Imarticus Learning

Project on:

Credit Risk Analysis

Submitted from DSP Batch 27 by,
Nidhi Patel
Sanyukta Shenoy
Wasim Akram

Credit Risk Analysis

What is Credit Risk Analysis?

Why Credit Risk Analysis is used?

Where Credit Risk Analysis is used?

Data Analysis

- Data Analysis refers to extracting knowledge and insight from large amount of Data.

Process of extracting Gold :

Searching Gold Mine

Removing Soil and Rock
Or
Extracting Gold

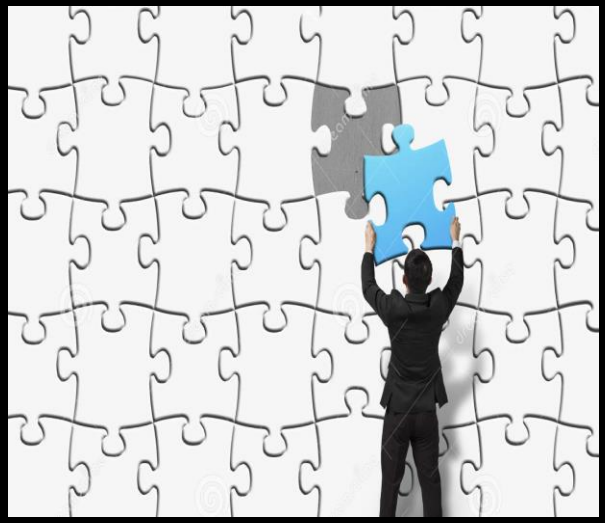
Required Gold

Our Data Analysis

Process of Extracting Data for Credit Risk Analyse:

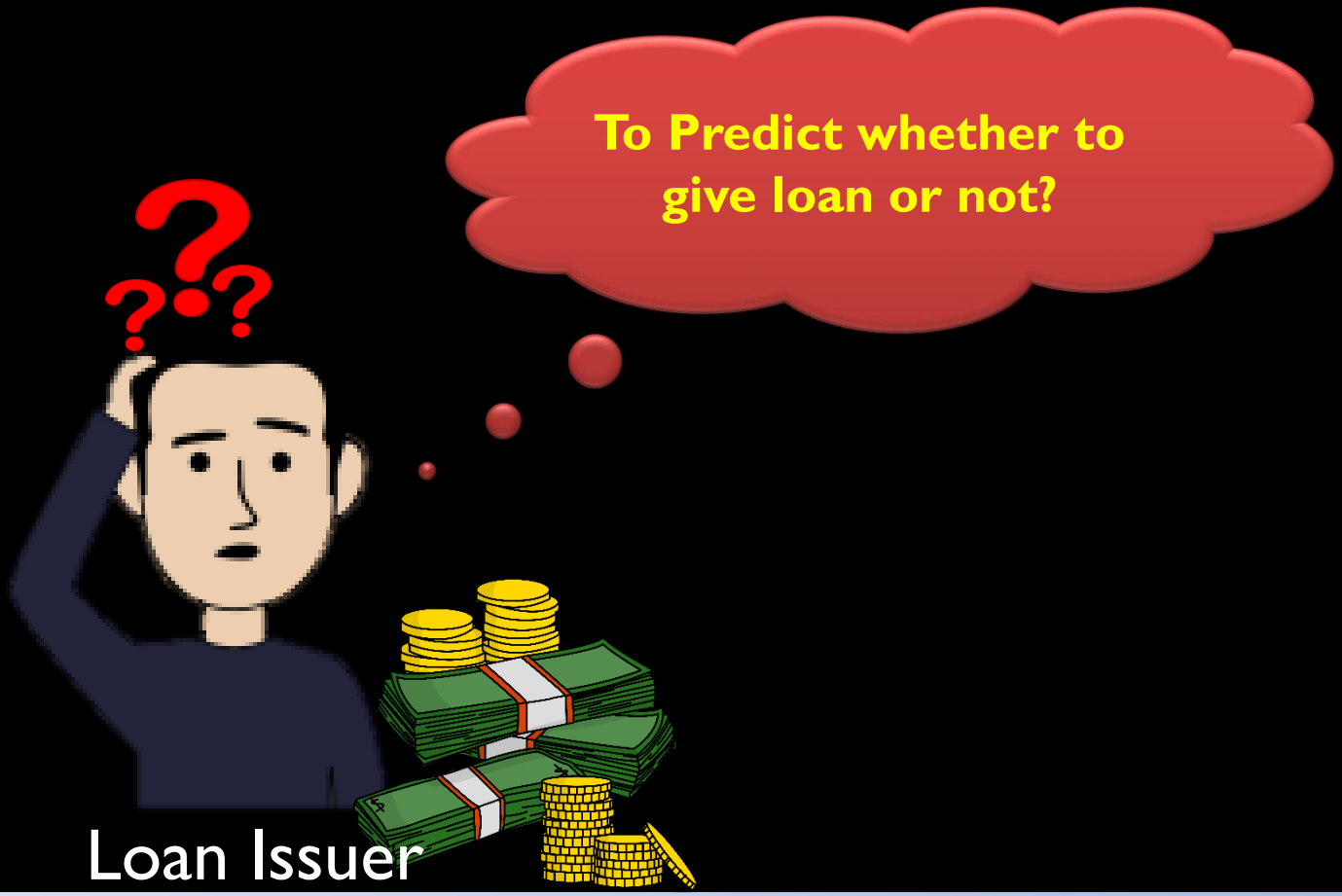
Data Taken from XYZ Co-operation

Data Cleaning and Feature Selection

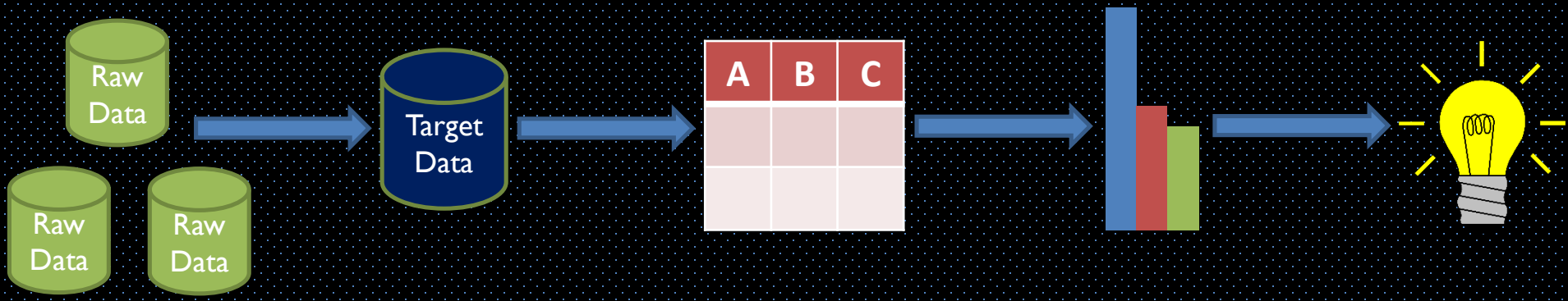


Required Data for Analysis

OBJECTIVE



Steps used for Credit Risk Analysis



Extraction of data from repositories

Data Cleansing and loading into Database

Data Transformation

Pattern Discovery using Algorithm such as regression, Classification, etc

Data Visualisation and Result

Extraction of Data from XYZ Co-operation

Data Contains

1. All the details of the people who applied for loan between June 2007 and December 2015
2. The data contains the indicator of default, payment information, Credit history, etc.
3. Columns – 73
   Rows - 855696

Pre - processing of Data

Data Cleaning

Pre Processing the Data

Find Missing Value and Calculating Missing Value Percentage.

In [11]:
```
Count_Null=Credit_DS.isnull().sum()
print(Count_Null)
```

...

In [13]:
```
Per_Null =round(Credit_DS.isnull().sum()/len(Credit_DS) * 100,2)
print(Per_Null)
```

...

