

Performance Analysis of Support Vector Machine Linear Kernel towards Machine Learning

Tafzil Anjum Ansari¹ Dr. Sanjay Thakur²

¹M.Tech. Student ²Professor

^{1,2}Department of Computer Science and Engineering

^{1,2}L.N.C.T. (Bhopal) Indore Campus Indore, India

Abstract— Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for classification as well as regression problems. The important concepts in SVM includes Support Vectors, data points that are closest to the hyper plane is called support vectors. Separating line will be defined with the help of these data points. Margin, it may be defined as the gap between two lines on the closet data points of different classes. It can be calculated as the perpendicular distance from the line to the support vectors. Large margin is considered as a good margin and small margin is considered as a bad margin. There are some problem needs to consider, while implementing SVM for a data set. There are multiple hyper planes that separate the data exists, which one to choose. Target classes are overlapping. In cases where the number of features for each data point exceeds the number of training data samples. In this paper we used linear kernel based SVM and develop to models first model used heart disease data with 14 attribute. Last attribute is classify the records based on SVM model that patient belong to yes or no class. In second model we used Social data set which has five attribute customers ID, Gender, Age Salary, purchased, the last class is a binary class. Based on salary and age we need to classify the into purchased class (Yes or No). We used R language to implement both the model. We divided the data set into training and testing data set. First we create model using training data set. After that we apply the testing data set. We create Confusion Matrix for both models after allaying testing data set to find out the accuracy of the models. We calculate the accuracy and use graph to the show the number of records classifies correctly by both model. Real life data set is used in both the models.

Keywords: Support Vectors, Supervised Learning, Margin, Hyper Planes, Kernel, Confusion Matrix

I. INTRODUCTION

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for classification as well as regression problems. SVM was developed by Vladimir Vapnik in the 1970s. It was developed as part of a bet where Vapnik envisaged that coming up with a decision boundary that tries to maximize the margin between the two classes will give great results and overcome the problem of over fitting. Primarily it is used for classification problems in Machine Learning. The goal of the SVM algorithm is to create the best line or decision boundary that can segregate n-dimensional space into classes so that can easily put the new data point in the correct category in the future. This best decision boundary is called a hyper plane. SVM chooses the extreme points/vectors that help in creating the hyper plane. These extreme cases are called as support vectors, and hence algorithm is termed as Support Vector Machine. Being a linear algorithm at its core can be imagined

almost like a Linear or Logistic Regression. For example, an SVM classifier creates a line (plane or hyper-plane, depending upon the dimensionality of the data) in an N-dimensional space to classify data points that belong to two separate classes. The original SVM classifier had this objective and was originally designed to solve binary classification problems, linear regression that uses the concept of line of best fit, which is the predictive line that gives the minimum Sum of Squared Error (if using OLS Regression), or Logistic Regression that uses Maximum Likelihood Estimation to find the best fitting sigmoid curve, Support Vector Machines uses the concept of Margins to come up with predictions[1,2,3].

II. HOW DOES SUPPORT VECTOR MACHINE ALGORITHM WORK

Support vector machines are among supervised machine learning algorithms, which means they need to be trained on labeled data. To understand support vector machines, consider the following dataset, composed of a set of points from two different classes (yellow and blue)[4,5,6].

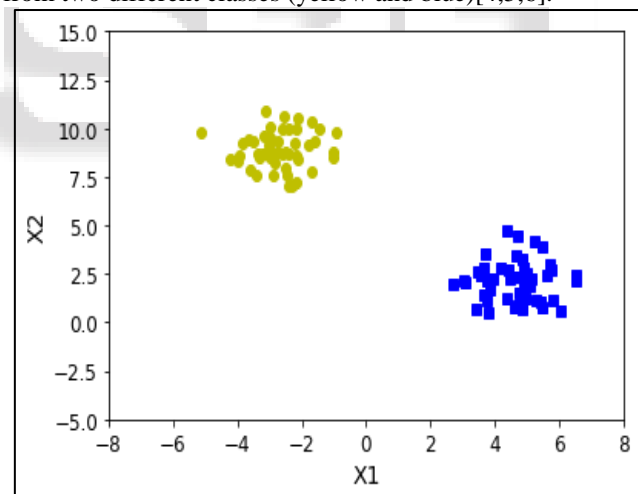


Fig. 1: Set of points from two different classes (yellow and blue)

A machine learning model can learn to classify the points into the two classes through linear separation.

