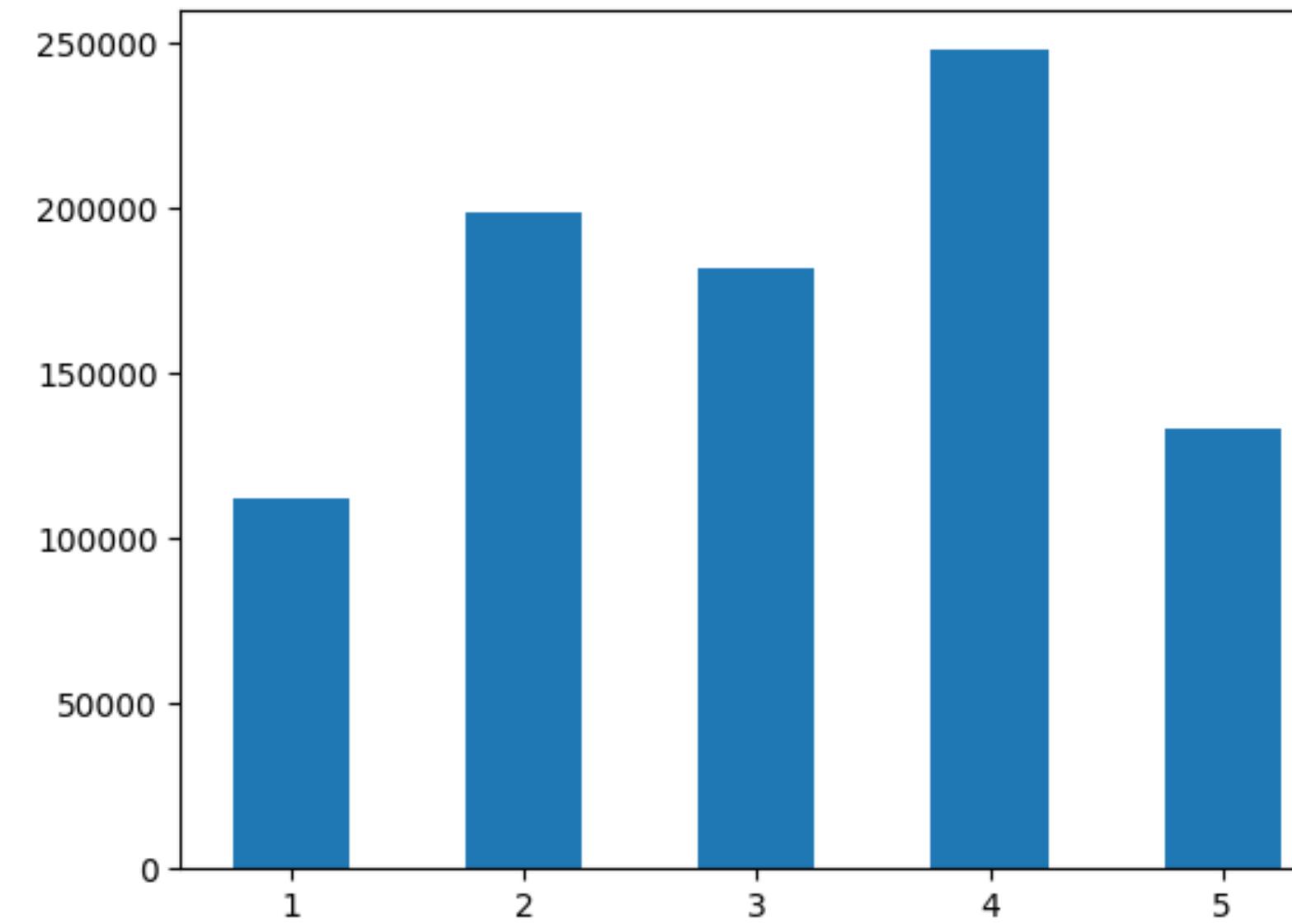
1. Self-Attention and Intersample Attention Transformer (SAINT)
2. Extreme Gradient Boosting (XGBoost)
3. Long Short Term Memory (LSTM)
4. Gated Recurrent Unit (GRU)

Methodology



Why we resampled the data

The responses to the "I get irritated easily" are not spread evenly



Response "4" occurred almost two times more than "1" and "5".

