

## **Technical Report: Airline Passenger Satisfaction Prediction**

**Data source (Kaggle):** <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction>

### **Business Overview:**

Airline industry heavily relies upon customer satisfaction for the rate of retention. This project determines and predicts whether a customer is satisfied or not from the quality of services provided by the airline.

### **Data Source and Preprocessing:**

This model uses a dataset from Kaggle called “*Airline Passenger Satisfaction*” which contains more than 100,000 records. The dataset consists of passenger information such as age, gender, rating for various services, type of travel, etc. An airline can benefit in many ways by predicting a customer’s satisfaction and work on improving their services. This dataset was cleaned by filling in the missing values in “Arrival Delay” column. One-hot encoding was also used to convert the categorical features like class and gender into numerical form. The data was split into 70/30 as per the requirement.

### **Model Selection and their Performance:**

For customer satisfaction predictions, several models were used from the best to least performing models are listed below:

1. Random Forest = 96.08% most accurate
2. SVM = 94.98%
3. Decision Tree = 94.32%
4. KNN (k=7) = 92.29%
5. Logistic Regression = 87.55%
6. Naïve Bayes = 85.03%
7. K-Means Clustering = 83.40%

From the above models, Random Forest was the best performing model where cross-validation and confusion matrices was used in confirming the results.

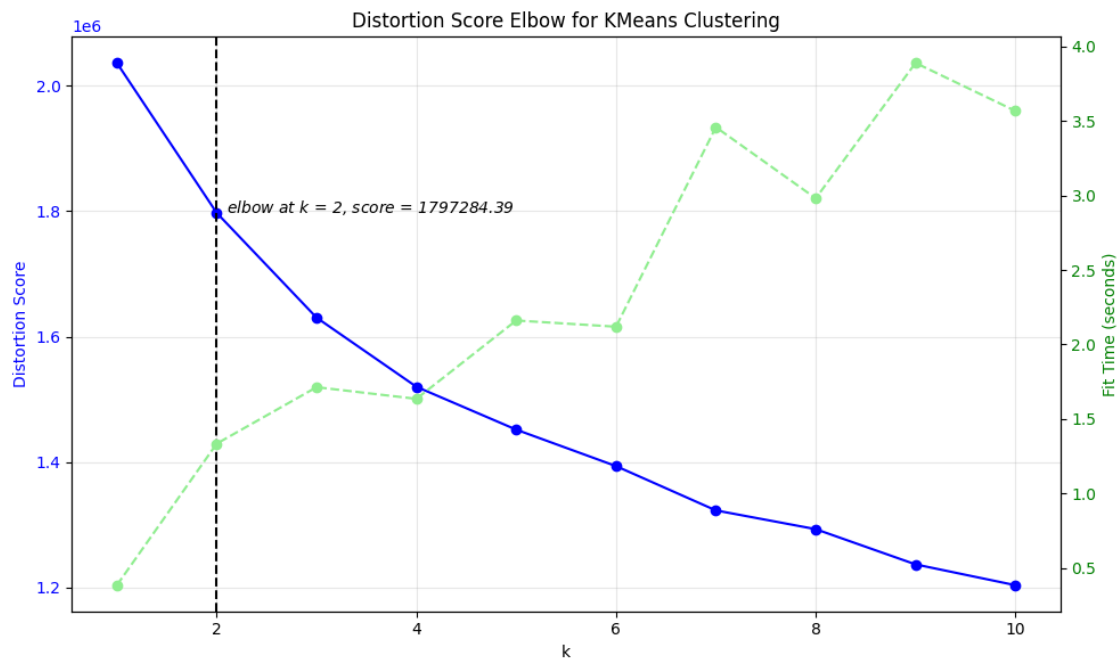
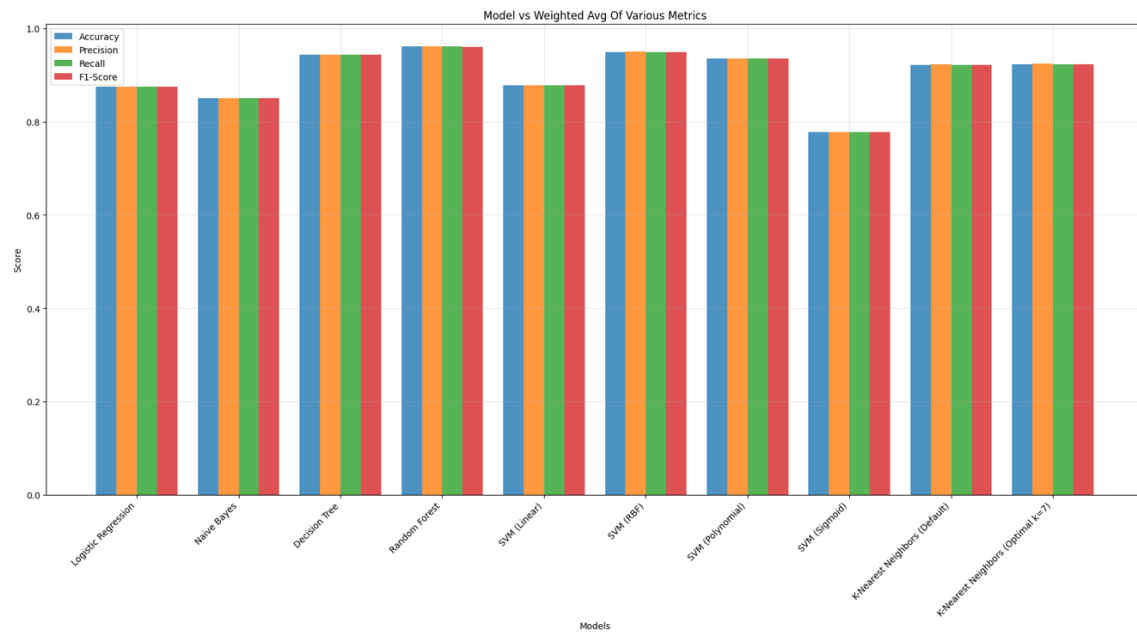
### **Key Findings and Recommendations:**

