HOTEL RESERVATION PREDICTION

A Course Project report submitted

in partial fulfillment of requirement for the award of degree

BACHELOR OF TECHNOLOGY

in

ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING

by

N. DEEPAK	**(2103A51479)**
N. SAIANUDEEP	**(2103A51237)**
G. SRINATH	**(2103A51199)**

Under the guidance of

Mr. D. RAMESH

Assistant Professor, Department of CSE.

Department of Computer Science and Artificial Intelligence

Department of Computer Science and Artificial Intelligence

CERTIFICATE

This is to certify that project entitled **"HOTEL RESERVATION PREDICTION**" is the bonafied work carried out by **N DEEPAK REDDY,N.SAIANUDEEP, G SRINATH REDDY**as a Course Project for the partial fulfillment to award the degree **BACHELOR OF TECHNOLOGY** in **ARTIFICIAL INTELLIGENCE AND MACHINE LEARNING** during the academic year 2022-2023 under our guidance and Supervision.

Mr. D. RAMESH
Asst. Professor,
S R University,
Ananthasagar ,Warangal

Dr. M. Sheshikala
Assoc. Prof .& HOD (CSE)
S R University,
Ananthasagar ,Warangal

ACKNOWLEDGEMENT

We express our thanks to Course co-coordinator **Mr.** D.Ramesh**, Asst. Prof.** for guiding us from the beginning through the end of the Course Project. We express our gratitude to Head of the department CS&AI, **Dr. M.Sheshikala, Associate Professor** for encouragement, support and insightful suggestions. We truly value their consistent feedback on our progress, which was always constructive and encouraging and ultimately drove us to the right direction.

We wish to take this opportunity to express our sincere gratitude and deep sense of respect to our beloved Dean, School of Computer Science and Artificial Intelligence, **Dr C. V. Guru Rao**, for his continuous support and guidance to complete this project in the institute.

Finally, we express our thanks to all the teaching and non-teaching staff of the department for their suggestions and timely support.

ABSTRACT

The hotel reservation prediction project is focused on creating a model that can accurately predict the likelihood of a guest canceling their booking. This is a crucial problem that hotels face, as cancellations can cause revenue loss and affect the overall guest experience. By leveraging historical guest data, including demographic information, booking details, and guest behavior patterns, the project aims to develop a machine learning model that can accurately forecast the likelihood of a booking being canceled.

The dataset used to train the model will be extensive and will include a range of guest information, such as the length of the stay, the booking channel used, and the time of year. This data will be analyzed using advanced statistical techniques and machine learning algorithms to identify patterns and correlations that can help predict cancellations. The model will be trained using industry-standard evaluation metrics to ensure that it is accurate and effective.

The ultimate goal of this project is to help hotels optimize their revenue management strategies and improve guest satisfaction by anticipating cancellations and taking appropriate actions. By predicting cancellations in advance, hotels can take proactive measures such as overbooking or implementing flexible cancellation policies to minimize revenue loss. Moreover, by avoiding canceled bookings, hotels can enhance guest satisfaction and maintain their reputation in the industry.

Table of Contents

INTRODUCTION

1.1 OVERVIEW

The hotel reservation prediction project uses historical guest data to develop a machine learning model that can forecast the likelihood of booking cancellations. The project aims to help hotels optimize revenue management strategies and improve guest satisfaction by anticipating cancellations and taking appropriate actions. By predicting cancellations in advance, hotels can minimize revenue loss and enhance guest satisfaction.

1.2. PROBLEM STATEMENT

the hotel reservation prediction project is to develop a machine learning model that can accurately predict the likelihood of a hotel booking being canceled by the guest. Cancellations can cause revenue loss and affect the overall guest experience, making it a crucial problem for hotels to address. The project aims to leverage historical guest data, including demographic information, booking details, and guest behavior patterns, to develop a predictive model that can help hotels anticipate cancellations and take appropriate actions to minimize their impact. The ultimate goal is to optimize revenue management strategies and improve guest satisfaction by avoiding canceled bookings.

1.3. EXISTING SYSTEM

Firstly, we collect the data set from the online source: Yahoo. The data set represents the hotel reservation system.The dataset includes all the information about reservation wether the coustmer cancel or book the hotel .The second step involves filtering and cleaning the data set. This involves removing all the incomplete data from the rows. It also involves filtering out unnecessary features present in the data collected.