In [16]:
```
Missing_Val.to_csv(r'D:\Desktop\MissingValues.csv')
```

Pre - processing of Data

Column Name	count	% Count	Column Name	count	% Count
annual_inc_joint	855527	99.95	inq_last_12m	842681	98.45
dti_joint	855529	99.95	desc	734157	85.77
verification_status_joint	855527	99.95	mths_since_last_record	724785	84.67
il_util	844360	98.64	mths_since_last_major_derog	642830	75.1
mths_since_rcnt_il	843035	98.49	mths_since_last_delinq	439812	51.38
open_acc_6m	842681	98.45	next_pymnt_d	252971	29.55
open_il_6m	842681	98.45	tot_coll_amt	67313	7.86
open_il_12m	842681	98.45	tot_cur_bal	67313	7.86
open_il_24m	842681	98.45	total_rev_hi_lim	67313	7.86
total_bal_il	842681	98.45	emp_title	49443	5.78
open_rv_12m	842681	98.45	emp_length	43061	5.03
open_rv_24m	842681	98.45	last_pymnt_d	8862	1.04
max_bal_bc	842681	98.45	revol_util	446	0.05
all_util	842681	98.45	last_credit_pull_d	50	0.01
inq_fi	842681	98.45	collections_12_mths_ex_med	56	0.01
total_cu_tl	842681	98.45			

Data Transformation

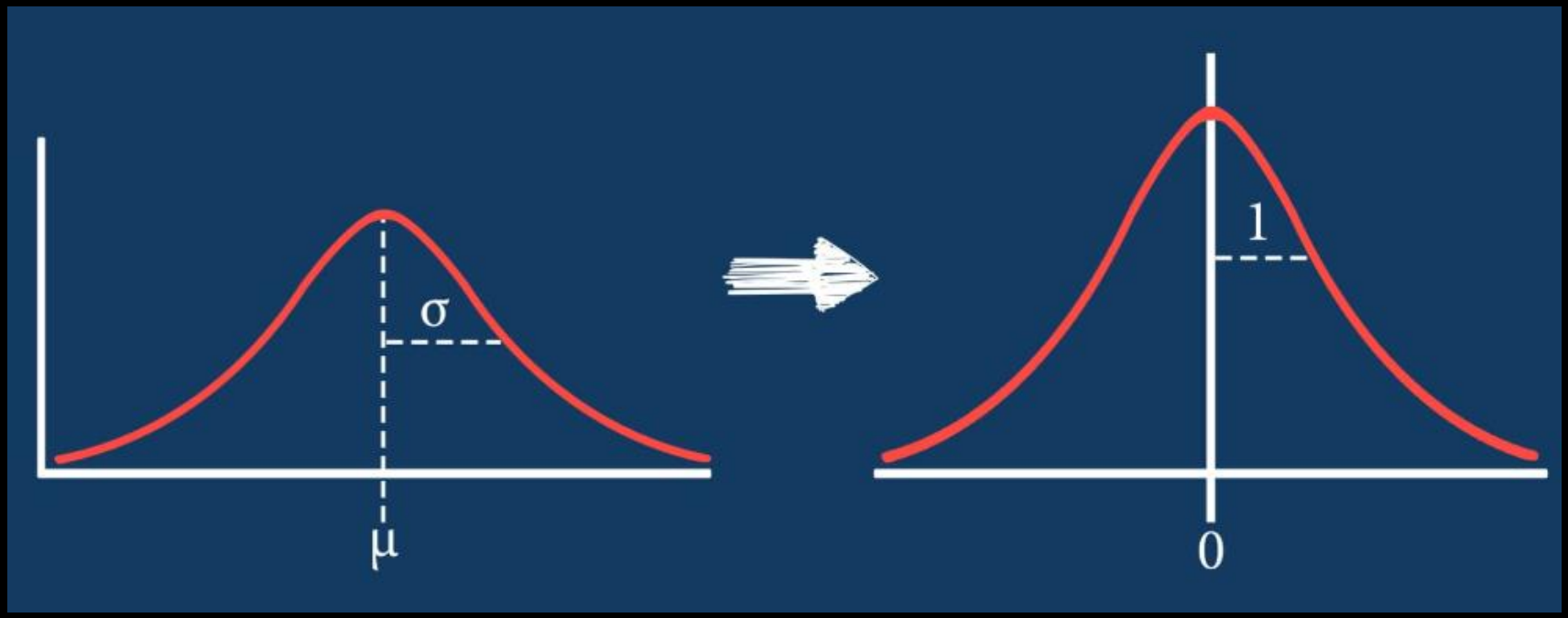
Converting Categorial data to Numeric

Column Name
Term
Grade
Emp_Length
Home_ownership
Verification_status
Purpose
Initial_List_Status
Application_type
Open_account

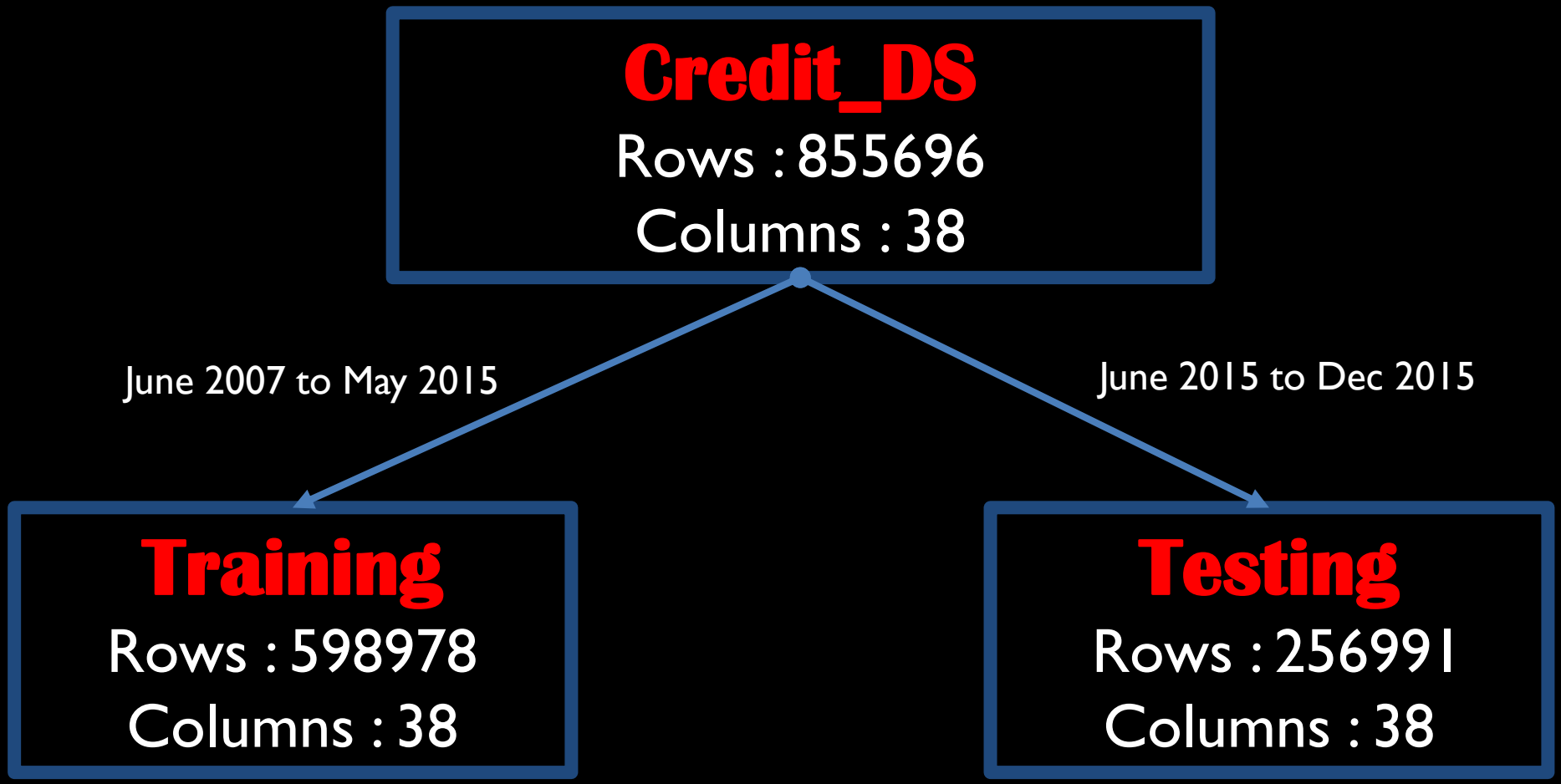
For Example:

```
In [30]: Credit_DS.grade.value_counts()
         grade_final={'A':6,'B':5,'C':4,'D':3,'E':2,'F':1,'G':0}
         Credit_DS.grade=[grade_final[item]for item in Credit_DS.grade]
         print(Credit_DS.grade)
```

