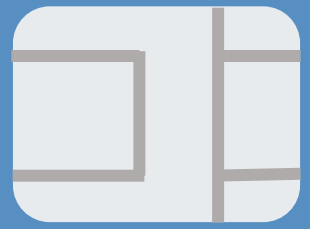
CREDIT CARD APPROVAL

1

Ditablan, Gil Jeremy P.
Reyes, John Ramil C.
Tajanlangit, Renzel E.

Group 1



BACKGROUND

Credit Card

Used by the card holders to purchase goods and services that accept cards as payment by borrowing funds from banks or other similar services.

Criteria:

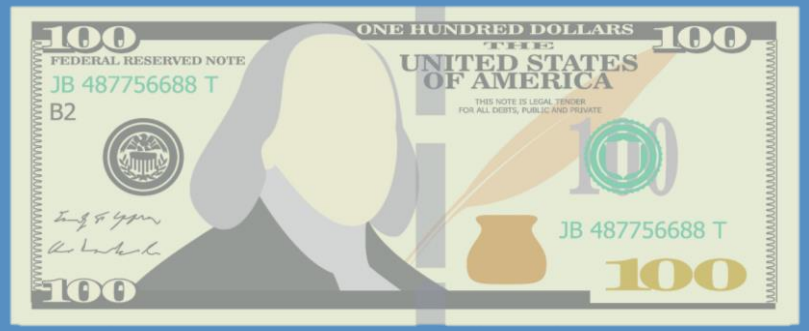
Credit Score

Income

Occupation

Location

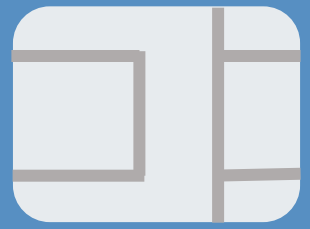
Current Situation



In the US payment ecosystem, the pandemic has caused credit card market stats to fall but eventually have seen growth within two years. Credit card growth is continually stabilizing specially when the consumer has embraced e-commerce.

This is also the case in our country where the credit card stats has fallen but eventually continues to rise within the past quarters because of economic revival and the rise in consumer spending.

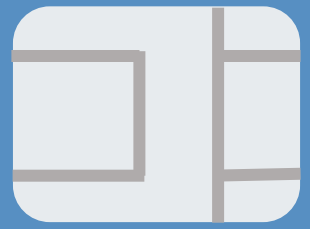
Group 1



PROBLEM DEFINITION

The main problem is the cost and time when dealing with multiple applications for a credit card that needs to be reviewed and multiple factors or variables that needs to be consider.

Group 1



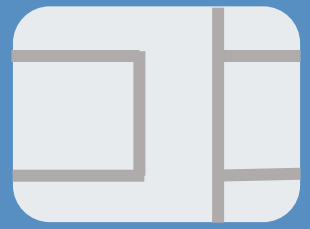
DATASET

The credit card approval dataset has **690** data entries.

Features	Description	Variable Type	Value
Debt	Amount of applicant's current debt.	Independent	Numerical range
BankCustomer	If the applicant has a bank account or not.	Independent	1 (Yes) or 0 (No)
YearsEmployed	Number of years the applicant has been employed.	Independent	Numerical range
Employed	Applicant's current employment status.	Independent	1 (Yes) or 0 (No)
CreditScore	A numerical score that depicts the applicant's worthiness given by the bank.	Independent	Numerical range
Income	Applicant's Monthly income.	Independent	Numerical range
Approved	Whether the credit card application is approved or rejected.	Dependent	1 (Yes) or 0 (No)

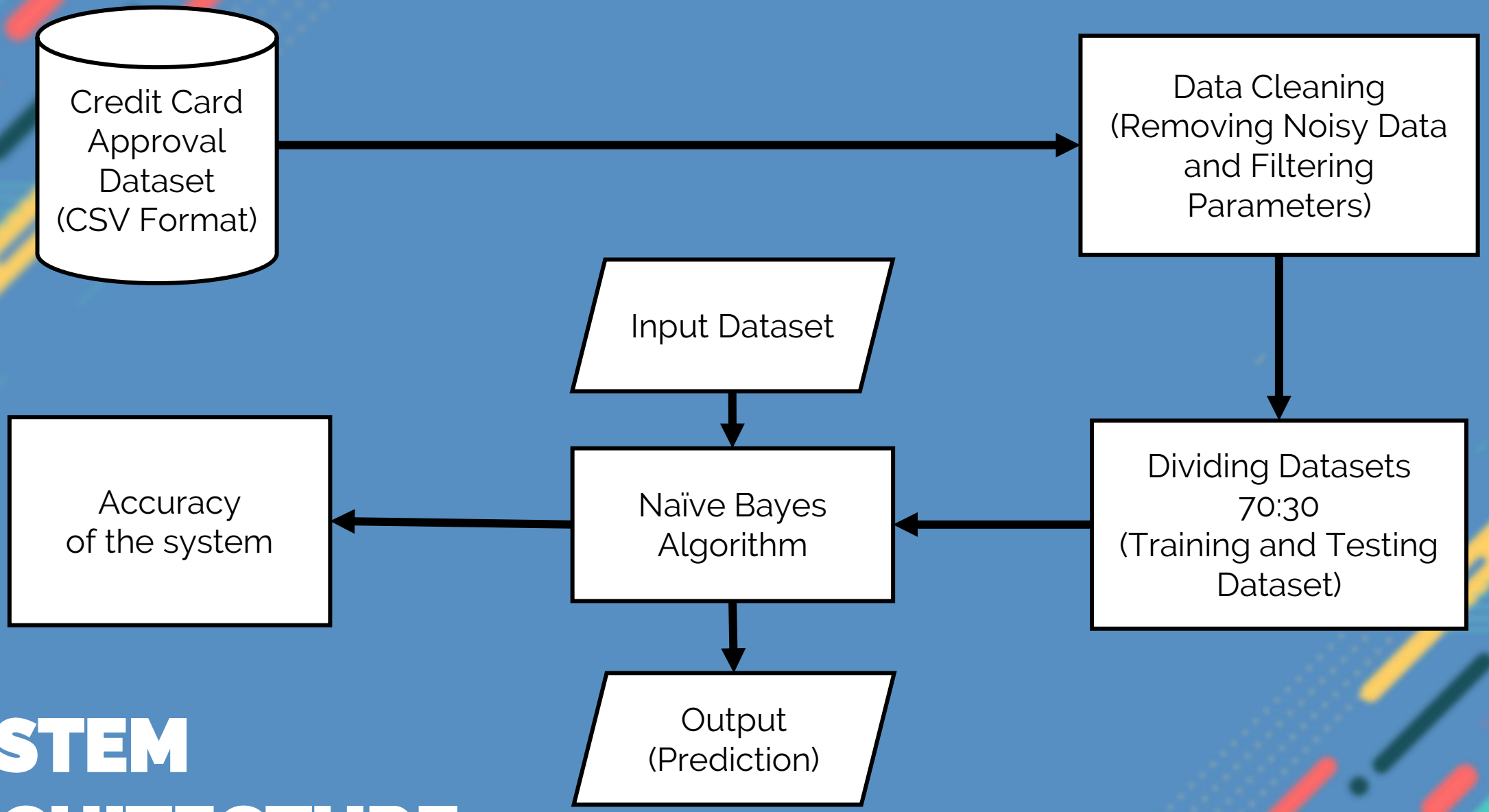
Gender	Age	Debt	BankCustomer	YearsEmployed	Employed	CreditScore	Income	Approved
1	30.83	0	1	1.25	1	1	0	1
0	58.67	4.46	1	3.04	1	6	560	1
0	24.5	0.5	1	1.5	0	0	824	1
1	27.83	1.54	1	3.75	1	5	3	1
1	20.17	5.625	1	1.71	0	0	0	1
1	32.08	4	1	2.5	0	0	0	1
1	33.17	1.04	1	6.5	0	0	31285	1
0	22.92	11.585	1	0.04	0	0	1349	1
1	54.42	0.5	0	3.96	0	0	314	1
1	42.5	4.915	0	3.165	0	0	1442	1
1	22.08	0.83	1	2.165	0	0	0	1
1	29.92	1.835	1	4.335	0	0	200	1
0	38.25	6	1	1	0	0	0	1
1	48.08	6.04	1	0.04	0	0	2690	1
0	45.83	10.5	1	5	1	7	0	1
1	36.67	4.415	0	0.25	1	10	0	1
1	28.25	0.875	1	0.96	1	3	0	1
0	23.25	5.875	1	3.17	1	10	245	1
1	21.83	0.25	1	0.665	0	0	0	1
0	19.17	8.585	1	0.75	1	7	0	1
1	25	11.25	1	2.5	1	17	1208	1
1	23.25	1	1	0.835	0	0	0	1
0	47.75	8	1	7.875	1	6	1260	1
0	27.42	14.5	1	3.085	1	1	11	1
0	41.17	6.5	1	0.5	1	3	0	1
0	15.83	0.585	1	1.5	1	2	0	1

