Heart Attack Prediction

[Report]

SKY TEAM

1. Ashraf Abdulkhaliq Bassunoi Abdulaziz [1]
2. Ahmed Elsayed Mohamed Abo-elkher [1]
3. Ahmed Mostafa Ragab Mostafa [1]
4. Saif El-den Emad Abdulhamid [2]
5. Abdulsalam Othman Abdulsalam El-khedr [2]

Overview:

Heart attacks, also known as myocardial infarctions, are a leading cause of death and disability worldwide. Early prediction and prevention of heart attacks can save lives and improve patient outcomes. In this report, we present the results of our evaluation of a heart attack prediction model.

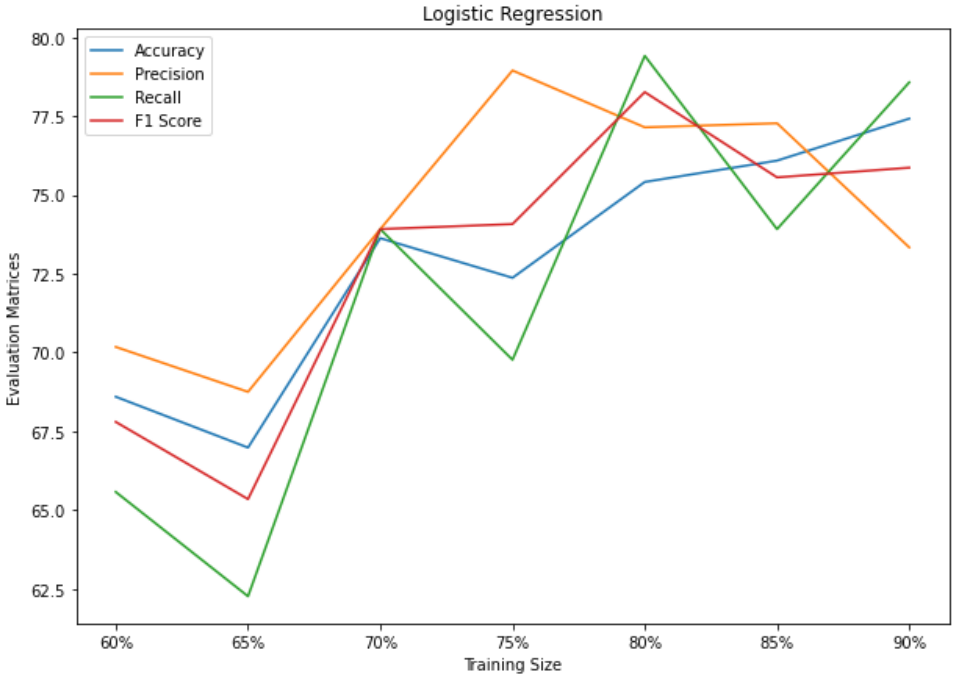
The purpose of this model is to predict the likelihood of a heart attack occurring within the next year in each patient, based on their medical history and other factors. We describe the methodology used to develop and evaluate the model and present the results in the form of various graphs and statistics.

Our goal is to assess the performance and utility of the model in a clinical setting, and to identify any areas for improvement.

Dataset Description:

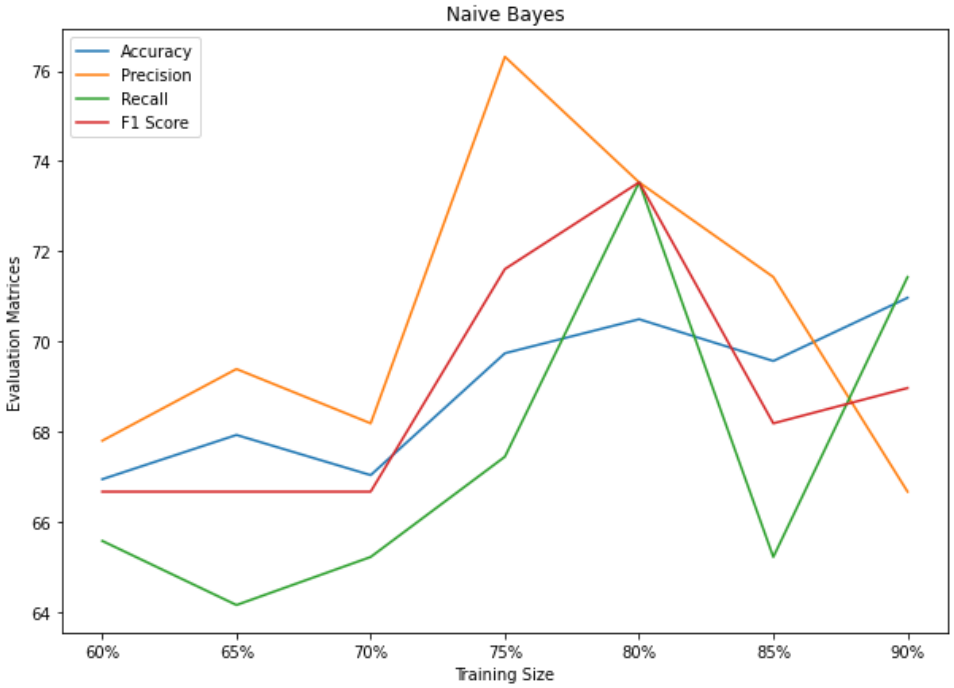
age	The number of years a person has lived.
sex	Male or Female.
CP	**Cerebral palsy (CP)** is a group of disorders that affect a person's ability to move and maintain balance and posture.
trestbps	**The person's resting blood pressure** (mm Hg on admission to the hospital).
chol	The person's **cholesterol** measurement in mg/dl.
FBS	**Fasting blood sugar (FBS)** measures blood glucose after you have not eaten for at least 8 hours.
restecg	**resting electrocardiographic** results (0 = normal; 1 = having ST-T; 2 = hypertrophy)
thalach	The person's maximum **heart rate** achieved.
exang	**Exercise induced angina** (1 = yes; 0 = no)
oldpeak	**ST depression** induced by exercise relative to rest ('ST' relates to positions on the ECG plot. See more here)
slope	**slope of the peak exercise** ST segment — 0: downsloping; 1: flat; 2: upsloping
ca	The number of major vessels (0–3)
thal	An inherited blood disorder that causes your body to have less hemoglobin than normal

Logistic Regression Model:



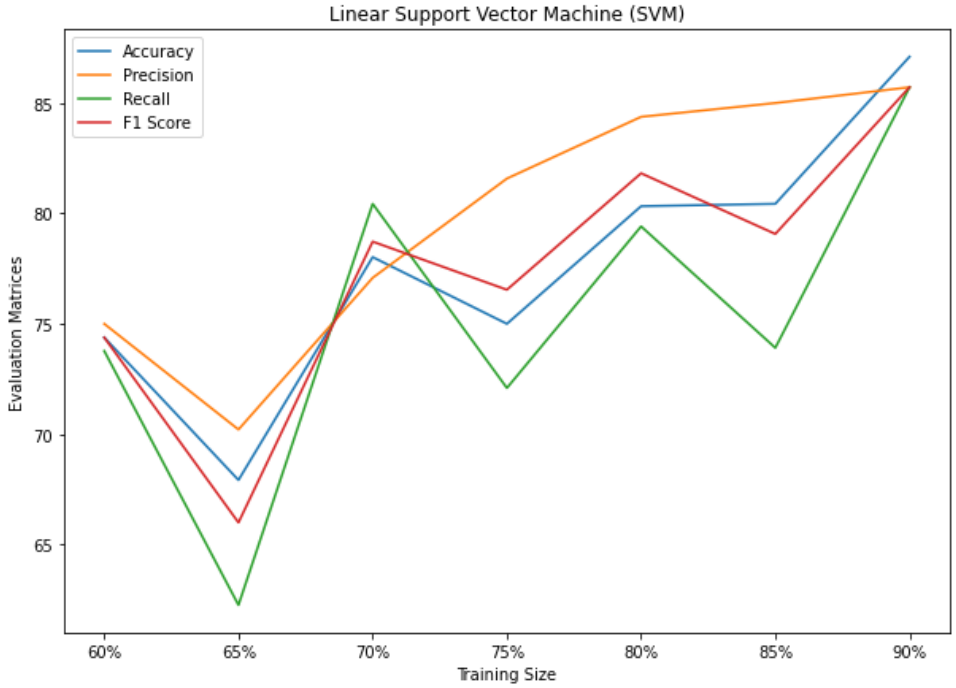
Dataset		Accuracy	Lose	Precision	Recall	F1 Score
Training	Testing					
60%	40%	68.59%	31.40%	70.17%	65.57%	67.79%
65%	35%	66.98%	33.01%	68.75%	62.26%	65.34%
70%	30%	73.98%	26.37%	73.91%	73.91%	73.91%
75%	25%	72.36%	27.63%	78.94%	69.76%	74.07%
80%	20%	75.40%	24.59%	77.14%	79.41%	78.26%
85%	15%	76.08%	23.91%	77.27%	73.91%	75.55%
90%	10%	77.41%	22.58%	73.33%	78.57%	75.86%

Naïve Bayes Model:



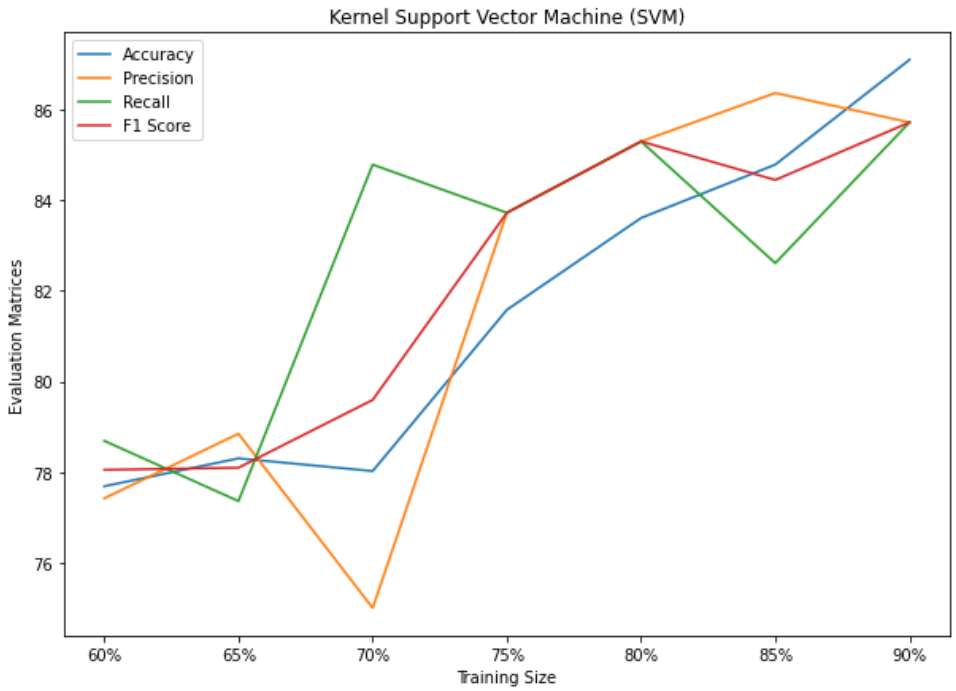
Dataset		Accuracy	Lose	Precision	Recall	F1 Score
Training	Testing					
60%	40%	66.94%	33.05%	67.79%	65.57%	66.66%
65%	35%	67.92%	32.07%	69.38%	64.15%	66.66%
70%	30%	67.03%	32.96%	68.18%	65.21%	66.66%
75%	25%	69.73%	30.26%	76.31%	67.44%	71.60%
80%	20%	70.49%	29.50%	73.52%	73.52%	73.52%
85%	15%	69.56%	30.43%	71.42%	65.21%	68.18%
90%	10%	70.96%	29.03%	66.66%	71.42%	68.96%

