An Analysis on Predicting Social Media Ads using Kernel SVM Function

Swathi Jayaprakash [1], Dutta Yasaswi[2] and Pattabiraman V[3]
[1] Vellore Institute of Technology Chennai
swathijayaprakash2000@gmail.com
[2] Vellore Institute of Technology Chennai
yasaswidutta666@gmail.com
[3] Vellore Institute of Technology Chennai
pattabiraman.v@vit.ac.in

Abstract. This paper predicts the social media users who will buy a car based on the previously observed dataset. The dataset is extracted from user who has undergone the search of cars and advertised by it in the internet. The particular dataset has attributes such as User ID, Gender, Age, Estimated Salary, and Purchased car. Using certain data mining methods, we predict the customer who's most likely to buy a car and our focused gets on entirely on respective users. Ignoring those aren't really ready to purchase a car at that time. This helps marketing department where to relay on hence, saving of money and time on par with having an analysis of going whether to car or not.

This gives an idea to monitor on certain users to look up for. Push on the ads to these users who are most likely to purchase. Concentrating less on those are unlikely to buy based on dataset, we have extracted. Based on the analysis of the data extracted, we used different kernel svm techniques and already existing data mining models, according to this data we say which algorithm gave an effective result, and conclude the best algorithm. The approach we going to possess is via confusion matrix and get concluded on accuracy, precision, specificity etc.

Keywords: svm, Kernel svm, accuracy, RBF, polynomial kernel.

1 Introduction

Now-a-days, social media has become much commercial than ever due to its usage in the current world. It has become boon to e-commerce companies. As, it working as platform to reach out their products to the people in much easier way possible and get to increase their market with lesser time and money. On that note, we have come with a paper called social media ads prediction system. The application has designed completely for marketing business management. Our work gives an accurate result of user's data and tells whether user has particular user has brought the product or not. User's data contains user id, estimated salary, age, gender, purchased or not. If user seems on affordable basis, the social media ads get sent to respective user. If the user has not purchased the product and looking for buying the particular product. The required social media ad gets sent to respective user.

The application that we have made would reduce the money that has to spend on marketing. Companies need not to suggest their product based on the user's dataset which saves the time. Our application performs on nine different models of data mining techniques which are linear kernel SVM, naive bayes classifier, logistic regression, polynomial kernel SVM, Radial type kernel SVM, Random Forest classifier, sigmoidal type SVM, KNN and decision tree classifier. All these nine models' give's a different result. We compare these nine data mining models and conclude which model would be best suitable in predicting the social media ads prediction. The comparison is done based on the confusion matrix, the data visualization of graphs for both training and testing data. Eventually, E-commerce companies can totally rely on the work which brings them with better results.

There are four major kernel function we used in this paper which are namely linear type, radial basis type, polynomial type and sigmoid type function. SVM Algorithm is actually an Algorithm which is very powerful when it comes to classification in data mining and machine learning. Svm has a good Math Formula which is required to build the basic difference between the groups of classes. The software we used to implement was RStudio using R Programming. ElemStatLearn, CaTool, e1071 are few packages used in implementation. E1071 is an important package used in R which helps us to use some of stats and classification functions like svm, naïve bayes, Fourier transforms etc.

There are several packages used for implementing svm, but we have used e1071 package as it delivers a powerful interface. As we are using the different kernel functions and implementing the svm. This package provides all the kernel function we used in this paper. On the other hand, CaTools is the package which provides the fundamental functions in R like quick calculations, error round off, LogitBoost etc. The ElemStatLearn package provides a better visualization of data and also has many statistical functions included in it.

2 Literature Survey

[1] This paper basically inspects about svm and various modified methods of svm. This paper also speaks about the problem faced in svm such as uneven distribution, sensitive to noise, the presence of outliers. [4] Deals with the svm Algorithm, wherein they apply svm and other data mining Algorithm for about 4 different set of datasets say diabetics dataset, satellite dataset, shuttle dataset and heart dataset. They concluded that applying the type of kernel depends on the dataset used and the best results were shown by the RBF as the data is multi class.

