Ground Water Management

Water Quality Prediction

CS304N Minor Project

Batch 11

Team Members:

Anjali Gupta

190001004

Deepali Sukhija

190001009

Deepika Sukhija

190001010

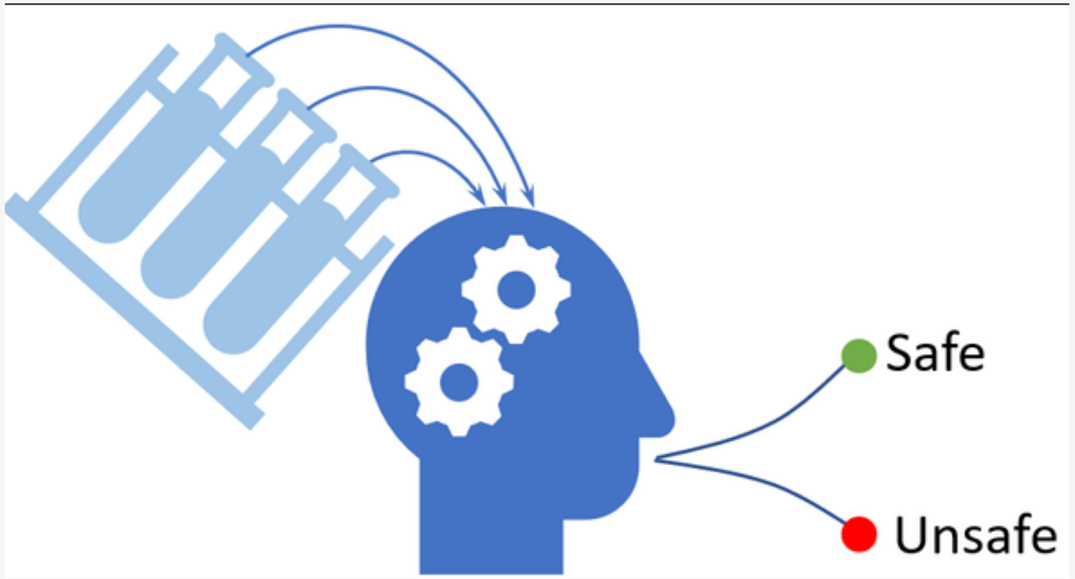
Submitted to :

Dr. Aruna Tiwari

Agenda

- Introduction
- Dataset
- Data Visualisation
- ML Model
- Deep Learning
- Conclusion

INTRODUCTION



In rural areas, groundwater is the primary source of water. Groundwater is used for drinking, agriculture, and many more purposes. So, the villagers should know whether the water is safe or unsafe. If it is hazardous, people could use it for purposes other than drinking water as it will be dangerous for their health. So, we thought of developing an Artificial Intelligence Model that will predict the water quality, whether it is safe or unsafe, based on the parameters like ph, hardness, Chloramines, Sulfate, etc. present in the groundwater.