Linear SVM Model:



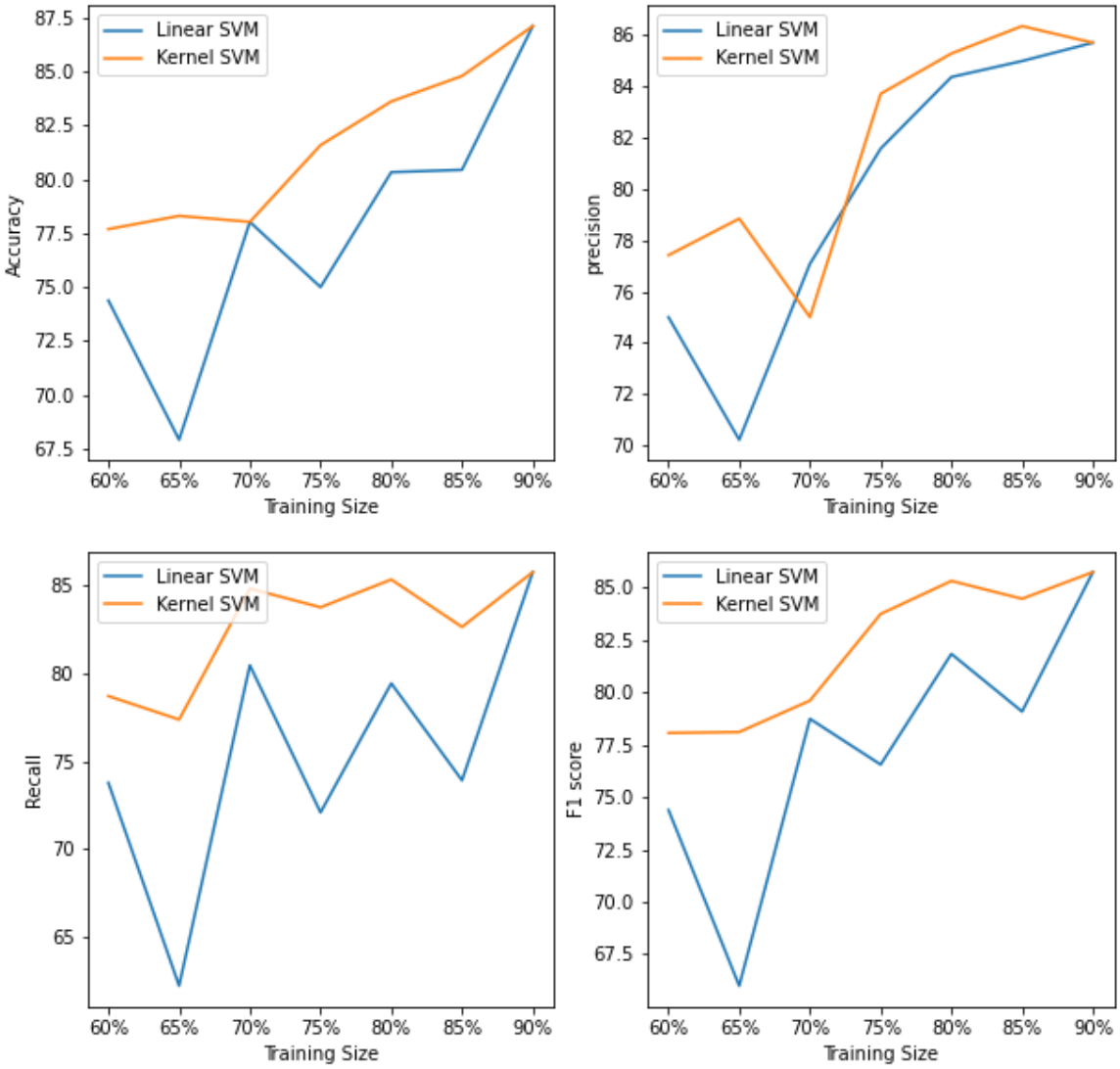
Dataset		Accuracy	Lose	Precision	Recall	F1 Score
Training	Testing					
60%	40%	74.38%	25.61%	75.00%	73.77%	74.38%
65%	35%	67.92%	32.07%	70.21%	62.26%	66.00%
70%	30%	78.02%	21.97%	77.08%	80.43%	78.72%
75%	25%	75.00%	25.00%	81.57%	72.09%	76.54%
80%	20%	80.32%	19.67%	84.37%	79.41%	81.81%
85%	15%	80.43%	19.56%	85.00%	73.91%	79.06%
90%	10%	87.09%	12.90%	85.71%	85.71%	85.71%