Data Standardisation



Data Splitting

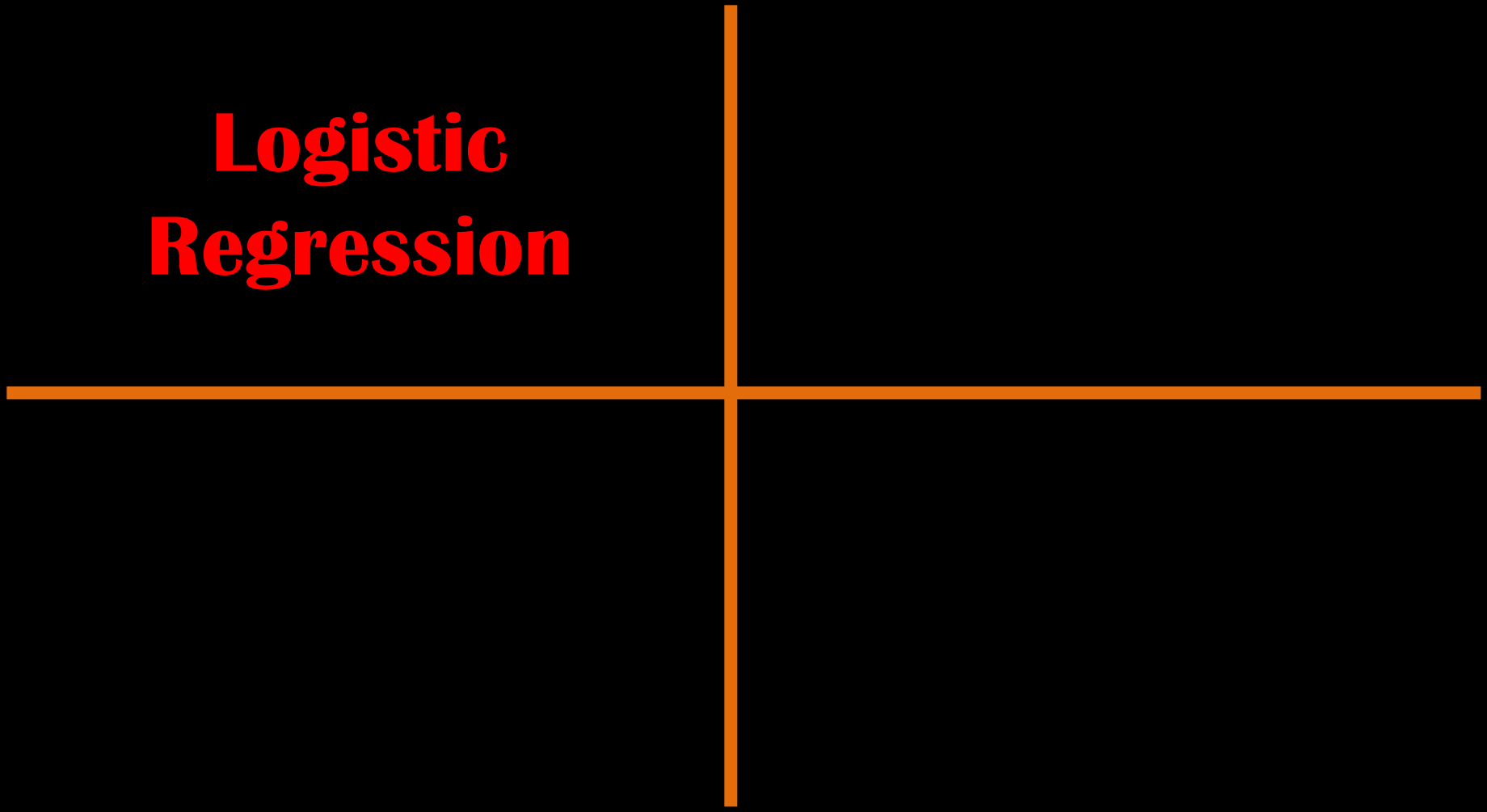


Models used for Analysing

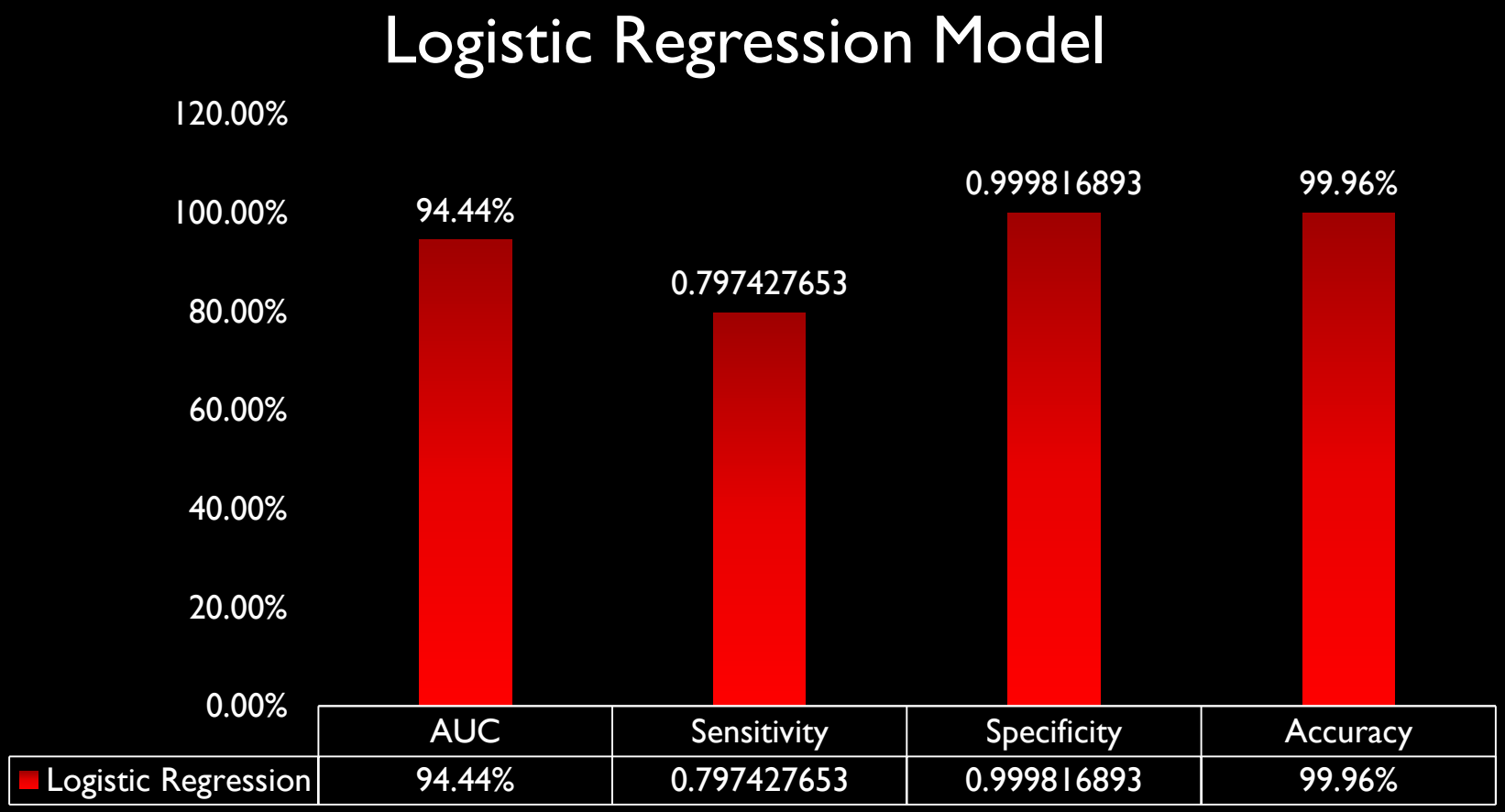
- Logistic Regression
- Decision Tree
- Adaptive Boost Classifier
- Extra Trees Classifier

