PREDICTING SHOPPING SATISFACTION OF AMAZON CONSUMERS: A DATA SCIENCE APPROACH

YOUSSAF MENACER, PHD

MENTOR: TONY PAEK

7.20.23

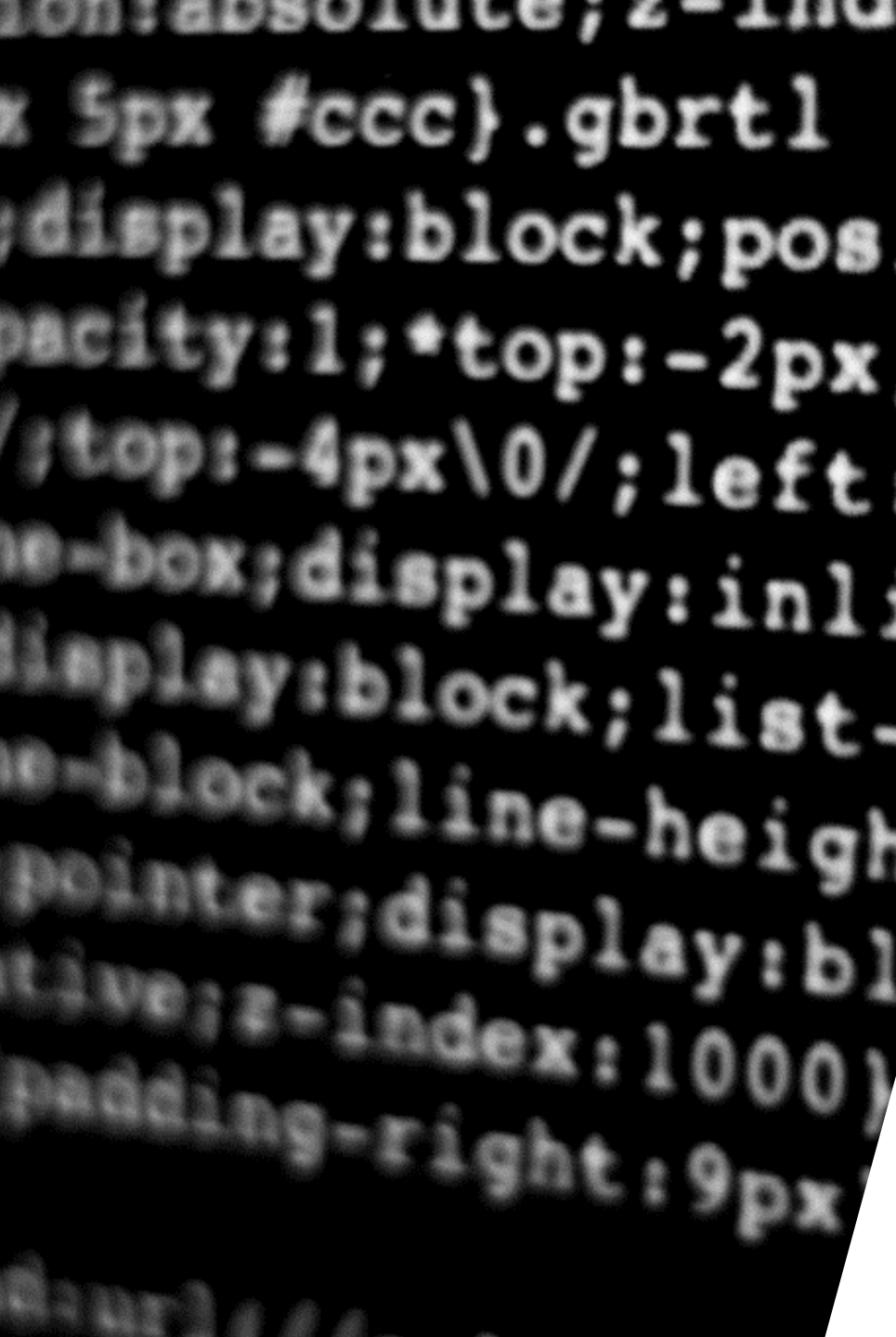
AGENDA

* INTRODUCTION
* Problem Statement
* Data Collection, Exploration, and Preprocessing
* Model Development and Evaluation
* Conclusion:
* Q&A and Discussion:

INTRODUCTION

How important is Customer satisfaction?

