AIRLINE CUSTOMER SATISFACTION

P Sai Varshith – PES1UG19CS320

P J Subramanya Hande –PES1UG19CS316

Y Nikhil Bhardwaj-PES1UG19CS586

24Slides

INTRODUCTION

The driving factor to any successful service organization is customer satisfaction and his feedback. Customer satisfaction has a strong impact and influence on the customer retention and business effectiveness in the firm
The strategic factors include :
structure, culture, planning and forecasting, customer satisfaction marketing and branding.
This feedback helps us to get insights in which area they can improve to get more profits
Understanding Customer feedback is the driving approach to success

We have used different models to get insight about customer rating

24Slides

DESCRIPTION OF DATASET

01

Attribute types:
Total features-23
Categorical-5
Class Labels-1
Numerical-17

The data given is about airline organization which contains the details of the customers used their airlines and their rating in different aspects.

02

Entries:
There are total of around 1.3 lakh entries.
Inconsistent/missing data:393

PROBLEM TO BE SOLVED

The goal of our project is to predict whether a customer is satisfied with their service based on the details given in dataset,
Also the airlines need to know on which aspect of the services offered by them have to be emphasized more to generate more satisfied customers.

Airlines Passenger Satisfaction

Solution approach

We have started of with encoding the categorical data into numerical data , We have taken care of the NaN values and through co-relation analysis we removed redundant attributes , We used some visualization techniques such as bar graph and pie chart to get better inferences , We have some of the well known classification models until we got the best accuracy in the prediction.

Classification Models

01 KNN

K-NN algorithm stores all the available data and classifies a new data point based on the similarity , It can be used for both regression and classification , in our case we used it for classification and got an accuracy of 70.8% , The disadvantage of KNN is with large data the prediction stage might be slow

02 Logistic Regression

Logistic regression is a more efficient method for binary and linear classification problems , It makes use of a logistic function to predict the class of a attribute , we got an accuracy of 74.03% by using this method. The major limitation of Logistic Regression is the assumption of linearity between the dependent and the independent variables

03 Naïve Bayes Classifier

The crux of this classifier is based on the Bayes theorem , which gives the probability of a data point belonging to hypothesis , We got an accuracy of 81.2% using this approach , The main disadvantages of this approach is that we make an assumption that all the attributes are independent of each other

Classification Models

04 Support Vector Machines

In SVM, we plot each data item as a point in n-dimensional space . Classification is done by plotting a hyperplane that differentiates the two classes of data points. The hyper plane should have maximum distance from the two classes . SVM does not perform well with respect to large datasets . It got an accuracy of 81.2%.

05 Adaboost

It is one of the ensemble based models wherein the predictions of several weak classifiers are combined together to give a better result with a high degree of confidence. In these types of models, all data instances are not given equal weights. . The disadvantage being, it is sensitive to outliers .It got an accuracy of 89.4%
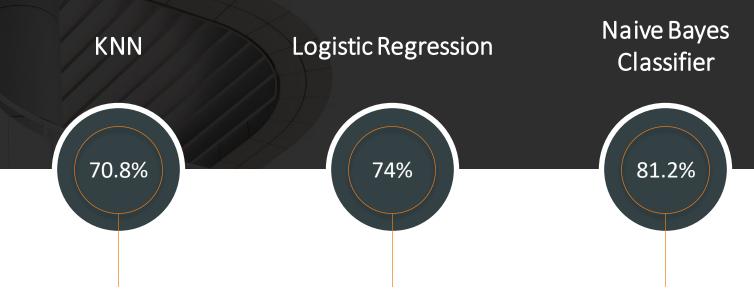
06 Random Forest

It follows the principle of bagging. Over here multiple weak classifiers classify the different datasets parallelly and then are combined together to provide a result. Bagging decreases variance and solves overfitting issues in a model. It gave an accuracy of 95.1%

Performance Analysis

Accuracy Table:

SNo	Model	Accuracy
1	KNN model	70.8
2	Logistic Regression	74.02
3	Naïve Bayes model	81.2

KNN

70.8%

Due to high bias , we do not get great accuracy in KNN

Logistic Regression

74%

Logistic regression tends to underperform when there are multiple or non-linear decision boundaries.

Naive Bayes Classifier

81.2%

We are assuming the attributes , but that is not the case in real world problems.

Performance Analysis

Accuracy Table:

SNo	Model	Accuracy
1	KNN model	70.8
2	Logistic Regression	74.02
3	Naïve Bayes model	81.2
4	SVM model	82.08
5	Adaboost	89.39
6	Random Forest	95.12

SVM

82.1%

SVMs perform poorly in imbalanced datasets

Adaboost

89.4%

It performed comparatively well because there is no overfitting . But as it is sensitive to outliers, it didn't give the best result.

Random Forest

95.1%

There is no overfitting in this and produced one of the best accuracies.

Conclusion

- The 2 most important factors to be considered are Inflight entertainment and seat Comfort
- Onboard Service is the most important flight service attribute and Online Boarding is the most important pre-flight service attribute for passengers
- The flight arrival delay of some less time is positively perceived by passengers but Departure/Arrival Time Convenience is considered a less important attribute by passengers
- The most satisfied group of customers are mostly female business passenger's traveling in the business class

THANK YOU

24Slides