Understanding Model Implementation

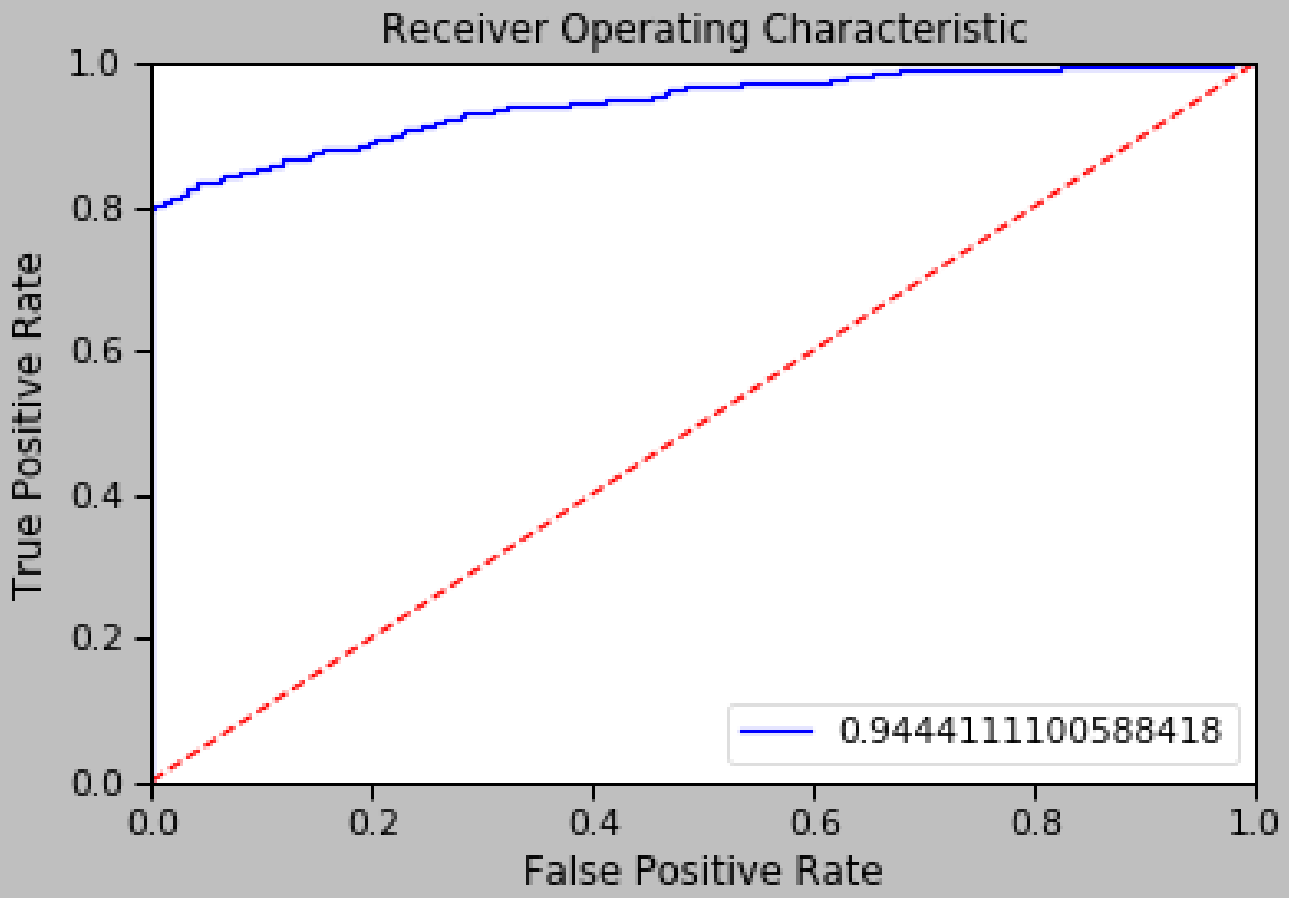
Logistic Regression



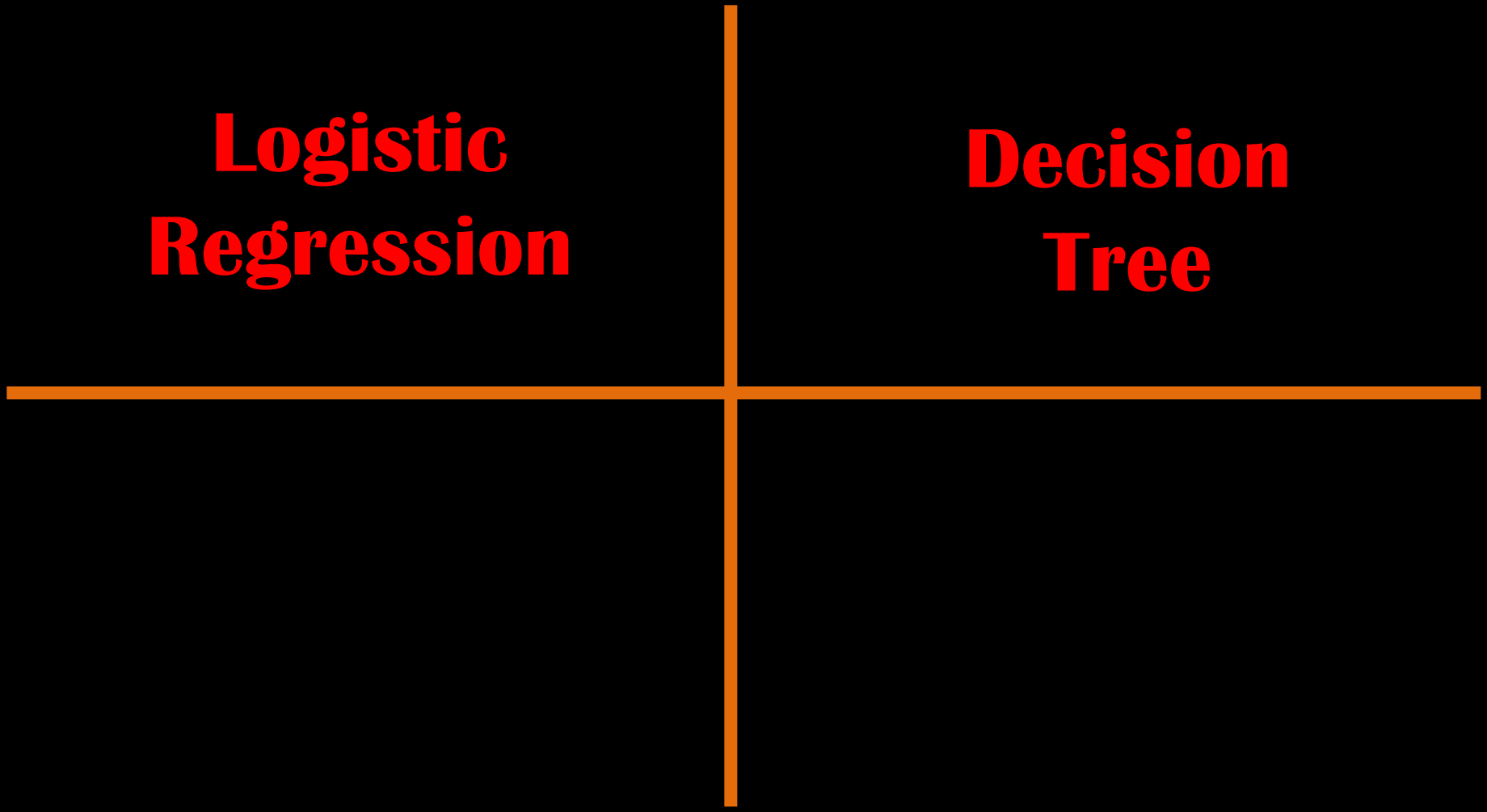
Model – 1 – Logistic Regression



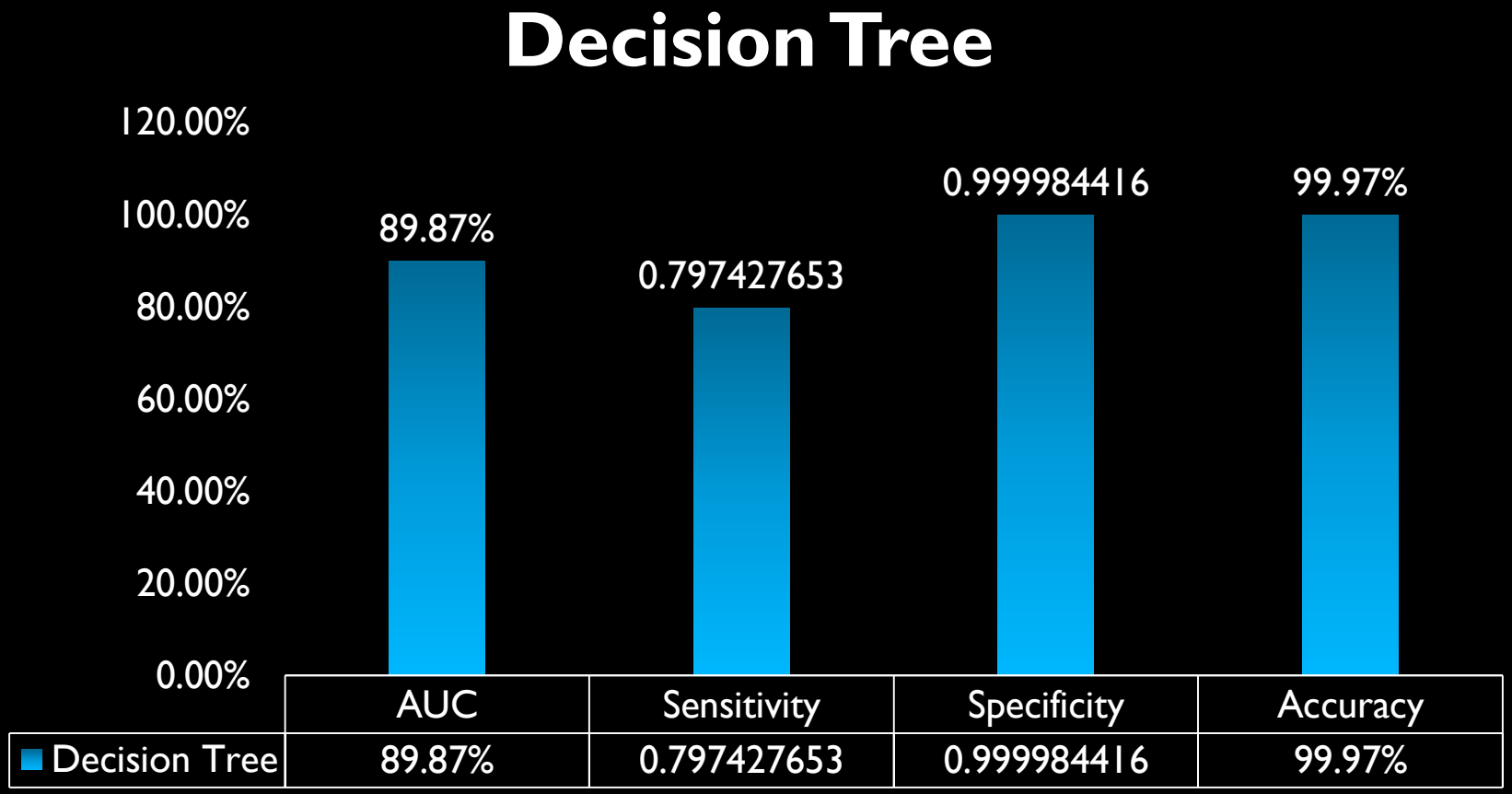
Receiver Operating Characteristic



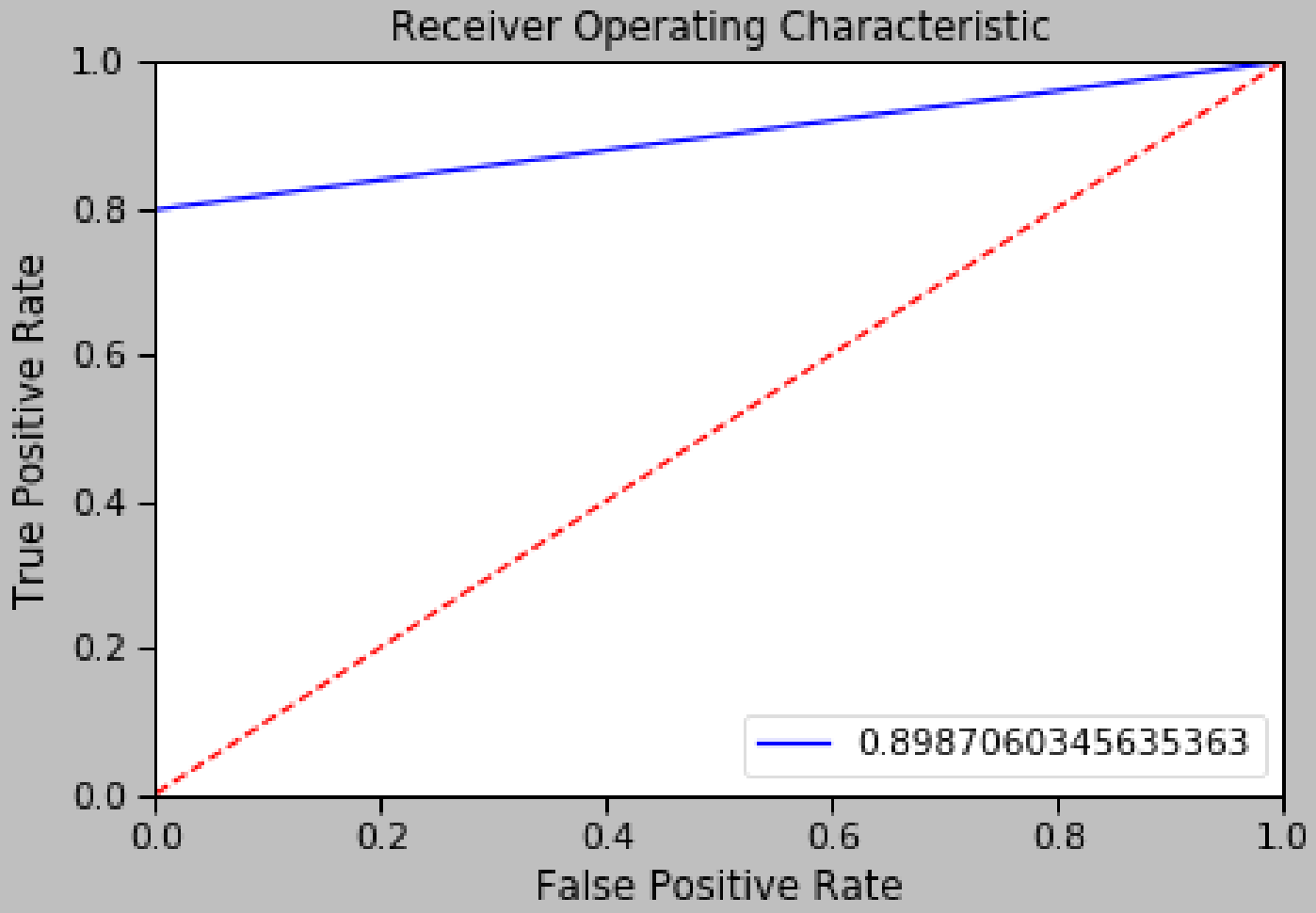
Understanding Model Implementation



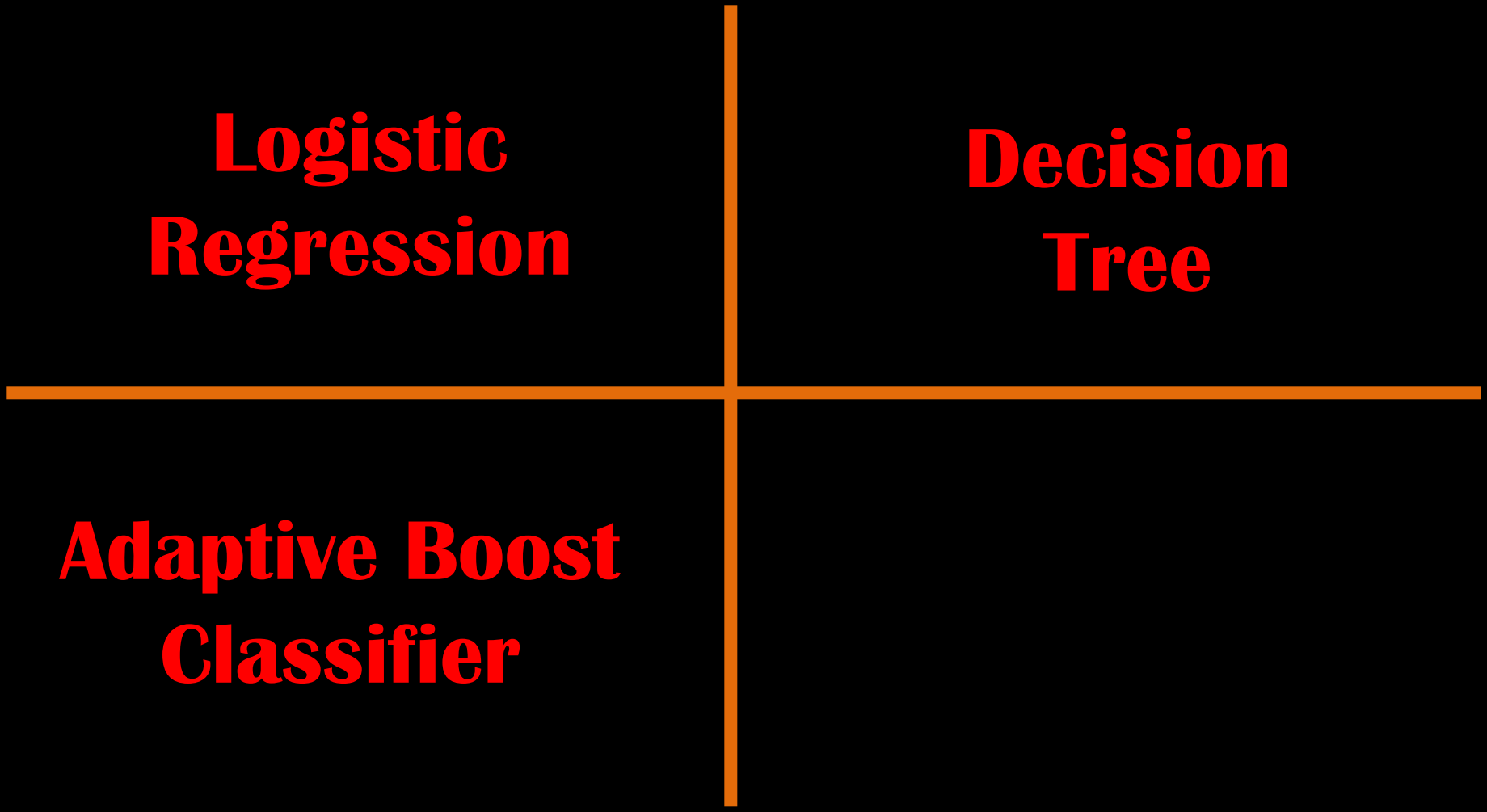
Model – 2 Decision Tree