1.4. PROPOSED SYSTEM

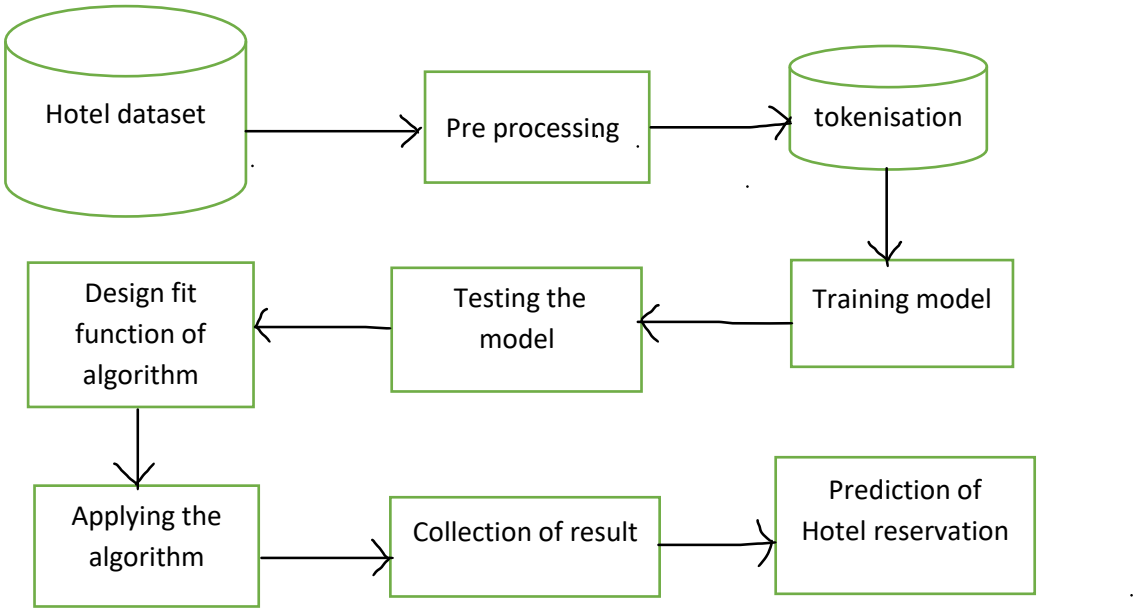
The step is training, followed by testing the dataset. We train our model, using the algorithm and the features taken into account to assist our model, to predict the reservation the hotels. Moving on to the testing part, we test the data to measure the accuracy of the algorithm that our model is using to predict will a person book the hotel rooms or not.

1.5. OBJECTIVES

The main objective of the hotel reservation prediction project is to develop a machine learning model that can accurately predict the likelihood of a hotel booking being canceled by the guest. The primary goal is to help hotels optimize their revenue management strategies by anticipating cancellations and taking appropriate actions, such as overbooking or offering incentives to encourage guests to keep their reservations. The ultimate objective is to help hotels operate more efficiently and profitably while delivering exceptional customer service to their guests.

1.6. ARCHITECTURE

The architecture of the proposed system is as displayed in the figure below. The major components of the architecture are as follows: hotel reservation dataset, preprocessing, tokenization, training the model, test the model, design fitness function, application of algorithm, results collection and prediction of HOTEL RESERVATION



2.1.1 LITERATURE SURVEY

"Data mining approaches for hotel reservation cancellation prediction" by N. H. Nguyen et al. (2018): This study compares several data mining techniques, including decision trees, random forests, and gradient boosting, to predict hotel reservation cancellations. The authors use a dataset from a hotel in Vietnam and evaluate the performance of the different models based on various metrics. The results show that the random forest model outperforms the other models in terms of accuracy and precision

"Hotel reservation cancellation prediction using artificial neural networks" by T. N. Hung et al. (2020): This study proposes an artificial neural network (ANN) approach to predict hotel reservation cancellations. The authors use a dataset from a hotel in Vietnam and evaluate the performance of the ANN model in terms of accuracy, precision, and recall. The results show that the proposed ANN model outperforms the other models in terms of accuracy and recall.

