Alumni Opportunities Using Machine Learning

Bachelor in Computer Science and Engineering

Under the supervision of

Dr. JiaUddin

And

The co-supervision of

Dr. Amitabha Chakrabarty

By

Arshif Sharlimin Siaum (13101063)

Md. Jawad Amin (13101045)

Kamrul Islam (13121013)

School of Engineering and Computer Science

BRAC University, Dhaka, Bangladesh

Submitted on : 23rd November, 2018

DECLARATION

We hereby declare that this thesis is based on results obtained from our own work. Due acknowledgement has been made in the text to all other material used. This thesis, neither in whole nor in part, has been previously submitted to any other University or Institute for the award of any degree or diploma.

Signature of Supervisor **Signature of Authors**

Dr. Jia Uddin

Kamrul Islam(13121013)

Arshif Sharlimin Siaum(13101063)

Md. Jawad Amin(13101045)

ACKNOWLEDGEMENT

First of all, we would like to thank Almighty Allah for enabling us to put our best efforts and successfully complete the research.

Second of all, we submit our heartiest gratitude to our respected Supervisor Dr. Jia Uddin for his contribution, guidance and support in conducting the research and preparation of the report. Every last involvement of his, starting from instilling in us the deadliest of fears to the kindest words of inspiration has permitted us to effectively complete the paper. We are truly grateful to him.

We are grateful for love and moral support of our parents and our friends. They helped us with their direct and indirect suggestions which assisted in achieving our goal. We would also like to acknowledge the assistance we have received from numerous resources over the internet especially from the research community.

Finally, we thank BRAC University for providing us the opportunity of conducting this research and for giving us the chance to complete our Bachelor degree.

Table of Contents

List of Figures

List of Tables

ABSTRACT

In the present time data science and artificial intelligence is taking charge of every aspect in our life to make our lives more easy and comfortable. Machine learning is an approach to achieve artificial intelligence and train machine to predict based on the given set of data. We propose and demonstrate a model using machine learning approach and apply it on the employment sector to connect proper graduates and employers with a good match. The proposed model enables analyzing the existing vacancies and unemployed graduates according to the education, work experience, and job requirement, employment type to build correlation among graduates and employers. This will allow us to predict the behavior of different type of graduates towards different type of vacancies. We used some existing machine learning algorithm to train and evaluate our model and predict the probability of finding a good match between graduates and employers.

CHAPTER 1

INTRODUCTION

1.1 Motivation

In recent times with increasing population and advancement of technology, we still fail to decrease unemployment rate in Bangladesh. Around 70% of unemployed people in Bangladesh are youth. On the other hand, industry lacks sufficient skilled employees. Why is this gap? We aspire to bridge this gap by connecting graduates and potential employers. In this job portal, there will be the whole database of the particular university students. With Basic, educational and job information. An employer can search, sort, view and mark/unmark students and call for interview from there.

1.2 Thesis Outline

Chapter 2 provides the Background study in details including the algorithms and techniques used in the system Chapter 3 describes the proposed model and work flow along with implementation details. Chapter 4 presents the results of the experiment along with performance analysis and comparisons. Chapter 5 concludes the paper specifying the limitations and challenges while planning future development of the project.

CHAPTER 2

History & Background Information

2.1 Fundamental of Machine Learning

We are going to give a small brief about machine learning and some algorithm we used in our model. In this present era everyone is familiar with machine learning. As we know AI (Artificial Intelligence) is taking over modern technology very rapidly and machine learning is one of the major factors of AI. Machine Learning is a field of computer science where the system has the ability to learn from dataset and predict the required result. In other words machine learning is a sub-set of AI where computer algorithms are used to learn autonomously from data and information. Data mining on the other hand, is the process of examining great volume of data to find hidden relationships and patterns. English computer scientist, mathematician Alan Turing created the “Turing Test” to determine if a computer has real intelligence. To pass the test, a computer must be able to fool a human into believing it is also human. The term machine learning was first introduced by Arthur Samuel, one of the pioneers of machine learning [1]. In 1959 while at “IBM” he developed a program that learned how to play checkers better than him. Tom Mitchell, another well regarded machine learning researcher, proposed a precise definition in 1998, Well posed Learning Problem: A computer program is said to learn from experience E with respect to some task T and some performance measure P, if its performance on T, as measured by P, improves with experience E [2]. The first successful neuro-computer (the Mark I perceptron Figure 2.1) was developed during 1957 and 1958 by Frank Rosenblatt, Charles Wightman, and others. The “nearest neighbor” algorithm was written, allowing computers to begin using very basic pattern recognition. This could be used to map a route for traveling salesmen, starting at a random city but ensuring they visit all cities during a short tour.

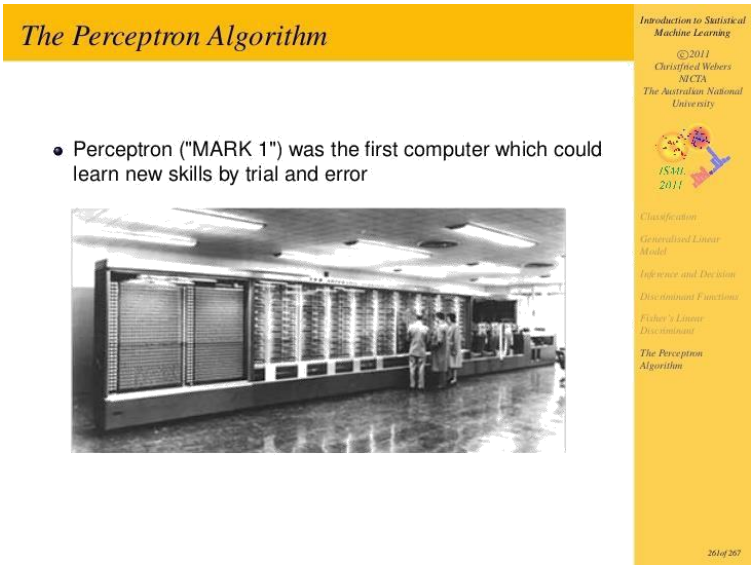

Figure 2.1: Mark 1 Perceptron (first neuro computer)

In 1985 Terry Sejnowski invents Net Talk, which learns to pronounce words the same way a baby does .Then in 1997 IBM developed deep blue which beat world champion at chess. In 2006 Geoffrey Hinton used the term "deep learning" to explain new algorithms that let computers "see" and distinguish objects and text in images and videos. And in 2010 The Microsoft developed Kinect which can track 20 human features at a rate of 30 times per second, allowing people to interact with the computer via movements and gestures [3-5].

There were also some works which made the road of machine learning more fluent. For example in 1979 students of Stanford university invented the "Stanford cart" which had the capability to navigate obstacles in a room on its own [6].It was a remotely controlled TV-equipped mobile robot.

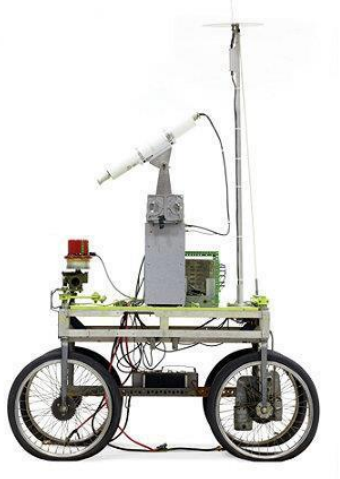

Figure 2.2: Stanford Cart

Explanation based learning was first introduced by Gerald Dejong in 1986.It became an important sector of machine learning because of its capability of using prior knowledge to generalize more correctly from fewer training data sets [6].

There are mainly two types of techniques used in machine learning algorithm.

A. Supervised Learning: Trains a model on known input and output data so that it can predict future outputs.

B. Unsupervised Learning: This finds hidden patterns or intrinsic structures in input data.

There is also another type of learning which is known as reinforcement learning.

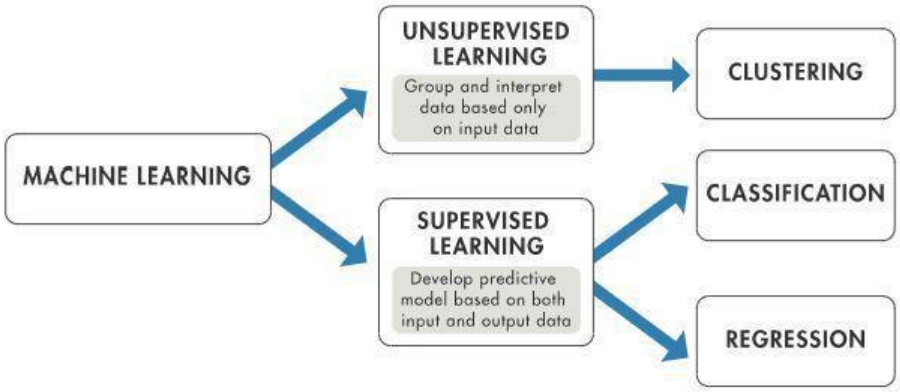

Figure 2.3 : Machine Learning Techniques

A. Supervised Learning: Supervised machine Learning develops a model which makes prediction based on evidence in the presence of uncertainty. A supervised learning algorithm takes a known set of input data and known responses to the data (output) and trains a model to generate reasonable predictions for the response to new data.

Supervised learning uses classification and regression techniques to develop predictive models. The machine attempts to learn the relationship between different parameters from scratch, by running labeled training data through a learning algorithm [8]. Algorithms used for supervised learning are linear regression for regression problems, Random forest for classification and regression problems and support vector machine for classification problems.

Classification techniques predict discrete responses. For example, whether an email is genuine or spam. Classification models classify input data into categories. Common algorithms: support vector machine (SVM), boosted and bagged decision trees, k-nearest neighbor, Naïve Bayes, discriminant analysis, logistic regression, and neural networks.

B. Unsupervised Learning

Unsupervised learning finds hidden patterns or intrinsic structures in data. It is used to draw inferences from datasets consisting of input data without labeled responses**.** The goal for unsupervised learning is to model the structure or re-arrange the data set to learn more about the data. Unsupervised learning can be divided in to two types according to their work.

i) Clustering

ii) Association

Clustering: A clustering unsupervised learning is when the goal is to discover the inherent grouping in data set [9]. For example: grouping people.

Association: Association rules unsupervised learning is when the goal is to find the group that represents large amount of data in the data sets [9].

Common Algorithms: k-means algorithm, Hierarchical clustering.

In our project we are going to use different machine learning algorithms to increase the accuracy of our system. The algorithms we are going to use are decision tree algorithm, logistic regression, k-nearest neighbor algorithm, neural network algorithm.

As we trained our model using supervised machine learning approach, we are going to discuss more about those algorithms which are used in supervised learning.

Decision tree:

Decision trees are a classic machine learning technique. The basic intuition behind a decision tree is to map out all possible decision paths in the form of a tree. Decision tree algorithm is a commonly used data mining method for establishing classification systems based on multiple covariates or for developing prediction algorithms for a target variable [10]. First introduced in 1960's, decision tree is one of the most effective algorithms for data mining; they have been frequently used in several sectors because they are easy to be used, free of ambiguity, and robust even in the presence of missing values. A decision tree is a structure like flow chart, where each internal node denotes a test on an attribute, each branch represents the outcome of a test, and each leaf (or terminal) node holds a class label. The topmost node in a tree is the root node [8]. There are many specific decision-tree algorithms. Decision Tree algorithm belongs to the supervised learning algorithms family. Unlike other supervised learning algorithms, decision tree algorithm can be used for solving regression and classification problems too [16]. In figure 2.4 we demonstrate how decision tree works for predicting an output.

Figure 2.4: Example of how Decision Tree Works

Logistic Regression:

Logistic regression is another re-known algorithm for supervised machine learning. The ability to predict the odds has made the logistic regression model a popular method for statistical analysis [10].It can be used for prospective, retrospective or cross-sectional data. The result is the impact of each variable on the odds ratio of the observed event of interest. The main advantage is to avoid confounding effects by analyzing the association of all variables together [11]. Logistic regression uses an equation as the representation, very much like linear regression.