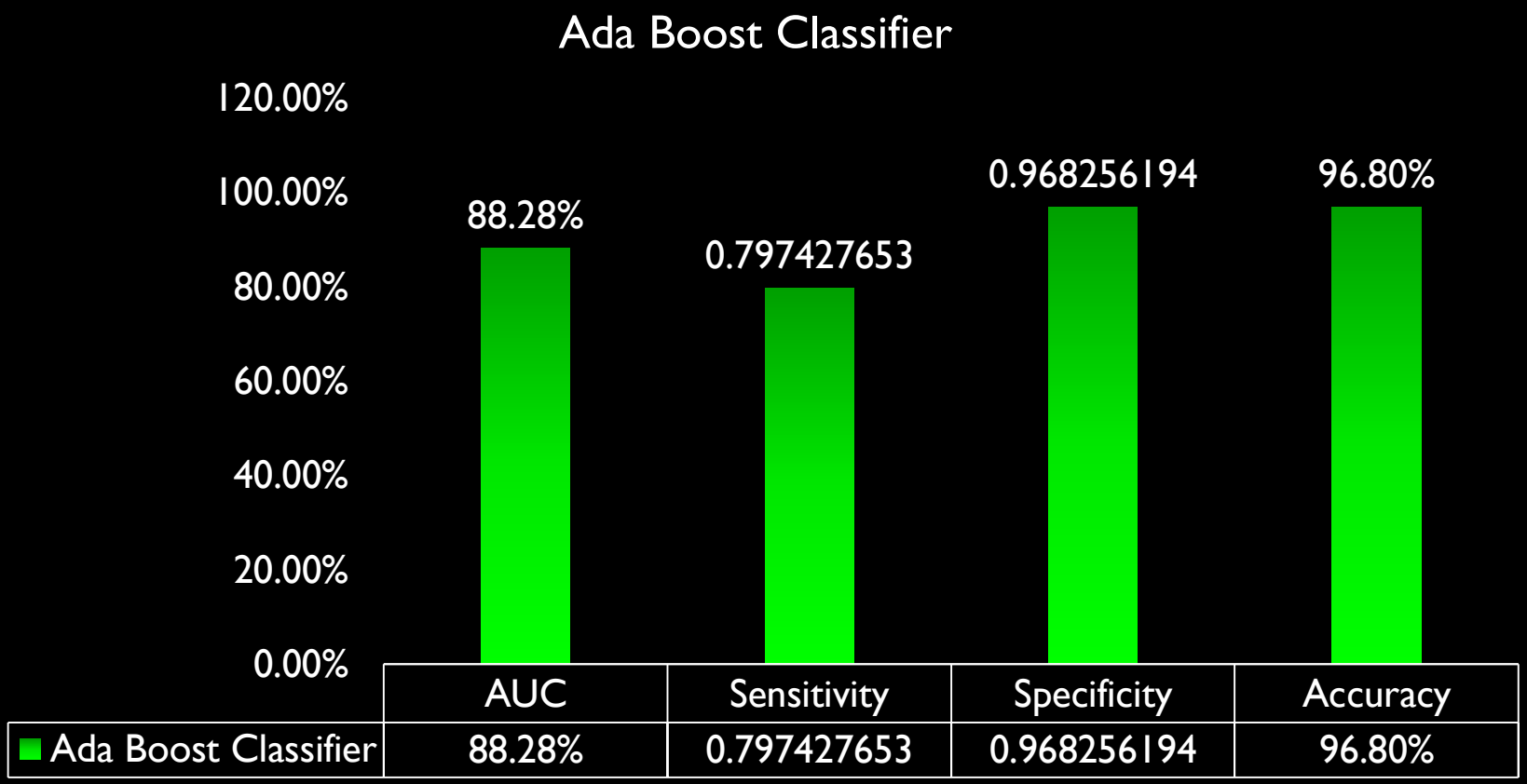
Receiver Operating Characteristic



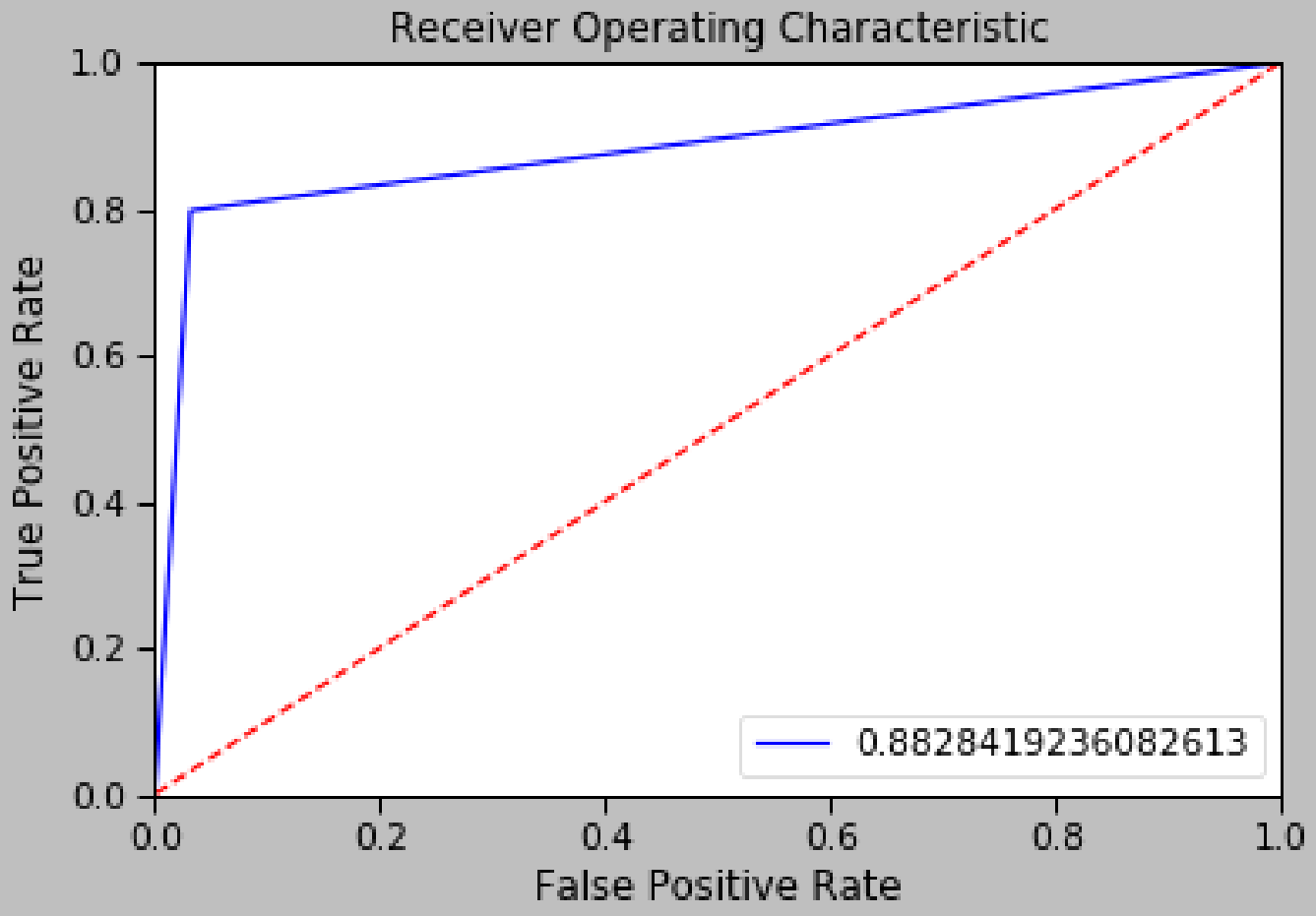
Understanding Model Implementation



Model – 3 Adaptive Boosting Classifier



Receiver Operating Characteristic



Understanding Model Implementation

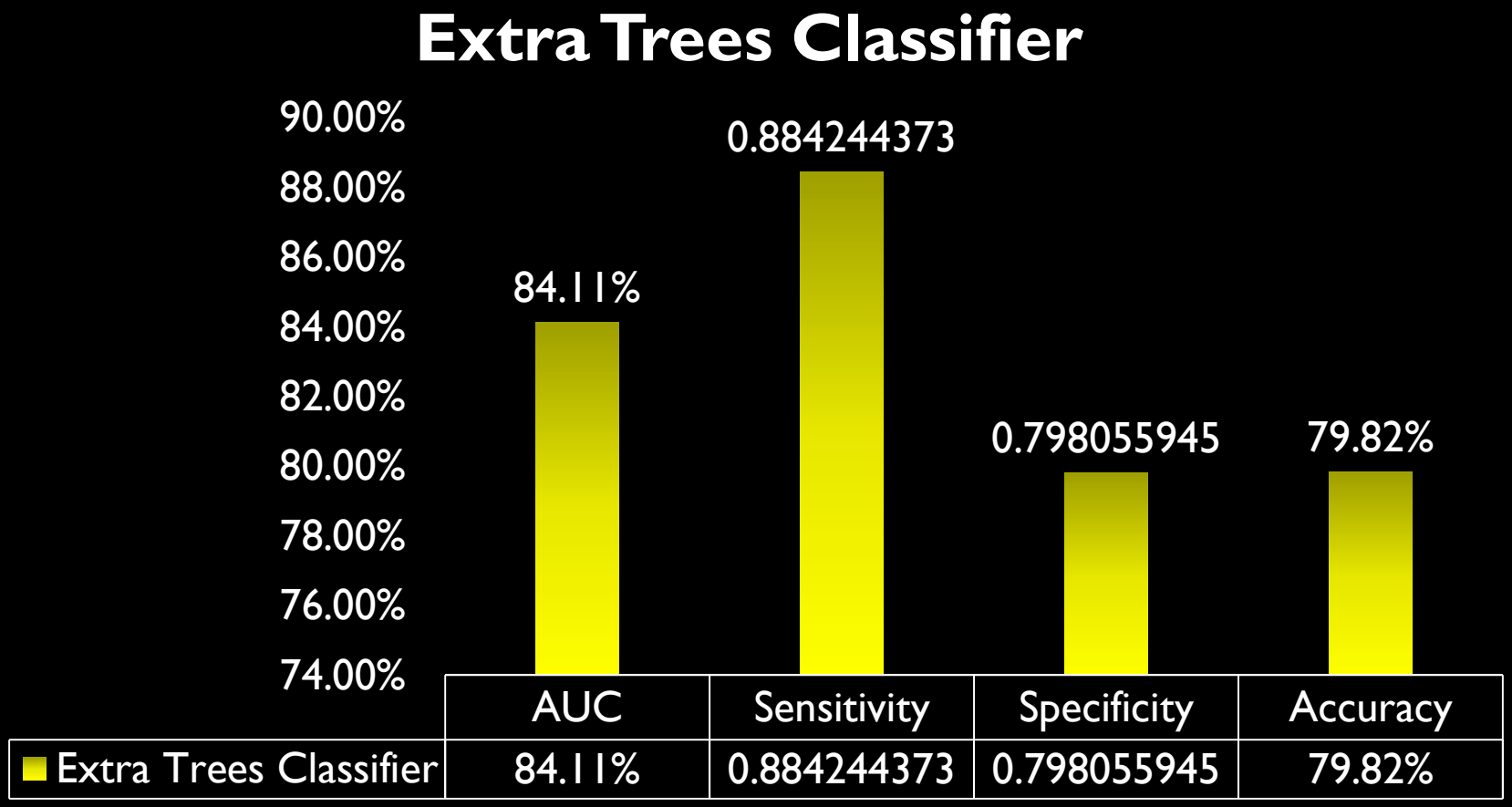
Logistic Regression

Decision Tree

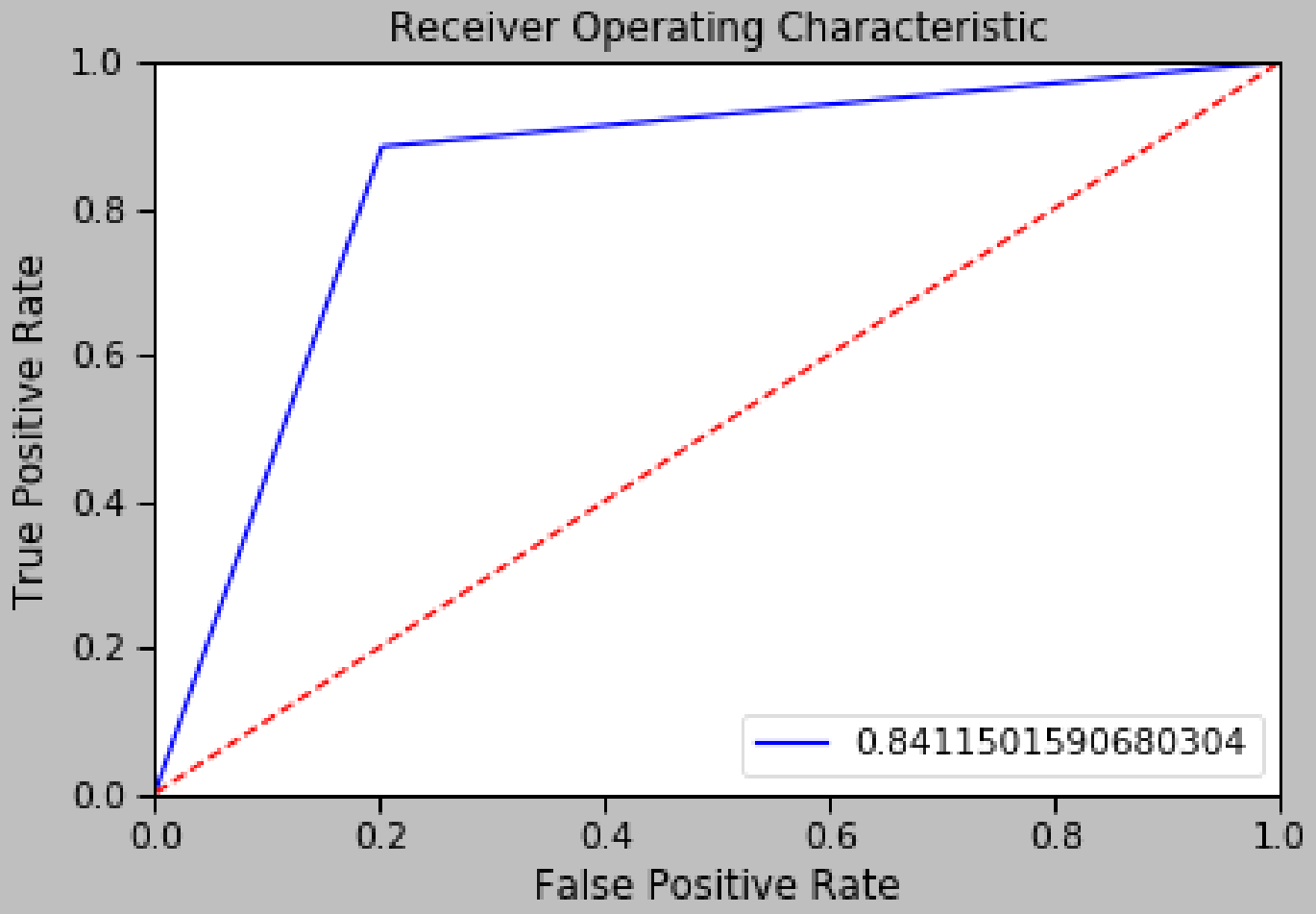
Adaptive Boost Classifier

Extra Trees Classifier