For solving the svm [7], the authors came with 3 iterative approaches and suggested approaches made use of two functions namely convex Huberloss and robust for finding the error. Dinesh et al. [2] briefs on merits, by using the Algorithms for prediction in the field of education. In addition to that they state about applying the prediction Algorithm to collected data makes huge difference. Hachesu et al. [3] the authors have stressed on prediction of length of stay in respective hospitals Prediction was based on cardiac patients who were admitted in hospitals. Based on highest accuracy rate, the length of stay was predicted which in return reduces the burden of hospital management. Charles et al. [5] deals with predicting the type of soil suitable for agriculture process. The authors have implemented the prediction Algorithms in order to come up with best soil for better results.

3 Datasets Used

We have used a social media advertisement data set for predicting It basically checks, whether users have purchased a product by clicking on the advertisements shown to them. So, the dataset has got 5 variables namely user id, gender, age, estimated salary, purchased [it’s a Boolean value, 0 representing the user have not purchased, 1

representing the user buying the product]. We have extracted this dataset from Kaggle and sample of first 5 data, from the dataset shown in Table 1. This dataset has about 400+ data.

Table 1. Social media Ads Prediction dataset

User Id	Gender	Age	EstimatedSalary	Purchased
15624510	Male	19	19000	0
15810944	Male	35	20000	0
15668575	Female	26	43000	0
15603246	Female	27	57000	0
15804002	Male	19	76000	0

4 Methodology

For implementation of the kernel svm and other Algorithms, we used a software called R Studio. Even though the svm type of model works well for regression analysis, the svm is widely used for classification analysis. We apply this just by plotting out the data in the n-dimensional space. And further on, we mined to discover the perfect hyper plane which separates between two group classes.

The points that's closer to the hyperplane will have an effect on the orientation, arrangement and location of the hyperplane, this kind of points is called as the "support vectors". Which is basically the coordinate representation of many observations and it's a segregation method for separating the two class. There are namely 3 main steps to follow in svm:

1. Start with a data based on low dimension.
2. Then we should supposed to move the data into higher dimension. In our case we are already dealing with 2-Dimension data. [x axis= age, y axis= estimated salary]. So, we should convert the 2D data into 3D data.
3. Find the support vector classifier which separates the higher dimension data into two groups.

For conversion of two-dimension data to three-dimension data, you may wonder how we decide, how to transform the data from 2D to3D. Only in order to make the mathematics possible, the svm makes use of something called Kernel functions to comprehensively discover support vector classifiers in higher-dimension.

There are namely many types of kernels. Few of the kernel svm present in R Studio are namely Radial, Linear, sigmoid, polynomial. We are exploring the all above mentioned models and find which type of kernel svm gives a good classification result.

Each of these kernels has a unique mathematical equation which helps in forming the hyper plane, separating 2 class of data. The Algorithm is just the same and the only difference for the different type of kernel svm is their kernel function used. These kernel functions are difficult to solve manually, so we use R Studio to do that. The Algorithm of Kernel svm is as follows:

1. First generate the data in 2 dimensions, and separate into 2 matrix.
2. Choose two hyper planes (in 2D) that can separate the data with no points between them.
3. Try to maximize their distance (the margin).
4. The average line will be the decision boundary.
5. Now we should load the package e1071 which contains the svm function.
6. As we are using 2 variables, let's take y as the response variable and other variables as the predictors.
7. Fit the model using "svmfit" function present in e1071.
8. And plot the confusion matrix to find the accuracy of the model.

The Different kernel functions used are:

4.1 Polynomial Type Kernel Svm

The polynomial kernel is also a svm kernel feature which in reality represents the similarity among the vectors, allowing it to understand and learn the non-linear models.