Methodology

Random Oversampling

- Duplicate the samples until every class has the same occurrences as the most represented class.
- Applied on XGBoost, LSTM, and GRU

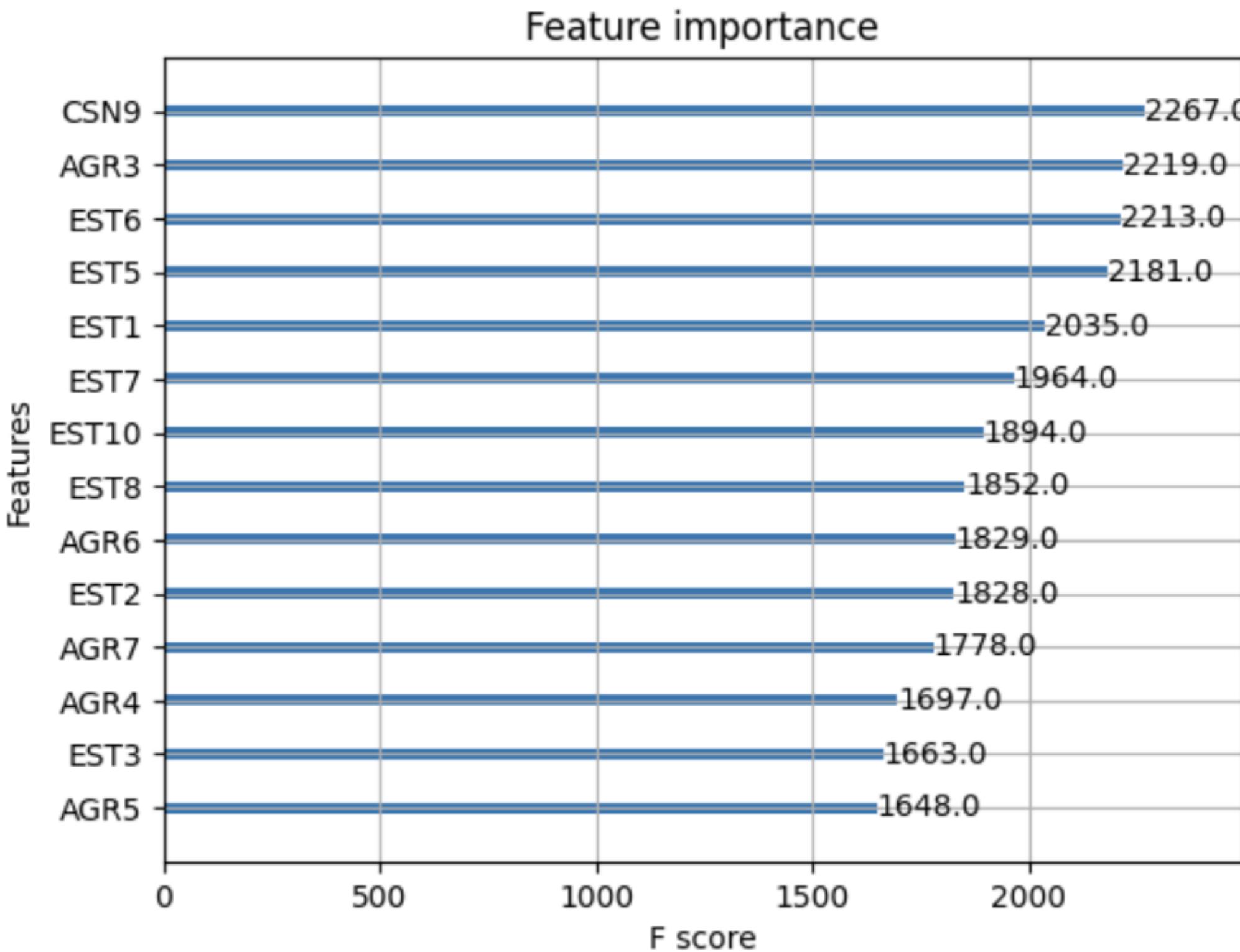
Random Undersampling

- Delete the samples until every class has the same occurrences as the least represented class.
- Applied on SAINT



Methodology

XGBoost Feature Selection



14 of 49 features with the highest feature importance.



Result & Analysis



Model	Accuracy [%]	Precision [%]	Recall [%]	F1-score [%]
XGBoost	52.53	52.57	52.53	52.13
XGBoost with Resampling	55.45	55.00	55.45	55.17
XGBoost with Feature Selection	52.34	52.39	52.40	51.95
LSTM	51.29	53.38	51.16	51.54
LSTM with Resampling	52.93	51.80	52.92	51.97
LSTM with Feature Selection	51.16	52.68	50.98	50.88
GRU	51.31	54.34	50.17	50.76
GRU with Resampling	53.17	52.8	53.15	52.75
GRU with Feature Selection	51.22	52.13	51.46	51.49
GRU with Resampling and Feature Selection	52.97	52.64	52.93	52.70
SAINT	47.08	47.43	47.08	46.56
SAINT with Resampling	52.79	52.43	52.79	52.48

Conclusion



- All four of the algorithms are shown unable to produce any model that can reliably predict anger proneness.
- Despite the efforts done in preprocessing the data, the results didn't seem to be noticeably affected.
- It leads the study to the conclusion that the used dataset has too many inconsistencies in the data for it to be inferred and predicted.
- In other words, the results of this study suggest the prominent inconsistencies in people's answers to the particular personality test consisted in the dataset.

Thank You