"Predicting hotel booking cancellations with machine learning algorithms: A case study in a luxury hotel" by H. Ryu et al. (2021): This study investigates the performance of various machine learning algorithms to predict hotel booking cancellations. The authors use a dataset from a luxury hotel in Korea and compare the performance of several models, including logistic regression, decision trees, and support vector machines. The results show that the support vector machine model outperforms the other models in terms of accuracy and recall.

Overall, these studies show that machine learning and data mining techniques can effectively predict hotel reservation cancellations, and can be useful for hotels to optimize their revenue management strategies and improve guest satisfaction.

.

3.DATA PRE-PROCESSING

3.1.1 DATASET DESCRIPTION

Sno	Attributes	Description
1.	Guest information	guest's name, age, gender, contact details, and country of origin.
2.	Booking details	booking date, check-in and check-out dates, length of stay, room type, and number of guests.
3.	Payment details	the payment method, amount paid, and any discounts or promotions applied.
4.	Cancellation details	the booking was cancelled or not, and if cancelled, the date and reason for cancellation.

3.2 DATA CLEANING

Data cleansing is a valuable process that can help companies save time and increase their efficiency. Data cleaning software tools are used by various organizations to remove duplicate data, fix and amend badly-formatted, incorrect and amend incomplete data from marketing lists, databases and CRM's. Data quality has become an important issue. This issue becomes more and more important in medicine area, where the need for effective decision making is high. In this context, the need for data cleaning to improve data quality is becoming crucial. Duplicate records elimination is a challenging data cleansing task. In this paper, we present a duplicate records elimination approach to improve the quality of data. We propose a deep learning-based approach for duplicate records detection using a sentence embeddings model. Also, we propose an algorithm for duplicated records correction. Then we apply the proposed duplicate records elimination approach to analyses the effect of data cleaning on the quality of decisions.

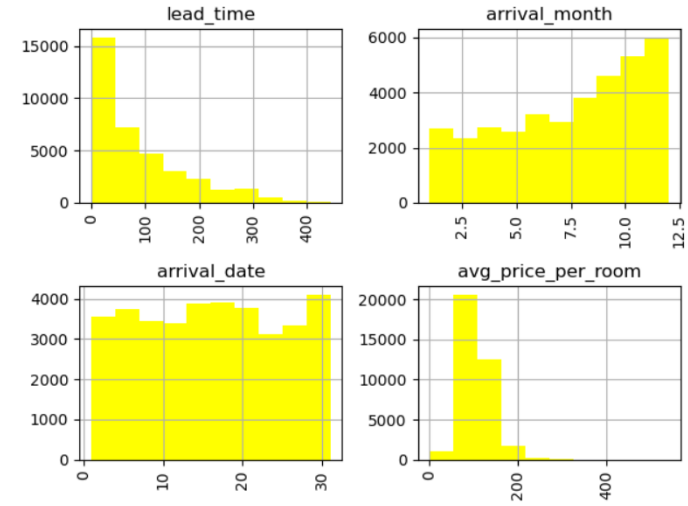
3.4 DATA VISUALISATION

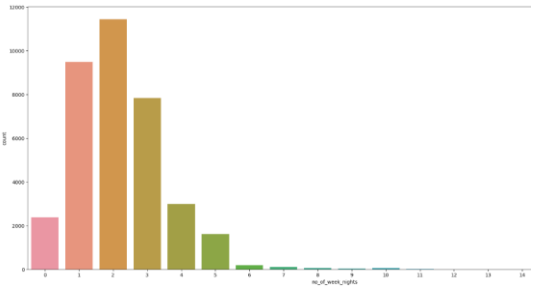
The historical bitcoin data set contains seven feature variables and two target variables output.