Additionally, the polynomial type kernel svm kernel now not only seems at the characteristic of Input samples to discover the similarity among variables, but also specializes in all the available combos. While we communicate this with respect to regression evaluation, those set of combinations are referred to as interaction features. While the Input functions are binary-valued (both 0 and 1), then we say the features correspond to logical co-prevalence of Input features. Xi and Xj are 2 different observations in the dataset. R stands for the coefficient of the polynomial.

$$(Xi, Xj)=(Xi.Xj+c)2 \quad \textbf{(1)}$$

c is simply a constant in (1).

The d in (1) stands for the degree of the polynomial, when d=1, the polynomial kernel finds the relation between each pair of observation in one dimension. Further on the calculated relationships are used to find a support vector classifier. When d=2, we get a second dimension based on square of the equation and polynomial kernel finds the relation between each pair of observation in two dimensions and so on. To conclude, the polynomial kernel comprehensively gets increasing its dimensions just by setting the value of "d". So, by applying polynomial kernel function to the dataset, it generates the following output. (see Fig.1. a) and (see Fig.1. b) shows the visual representation of the training and testing dataset with polynomial kernel SVM respectively and Table 2 signifies the confusion matrix generated.

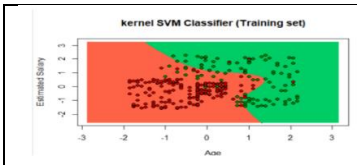

Fig. 1. a. Plot of training dataset using polynomial kernel

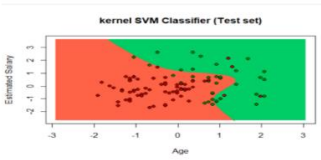

Fig.1. b. Plot of testing dataset using polynomial kernel

Table 2. Confusion matrix for polynomial type kernel svm

Y predictions	**0**	**1**
0	60	4
1	18	18

4.2 Radial Type Kernel Svm

Radial Kernel is one of the powerful kernel svm technique. This type of kernel svm is more useful when we deal with not so linear separable data. So, the process that we intake to resolve this type of dispute is just by applying a non-linear modification to the feature variable then, further on transforming those variables to a higher dimension (i.e., from Two-Dimension to three-Dimension space) which is often referred to as feature space. By doing the above-mentioned steps we are segregating the irregular/non-uniform data with a non-uniform partition. RBF is abbreviated as the radial basis function. It's basically used when the user doesn't have any prior knowledge about the dataset used. It's represented as:

$$(Xi,Xj)=e-\gamma(Xi-Xj)2 \quad \textbf{(2)}$$

In (2) xi and xj two different variables, the difference between the variables is squared which gives us the squared distance between the two variables. so, the number of impact one observation has on another is a function of the squared distance. The gamma determines the cross validation, which scales the squared distance which means it scales the influence. If we plug in the values for xi and xj. If the points are relatively closer than the value of the kernel svm will be higher. If we calculate for points which are distant then the value would be less.

The radial kernel finds support vector classifiers in infinite dimensions. In this type of model, the nearest data points have a lot of influence on how we classify the new observation. (see Fig.2a) and (see Fig.2b) shows the visual representation of the training and testing dataset with radial basis kernel SVM respectively and Table 3 signifies the confusion matrix generated for it.

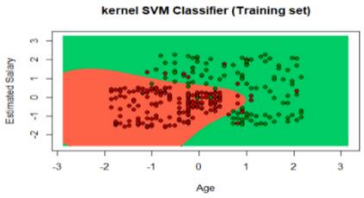

Fig. 2a. Plot of training dataset using RBF kernel

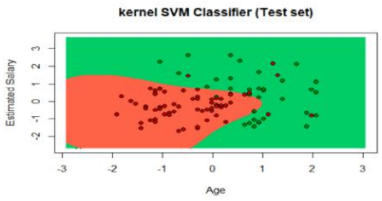

Fig.2b. Plot of testing dataset using RBF kernel

Table 3. Confusion matrix for Radial type kernel svm

Y predictions	0	1
0	58	6
1	4	32

4.3 Linear Type Kernel Svm

This kernel is one-dimensional and is the maximum simple form of kernel in SVM. The equation is:

$$K(Xi,Xj)=Xi.Xj \tag{3}$$

Linear kernel svm forms a liner hyper-plane. This model is effective when we work with uniform data. As we are dealing with non-uniform set of data this type of kernel svm doesn't suit better. This creates a hyper plane which is a straight line. And the points may be mislaid.