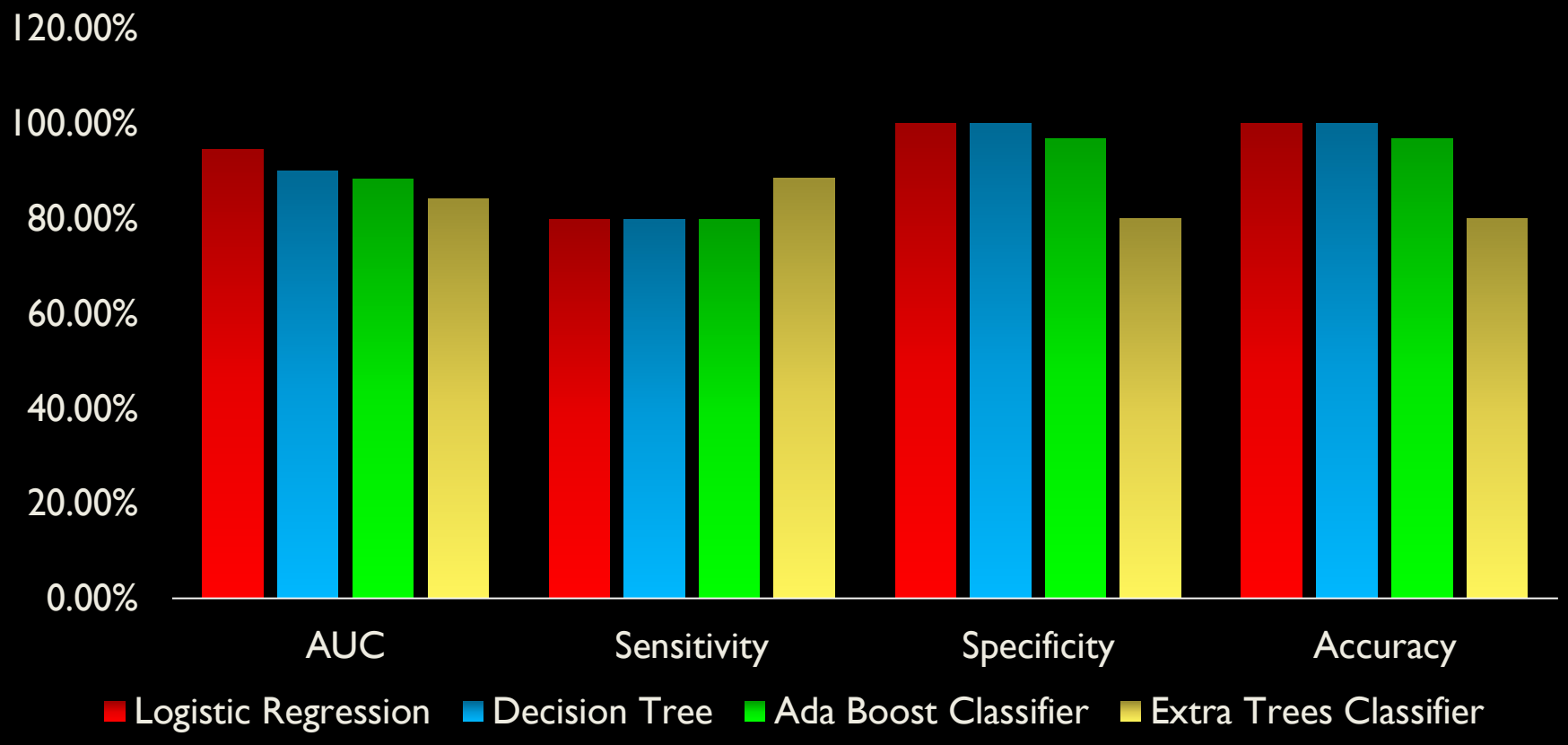
Model – 4 – Extra Trees Classifier



Receiver Operating Characteristic



Which Model is Performing Better?



Which Model is Performing Better?

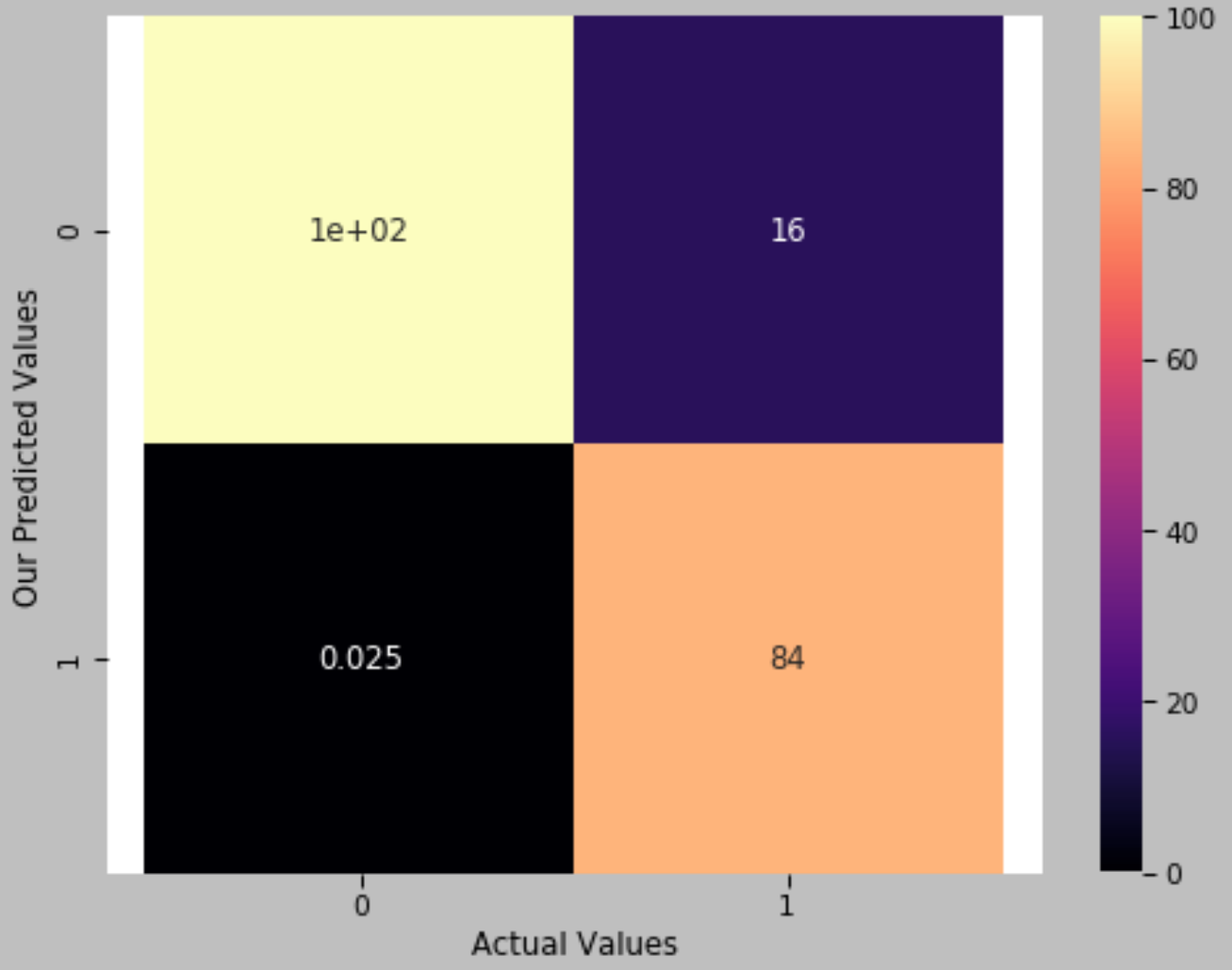
Logistic Regression

Decision Tree

Adaptive Boost Classifier

Extra Trees Classifier

Heat Map of Logistic Regression



Visualisation of Data

Conclusion

Credit Risk Analysis is a very crucial part of the banking sector and it plays an important role in the growth of the bank's profit. Using analysing techniques one can predict or analyse that a person applying for loan will repay the loan or not. So, multiples algorithms have been implemented to analyse a defaulter. The best model selected out of all models that have been tested is Logistic Regression model with an accuracy of 99.957%, which is cross verified with K-fold cross validation technique having the same accuracy of 99.956%. Further insights gained using visualizations from Tableau is that the bank is really working hard to come up from crisis of loss of revenue. To overcome this crisis, credit risk analysis will help them to grow and earn profit.

Documentation for the Credit Risk Analysis

Prepared by Nidhi Patel

Thank You !