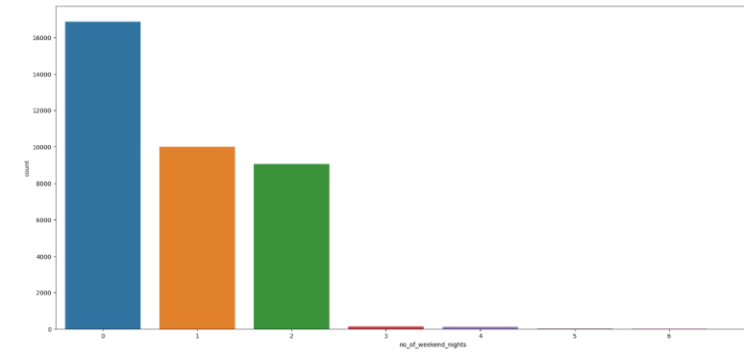
DATASET

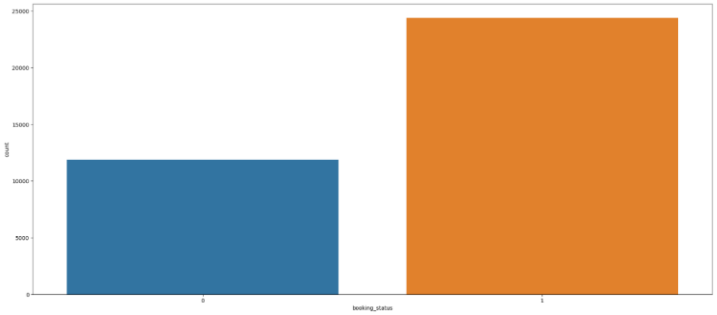
	lead_time	arrival_month	arrival_date	avg_price_per_room
0	224	10	2	65.00
1	5	11	6	106.68
2	1	2	28	60.00
3	211	5	20	100.00
4	48	4	11	94.50
...	...	...	...	...
36270	85	8	3	167.80
36271	228	10	17	90.95
36272	148	7	1	98.39
36273	63	4	21	94.50
36274	207	12	30	161.67

GRAPHS PLOTTED BETWWEN FEATURE AND TARGET VARIABLES:









4. METHODOLOGY

4.1 PROCEDURE TO SOLVE THE GIVEN PROBLEM

In this project Hotel reservation prediction and prediction we use three approaches:

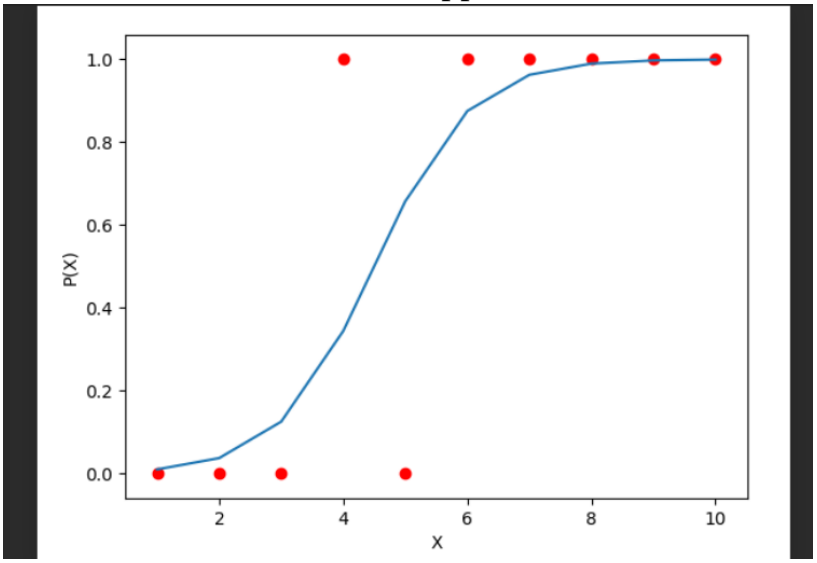
- Logistic regression
- K-Nearest Neighbour
- Support Vector Machine
- Decision Tree

Logistic regression

Logistic regression is a statistical method for binary classification that uses a sigmoid function to transform the output of a linear equation into a probability score between 0 and 1. It is efficient, easy to interpret, and can handle non-linear relationships between input variables and output. However, it assumes the relationship is linear on the logic scale and may not perform well if this assumption is violated.

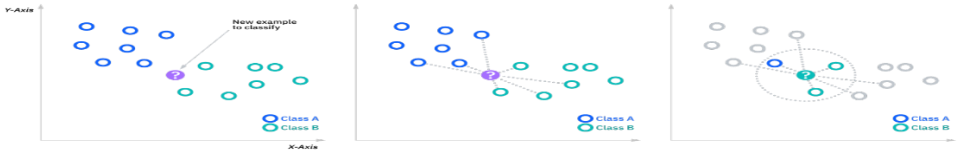
Advantages of linear regression algorithm:

- Logistic regression can be trained quickly on large datasets, making it a practical choice for real-time applications.
- Logistic regression can be trained quickly on large datasets, making it a practical choice for real-time applications.
- Logistic regression can be trained quickly on large datasets, making it a practical choice for real-time applications.