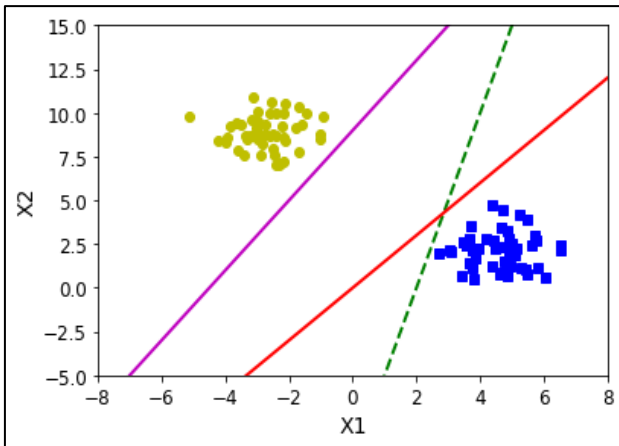


Fig. 2: Classify the points into the two classes through linear separation

However, we can create many different ML models to separate the two classes, and not all of them are of equal value. For example, if draw the line too close to one of the classes, model will become sensitive to small deviations from the norm on that class. Therefore, if a new point is a bit farther from the cluster used during training, the ML model will misclassify it.

SVMs use “support vectors” to find optimal solutions to classification problems. Support vectors (dashed lines in the figure below) are at equal distance of the main classification line (solid line). By placing the support vectors on the edge instances of each class, the SVM creates an ML model that is robust to unseen data that do not closely fit the training examples.

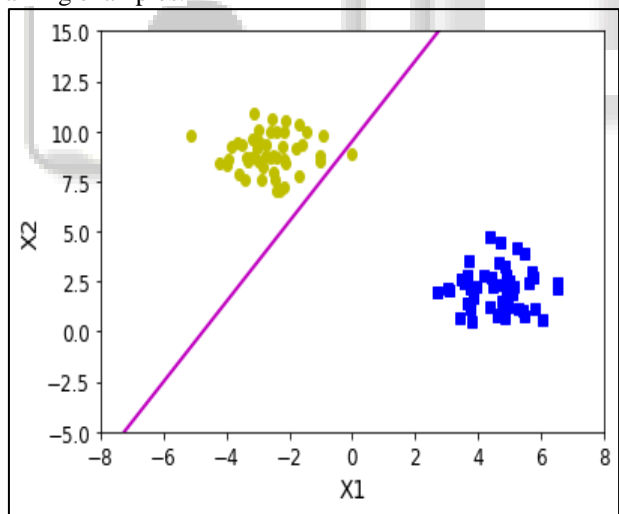


Fig. 1.3: ML model do not closely fit the training sample
The followings are important concepts in SVM –

- Support Vectors – Data points that are closest to the hyper plane is called support vectors. Separating line will be defined with the help of these data points.
- Hyper plane – As we can see in the above diagram, it is a decision plane or space which is divided between a set of objects having different classes.
- Margin – It may be defined as the gap between two lines on the closet data points of different classes. It can be calculated as the perpendicular distance from the line to the support vectors. Large margin is considered as a good margin and small margin is considered as a bad margin.

III. SVM MARGINS

Margin can be defined using the minimum distance (normal distance) from each observation to a given separating hyper plane. Let’s see how we can use Margin to find optimal hyper plane. The size of the Margin defines the confidence of the classifier; hence the widest margin is preferable. Let’s select two Hyper planes, based on the margin. The first one has much wider margin than the 2nd one, hence the first Hyper plane is more optimal than 2nd one[17,18].

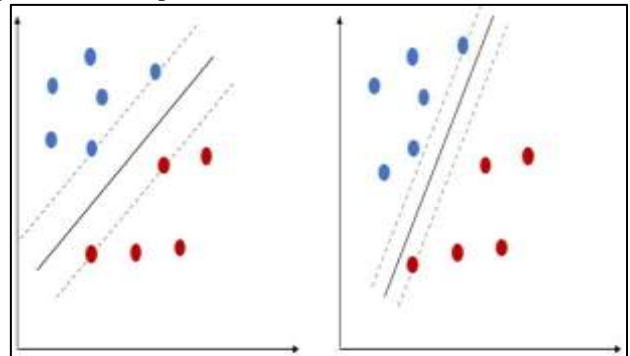


Fig. 4: First one has wider margin than the 2nd and more optimal than 2nd one

Finally, we can say, in Maximal Margin Classifier, in order to classify the data, we will use a separating hyper plane, which has the farthest (max) minimum (min) distance from the observations. The Margin to select the optimal separating hyper plane.

