Supervised Machine Learning of Workplace Mental Health Acceptance

27th Sep 2023

Project by: Haruka Kozen

Description of Dataset

Rows (Employees): 1434, Columns (Questions) : 63

Target Value	Do you think that team members / co-workers would view you more negatively if they knew you suffered from a mental health issue?
Work environment related Questions	Health benefit, Wellness Campaign, Remote Work, Protection of Anonymity
Personal Questions	Age, Gender, Country, Family History of Mental Health, Diagnoses
Asking Feeling / Thoughts Questions	Comfortableness discussing Mental Health with co-workers/boss, Easiness of asking for Medical Leave

Question

Can Machine predict human thoughts ?

Can Machine guess if employees think their co-workers view them negatively ?
(twist: by only asking work-environment related questions)

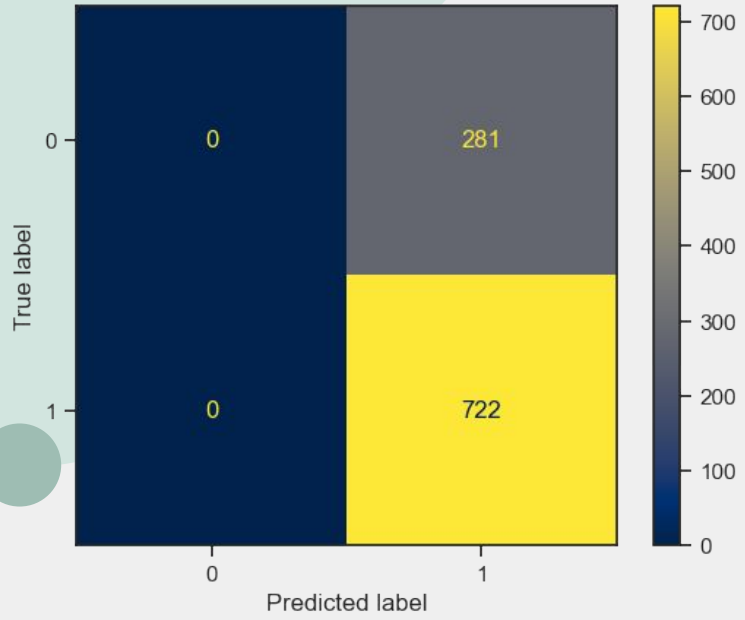
3 SVC Models performance for Training Data

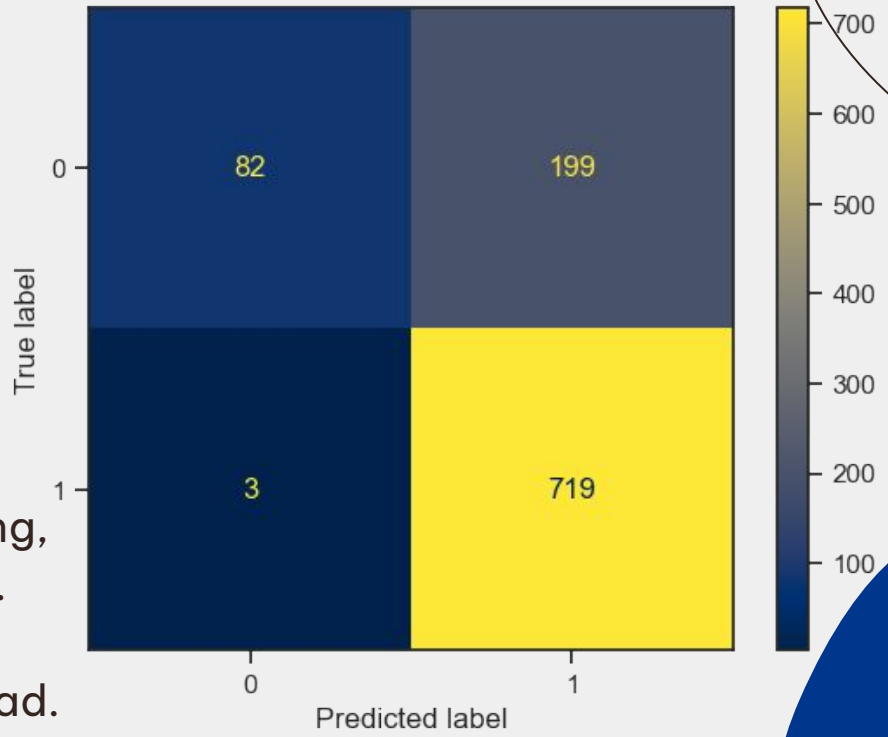
SVC performs better over
RandomForestClassifier/DecisionTreeClassifier/LogisticRegression

	Model 1 previous job data	Model 2 Current job data	Model 3 Previous & Current
Accuracy Score	78 % (Baseline: 69 %)	69 % (Baseline: 69 %)	79 % (Baseline: 71 %)
Precision Score	76 %	69 %	78 %
Recall Score	99.0 %	100 %	99.5 %

Target Value 1 means employee thinks they would be reviewed negatively by co-workers. Therefore it is ideal to optimize by Recall score, however...