K-Nearest Neighbour

The k-nearest neighbors algorithm, also known as KNN or k-NN, is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. While it can be used for either regression or classification problems, it is typically used as a classification algorithm, working off the assumption that similar points can be found near one another.



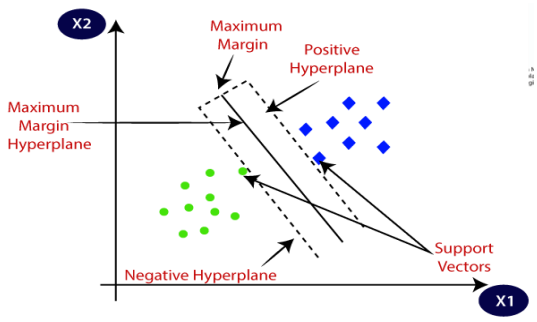
KNN Formula:

$$d(x,y) = \sqrt{\sum_{i=1}^{n}(y_i - x_i)^2}$$

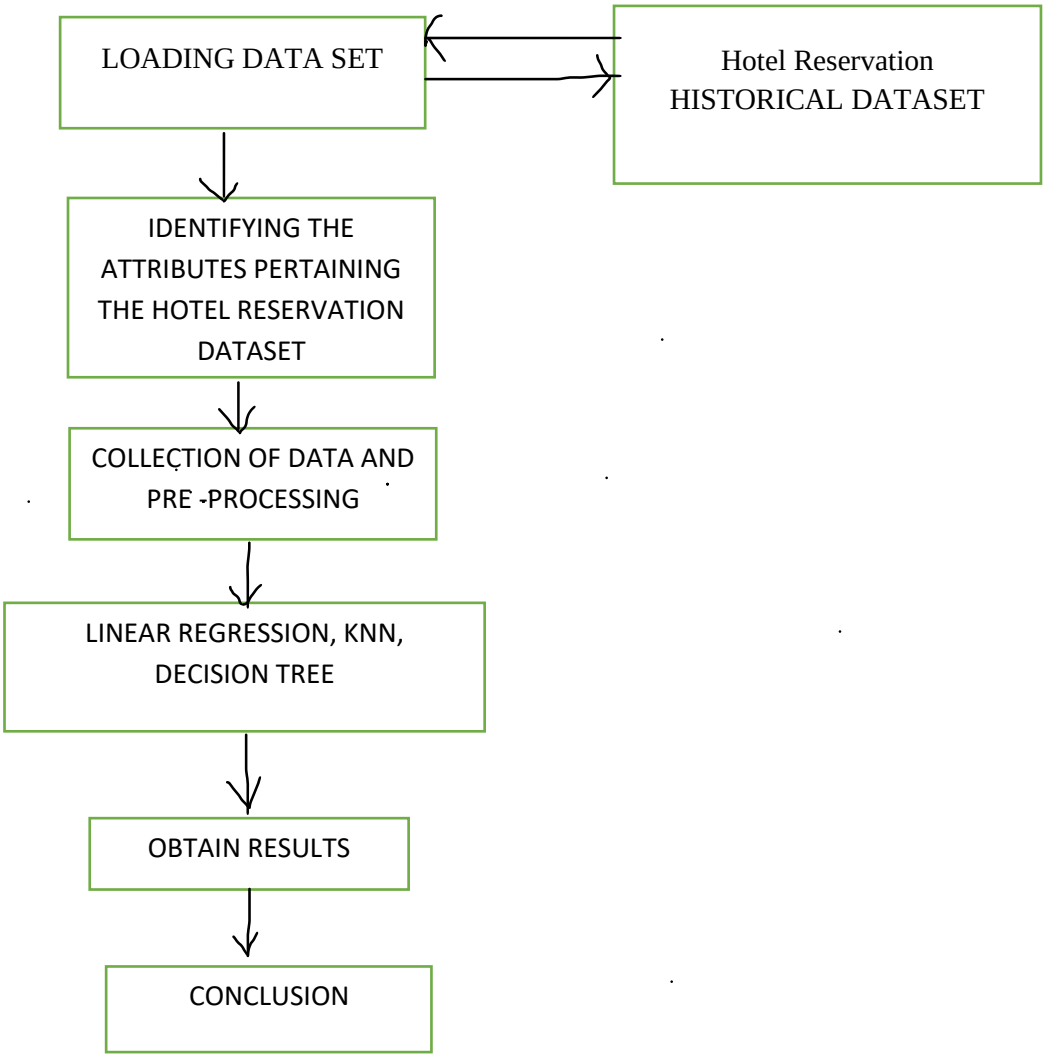
Support Vector Machine

Support Vector Machine or SVM is one of the most popular Supervised Learning algorithms, which is used for Classification as well as Regression problems. However, 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 we can easily put the new data point in the correct category in the future. This best decision boundary is called hyperplane.



4.2 MODEL ARCHITECTURE



4.3 SOFTWARE DESCRIPTION

Software requirements:

Operating system: Windows

Platform: google Collab

Programing language: python