Group 1

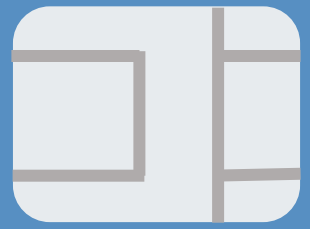


SYSTEM ARCHITECTURE

SYSTEM ARCHITECTURE



Group 1



MACHINE LEARNING ALGORITHM

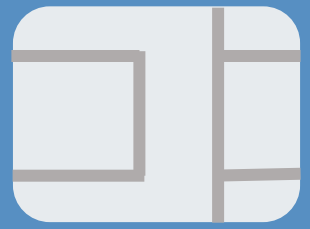
NAÏVE BAYES ALGORITHM

Predict the approval of credit card applications

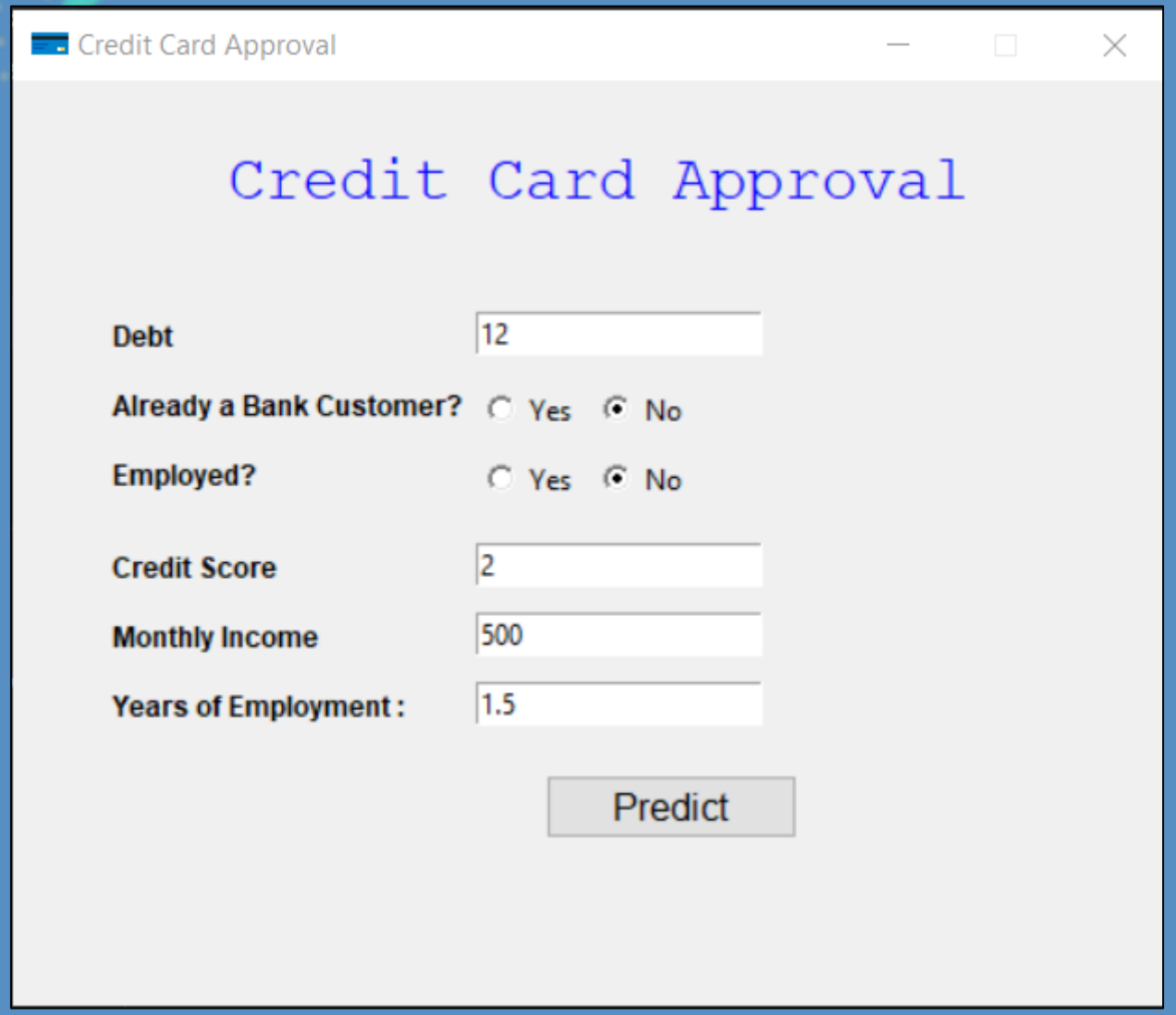
Classify different credit card applications according to their feature value

The feature variables that will be fitted into the Bayesian model are Debt, BankCustomer, YearsEmployed, Employed, CreditScore, and Income

Group 1



PYTHON PROGRAMMING DISCUSSION



Graphical User Interface

```
model.fit(features_train, target_train)
```

Transformation of data from Qualitative to Quantitative

Splitting for Training and Testing

	Training	Testing
Percentage	70%	30%
Quantity	483	207

```
features_train, features_test, target_train, target_test = train_test_split(credit_card[features],
credit_card[target], test_size = 0.30,
    random_state = 20)
```

Modeling based on the Training Set

The developer filtered the parameters of the training datasets into 7 parameters:
6 - independent Variable
1 – Dependent Variable

The developer used

Gaussian Naïve Bayes

because the data sets are numerical and Categorical.

The Developers will feed the dataset into the model to train and test the said model

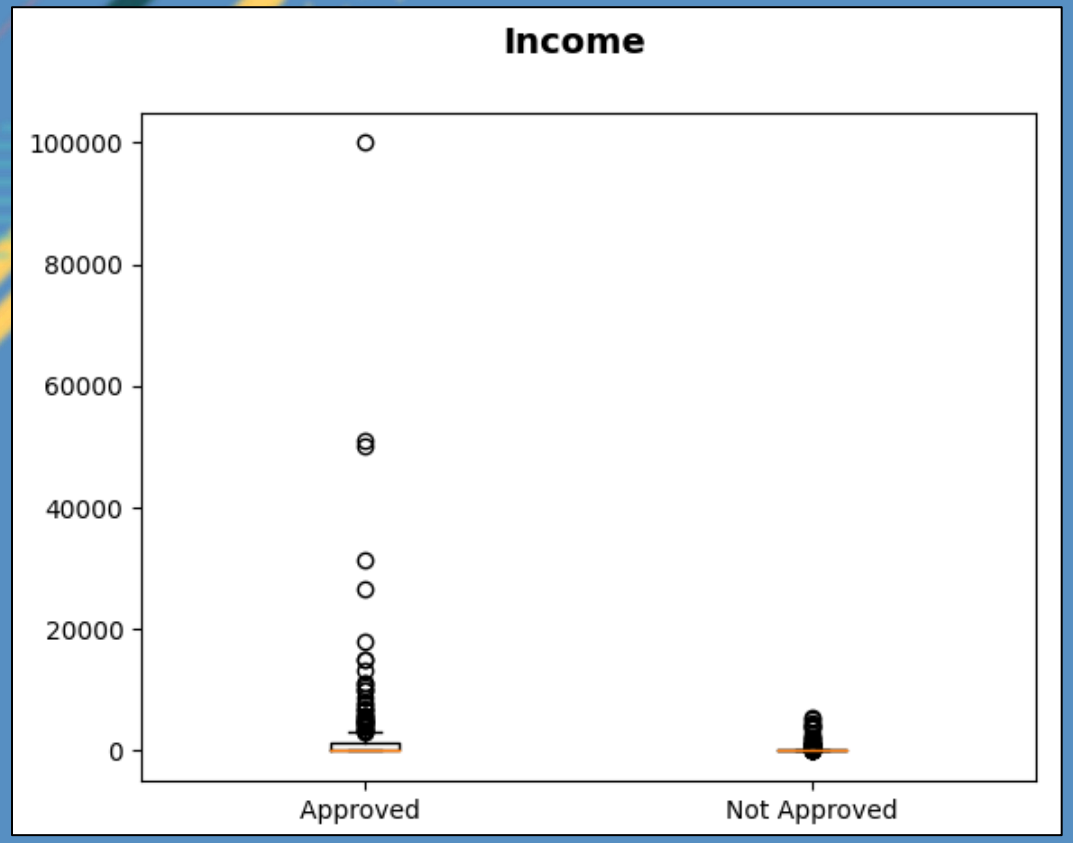
Accuracy of the Model

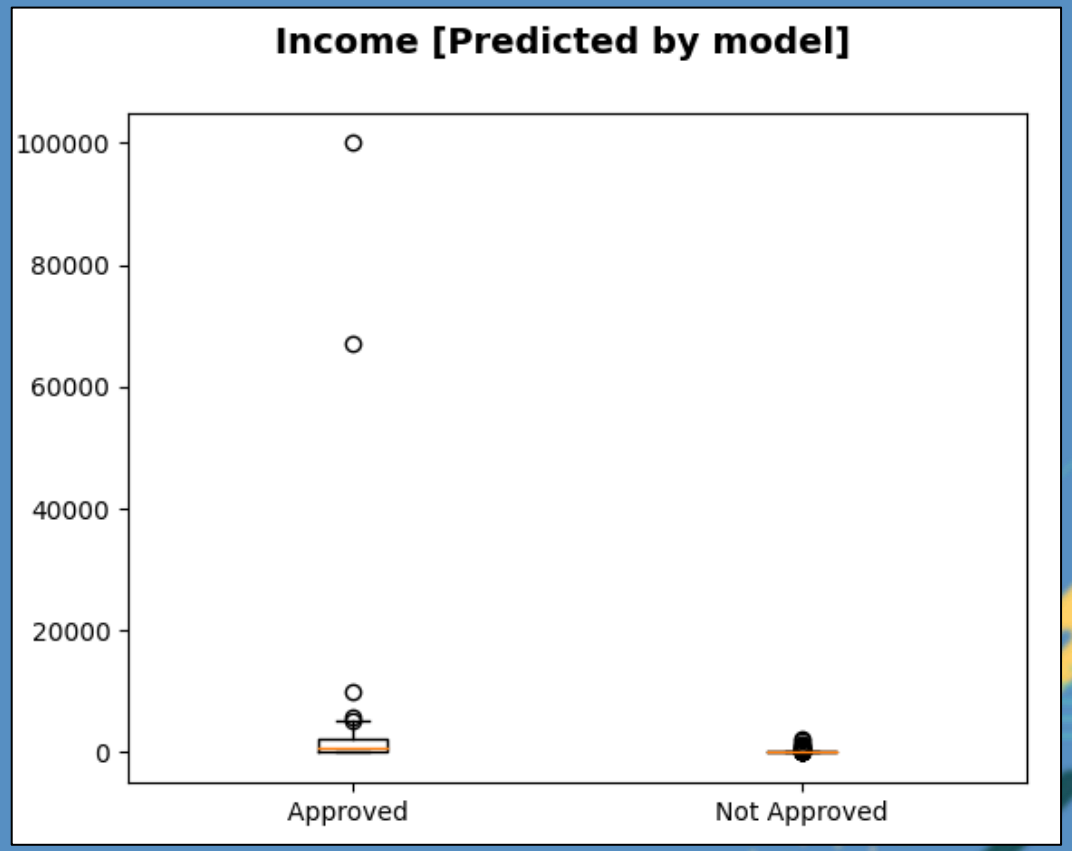
```
Model Accuracy =  85.02415458937197 %
[[114   8]
 [ 23  62]]
```

The Naive Bayes model trained on our dataset got **85.02%** of accuracy in predicting approval state of the credit card applications based on the testing dataset.