Input values (x) are combined linearly using weights or coefficient values (referred to as the Greek capital letter Beta) to predict an output value (y). A key difference from linear regression is that the output value being modeled is binary values (0 or 1) rather than a numeric value.[15] Logistic regression expressed/ calculated using the equation following equation [1]:

$$= \frac{\quad}{1+\ ^{\wedge}(\ 0+\ 1\cdot\ )} \qquad (1)$$

Here,

y is the predicted output

b0 is the bias or intercept term

b1 is the coefficient for the single input value (x).

Each column in our input data has an associated b coefficient (a constant real value) that must be learned from our training data.

Neural Network:

A neural network is a set of interconnected layers. The inputs are the first layer, and are connected to an output layer by an acyclic graph comprised of weighted edges and nodes.

Between the input and output layers you can insert multiple hidden layers. Most predictive tasks can be accomplished easily with only one or a few hidden layers. However, recent research has shown that deep neural networks (DNN) with many layers can be very effective in complex tasks such as image or speech recognition. The successive layers are used to model increasing levels of semantic depth. [6]

The relationship between inputs and outputs is learned from training the neural network on the input data. The direction of the graph proceeds from the inputs through the hidden layer and to the output layer. All nodes in a layer are connected by the weighted edges to nodes in the next layer. [3]

To compute the output of the network for a particular input, a value is calculated at each node in the hidden layers and in the output layer. The value is set by calculating the weighted sum of the values of the nodes from the previous layer. An activation function is then applied to that weighted sum.

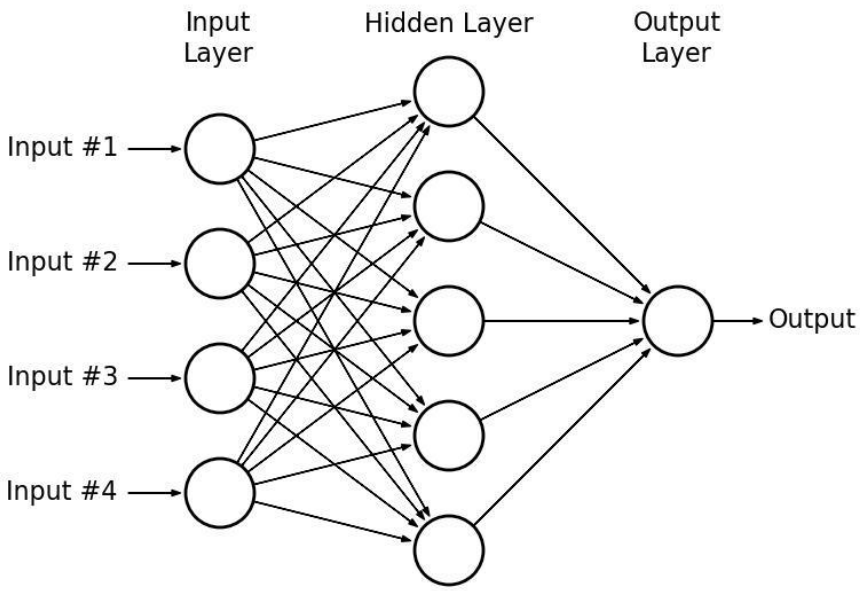

Figure 2.5: Neural Network Basic working technique

Support Vector Machine:

Kernel-classifiers comprise a powerful class of non-linear decision functions for binary classification. The Support vector machine is an example of a learning algorithm for kernel classifiers that singles out the consistent classifier with the largest margin [22]

Support vector machines (SVMs) are particular linear classifiers which are based on the margin maximization principle. They perform structural risk minimization, which improves the complexity of the classifier with the aim of achieving excellent generalization performance.[14] The SVM accomplishes the classification task by constructing, in a higher dimensional space, the hyper plane that optimally separates the data into two categories. In support vector machine algorithm we plot each item as a point in n-dimensional space (n-number of features) with the value of every feature being the value of a particular co-ordinate. After that we classify the features by finding the hyper plane that differentiates the two classes very well [2].

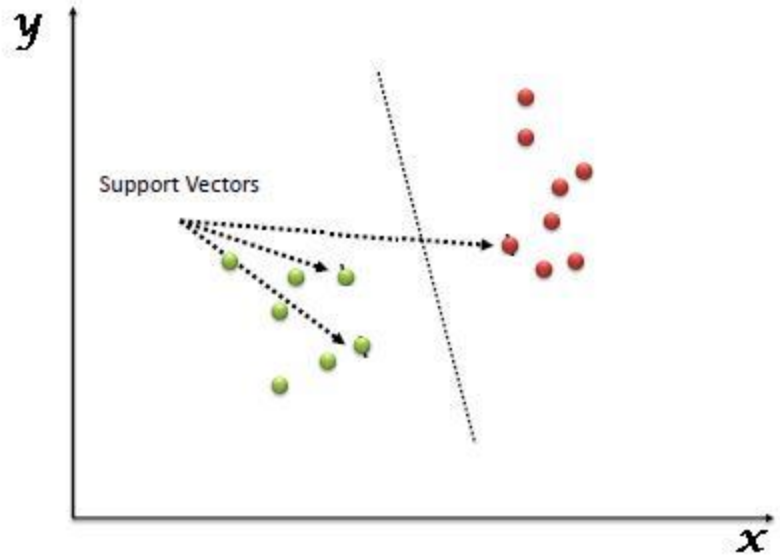

Figure 2.6: Support Vector Machine

Kernel

The learning of the hyper-plane in linear SVM is done by transforming the problem using some linear algebra. This is where the kernel plays role.

For linear kernel the equation for prediction for a new input using the dot product between the input (x) and each support vector (xi) is calculated as follows[15]:

$$() = 0 + \quad (\ * (, \ )) \ldots\ldots\ldots\ldots\ldots\ldots\ldots\ldots 2$$

This is an equation that involves calculating the inner products of a new input vector (x) with all support vectors in training data. The coefficients B0 and ai (for each input) must be estimated from the training data by the learning algorithm.

Gamma

The gamma parameter defines how far the influence of a single training example reaches, with low values meaning ‘far’ and high values meaning ‘close’. In other words, with low gamma, points far away from plausible separation line are considered in calculation for the separation line. Whereas high gamma means the points close to plausible line are considered in calculation. [22]

Margin

And finally last but very important characteristic of SVM classifier. SVM to core tries to achieve a good margin. A margin is a separation of line to the closest class points. A good margin is one where this separation is larger for both the classes. A good margin allows the points to be in their respective classes without crossing to other class [22].

Decision Forest:

This decision forest algorithm is an ensemble learning method intended for classification tasks. Ensemble methods are based on the general principle that rather than relying on a single model, we can get better results and a more generalized model by creating multiple related models and combining them in some way. Generally, ensemble models provide better coverage and accuracy than single decision trees [9]. There are many ways to create individual models and combine them in an ensemble. This particular implementation of a decision forest works by building multiple decision trees and then voting on the most popular output class. Voting is one of the better-known methods for generating results in an ensemble model [7].

Many individual classification trees are created, using the entire dataset, but different (usually randomized) starting points. This differs from the random forest approach, in which the individual decision trees might only use some randomized portion of the data or features. Each tree in the decision forest tree outputs a non-normalized frequency histogram of labels. The aggregation process sums these histograms and normalizes the result to get the “probabilities” for each label. The trees that have high prediction confidence will have a greater weight in the final decision of the ensemble.[18]

Decision trees in general have many advantages for classification tasks: They can capture non-linear decision boundaries. We can train and predict on lots of data, as they are efficient in computation and memory usage. Feature selection is integrated in the training and classification processes. Trees can accommodate noisy data and many features. They are non-parametric models, meaning they can handle data with varied distributions.

Bayes Point Machine:

The algorithm in this module uses a Bayesian approach to linear classification called the "Bayes Point Machine". This algorithm efficiently approximates the theoretically optimal Bayesian average of linear classifiers (in terms of generalization performance) by choosing one "average" classifier, the Bayes

Point. Because the Bayes Point Machine is a Bayesian classification model, it is not prone to over fitting to the training data. [17]

Chapter 3

Methodology

To get our desired result we have to follow a workflow following from data collection to evaluating our model. Our proposed model will have the following steps.

- ✓ Collect data
- ✓ Pre-process data
- ✓ Split data
- ✓ Train model using different algorithm
- ✓ Evaluate model with test data
- ✓ Choose the best algorithm with highest accuracy

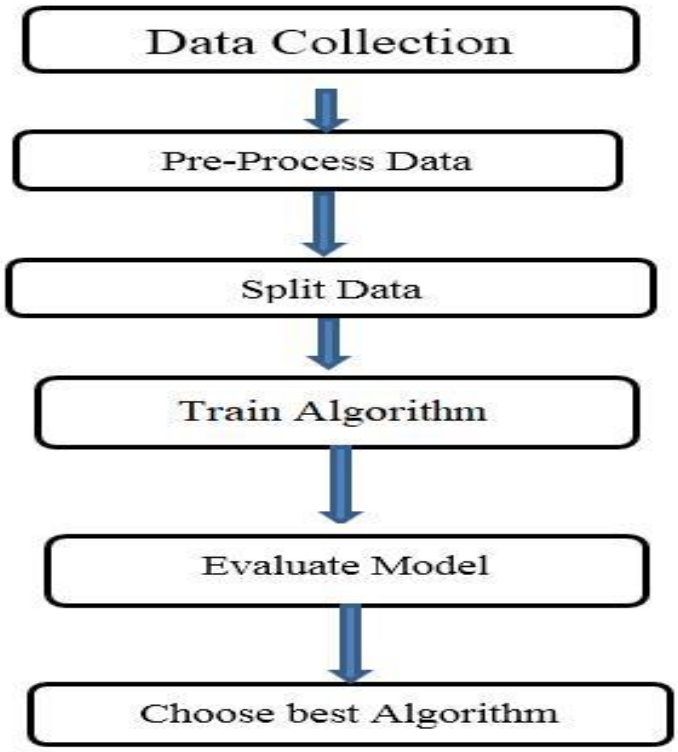

Figure 3.1: Workflow of proposed Model

3.1 Data Collection

To apply machine learning algorithm on the data set we needed to have information of people including Education (considering university, subject, major, etc.), job experience (their previous experience). Also, we needed to have employer information of different type such as Organization Industry, Department, Experience, Salary, Education, Department, etc. Data table for graduates and employer success probability would look like the given tables.

Table 3.1: Information of Graduate

Name	Sex	Age	Education	Job Experience	Salary	Preferred Sector	Preferred Job
Graduate 1	M/F	Age	Undergrad/Masters	Past organizations	Expected	NGO/Private	Organization name

Graduate Information table

Table 3.2: Information of vacancy

Organization	Industry	Department	Experience	Salary	Education
Org name	NGO/ Private/ Multinational	Dept. Name	Undergrad/ Masters	Offered	Undergrad/ Masters

Feature & Class selection

After merging all the data set we need to select feature and class which will help the algorithm to learn.

Feature

In machine learning and pattern recognition, a feature is an individual measurable property or characteristic of a phenomenon being observed. Choosing informative, discriminating and independent feature**s** is a crucial step for effective algorithms in pattern recognition, classification and regression.

Class

Class is the output category of our data. We can call these categories as well. The labels on our data will point to one of the classes.

Label

Labeled data is a group of samples that has been tagged with one or more labels. Labeling typically takes a set of unlabeled data and augments each piece of that unlabeled data with meaningful tags that are informative

After merging the two data table we will have a final dataset from which our model will be trained. The new table will look like this

Table 3.3: Final Dataset Table

Job	Sector	Salary	Department	Job experience	Education	Age	Job type	Area	Probability

From the table we can see that probability is the label of our dataset which has two classes true (positive ROI and false (Negative ROI). The other columns contain features to predict the class.

3.2 Data Pre-processing

Formatting: Most of the data where we ran algorithms were generated with given specification. Therefore, we can skip this part of data preprocessing.

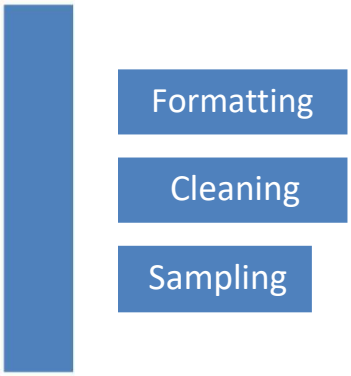

Figure 3.3: Data Pre-process

Cleaning: Cleaning data is the removal or fixing of missing data. Some of the data instances were incomplete and did not carry the data we need to address the problem those instances needed to be removed.

