A Comparative Study of Optimal Kernel Selection for Support Vector Machines

Ch. Chandu Priya

PG Scholar, Dept. of Computer Science Sri Venkateswara University, Tirupati

Abstract— *Support vector machine (SVM) is prepared to do outcompeting each and every learned model calculation as far as exactness and other superior execution measurements by its high layered information projection for grouping. By and by, the exhibition of the SVM is enormously impacted by the decision of the part capability which helps in the exactly. This paper talks about the working of SVM and its reliance on the bit capability, alongside the clarification of the kinds of parts. The attention is on picking the ideal bit for three unique kinds of information that shift on volume of highlights and classes to close the ideal decision of the piece for each sort of the three datasets. For execution measures, we utilized measurements like exactness, kappa, explicitness what's more, responsiveness. This concentrate measurably analyzes and looks at each kind of bit against the referenced measurements. In this research paper, we employ Support Vector Machine (SVM) on the Indian Liver Patient dataset to analyze the performance of different kernel choices. SVM is a powerful machine learning technique that has shown significant success in various domains. Our study focuses on two SVM kernels: polynomial and radial basis function (RBF). Through our experiments, we demonstrate that the polynomial kernel achieves the highest precision*.

I. INTRODUCTION

Artificial Intelligence has turned into a focal region of the planet progressing and driving organizations like Google, Facebook, YouTube, etc. It is a space of study which permits models and calculations to gain from encounters or examples from prepared information to execute such procured information into anticipating future results without being expressly modified [5]. Prescient examination of AI has plentiful applications in true situations with the exception of that calculations are perplexing and the interpretability is restricted denying us to know about which model or which boundaries of the picked model to use for a specific issue [7]. One of such elite execution models is SVM which when conveyed with the right piece capability gives more exact outcomes contrasted with different models. In any case, there is no immovable rule to know the right part that could be utilized with the SVM, leaving us with the preliminary and mistake strategy.

In this research paper, we employ Support Vector Machine (SVM) on the Indian Liver Patient dataset to analyze the performance of different kernel choices.

In the field of medicine, liver cancer is one of the most prevalent and fatal malignancies [6]. Early detection of liver diseases is challenging due to various risk factors [8]. In this research paper, we employ Support Vector Machine (SVM) on the Indian Liver Patient dataset to analyze the performance of different kernel choices. SVM is a powerful machine learning technique that has shown significant success in various domains. Our study focuses on two SVM kernels: polynomial and radial basis function (RBF). Through our experiments, we demonstrate that the polynomial kernel achieves the highest precision.

II. CLASSIFICATION

Classification is a process of developing models or decisions that categorize and identify data classes or patterns. It involves a learning step where a classifier is created based on a labeled training set, and a testing step where the accuracy of the classifier is evaluated on unseen data. Technique is the way toward finding a model or an end that depicts and sees data classes and contemplations, to use the model to expect the classes of things whose class mark isn't known. Data arrangements should be visible as a two-stage measure: learning step in which a classifier is made depicting a destined design of classes or thoughts by detaching the status set contained instructive synopsis tuples and their associated names [3[5]. In the subsequent movement model is used for request by first evaluating the sensible accuracy of classifier worked during the secret turn of events. It is done using the test data. The precision of classifier on a given test set tuples is level of tuples that are evidently alluded to by the classifier. If the precision is over some alright level, the classifier can be used to expect future tuples whose class mark isn't known.

Portrayal is a kind of data evaluation that can be used to make models portraying epic data classes. Structure is a data mining thinking used to expect pack pay for data models. It is one of the fundamental structures in data mining and is used in various applications, for instance, plan interest, torment assertion, client relationship the pioneers, and surrendered out appearing. The goal of the portrayal assessments is to collect a model from a colossal stack of orchestrating data whose target class names are known and reasonably this model is used to pack covered cases [4][6].