Simulation with Test Data

Visualization of the Results

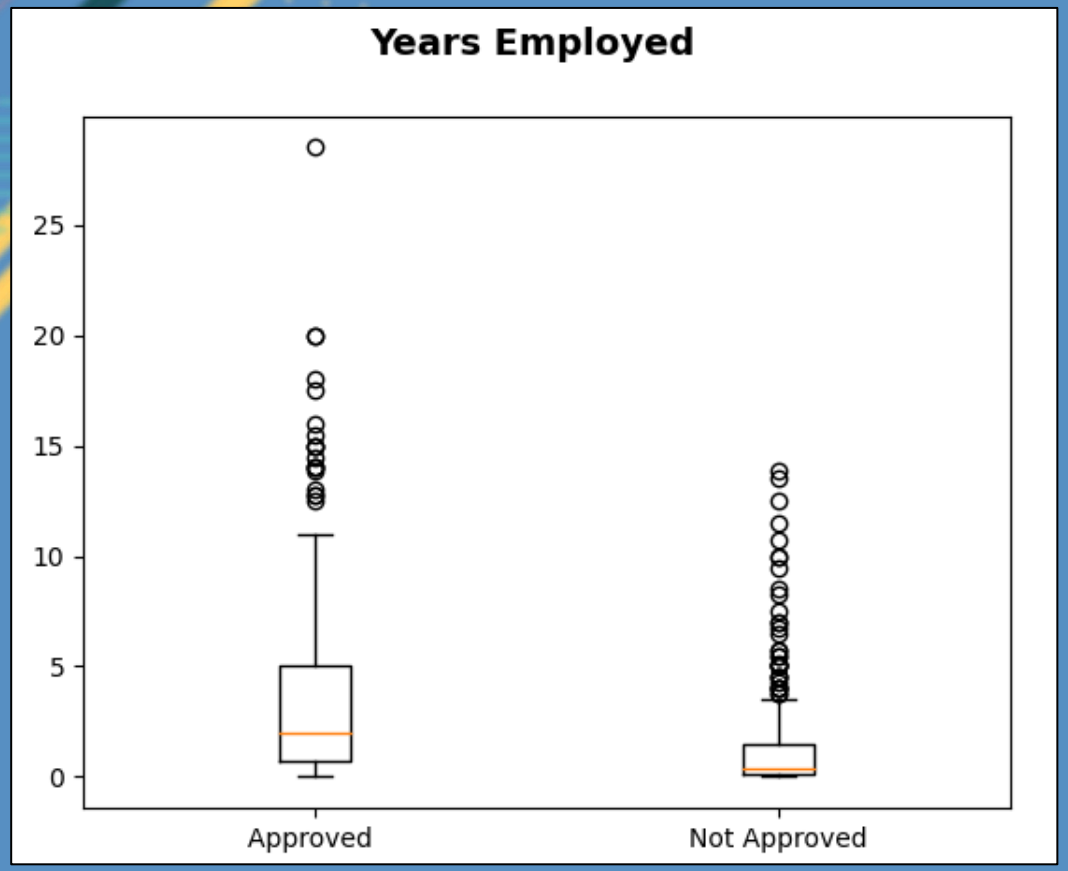


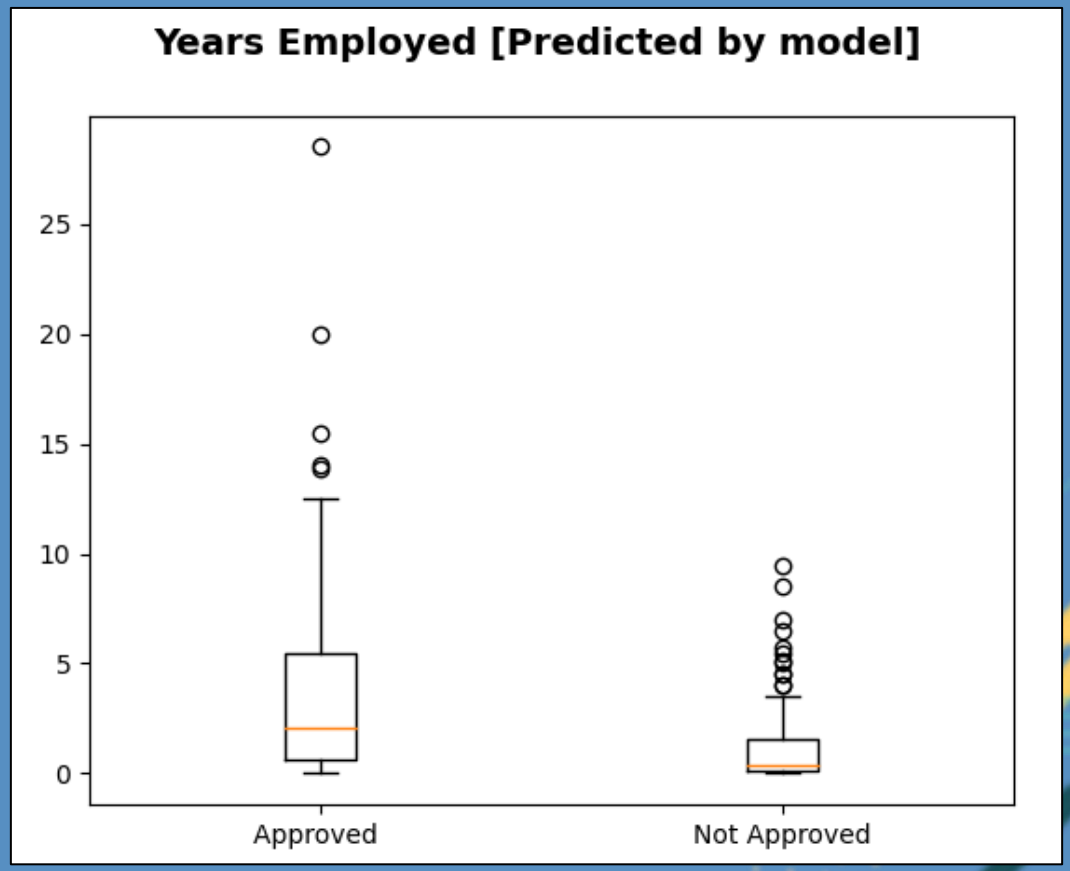


Box Plot (Training Dataset)

Box Plot (Testing Dataset)

