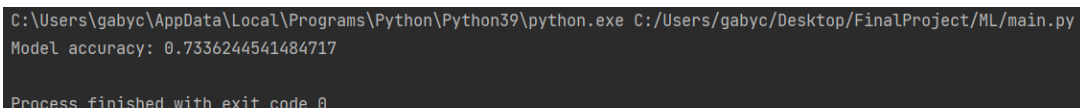# Machine Learning Report

The main objective of this project is to predict a user's choice between two different rating distributions, using machine learning. In this project we used the machine learning algorithm - Decision Tree.

The Decision Tree algorithm is a supervised learning technique that can be used for classification. A tree can be "learned" by splitting the source set into subsets based on an attribute value test. This process is repeated on each derived subset in a recursive manner called recursive partitioning.

The attributes in our decision tree are set to be the rating stars percentage. Meaning the 1st attribute is the percentage of people who rated it 1 star on the left distribution. The 2nd attribute is the percentage of people who rated it 2 stars on the left distribution and so forth for all the ratings in both distributions (10 total attributes, left and right). The labels in our model are set to be the decisions made by the people who filled the survey. The leaves are the choices - the right or left distribution.

The decision tree is a greedy algorithm that proceeds its course using the path with the most information gain. Which means it chooses the one with the lowest entropy. A decision tree algorithm does not require hyper parameters, because of that it is well adjusted for exploratory knowledge discovery. In our project we collected the answers of 73 different persons. Which means 73 * 8 = 584 instances (two of the tests were tipped to the extremes of the graphs to check for the user's credibility). We split the data so that 20 percent of it was used to train the model, the rest of the data was used to test it. The results were around 70% consistency with the "correct answer", the "correct answer" being the distribution with the higher rating average value. In the image below you can see the accuracy of the model.

```
C:\Users\gabyc\AppData\Local\Programs\Python\Python39\python.exe C:/Users/gabyc/Desktop/FinalProject/ML/main.py
Model accuracy: 0.7336244541484717

Process finished with exit code 0
```

Future relevant experiments to better understand the process of decision making by a human –

First of all, different machine learning algorithms could grant new perspectives on certain questions. Linear regression could render interesting results. SVM is useful for classification type problems as well as Perceptron (although the perceptron algorithm won't always stop). When there are enough inputs, it is possible to implement a neural network to make a more accurate decision.

Secondly, adding a products picture could maybe change the perception of a user while the survey is being filled. A different model could learn based on the decisions made by people who were given the image, and another model would learn based on people without said information. Afterwards we could analyze the difference in

the choices and see to what extent the pictures make an impact on a user's pick. Similarly, we could do the same by adding the price of the products and analyzing the differences.