Plan is the most conventional and most wellknown data mining techniques. Outline maps data into predefined gatherings or classes. It is typical proposed as facilitated getting contemplating how the classes are settled going prior to taking a gander at the data [5]. Method is the way toward finding a model that sees data classes, to use the model to predict the class of things whose class name is dull. The picked model relies upon the assessment of a gigantic load of orchestrating data. Instructive arrangements are rich with covered information that can be used for cautious dynamic.

III. METHODOLOGY

SVM, a stunning machine system produced using legitimate learning and has made huge accomplishment in some field. Presented during the 90's, they incited an effect of pay in PC based knowledge. The groundworks of SVM have been made by Vapnik and are getting comprehensiveness in field of man-made brainpower by virtue of different engaging highlights and promising observational execution.

3.1 Support Vector Machine

Support Vector Machines (SVM) is a man-made reasoning estimation that is in general around used for request issues. SVM estimation is potentially the most basic portrayal systems that were actually applied to various authentic issues [1][2]. SVM depend in the wake of sorting out server homesteads to a high layered piece space where a segregating hyper-plane can be found. The standard reasoning used by SVM for data request is to drawn ideal hyper-plane which goes probably as a separator between the two classes. The vectors near the hyper-plane are called help vectors. This organizing can be carried on by applying the piece stunt which obviously changes the data space into another high layered part space. The hyper-plane is managed by reinforcing the distance of the closest plans, i.e., edge support, avoiding the issue of overfitting [10].

Consider the two-class issue where the classes are straightly distinct. Let the dataset D be given as (x1, y1), (x2, y2) … . (xn, yn) ε Rn, where xi is the course of action of preparing tuples with related class marks, yi. Every yi can take one of the two qualities, either +1 or - 1 [11]. The information is clearly distinct considering the way that numerous amounts of straight lines can restrict the server farms into two explicit classes where, in class 1, y = +1 and in class 2, y = - 1. The best separating hyperplanes will be the one which have the maximal edge between them.

3.2 Kernel decision of SVM

Part assurance expects a critical part in SVM planning and game plan. A properly arranged part ability can restrict hypothesis botch, accelerate mix speed, and augmentation assumption precision. There are two ordinary improvement methods, adding limits and part plan. Adding limits is a procedure for putting additional limits in the piece and redesigning those limits to deal with the show. There are four-piece techniques are open:

- Straight piece: $K(x_i, x_j) = x_i^T x_j$.
- Polynomial piece: $K(x_i, x_j) = (\gamma x_i^T x_j + r)^d$, $\gamma > 0$
- RBF segment: $K(x_i, x_j) = \exp(-\gamma \|x_i - x_j\|^2)$, $\gamma > 0$
- Sigmoid piece: $K(x_i, x_j) = \tanh(\gamma x_i^T x_j + r)$ Here, γ, r and d are bit limits.

One of the most extraordinary acknowledged systems is the SVM, a part based methodology which has found applications in numerous model affirmation issues [10][11]. Simply polynomial and RBF parts were bankrupt down in this work.

Support Vector Machines (SVM) have gained popularity in the field of artificial intelligence due to their attractive features and promising performance. SVM is based on the concept of mapping data into a higher-dimensional space, where a separating hyperplane can be found. The choice of the kernel function plays a vital role in SVM training and classification. The kernel function determines the similarity measure between data points in the transformed feature space. In this study, we compare the performance of two common SVM kernels: polynomial and radial basis function (RBF).

IV. EXPERIMENTAL RESULTS

To assess the presentation of various portions, we led tests utilizing the Indian Liver Patient dataset got from the UCI AI Vault [9]. We have utilized the Python Language to test our proposed calculations. The Python Scikit-learn is a pack for information strategy, break faith, social occasion and depiction. The dataset comprises of 576 occurrences with 11 ascribes and two class names: Liver sickness present and Missing class. We split the information into a preparation set (70%) and a test set (30%).

We executed SVM arrangement utilizing polynomial and RBF portions, changing the part boundaries to track down the ideal qualities.

