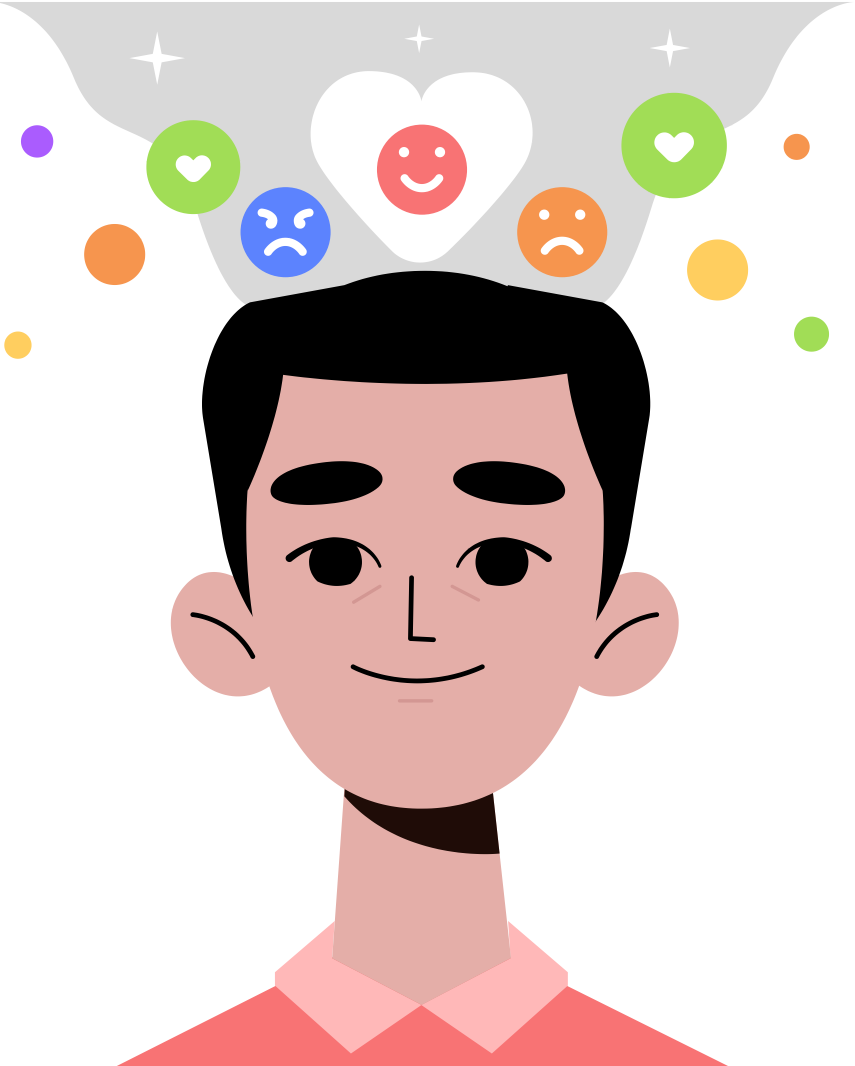


DEPRESSION PREDICTION

By Nouha BEL HAJ YOUSSEF
Date: January, 11th 2024

Plan

01 Depression

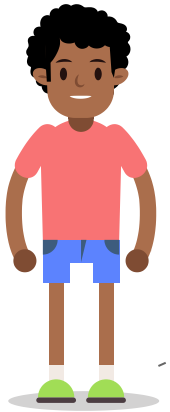
02 Social media and artificial intelligence

03 Dataset

04 Steps of the project

05 Detailed Steps

Depression



Kids at school

Depression affects

Physical health

Mental health

Adults at work

Social Media & AI

Social Media
& AI ?

Social Media

- ✓ Real-time data
- ✓ Rich Source of Data (thoughts, feelings, and daily experiences)

Machine Learning

- ✓ Early Detection

Dataset

Dataset name: Sentimental Analysis for Tweets

Key features: message to examine / label

Size: 10314 rows

Steps of the project

BERT

Random Forest

SVM

Dataset

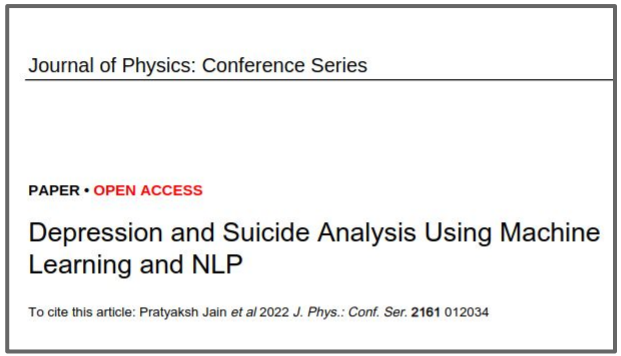
Sentimental Analysis for Tweets

From kaggle.com

Scientific paper

“Depression and Suicide Analysis Using Machine Learning and NLP”

