Student ID: 630615036
Name: Pheerawadee Chantakhat

Report

Drug Classification Project

Abstract

In this project, I classified certain types of drugs based on several features, including age, sex, blood pressure levels, cholesterol levels, and the Na to Potassium ratio. The dataset consisted of 200 records, with the target feature being the type of drug. I split the data into training and testing sets and applied different models, including Logistic Regression, K Neighbors Classifier, SVM(Linear), SVM(Poly), and Decision Tree, to classify drugs based on the given features. Finally, I evaluated the performance of each model using metrics such as accuracy, precision, recall, F1 score, and AUC score. The best-performing model was then selected to classify the drug types based on the given features.

Student ID: 630615036
Name: Pheerawadee Chantakhat

Classification Problem

The classification problem is to predict the type of drug based on features such as age, sex, blood pressure levels, cholesterol levels, and Na to Potassium ratio. This is a multi-class classification problem, with the target variable being the type of drug and the feature set consisting of age, sex, blood pressure levels, cholesterol levels, and Na to Potassium ratio. The provided code performs exploratory data analysis and prepares the data for model training. To solve the problem, I used various classification algorithms, such as logistic regression, k-nearest neighbors, decision tree, SVM(Poly), and SVM(Linear). The dataset was preprocessed by converting categorical columns into numerical values. The goal was to find a model that could accurately classify the drug type with high accuracy and F1-score.

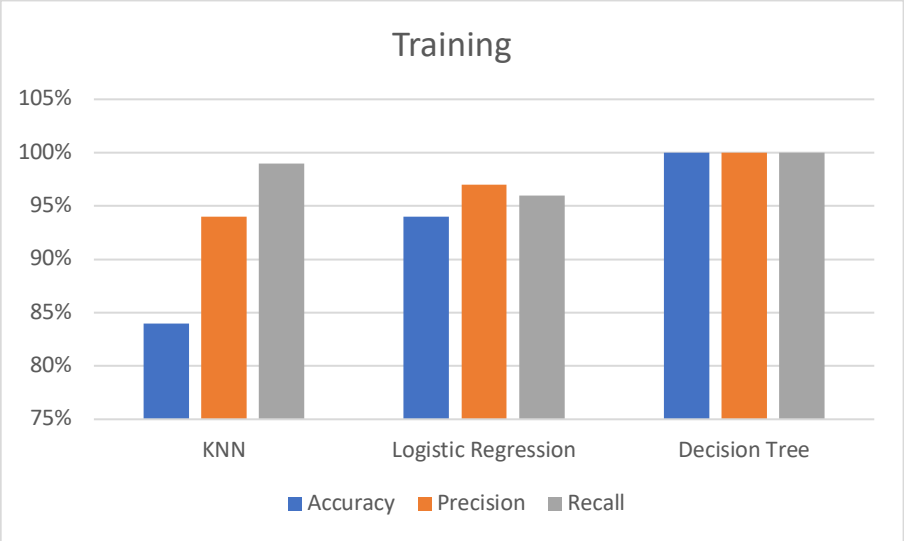
\

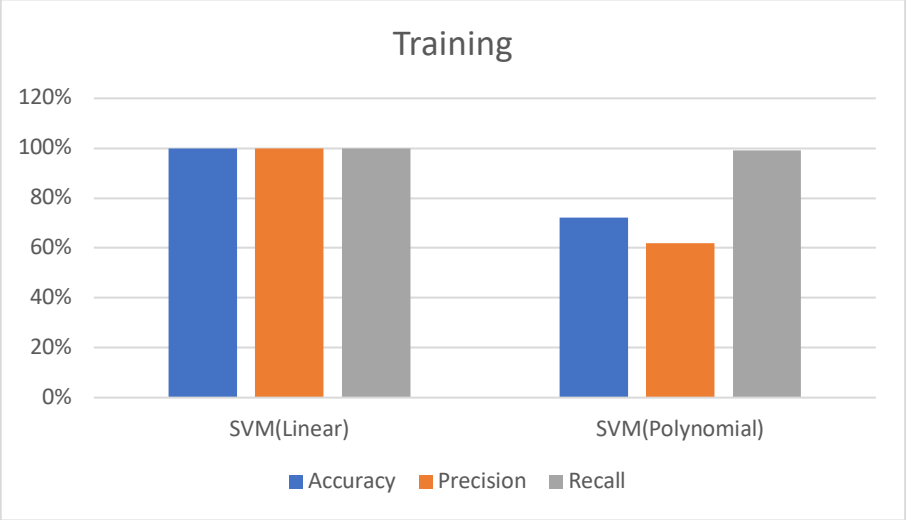
Student ID: 630615036
Name: Pheerawadee Chantakhat