So, (see Fig.3a) and (see Fig.3b) shows the visual representation of the training and testing dataset of linear kernel SVM respectively and Table 4 signifies the confusion matrix generated for it.

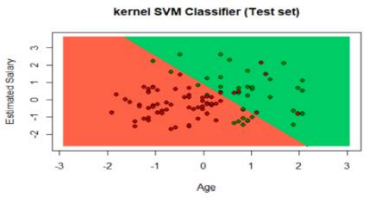

Fig.3a. Plot of testing dataset using linear kernel [Left]

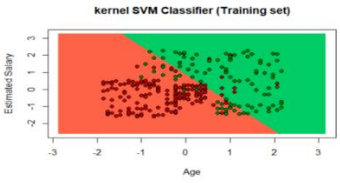

Fig.3b. Plot of training dataset using linear kernel [Right]

Table 4. Confusion matrix for linear type kernel svm

Y predictions	0	1

0	57	7
1	13	23

4.4 Sigmoid Type Kernel Svm

The Sigmoid Kernel occurs to arise from Neural Network area, wherein the bipolar [that is -1 and +1] sigmoid type kernel function is frequently applied like an activation feature for AN's. A svm model that makes use of the sigmoid kernel characteristic is just equal to the 2-layer, perceptron neural network. We are able to constitute this as:

$$K(Xi, Xj) = \tanh(\propto Xi^{T}Xj + c) \quad \textbf{(4)}$$

So, (see Fig.4a) and (see Fig.4b) shows the visual representation of the training and testing dataset of sigmoid kernel SVM respectively and Table 5 signifies the confusion matrix generated.

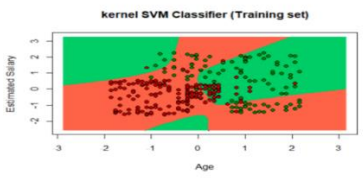

Fig. 4a. Plot of training dataset using sigmoid kernel

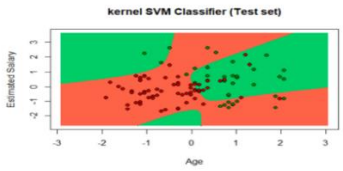

Fig.4b. Plot of testing dataset using sigmoid kernel [Right]

Table 5. Confusion matrix for linear type sigmoid svm

Y predictions	**0**	**1**
0	53	11
1	14	22

5 Observation

So, according to the social media ads data set. We have worked on with several Algorithms and here is the Table 6 and Table 7, which compares the correctness of the model. The accuracy of our model is calculated using the confusion matrix. Which generates 4 values namely TruePositive [Tp], FalsePositive [Fp], TrueNegative [Tn] and FalseNegative [Fn].
The correctness of any system is basically calculated using the below formula:

Table 6. We are calculating the accuracy for all the Algorithms used.

S.NO	**Algorithm Used**	**Confusion matrix Result Generated**				**Accuracy in percentage**
		Tp	**Fp**	**Tn**	**Fn**	

1.	Logistic Regression	57	7	26	10	83
2.	KNN	59	5	30	6	89
3.	Decision tree classifier	53	11	30	6	83
4.	Random Forest Classifier	53	11	30	6	86
5.	Naive Bayes	57	7	29	7	86
6.	Kernel SVM using Radial kernel	58	6	32	4	90
7.	Kernel SVM using Linear kernel	57	7	23	13	80
8.	Kernel SVM using Sigmoid kernel	53	11	22	14	75
9.	Kernel SVM using polynomial kernel	60	4	18	18	78

Table 7. We are measuring of correctness for existing Algorithm.