Sampling: The amount of data we first generated was way more. More data can result in much longer running times for algorithms and larger computational and memory requirements. Therefore, we took smaller representative sample of the selected data that may be much faster for exploring and prototyping solutions before considering the whole dataset.

3.4 Split Data

After cleaning our dataset, we needed to split the data set between training data and test data. For splitting test data and training data, we take 75% of our whole data for training our model and the rest 25% data were kept aside as test data

Training data: Training data were used to train our model. Mainly our model learns from this training data. In the training data set we will run different algorithms.

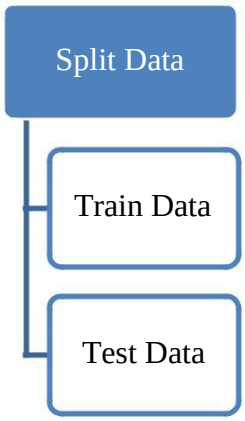

Fig 3.4: Split Dataset.

Test data: After the training we need to evaluate how much our model learn and how accurate the model is. To complete this process, we use test data to get the accuracy, as test data already know that which observation falls into which class.

3.4 Evaluate Model

We have to be careful choosing algorithm. There is list of algorithms in machine learning. Some are good for numeric data; some are good for categorical data. Some are good for supervised learning, some are good for unsupervised, and some are good for reinforcement learning. Among all of these algorithms we choose classification algorithms as our predicting model is based on classification.

3.4.1 Logistic Regression

Logistic Regression is one of the best algorithms for classification problems. We discussed earlier how logistic regression works. Now we apply logistic regression to train our model. In figure 3.5 we can see training model where logistic regression is applied

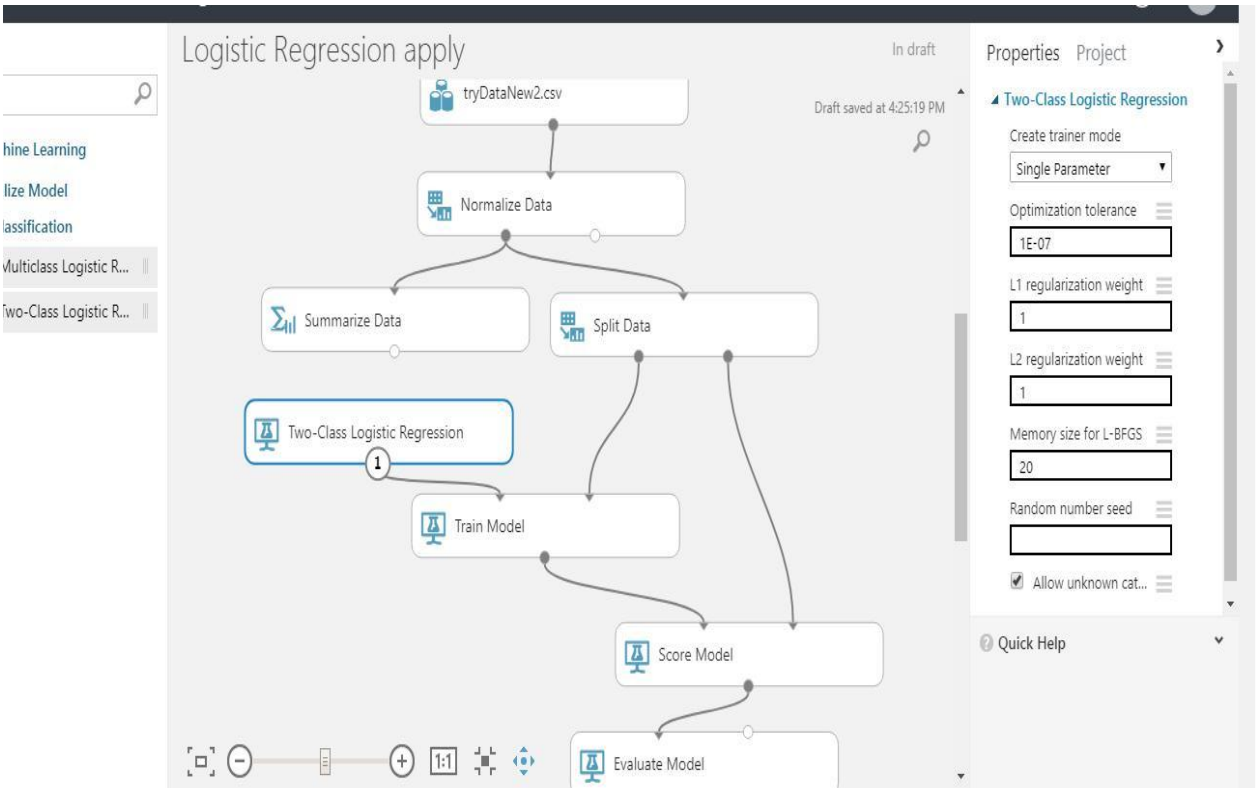

Figure 3.5: Logistic Regression

We also get the ROC curve and accuracy by applying this algorithm.

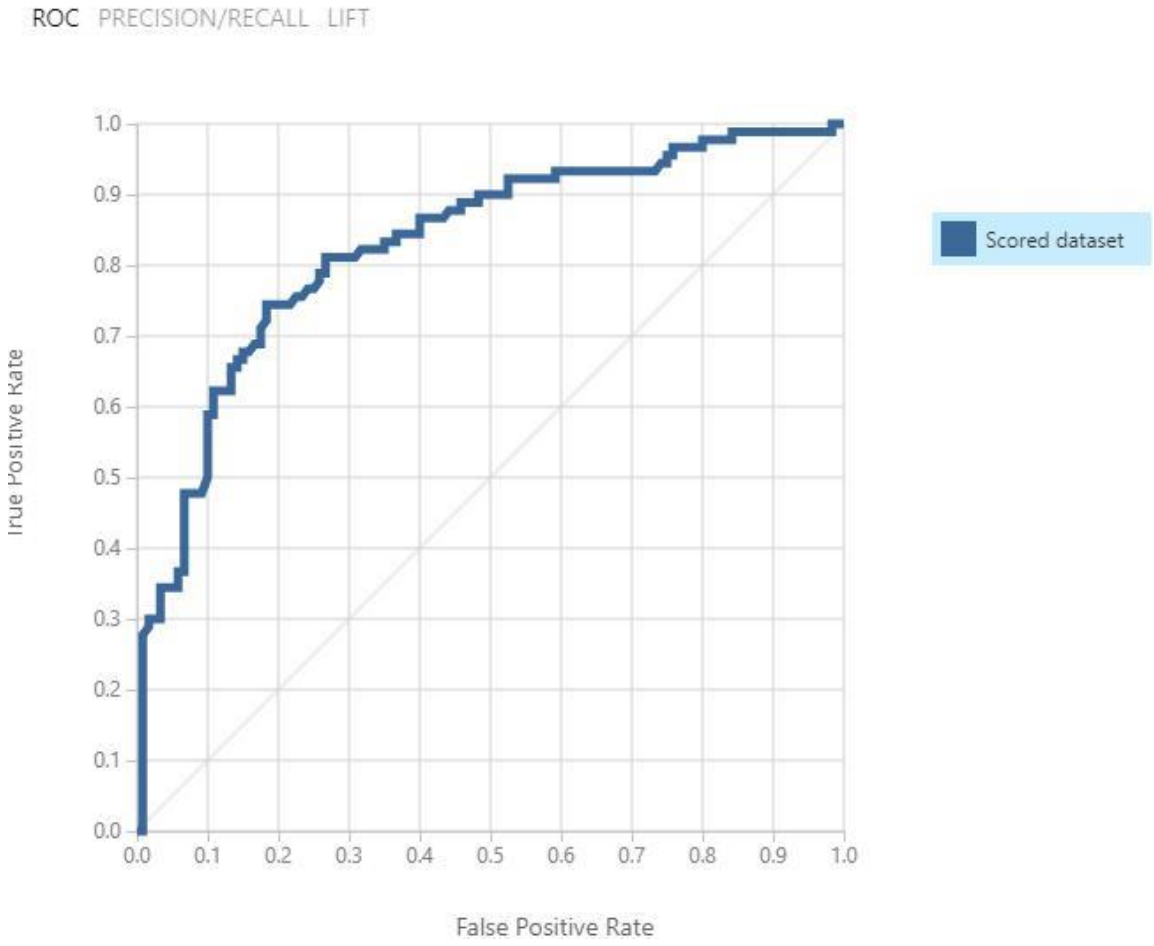

Figure 3.6 : ROC Curve

Figure 1ROC Logistic Regression

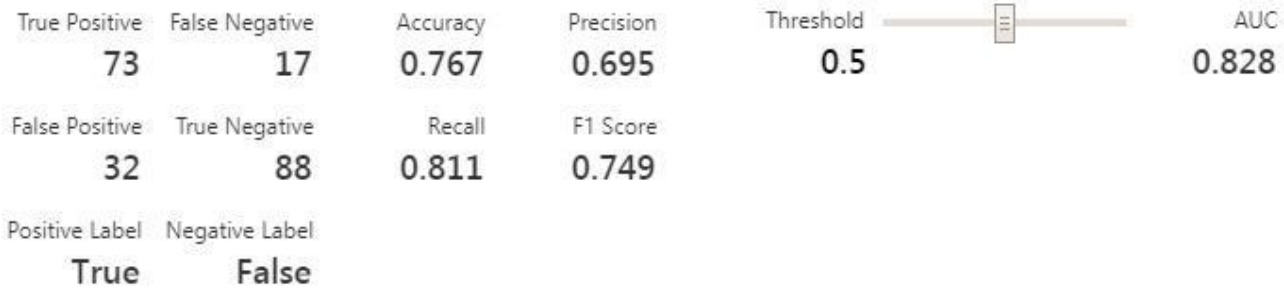

Figure 3.7: Accuracy Logistic Regression

3.4.2 Decision Tree

In figure 3.8 Applying decision tree in our dataset to train our model. We also get the ROC curve in figure 3.9 and Accuracy of Decision Tree in figure 3.10 after training the model.