A. Hard and soft margins

SVMs can have hard and soft margins. If we think as support vectors and the main classification line as a street, a hard margin SVM will try to place all instances off or at the edge of the street. This is not a bad strategy if the instances of all classes are divided into neat clusters. But if the data contains outliers, then the SVM will draw a bad boundary. And if there is overlap between the classes, then a hard-margin SVM won’t work at all[19,20,21].

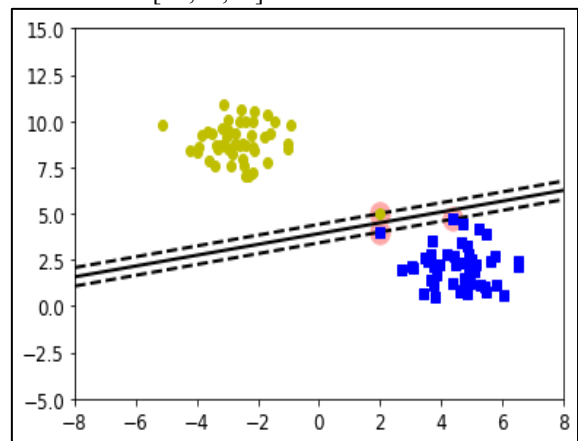


Fig. 5: Overlap between the classes, then a hard-margin SVM

To address these problems, SVMs support “soft margins,” a hyper parameter that can be adjusted before training the model. Soft margins allow a number of instances to violate the support vector boundaries to choose a better classification line. The lower the soft margin number (usually specified by the Greek letter zeta, or “C” in programming

libraries), the more boundary violations the model allows and the wider the SVM Street becomes. Choosing the right soft margin value will ensure the SVM will have the right balance between clean boundaries and generalizability. In the example below, even though one outlier in the training data has been misclassified, the soft margin has resulted in a better classification line.

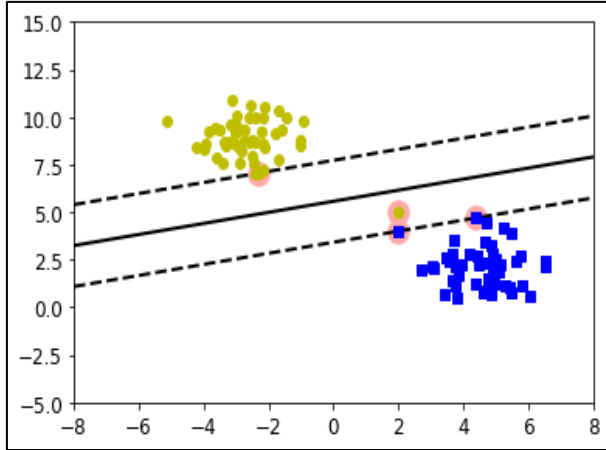


Fig. 6: Soft margins violate the support vector boundaries

IV. LITERATURE SURVEY

In 2015 Dr. S. Vijayarani et al proposed “Kidney Disease Prediction Using SVM and ANN Algorithm”. Advanced data mining techniques can help and provides remedy to handle this situation. Data mining is the process of extraction of hidden information from the voluminous dataset. Various data mining techniques are clustering, classification, association analysis, regression, summarization, time series analysis and sequence analysis, etc. Their objective was to predict kidney diseases by using Support Vector Machine (SVM) and Artificial Neural Network (ANN). They compared the performance of these two algorithms on the basis of its accuracy and execution time. From the experimental results it is observed that the performance of the ANN is better than the other algorithm. They worked for classification process to classify four types of kidney diseases. Comparisons of Support Vector Machine (SVM) and Artificial Neural Network (ANN) algorithms are done based on the performance factors classification accuracy and execution time. They concluded that the ANN achieves increased classification performance, yields results that are accurate; hence it is considered as best classifier when compared with SVM classifier algorithm. Feasibly, SVM classifier classifies the data with minimum execution time [6].

In 2016 Nick Guenther et al proposed “Support vector machines are statistical and machine learning techniques with the primary goal of prediction”. They can be applied to continuous, binary, and categorical outcomes analogous to Gaussian, logistic, and multinomial regression. They introduced a new command for this purpose, svmachines. This package is a thin wrapper for the widely deployed labs. They illustrated svmachines with two examples. SVM can be a powerful tool, but with great power comes great responsibility. Several issues should be remembered: First, the tuning of parameters is very

important. Alternatively, computer scientists tend to standardize the variables to a range such as. Either method should work fine.. A linear kernel might be suitable when both the number of observations and the number of variables are very large (Hsu, Chang, and Lin 2003). This problem occurs, for example, in document classification. SVM is among the best-performing statistical-learning or machine-learning algorithms in terms of accuracy but there is no guarantee it will perform better than other learning techniques. Particularly for binary outcomes, logistic regression and SVM will tend to give similar results because their implied loss function is similar [7].

