File Mon May 10 21, 21:55:42

File

File name: D:/Documents/COLLEGE/RCSS - PG/SEMESTER 2/CSDA207 DM LAB/satisfaction.xlsx
Format: Microsoft Excel spreadsheet

Data

Data instances: 129880
Features: id, satisfaction_v2, Gender, Customer Type, Age, Type of Travel, Class, Flight Distance, Seat comfort, Departure/Arrival time convenient, Food and drink, Gate location, Inflight wifi service, Inflight entertainment, Online support, Ease of Online booking, On-board service, Leg room service, Baggage handling, Checkin service, Cleanliness, Online boarding, Departure Delay in Minutes, Arrival Delay in Minutes (total: 24 features)

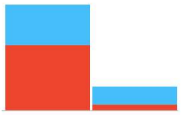
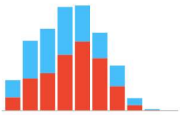
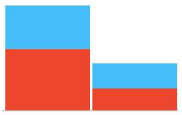
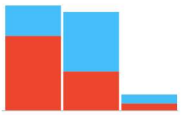
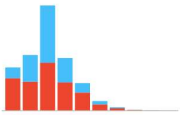
Data Info Mon May 10 21, 21:55:51

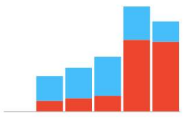
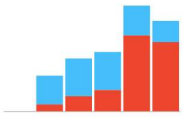
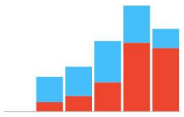
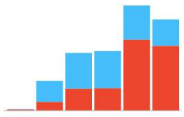
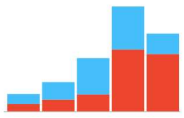
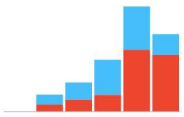
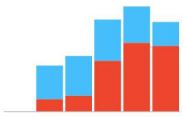
Name: satisfaction
Rows: 129880
Features: 5 categorical, 19 numeric

Preprocess Mon May 10 21, 21:55:59

Settings

Normalize Features: Standardize to μ=0, σ²=1
Impute Missing Values: Average/Most frequent

	Name	Distribution	Center	Dispersion	Min.	Max.	Missing
N	id		64940.50	0.58	1	129880	0 (0%)
C	satisfaction_v2		satisfied	0.689			0 (0%)
C	Gender		Female	0.693			0 (0%)
C	Customer Type		Loyal Customer	0.476			0 (0%)
N	Age		39.43	0.38	7	85	0 (0%)
C	Type of Travel		Business travel	0.619			0 (0%)
C	Class		Business	0.902			0 (0%)
N	Flight Distance		1981.41	0.52	50	6951	0 (0%)
N	Seat comfort		2.84	0.49	0	5	0 (0%)
N	Departure/Arrival time convenient		2.99	0.51	0	5	0 (0%)
N	Food and drink		2.85	0.51	0	5	0 (0%)
N	Gate location		2.99	0.44	0	5	0 (0%)
N	Inflight wifi service		3.25	0.41	0	5	0 (0%)
N	Inflight entertainment		3.38	0.40	0	5	0 (0%)

N	Online support		3.52	0.37	0	5	0 (0%)
N	Ease of Online booking		3.47	0.38	0	5	0 (0%)
N	On-board service		3.47	0.37	0	5	0 (0%)
N	Leg room service		3.49	0.37	0	5	0 (0%)
N	Baggage handling		3.70	0.31	1	5	0 (0%)
N	Checkin service		3.34	0.38	0	5	0 (0%)
N	Cleanliness		3.71	0.31	0	5	0 (0%)
N	Online boarding		3.35	0.39	0	5	0 (0%)
N	Departure Delay in Minutes		14.71	2.59	0	1592	0 (0%)
N	Arrival Delay in Minutes		15.09	2.55	0	1584	393 (0%)

Select Columns

Mon May 10 21, 21:56:22

Input data

Features: id, satisfaction_v2, Gender, Customer Type, Age, Type of Travel, Class, Flight Distance, Seat comfort, Departure/Arrival time convenient, Food and drink, Gate location, Inflight wifi service, Inflight entertainment, Online support, Ease of Online booking, On-board service, Leg room service, Baggage handling, Checkin service, Cleanliness, Online boarding, Departure Delay in Minutes, Arrival Delay in Minutes (total: 24 features)