By Pratyaksh Jain, 2022


Journal of Physics: Conference Series

PAPER • OPEN ACCESS

Depression and Suicide Analysis Using Machine Learning and NLP

To cite this article: Pratyaksh Jain *et al* 2022 *J. Phys.: Conf. Ser.* **2161** 012034

Scientific Paper

Pre-processing

Lowercasing / Tokenization
Remove : punctuation, numbers
Remove stop words / Stemming

Metrics

Precision
Recall
f1-score

Naïve bayes

“Depression and Suicide Analysis Using Machine Learning and NLP”

SVM

Random Forest

Logistic Regression

Preprocessing

01

Lowercasing / Removing punctuation and numbers / Stopwords removals / Stemming

02

K-Fold Cross-Validation (generalization)

Train and validating the model on K different combinations

03

Word Embeddings

Use GloVe to convert words into vector representations

04

Scaling (Standardization: Xstd=std(X)X−mean(X))

Prevent certain features from dominating others during the training process.

Model 1

SVM

Objective:

Find the optimal **hyperplane** that best **separates** the data into different classes.

Kernel = 'Linear'

Kernel = 'RBF'

non-linear

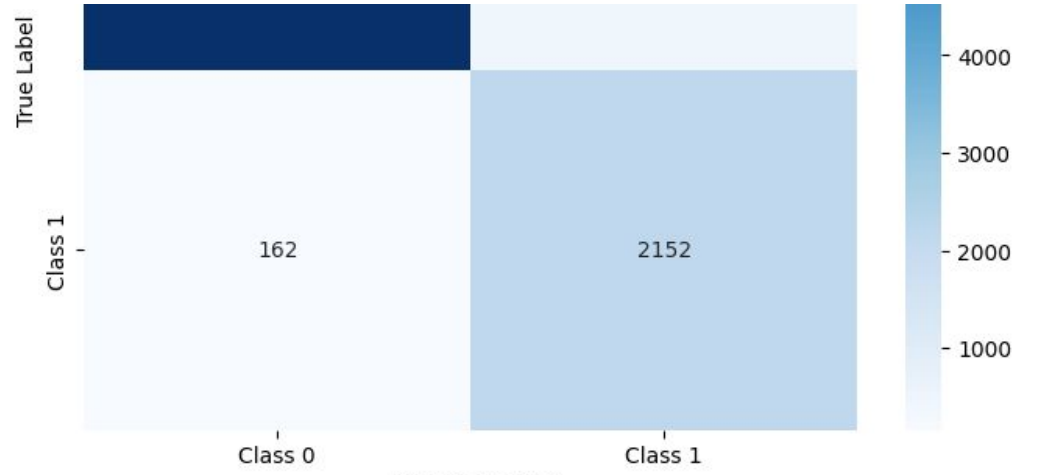
Kernel : transform the input data into a higher dimensional space. The choice of kernel determines the decision boundary of the SVM model.

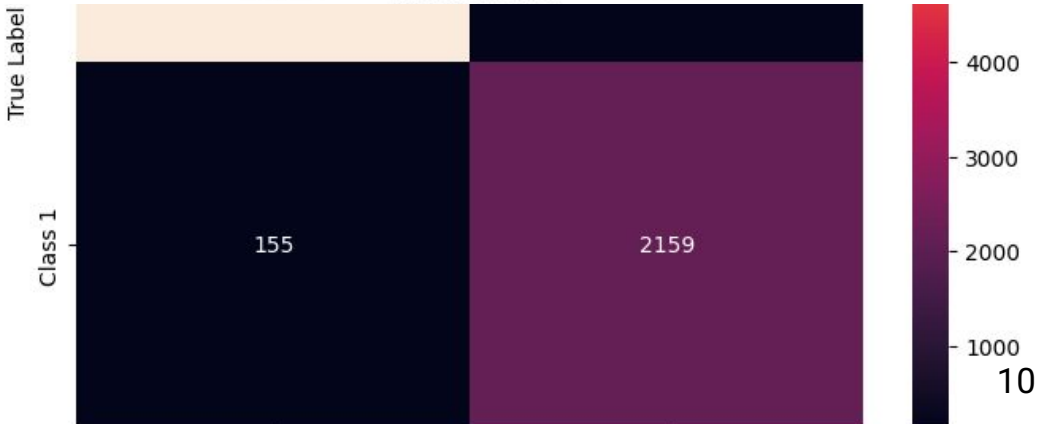
Model 1

Kernel = ‘Linear’

Kernel : transform the input data into a higher dimensional space. The choice of kernel determines the decision boundary of the SVM model.