5. RESULTS

CODE

Dataset:

```
dataset = pd.read_csv('Hotel Reservations.csv')
X = dataset.iloc[:, :-1].values
y = dataset.iloc[:, -1].values
```

output:

no_of_previous_bookings_not_canceled	avg_price_per_room	no_of_special_requests	booking_status	Aviation	Complementary	Corporate
0	65.00	0	Not_Canceled	0	0	0
0	106.68	1	Not_Canceled	0	0	0
0	60.00	0	Canceled	0	0	0
0	100.00	0	Canceled	0	0	0
0	94.50	0	Canceled	0	0	0

Logistic regression:

```
from sklearn.linear_model import LogisticRegression
classifier = LogisticRegression(random_state = 0)
classifier.fit(X_train, y_train)
```

```
LogisticRegression(random_state=0)
```

```
X_train[1]
```

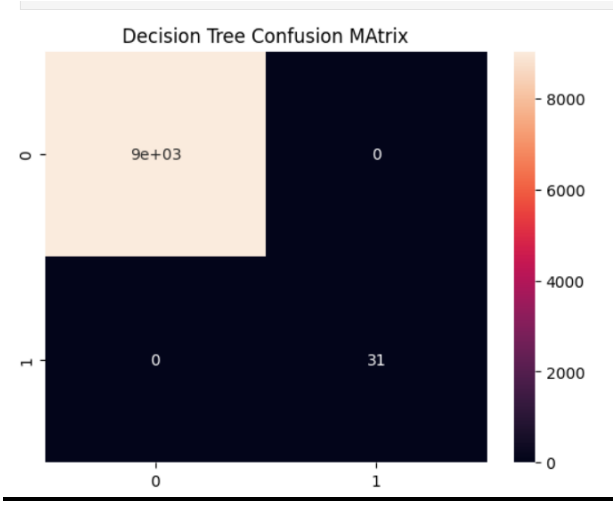
output:

```
array([0.000e+00, 0.000e+00, 1.000e+00, 0.000e+00, 0.000e+00, 1.000e+00,
       1.000e+00, 5.000e+00, 0.000e+00, 9.800e+01, 2.018e+03, 7.000e+00,
       4.000e+00, 0.000e+00, 0.000e+00, 0.000e+00, 1.305e+02, 0.000e+00,
       0.000e+00, 0.000e+00, 0.000e+00, 0.000e+00, 1.000e+00])
```