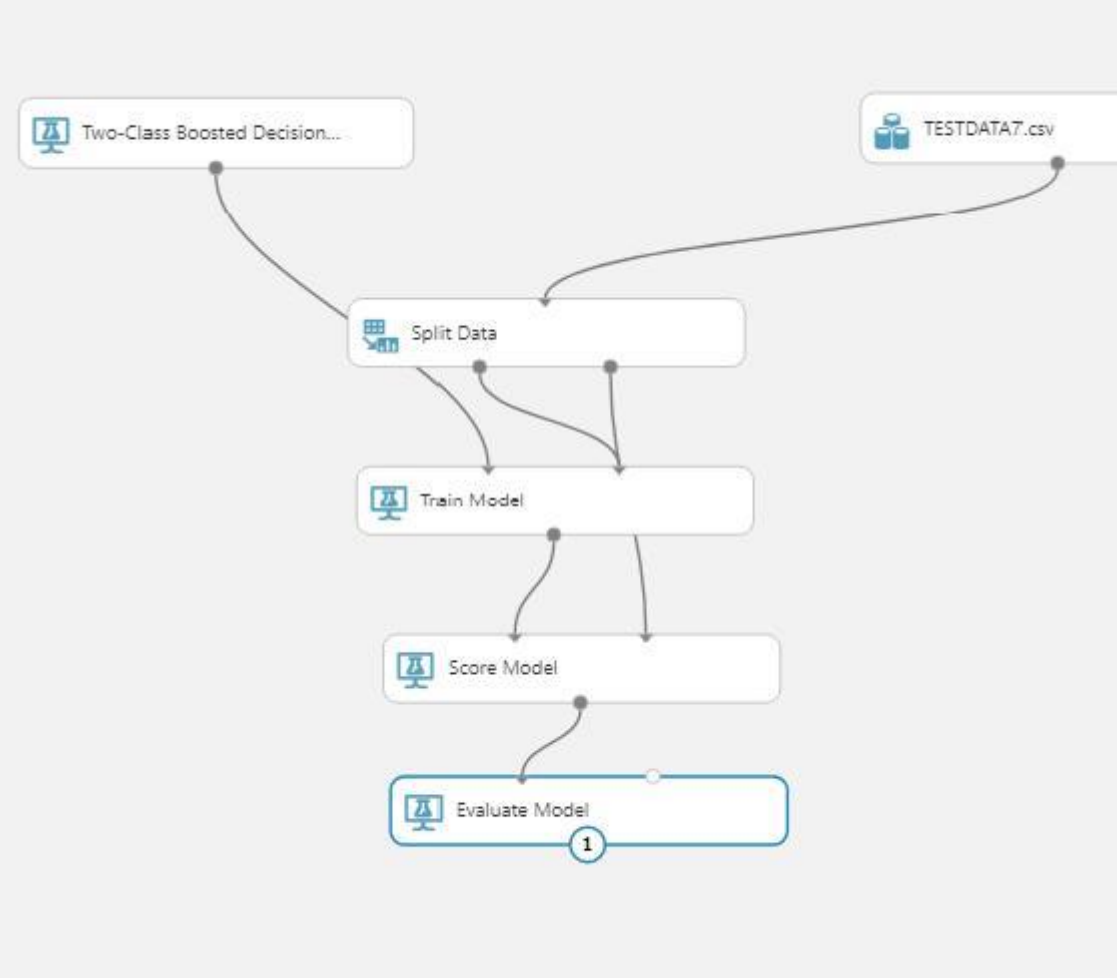

Figure 3.8: Two Class Boosted Decision Tree

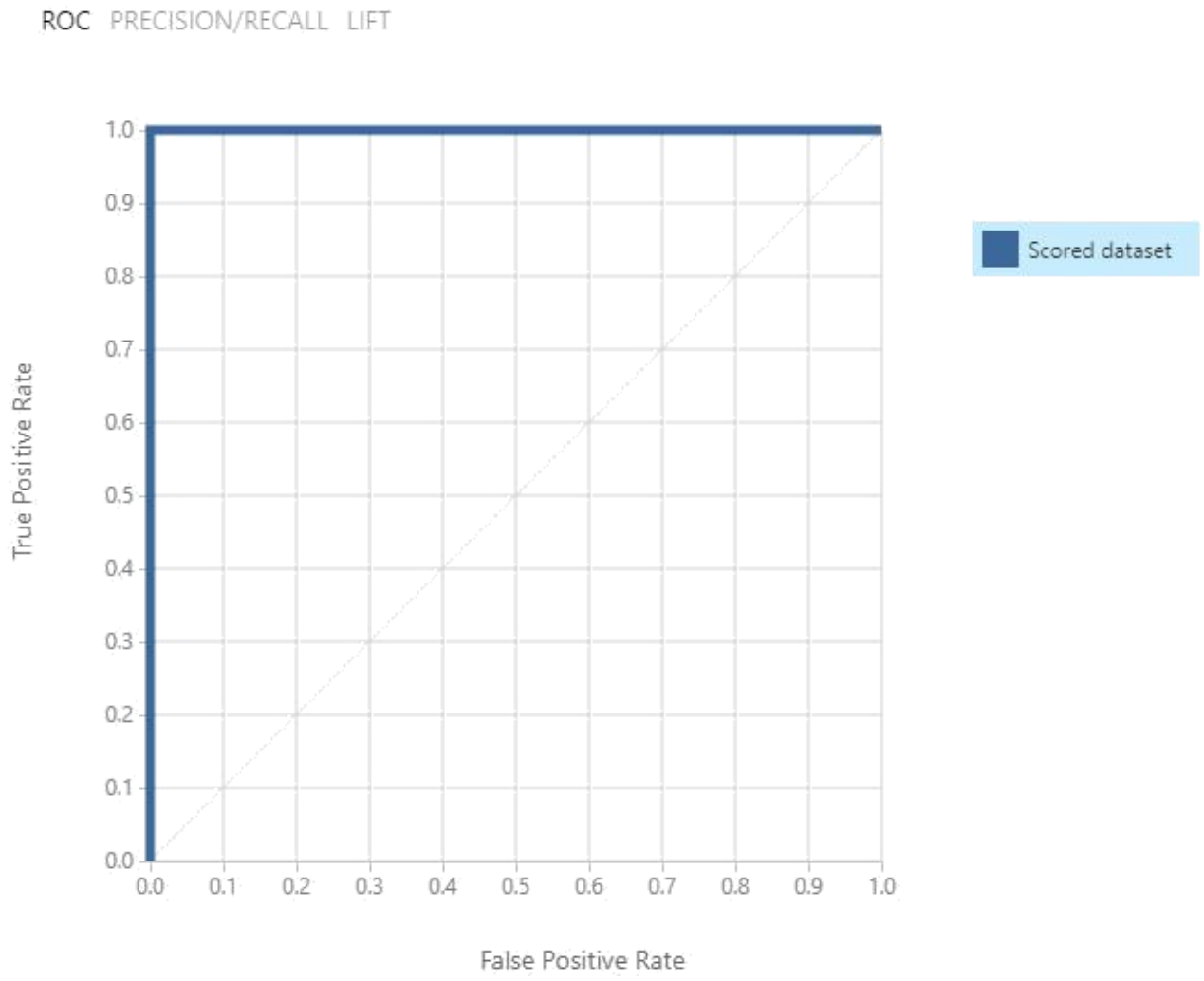

Figure 3.9: Roc curve Decision tree

Figure 3.10: Accuracy Decision Tree

3.4.3 Support Vector Machine

Next, we apply SVM in our dataset to train the model. Figure 3.11 show our model while applying SVM. Also figure 3.12 & figure 3.13 shows the ROC curve and Accuracy of SVM model.

