IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁒ Impact Factor 8.102 ⁒ Peer-reviewed / Refereed journal ⁒ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12640

Flight Delay Prediction System in Machine Learning using Support Vector Machine Algorithm

Prof. Bharti Sahu[1], Kunal Desale[2], Ashish Patil[3], Prithvi Laishetty[4], Bhuvaneshwar Patil[5]

[*1]Assistant Professor, Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pune, Maharashtra, India.

[*2,3,4,5]Student, Department of Computer Engineering, Dr. D. Y. Patil Institute of Technology, Pune, Maharashtra, India.

Abstract: Flight delays have been extensively studied in recent years. The rising demand for air travel has led to a rise in flight delays. Commercial scheduled flights regularly encounter delays as a result of clogged airspace, a rise in passengers each year, maintenance and safety concerns, unfavorable weather, and the delayed arrival of the aircraft that will be used for the next flight. In order to considerably reduce expenses, academics are looking at how to anticipate and analyze flight delays because it has become a serious problem in the US. The recommended approach therefore makes use of machine learning to predict flight arrival and delay. We have developed a model that implements different machine learning algorithms to predict whether a flight will be delayed or not based on certain characteristics. These characteristics include weather data, past flight data and flight details. We have analyzed numerous algorithms based on past research and settled on the Support Vector Machine algorithm. The SVM algorithm is a supervised machine learning algorithm which is majorly used for classification as well as regression problems. We also aim to help passengers in their stay in the vicinity of the airport in situations where their flights are delayed.

Keywords: Flight delay prediction, Supervised Machine Learning, Classification, Prediction, Support Vector Machine, Air traffic management, predictive analytics.

I. INTRODUCTION

When an airline flight departs or touches down after its planned time, it has experienced a flight delay. An important issue that has an international influence on the aviation sector is flight delays. Every year, it costs the industry billions of dollars. Adverse weather conditions, air traffic jams, late arriving aircraft from earlier flights, maintenance problems, and security concerns are notable causes of delays for commercially scheduled flights. In recent years, there has been a lot of research into flight delays. Flight delays have increased as a result of the increased demand for air travel. Since it became a significant issue in the US, researchers are looking at how to predict and analyse aircraft delays in order to cut expenses significantly. Therefore, using machine learning techniques, we have forecasted flight arrival and delay in the suggested system. All parties concerned, including passengers and airline corporations, are impacted by delayed flights. Although cost is the most important issue in commercial airlines, this influence might have an impact on a wide range of variables. Airlines and customers are equally impacted by flight delays. Due to delays, flights' primary goals of being quick, inexpensive, and safe have failed. Leisure travel and those sectors that depend on air transportation for their business are impacted as airlines charge passengers more as their expenses rise. The subsequent spill-over on the rest of the economy from this indirect effect drives up the already high cost of flights. For U.S. passenger carriers, the average cost of an aircraft block (taxi plus airborne) time in 2021 was $80.52 per minute. The biggest line item, labour cost, increased to $28.14 per minute from 2019 levels by 14.5 percent. The second-largest line item, fuel expenditures, decreased 11% to $22.50 per minute. The expenditures of owning an airplane increased by 26% and 23%, respectively, while all other expenses increased by 5%. At least one out of every four flights into or out of India's four major airports—Delhi, Mumbai, Hyderabad, and Bangalore had a delay in January 2017.

II. LITERATURE REVIEW

Different researchers have studied the development of flight delay prediction systems. Some of these researches focussed on implementing Artificial Neural Networks to predict flight delays. For example, Cheevachaipimol et al. (2021) used Gradient Boosted Trees and Hybrid Deep Learning taking parameters such as internal factors like gate traffic or

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁙ Impact Factor 8.102 ⁙ Peer-reviewed / Refereed journal ⁙ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12640

availability issues and external factors like weather, passengers/luggage and late onboarding. It achieves 79% using GBDT, 51.43% using ANN and 74.38% accuracy using hybrid deep learning [1]. Other recent papers have Example, Gui G. et al. (2022) takes parameters such as bad weather conditions, seasonal and holiday demands, airline policies, technical problems such as problems with airport facilities, baggage handling and mechanical equipment, and the accumulation of delays from previous flights to predict delays using random forest, k-nearest neighbour and support vector machine algorithm [2]. Another research from Lu. M. et al. (2021) achieves 82.87% accuracy using the Gradient Boosting algorithm to predict flight delays [3]. The paper by Tang. Y. et al. (2021) implemented the main parameters and were used in this paper which affected the airline network of the US and were Visibility, wing and departure time and also the aircraft type. It achieved 0.9777 f1 score [4]. Devansh Shah et al. (2020) implemented many algorithms making it like a hybrid system. The algorithm followed Random Forest, Support Vector Machine, K- Nearest Neighbours and Long Short-Term Memory. The parameters used were arrival and departure time of a flight. This paper was able to achieve up to 92% accuracy [5]. Paper by Author SP Laxminarayan et al. (2020) implemented the supervised learning algorithm. The parameters used were the scheduled arrival or departure time. 70% of data was used from the dataset. This paper was able to gain 92.013% accuracy [6]. Borse, Y. et al. (2020) implemented Naive Bayes Classification and Decision Tree Algorithm. The parameters used were Weather, Temperature, Humidity, Rain (calculated in mm), Visibility. This paper was able to achieve 75% accuracy using these algorithms but the Naïve Bayes assumes that all predictors are independent or rarely happening which limited the applicability of this algorithm in real world use cases [7]. Author Bo Zang et al. (2020) implemented the Categorical Boosting Algorithm to predict flight delays. It took parameters such as DEP_DELAY, MONTH, QUARTER, Precipitation, Visibility, Wind Speed etc. It was able to achieve 77% accuracy using the US DOT Airline On-Time Performance dataset [8]. The paper published by Author Garg. R. et al. (2020) implemented the Random Forest Algorithm to predict flight delays. It took parameters such as Departure airport, destination airport, Scheduled timings, weather condition, wind direction, traffic flow. It was able to achieve 90.2% accuracy when implementing binary classification and 70% accuracy when classifying into four categories [9]. Another research by SP Laxminarayan et al. (2020) paper implements the Gradient Boosting Regression to predict flight delays. It predicts using the Atlanta International Airport Dataset. It achieves 79.7% with gradient booster as a classifier with a limited data set [10]