Output data

Features: Flight Distance, Seat comfort, Departure/Arrival time convenient, Food and drink, Gate location, Inflight wifi service, Inflight entertainment, Online support, Ease of Online booking, On-board service, Leg room service, Baggage handling, Checkin service, Cleanliness, Online boarding, Departure Delay in Minutes, Arrival Delay in Minutes (total: 17 features)
Meta attributes: Age, Customer Type, Gender, Class, Type of Travel
Target: satisfaction_v2

Removed: 1 (id)

Rank

Mon May 10 21, 21:56:35

Input

Features: Flight Distance, Seat comfort, Departure/Arrival time convenient, Food and drink, Gate location, Inflight wifi service, Inflight entertainment, Online support, Ease of Online booking, On-board service, Leg room service, Baggage handling, Checkin service, Cleanliness, Online boarding, Departure Delay in Minutes, Arrival Delay in Minutes (total: 17 features)
Meta attributes: Age, Customer Type, Gender, Class, Type of Travel
Target: satisfaction_v2

Ranks

	#	Info. gain	Gain ratio	Gini	χ^2
Inflight entertainment		0.3175380812411843	0.16104841019038316	0.19326257382391537	36484.71992814663
Ease of Online booking		0.15253823616125317	0.0773859570964429	0.10114381524332988	19767.975600710124
Online support		0.13974080435125869	0.07120928862267918	0.0929334400082158	16550.45145342652
Seat comfort		0.10919138181524901	0.0557470896490972 4	0.07191501997161254	4611.235193497694
On-board service		0.0961107065033727 2	0.04846297242131937	0.06411770476307566	12578.969133399718
Online boarding		0.0908731815530095	0.04550901426273333	0.06081822676490806	12205.452226199348
Leg room service		0.0800198047125794 1	0.040583875500444254	0.054124013151734596	10113.053463800024
Baggage handling		0.0710722663290956 9	0.03691084258099109	0.047666532147602414	6034.653200228009
Cleanliness		0.0685802108493984 7	0.03571543791272118	0.04604817191552868	5820.727816078015
Checkin service		0.0582510768996391 54	0.029273101898804645	0.03910185063872923	7008.150145120008
Food and drink		0.0339849179378338 7	0.017409426656152197	0.023100550345202975	1115.3092046807521
Inflight wifi service		0.0318284515864215 6	0.016080322404131642	0.021674575419346587	5282.837627330566
Flight Distance		0.0228032271670106 65	0.011401613749285873	0.015512665132439485	228.34005383251355
Gate location		0.0081325504931458 74	0.0042188500153132686	0.005590487671158217	161.3033253037877
Arrival Delay in Minutes		0.0075130719524247 88	0.004454501564981297	0.005170137980440415	1937.201399460946
Departure Delay in Minutes		0.0042899494933141 82	0.0025572036923191496	0.002956718936324265	1062.2206441134076
Departure/Arrival time convenient		0.0008803348814527823	0.00046027865629912903	0.0006037160656577889	25.104235981258043

Output

Features: Inflight entertainment, Ease of Online booking, Online support, Seat comfort, On-board service, Online boarding, Leg room service, Baggage handling, Cleanliness, Checkin service (total: 10 features)
Meta attributes: Age, Customer Type, Gender, Class, Type of Travel
Target: satisfaction_v2

Random Forest Mon May 10 21, 21:56:50

Name: Random Forest

Model parameters

Number of trees: 10
Maximal number of considered features: unlimited
Replicable training: No
Maximal tree depth: unlimited
Stop splitting nodes with maximum instances: 5

Data

Data instances: 129880
Features: Inflight entertainment, Ease of Online booking, Online support, Seat comfort, On-board service, Online boarding, Leg room service, Baggage handling, Cleanliness, Checkin service (total: 10 features)
Meta attributes: Age, Customer Type, Gender, Class, Type of Travel
Target: satisfaction_v2

Naive Bayes Mon May 10 21, 21:56:57

Name: Naive Bayes

Tree Mon May 10 21, 21:57:06

Name: Tree

Model parameters

Pruning: at least two instances in leaves, at least five instances in internal nodes, maximum depth 100
Splitting: Stop splitting when majority reaches 95% (classification only)
Binary trees: No