DATASET

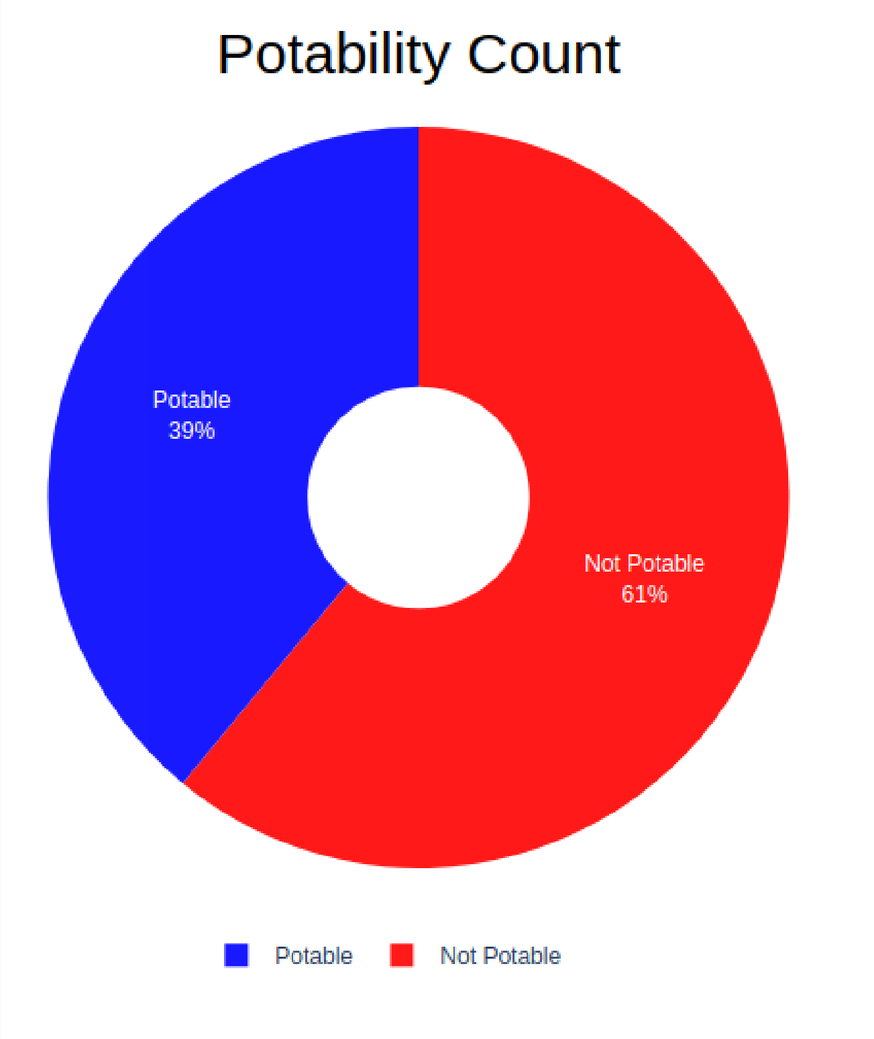
Column in Dataset Description:

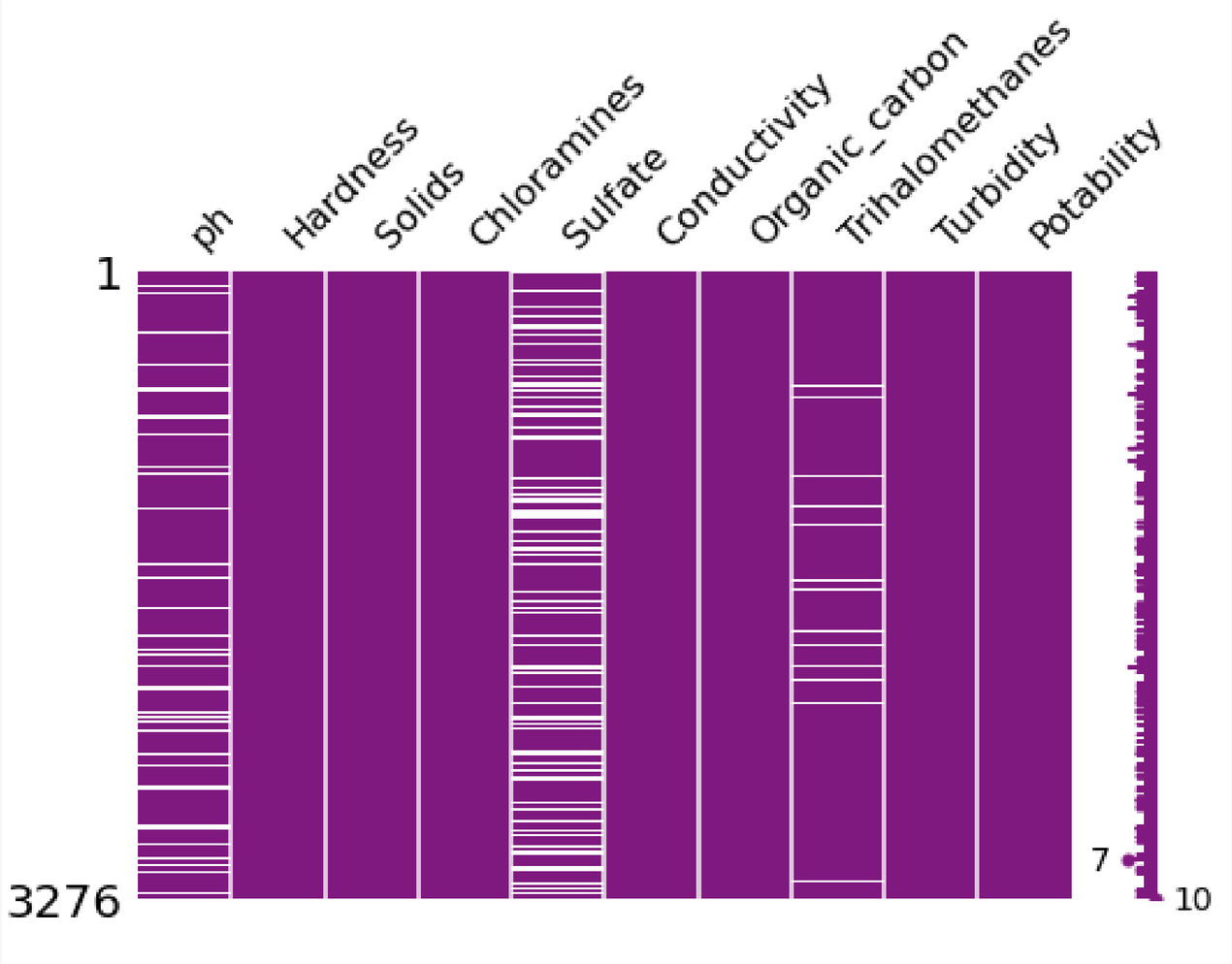
1) **ph:** pH of water is one of the important parameter for evaluating acid-base balance.
2) **Hardness:** Capacity of water to precipitate soap caused by Calcium(Ca) and Magnesium(Mg) in mg/L.
3) **Solids:** Total dissolved solids in ppm.
4) **Chloramines:** Amount of Chloramines in ppm.
5) **Sulfate:** Amount of Sulfates dissolved in mg/L.
6) **Conductivity:** Electrical conductivity of water in µS/cm.
7) **Organic_carbon:** Amount of organic carbon in ppm.
8) **Trihalomethanes:** Amount of Trihalomethanes(chemicals found in water treated with chlorine) in µg/L.
9) **Turbidity:** Measure of light emitting property of water in NTU.
10) **Potability:** Indicates if water is safe for human consumption where 1 means Potable and 0 means not potable.