It has been recommended that accuracy secured by the SVM relies overall on the part picked and the cutoff points. The assessment zeroed in on polynomial and expanded premise piece (RBF). The polynomial and RBF fragments have limit "C" that interfaces with the discipline for misclassification. The higher the worth of "C" is, the more the discipline, driving the approach model to over-fit. Then again, more unpretentious worth of "C" prompts a more summed up model that can not orchestrate the dull information precisely. In this paper, "C‟ was wavered from 1 to 50 for polynomial and 1 to 50 for the RBF fragment and subsequently the best worth of "as yet hanging out there close by calculating the presumption rate. Taking into account the run of the mill supposition rate obtained by moving the alluded to limits they proposed ideal qualities for "C‟ as 1 for polynomial piece and 10 for RBF. SVM assessment was assessed involving two portions for our dataset. For SVM utilizing polynomial part, the worth of p was changed from 1 to 3. Just 1, 2 and 3 were picked in this assessment. For SVM utilizing RBF part, the worth of γ was changed from 1 to 3. Just 1, 2 and 3 were picked in this examination. Notwithstanding, in our dataset the best furthest reaches of p=2 and γ=2 shows the most raised accuracy are dropped by our outcome as displayed in figure-1 and same displayed in the table-1.

Table 1
Results of SVM Kernels

SVM Kernel	Accuracy	Precision	Recall
Polynomial	95.76	95.7	95.7
RBF	93.58	93.59	93.67

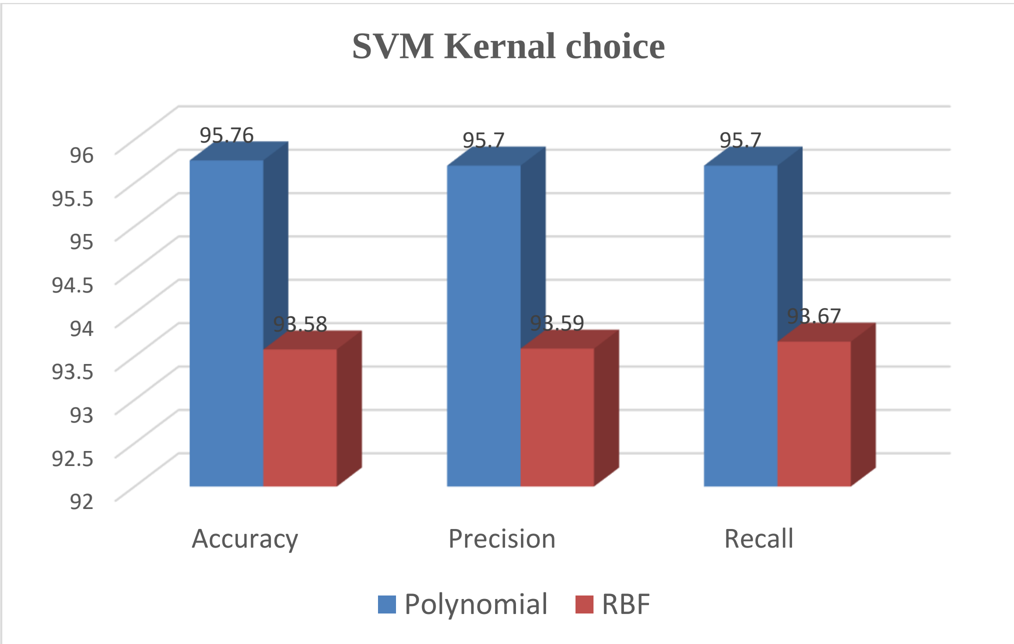

Figure-1: SVM Kernel results

V. RESULTS AND DISCUSSION

The results indicate that both SVM kernels achieved high accuracy in predicting liver disease expectancy. The polynomial kernel outperformed the RBF kernel, achieving an accuracy of 95.76% compared to 93.58% for the RBF kernel. This suggests that the polynomial kernel provides a better fit to the underlying data distribution in the Indian Liver Patient dataset.