III. PROPOSED SYSTEM

Flight delay detection is a critical task that can benefit greatly from machine learning (ML) techniques. A proposed system for flight delay detection in ML would involve several key components. First, the system would need to gather data on flights, including historical flight data as well as real-time data on current flights. This data would include information such as the departure and arrival times, the weather conditions at both the departure and arrival airports, and any other relevant information that could affect the likelihood of a delay. Next, the system would need to pre-process and clean the data, removing any irrelevant or erroneous information and transforming the remaining data into a format that can be used by ML algorithms. The ML algorithms used for flight delay detection could include a variety of techniques, such as decision trees, random forests, and neural networks. These algorithms would be trained on the pre-processed data to predict the likelihood of a flight delay based on the available information. Finally, the system would need to provide actionable insights to airline operators and other stakeholders. This could include alerts when a flight is at risk of being delayed, recommendations for how to mitigate the risk of delay, and visualizations of the data to help operators better understand the patterns and trends that affect flight delays.

IV. USER INTERFACE

When the user first lands on the interface, they are greeted with a simple search bar. Here, they can search for a specific flight using the flight number or departure/arrival airport codes. Once the user enters their search query, the interface will display a page showing the flight details such as flight number, departure time, and arrival time. It will also display a prediction of the likelihood of the flight being delayed and the expected length of the delay. To make the prediction, the interface will use a combination of historical flight data, current weather conditions, and other relevant factors. The prediction will be displayed using a simple graph or bar chart, showing the probability of delay and the expected length of the delay. To help users better understand the prediction, the interface may also provide additional information or explanations about the factors that are contributing to the prediction. This could include Further Assistance.

V. SOFTWARE INTERFACE

Users will receive precise and timely estimates of flight delays using the software interface for machine learning-based flight delay prediction. It makes predictions about the chance of a flight being delayed and the anticipated length of the delay by analysing numerous data sources, including past flight data and other pertinent parameters, using a machine

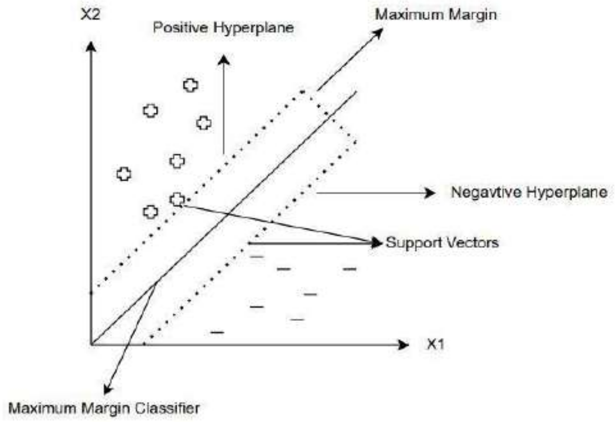
DOI: 10.17148/IJARCCE.2023.12640

learning algorithm like SVM. Users may quickly and simply comprehend the forecast by using the interface, which presents this information in a user-friendly manner together with additional visualisations. Additionally, it offers more details and justifications on the variables influencing the forecast, assisting users in making well-informed choices regarding their trip arrangements.

VI. ALGORITHM AND RELATED MATHEMATICAL MODELS

SVM (Support Vector Machine) Algorithm:

A machine learning approach called Support Vector Machines (SVM) is utilised for classification, regression, and outlier detection. It is frequently applied to a variety of disciplines, including bioinformatics, natural language processing, and image recognition. The program finds the best hyperplane for classifying data into distinct groups. SVM looks for a line that divides the data points into two classes in a two-dimensional space. The ideal hyperplane is the line that maximises the separation between the nearest points of each class. If the data points cannot be separated linearly, SVM uses a kernel function to map them onto a higher-dimensional space. There is a hyperplane that divides the data points into various classes in this higher-dimensional space. The capacity to tackle non-linear classification issues is another feature of SVM. To achieve this, the input data is transformed into a higher-dimensional feature space where a linear boundary can be discovered for dividing the classes. The inner product of two feature vectors in the converted space is calculated using the kernel function. SVMs are effective when dealing with high-dimensional data and can handle many features. SVMs are also robust to overfitting and can generalize well to new data. However, SVMs can be computationally expensive when dealing with large datasets and can be sensitive to the choice of hyperparameters.



The formula for SVM can be broken down into the following steps:

1) Given a set of training data points (x1, y1), (x2, y2), ..., (xn, yn), where xi represents the feature vector and yi represents the class label (-1 or 1).

2) The SVM algorithm finds a hyperplane that maximizes the margin between the two classes.

3) The hyperplane is defined by the equation w.x + b = 0, where w is the normal vector to the hyperplane and b is the bias term.

4) The margin is the distance between the hyperplane and the closest data points from each class.

5) The optimization problem for finding the hyperplane can be formulated as:

6) Minimize (1/2) * ||w||^2 subject to yi(w.xi + b) >= 1 for all I This is a quadratic optimization problem with linear constraints.

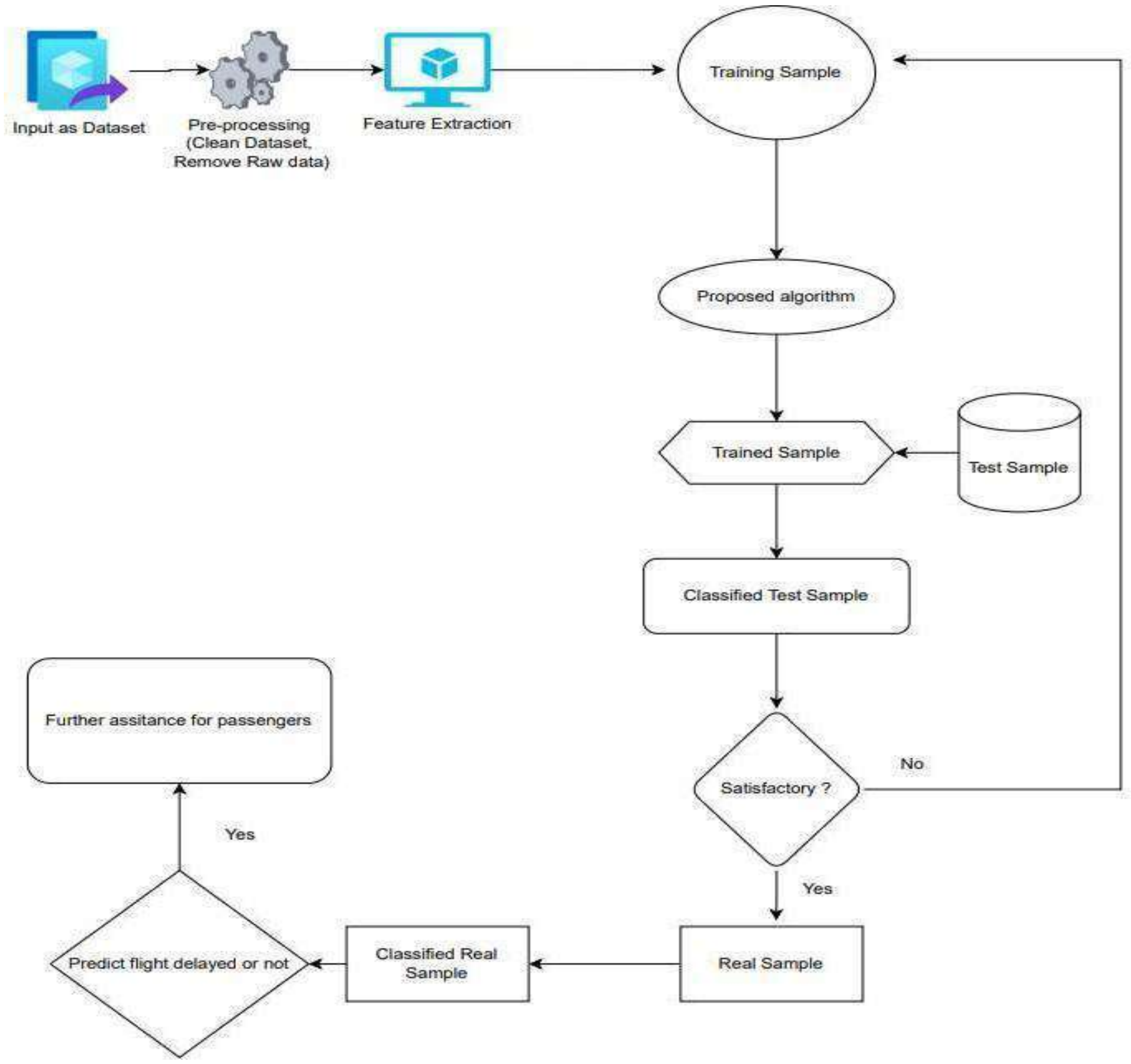
7) Once the hyperplane is found, new data points can be classified by computing w.x + b and checking the sign of the result.

The SVM algorithm can be further extended to handle non-linearly separable data by using kernel functions to transform the data into a higher-dimensional space where a linear hyperplane can separate the classes.