Visualization of the Results

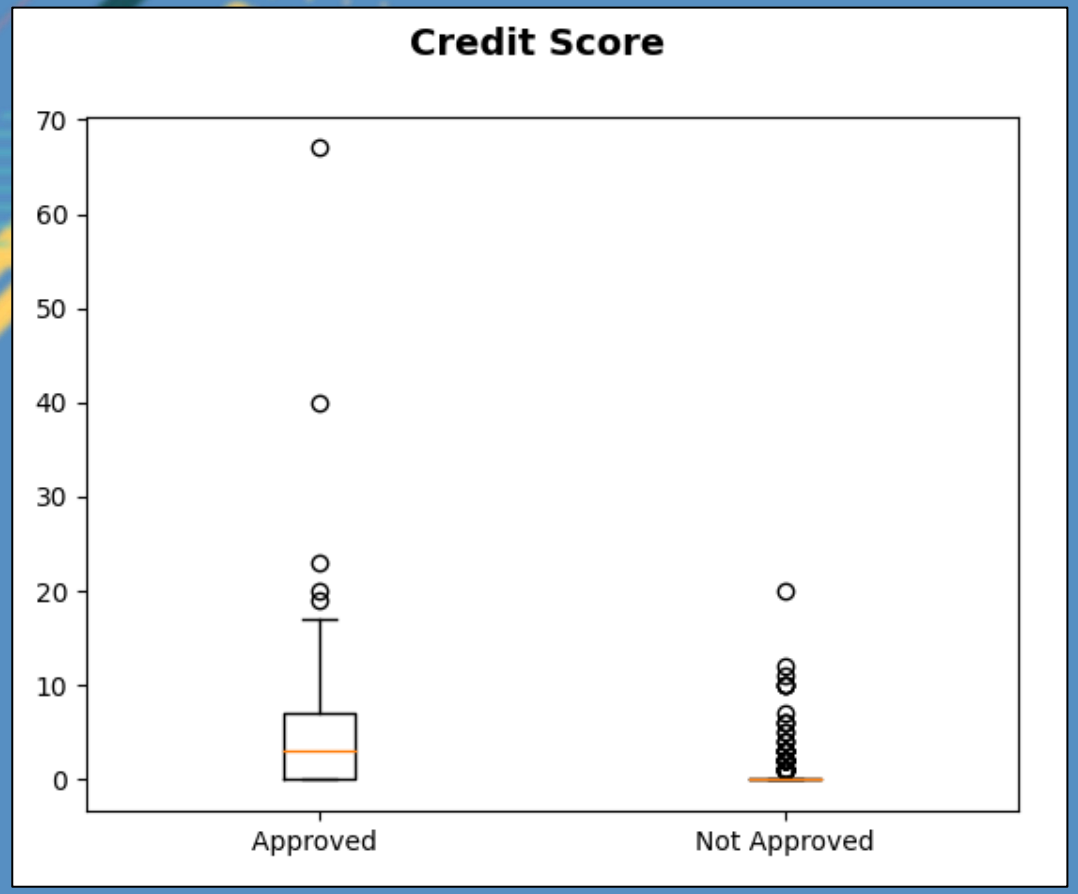


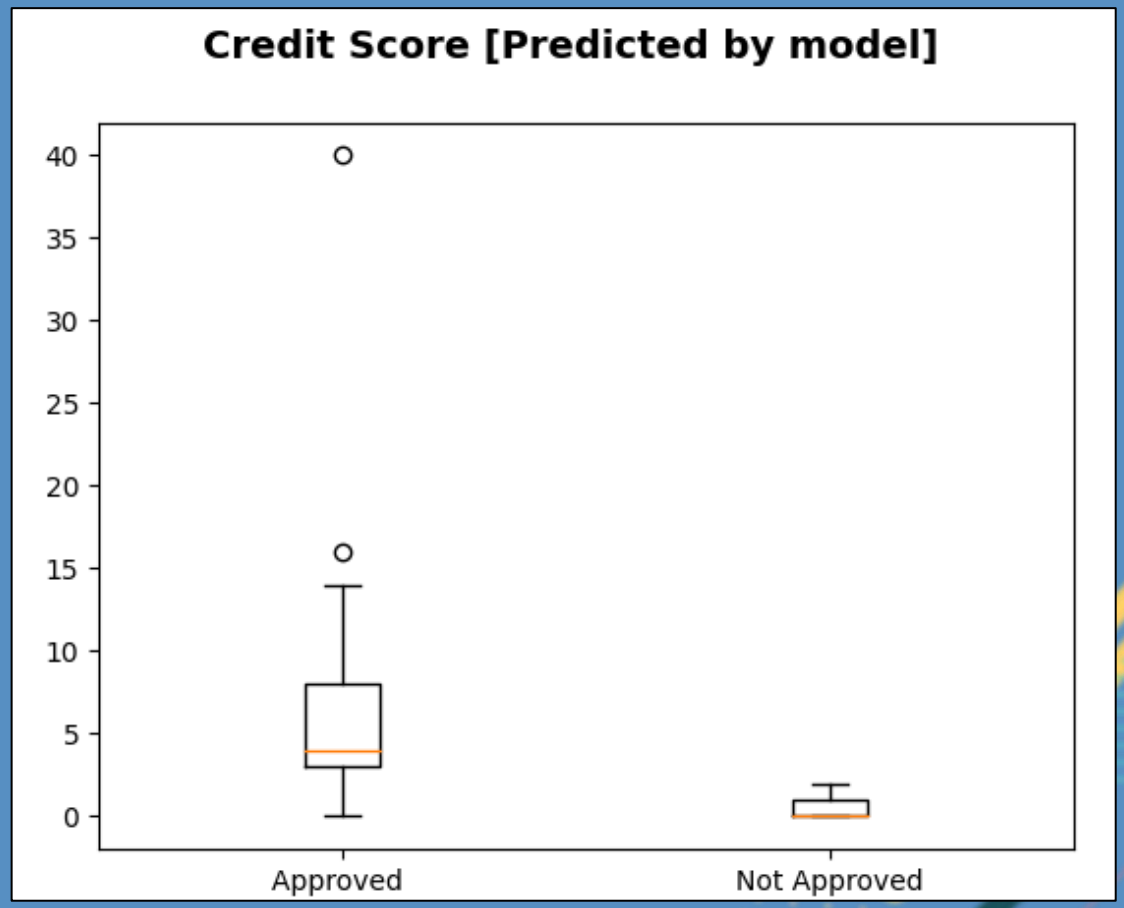


Box Plot (Training Dataset)

Box Plot (Testing Dataset)

