Report: Diabetes Prediction using Random Forest Classifier

**I. Introduction**

Diabetes is a chronic disease affecting millions worldwide, characterized by high blood sugar levels. Early detection is crucial for effective management and prevention of complications. This project aims to predict diabetes using the Pima Indians Diabetes Dataset.

**II. Methodology**

1. **Data Preprocessing**: The dataset was preprocessed by scaling features using StandardScaler and transforming data using PolynomialFeatures.
2. **Feature Engineering**: Polynomial features were used to capture interactions between variables.
3. **Model Selection**: RandomForestClassifier was chosen for its robustness and handling of non-linear relationships.
4. **Hyperparameter Tuning**: Hyperparameters were tuned using grid search and cross-validation.

**III. Results**

* **Model Performance Metrics**:

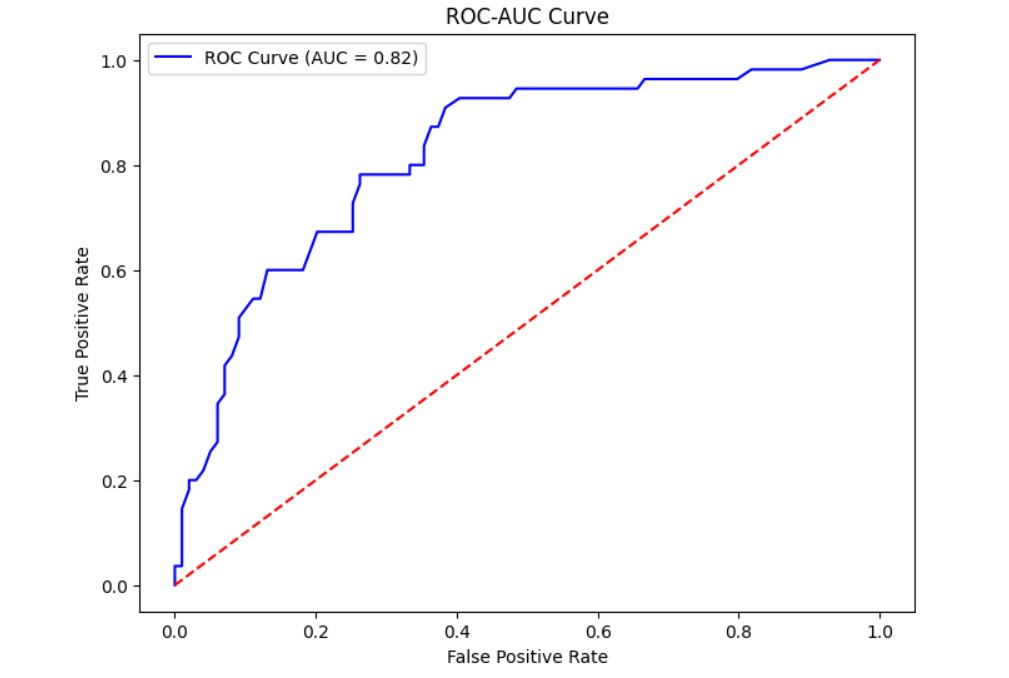
\* Accuracy: 73.38%

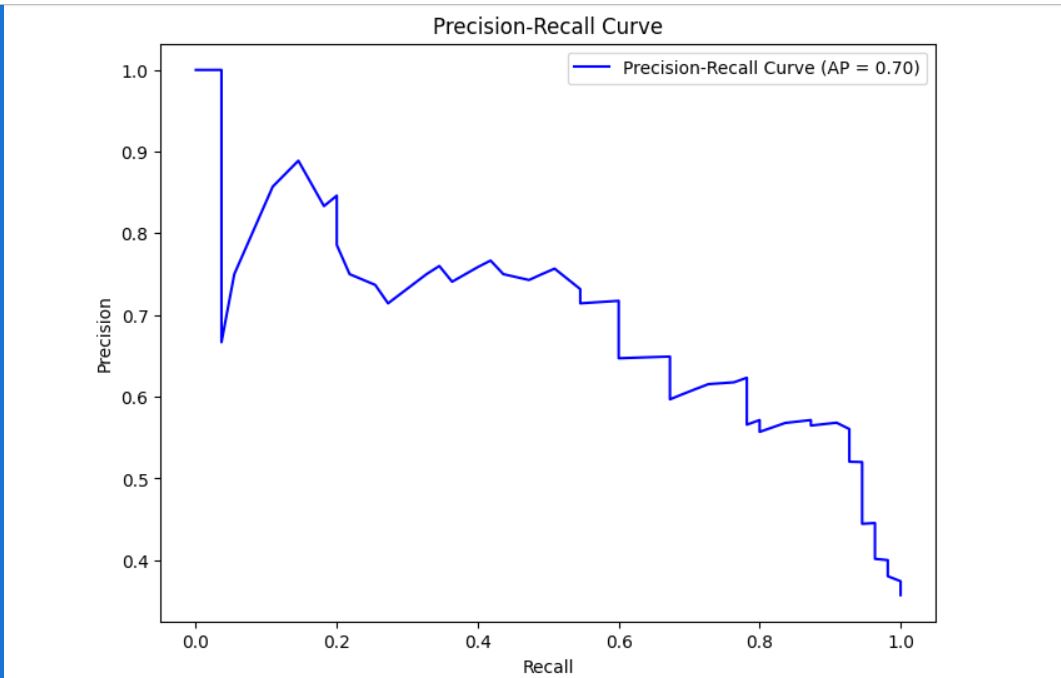
\* Precision (Class 0): 0.81, (Class 1): 0.62

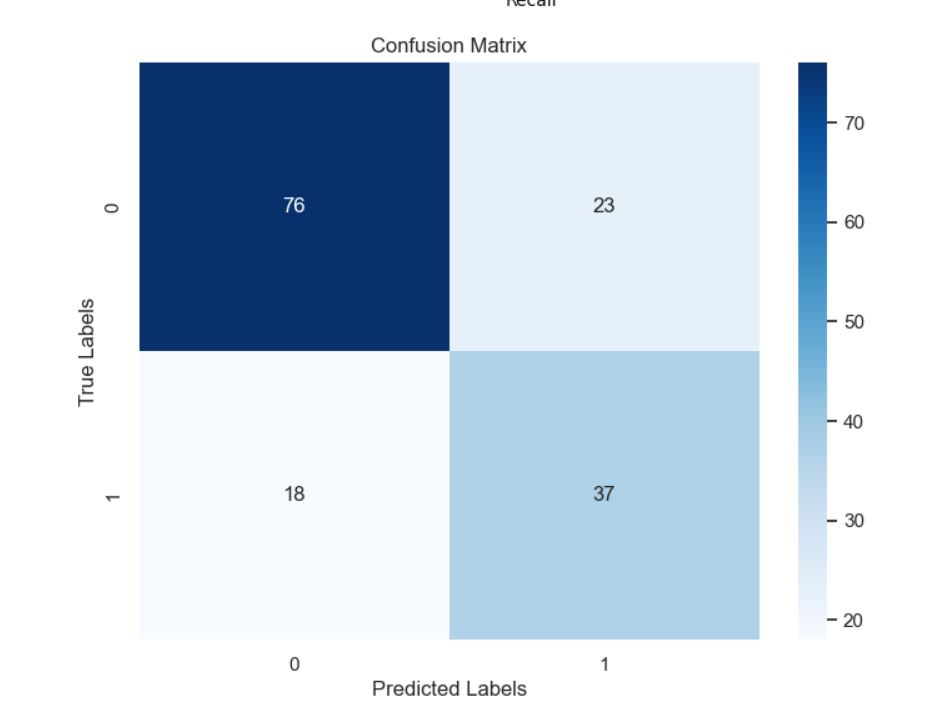
\* Recall (Class 0): 0.77, (Class 1): 0.67

\* F1-score (Class 0): 0.79, (Class 1): 0.64

* **Visualizations**:







**IV. Discussion**

The results indicate that the model effectively predicts diabetes with an accuracy of 73.38%. However:

* Class imbalance affects model performance, with lower precision for Class 1.
* Feature engineering techniques improved model performance.

**V. Conclusion**

In conclusion:

* The RandomForestClassifier accurately predicts diabetes using the Pima Indians Diabetes Dataset.
* Future work could explore incorporating additional features, handling class imbalance, and comparing models.