DOI: 10.17148/IJARCCE.2023.12640

VII. MODELLING AND ANALYSIS

SYSTEM ARCHITECTURE:



VIII. ALGORITHM PSEUDO CODE

```
def ModelTraining() :
data = pd.readcsv(”dataset1.csv”)
data.head()
data = data.dropna()
”””One Hot Encoding”””
le = LabelEncoder()
data[’MONTH’] = le.fittransform(data[′MONTH′])
data[’DAY’] = le.fittransform(data[′DAY′])
```

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12640

```
data[’AIRLINE’] = le.fittransform(data[′AIRLINE′])
data[’FLIGHTNUMBER′] = le.fittransform(data[′FLIGHTNUMBER′])
data[’TAILNUMBER′] = le.fittransform(data[′TAILNUMBER′])
data[’ORIGINAIRPORT′]= le.fittransform(data[′ORIGINAIRPORT′])
data[’DESTINATIONAIRPORT′]= le.fittransform(data[′DESTINATIONAIRPORT
data[’DISTANCE’] = le.fittransf orm(data[′DISTANCE′])
data[’SCHEDULEDARRIVAL′]= le.fittransform(data[′SCHEDULEDARRIVAL′])
data[’ARRIVALTIME′] = le.fittransform(data[′ARRIVALTIME′])
”””Feature Selection =¿ Manual”””
x = data.drop([’YEAR’,’Delay’], axis=1)
data = data.dropna()
print(type(x))
y = data[’Delay’]
print(type(y))
x.shape
from sklearn.modelselectionimporttraintestsplit
xtrain, xtest, ytrain, ytest = traintestsplit(x, y, testsize = 0.30, randomstate = 4)

from sklearn.svm import SVC
svcclassifier = SVC(kernel=’linear’)
svcclassifier.fit(xtrain, ytrain)
ypred = svcclassifier.predict(xtest)
print(ypred)
print(”=” * 40)
print(”==========”)
print(”Classification Report : ”,(classificationreport(ytest, ypred)))
print(”Accuracy : ”, accuracyscore(ytest, ypred) ∗ 100)
accuracy = accuracyscore(ytest, ypred)
print(”Accuracy : ACC = (accuracyscore(ytest, ypred) ∗ 100)
repo = (classificationreport(ytest, ypred))
label4=tk.Label(root,text=str(repo),width=45,height=10,bg=’khaki’,fg=’black’,font=(”TSanc ITC”,14))
label4.place(x=205,y=200)
label5=tk.Label(root,text=”Accuracy:”+str(ACC)+”label5.place(x=205,y=420)
from joblib import dump
dump (svcclassifier,”Fmodel.joblib”)
print(”ModelsavedasFmodel.joblib”)
```

.

IJARCCE ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁘ Impact Factor 8.102 ⁘ Peer-reviewed / Refereed journal ⁘ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12640

