Test and Score — Tue Mar 26 24, 19:38:57

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	MCC
Logistic Regression	3.613	0.043	0.932	0.891	0.889	0.887	0.891	0.625
Random Forest	1.304	0.169	0.977	0.954	0.953	0.953	0.954	0.844
kNN	0.992	2.460	0.859	0.860	0.844	0.848	0.860	0.468
Neural Network	58.082	0.145	0.972	0.939	0.938	0.938	0.939	0.793
Naive Bayes	0.110	0.023	0.916	0.874	0.869	0.867	0.874	0.553

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, 19:39:19

Confusion matrix for Logistic Regression (showing proportion of predicted)

		Predicted		
		Loyal Customer	**disloyal Customer**	**∑**
Actual	**Loyal Customer**	92.3 %	26.9 %	**21,177**
	disloyal Customer	7.7 %	73.1 %	**4,799**
	∑	**21,702**	**4,274**	**25,976**

Confusion Matrix — Tue Mar 26 24, 19:39:28

Confusion matrix for Random Forest (showing proportion of predicted)

		Predicted		
		Loyal Customer	**disloyal Customer**	**∑**
Actual	**Loyal Customer**	96.8 %	11.2 %	**21,177**
	disloyal Customer	3.2 %	88.8 %	**4,799**
	∑	**21,343**	**4,633**	**25,976**

The confusion matrix related to Random forest shows us that only 11.2 % and 3.2 % of the data are false positive and false negatives respectively.

Confusion Matrix Tue Mar 26 24, 19:39:35

Confusion matrix for kNN (showing proportion of predicted)

		Predicted		
		Loyal Customer	**disloyal Customer**	∑
Actual	**Loyal Customer**	87.7 %	28.4 %	**21,177**
	disloyal Customer	12.3 %	71.6 %	**4,799**
	∑	**23,255**	**2,721**	**25,976**

Confusion Matrix Tue Mar 26 24, 19:40:25

Confusion matrix for Neural Network (showing proportion of predicted)

		Predicted		
		Loyal Customer	**disloyal Customer**	∑
Actual	**Loyal Customer**	95.6 %	14.2 %	**21,177**
	disloyal Customer	4.4 %	85.8 %	**4,799**
	∑	**21,484**	**4,492**	**25,976**

The confusion matrix related to Random forest shows us that only 14.2 % and 4.4 % of the data are false positive and false negatives respectively.

Confusion Matrix Tue Mar 26 24, 19:40:47

Confusion matrix for Naive Bayes (showing proportion of predicted)

		Predicted		
		Loyal Customer	**disloyal Customer**	∑
Actual	**Loyal Customer**	90.5 %	30.4 %	**21,177**
	disloyal Customer	9.5 %	69.6 %	**4,799**
	∑	**22,086**	**3,890**	**25,976**

Input

Features: Gender, 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, satisfaction (total: 22 features)
Target: Customer_Type

Ranks

		#	χ^2
1	N Age		9346.838
2	C Type_of_Travel	2.0	6812.196
3	N Online_boarding		3289.592
4	N Flight_Distance		3168.534
5	N Departure/Arrival_time_convenient		2535.154
6	C satisfaction	2.0	2073.017
7	N Seat_comfort		1984.458
8	N Inflight_entertainment		996.033
9	N Cleanliness		552.577
10	N On-board_service		274.039
11	N Leg_room_service		246.164
12	N Food_and_drink		222.279
13	C Class	3.0	122.530
14	N Checkin_service		80.408
15	N Gate_location		79.312
16	C Gender	2.0	53.091
17	N Baggage_handling		28.617
18	N Inflight_service		26.610
19	N Ease_of_Online_booking		5.886
20	N Arrival_Delay_in_Minutes		4.714
21	N Departure_Delay_in_Minutes		1.594
22	N Inflight_wifi_service		0.094

Output

Features: Age, Type_of_Travel, Online_boarding, Flight_Distance, Departure/Arrival_time_convenient, satisfaction, Seat_comfort, Inflight_entertainment
Target: Customer_Type