Measures	Logistic Regression	KNN	Decision tree	Random Forest	Naive Bayes	Radial Kernel	Linear kernel	Sigmoid kernel	polynomial kernel
Sensitivity Tp/(Tn+Fn)	85	90	89	86	89	93	81	79	76
Specificity Tn/(Tn+Fp)	78	85	73	84	80	84	76	66	81
Precision Tp/(Tp+Fp)	89	92	82	92	89	90	89	82	93
False Positive Rate (Fp/(Fp+Tn)	21	14	26	15	19	15	23	33	18
False Negative Rate Fn/(Fn+Tp)	14	9	10	13	10	06	18	20	23

F1 Score 2Tp/ (2Tp+Fp+Fn)	87	91	86	89	89	92	85	80	84
False Discov-eryRate Fp/ (Fp+Tp)	10	7	17	7	10	9	10	17	6

6 Result

The Output got to be an analysis on predicting a social media user can purchase a car or not. Our analysis is based on nine different data mining Algorithms. We get to find confusion matrix for all of the nine methods and data visualization way of representation. This gives us how these nine data mining methods are efficient from each other and their respective outcome. We have further calculated the other correctness measures to find the efficient Algorithm.

The Table 6 and Table 7, Signifies the various calculation made out of the results of confusion matrix and the values were based on percentage. Polynomial kernel svm has highest rate of precision value (see Fig.8). Internally, it depicts the measure of quality. KNN here again has the highest specificity (see Fig.7) measure. When it comes to decision tree classifier, the measure of sensitivity tops the list that other measures. The sensitivity sees through the actual positive that get predicted as positive. Another classifier that we have implemented is Random Forest Classifier has the highest percentage of precision which in turn speaks about quality. The third classifier, Naïve Bayes has got to have similar percentage values of sensitivity, Precision and F1 score [Table 7]. The innovative work in this paper deals with kernel SVM, the radial kernel has the highest rate of sensitivity (see Fig.6) which predicts positivity. Linear kernel excels at the quality aspect, Precision. Sigmoidal kernel is followed by linear kernel which is precision. The last type of kernel SVM, Polynomial kernel has the highest rate measure of precision (see Fig.8). The no of highest false positive rate present is in sigmoidal kernal svm (see Fig 9). When it comes to the highest percentage of false negative rate, polynomial kernel svm tops the list (see Fig.10).

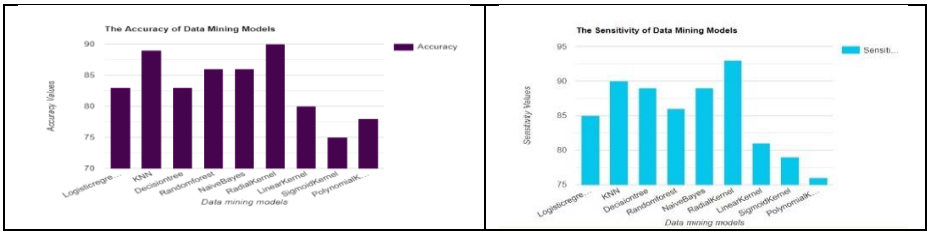

Fig.5. Graphical representation of accuracy of the Algorithms used.	**Fig.6.** Graphical representation of sensitivity measure of the Algorithms used.

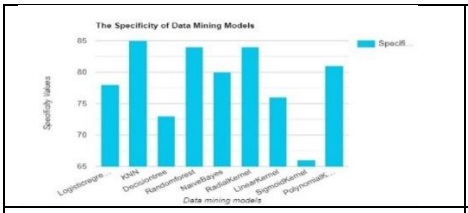	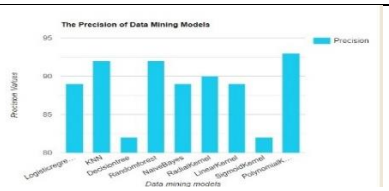
Fig.7. Graphical representation of specificity measure of the Algorithms used.	**Fig.8.** Graphical representation of precision measure of the Algorithms used.