IX. RESULTS

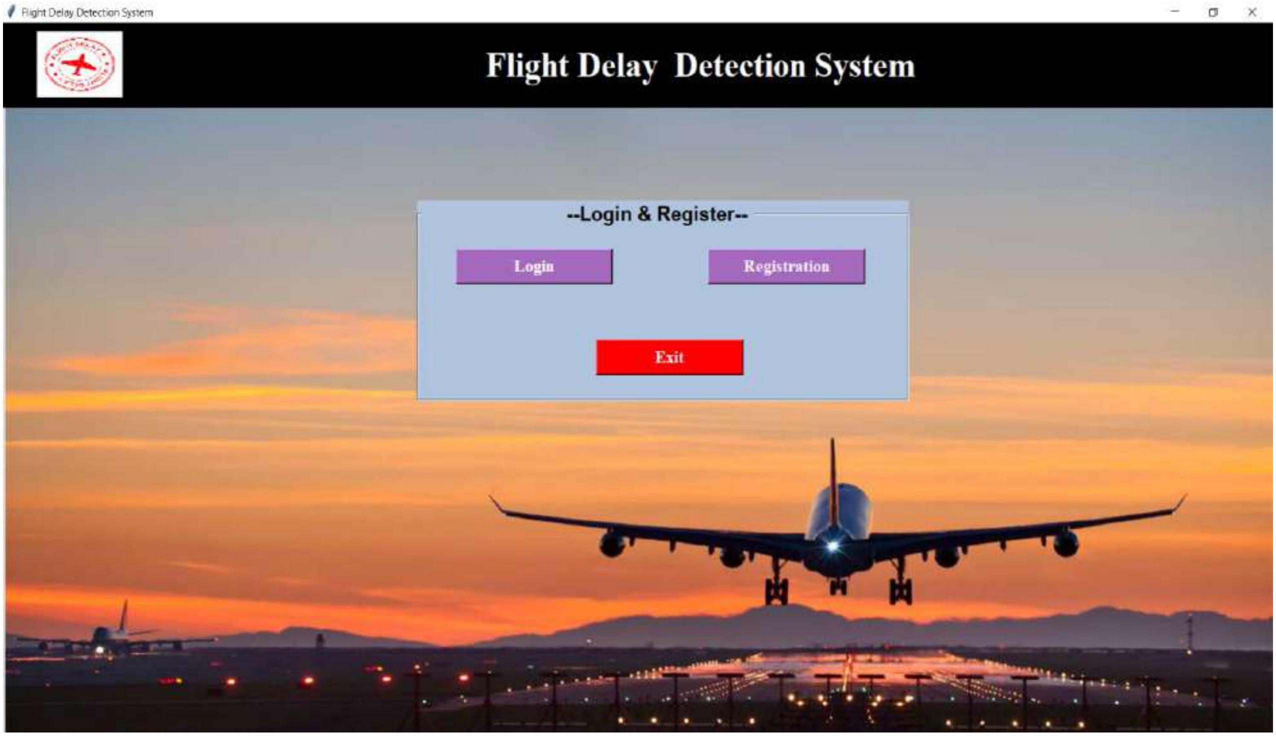


Figure 1: Dashboard

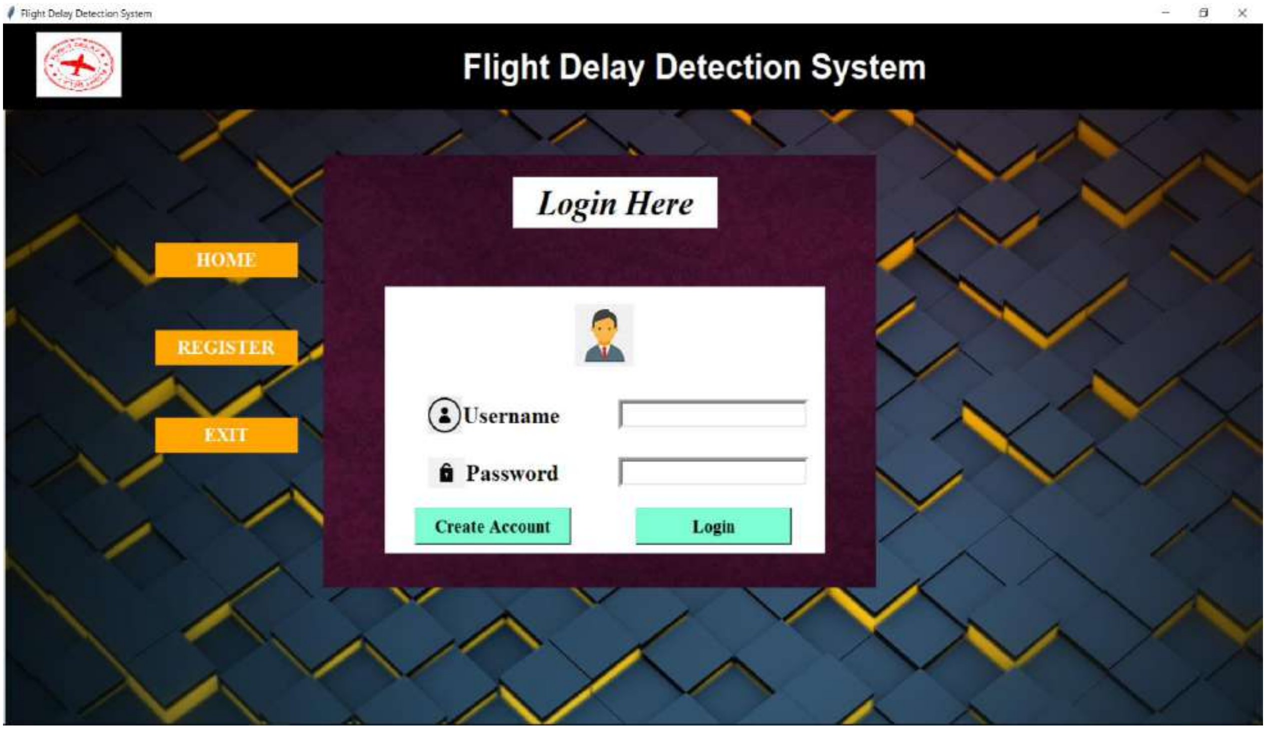


Figure 2: Login Page

ISSN (O) 2278-1021, ISSN (P) 2319-5940

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁜ Impact Factor 8.102 ⁜ Peer-reviewed / Refereed journal ⁜ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12640