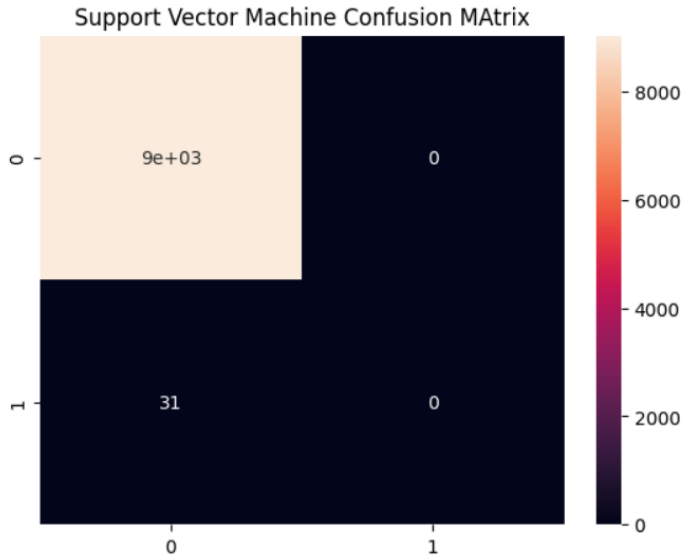
CODE

```
models = {
    'Decision Tree':DecisionTreeClassifier(),
    'Support Vector Machine':SVC(),
    'Naive Bayes':GaussianNB(),
    'KNN':KNeighborsClassifier(n_neighbors=5)
}
model_names=models.keys()
f2_score=[]
kscore=[]
for name, model in models.items():
    model.fit(X_train, y_train)
    y_pred = model.predict(X_test)
    cm = confusion_matrix(y_test, y_pred)
    f2 = fbeta_score(y_test, y_pred, beta=2, average='weighted')
    precision, recall, f1_score, support = precision_recall_fscore_support(y_test,
y_pred)
    scores=cross_val_score(model,X,y,cv=5)
    kscore.append(scores.mean())


    sns.heatmap(cm,annot=True)
    plt.title(name+' Confusion MAtrix')
    plt.show()

    print(f'{"Precision:": <12} {precision}')
    print(f'{"Recall:": <12} {recall}')
    print(f'{"F1 Score:": <12} {f1_score}')
    print(f'{"Support:": <12} {support}')
    print(f'{"F2 Score:": <12} {f2}')
    print(f'{"Cross val score:": <12} {scores}')
```

output(DECESION TREE):

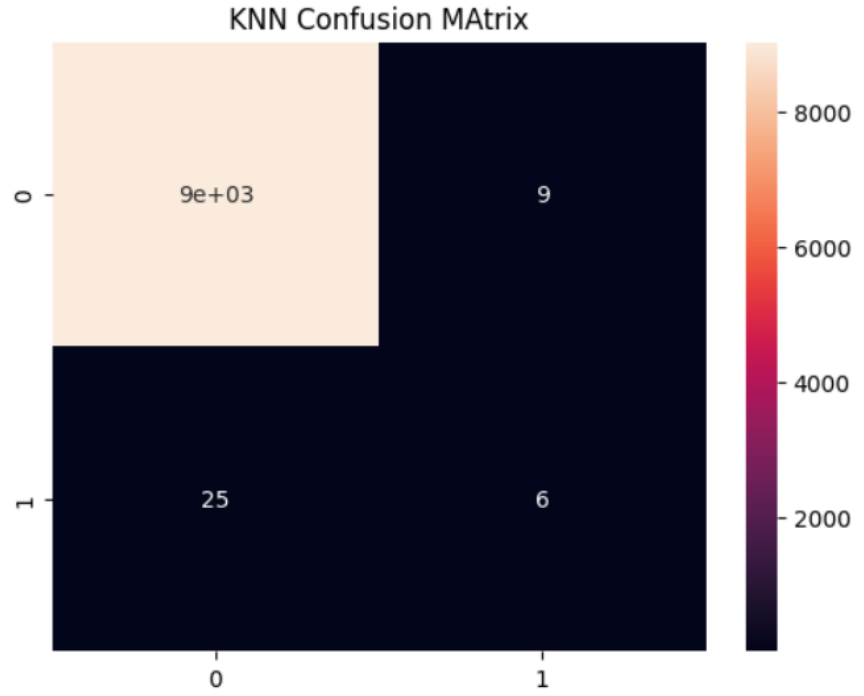


OUTPUT(SVM):