Visualization of the Results

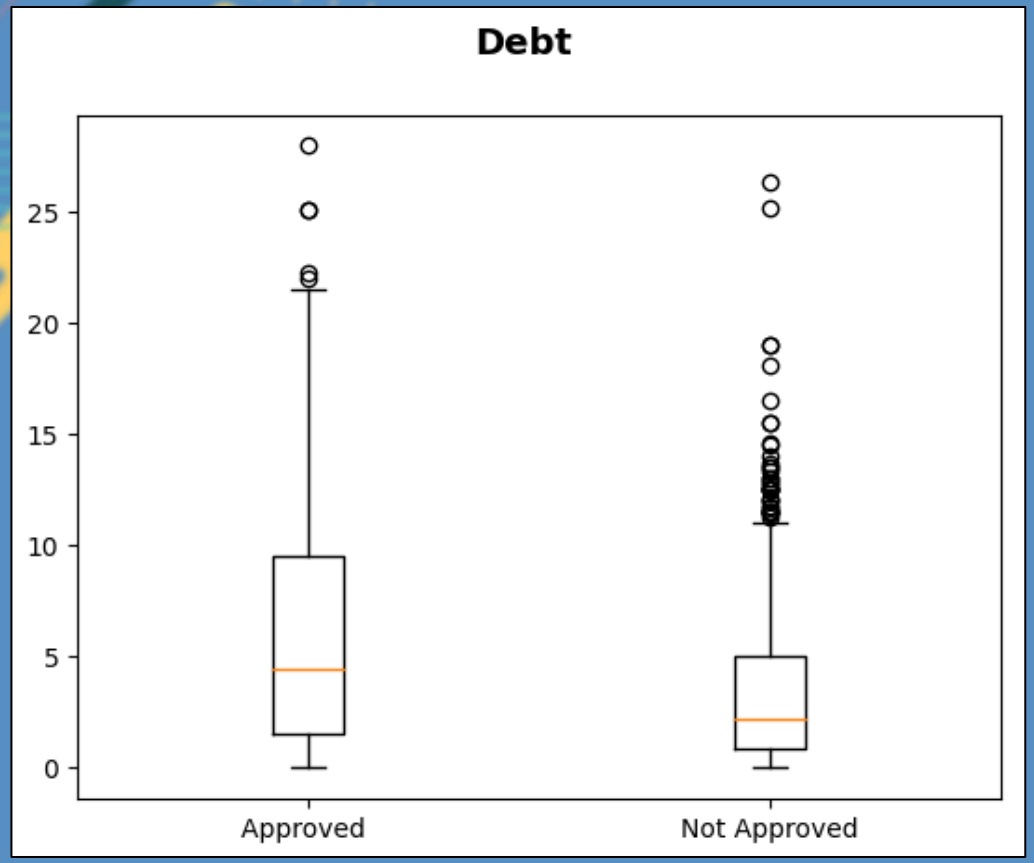


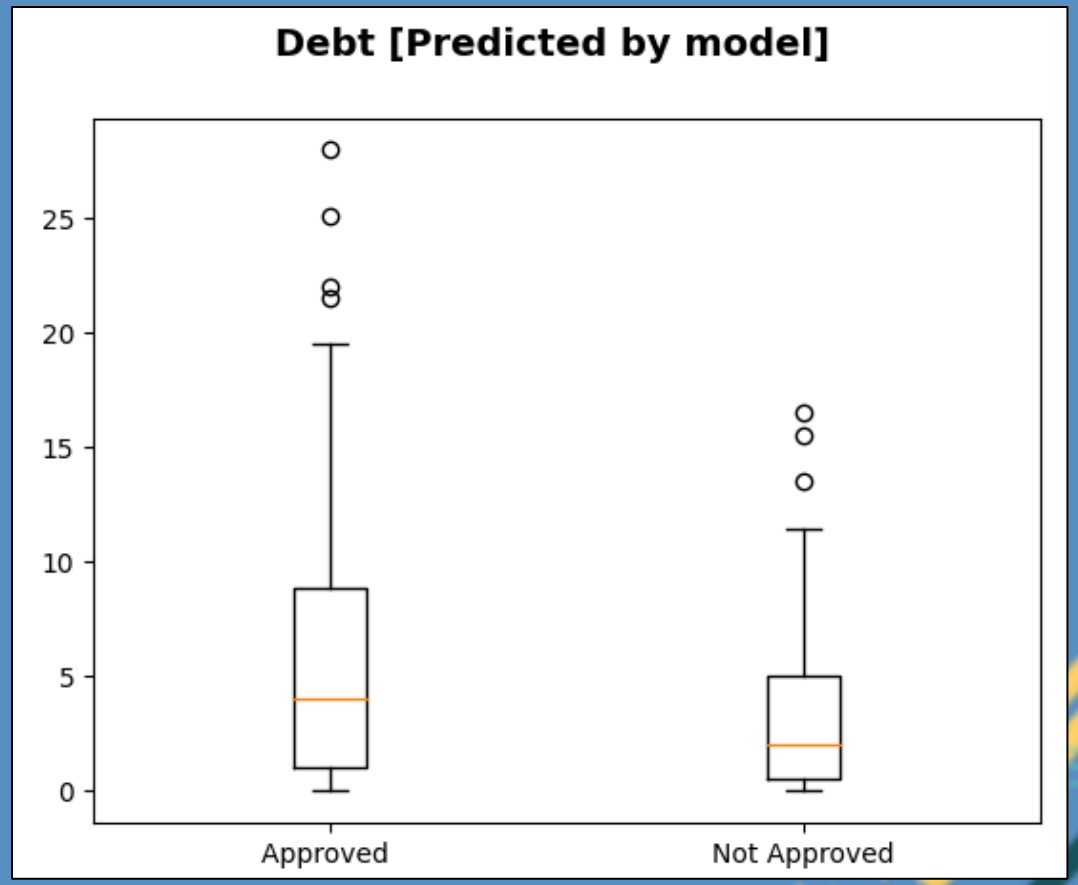


Box Plot (Training Dataset)

Box Plot (Testing Dataset)