Figure 3: Check Flight Interface

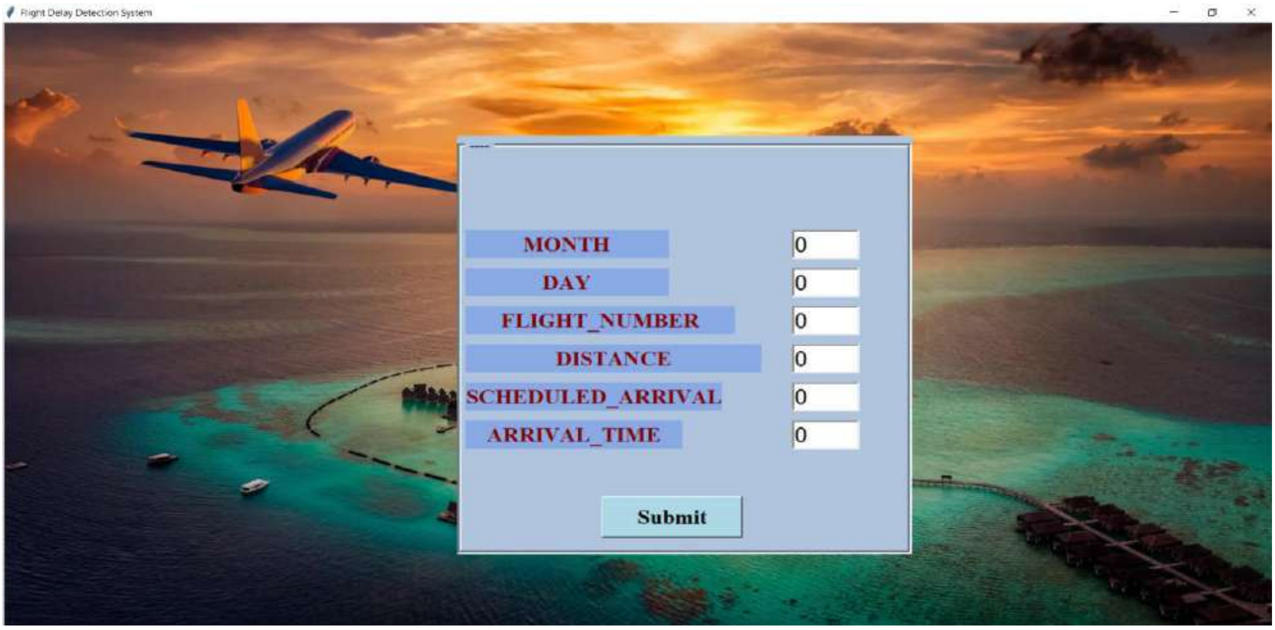


Figure 4: Final Prediction System

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⋇ Impact Factor 8.102 ⋇ Peer-reviewed / Refereed journal ⋇ Vol. 12, Issue 6, June 2023
DOI: 10.17148/IJARCCE.2023.12640

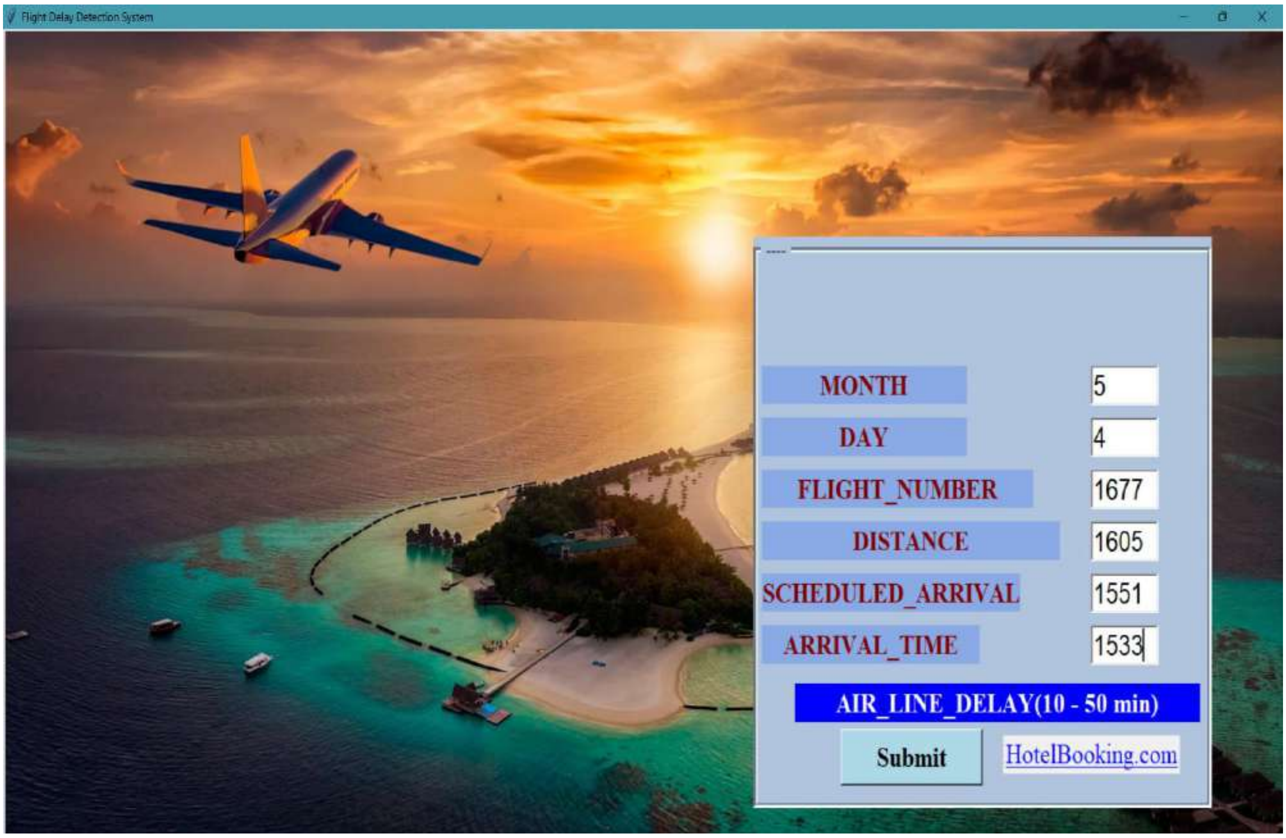


Figure 5: Final Output(1)