Precision and recall, which are important performance metrics in classification tasks, were also evaluated for both kernels. Precision measures the proportion of true positive predictions out of all positive predictions, while recall measures the proportion of true positive predictions out of all actual positive instances.

For the polynomial kernel, we observed a precision of 95.7% and a recall of 95.7%. This indicates that the model achieved a high level of precision in correctly identifying liver disease cases, as well as a high recall in capturing the majority of positive instances.

Similarly, the RBF kernel demonstrated good precision and recall values, with precision of 93.59% and recall of 93.67%. While slightly lower than the polynomial kernel, these metrics still indicate a strong performance in correctly classifying liver disease cases.

Overall, the experimental results confirm the effectiveness of SVM in predicting liver disease expectancy. Both kernels achieved high accuracy rates, with the polynomial kernel exhibiting slightly better performance. These findings suggest that the choice of kernel function can significantly impact the predictive accuracy of the SVM model in the context of liver disease prediction.

It is worth noting that the performance of the SVM algorithm depends on various factors, including the choice of hyperparameters, dataset characteristics, and the presence of class imbalances. Further research and experimentation could focus on optimizing the SVM model by fine-tuning the hyperparameters and exploring other kernel functions to potentially enhance the accuracy and reliability of liver disease prediction.

In our experiment, the study focused on SVM algorithm was evaluated using two kernels polynomial and RBF kernels. We find in the Figure-3, the introduction of the SVM with polynomial kernel estimation has accomplished 94.64% accuracy, while SVM with RBF kernel has achieved 92.84%. It has been suggested that accuracy obtained by the SVM depends on the kernel selected and the parameters.

VI. CONCLUSION

In this study, we conducted a comparative analysis of SVM kernels for predicting liver disease expectancy. Our experimental results demonstrate that the choice of the kernel significantly affects the performance of the SVM classifier. In conclusion, our experimental results highlight the importance of kernel selection in SVM-based models for liver disease expectancy prediction. The polynomial kernel demonstrated superior performance in terms of accuracy, precision, and recall, emphasizing its potential as a reliable choice for this specific application.

REFERENCES

[1] Bernhard Schölkopf and Alex Smola, Learning with kernels. MIT Press, Cambridge, MA, 2002.

[2] G. Bo and H. Xianwu, "SVM multi-class classification," Journal of Data Acquisition & Processing, vol. 21, pp. 334-339, 2006.

[3] G. Ravi Kumar, S. Rahamat Basha, Surya Bhupal Rao, "A Summarization on Text Mining Techniques for Information Extracting from Applications and Issues", Journal of Mechanics of Continua and Mathematical Sciences, Special Issue, No.-5, January (2020) pp 324-332, ISSN:0973-8975.

[4] Dr. G. Ravi Kumar, K. Tirupathaiah and Prof. B. Krishna Reddy, "Client Churn Prediction of Banking and fund industry utilizing Machine Learning Techniques", International Journal of Computer Sciences and Engineering, Volume-7, Issue-6, e-ISSN: 2347 — 2693, PP: 871-875, June 2019,

[5] Han J, Kamber M. Data Mining: Concepts and Techniques[J]. Data Mining Concepts Models Methods &Algorithms Second Edition, 2011, 5 (4)

[6] https://www.medicinenet.com/liverdisease/article.htm

[7] Ian H. Witten and Eibe Frank. Data Mining: Practical machine learning tools and techniques. 2nd ed. San Francisco: Morgan Kaufmann, 2005.

[8] Lam, Yee Hong Brian, Proteomic Classification of Liver Cancer using Artificial Neural Network", 2005

[9] UCI machine learning repository. http://archive.ics.uci.edu/ml/.

[10] Vapnik, V.N. Statistical Learning Theory. John Wiley and Sons, New York, USA, 1998.

[11] Vapnik, V.N. The Natural of Statistical Learning theory. Springer–Verleg, NewYork ,USA1995.