Logistic Regression Mon May 10 21, 21:57:29

Name: Logistic Regression

Model parameters

Regularization: Ridge (L2), C=1, class weights=False

SVM Mon May 10 21, 21:57:36

Name: SVM

Model parameters

SVM type: SVM, C=1.0, ε=0.1
Kernel: RBF, exp(-auto|x-y|²)
Numerical tolerance: 0.001
Iteration limt: 100

kNN Mon May 10 21, 21:57:52

Name: kNN

Model parameters

Number of neighbours: 5
Metric: Euclidean
Weight: Uniform

Test and Score Mon May 10 21, 22:01:11

Settings

Sampling type: Stratified 10-fold Cross validation
Target class: Average over classes

Scores

Model	AUC	CA	F1	Precision	Recall
kNN	0.9579291666019916	0.9041499846011704	0.904227819531647	0.9044682659988468	0.9041499846011704
Tree	0.9459417128206549	0.9140668309208501	0.9141154107004696	0.9142381452061022	0.9140668309208501
SVM	0.5209387644854975	0.620364952263628	0.6111195837144696	0.6607771448163798	0.620364952263628
Random Forest	0.9775003783535181	0.9234909146904835	0.9235020355921382	0.923517611024021	0.9234909146904835
Naive Bayes	0.8910809063653667	0.8034724360948567	0.8034162479772173	0.8033736720396056	0.8034724360948567
Logistic Regression	0.8761775317485826	0.8004542654758239	0.8002611542635204	0.8002050008494483	0.8004542654758239

Target: Average Over Classes

Test and Score Mon May 10 21, 22:01:51

Settings

Sampling type: Stratified 10-fold Cross validation
Target class: satisfied

Scores

Model	AUC	CA	F1	Precision	Recall
kNN	0.9579288967593307	0.9041499846011704	0.9116747667529888	0.9196942368805291	0.9037939426336742
Tree	0.9454641748774704	0.9140668309208501	0.9210394130839271	0.92644672796106	0.9156948527860228
SVM	0.5412503410863709	0.620364952263628	0.5711502500543596	0.7481316077287641	0.46188473279221237
Random Forest	0.9774968094776723	0.9234909146904835	0.9300078183880033	0.9313263549925233	0.9286930099736942
Naive Bayes	0.8910668007360968	0.8034724360948567	0.8209412771569474	0.8187758871599686	0.8231181509980728
Logistic Regression	0.8761866653654409	0.8004542654758239	0.8192312252826582	0.8124481212993193	0.826128546710369