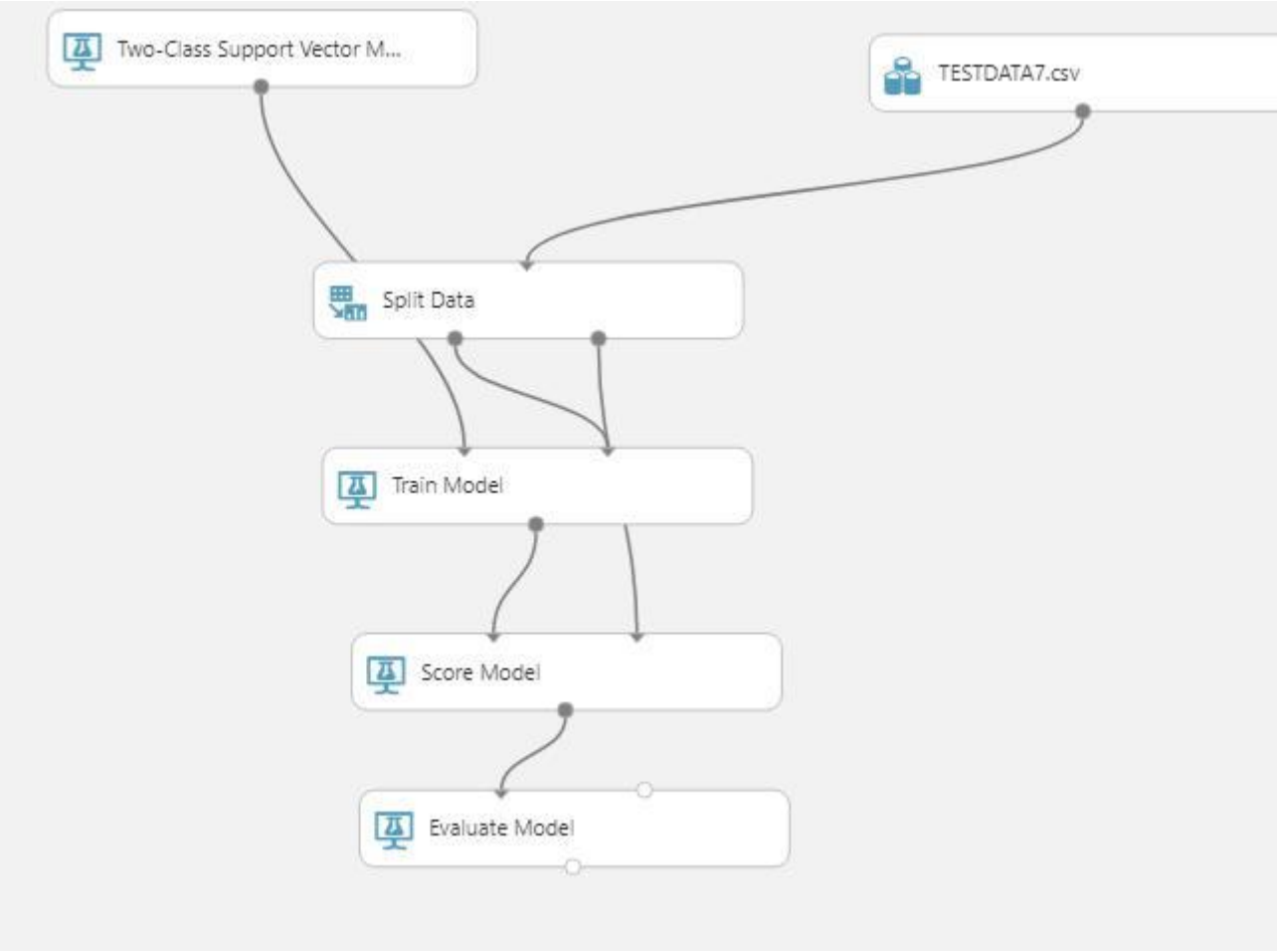

Figure 3.11: SVM model

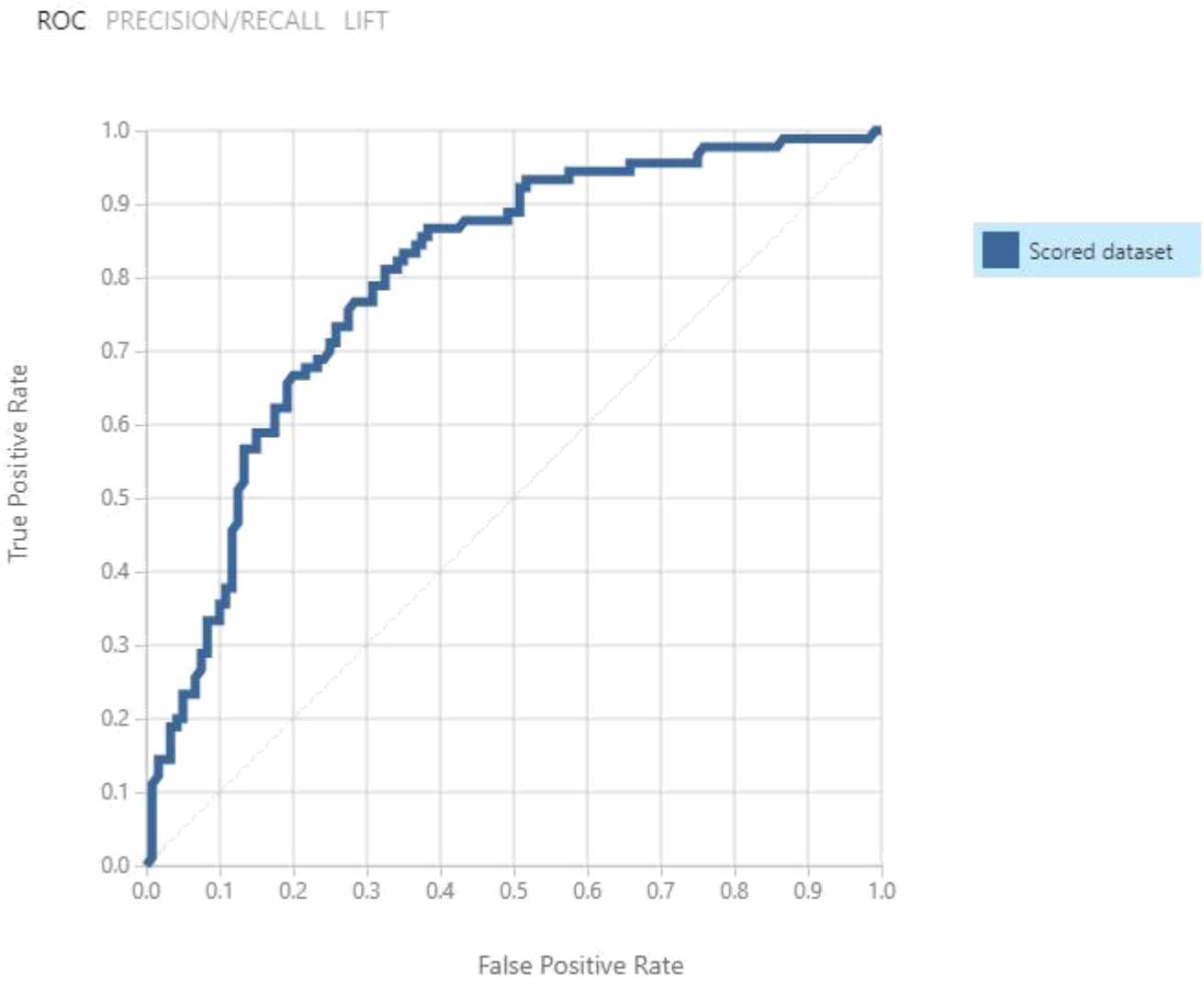

Fig 3.12: ROC SVM

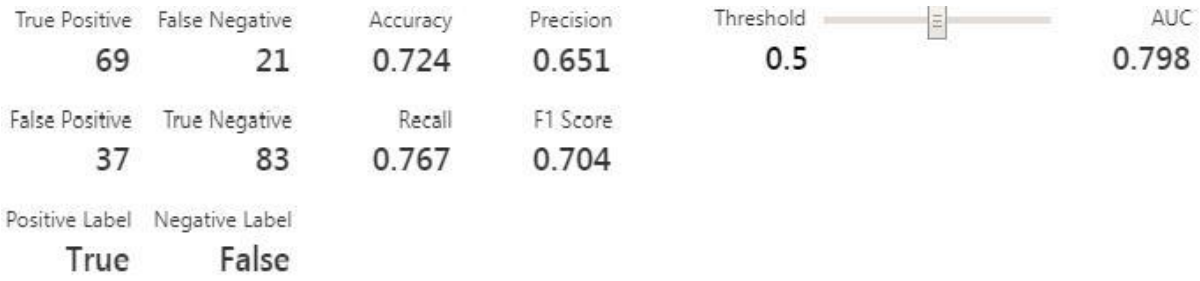

Figure 3.13: Accuracy SVM

3.4.4 Neural Network

In our model we apply neural network which is shown in figure 3.14. Also, the ROC curve and accuracy we get in figure 3.15 & figure 3.16

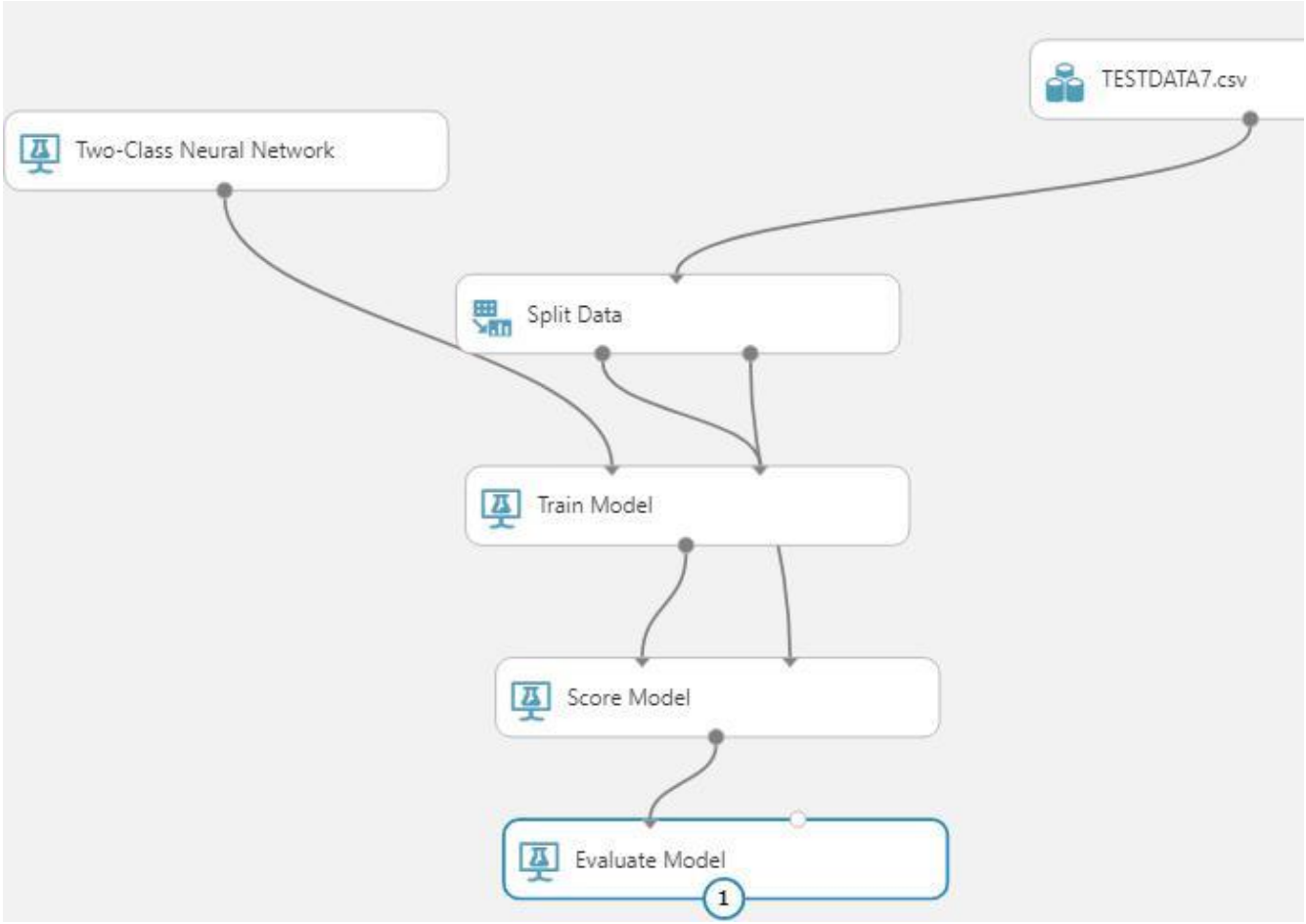

Figure 3.14: Neural Network

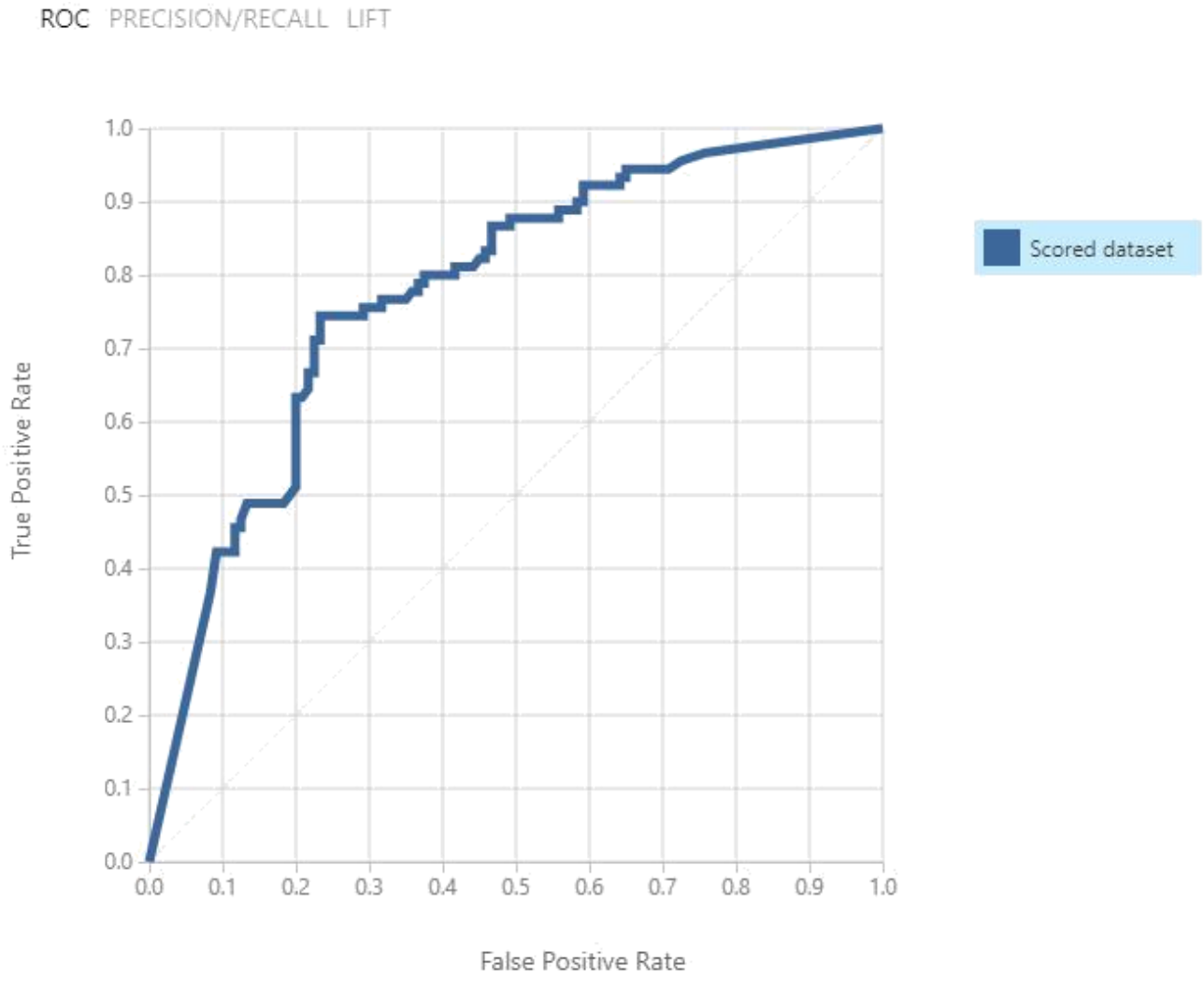

Figure 3.15: ROC Neural Network

Figure 3.16: Accuracy: Neural Network

3.4.5 Decision Forest

In figure 3.17 we show Decision forest applying in our dataset & in figure 3.18 and in figure 3.19 we get the ROC curve and Accuracy of decision forest based on our dataset.