Note:

ppm: parts per million
µg/L: microgram per litre
mg/L: milligram per litre

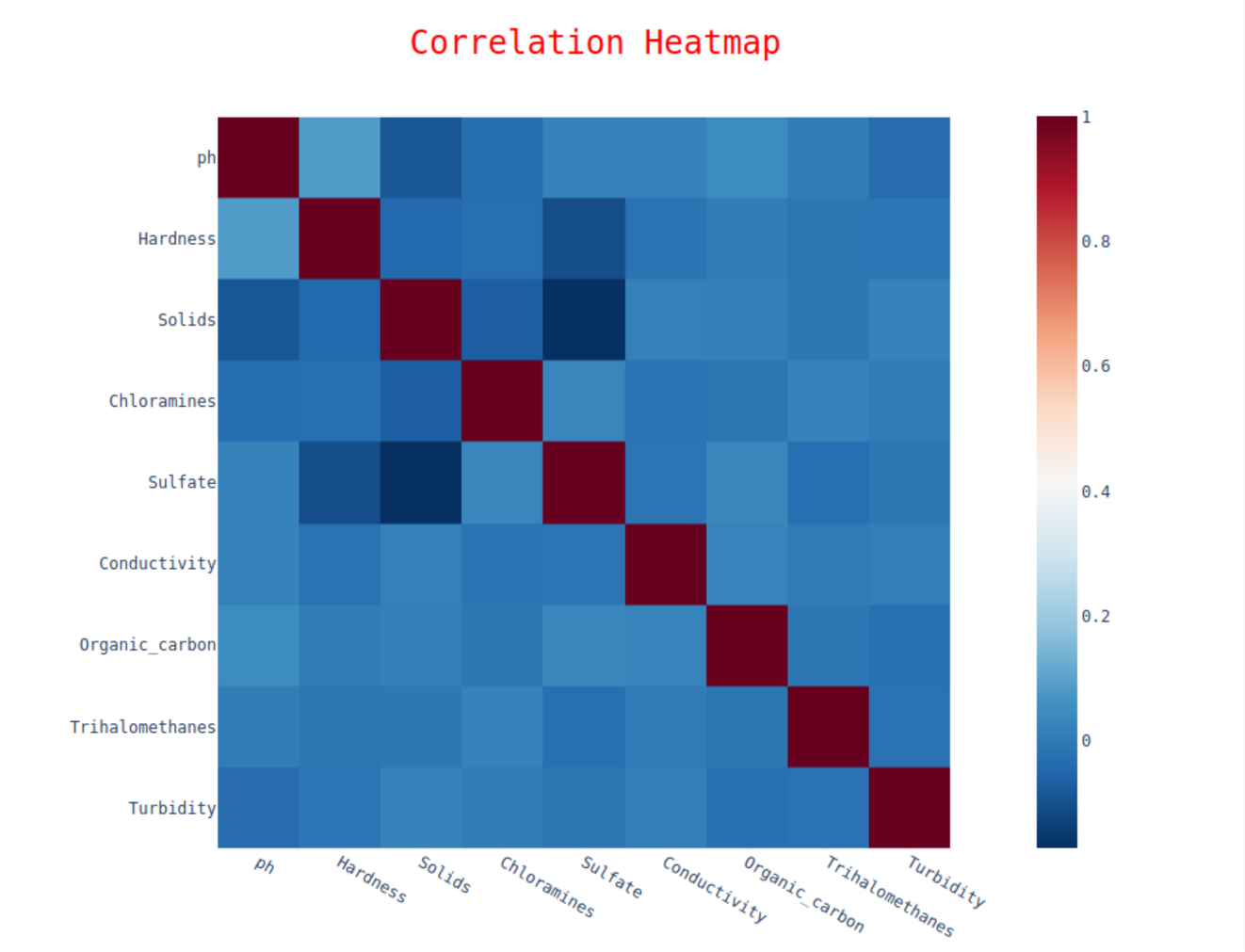
DATA VISUALISATION





DATA VISUALISATION

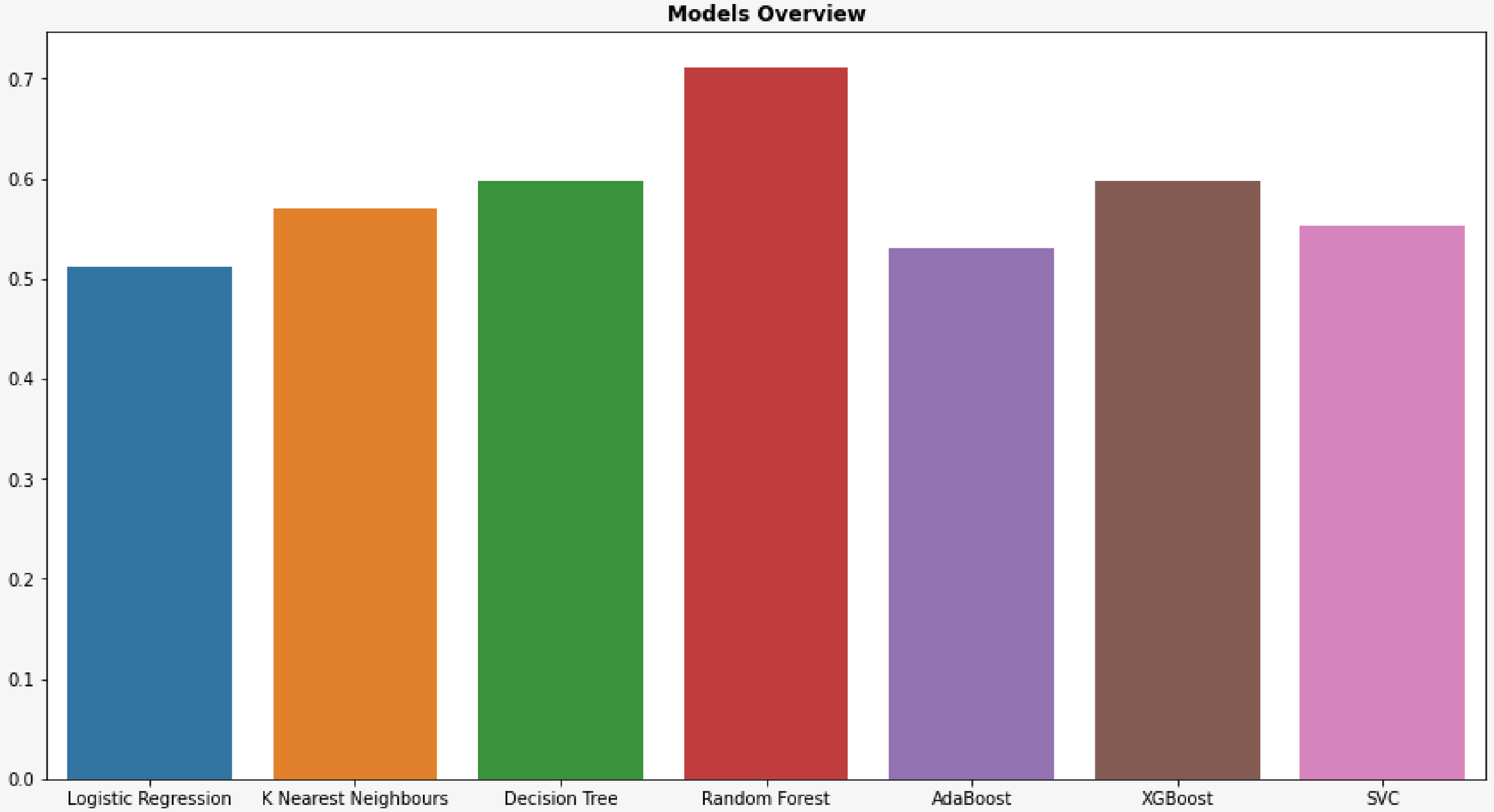
continued



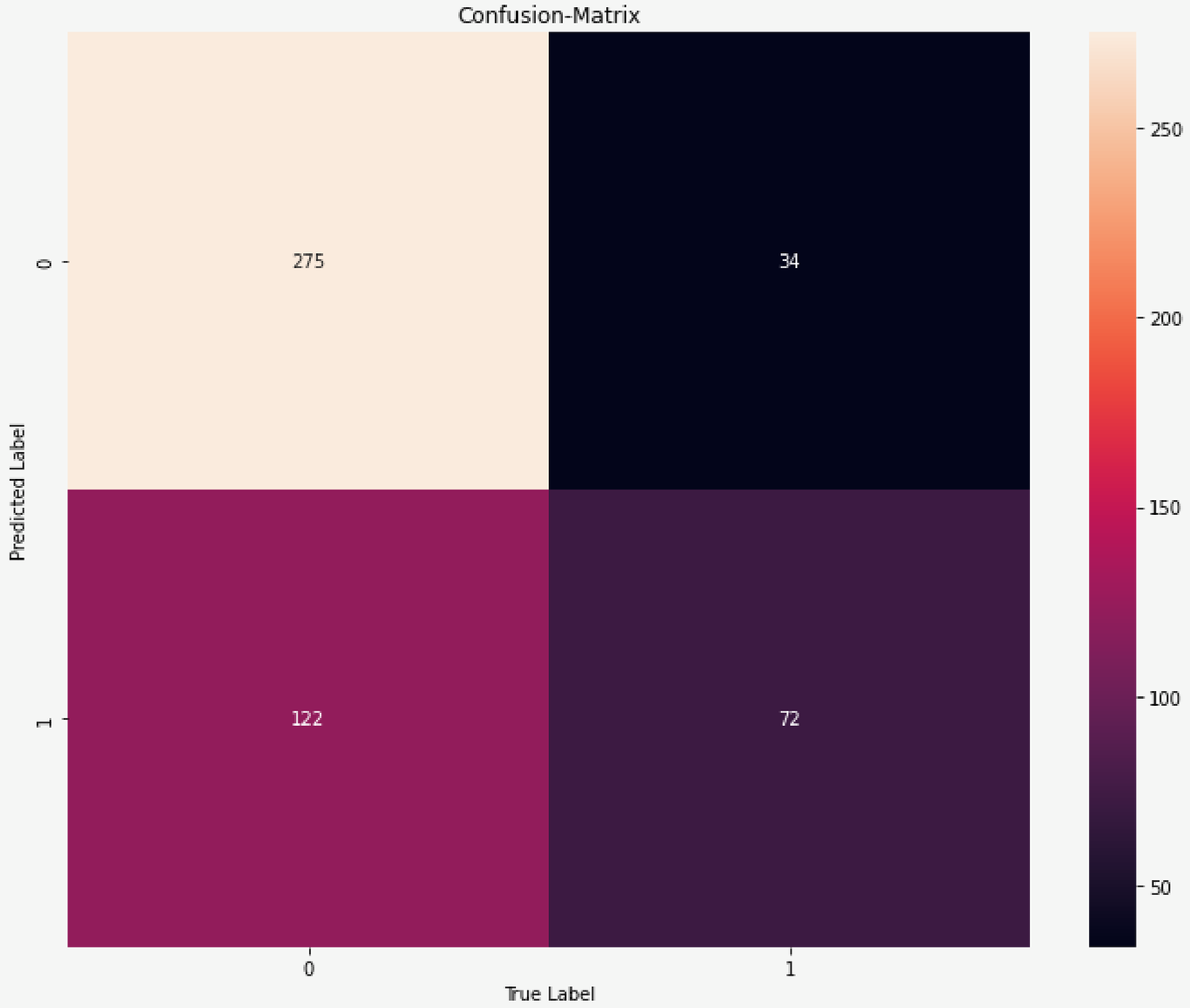
APPLIED ML MODEL

1. **Logistic Regression**
2. **K-Neighbors Classifier**
3. **Decision Tree Classifier**
4. **Random Forest Classifier**
5. **Ada Boost Classifier**
6. **XGB Classifier**