Target: satisfied

Test and Score Mon May 10 21, 22:05:07

Settings

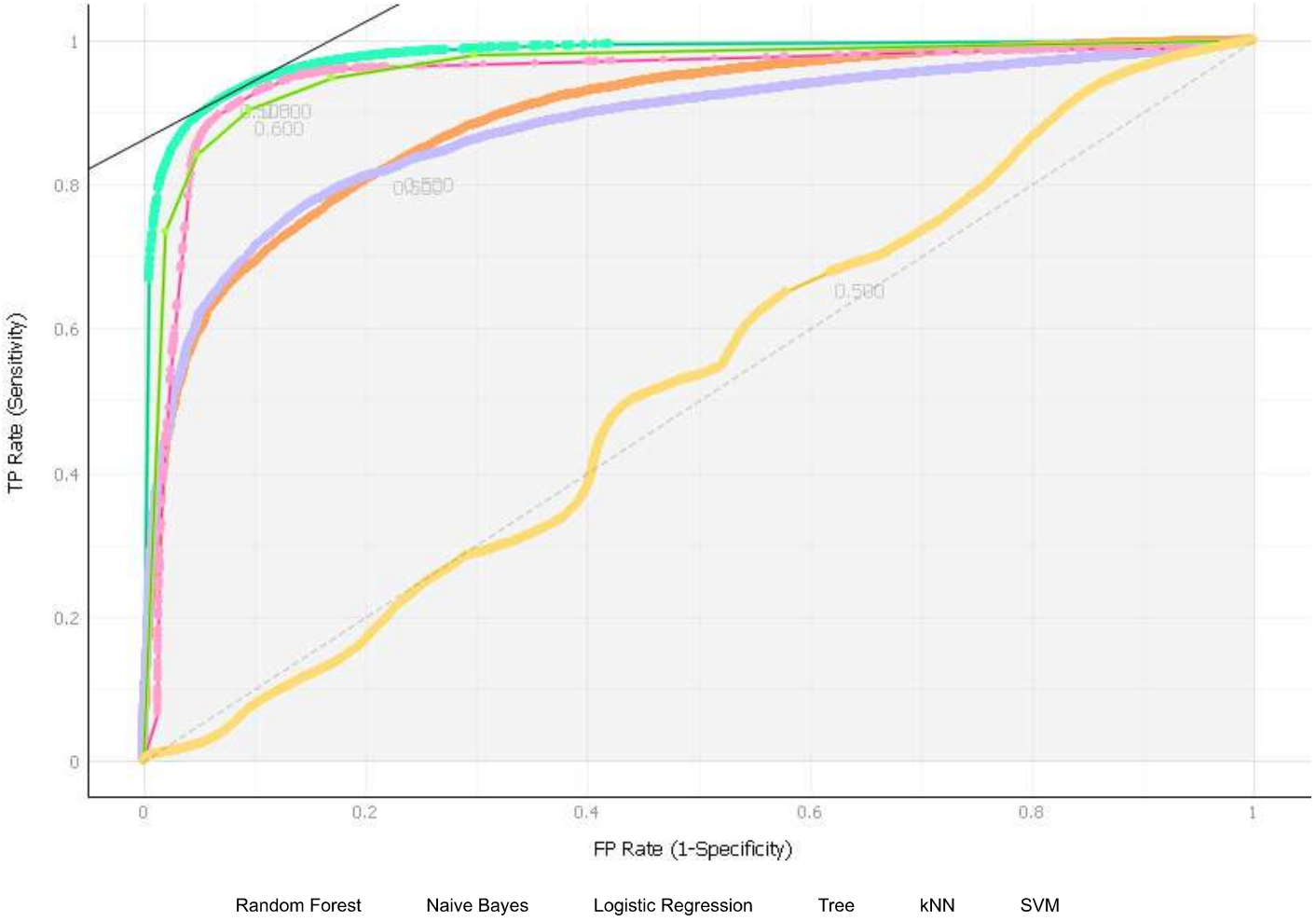
Sampling type: Stratified 10-fold Cross validation
Target class: neutral or dissatisfied

Scores

Model	AUC	CA	F1	Precision	Recall
kNN	0.9579288967593309	0.9041499846011704	0.8952236670454068	0.8860584452367465	0.904580477267702
Tree	0.9454641748774704	0.9140668309208501	0.9057435542306034	0.8994766681203662	0.9120983790587315
SVM	0.5412503410863709	0.620364952263628	0.659446765894257	0.5551562936087078	0.8119844199139353
Random Forest	0.9774964529732959	0.9234909146904835	0.915635851155051	0.9140760077295996	0.9172010273331859
Naive Bayes	0.8910668007360968	0.8034724360948567	0.782226620822633	0.7847507532182963	0.7797186739918017
Logistic Regression	0.8761866653654409	0.8004542654758239	0.7773243175901502	0.7854017640113897	0.7694113244774038

Target: neutral or dissatisfied

Target class: satisfied
Costs: FP = 500, FN = 500
Target probability: 55.0 %



Confusion matrix for Random Forest (showing number of instances)

		Predicted		
		neutral or dissatisfied	**satisfied**	∑
Actual	**neutral or dissatisfied**	53925	4868	**58793**
	satisfied	5069	66018	**71087**
	∑	**58994**	**70886**	**129880**

Confusion Matrix Mon May 10 21, 22:12:48

Confusion matrix for Naive Bayes (showing number of instances)

		Predicted		
		neutral or dissatisfied	**satisfied**	**∑**
Actual	**neutral or dissatisfied**	45842	12951	**58793**
	satisfied	12574	58513	**71087**
	∑	**58416**	**71464**	**129880**

Confusion Matrix Mon May 10 21, 22:13:04

Confusion matrix for Logistic Regression (showing number of instances)

		Predicted		
		neutral or dissatisfied	**satisfied**	**∑**
Actual	**neutral or dissatisfied**	45236	13557	**58793**
	satisfied	12360	58727	**71087**
	∑	**57596**	**72284**	**129880**

Confusion Matrix Mon May 10 21, 22:13:15

Confusion matrix for Tree (showing number of instances)

		Predicted		
		neutral or dissatisfied	**satisfied**	**∑**
Actual	**neutral or dissatisfied**	53625	5168	**58793**
	satisfied	5993	65094	**71087**
	∑	**59618**	**70262**	**129880**