Visualization of the Results

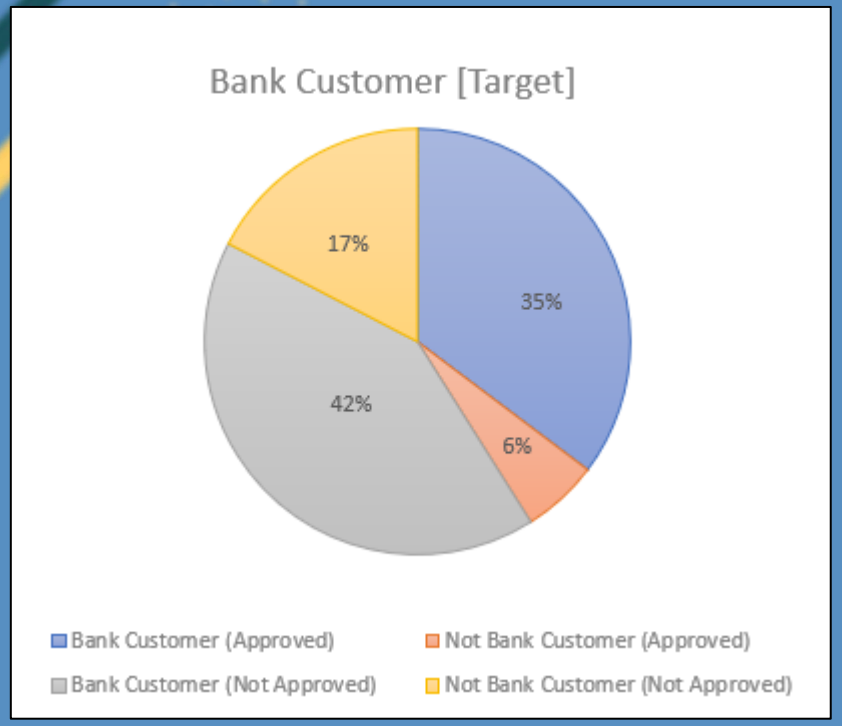


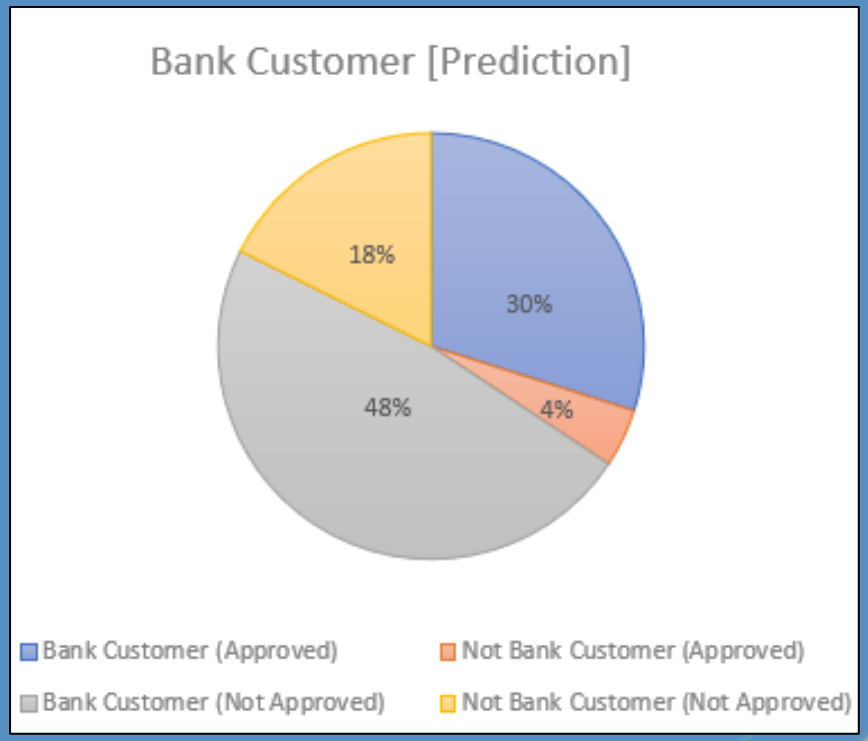


Box Plot (Training Dataset)

Box Plot (Testing Dataset)