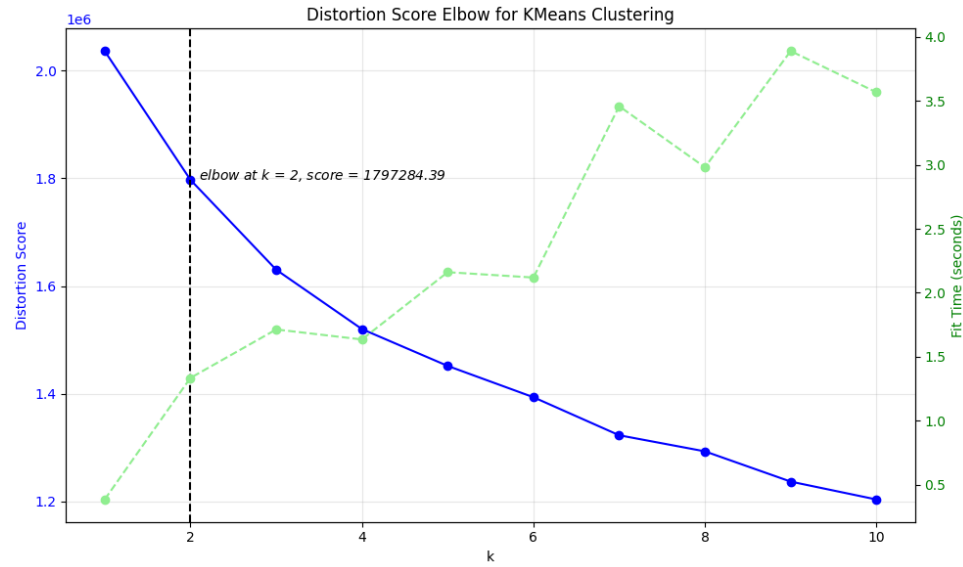
This model has helped in determining key factors that affects the satisfaction level of airline passengers. Features like in-flight wi-fi, online boarding and travel class were highly subjected to affect the customer satisfaction. So, improving these features will help the airline in maintaining customer retention and boost satisfaction. This model will help the airline to flag dissatisfied customers in real time and offer proactive support.

