Diabetes Prediction Using Machine Learning

Submitted by:-

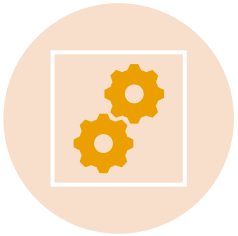
ANN BIJU MARIYAM UIN:678591542

NITHYASHREE BABU REDDY UIN:660354535

OVERVIEW

- INTRODUCTION

- DATASET

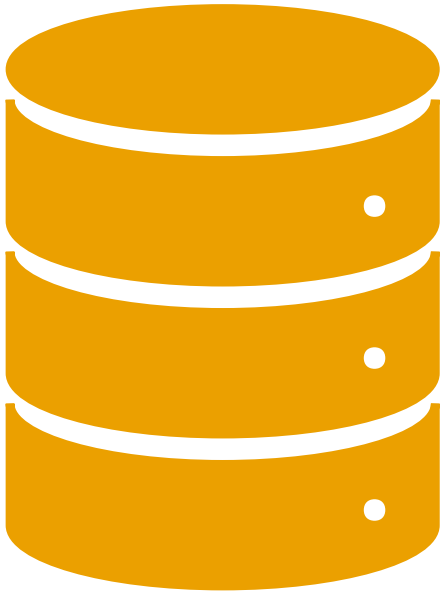


- METHODOLOGY

- RESULTS

- CONCLUSION

DATASET



Diabetes Prediction Dataset Overview

- **Source:** https://www.kaggle.com/datasets/iammustafatz/diabetes-prediction-dataset
- **Dataset Summary:**
  - **Samples:** 768
  - **Features:** 9
- **Features:**
  - **Pregnancies:** Number of pregnancies
  - **Glucose:** Plasma glucose concentration
  - **Blood Pressure:** Diastolic blood pressure (mm Hg)
  - **Skin Thickness:** Triceps skinfold thickness (mm)
  - **Insulin:** 2-hour serum insulin (mu U/ml)
  - **BMI:** Body mass index (kg/m^2)
  - **Diabetes Pedigree Function:** Genetic risk of diabetes
  - **Age:** Age in years
  - **Outcome (Target Variable):**
    - 0 = Non-diabetic
    - 1 = Diabetic

MODELS USED

NEURAL NETWORK

LOGISTIC REGRESSION

SVM

RANDOM FOREST

METHODOLOGY

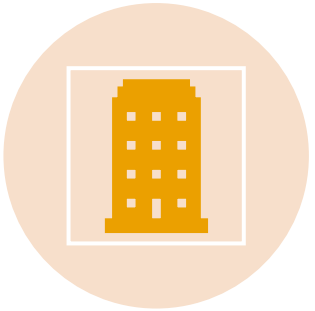
- Preprocessing: Missing values, encoding, scaling
- Models used: Neural Network, Logistic Regression, Random Forest, SVM
- Train-Test Split: 80-20
- Evaluation: Accuracy, Precision, Recall, F1-score, ROC-AUC

