A Comparative Analysis of Machine Learning Models for Classification Tasks

By:-

Mohammed Zubaidi Shaik - 16353803

Introduction

- Introduction to the study
- Importance of selecting the right classification model
- Overview of the models analyzed

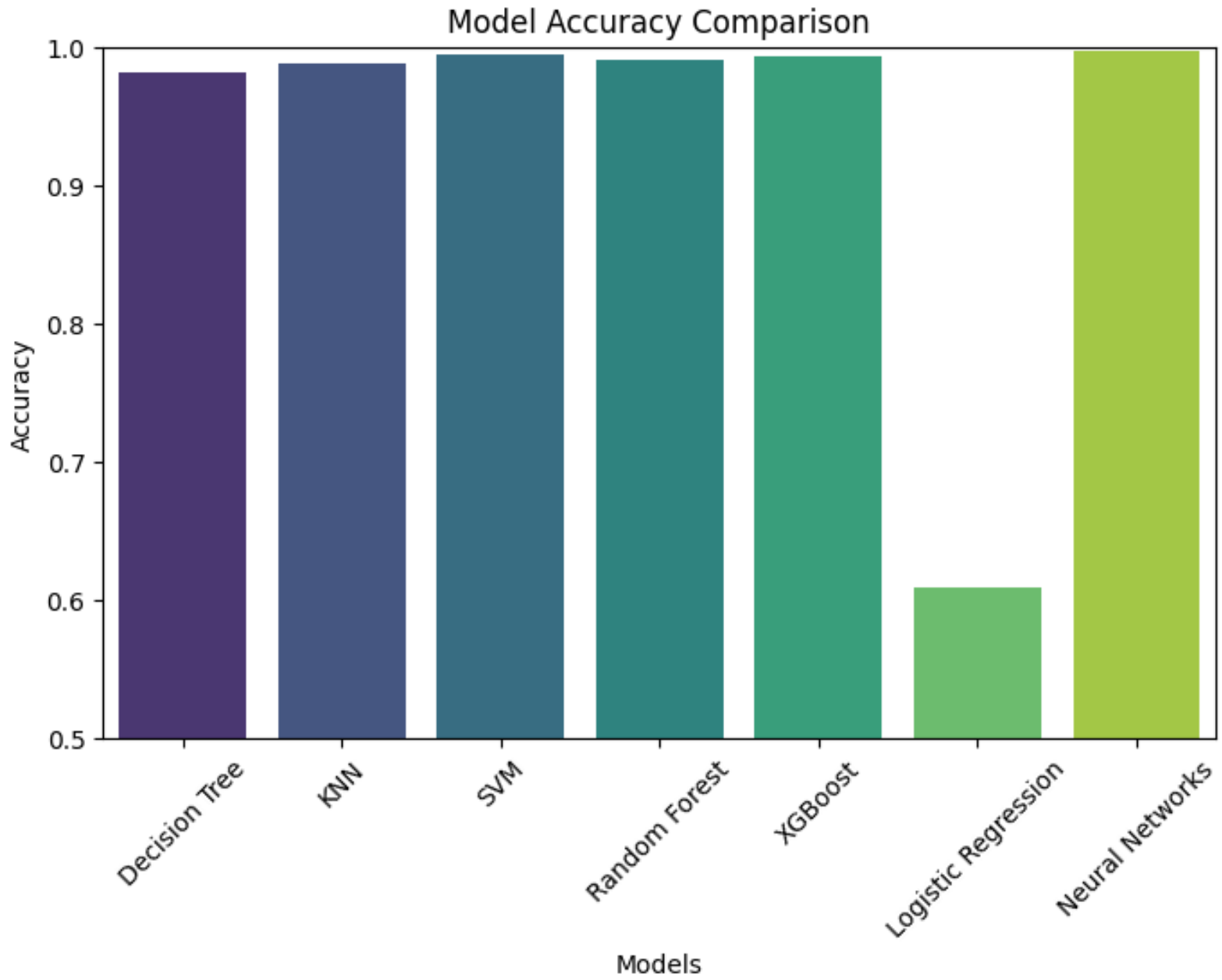
Evaluation Metrics & Methodology

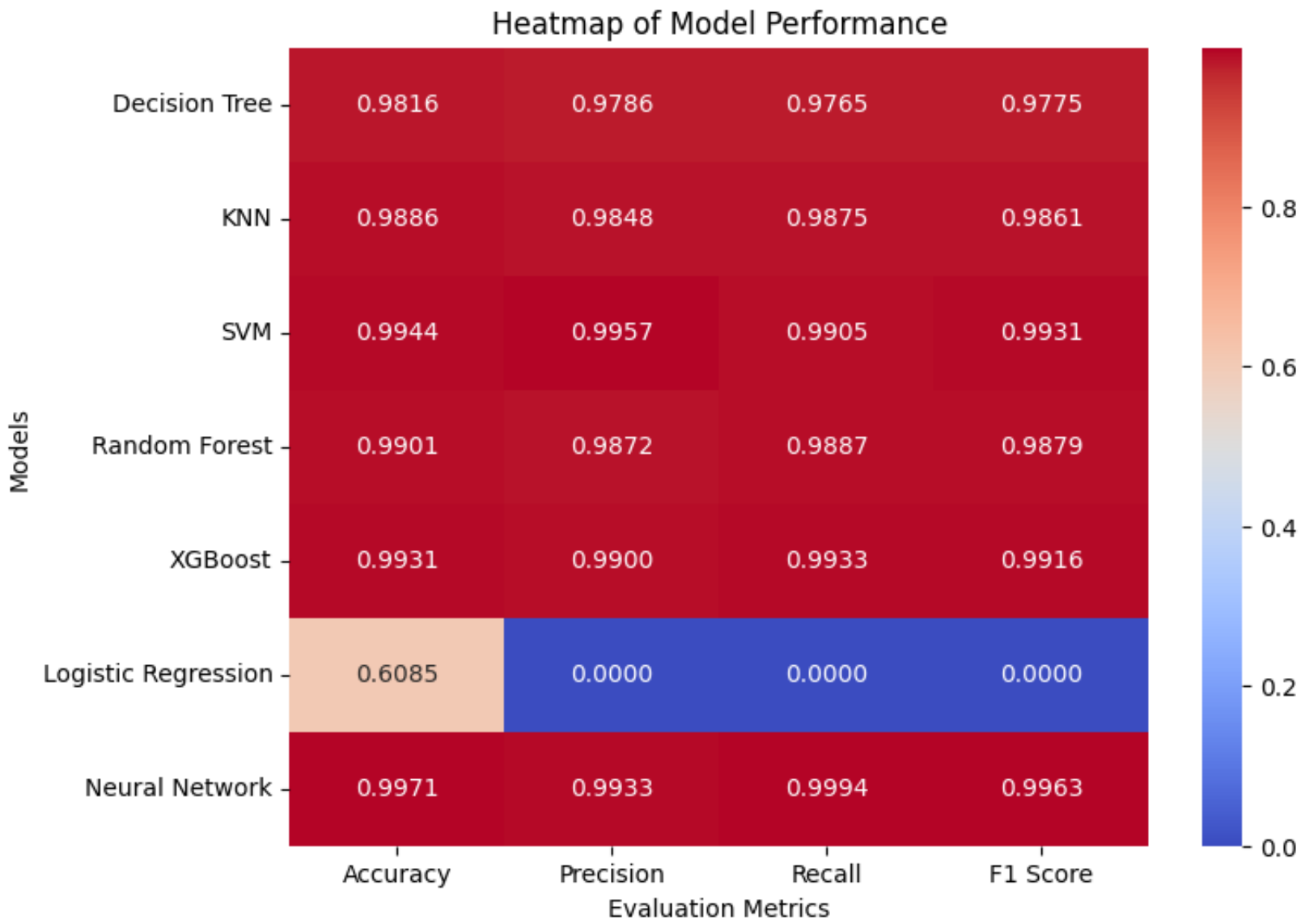
- Dataset used and evaluation criteria
- Models implemented using Scikit-learn & TensorFlow
- Key metrics: Accuracy, Precision, Recall, F1 Score, AUC

Model Performance Results

- Comparison of model performances (summary of table)
- Neural Networks achieved the highest accuracy (99.71%)
- XGBoost and Random Forest performed well
- Logistic Regression performed poorly

Model	Accuracy	Precision	Recall	F1 Score	AUC
Decision Trees	0.9816	0.9786	0.9765	0.9775	0.9808
KNN	0.9886	0.9848	0.9875	0.9861	0.9994
SVM	0.9944	0.9957	0.9905	0.9931	N/A
Random Forest	0.9901	0.9872	0.9887	0.9879	0.9997
XGBoost	0.9931	0.9900	0.9933	0.9916	0.9998
Logistic Regression	0.6085	0.0	0.0	0.0	0.4983