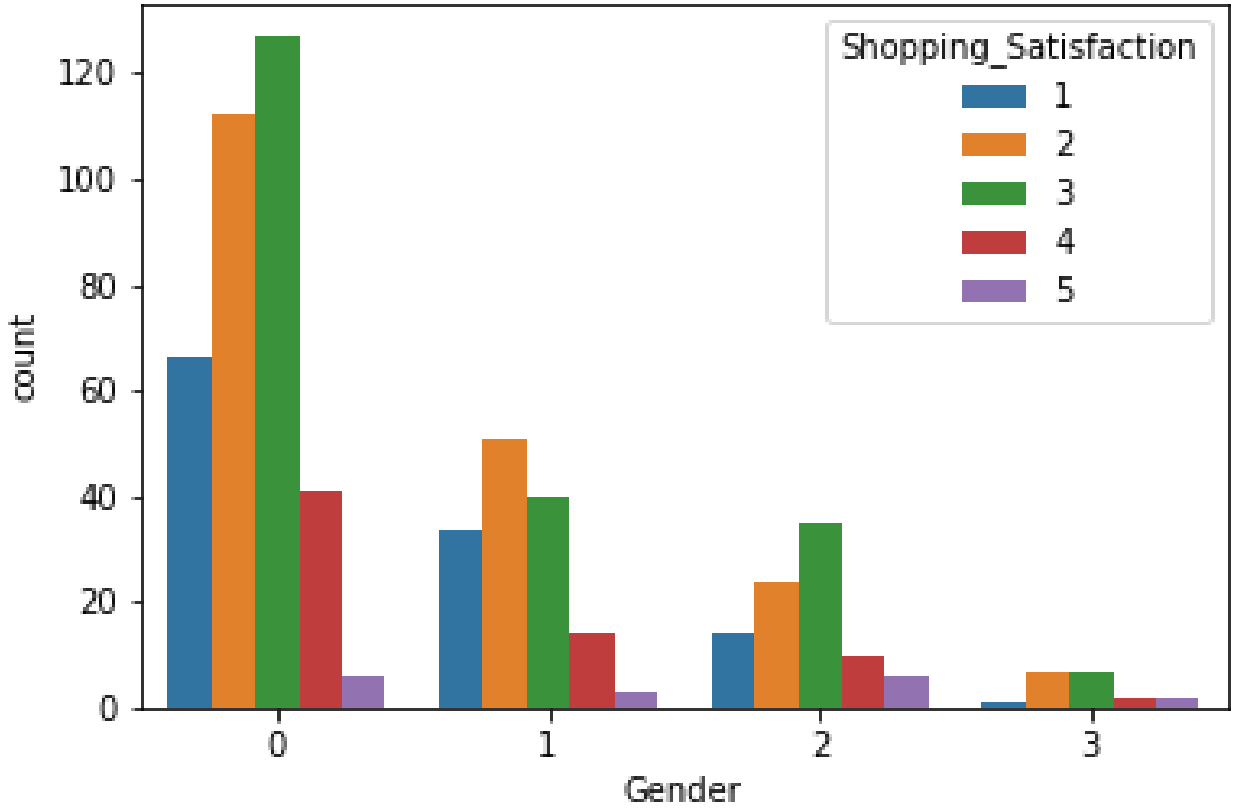
Customer satisfaction is a critical factor for any successful e-commerce platform, and Amazon, being a global leader, is no exception. Understanding and predicting shopping satisfaction among Amazon consumers is of paramount importance for the company's growth and competitive edge. By accurately predicting shopping satisfaction, Amazon can proactively address customer concerns, optimize their shopping experience, and tailor personalized solutions to enhance overall satisfaction levels.