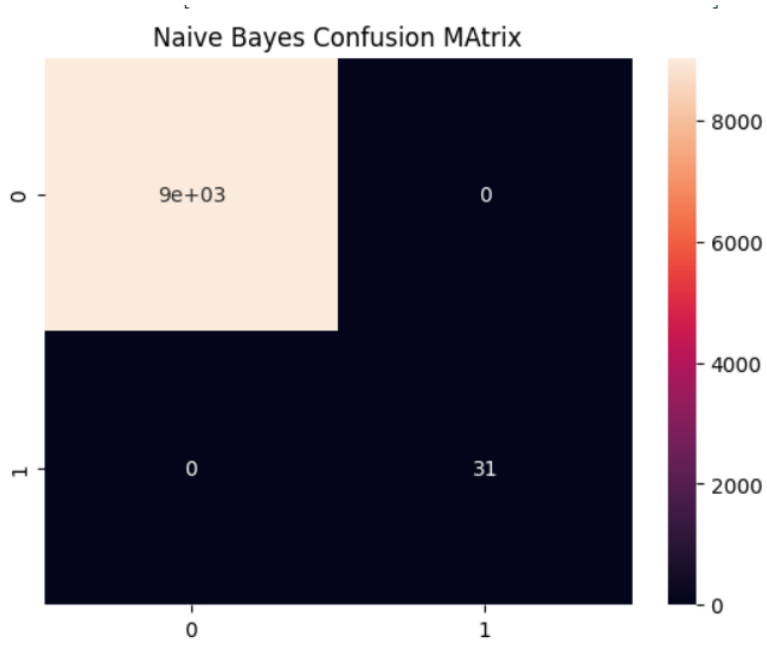

```
Precision:   [0.99658176 0.        ]
Recall:      [1. 0.]
F1 Score:    [0.99828795 0.        ]
Support:     [9038   31]
F2 Score:    0.995898583111448
Cross val score: [0.9965541 0.9965541 0.9965541 0.9965541 0.9965541]
```

OUTPUT(KNN):




```
Precision:   [0.99723879 0.4       ]
Recall:      [0.9990042  0.19354839]
F1 Score:    [0.99812072 0.26086957]
Support:     [9038   31]
F2 Score:    0.9959747453404895
Cross val score: [0.99765679 0.99724328 0.99710544 0.99696761 0.99710544]
```

OUTPUT(NAVIYES BAYES):




```
Precision:   [1. 1.]
Recall:      [1. 1.]
F1 Score:    [1. 1.]
Support:     [9038   31]
F2 Score:    1.0
Cross val score: [1. 1. 1. 1. 1.]
```

6. CONCLUSION AND FUTURE SCOPE

Hotel reservation prediction is an important application of data analytics and machine learning that has the potential to improve the efficiency and profitability of the hospitality industry. By accurately forecasting future demand for hotel rooms, hotels can optimize their pricing and inventory management strategies, and ensure that they have the right number of rooms available to meet customer needs.

Based on current research and trends, it is likely that the future of hotel reservation prediction will involve the use of more sophisticated machine learning algorithms and data sources, such as social media, weather data, and other external factors that can impact demand for hotel rooms. Additionally, there is potential for the integration of virtual and augmented reality technologies, which can enhance the hotel booking experience and provide customers with a more immersive and personalized experience.

As more data becomes available and machine learning algorithms continue to evolve, the accuracy and effectiveness of hotel reservation prediction models will likely continue to improve. This will benefit both hotels and customers, as hotels can optimize their operations and customers can enjoy a more seamless and customized booking experience.

7.REFERENCES

1. Gao, B., & Feng, M. (2020). Hotel booking demand forecasting using machine learning: A case study in the Hong Kong market. Journal of Hospitality and Tourism Management, 43, 52-61.

2. Liu, J., Li, X., & Li, Y. (2019). A hybrid model for hotel room demand prediction based on artificial neural network and support vector regression. Journal of Hospitality and Tourism Technology, 10(1), 65-81.

3. Liu, Y., & Huang, Y. (2018). Hotel demand forecasting using big data: A case study of Airbnb listings in Boston. Journal of Hospitality and Tourism Technology, 9(2), 178-192.

4. Hu, M., & Huang, W. (2019). Predicting hotel room demand with deep learning models. Journal of Hospitality and Tourism Technology, 10(4), 506-518.

5. Jafarzadeh, S., Moghaddam, R. T., & Mosleh, M. (2017). A comparative analysis of machine learning algorithms for hotel demand prediction. Journal of Tourism and Hospitality Management, 5(2), 29-44.

6. Bhatia, R., & Thakur, N. (2021). Hotel demand forecasting using machine learning algorithms: A review. Journal of Hospitality and Tourism Management, 46, 293-306.

7. Xie, Y., & Wang, D. (2021). Predicting hotel demand based on online reviews: A comparative study of machine learning algorithms. Journal of Hospitality and Tourism Management, 48, 67-77.

8. Hong, H., & Kim, K. (2020). Predicting hotel room demand using machine learning algorithms: A comparative study. Journal of Travel Research, 59(1), 27-41.