RESULTS

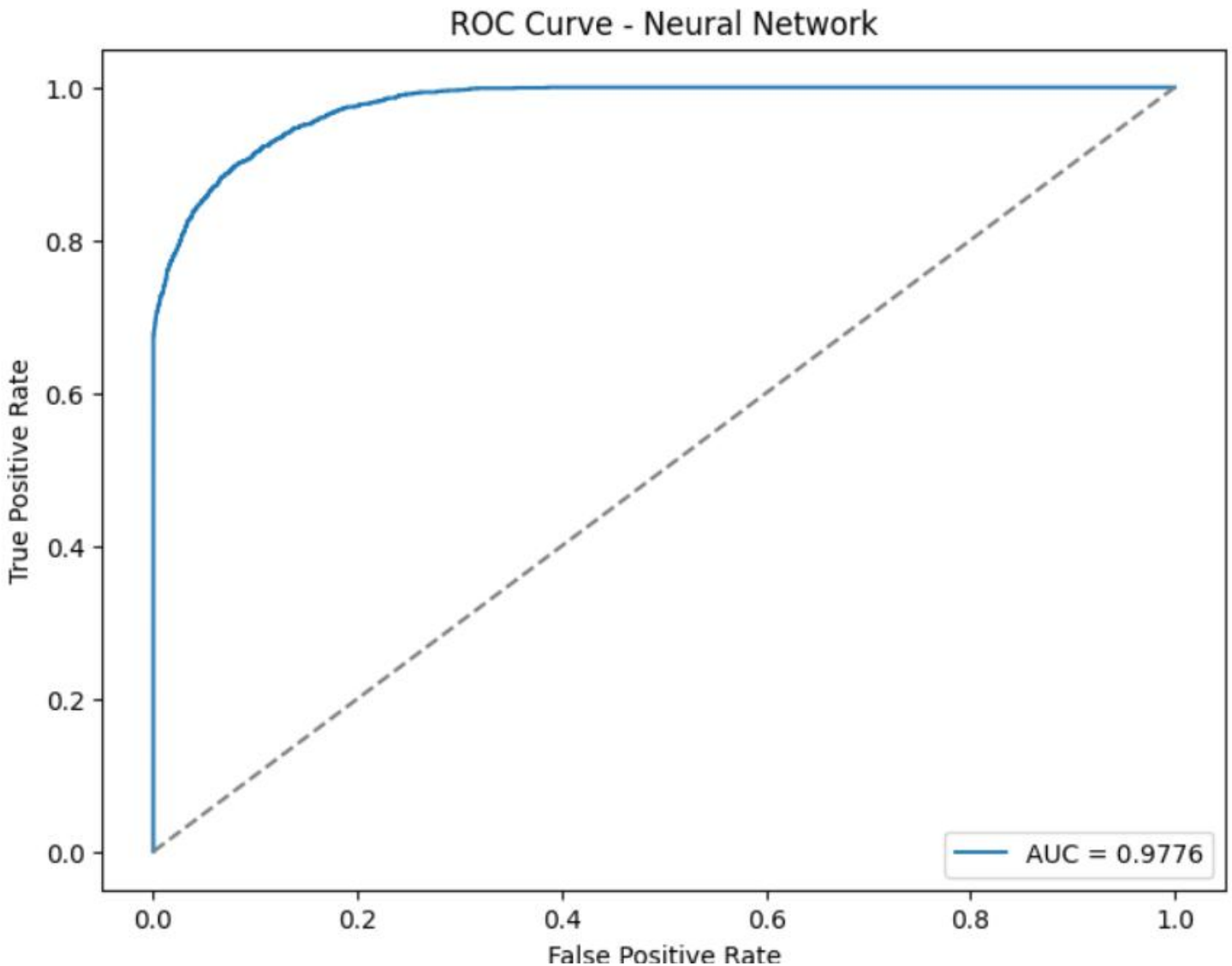
- PERFORMED BEST : NEURAL NETWORK

- ACCURACY: 97.20%

- PRECISION: 98.81%



- ROC-AUC: 97.76%



Confusion Matrix for Neural Network:

```
[[18278    14]
 [  546  1162]]
```

MODEL COMPARISON

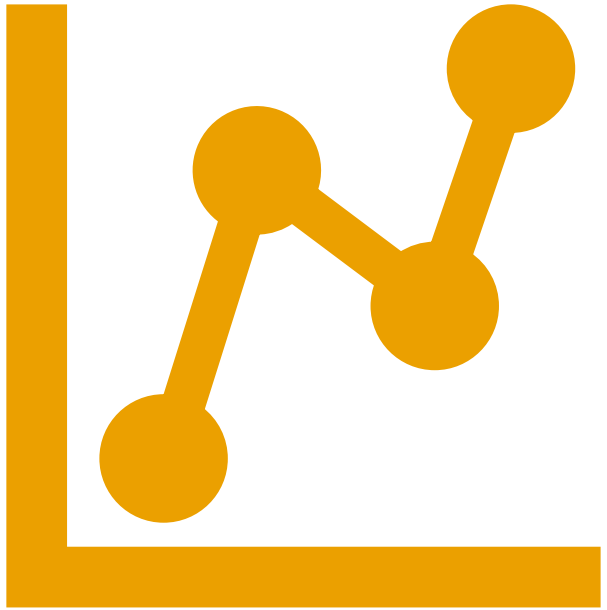
Metric	Neural Net	Logistic Reg	Random Forest	SVM
Accuracy	0.9720	0.9587	0.9708	0.9624
Precision	0.9881	0.8639	0.9532	0.9761
Recall	0.6803	0.6130	0.6915	0.5732
F1-Score	0.8058	0.7171	0.8015	0.7222
AUC-ROC	0.9776	0.9612	0.9637	0.9319

MODEL INSIGHTS

- Neural Network had the highest accuracy, precision, F1-score, and AUC-ROC.
- Random Forest had the highest recall, making it more sensitive to detecting positive cases.
- Logistic Regression performed well and is simpler to interpret.
- Random Forest was close to NN in performance but more explainable.

→ Neural Network is the best overall performer for this dataset.

CONCLUSION



- Machine learning models can predict diabetes effectively
- Random Forest had highest performance
- Preprocessing plays a key role
- Useful in early detection and healthcare support

THANK YOU…