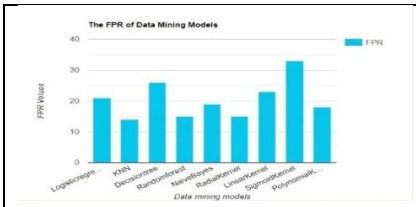	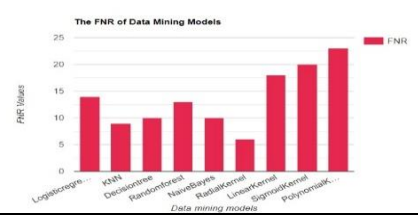
Fig.9. Graphical representation of fpr measure of the Algorithms used.	**Fig.10.** Graphical representation of fnr measure of the Algorithms used.

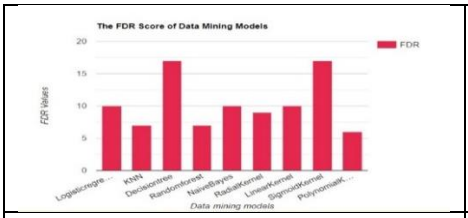	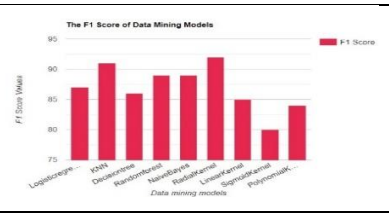
Fig.11. Graphical representation of fdr measure of the Algorithms used.	**Fig.12.** Graphical representation of f1 measure of the Algorithms used.

Conclusion

For the Social Media Ads Prediction data set, on total we have used about 9 Algorithms and found the accuracy and other measures of each and every Algorithm Table 6 and Table 7. Out of the 9 Algorithms. The 4 kernel functions in svm algorithm are claimed to be our novelty in this work which are namely radial kernel svm, Linear

Kernel svm, sigmoidal kernel svm, and polynomial Kernel svm. We observe that the accuracy (see Fig.5) of the system with these 4 kernel svm and the already existing data mining/ machine learning Algorithm, we conclude that polynomial kernel SVM has got the highest rate of precision value (see Fig.8). The precision precisely deals with the quality of measure where the ratio is driven with predicted positive notions to the total predicted positives notions.

The kernel svm using radial basis function has given the maximum accuracy (see Fig.5) and F1 Score (see Fig.10) for the system. Which in turn signifies the performance of the model. So, thereby we declare the kernel svm using radial kernel function is the best Algorithm compared to all other existing algorithm, with respect the Social Media Ads Prediction data set.

References

[1] Babacar Gaye, Dezheng Zhang, Aziguli Wulamu, "Improvement of Support Vector Machine Algorithm in Big Data Background", Mathematical Problems in Engineering, vol. 2021, Article ID 5594899, 9 pages, 2021.

[2] A.Dinesh Kumar1 , R.Pandi Selvam2 , K.Sathesh Kumar3 " Review on Prediction Algorithms in Educational Data Mining ", International Journal of Pure and Applied Mathematics, Volume 118 No. 8, 2018.

[3] Hachesu PR, Ahmadi M, Alizadeh S, Sadoughi F, "Use of data mining techniques to determine and predict length of stay of cardiac patients." Healthcare Informatics Research. 2013, PMCID: PMC3717435

[4] durgesh k. srivastava, lekha bhambhu, "data classification using support vector machine", journal of theoretical and applied information technology, 12(1):1-7, 2010.

[5] S.S.Baskar, L.Arockiam, S.Charles, "Applying Data Mining Techniques on Soil Fertility Prediction", International Journal of Computer Applications Technology and Research, Volume 2– Issue 6, 660 - 662, 2013, ISSN: 2319–8656

[6] M. Malvoni, M. G. De Giorgi, and P. M. Congedo, "Data on support vector machines (SVM) model to forecast photovoltaic power," *Data in Brief*, vol. 9, no. C, pp. 13–16, 2016.

[7] P. Borah and D. , "Functional iterative approaches for solving support vector classification problems based on generalized Huber loss," *Neural Computing and Applications*, vol. 32, no. 1, pp. 1135–1139, 2020.