Input

Features: Gender, Customer_Type, Age, Type_of_Travel, Class, Flight_Distance, Inflight_wifi_service, Departure/Arrival_time_convenient, Ease_of_Online_booking, Gate_location, Food_and_drink, Online_boarding, Seat_comfort, Inflight_entertainment, On-board_service, Leg_room_service, Baggage_handling, Checkin_service, Inflight_service, Cleanliness, Departure_Delay_in_Minutes, Arrival_Delay_in_Minutes (total: 22 features)
Target: satisfaction

Ranks

		#	χ^2
1	N Online_boarding		29436.298
2	N Inflight_entertainment		14821.618
3	C Type_of_Travel	2.0	14445.749
4	C Class	3.0	13606.876
5	N Seat_comfort		11319.140
6	N Leg_room_service		9436.536
7	N On-board_service		8999.430
8	N Cleanliness		8438.256
9	N Flight_Distance		5296.184
10	N Inflight_wifi_service		5198.837
11	N Baggage_handling		4567.717
12	N Checkin_service		4555.473
13	N Inflight_service		4366.103
14	N Food_and_drink		3825.385
15	C Customer_Type	2.0	2989.976
16	N Age		2129.704
17	N Ease_of_Online_booking		1981.331
18	N Arrival_Delay_in_Minutes		1478.244
19	N Departure_Delay_in_Minutes		755.365
20	N Departure/Arrival_time_convenient		175.601
21	N Gate_location		75.499
22	C Gender	2.0	7.862

Output

Features: Online_boarding, Inflight_entertainment, Type_of_Travel, Class, Seat_comfort, Leg_room_service, On-board_service, Cleanliness, Flight_Distance, Inflight_wifi_service, Baggage_handling, Checkin_service (total: 12 features)
Target: satisfaction

Above highlighted attributes were chosen as features for the model by employing chi square methods

Test and Score — Tue Mar 26 24, 19:46:31

Settings

Sampling type: No sampling, test on testing data
Target class: None, show average over classes

Scores

Model	Train	Test	AUC	CA	F1	Prec	Recall
Logistic Regression	6.237	0.056	0.914	0.854	0.854	0.854	0.854
Naive Bayes	0.215	0.042	0.929	0.865	0.865	0.865	0.865
kNN	1.519	5.353	0.924	0.856	0.854	0.860	0.856
Neural Network	60.959	0.233	0.982	0.932	0.932	0.932	0.932
Random Forest	1.413	0.190	0.985	0.946	0.946	0.946	0.946
SVM	22.047	2.269	0.514	0.591	0.490	0.645	0.591

The above scores are the evaluation of each classification model.
The best model interms of Accuracy - AUC are, Neural Network and Random Forest. Comparitevely Random Forest is better since it takes less time to train the model.

Confusion Matrix — Tue Mar 26 24, 19:46:37

Confusion matrix for Logistic Regression (showing proportion of predicted)

		Predicted		
		0	**1**	**∑**
Actual	**0**	86.2 %	15.6 %	**14,573**
	1	13.8 %	84.4 %	**11,403**
	∑	**14,905**	**11,071**	**25,976**

Confusion Matrix — Tue Mar 26 24, 19:46:41

Confusion matrix for Naive Bayes (showing proportion of predicted)

		Predicted		
		0	**1**	**∑**
Actual	**0**	87.4 %	14.7 %	**14,573**
	1	12.6 %	85.3 %	**11,403**
	∑	**14,790**	**11,186**	**25,976**

Confusion Matrix — Tue Mar 26 24, 19:46:45

Confusion matrix for kNN (showing proportion of predicted)

		Predicted 0	1	∑
Actual	0	83.3 %	10.6 %	14,573
	1	16.7 %	89.4 %	11,403
	∑	16,269	9,707	25,976

Confusion Matrix — Tue Mar 26 24, 19:46:48

Confusion matrix for Neural Network (showing proportion of predicted)

		Predicted 0	1	∑
Actual	0	93.1 %	6.8 %	14,573
	1	6.9 %	93.2 %	11,403
	∑	14,834	11,142	25,976

The confusion matrix related to NN shows us that only 6.8 % and 6.9 % of the data are false positive and false negatives respectively.

Confusion Matrix — Tue Mar 26 24, 19:46:50

Confusion matrix for Random Forest (showing proportion of predicted)

		Predicted 0	1	∑
Actual	0	94.2 %	5.0 %	14,573
	1	5.8 %	95.0 %	11,403
	∑	14,882	11,094	25,976

The confusion matrix related to Random forest shows us that only 5.0 % and 5.8 % of the data are false positive and false negatives respectively.

Confusion Matrix — Tue Mar 26 24, 19:46:53

Confusion matrix for SVM (showing proportion of predicted)

		Predicted 0	1	∑
Actual	0	58.1 %	27.4 %	14,573
	1	41.9 %	72.6 %	11,403
	∑	24,267	1,709	25,976