PROBLEM STATEMENT:

PREDICTING SHOPPING SATISFACTION IS CRUCIAL FOR AMAZON'S SUCCESS. IT ENABLES THE COMPANY TO OPTIMIZE THE SHOPPING EXPERIENCE, IMPROVE CUSTOMER RETENTION, AND STRENGTHEN ITS BRAND REPUTATION. BY PRIORITIZING CUSTOMER SATISFACTION, AMAZON DEMONSTRATES ITS COMMITMENT TO DELIVERING EXCEPTIONAL SERVICE AND MEETING CUSTOMER EXPECTATIONS.

Data Collection, Exploration, and Preprocessing:

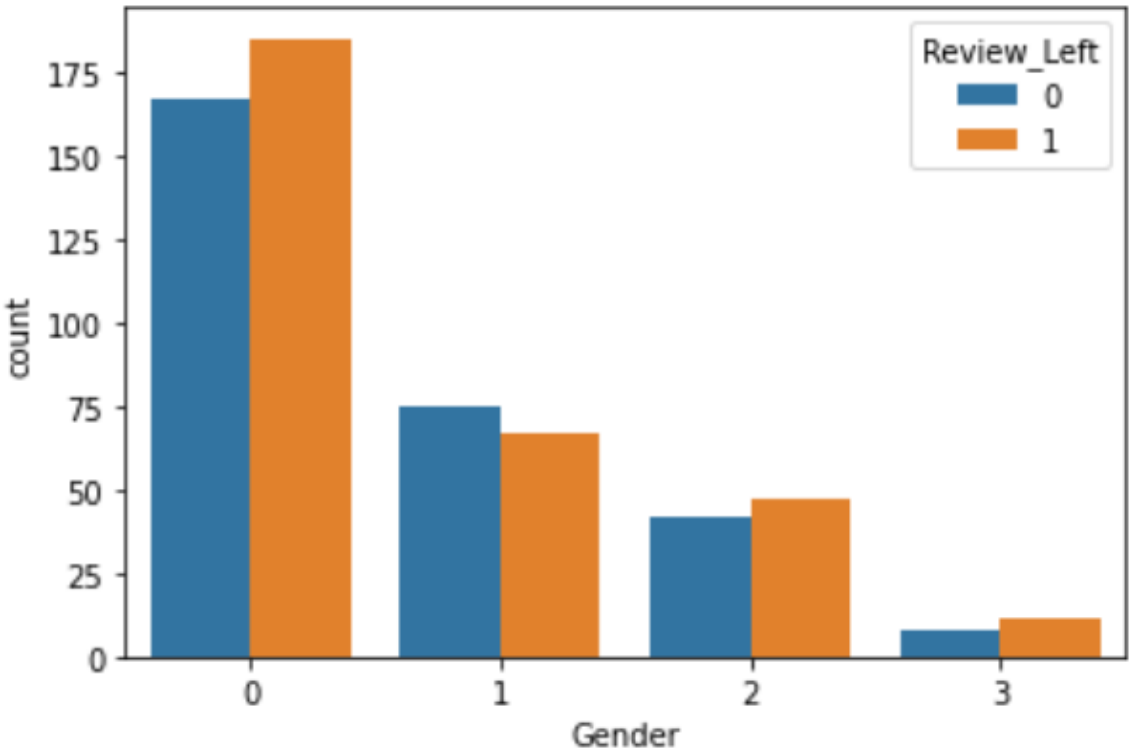


Here for shopping Satisfaction,
1 means “very satisfied”, 2 stands for “satisfied”, 3 designs “average”, 4 is for “Unsatisfied” and 5 represents “Very Unsatisfied”