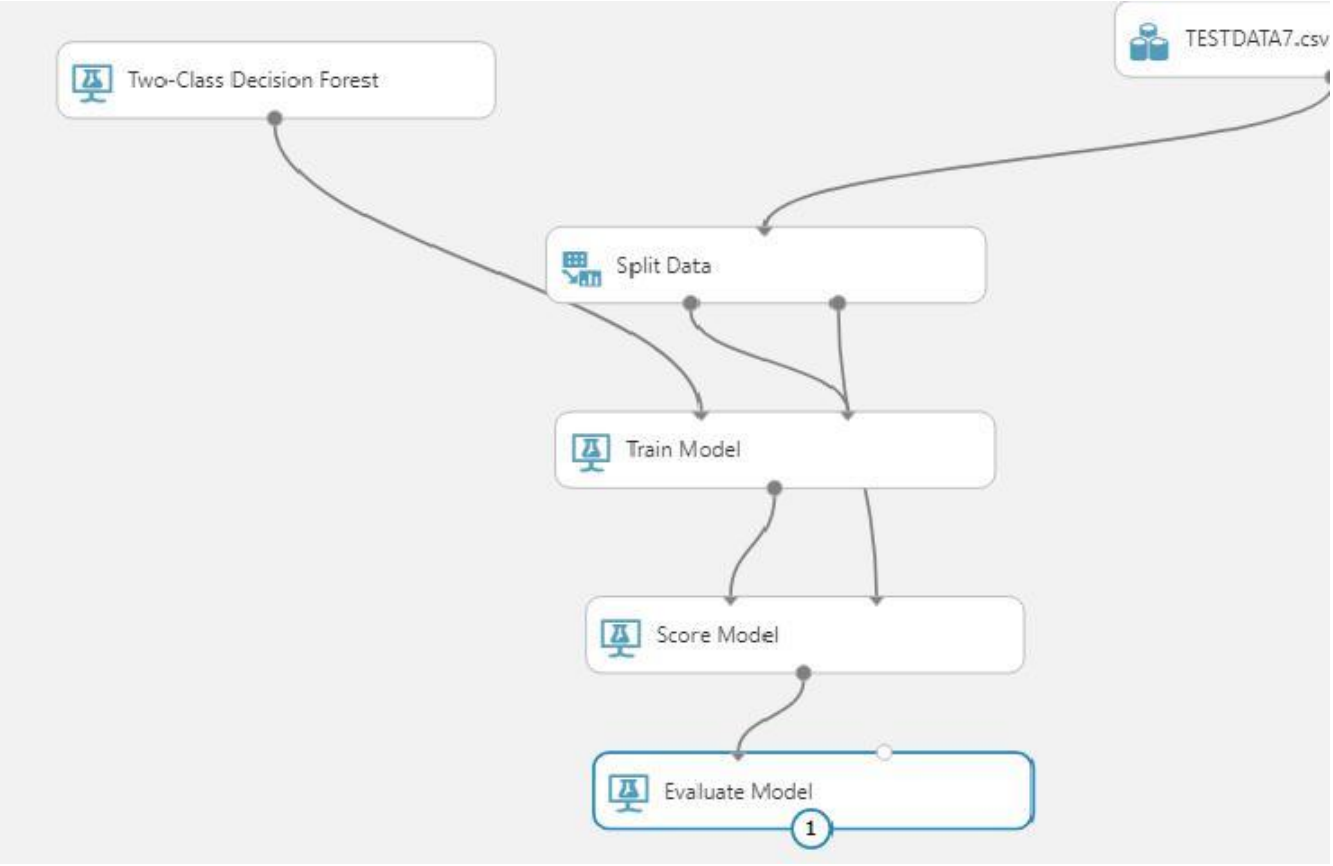

Figure 3.17: Decision Forest

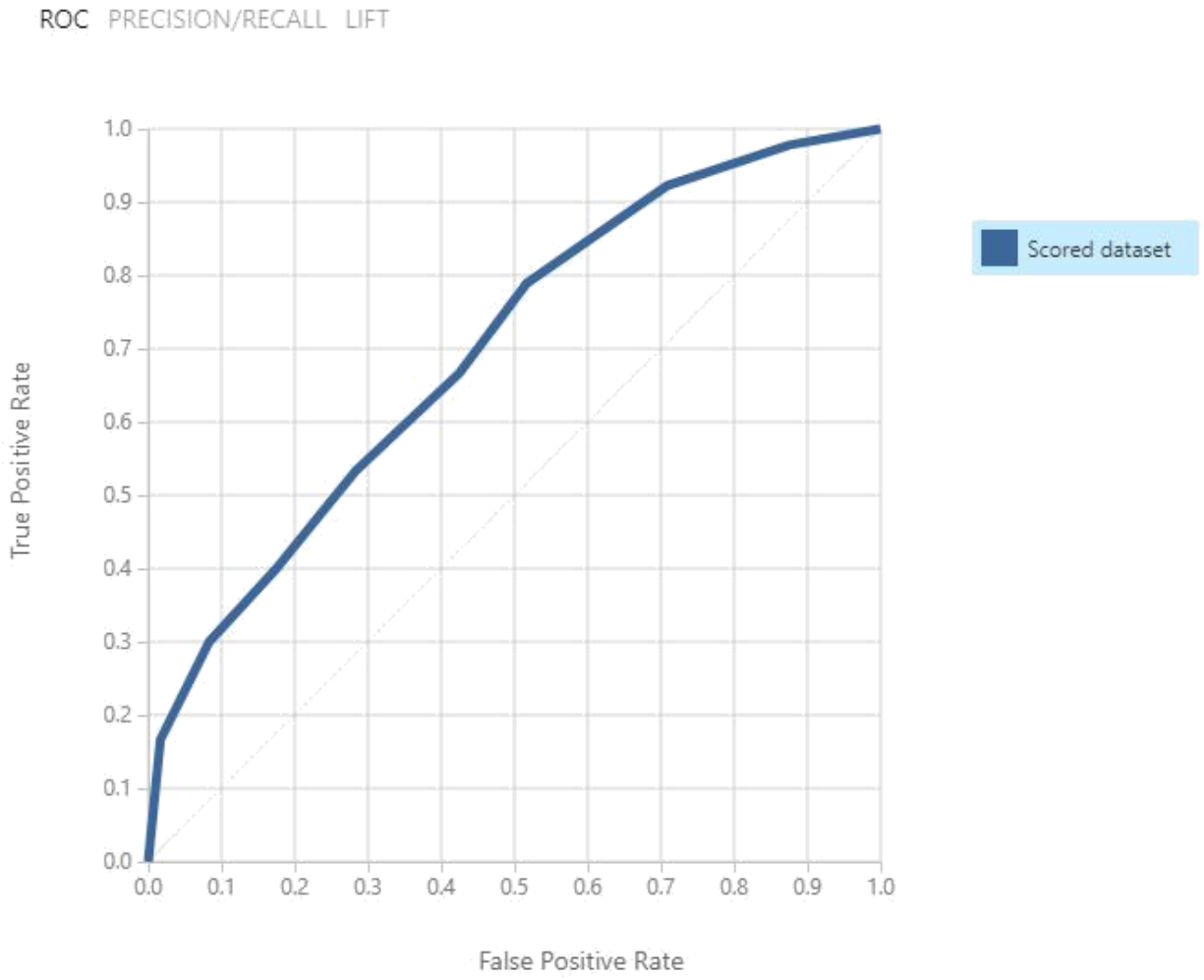

Figure 3.18: ROC Decision Forest

True Positive	False Negative	Accuracy	Precision	Threshold	AUC
48	42	0.638	0.585	0.5	0.696

False Positive	True Negative	Recall	F1 Score
34	86	0.533	0.558

Positive Label	Negative Label
True	False

Figure 3.19: Accuracy Decision Forest

3.4.6 Bayes Point Machine

Lastly, we applied Bayes Point Machine algorithm shown in figure 3.20 Also figure 3.21 & figure 3.22 gives the ROC curve and Accuracy of BPM algorithm.