Train and Test results without hyperparameter tuning

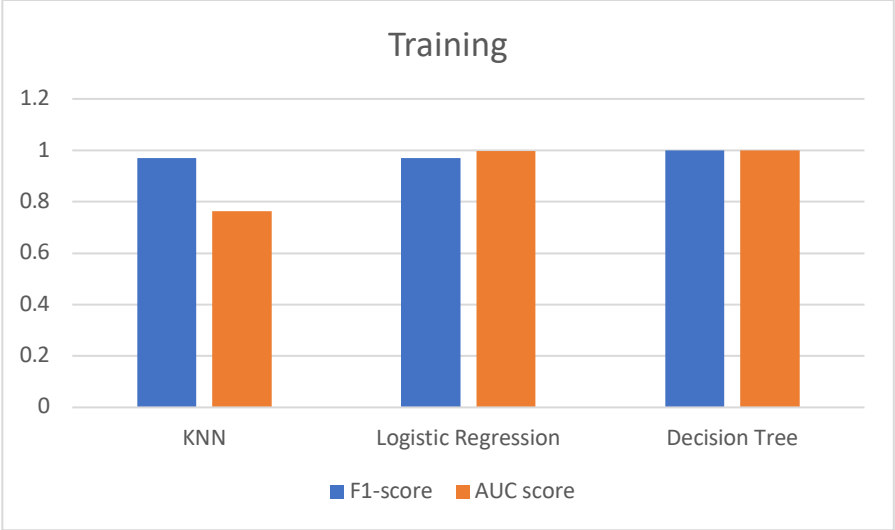
Training

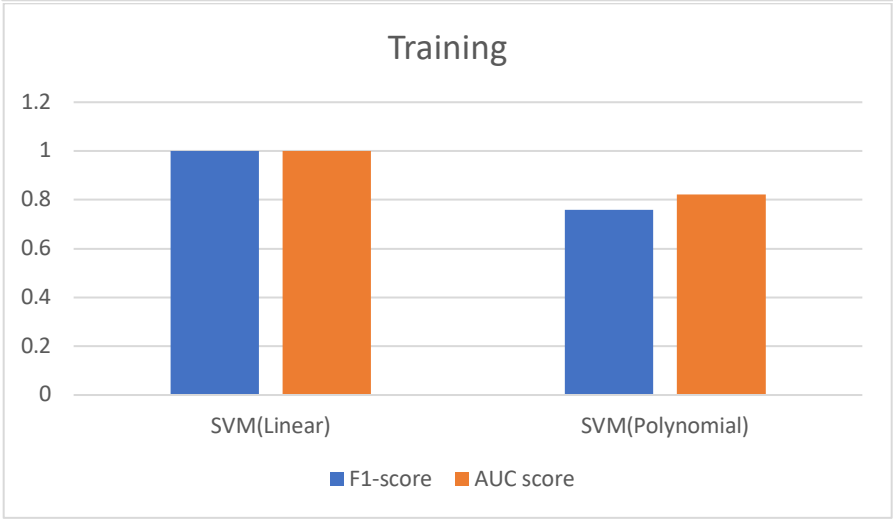
Model	Accuracy	Precision	Recall	F1-score	AUC score
KNN	79.29%	0.91	0.98	0.95	0.8379
Logistic Regression	94%	0.97	0.97	0.97	0.9960
SVM(Polynomial)	72.14%	0.62	0.99	0.76	0.8217
Decision Tree	100%	1.00	1.00	1.00	1.00
SVM(Linear)	100%	1.00	1.00	1.00	1.00