Kernel SVM Model:

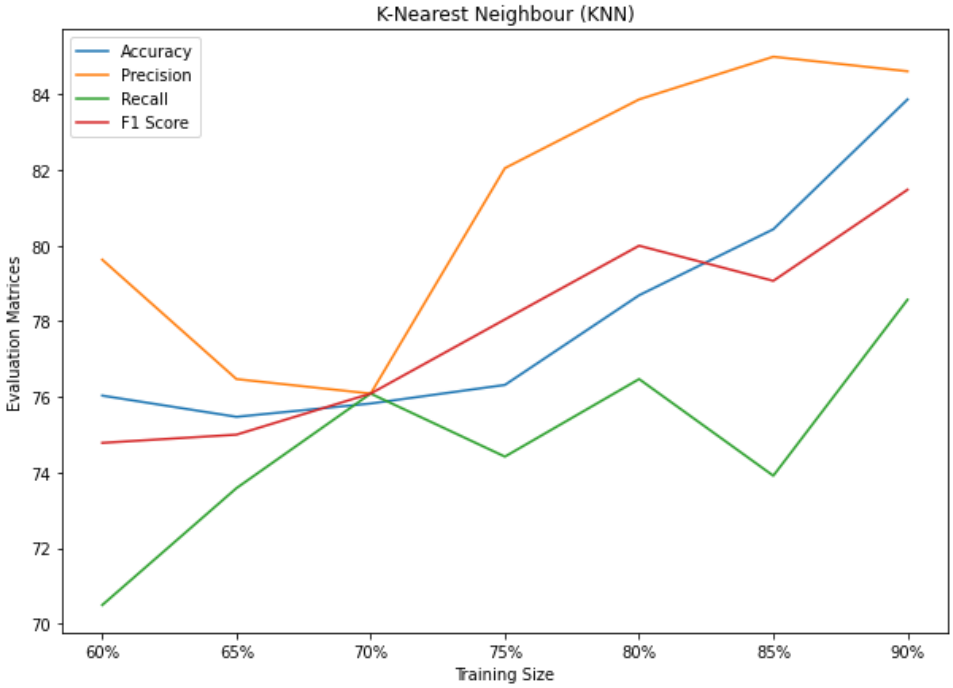


Dataset		Accuracy	Lose	Precision	Recall	F1 Score
Training	Testing					
60%	40%	77.68%	22.31%	77.41%	78.68%	78.04%
65%	35%	78.30%	21.69%	78.84%	77.35%	78.09%
70%	30%	78.02%	21.97%	75.00%	84.78%	79.59%
75%	25%	81.57%	18.42%	83.72%	83.72%	83.72%
80%	20%	83.60%	16.39%	85.29%	85.29%	85.29%
85%	15%	84.78%	15.21%	86.36%	82.60%	84.44%
90%	10%	87.09%	12.90%	85.71%	85.71%	85.71%