Neural Networks	0.9971	0.9933	0.9994	0.9963	0.9975


Model Accuracy Comparison
1.0
0.9
0.8
0.7
0.6
0.5
Accuracy
Decision Tree
KNN
SVM
Random Forest
XGBoost
Logistic Regression
Neural Networks
Models


Heatmap of Model Performance
Decision Tree
KNN
SVM
Random Forest
XGBoost
Logistic Regression
Neural Network
Models
0.9816
0.9786
0.9765
0.9775
0.9886
0.9848
0.9875
0.9861
0.9944
0.9957
0.9905
0.9931
0.9901
0.9872
0.9887
0.9879
0.9931
0.9900
0.9933
0.9916
0.6085
0.0000
0.0000
0.0000
0.9971
0.9933
0.9994
0.9963
Accuracy
Precision
Recall
F1 Score
Evaluation Metrics
0.8
0.6
0.4
0.2
0.0

Discussion & Key Findings

- Neural Networks outperformed other models
- Ensemble models (XGBoost, Random Forest) were highly effective
- SVM and KNN performed well with proper tuning
- Logistic Regression failed due to dataset complexity

Conclusion & Future Work

- Neural Networks and ensemble methods are best for complex tasks
- Simpler models may not be suitable for high-dimensional data
- Future work: Hyperparameter tuning, additional feature engineering
- Testing on other datasets