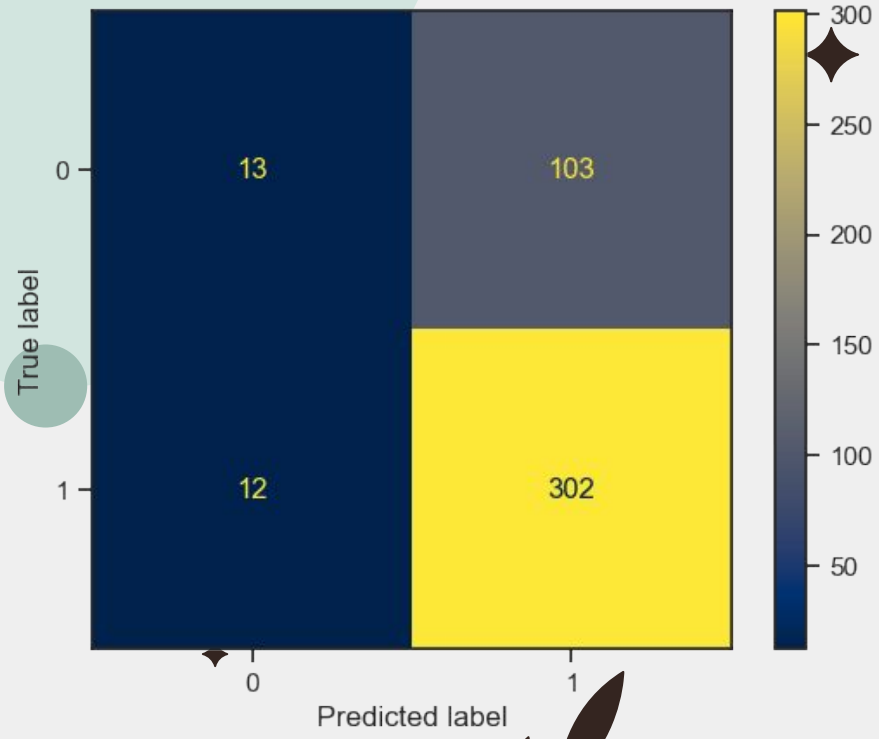
Problem appears on Confusion Matrix





When Optimizing by Recall, regardless of tuning, Model always predict 1 for every employees.

Therefore had to choose Accuracy score instead.



Test Data Evaluation

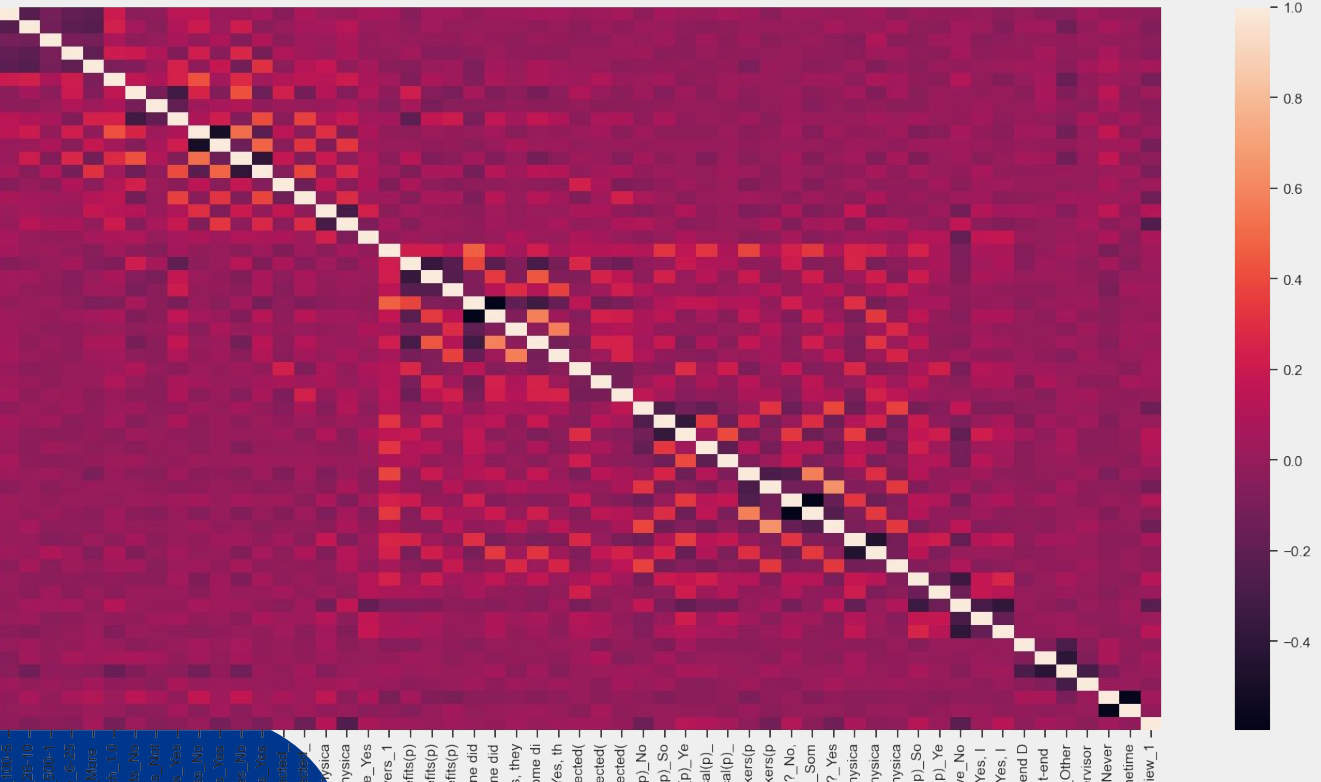
Accuracy: 73.2%
(Baseline: 73.0%)