Linear SVM vs Kernel SVM:

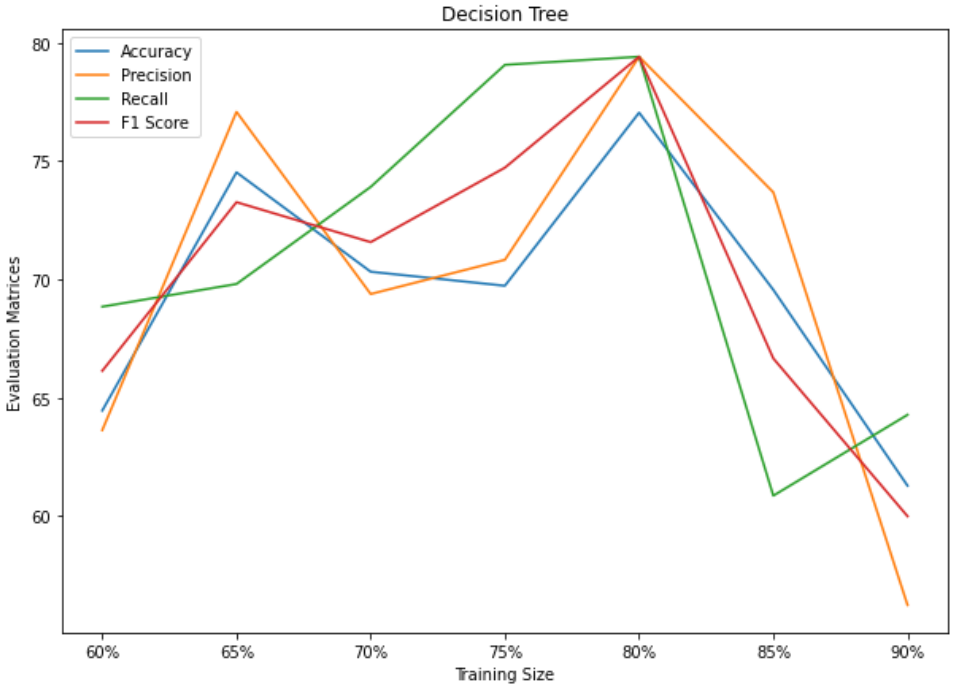


KNN Model:



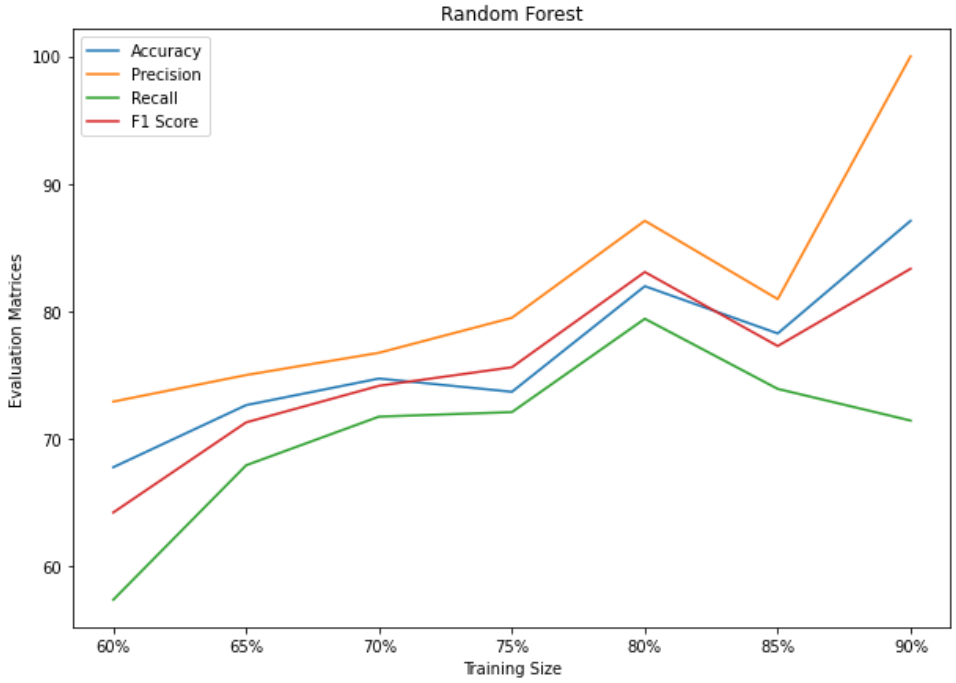
Dataset		Accuracy	Lose	Precision	Recall	F1 Score
Training	Testing					
60%	40%	76.03%	23.96%	79.62%	70.49%	74.78%
65%	35%	75.47%	24.52%	76.47%	73.58%	74.99%
70%	30%	75.82%	24.17%	76.08%	76.08%	76.08%
75%	25%	76.31%	23.68%	82.05%	74.41%	78.04%
80%	20%	78.68%	21.31%	83.87%	76.47%	80.00%
85%	15%	80.43%	19.56%	85.00%	73.91%	79.06%
90%	10%	83.87%	16.12%	84.61%	78.57%	81.48%

Decision Tree Model:



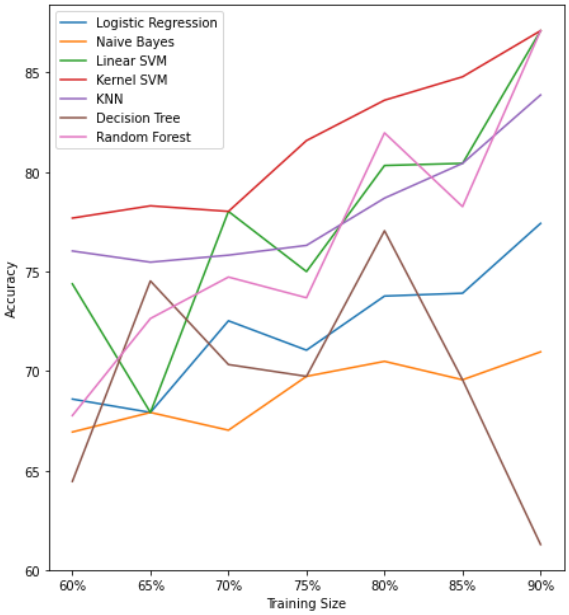
Dataset		Accuracy	Lose	Precision	Recall	F1 Score
Training	Testing					
60%	40%	64.46%	35.53%	63.63%	68.85%	66.14%
65%	35%	74.52%	25.47%	77.08%	69.81%	73.26%
70%	30%	70.32%	29.67%	69.38%	73.91%	71.57%
75%	25%	69.73%	30.26%	70.83%	79.06%	74.72%
80%	20%	77.04%	22.95%	79.41%	79.41%	79.41%
85%	15%	69.56%	30.43%	73.68%	60.86%	66.66%
90%	10%	61.29	38.70%	56.25%	64.28%	60.00%