In 2017 Nurhanna, A. A. et al proposed “Multi-class Support Vector Machine Application in the Field of Agriculture and Poultry: A Review”. They discussed and survey the implementation of multiclass support vector machine algorithms in the field of agriculture and poultry. They summarized the method used, their performance and accuracy to facilitate the process of choosing the best method for agriculture area, which involves other areas of interest. They discussed the important role and contributions of multiclass Support Vector Machine (SVM) in various research areas and mostly in the agriculture and poultry industries. Most of the findings show the most important characteristic of multiclass SVM which is its ability to generalize well even with a small dataset. In most cases, the selection of the feature extraction method and preprocessing of the data provide a higher contribution in the classification result. They have also found out that there are not many types of research done that demonstrate SVM capabilities in classifying texture images with similar background and foreground color[8].

In 2018 Shujun Huang proposed “Applications of Support Vector Machine (SVM) Learning in Cancer Genomics”. As advancements in high-throughput technologies lead to production of large amounts of genomic and epigenetic data, the classification feature of SVMs is expanding its use in cancer genomics, leading to the discovery of new biomarkers, new drug targets, and a better understanding of cancer driver genes. They reviewed the recent progress of SVMs in cancer genomic studies. We intend to comprehend the strength of the SVM learning and its future perspective in cancer genomic applications. Although SVMs with non-linear kernels are extremely powerful classifiers, they do have some downsides as following, finding the best model requires testing of various combinations of kernels and model parameters. It can be slow to train, particularly if the input dataset has a large number of features or examples. Their inner workings can be difficult to understand because the underlying models are based on complex mathematical systems and the results are difficult to interpret..The cancer genomic and epigenetic data are exponentially increased as the new generation of sequencing technologies advances[9].

In 2019 Rodrigo P. Monteiro et al proposed “Support Vector Machine Based Decision Stage to Improve the Fault Diagnosis on Gearboxes”. Rise of powerful graphical processing units spreads the use of deep learning-based solutions to many problems, which includes the fault diagnosis on gearboxes. Solutions usually require a significant amount of data, high computational power, and a

long training process. Training of deep learning-based systems may not be feasible when GPUs are not available. They proposed a solution to reduce the training time of deep learning-based fault diagnosis systems without compromising their accuracy. Solution is based on the use of decision stage to interpret all the probability outputs of a classifier whose output layer has the softmax activation function. Two classification algorithms were applied to perform the decision. They reduced the training time by almost 80% without compromising the average accuracy of the fault diagnosis system. They analyzed the use of a decision stage to interpret the outputs of a fault diagnosis system based on deep convolution neural networks. They analyzed the output distribution of the deep classifier to perform more reliable fault diagnosis[10].

In 2019 Sunil Kumar et al proposed “Rice Yield Forecasting Using Support Vector Machine”. Support Vector Machines (SVMs) have acquired considerable significance. These are widely used in making predictions, owing to their ability of generalization. They develop of SVM based classification models for the prediction of rice yield in India. MATLAB software has been used for experimentation in this work. In the past, a good amount of work has been done on the forecasting of rice yield. The development of soft computing models that provide an exact or even a good approximation to crop yield for next year will immensely contribute to the encouragement of farmers and other interested holders. These models will improve the quality of decision making in connection to domestic sustenance and international trade. The importance of rice in India and world cannot be underestimated. The efficacy of SVM in predicting the yield of rice crop has been demonstrated in this paper. SVM-based classification models have been developed for the forecasting of rice yield in India. SVM classification models have been tested using 3-fold, 4-fold and 5-fold cross validation methods, one-against-one multi classification method [11].

In 2019 Kai-Xu Han et al proposed “Application of Support Vector Machine (SVM) in the Sentiment Analysis of Twitter Data Set”. The Fisher kernel function based on the model is derived from the Probabilistic Latent Semantic Analysis model. By means of this method, latent semantic information involving the probability characteristics can be used as the classification characteristics, along with the improvement of the effect of classification for support vector machine, and the problem of ignoring the latent semantic characteristics in text sentiment analysis can be addressed. The results show that the effect of the method proposed in this paper, compared with the comparison method, is obviously improved. The Fisher kernel functions derived based on the probabilistic latent semantic analysis model, which fully takes the probability relationship and latent relationship among text, vocabulary and subject into consideration. The support vector machine, based on Fisher kernel function, can classify the text sentiment from the probability level of the generated model, such as the latent semantic level [12].