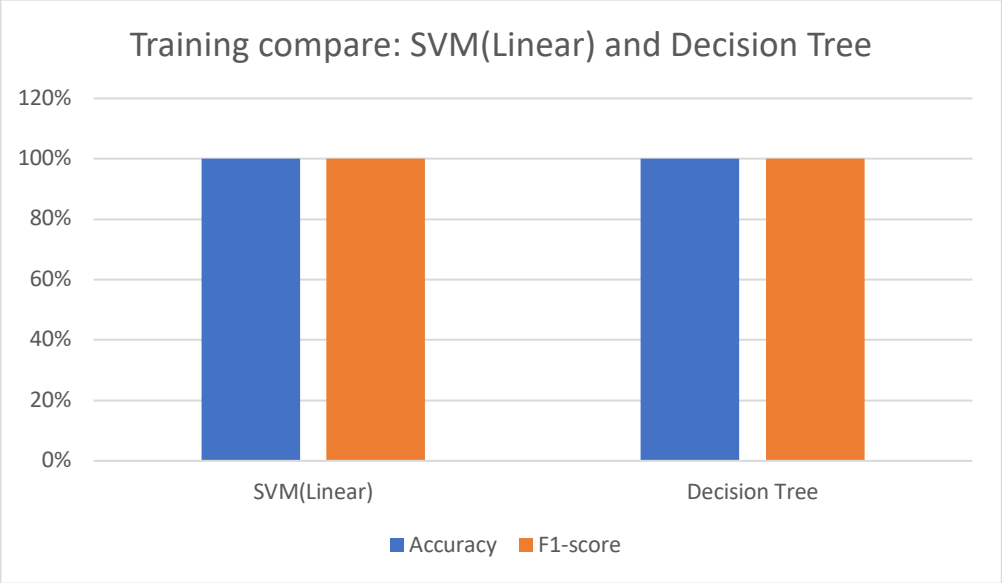
Student ID: 630615036
Name: Pheerawadee Chantakhat





Student ID: 630615036
Name: Pheerawadee Chantakhat

Compare F1-score and accuracy between SVM (linear) and decision tree:

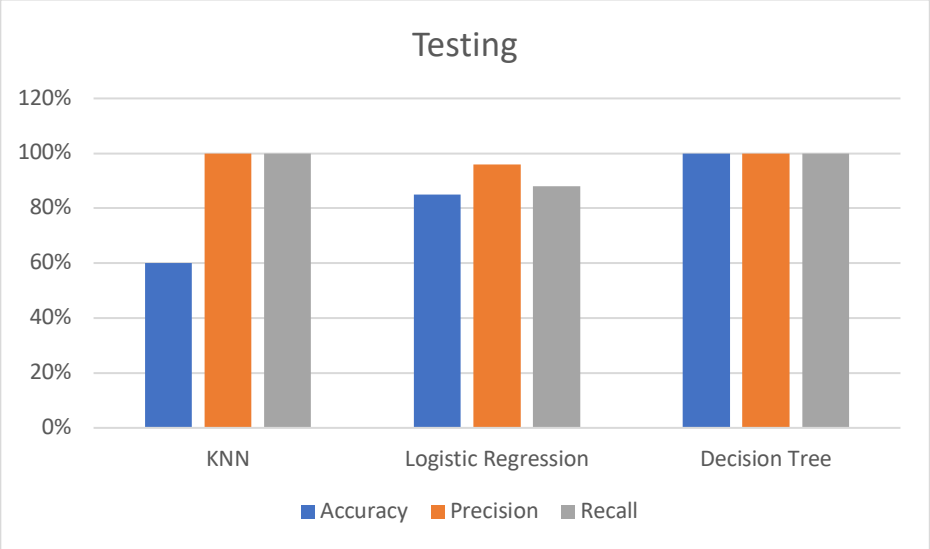


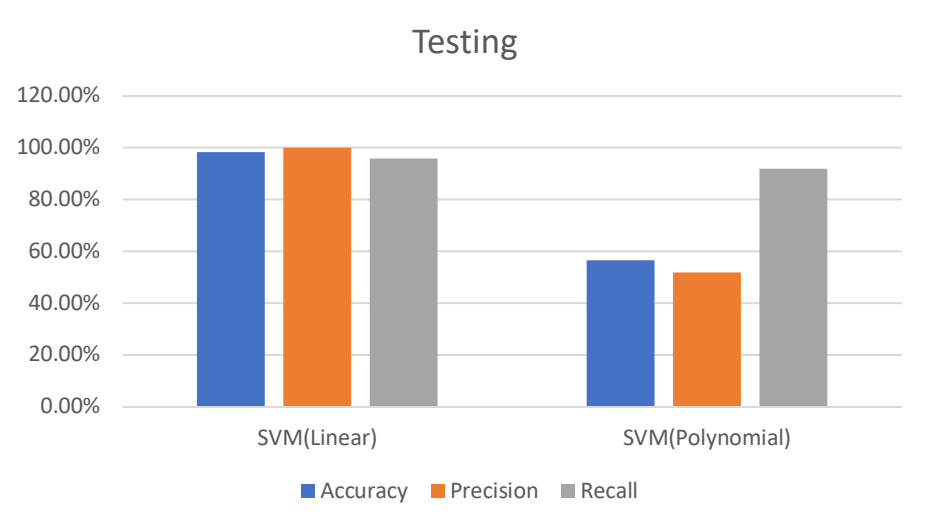
Best model for training is both the SVM (linear) and decision tree.

Student ID: 630615036
Name: Pheerawadee Chantakhat

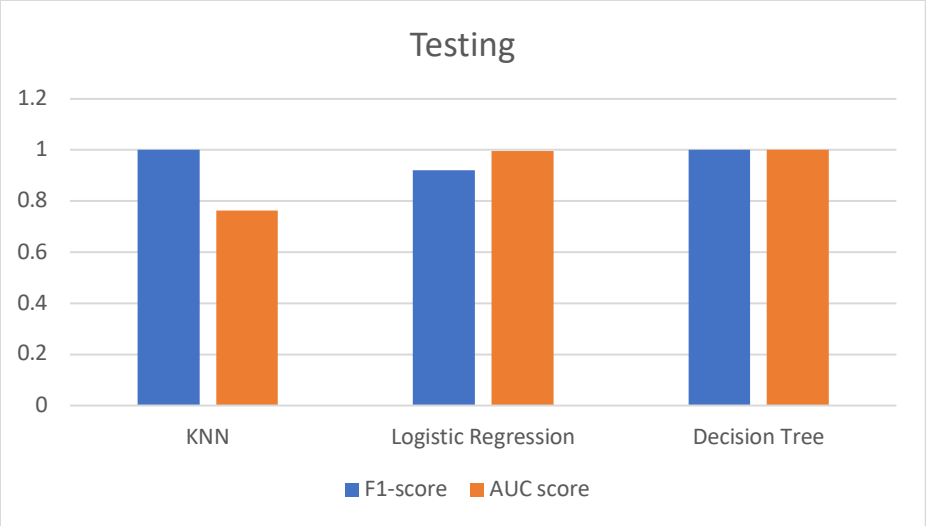
Testing

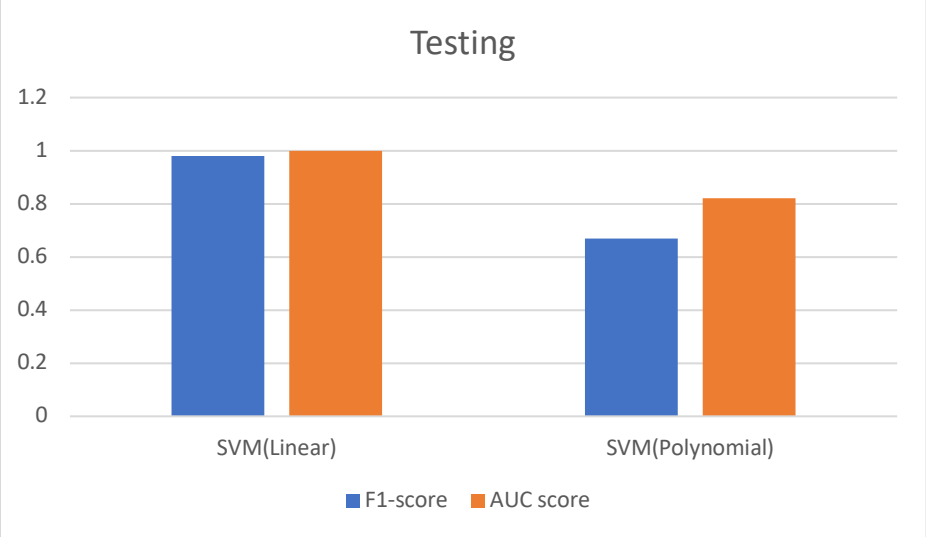
Model	Accuracy	Precision	Recall	F1-score	AUC score
KNN	63.33%	1.00	1.00	1.00	0.8379
Logistic Regression	85%	0.96	0.88	0.92	0.9894
SVM(Polynomial)	56.67%	0.52	0.92	0.67	0.8217
Decision Tree	100%	1.00	1.00	1.00	0.8217
SVM(Linear)	98.33%	1.00	0.96	0.98	1.00