For the x-axis,

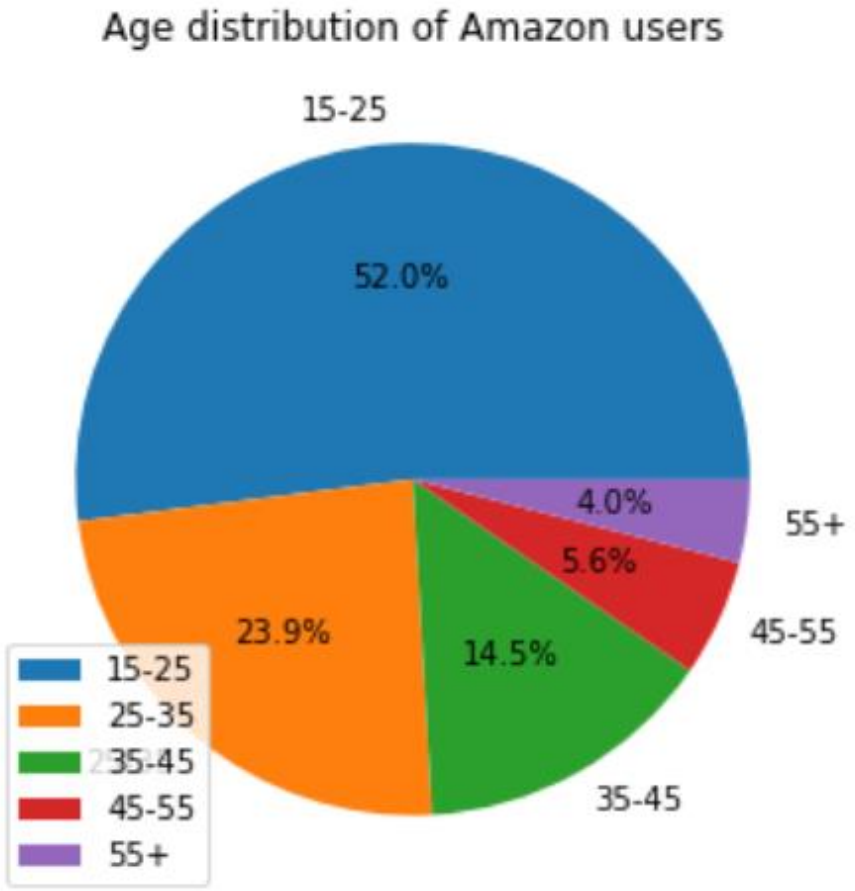
'Female': 0, 'Male': 1,'Prefer not to say': 2, 'Others': 3

DATA COLLECTION, EXPLORATION, AND PREPROCESSING:



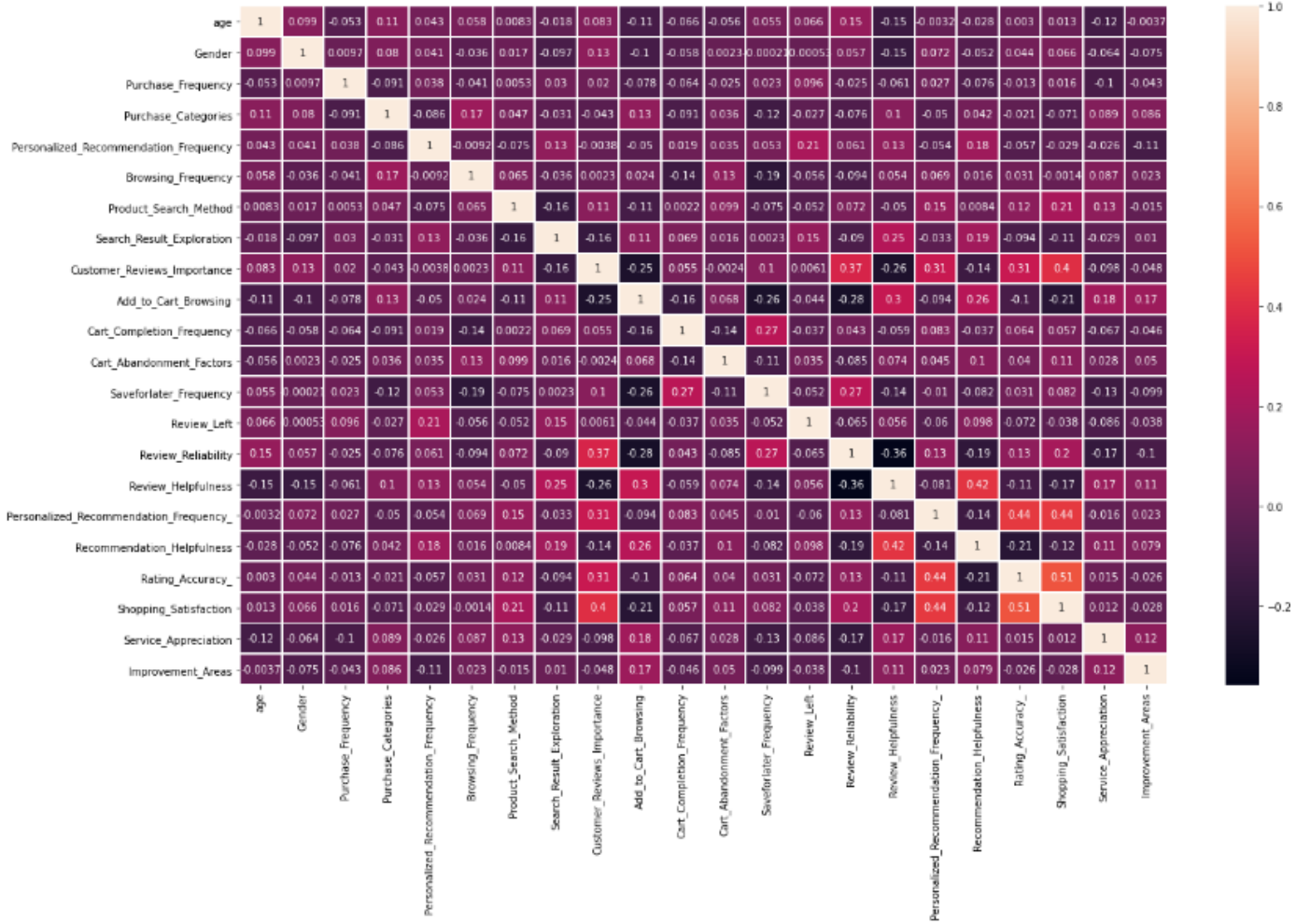
Here, we observe that most of the reviews left at Amazon platform are from female users.

DATA COLLECTION, EXPLORATION, AND PREPROCESSING:



-> Young consumers represent the majority

CHECKING THE CORRELATION BETWEEN FEATURES:



We observe that Shopping_Satisfaction has higher correlation rates comparing to other features.

Model Development and Evaluation

In order to build our model, we should select the best features for classification using ANOVA, focusing on those features that have the most significant impact on the target variable, which is “y=amazon_data['Shopping_Satisfaction’] ”.

1/ Scaling Data: we use here minmax() scaler.

2/ Splitting train/test: we use here test size of 30%.