7. **SVC**

ML MODEL ACCURACY GRAPH



ML MODEL RESULTS



ML MODEL RESULTS

```
print(classification_report(y_test, y_pred))
```

```
              precision    recall  f1-score   support

           0       0.69      0.89      0.78       309
           1       0.68      0.37      0.48       194

    accuracy                           0.69       503
   macro avg       0.69      0.63      0.63       503
weighted avg       0.69      0.69      0.66       503
```

DEEPLEARNING

Our deep learning model has 5 blocks.

The 1st block contains the dense layer of 1000 units followed by 'relu' activation function.

The next block contains the dense layer of 500 units, followed by 'relu' activation function and dropout of 0.15.

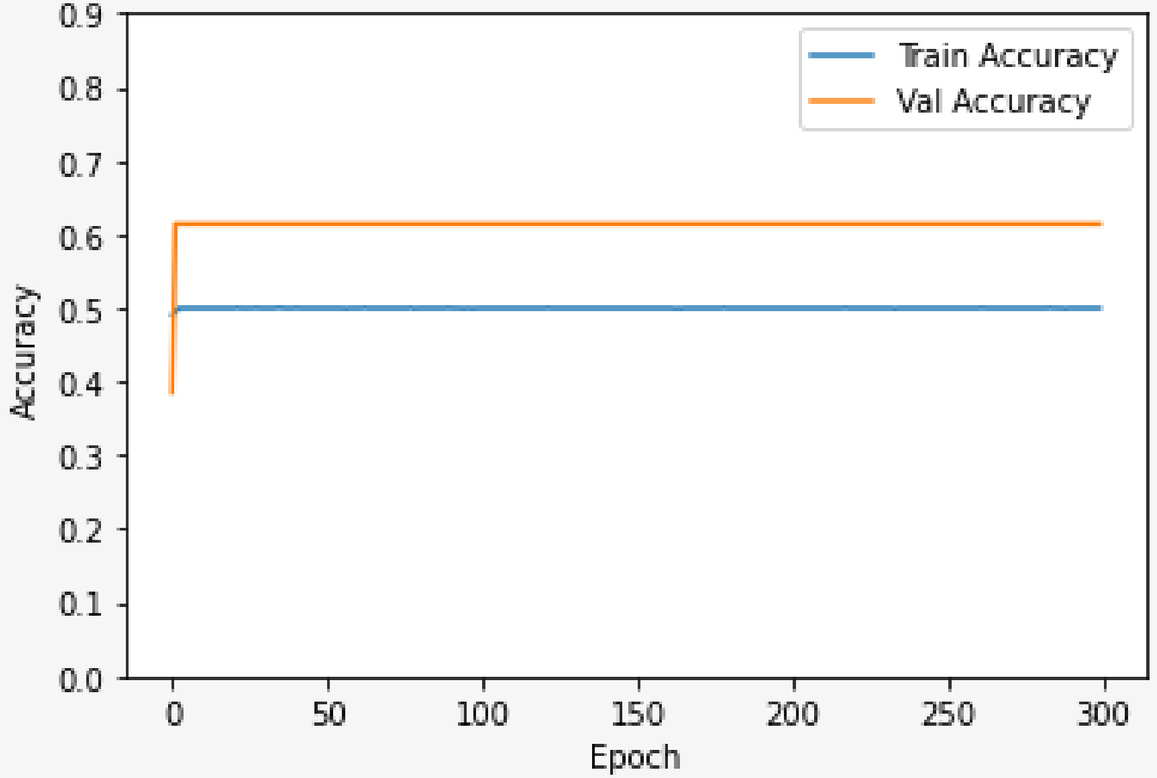
The next block contains the dense layer of 100 units, followed by 'relu' activation function and dropout of 0.15.

