TITLE:

Enhancing flight delays predictions using Support Vector Machine in Comparisons With Bayesian Prediction

T GANESH[1], Dr S Gomathi[2]

T Ganesh[1]
Research Scholar,
Department of Computer Science and Engineering,
Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences,
Saveetha University, Chennai, Tamil Nadu, India, Pin code: 602 105
thatimakulaganesh0352.sse@sasveetha.com

Dr S Gomathi[2]
Corresponding Author,
Department of Computer Science and Engineering,
Saveetha School of Engineering,
Saveetha Institute of Medical and Technical Sciences,
Saveetha University,Chennai,Tamil Nadu, India, Pincode: 602105.
gomathis.sse@saveetha.com

KEYWORDS: Support Vector Machine,Bayesian Predictions, Machine Learning ,Research Flight delay prediction.

ABSTRACT

Aim: The aim of the study is to investigate and compare the effectiveness of Flight delay Predictions using NOVEL SUPPORT VECTOR MACHINE with entropy in comparison With entropy in comparison with Bayesian predictions for better Accuracy.
Materials and Methods: In order to improve classification accuracy, the current study looked into and compared the efficacy of flight delay predictions made with a novel support vector machine (SVM) and entropy to the conventional Bayesian prediction technique. From Kaggle, a rich dataset of flight delays related to previous flight delays was acquired (Bianco, Dell'Olmo, and Odoni 2001). Design, data gathering, and feature extraction. To guarantee improvement, we used metrics to assess their accuracy. A sample size of 80 was used for each group's statistical parameters, with G Power = 0.80 for each group's ten iterations. The Statistical Package for Social Sciences (SPSS) was used to develop two algorithms: Bayesian Predictions and Support Vector Machines.
Result: Based on obtained results, SVM has significantly better accuracy (93.55%) compared to BP accuracy (16.82%) The statistically significant difference between SVM and the BP algorithm was found to be a p-value of 0.000 ($p<0.005$).
Conclusion: We have used the following algorithms, namely Novel Support Vector Machine (SVM) and Bayesian Predictions (BP), to predict the data. From the results, it is proven that the proposed Novel Support Vector Machine (SVM) works better than other algorithms in terms of accuracy

Key words: Support Vector Machine,Bayesian Predictions, Machine learning, research Flight delay prediction.

INTRODUCTION

The ability to anticipate and reduce flight delays is essential to maintaining both operational effectiveness and passenger pleasure in the ever-changing world of air travel. Airlines and airport authorities depend heavily on reliable prediction models because of the growing volume of air traffic and multiple factors that cause delays. Because they are easy to understand and straightforward, traditional regression approaches like Bayesian predictions have been extensively used in dictionaries (Bianco, Dell'Olmo, and Odoni 2001). However, sophisticated machine learning algorithms—like Support Vector Machine (SVM)—have become potent instruments for predictive analytics in recent years, with the potential to increase accuracy and robustness. The purpose of this study is to evaluate how well SVM and Bayesian predictions perform in forecasting aircraft delays**.**Throudelays. Using historical flight data with a variety of factors (weather, airport traffic(Molnar 2020), aircraft type, and airline schedules), we want to assess how well these two methods perform in predicting flight delays. Assuming a linear relationship between the input features and the target variable, Bayesian Predictions is a parametric approach. Although it offers simple readings of the coefficients, its effectiveness could be restricted when handling the complicated and non-linear data patterns that are a part of airline delay prediction(Gelman et al. 2013). However, by mapping input data into a higher-dimensional space and locating the ideal hyperplane that optimizes the margin between distinct classes, SVM, a non-parametric technique, excels at capturing non-linear correlations.Specifically, SVM is particularly promising for predicting the complex linkages causing flight delays because of its capacity to handle non-linear data structures(McCullagh 2018). We seek to evaluate the advantages and disadvantages of SVM and Bayesian predictions with respect to prediction accuracy, resilience against outliers, and generalization to unknown data. This study aims to advance the creation of more dependable and effective methods for controlling air transportation networks by determining the optimal strategy for airline delay prediction.

To sum up, the goal of this research is to determine whether Support Vector Machines can be used to anticipate aircraft delays instead of Bayesian Predictions. We want to offer important insights into the efficacy of these techniques and direct future work toward enhancing delay prediction models for the aviation sector by assessing their performance on actual flight data.

MATERIALS AND METHODS

The Saveetha Institute of Medical and Technical Sciences in Chennai is home to the Saveetha School of Engineering, where the research work was carried out in the Data Analytics lab.For the purpose of predicting flight delays, two groups were chosen to use the Novel Support Vector Machine [SVM] and Bayesian Predictions (BP) method(Mirkin 2011). A sample size of 80 was used for each group's statistical parameters, with G Power=0.80 for ten iterations for each group. SVM and BP, two algorithms, were implemented with the help of the Statistical Package for Social Sciences (SPSS)(Barnes 2015). SVM and BP are our two independent variables for forecasting flight delays and their efficiency.

Materials and Methods:
Data collection: Reliable sources such as airline databases, flight tracking services, and weather databases are used to gather historical flight data covering a sizable time period. This data includes features like departure and arrival times, airline, airport, weather, aircraft type, and previous delay information. The dataset is cleaned and prepared using data preparation

procedures, which include normalizing numerical features, managing missing values, and encoding categorical variables(Ghavamzadeh et al. 2015). Choosing Features: After doing an exploratory data analysis and applying domain expertise, relevant features for delay prediction are found.To choose the most informative features for modeling, feature selection approaches like correlation analysis, feature importance ranking, and domain expertise are used. Model Development and Assessment: The models for Bayesian Predictions and Support Vector Machine (SVM) are implemented using the relevant libraries or frameworks (e.g., scikit-learn in Python). Using methods like cross-validation, the dataset is divided into training and testing sets to provide an objective assessment of the model's performance. Using performance criteria like Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and R-squared (R^2) on the testing set, the SVM and Bayesian Predictions models are trained on the training data. Adjusting Hyperparameters: Hyperparameters for the SVM model include the kernel function selection ,To maximize performance, regularization parameter (C) and kernel-specific parameters (such as gamma for the RBF kernel) are adjusted using methods like grid search and random search. Similar to this, regularization methods like Lasso or Ridge regression can be used to prevent overfitting in Bayesian predictions, and the hyperparameters of each are adjusted appropriately. Model Comparison: Using a variety of assessment criteria gleaned from the testing set, the SVM and Bayesian Predictions models' performances are contrasted. The statistical significance of the performance difference between the two models is assessed using tests like Wilcoxon signed-rank tests and paired t-tests. Sensitivity analysis is used to evaluate how resilient SVM and Bayesian prediction models are to modifications in input data and model assumptions. .To find out about outliers and lessen their negative effects on model performance, one can use robust regression approaches and outlier detection techniques. Interpretability: Both SVM and Bayesian Predictions models' interpretability is evaluated, taking into account elements like the size and importance of the coefficients in Bayesian Predictions and the decision boundaries and support vectors in SVM. Model prediction interpretation is aided by the use of visualization tools like decision boundaries and feature importance plots. Cross-validation: Methods like k-fold cross-validation or leave-one-out cross-validation can be used to evaluate the stability of model performance estimates across various data subsets in order to guarantee the validity of the results.

.Our objective is to perform a thorough comparative analysis of Bayesian predictions and Support Vector Machine predictions for airline delays using these materials and methods. By doing so, we want to determine the best strategy for improving forecast accuracy in air transportation systems.