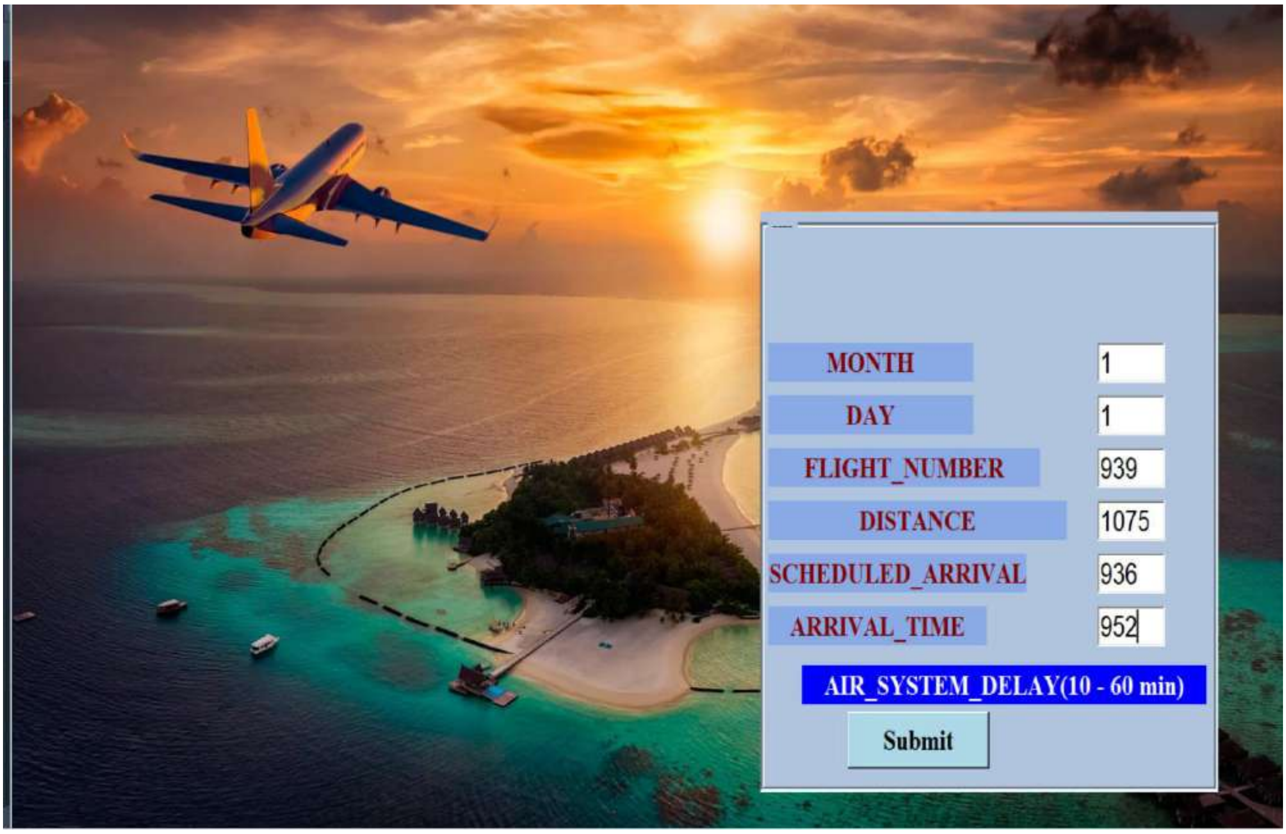


Figure 6: Final Output(2)

International Journal of Advanced Research in Computer and Communication Engineering
ISO 3297:2007 Certified ⁒ Impact Factor 8.102 ⁒ Peer-reviewed / Refereed journal ⁒ Vol. 12, Issue 6, June 2023
DOI: 10.17148/IJARCCE.2023.12640

