

# Data mining: Assignment 2

Arno Deceuninck

April 27, 2022

## 1 Historical performance

In the historical dataset, 36382 contracts were accepted (36%) and 63618 contracts were rejected (64%). Of the accepted contracts, 29649 (81%) were high value contracts and 6733 (19%) were low value contracts. This makes a total income of \$2,090,800 (given a high value contracts).

You can see the contracts high/low and accepted/rejected contracts in the table below.

	Low Value	High Value	
Accepted	45259	18359	63618
Rejected	6733	29649	36382
	51992	48008	100000

## 2 Percentage high contracts

The company has taken 62% of the high value contracts (29,649 of the 48,008). This is not higher because you need to find a good balance to lower your chances of getting a low value contract (false positive), but still getting as much high value contracts as possible (true positives).

## 3 Monopoly Profits

If we just base ourselves on the estimated percent of contracts the company takes (30% high contracts, 7% low contracts), you would expect an income of \$418,160. This was just a rough guess. To make a better guess and see which contracts were actually taken, I also trained a model with sklearn. For this, I used different models and GridSearchCV to find the optimal hyperparameters for each of the options and select the model that gave the best performance given those hyperparameters. This was a RandomForest. I fitted both a model for highValue prediction and a model for the prediction column prediction. We can then run this model on the current data, to get a highValue prediction and a prediction whether a contract would be taken and based on those parameters, I calculated an expected profit of \$1,014,750 for the current company.

## 4 Maximize Profit

In order to maximize your own profit, you want to take as much highvalue contracts as possible. Some of those contracts will be shared with the other company, but will still give us a positive income. So we can just use the model of above that predicts all highValue accommodations. You can find the selection in selections.csv.

## 5 Expected Earnings

We'll take all contracts that we predict as highValue contracts, regardless whether the other company will also take it. If the other company also takes it, our income gets split, but we still have a positive income. This gives us the expected earnings of \$1,239,150