Support Vector Machine (SVM)

Support Vector Machines (SVMs) are an effective class of supervised learning algorithms that are mostly applied to regression and classification problems. SVM works by locating the best hyperplane in a high-dimensional space to divide data points into distinct classes. The distance between the hyperplane and the closest data point of each class is called the margin, and it is this distance that determines the hyperplane. The data points that are closest to the decision boundary and are vital in identifying the ideal hyperplane are known as support vectors, and the SVM model finds them(Consoli, Recupero, and Saisana 2021). SVM is well-suited for situations where it is necessary to identify complex decision boundaries since the algorithm strives to increase the margin while minimizing the classification error.

In the mathematical formulation, non-linear correlations between features are handled by using different kernel functions to solve a convex optimization problem. SVM has shown promise in a variety of domains, including flight forecasting.
in the actual world.

Procedure for Support Vector Machine(SVM)
Step 1:
Collect and Prepare Data Assume X_train is the feature matrix for training data, y_train is the corresponding labels.
Step 2:
Choose the SVM Kernel Assume a 'linear' kernel for simplicity.
Step 3:
Define SVM Model 4/16 model = SVM(kernel='linear', C=1.0)
Step 4:
Train the SVM Model model.train(X_train, y_train)
Step 5:
Make Predictions Assume X_test is the feature matrix for test data. predictions = model.predict(X_test)
Step 6:
Evaluate Model Performance (Optional) You can use metrics like accuracy, precision, recall, and F1-score. accuracy = calculate_accuracy(predictions, y_test)
Step 7:
Visualize Results.

Bayesian Predictions(BP)
Through the examination of flight delays, Bayesian Predictions has proven to be an invaluable tool in the world of aviation for the accurate delay prediction of flights. BP is predicated on the idea that prediction patterns that are similar are probably members of the same class. Through

the evaluation of case closeness in the feature space, BP efficiently detects commonalities and facilitates the categorization of Flight delay prediction by comparing them to known cases. However, choosing the right parameters is essential to BP's success.,21 In particular, the careful inclusion of pertinent elements and the selection of 'L', the number of Linear examined(Cristianini, Shawe-Taylor, and Department of Computer Science Royal Holloway John Shawe-Taylor 2000). Although BP has shown effectiveness in aviation applications, such as flight delays, its performance may differ depending on the intricacy of the dataset and the particulars of the flight delays predictions.

Procedure for BP:-

Step 1:

Collect and Prepare Data

Assume X_train is the feature matrix for training data, y_train is the corresponding labels.

Step 2:

Choose BP Parameters Assume 'L' is chosen, either based on cross-validation or prior knowledge.

Step 3:

Define BP Model 5/16 model =BP(L=5) # Assume L=5 for simplicity

Step 4:

Train the BP Model (Note: BPis a lazy learner and doesn't explicitly train) In BP, training involves storing the training data.

Step 5:

Make Predictions Assume X_test is the feature matrix for test data. predictions = model.predict(X_test, X_train, y_train)

Step 6:

Evaluate Model Performance (Optional) You can use metrics like accuracy, precision, recall, and F1-score. accuracy = calculate_accuracy(predictions, y_test)

Step 7:

Visualize Results (Not applicable for BP) BP doesn't have a decision boundary like SVM; visualization is more challenging.

Step 8:

End

STATISTICAL ANALYSIS

The analysis was prepared through IBM SPSS version 21. Independent variables and impactful values are considered for both proposed and existing algorithms, Iterations were done with a maximum of 80 samples and for each iteration the recorded accuracy was noted for necessary analysis. The Dependent Variables are indicated as previous data ,airport data,weather conditions and Independent Variables are flight date,origin city,destination city With the corresponding value that is obtained from the iterations, the Independent sample T-test was performed.