Confusion Matrix Mon May 10 21, 22:13:31

Confusion matrix for kNN (showing number of instances)

		Predicted		
		neutral or dissatisfied	**satisfied**	**∑**
Actual	**neutral or dissatisfied**	53183	5610	**58793**
	satisfied	6839	64248	**71087**
	∑	**60022**	**69858**	**129880**

Confusion Matrix

Mon May 10 21, 22:13:42

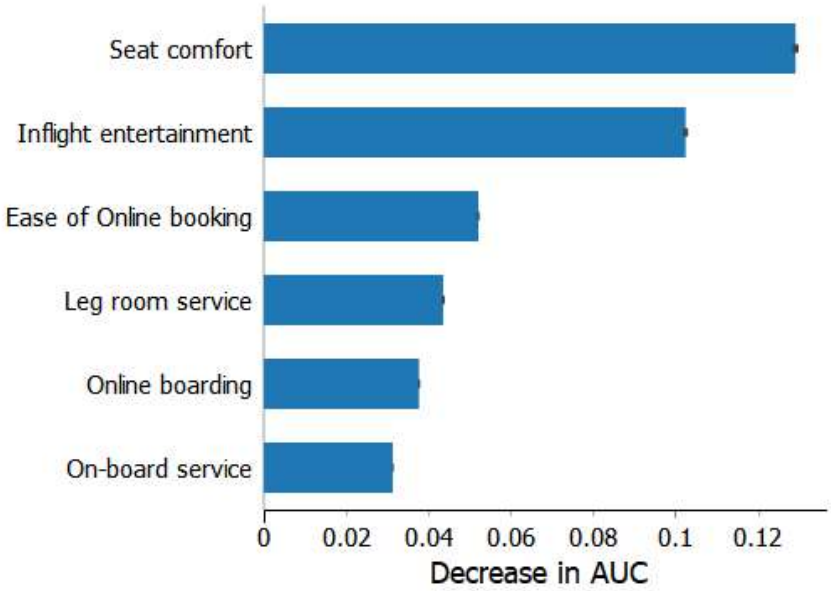
Confusion matrix for SVM (showing number of instances)

		Predicted		
		neutral or dissatisfied	**satisfied**	∑
Actual	**neutral or dissatisfied**	47739	11054	**58793**
	satisfied	38253	32834	**71087**
	∑	**85992**	**43888**	**129880**

Feature Importance

Mon May 10 21, 22:15:49

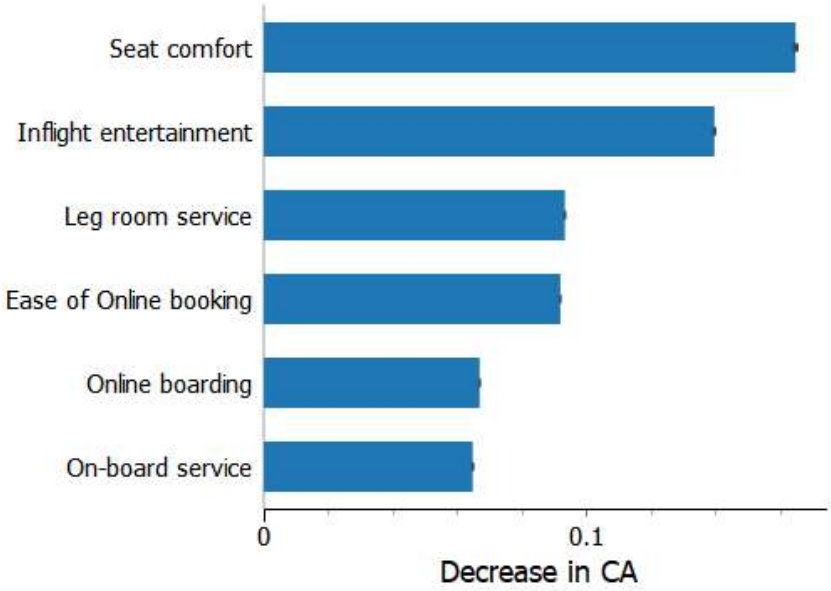
Score: AUC
Permutations: 10



Feature Importance

Mon May 10 21, 22:18:24

Score: CA
Permutations: 10



Explain Model

Mon May 10 21, 22:20:32

Target class: satisfied