In 2020 A. Ravi Kumar et al proposed “Support Vector Machine for Text Categorization using Principle Component Analysis in Data Mining”. The major issues in text categorization are classification accuracy and computation time. To overcome these issues, an efficient

classification method is needed for high differentiation exactness as fine as minimizing the computation period. They proposed the classification of data using support vector machine for text categorization along with principle component analysis. Principle Component Analysis (PCA) is the feature removal technique is used towards mine the features within the text. Chi-Square is a further assortment technique it is used to selecting the features from removed features. Proposed work, the classification accuracy also computation period is improved than other existing algorithms in many applications They used use principle/component/ analysis/(P/C/A) as a feature removal method to decrease the great dimensionality of the dataset. PCA eliminates the irrelevant, noisy and redundant parameters commencing the feature vector also thus develop the presentation of the classifier for text classification. They processed s the documents where the feature vector is attained over dissimilar steps similar stop words elimination, stopping also indexing. [13].

In 2020 Soham Guhathakurata et al “A novel approaches to predict COVID-19 using support vector machine”. They proposed as study in which ongoing outbreak of Corona virus disease (COVID-19) caused by severe acute respiratory syndrome Corona virus 2 was first reported in December 2019 in Wuhan City, Hubei province, China. They explained the disease has spread globally at an exponential rate infecting more than 22, 00,000 and killing more than 1, 53,000 and still counting. The Study has shown that transmission takes place mainly via respiratory droplets and close contacts. The virus can survive for 72 h on the surface. A person infected with COVID-19 generally starts to show signs and symptoms within 5e6 days after the infection. The occurrence of the first symptom of COVID-19 to death is estimated to be in a rang. Their aims was to develop a model that can predict whether a person is affected by COVID-19 or not using SVM classification. All the possible conditions of infection based on the symptoms have been meticulously looked upon, and accordingly the dataset has-been framed. It’s hard to reflect accurate prediction since persons without any signs or symptoms of COVID-19 are also getting affected[14].

In 2020 Nitul Duttaer al proposed “SVM-based Analysis for Predicting Success Rate of Interest Packets in Information Centric Networks”. They used a machine learning (ML) algorithm, particularly a Support Vector Machine (SVM) to forecast the success of the Interest packet. A CR can then send an Interest packet in the outgoing interface which is forecasted successful. The objective is to maximize the success rate which in turn minimizes content search time and maximizes throughput. The dataset used in is generated from a simulation topology designed in *ndn.Sim* comprising 10 K data points having 10 features. The linear, RBF and the polynomial kernel (with degree 3) are used to analyze the dataset. The polynomial kernel shows the best behavior with 98% accuracy. They found that yes, it is possible to enhance the performance of the content search process by integrating Interest forwarding with mL techniques. Their claims supported by applying SVM into a dataset collected from ns-3-based ICN tested (Section). A dataset with 10 K records and 10 features are used with training-testing ratio of 70–30. A dataset with 10 K records

and 10 features are used with training-testing ratio of 70–30. Three variations of SVM model are applied on the available dataset and their prediction performance is evaluated during the research [15].

In 2021 C. Karthikeyan et al proposed “Prediction of Climate Change using SVM and Naïve Bayes Machine Learning Algorithms”. They analyzed the different machine learning algorithms, such as, Random forest, Linear Regression, k-means and naïve-bayes algorithms. From machine learning, a quality classifier called support vector machine (SVM) classification is used to predict and quantify the failures probability as a function of the values of POP2 parameters. Machine learning types are used to define in aggregating failure probabilities of listed parameters. Sensitivity analysis utilized to analyze various parameters and modules to explain the simulation failures. They presented probabilistic networks and show their relevance for nearby climate forecasting and down scaling. They discussed the fundamental results show how such models can be fabricated and how they can use for performing derivation. Further analysis is as yet required for deciding the pragmatic employable productivity of these models; first tests are being promising [16].