RESULTS

Table 1 Shows the various iterations of the Support Vector Machine (SVM) and Bayesian Predictions(BP) efficiency values are compared. Table 2 Shows the Group Statistics Results: An Novel Support Vector Machine (SVM) and Bayesian Predictions for Testing Independent Samples Statistically Among SVM and BP Methods SVM has a mean accuracy of 93.5550 and a BP of 16.8220. SVM has a standard deviation of 1.09526 and a BP of 3.20560. The SVM standard error mean (.34635) and BP of (1.01370) were compared using the T-test. In Table 3, The 2- significant value smaller than 0.000 ($p<0.05$) impacted that our hypothesis holds good for further consideration. Figure 1 shows bar graph comparison on mean accuracy of Support Vector Machine (SVM) and Bayesian Predictions(BP). In x-axis SVM and BP methods Error Bars: +/-2 SD and 95% CI of Error Bars are shown, In y-axis mean accuracy is shown.

DISCUSSION

The primary goal of the project is to determine precise forecasts of flight delays under challenging circumstances. In order to do that, I divided the Aviation manifestations of Flight Delay dataset into ten samples, numbered 1-2000, 1-4000, 1-6000,...,.1-20000, and determined the precise accuracy values for each sample. Additionally, we observed that accuracy values and tests were conducted using SPSS's independent sample T-Test(Bianco, Dell'Olmo, and Odoni 2001). The results showed that SVM's accuracy (93.55%) was significantly higher than BP's (16.82%). The difference between the SVM and BP algorithms was determined to be statistically significant, with a p-value of $p=0.024$ ($p<0.05$)(Mirkin 2011). We made an effort to increase accuracy for every phase in an effective way.

.In this case, Support Vector Machine (SVM) outperforms Bayesian Predictions (BP) in terms of accuracy. The accuracy of flight delays has improved recently because of the combination of machine learning and the aviation industry, especially in the analysis of flight delays. One interesting method is to use Support Vector Machines (SVM) with entropy(Ghavamzadeh et al. 2015). This is because SVMs, which are well-known for their ability to identify the best hyperplanes for data separation, can be enhanced with entropy in order to extract the subtle information that is hidden in battle delay data..By adding entropy, the model becomes more capable of identifying complex patterns related to flight delays, which makes it a reliable instrument for classifying delay predictions. In contrast, the traditional Bayesian Predictions (BP) method(Consoli et al. 2021), which is frequently employed in aviation studies, is not always able to handle the intricate relationships that exist in the aviation sector with regard to flight delays. The sensitivity to irrelevant features and dependence on proximity in feature space

of Bayesian Predictions may make it less effective in identifying subtle patterns that are essential for precise diagnosis.
The potential superiority of SVM with entropy is highlighted by the comparison with BP. Combining the information-rich entropy with the non-linear capabilities of SVMs creates a more thorough and efficient framework for studying Flight Delaying forecasts(Starčević and Marinković 2020), which eventually improves the accuracy of aviation. This novel method of combining SVM with entropy has important ramifications for manual practice in addition to advancing machine learning applications for airport personnel. By using this methodology, which achieves higher accuracy than more conventional techniques like BP, early detection and prediction procedures might be revolutionized and flight outcomes improved in the event of delays and other aviation delay conditions.

CONCLUSION

Our study has demonstrated a substantial and statistically significant difference in accuracy between Novel Support Vector Machine (SVM) and Bayesian Predictions (BP) algorithms forFlight delay predictions in Aviation Industries. The SVM model achieved an impressive accuracy of 93.55%, surpassing theBP accuracy of 16.82%. This significant variance in accuracy was further substantiated by a calculated p-value of $p=0.000 (p<0.05)$, confirming that the superiority of SVM in Prediction of flight delays is not merely a chance occurrence. These findings underscore the potential of SVM as a more reliable and precise tool for Aviation craft Delays prediction, emphasizing the importance of incorporating advanced machine learning techniques to enhance the accuracy and effectiveness of Flight Delay Prediction models. This study contributes to the Rapid growth of traveling research supporting the adoption of SVM in Aviation manifestations, with the goal of improving our ability to provide more accurate and timely Flights Delaying to their Destinations .

DECLARATIONS

Conflict of interests

No conflict of interest in the manuscript.