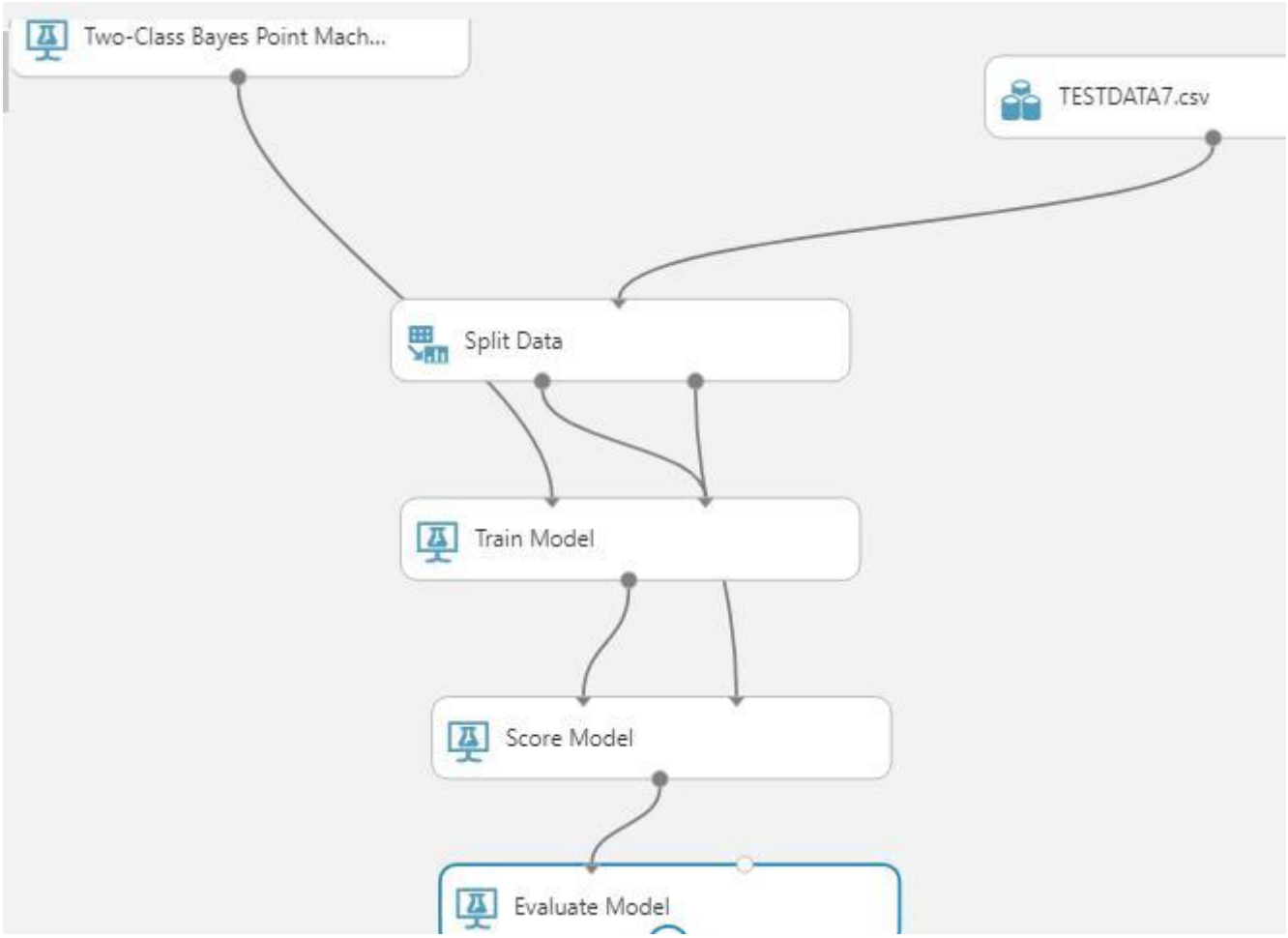

Figure 3.20: BPM MODEL

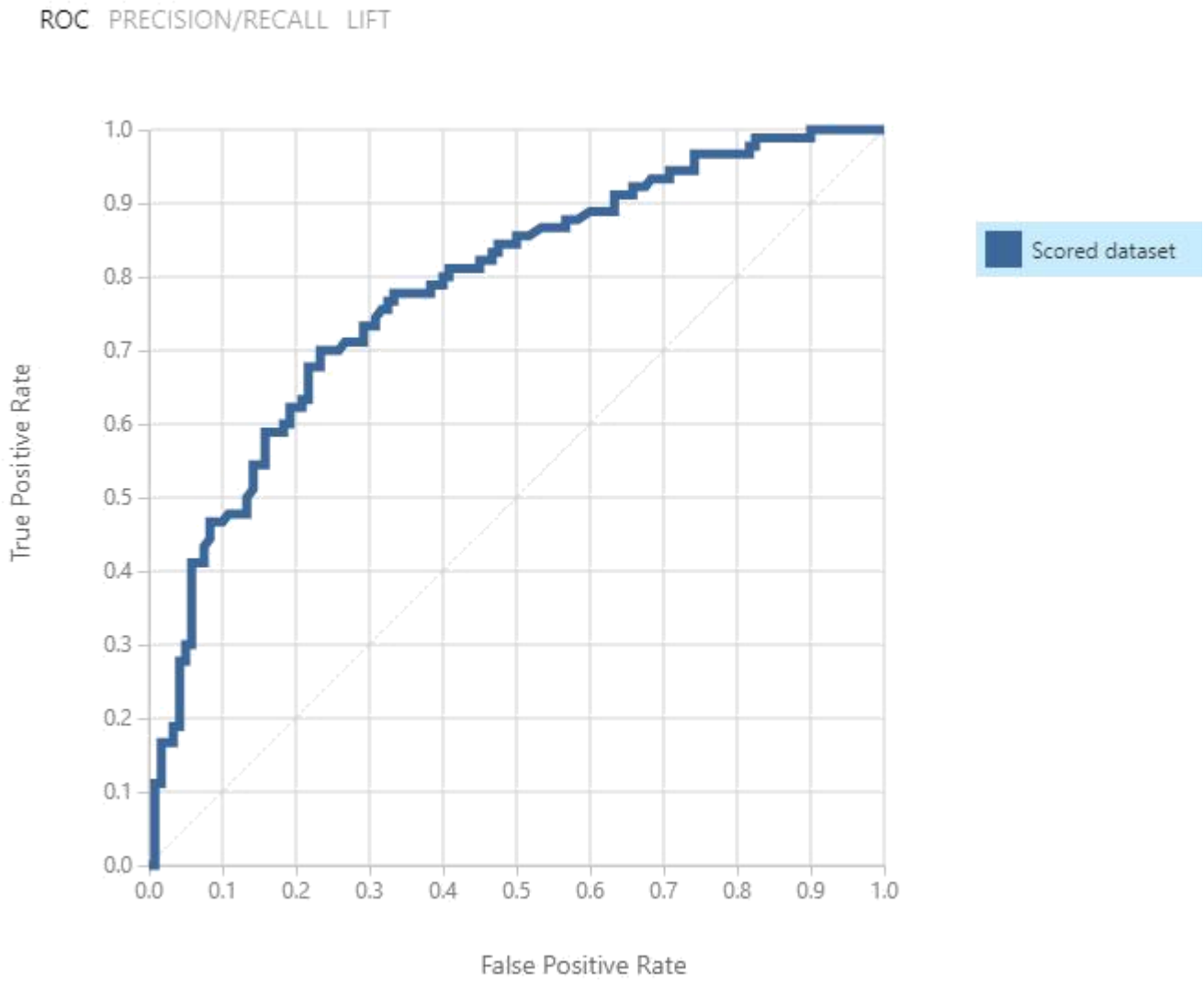

Figure 3.21: ROC BPM

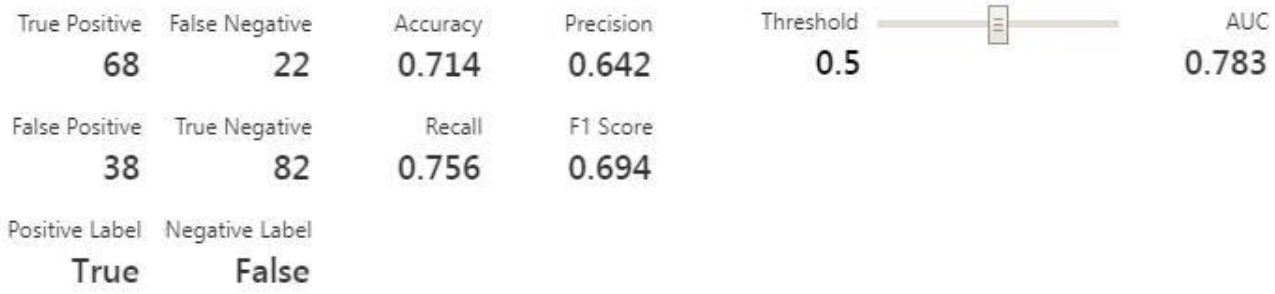

Figure 3.22: Accuracy BPM

Chapter 4

Result Analysis

At first, we need to get familiar with some term to analyze the results.

Accuracy: A number of metrics are used in ML to measure the predictive accuracy of a model. The choice of accuracy metric depends on the ML task. It is important to review these metrics to decide if our model is performing well.

The final formula appears as (measured value – accepted value) ÷ accepted value x 100 = % error.

True Positive: Sensitivity (also called the true positive rate, the recall, or probability of detection in some fields) measures the proportion of positives that are correctly identified as such.

True Negative: A false positive test result is one that detects the condition when the condition is absent.

False Positive: A test result which wrongly indicates that a particular condition or attribute is present.

False Negative: A false negative, is a test result that indicates that a condition does not hold, while in fact it does.

Precision & Recall: The precision-recall curve shows the tradeoff between precision and recall for different threshold. A high area under the curve represents both high recall and high precision, where high precision relates to a low false positive rate, and high recall relates to a low false negative rate. High scores for both show that the classifier is returning accurate results (high precision), as well as returning a majority of all positive results (high recall).

AUC: AUC is an abbreviation for area under the curve. It is used in classification analysis in order to determine which of the used models predicts the classes best. An example of its application is ROC curves. Here, the true positive rates are plotted against false positive rates

ROC: The ROC curve is a fundamental tool for diagnostic test evaluation. In a ROC curve the true positive rate (Sensitivity) is plotted in function of the false positive rate (100-Specificity) for different cut-off points of a parameter.

We get different accuracy level for different algorithms and choose the algorithm with highest accuracy. Table 4.1 shows the accuracy of different algorithm.

Table 4.1: Accuracy of different Algorithms

Algorithm	Accuracy
1. Logistic Regression	82.8%
2. Decision tree	91.1%%
3. SVM	79.8%
4. Neural Network	77.4%
5. Decision Forest	69.6%
6. BPM	78.3%

To compare between different algorithms and measure them by their ROC curve we need to create a comparing model. In figure 4.1 we show model for different training algorithm. In figure 4.2, figure 4.3 and in figure 4.4 we can see the comparison of ROC curve among Logistic regression, decision tree and neural network.