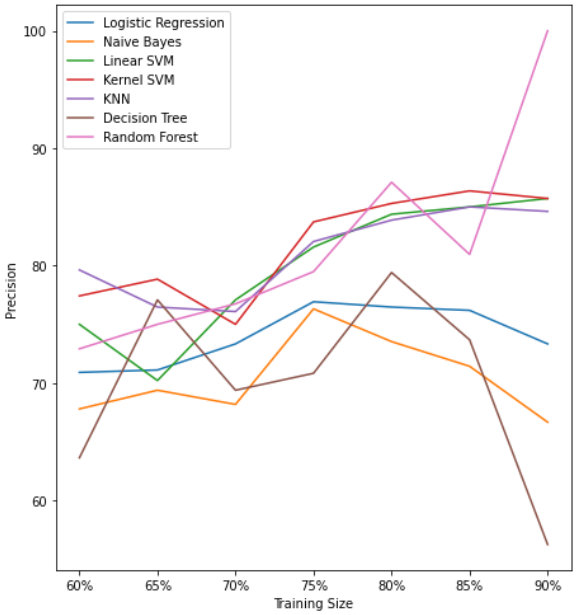
Random Forest Model:

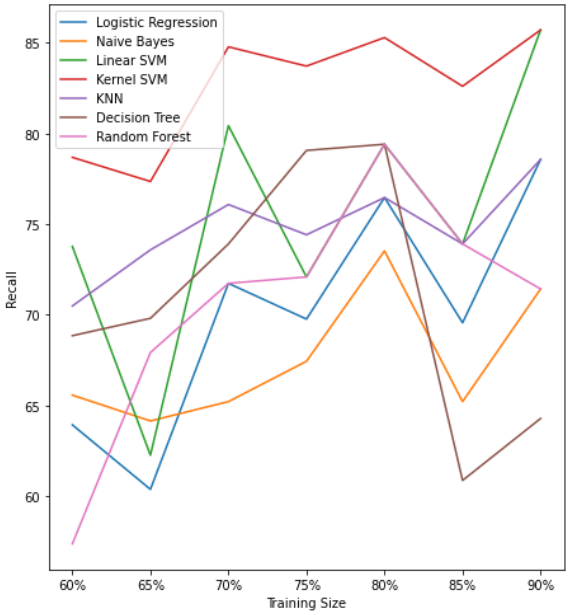


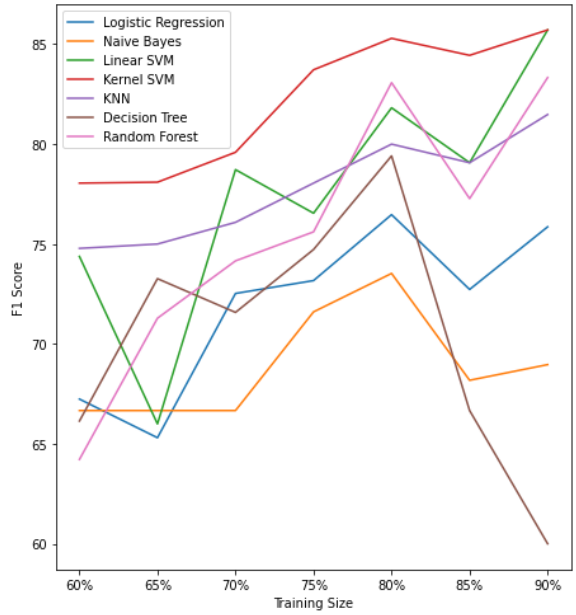
Dataset		Accuracy	Lose	Precision	Recall	F1 Score
Training	Testing					
60%	40%	67.76%	32.23%	72.91%	57.37%	64.22%
65%	35%	72.64%	27.35%	75.00%	67.92%	71.28%
70%	30%	74.72%	25.27%	76.74%	71.73%	74.15%
75%	25%	73.68%	26.31%	79.48%	72.09%	75.60%
80%	20%	81.96%	18.03%	87.09%	79.41%	83.07%
85%	15%	78.26%	21.73%	80.95%	73.91%	77.27%
90%	10%	87.09%	12.90%	100.00%	71.42%	83.33%

Models Evaluation:









Conclusion:

After training and testing the dataset on all these algorithms to choose which of them are better to our problem with low lose and high accuracy.

SVM with Gaussian kernel has did well with our problem, which has good evaluations with confusion matrices.