Recall: 96%
Precision: 74%

103 Errors on False Positive and Accuracy almost same as Baseline indicate the model is not reliable.

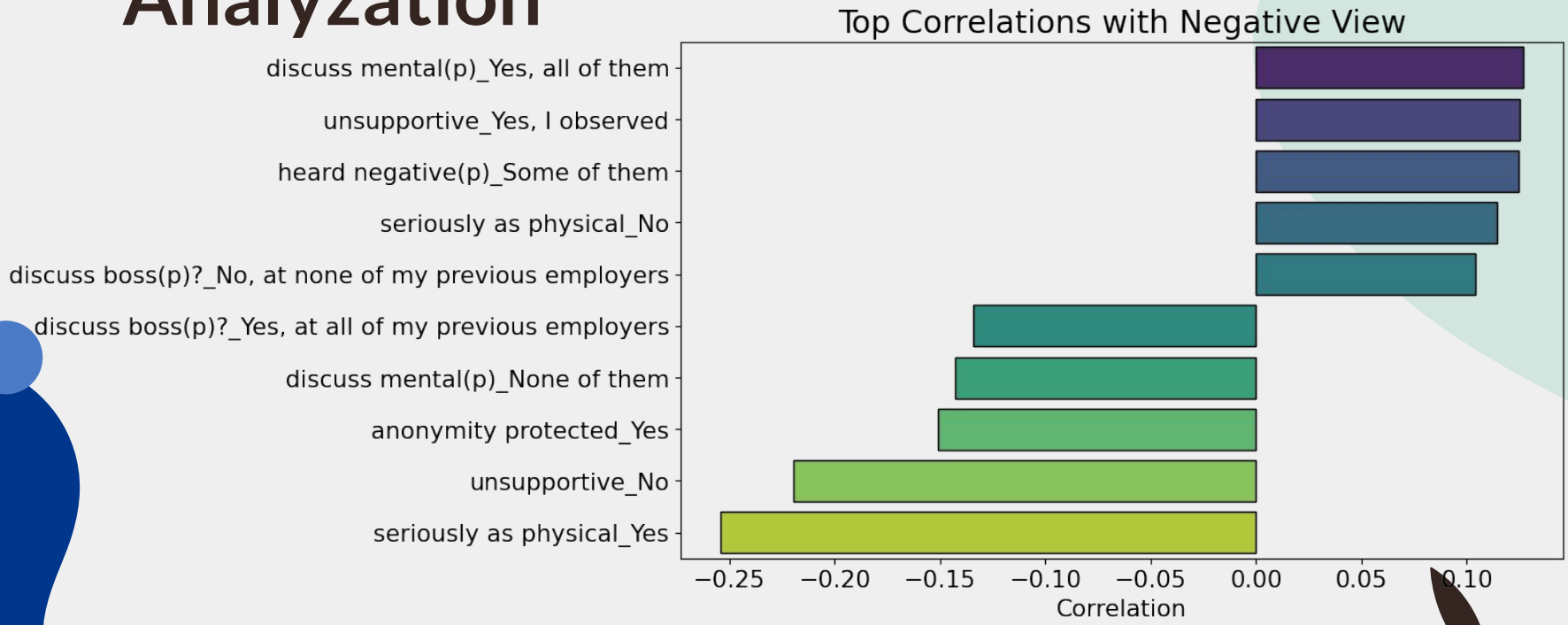
Post Test Analyzation

↓ Target



Issue :
Low correlation of every features with the Target Value (the rightmost column) indicates Target Value was not suitable for prediction

Post Test Analyzation



There are some correlation but might not be significant enough for prediction

Summarize

Previous Job influences Employee's view of the current Workplace regarding Mental Health

Models with previous job features performs well for predicting training data

Current Workplace environment such as being supportive or treat Mental Health as serious as physical health could relieve Employees

Those features have Negative Correlationship with the Target Value

Learning

01

Human thoughts is complex

Might be challenging to predict by ML with limited information

02

Visualize & Analyze before building Model

Both human and machine needs to work

03

Some Correlation is not Prediction

Target Value needs to be independent but needs to be predicted

04

Confusion Matters as much as high scores

Even precision/recall scores are high, Confusion Matrix may reveal the bad prediction

Thank you!

OSMH/OSMI Mental Health in Tech Survey 2016

https://osmhhelp.org/research.html