V. PROBLEM STATEMENT

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms. There some problem, when we used SVM for a data set [19,22,23]

- There are multiple hyper planes that separate the data exists, which one to choose?
- Finding maximum margin hyper planes for given data set.
- Sometimes the linear decision boundary is not good. How can we learn non-linear decision boundaries with the SVM?
- SVM algorithm is not suitable for large data sets.
- Target classes are overlapping. In cases where the number of features for each data point exceeds the number of training data samples

VI. PROPOSED APPROACH

We need to define a line as:

$$w^T x + b = 0$$

Drawing this on a 2-dimensional space gives us all the possible solutions which satisfy this equation. w defines the orientation and b the position in relation to the origin. Note that the parameter vector w is perpendicular to the line itself. If a point, x , satisfies this condition then we know it lies on the boundary itself. If the result is larger than 0 then it's on the positive side of the boundary, if it's less than 0 then it's on the negative side. This is now our decision rule.

Now, let's add a few constraints. First, if we have two classes they're denoted as $y_i = +1$ or $y_i = -1$. Second, I want every sample to have a value indicative of its class:

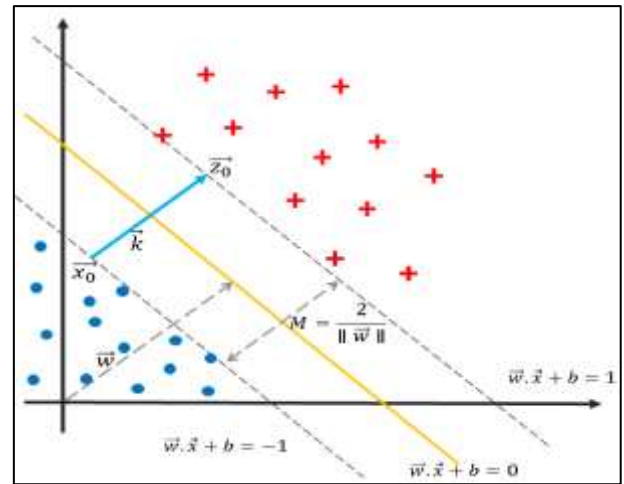


Fig. 7: Boundary for positive and negative sample
 $w^T x_+ + b \geq 1$ (for all positive samples)
 $w^T x_- + b \leq -1$ (for all negative samples)

This allows us to squeeze the above two expressions into one:

$$y_i(w^T x_i + b) - 1 \geq 0$$

If the above happens to give us 0 then we know x lies exactly at margin's length from the boundary:

$$y_i(w^T x_i + b) - 1 = 0 \quad y_i(w^T x_i + b) = 1$$

As a matter of fact, if we're not interested in the sign we can re-write this as:

$$|w^T x_i + b| = 1$$

This will come in handy later on.

Next we need to get a handle on the distance for the absolute nearest point x_i to the boundary, otherwise the machinery of the method would break down. In order to do this we need to imagine this line being comprised of some set of points in the feature space. Take any point, x , from this set and then take the vector connecting it with x_i . This gives us the vector $x - x_i$. Bear in mind that the Euclidean distance between these points $\|x - x_i\|$ isn't what we want; this is just helping us get there. What we want is the length of the component in the direction of w , in other words the projection of $(x - x_i)$ on w . This is given by:

$$d = (x - x_i)^T \cdot \frac{w}{\|w\|}$$

VII. IMPLEMENTATION ENVIRONMENT

We evaluate the performance of linear SVM, we used Python language design user interface and to used CSV files to store data set. The experiments were performed on Intel Core i3processor 2 GB main memory and RAM: 4GB Inbuilt HDD: 400GB OS: Windows7. Synthetic datasets are used to evaluate the performance of the algorithms. Description about the Dataset for first model Heart Disease Recognition Data Set Description Heart Disease data set consists of 14 attributes data. All the attributes consist of numeric values. First 13 variables will be used for predicting 14th variables. The target variable is at index 14.