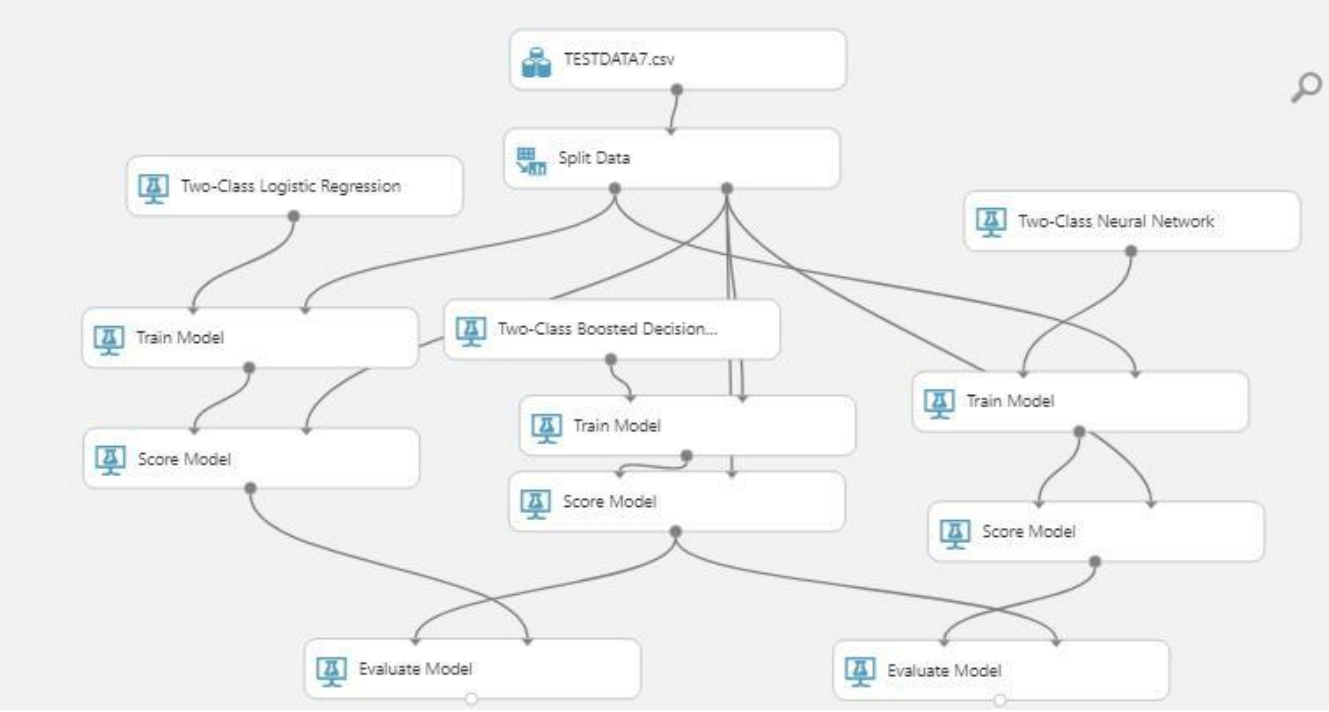

Figure 4.1: Compare between Algorithm

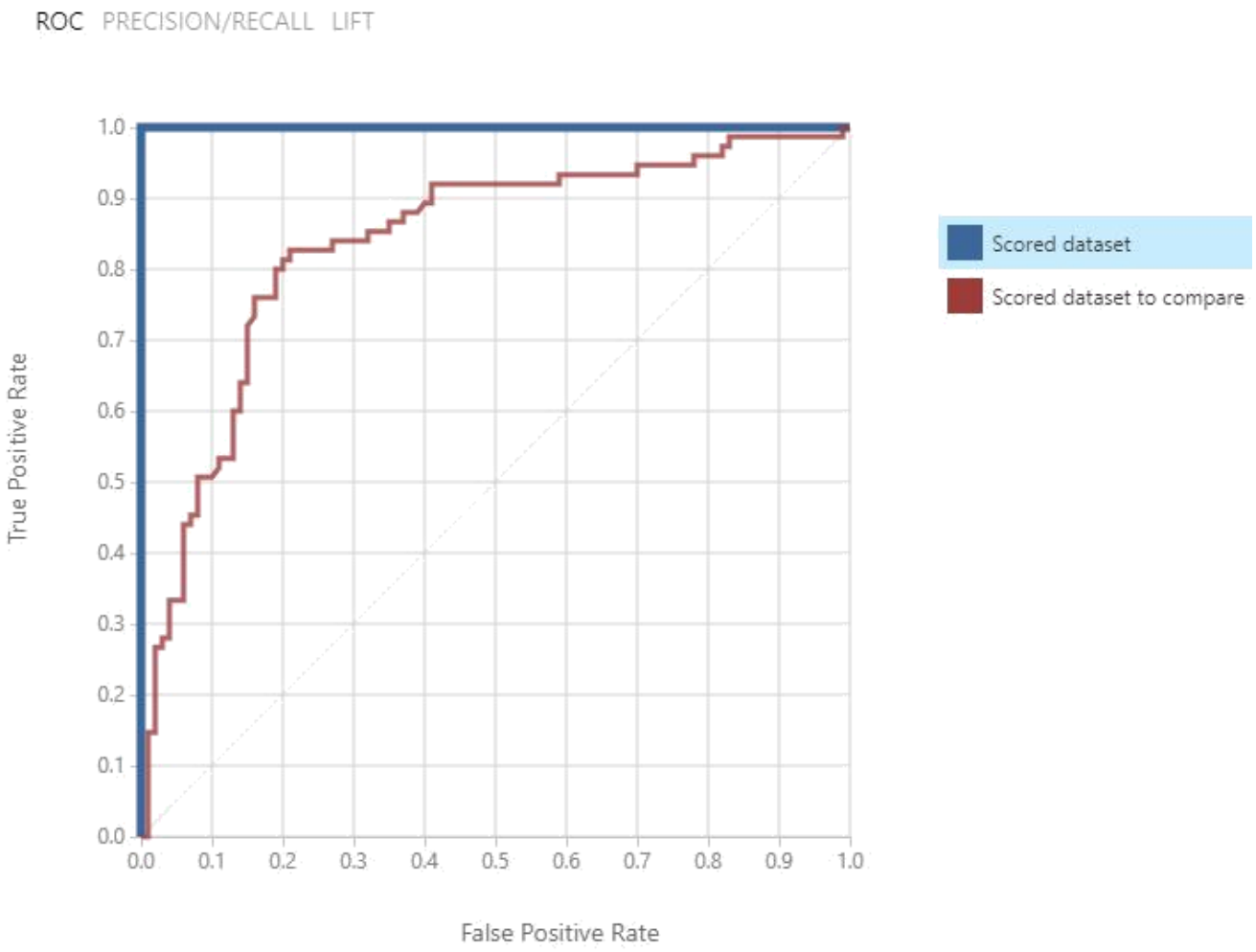

Figure 4.2: Blue: Decision Tree; RED: Logistic Regression

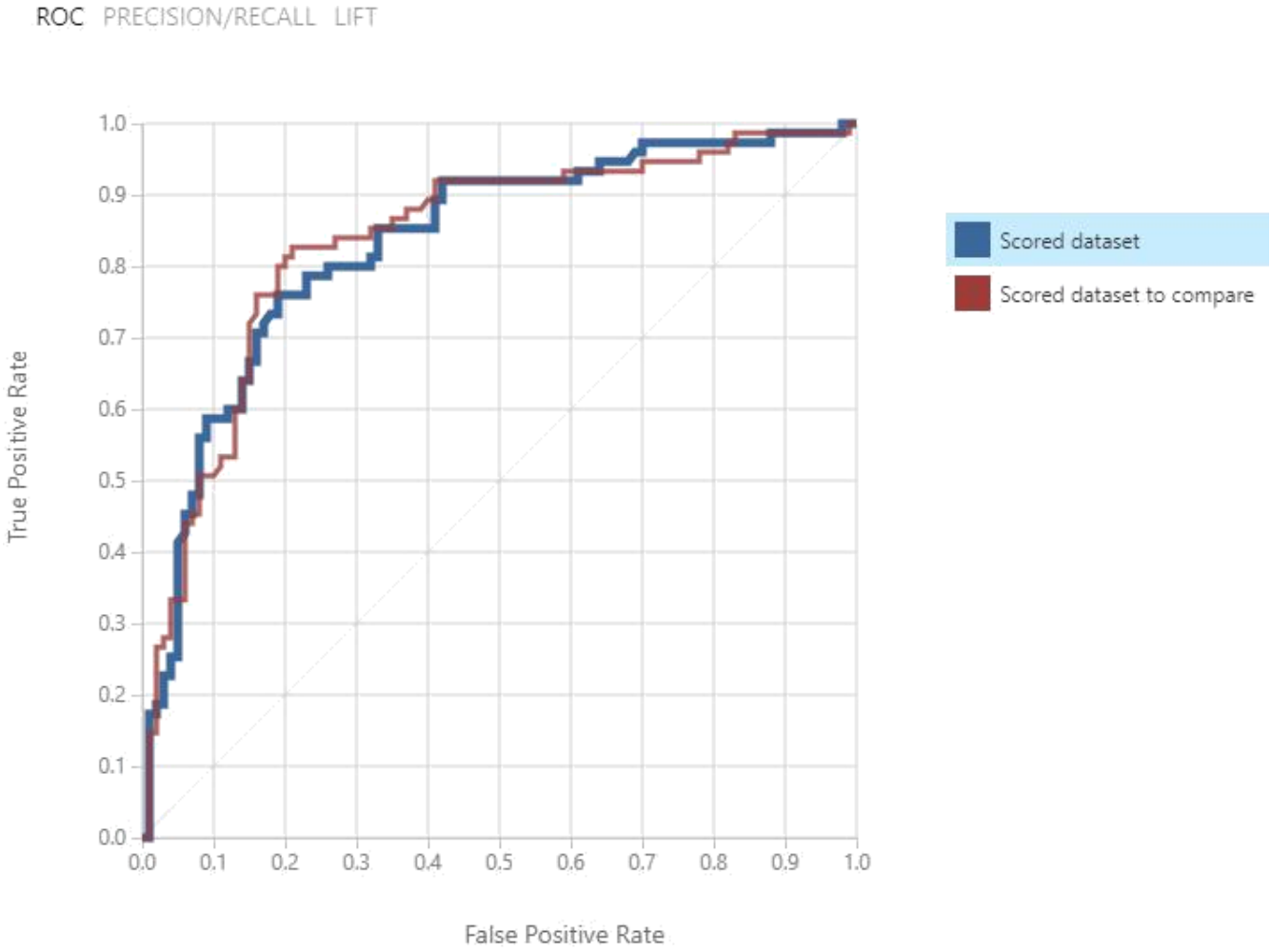

Figure 4.3: Blue: SVM; Red: Logistic Regression

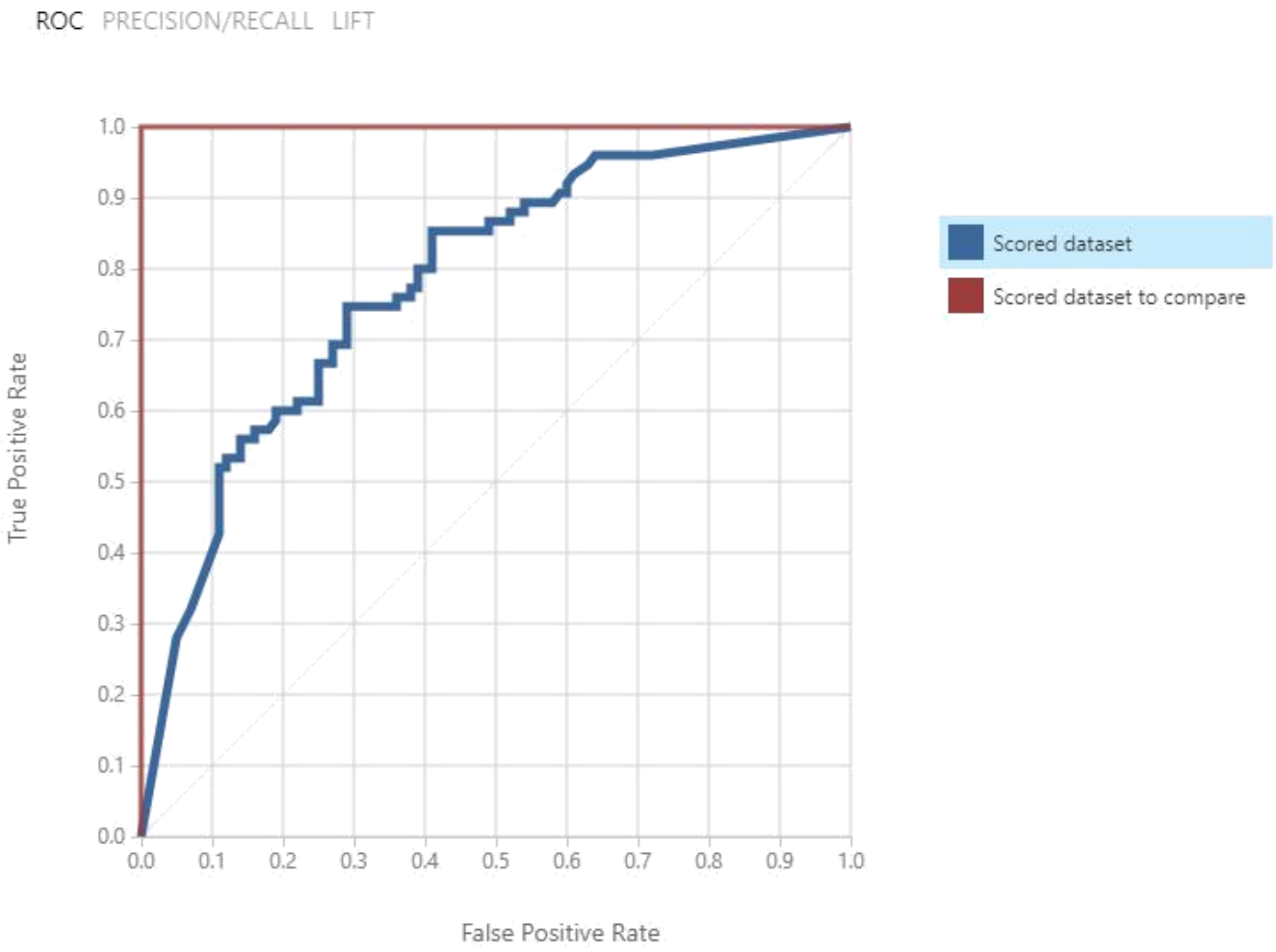

Figure 4.4: Blue: Neural Network; Red: Decision Tree

By comparing ROC curve and accuracy of different algorithm we can see that in every case decision tree gives the highest accuracy. Therefore, we finally choose decision tree to train our model. To show how our algorithm predicts, we deploy our trained model and allow user to give input. Based on the input our model gives an output.

Chapter 05

Conclusion

For concluding we would like to say that this system will make a huge impact on the job field. It will lessen the struggle of graduates and bring out the proper opportunity. With some improvement and change in the parameter we will be able to use our algorithm in different types of scenario for different type of purpose.

Future Work

We will be able to find hidden opportunities in different sectors. We will also be able to create opportunity, like suggesting graduates that which skills the employers are looking for by analyzing data we learnt. Also, in many other concepts such as predicting upcoming openings, tentative salary, etc.

References

[1] A.L. Samuel (1959) *Some Studies in Machine Learning Using the Game of Checkers*. IBM Journal of Research and Development, United States of America

[2]Tom M. Mitchel (1997). *Machine Learning.* McGraw hill science/engineering, United states of America.

[3]Menard , S. (2002) . *Applied Logistic Regression Analysis* . Sage University . Cali-fornia , United States of America.

[4]Coadou , y. (2016) , *Boosted decision trees* , School of Statistics , Autrans , France.

[5] Freund , y Mason , L (2008) , *The alternating decision tree learning algorithm* , Australian National University , ACT , Australia .

[6] Leo Breiman .(2006) *Bias , Variance and arcing classifiers technical report 460* , statistical department , university of California , Berkley , United States of America .

[7] C. J.C. Burges.(1998). *A tutorial on support vector machines for pattern recognition. Data Mining and Knowledge Discovery,* 2:121–167.

[8] Simon , T (2001) , *Support Vector Machine Active Learning with Applications to Text Classification, Computer Science Department Stanford University , Stanford CA 94305-9010, USA*

[9] A.K. Jain (1996 *), Artificial neural networks* , IEEE , IEEE Computer Society .

[10] S. Haykin .(1994), Neural Networks: *A Comprehensive Foundation*, Cambridge University, UK.

[11] E. Levin, N. Tishby, S. Solla,(1990) "*A statistical approach to learning and gen-eralization in layered neural networks"*, Proc. IEEE.

[12] R. P. Lippmann, (1987)"*An introduction to computing with neural nets*", IEEE ASSP Mag., pp. 4-22.

[13] Herbrich , R (2001) , *Bayes Point Machines* , Technical University of Berlin, Franklinstr. 28/29, 10587 Berlin, Germany .

[14] W. Buntine.(1992) *A Theory of Learning Classification Rules*. PhD thesis, University of Technology, Sydney, Australia.

[15] Edelman, G. M. (1987). *Neural Darwinism: The theory of neuronal group selection*. New York: Basic Books.

[16] Goldberg, D. E. (1989). *Genetic algorithms in search, optimization, and machine learning*. Reading, MA: Addison-Wesley.

[17] Goldberg, D.E.(1988) , *Genetic Algorithms and Machine Learning*, University of Alabama , Tuscaloosa , USA.

[18] Han, J. Kamber, M. Pei, J. (2012). *Data Mining: Concepts and Techniques.* Wal-tham, Elsevier Inc.

[19] Michalski, S. Bratko, I. Puget, J. (1998). *Machine Learning and Data Mining; Methods and Applications.* New York, John Wiley & Sons, Inc.

[20] Gelman, A. Carlin, J. Stern, H. Rubin, D. (2003). *Bayesian Data Analysis*, Florida, Chapman & Hall.

[21]Quinlan, J. Mach Learn (1986). *Induction of Decision Trees.* Kluwer Academic Publishers-Plenum Publishers.

[22] Pujari, A. (2001). *Data Mining Techniques.* Hyderabad, India. Universities press private ltd.

[23] Linoff, G. Berry, M. (2011). *Data Mining Techniques for Marketing, Sales and Customer Relationship Management.* Indiana. Wiley Publishing Inc.

[24]Sammut, C. Webb, G. (2011). *Encyclopedia of Machine Learning*. New York, Springer.