### **Limitations and Future Improvements:**

This model is limited to internal factors and not external factors like weather. Updating the model with real-time data will help in more accurate predictions. Exploring deep learning is highly beneficial for better performance of the model in the future.

## Appendix:





=== Logistic Regression ===				
	precision	recall	f1-score	support
neutral or dissatisfied	0.88	0.91	0.89	17664
satisfied	0.87	0.84	0.85	13508
accuracy			0.88	31172
macro avg	0.87	0.87	0.87	31172
weighted avg	0.88	0.88	0.88	31172
=== Naive Bayes ===				
	precision	recall	f1-score	support
neutral or dissatisfied	0.86	0.87	0.87	17664
satisfied	0.83	0.82	0.83	13508
accuracy			0.85	31172
macro avg	0.85	0.85	0.85	31172
weighted avg	0.85	0.85	0.85	31172
=== Decision Tree ===				
	precision	recall	f1-score	support
neutral or dissatisfied	0.95	0.95	0.95	17664
satisfied	0.93	0.94	0.93	13508
accuracy			0.94	31172
macro avg	0.94	0.94	0.94	31172
weighted avg	0.94	0.94	0.94	31172
=== Random Forest ===				
	precision	recall	f1-score	support
neutral or dissatisfied	0.96	0.98	0.97	17664
satisfied	0.97	0.94	0.95	13508
accuracy			0.96	31172
macro avg	0.96	0.96	0.96	31172
weighted avg	0.96	0.96	0.96	31172
=== SVM (Linear) ===				
	precision	recall	f1-score	support
neutral or dissatisfied	0.87	0.92	0.89	17664
satisfied	0.88	0.83	0.85	13508
accuracy			0.88	31172
macro avg	0.88	0.87	0.87	31172
weighted avg	0.88	0.88	0.88	31172

=== SVM (RBF) ===				
	precision	recall	f1-score	support
neutral or dissatisfied	0.95	0.96	0.96	17664
satisfied	0.95	0.93	0.94	13508
accuracy			0.95	31172
macro avg	0.95	0.95	0.95	31172
weighted avg	0.95	0.95	0.95	31172
=== SVM (Polynomial) ===				
	precision	recall	f1-score	support
neutral or dissatisfied	0.93	0.95	0.94	17664
satisfied	0.94	0.91	0.92	13508
accuracy			0.94	31172
macro avg	0.94	0.93	0.93	31172
weighted avg	0.94	0.94	0.94	31172
=== SVM (Sigmoid) ===				
	precision	recall	f1-score	support
neutral or dissatisfied	0.80	0.81	0.80	17664
satisfied	0.75	0.74	0.74	13508
accuracy			0.78	31172
macro avg	0.77	0.77	0.77	31172
weighted avg	0.78	0.78	0.78	31172
=== K-Nearest Neighbors (Default) ===				
	precision	recall	f1-score	support
neutral or dissatisfied	0.91	0.96	0.93	17664
satisfied	0.94	0.87	0.91	13508
accuracy			0.92	31172
macro avg	0.93	0.92	0.92	31172
weighted avg	0.92	0.92	0.92	31172
=== K-Nearest Neighbors (Optimal k=7) ===				
	precision	recall	f1-score	support
neutral or dissatisfied	0.91	0.96	0.93	17664
satisfied	0.95	0.87	0.91	13508
accuracy			0.92	31172
macro avg	0.93	0.92	0.92	31172
weighted avg	0.92	0.92	0.92	31172