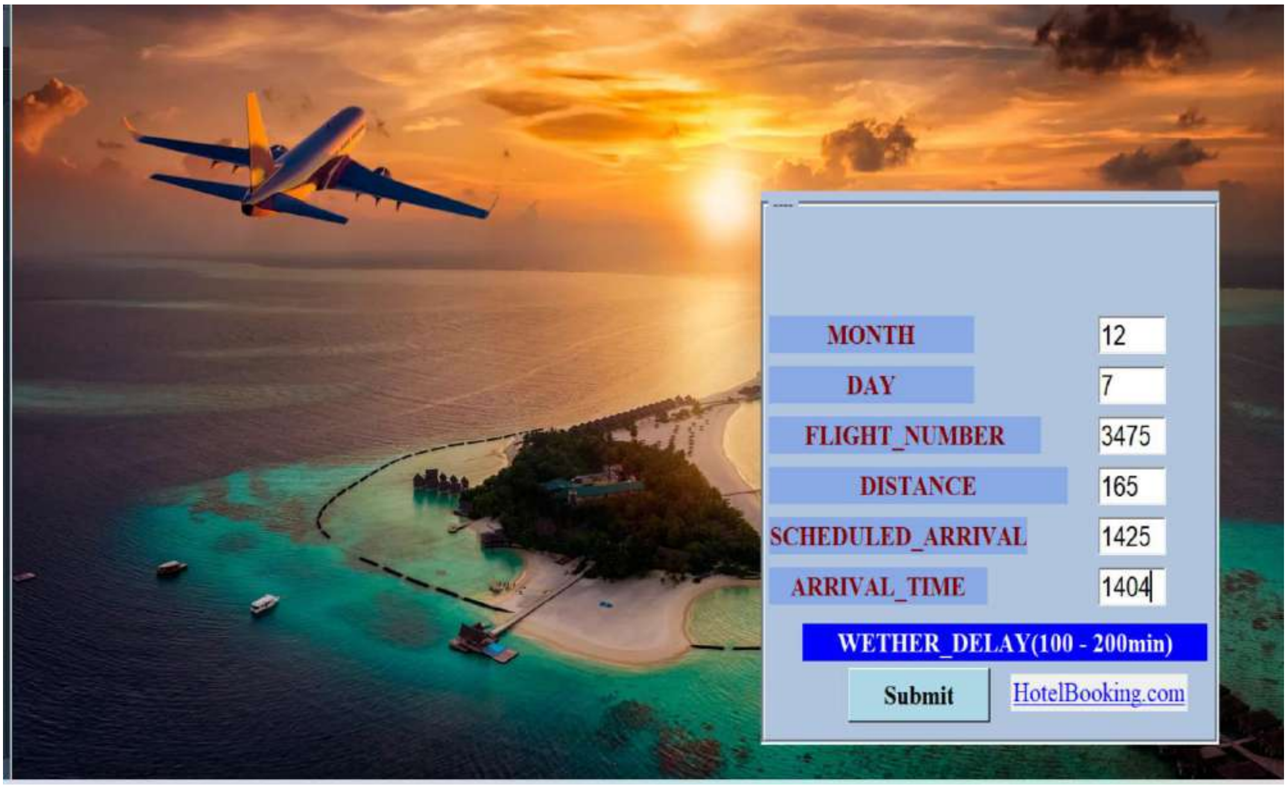


Figure 7: Final Output(3)

X. Conclusion

In conclusion, the project focused on harnessing the power of machine learning, specifically the Support Vector Machine (SVM) algorithm, to predict flight delays. By utilizing historical flight data and various relevant features, the SVM model was trained to accurately classify flights as delayed or on time. Machine learning methods were applied in stages to anticipate flight arrival and delay. An SVM model was produced. The suggested approach employs Support Vector Machines to categorize the dataset. Using the SVM model, to determine and forecast whether a specific flight's arrival will be delayed or not.

XI. Future Scope

The future scope of this paper includes enhanced feature engineering, big data integration, multi-class classification, real-time prediction, ensemble methods, a user-friendly interface, external factor incorporation, and generalization to other domains. These areas offer opportunities to improve accuracy, usability, and applicability. Enhanced feature engineering can incorporate diverse features, while big data integration can leverage large-scale datasets for better insights. Multi-class classification can categorize delays, and real-time prediction can provide up-to-the-minute insights. Ensemble methods can improve accuracy by combining models, and a user-friendly interface can enhance accessibility. Incorporating external factors can enhance performance, and generalization to other domains can extend applicability. Future researchers could also help air travelers in notifying them about the delay and assisting them further in case of delays. Overall, the future scope involves diverse improvements to enhance the model's effectiveness and broaden its impact.

References

[1]. Cheevachaipimol, W., Teinwan, B. and Chutima, P., 2021. Flight Delay Prediction Using a Hybrid Deep Learning Method. Engineering Journal, 25(8), pp.99-112.

[2]. Gui, G., Liu, F., Sun, J., Yang, J., Zhou, Z. and Zhao, D., 2019. Flight delay prediction based on aviation big data and machine learning. IEEE Transactions on Vehicular Technology, 69(1), pp.140-150.

International Journal of Advanced Research in Computer and Communication Engineering

ISO 3297:2007 Certified ⁑ Impact Factor 8.102 ⁑ Peer-reviewed / Refereed journal ⁑ Vol. 12, Issue 6, June 2023

DOI: 10.17148/IJARCCE.2023.12640

[3]. Lu, M., Peng, W., He, M. and Teng, Y., 2021. Flight delay prediction using gradient boosting machine learning classifiers. Journal of Quantum Computing, 3(1), p.1.

[4]. Tang, Y., 2021, October. Airline Flight Delay Prediction Using Machine Learning Models. In 2021 5th International Conference on E-Business and Internet (pp. 151-154).

[5]. Borse, Y., Jain, D., Sharma, S., Vora, V. and Zaveri, A., 2020. Flight Delay Prediction System. Int. J. Eng. Res. Technol, 9(3), pp.88-92.

[6]. Meel, P., Singhal, M., Tanwar, M. and Saini, N., 2020, February. Predicting flight delays with error calculation using machine learned classifiers. In 2020 7th International Conference on Signal Processing and Integrated Networks (SPIN) (pp. 71-76). IEEE.

[7]. Devansh Shah, Ayushi Lodaria, Danish Jain, Lynette D'Mello, Airline Delay Prediction using Machine Learning and Deep Learning Techniques, International Journal of Recent Technology and Engineering (IJRTE) ISSN: 2277- 3878 (Online), Volume-9 Issue-2, July 2020.

[8]. Bo Zhang, School of Traffic and Transportation, Beijing Jiaotong University Beijing, China, Flight Delay

[9]. Prediciton at An Airport using Machine Learning, 2020 5th International Conference on Electromechanical Control Technology and Transportation (ICECTT).

[10]. Garg, R., Gosavi, S., Choulwar, T., Vallal, O. and Aundhakar, S., FLIGHT DELAY PREDICTION BASED ON AVIATION BIG DATA AND MACHINE LEARNING.

[11]. SP Laxminarayan, S Sharbini, R Priyanka and Mohit Kumar., FLIGHT DELAY PREDICTION USING SUPERVISED LEARNING