AUTHORS CONTRIBUTIONS

TG was responsible for collecting data,conducting data analysis,and writing the manuscript.
SG contributed to the conceptualization ,validated the data ,and performed a critical review of the manuscript.

Acknowledgements

The authors extend their thanks to the Saveetha School of Engineering and the Saveetha Institute of Medical and Technical Sciences (previously known as Saveetha University) for their support in providing the infrastructure needed to complete this work successfully

Funding

We thank the following organizations for providing financial support that enabled us to complete the study.
1)Infoziant IT Solutions Pvt. Ltd., Chennai.
2)Saveetha University.
3)Saveetha School of Engineering.
4)Saveetha Institute of Medical and Technical Sciences.

REFERENCES

Barnes, Jeff. 2015. *Microsoft Azure Essentials Azure Machine Learning*. Microsoft Press.

Bianco, Lucio, Paolo Dell'Olmo, and Amedeo R. Odoni. 2001. *New Concepts and Methods in Air Traffic Management*. Springer Science & Business Media.

Consoli, Sergio, Diego Reforgiato Recupero, and Michaela Saisana. 2021. *Data Science for Economics and Finance: Methodologies and Applications*. Springer Nature.

Cristianini, Nello, John Shawe-Taylor, and Department of Computer Science Royal Holloway John Shawe-Taylor. 2000. *An Introduction to Support Vector Machines and Other Kernel-Based Learning Methods*. Cambridge University Press.

Gelman, Andrew, John B. Carlin, Hal S. Stern, David B. Dunson, Aki Vehtari, and Donald B. Rubin. 2013. *Bayesian Data Analysis, Third Edition*. CRC Press.

Ghavamzadeh, Mohammad, Shie Mannor, Joelle Pineau, and Aviv Tamar. 2015. *Bayesian Reinforcement Learning: A Survey*.

McCullagh, P. 2018. *Tensor Methods in Statistics: Monographs on Statistics and Applied Probability*. CRC Press.

Mirkin, Boris. 2011. *Core Concepts in Data Analysis: Summarization, Correlation and Visualization*. Springer Science & Business Media.

Molnar, Christoph. 2020. *Interpretable Machine Learning*. Lulu.com.

Starčević, Dušan, and Sanja Marinković. 2020. *Proceedings of the XVII International Symposium Symorg 2020: BUSINESS AND ARTIFICIAL INTELLIGENCE*. FON.

TABLES AND FIGURES

SVM	Bayesian Predictions(BP)
91.50%	24.00%
93.12%	19.12%
92.42%	17.67%
93.25%	17.56%
92.85%	17.90%
92.46%	15.17%
92.46%	14.75%
94.47%	13.66%
94.81%	13.06%
94.23%	15.38%

Table 1. The various iterations of the Support Vector Machine (SVM) and Bayesian Predictions(BP) efficiency values are compared.

Table 2. Group Statistics

Results: Support Vector Machines (SVM) and Bayesian Predictions for Testing Independent Samples Statistically Among SVM and BP Algorithms SVM has a mean accuracy of 93.5550 and a BP of 16.8270. SVM has a standard deviation of 1.09526 and a BP of 3.20560. The SVM standard error mean (.34635) and BP standard error mean (1.01370) were compared using the T-test.

	Groups	**N**	**Mean**	**Std.Deviation**	**Std.Error Mean**
Accuracy	SVM	10	93.5550	1.09526	.34635
	BP	10	16.8270	3.20560	1.01370

Group Statistics

Table 3. Independent independent sample T-Test is applied for the sample collections with a confidence interval as 95%. After applying the SPSS calculation, it was found that the least squares Bayesian Predictions (BP) have a statistical significance value of 0.000(P<0.05) that shows they are Statistically significant.