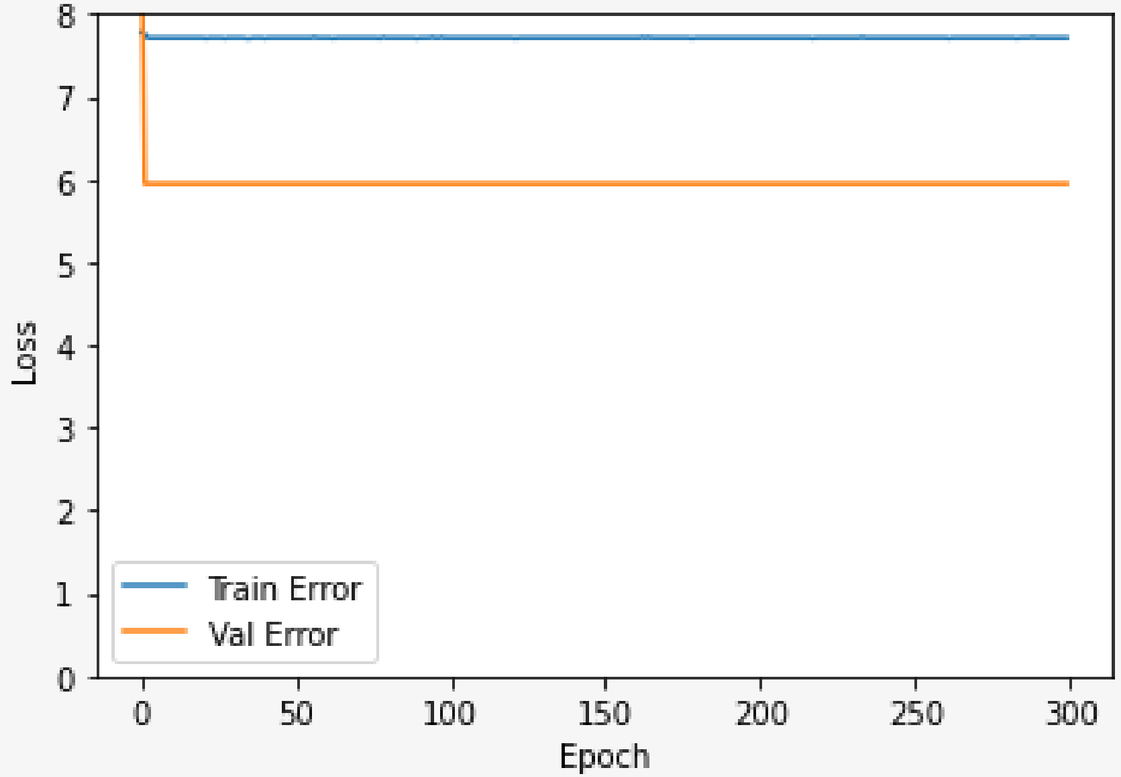
The next block contains the dense layer of 50 units, followed by 'relu' activation function and dropout of 0.15.

The next block contains the dense layer of 1 units, followed by 'relu' activation function.

We have used the 'adam' optimizer and 'binary_crossentropy' and received a accuracy of 0.6143.

DEEP LEARNING GRAPHS





CONCLUSIONS

1. **The correlation coefficients between the features were very low.**
2. **We received almost similar performance for for both the classes- potable and non potable. For potable (1), we got an Precision of 0.68 , Recall of 0.37 ,F1-score of 0.48.**
3. **For non-potable (0), we got an Precision of 0.69 ,Recall of 0.89 ,F1-score of 0.78.**
4. **We also trained neural network models and experimented on many parameters.**
5. **We received the best accuracy of 0.71 on Random Forest.**

Thank you!