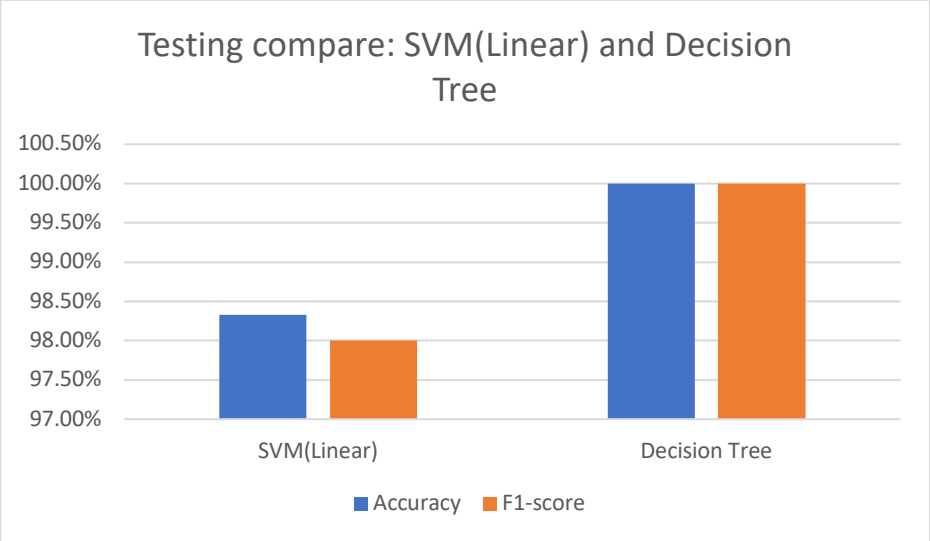
Student ID: 630615036
Name: Pheerawadee Chantakhat





Student ID: 630615036
Name: Pheerawadee Chantakhat

Compare F1-score and accuracy between SVM (linear) and Decision trees:



Best model for testing is Decision Tree

Student ID: 630615036
Name: Pheerawadee Chantakhat

Summary train and test without hyperparameter tuning

Note that some models have hyperparameter tuning, all are initial values no modification.

Default value of hyperparameter(for model that has)

KNN: n_neighbors = 5

SVM(Poly): degree = 23

Decision tree: max_leaf_nodes=20

Best model for training:

Model	Accuracy	Precision	Recall	F1-score	AUC score
Decision Tree	100%	1.00	1.00	1.00	1.00
SVM(Linear)	100%	1.00	1.00	1.00	1.00

Confusion matrix:

```
[[65  0  0  0  0]
 [ 0 16  0  0  0]
 [ 0  0 13  0  0]
 [ 0  0  0 10  0]
 [ 0  0  0  0 36]]
```

Best model for testing:

Model	Accuracy	Precision	Recall	F1-score	AUC score
Decision Tree	100%	1.00	1.00	1.00	0.8217

Confusion matrix:

```
[[26  0  0  0  0]
 [ 0  7  0  0  0]
 [ 0  0  3  0  0]
 [ 0  0  0  6  0]
 [ 0  0  0  0 18]]
```