Visualization of the Results



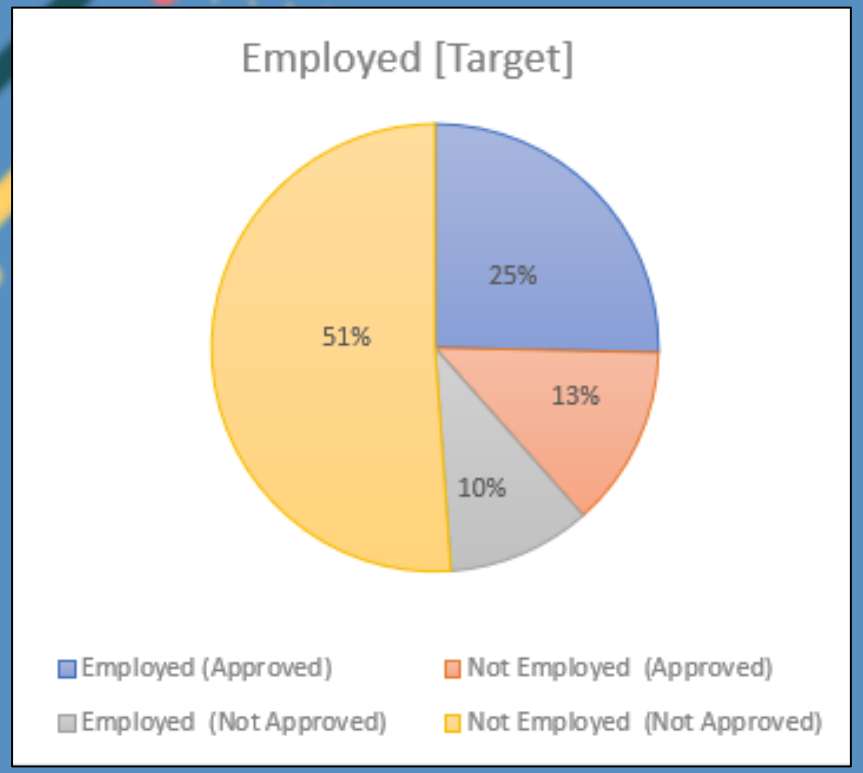


Bank Customer	Approved	Not Approved
1 (has a bank account)	73	86
0 (has no bank account)	12	36

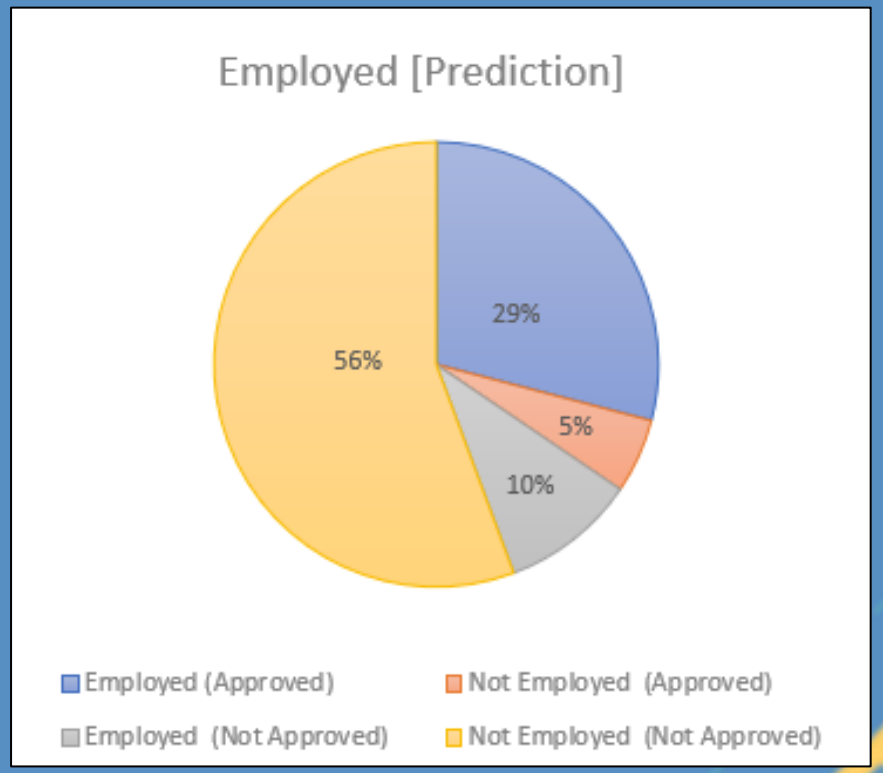
Bank Customer	Approved	Not Approved
1 (has a bank account)	61	98
0 (has no bank account)	9	39

Bank Customers

Visualization of the Results



Employed	Approved	Not Approved
1 (Is employed)	56	23
0 (Is not employed)	29	99



Bank Customer	Approved	Not Approved
1 (Is a bank customer)	59	20
0 (Is not a bank customer)	11	117

Employed

Added Discussion

```
Model Accuracy =  74.39613526570048 %
[[108   7]
 [ 46  46]]
```

Accuracy when Age and Gender have added

Added Discussion

```
Model Accuracy =  73.429951690082126 %
[[108   7]
 [ 48  44]]
```

Accuracy when Gender was added

Added Discussion

```
Model Accuracy =  74.39613526570048 %
[[108   7]
 [ 46  46]]
```

Accuracy when Age was added

Snippet of Codes

```
#import the necessary libraries
import pandas as pd
from sklearn.naive_bayes import GaussianNB
from sklearn.preprocessing import LabelEncoder
from sklearn.model_selection import train_test_split
from sklearn.metrics import accuracy_score, classification_report
from sklearn import metrics
```

Important libraries

Snippet of Codes

```
#read csv
credit_card = pd.read_csv("original dataset clean.csv")
print(credit_card)

#Store the independent variable to features
features = ["Debt", "BankCustomer", "YearsEmployed", "Employed", "CreditScore", "Income"]
#Store the dependent variable to target
target = ['Approved']
```

Read CSV, Features and Target

Snippet of Codes

```
#Splits the training and testing dataset with random = 20 and test size 30%
features_train, features_test, target_train, target_test = train_test_split(credit_card[features],
credit_card[target], test_size = 0.30,
   random_state = 20)

# Displaying the split datasets
print('\tTraining Features\n ',features_train)  #3 Print all of these
print('\tTesting Features\n ',features_test)
print('\tTraining Target\n ',target_train)
print('\tTesting Target\n ',target_test)
```

Data Splitting

Snippet of Codes

```
#create model
model = GaussianNB()

model.fit(features_train, target_train)

#prediction
pred = model.predict(features_test)
```

Model and Prediction

Snippet of Codes

```
#get accuracy
accuracy = accuracy_score(target_test, pred)

#print accuracy
print("Normal Accuracy",accuracy)
print("\nModel Accuracy = ",accuracy*100,"%")
print(metrics.confusion_matrix(target_test, pred))
print(metrics.classification_report(target_test, pred))
```

Getting the accuracy