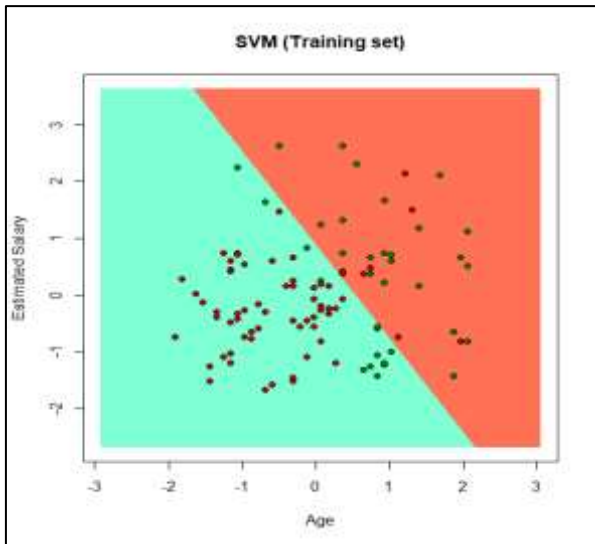


Fig. 8: Dividing data set using Linear SVM

VIII. ACCURACY WITH TRAINING DATA SET AND TESTING DATA SET

Data set	Sample size
Training data set	0.7677345
Testing data set	0.8333

Table 1: Comparing Accuracy with training data set and testing data set

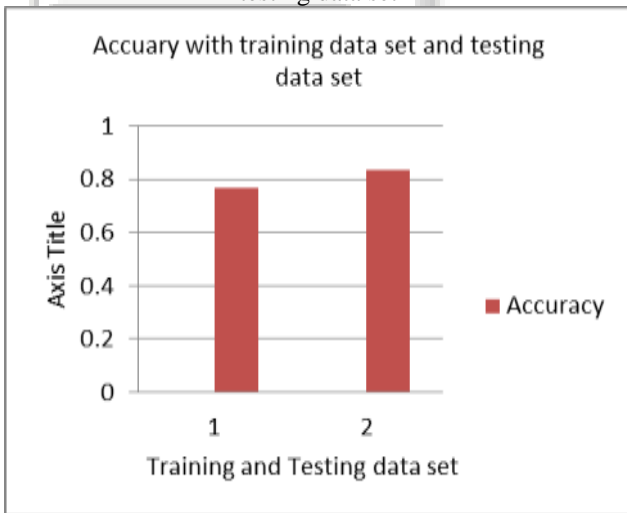


Fig. 9: Comparing Accuracy with training data set and testing data set

IX. CONCLUSION

In the proposed work we use linear kernel based SVM and develop to models first model used heart disease data with 14 attribute. Last attribute is classify the records based on SVM model that patient belong to yes or no class. In second model we used Social data set which has five attribute customers ID, Gender, Age Salary, purchased, the last class is a binary class. Based on salary and age we need to classify the into purchased class (Yes or No). We used python language to implement both the model. We divided the data set into training and testing data set. First we create model using training data set. After that we apply the testing data set. We create Confusion Matrix for both models after allying testing

data set to find out the accuracy of the models. We calculate the accuracy and use graph to the show the number of records classifies correctly by both model. Real life data set is used in both the models.

REFERENCE

- [1] Wei Yu , Tiebin Liu, Rodolfo Valdez, Marta Gwinn, Application of support vector machine modeling for prediction of common diseases: the case of diabetes and pre-diabetes National Office of Public Health Genomics, Coordinating Center for Health Promotion, Centers for Disease Control and Prevention, Atlanta, GA, USA Yu et al. BMC Medical Informatics and Decision Making 2010, 10:16 <http://www.biomedcentral.com/1472-6947/10/16> © 2010 Yu et al; licensee BioMed Central Ltd.
- [2] Lubor Ladický Locally Linear Support Vector Machines University of Oxford, Department of Engineering Science, Parks Road, Oxford, OX1 3PJ, UK Appearing in Proceedings of the 28 th International Conference on Machine Learning, Bellevue, WA, USA, 2011. Copyright 2011 by the author(s)/owner(s).
- [3] Yongli Zhang Support Vector Machine Classification Algorithm and Its Application C. Liu, L. Wang, and A. Yang (Eds.): ICICA 2012, Part II, CCIS 308, pp. 179–186, 2012. © Springer-Verlag Berlin Heidelberg 2012
- [4] Qing Wu and Wenqing Wang Piecewise-Smooth Support Vector Machine for Classification Hindawi Publishing Corporation Mathematical Problems in Engineering Volume 2013, Article ID 135149, 7 pages <http://dx.doi.org/10.1155/2013/135149>.
- [5] Gary Doran · Soumya Ray A theoretical and empirical analysis of support vector machine methods for multiple-instance classification Mach Learn (2014) 97:79–102 DOI 10.1007/s10994-013-5429-5 Case Western Reserve University, 10900 Euclid Ave, Cleveland, OH 44106, USA e-mail: gary.doran@case.edu.
- [6] Dr. S. Vijayarani, Mr.S.Dhayanand Kidney Disease Prediction Using SVM And ANN Algorithms International Journal of Computing and Business Research (IJCBR) ISSN (Online) : 2229-6166 Volume 6 Issue 2 March 2015.
- [7] Nick Guenther and Matthias Schonlau Support vector machines The Stata Journal (2016) 16, Number 4, pp. 917–937.
- [8] Nurhanna, A. A. and Othman, M. F Multi-class Support Vector Machine Application in the Field of Agriculture and Poultry: A Review Malaysian Journal of Mathematical Sciences 11(S) February: 35–52 (2017) Special Issue: Conference on Agriculture Statistics 2015 (CAS 2015).
- [9] Shujun Huang, Nianguang Cai, Pedro Penzuti Pacheco, Applications of Support Vector Machine (SVM) Learning in Cancer Genomics Cancer Genomics & Proteomics 15: 41-51 (2018) doi:10.21873/cgp.2006.
- [10] Rodrigo P. Monteiro , Mariela Cerrada, Using a Support Vector Machine Based Decision Stage to Improve the Fault Diagnosis on Gearboxes Hindawi Computational Intelligence and Neuroscience Volume 2019, Article ID