Kernel = ‘RBF’





Model 2

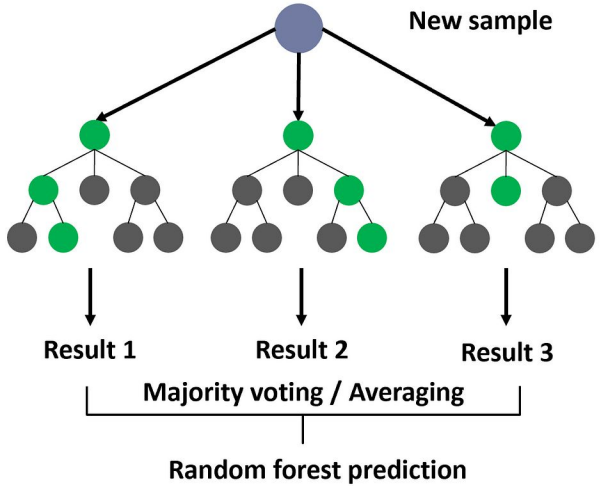
Random Forest

Objective:

builds multiple decision trees during training and outputs the **majority voting** of the classes

Decision Tree

partitions the data into subsets based on the values of input features



Hyperparameters

```
n_estimators=100
```

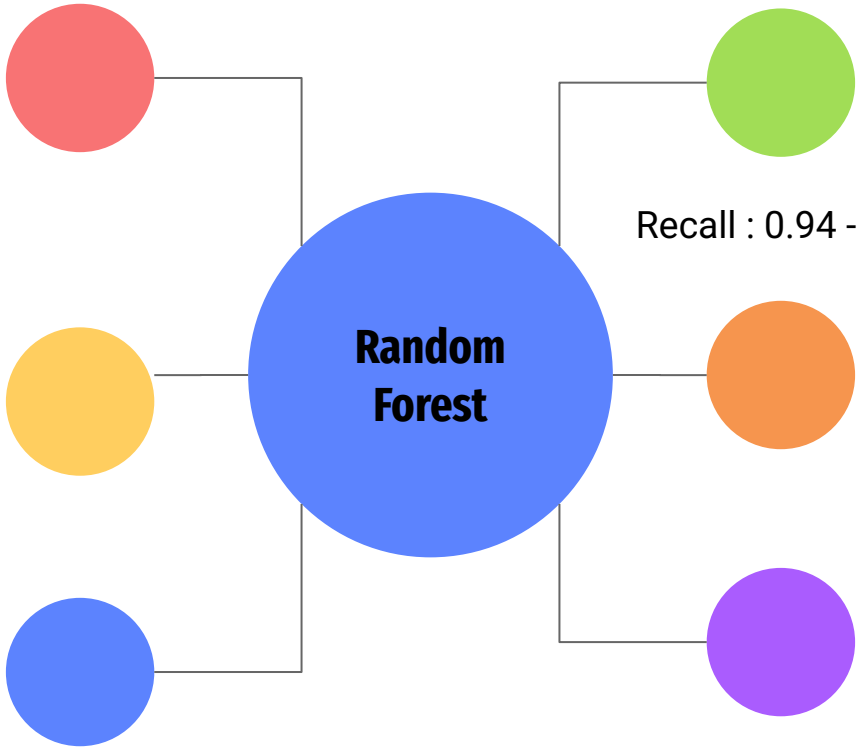
[[7932 68]
[488 1826]]

```
n_estimators=100,
max_features='sqrt'
n_estimators=150
```

[[7925 75]
[487 1827]]

```
n_estimators=150,
max_features='sqrt',
max_depth=20
```

[[7915 85]
[460 1854]]



Recall : 0.94 - 0.95

```
n_estimators=150,
max_features='sqrt',
max_depth=10
```

[[7827 173]
[345 1969]]

```
n_estimators=100,
bootstrap=True
```

[[7932 68]
[488 1826]]

```
n_estimators=100,
bootstrap=True,
min_samples_split=2
```

[[7932 68]
[488 1826]]

Using a pre-trained BERT model without fine-tuning

BERT, or Bidirectional Encoder Representations from Transformers

BERT is built on the Transformer architecture, which uses self-attention mechanisms to weigh the importance of different words in a sentence. This architecture allows BERT to capture long-range dependencies and relationships in text.

We used BERT to analyze and understand the sentiment in tweets.

Loading a pre-trained BERT model along with its tokenizer. The model we chose is specifically designed for sentiment analysis, which aligns well with our task

Created functions to predict the depression probability **for a given tweet** and to calculate the overall accuracy of our model.

To illustrate how our model works, we randomly selected a tweet predicted its depression probability, and compared it with the actual label.

Dataset link :
https://www.kaggle.com/datasets/gargmanas/sentimental-analysis-for-tweets/data
Article link :
https://iopscience.iop.org/article/10.1088/1742-6596/2161/1/012034/pdf

Thank you for your attention