K NEAREST NEIGHBORS

	precision	recall	f1-score	support
1	0.71	0.97	0.82	62
2	0.65	0.43	0.52	60
3	0.67	0.43	0.52	65
4	0.78	0.90	0.84	71
5	0.86	1.00	0.93	56
accuracy			0.75	314
macro avg	0.73	0.75	0.72	314
weighted avg	0.73	0.75	0.72	314

SVC

	precision	recall	f1-score	support
1	0.91	0.82	0.86	62
2	0.51	0.78	0.62	60
3	0.49	0.38	0.43	65
4	0.98	0.79	0.88	71
5	0.97	1.00	0.98	56
accuracy			0.75	314
macro avg	0.77	0.76	0.75	314
weighted avg	0.77	0.75	0.75	314

GUASSIAN NB

	precision	recall	f1-score	support
1	0.56	0.68	0.61	62
2	0.37	0.35	0.36	60
3	0.44	0.57	0.49	65
4	0.49	0.32	0.39	71
5	0.66	0.59	0.62	56
accuracy			0.50	314
macro avg	0.50	0.50	0.50	314
weighted avg	0.50	0.50	0.49	314

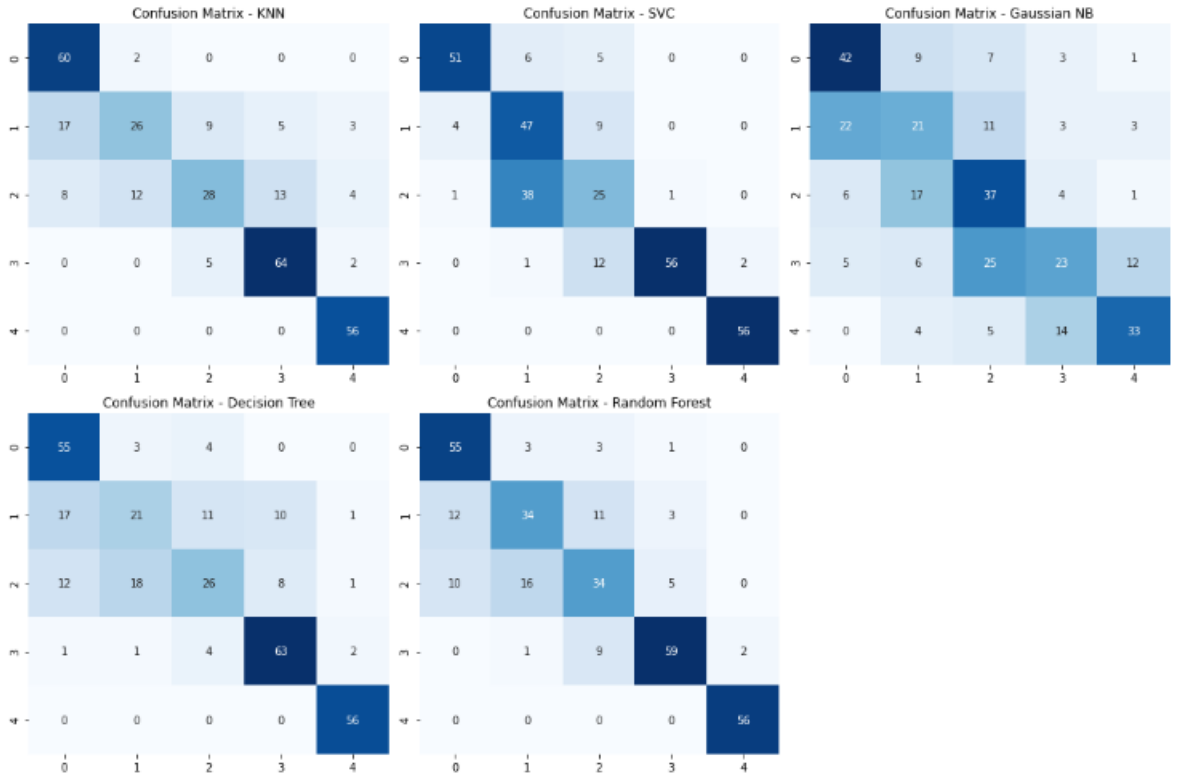
DECISION TREE:

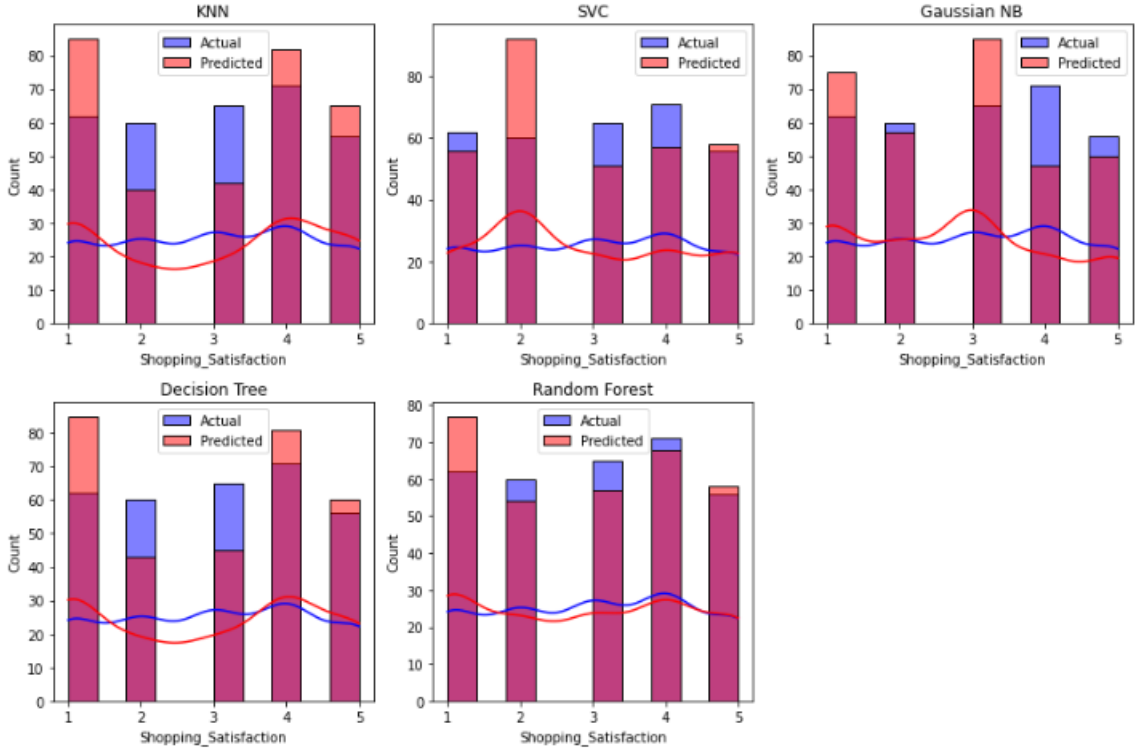
	precision	recall	f1-score	support
1	0.65	0.89	0.75	62
2	0.49	0.35	0.41	60
3	0.58	0.40	0.47	65
4	0.78	0.89	0.83	71
5	0.93	1.00	0.97	56
accuracy			0.70	314
macro avg	0.68	0.70	0.68	314
weighted avg	0.68	0.70	0.68	314

RANDOM FORESTS:

	precision	recall	f1-score	support
1	0.71	0.89	0.79	62
2	0.63	0.57	0.60	60
3	0.60	0.52	0.56	65
4	0.87	0.83	0.85	71
5	0.97	1.00	0.98	56
accuracy			0.76	314
macro avg	0.75	0.76	0.76	314
weighted avg	0.75	0.76	0.75	314

COMPARING THE MODELS





Confusion matrix for each model prediction and the ground truth

Histogram plots for each prediction/ground truth

SUMMARY

-> For this classification problem, random forest model gives the best result to predict the shopping satisfaction for Amazons consumers.

-> For future work on this project, we can do

. Incorporating sentiment analysis, performing longitudinal analysis, and deploying and monitoring the model.

. These avenues provide opportunities to further optimize the model's performance, explore alternative methods, gain insights into customer behavior, and ensure the model's practical implementation.

THANK YOU

Youssaf Menacer

Youssaf.menacer@gmail.com