Salary & Classification Modeling

Rahkeem D. George

Overview

Dataset

Methodology/Procedures

Classification Modeling

Results

Dataset

Website for dataset: Adult Dataset

Colum to predict: **Class**

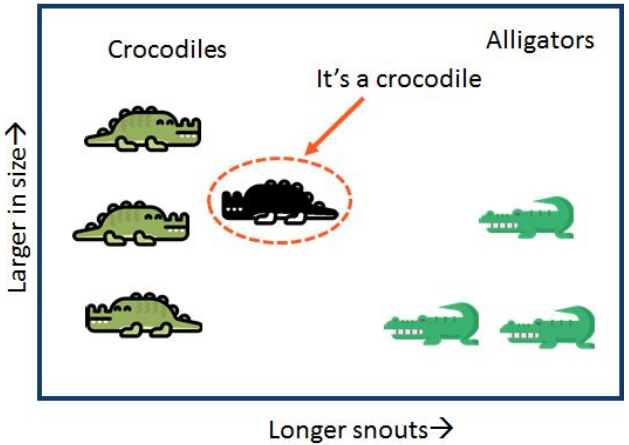
Columns to use for prediction:

Age	workclass	fnlwgt	education	"education-num
marital-status	occupation	race	sex	"capital-gain
"capital-loss	hours-per-week	native-country	relationship	

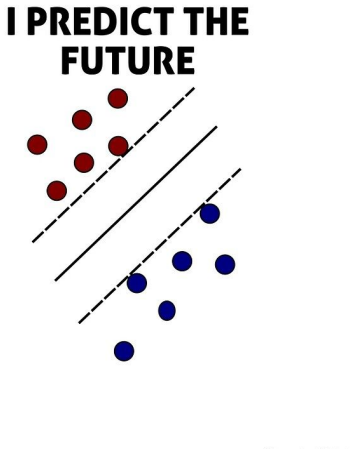
Methodology/Procedures

1. Clean our Data
2. Deal with Categorical values
3. Split data for train & testing phase
4. Create & Train our Models
5. Measure Performance

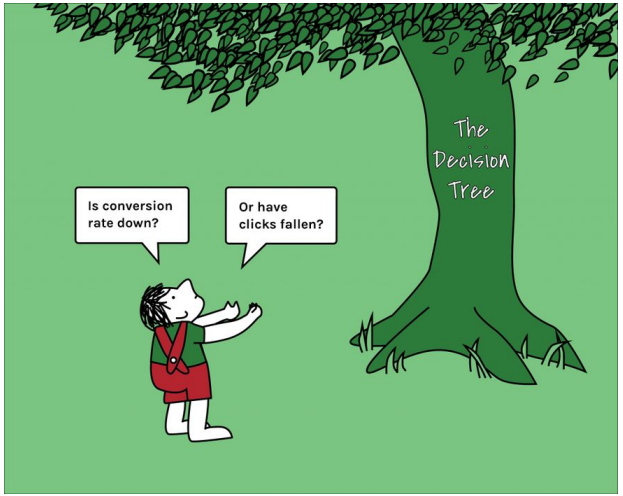
Classification Modeling Techniques



K-Nearest Neighbors



Support Vector Machines



Decision Tree

Results

Decision Trees Wins!!!

Future Works

- Data for more recent periods
- API to interface with the website
- Use of Random Forests in place of Decision Trees

References

1. Alligator vs Crocodile KNN illustration
2. Decision Tree illustration
3. Done list
4. Support Vector Machine
5. Ron Kohavi, "Scaling Up the Accuracy of Naive-Bayes Classifiers: a Decision-Tree Hybrid", Proceedings of the Second International Conference on Knowledge Discovery and Data Mining, 1996
6. Adult Dataset

Thank You