- 1383752, 13 pages
<https://doi.org/10.1155/2019/1383752>.
- [11] Sunil Kumar, Vivek Kumar, R. K. Sharma Rice Yield Forecasting Using Support Vector Machine International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-4, November 2019.
- [12] Kai-Xu Han, Wei Chien, Chien-Ching Chiu Application of Support Vector Machine (SVM) in the Sentiment Analysis of Twitter Data Set College of Electronic and Information Engineering, Beibu Gulf University, Qinzhou 535011, China; frog0696@163.com 23 December 2019; Accepted: 30 January 2020; published: 7 February 2020.
- [13] A. Ravi Kumar, G. Anil Kumar Support Vector Machine for Text Categorization using Principle Component Analysis in Data Mining International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277-3878, Volume-8 Issue-5, January 2020.
- [14] Soham Guhathakurata, Souvik Kundu, A A novel approach to predict COVID-19 using support vector machine Data Science for COVID-19. <https://doi.org/10.1016/B978-0-12-824536-1.00014-9> Copyright © 2021 Elsevier Inc. All rights reserved.
- [15] Nitul Dutta, Sudeep Tanwar, SVM-based Analysis for Predicting Success Rate of Interest Packets in Information Centric Networks Applied Artificial Intelligence An International Journal e: <https://doi.org/10.1080/08839514.2021.2020488>.
- [16] Nalindren Naicker, Timothy Adeliyi, Linear Support Vector Machines for Prediction of Student Performance in School-Based Education Hindawi Mathematical Problems in Engineering Volume 2020, Article ID 4761468, 7 pages <https://doi.org/10.1155/2020/4761468>.
- [17] C. Karthikeyana, Gurrarn Sunithab, J. Avanija Prediction of Climate Change using SVM and Naïve Bayes Machine Learning Algorithms Turkish Journal of Computer and Mathematics Education Vol.12 No.2 (2021), 2134 – 2139.
- [18] F R Lumbanraja,*, E Fitri, Ardiansyah Abstract Classification Using Support Vector Machine Algorithm (Case Study: Abstract in a Computer Science Journal) ICASMI 2020 Journal of Physics: Conference Series 1751 (2021) 012042 IOP Publishing doi:10.1088/1742-6596/1751/1/012042.
- [19] Dr. B.M. Gupta Dr. S.M.Dhawan Support Vector Machine (SVM) Research in India: A Scientometric Evaluation of India's Publications Output During 2002-1 Journal of Indian Library Association, Vol. 57 (3), July – September, 2021.
- [20] Seong-Uk Nam, Sangil Kim, HyunMin Kim, and YongBin Yu Comparative Study of The Performance of Support Vector Machines With Various Kernels East Asian Math. J. Vol. 37 (2021), No. 3, pp. 333–354 <http://dx.doi.org/10.7858/eamj.2021.023>.
- [21] Biplab Madhu, Md. Azizur Rahman, Arnab Mukherjee, A Comparative Study of Support Vector Machine and Artificial Neural Network for Option Price Prediction Journal of Computer and Communications, 2021, 9, 78-91 <https://www.scirp.org/journal/jcc> ISSN Online: 2327-5227 ISSN Print: 2327-5219.
- [22] M. Ramana Reddy Implementation of SVM machine learning Algorithm to predict lung And Breast Cancer Turkish Journal of Computer and Mathematics Education Vol.12 No.12 (2021), 3050-3060.
- [23] Dakhaz Mustafa Abdullah and Adnan Mohsin Abdulazeez Machine Learning Applications based on SVM Classification: A Review Technical College of Informatics, Akre Duhok Polytechnic University Duhok, Iraq dakhaz.abdullah@dpu.edu.kud May 2021