		Levene's test for equality of Variances		**T-test for equality means with 95% confidence interval**						
		F	Sig	t	df	sig(2-tailed)	Mean Difference	std .Error Difference	Lower	Upper
Accuracy	**Equal Variances Assumed**	5.590	.030	71.628	18	.000	76.73000	1.07124	74.47942	78.98058
	Equal Variances not assumed			71.628	11.073	.000	76.73000	1.07124	79.08588	79.08588

Graph:-

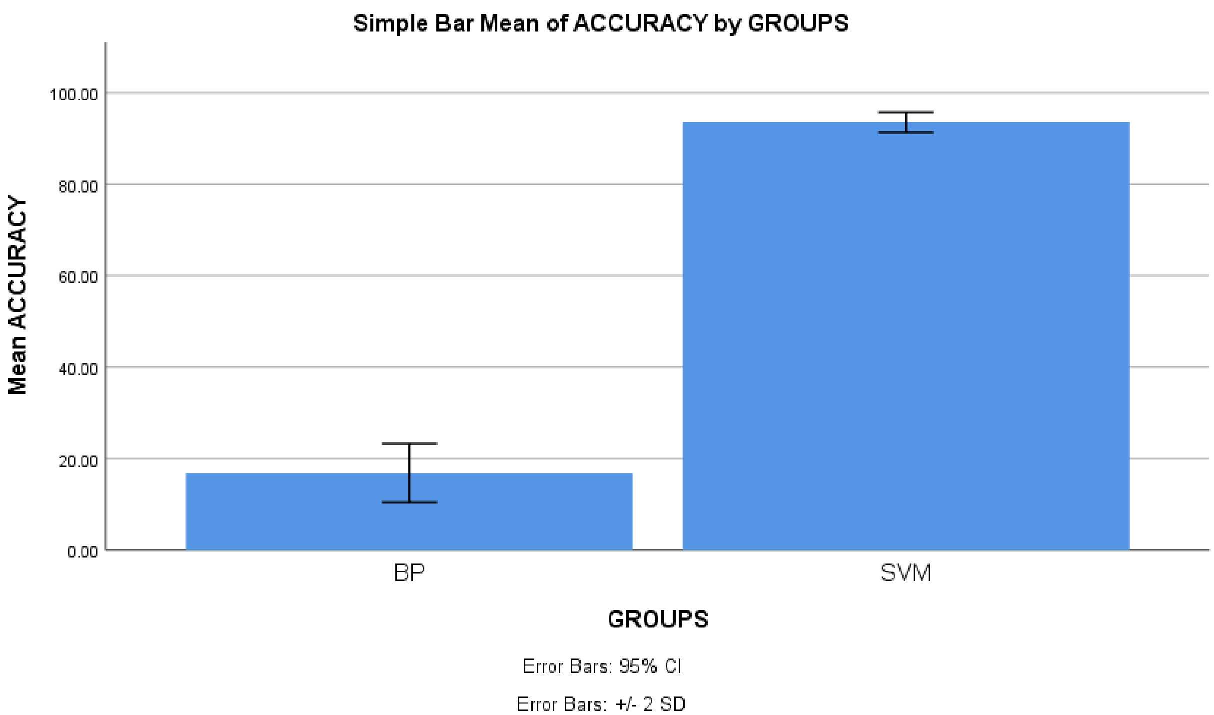
Simple Bar Mean of ACCURACY by GROUPS
Mean ACCURACY
100.00
80.00
60.00
40.00
20.00
0.00
BP
SVM
GROUPS
Error Bars: 95% CI
Error Bars: +/- 2 SD

Fig. 1

. The novel Support Vector Machine has a mean accuracy of 93.52%, where Bayesian Predictions(BP) has a mean of 16.82% in which the novel Support Vector Machine has better accuracy than Bayesian Predictions(BP). The SVM and BP Accuracy rates are shown along with the X-axis: novel Support Vector Machine and Bayesian Predictions Mean keyword identification Y-axis: Mean Accuracy, +/-2 SD, with a 95% Confidence Interval.