Machine Learning Design Challenge 1

Challenge: Graduate Admissions

This dataset contains parameters which are considered in MA admission processes of Indian MA programs. Tulsi is interested in applying and wants to find what her chances are for getting in with her current scores.

Predictable: Chance of Admit

Sample data:

Serial No., GRE Score, TOEFL Score, University Rating, SOP, LOR , CGPA, Research, Chance of Admit

1,337,118,4,4.5,4.5,9.65,1,0.92
2,324,107,4,4,4.5,8.87,1,0.76
3,316,104,3,3,3.5,8,1,0.72
4,322,110,3,3.5,2.5,8.67,1,0.8
5,314,103,2,2,3,8.21,0,0.65

Data set size:

400

Team members:

Implementation confusion matrix:

Solution alternatives:

Possible pitfalls:

Notes and thoughts:

Machine Learning Design Challenge 2

Challenge: Black Friday

This dataset contains information about customer behaviour at a retail store. The owner of the store is curious to know how much different customers are likely to purchase.

The store owner might also be interested in what kind of customers go for a certain product category.

Sample data:

User_ID, Product_ID, Gender, Age, Occupation, City_Category, Stay_In_Current_City_Years, Marital_Status, Product_Category_1, Product_Category_2, Product_Category_3, Purchase

1000001,P00069042,F,0-17,10,A,2,0,3,,,8370
1000001,P00248942,F,0-17,10,A,2,0,1,6,14,15200
1000002,P00285442,M,55+,16,C,4+,0,8,,,7969

Data set size:

537577

Team members:

Implementation confusion matrix:

Solution alternatives:

Possible pitfalls:

Notes and thoughts:

Machine Learning Design Challenge 3

Challenge: Mall Customers

This dataset describes a supermarket with memberships. Though these membership registrations you have acquired information about the customers and their spending habits. You want to understand the customers better in order to plan and create strategies for development, including creating a more personalized shopping experience for all kinds of customers.

What kinds of customers are there?

Sample data:

CustomerID, Gender, Age, Annual Income (k$), Spending Score (1-100)

1,Male,19,15,39
2,Male,21,15,81
3,Female,20,16,6
4,Female,23,16,77
5,Female,31,17,40

Data set size:

200

Team members:

Implementation confusion matrix:

Solution alternatives:

Possible pitfalls:

Notes and thoughts:

Machine Learning Design Challenge 4

Challenge: Diabetes Data

This dataset is originally from the National Institute of Diabetes and Digestive and Kidney Diseases. The objective of the dataset is to diagnostically predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset.

Predictable: Outcome
(1 = Diabetes, 0 = Not diabetes)

Sample data:

Pregnancies, Glucose, Blood Pressure, Skin Thickness, Insulin, BMI, Diabetes Pedigree Function, Age, Outcome

6,148,72,35,0,33.6,0.627,50,1
1,85,66,29,0,26.6,0.351,31,0
8,183,64,0,0,23.3,0.672,32,1
1,89,66,23,94,28.1,0.167,21,0
0,137,40,35,168,43.1,2.288,33,1

Data set size:

768

Team members:

Implementation confusion matrix:

Solution alternatives:

Possible pitfalls:

Notes and thoughts: