Addressing Critical Health Conditions through Data Science

According to the Centers for Disease Control and Prevention (CDC), every year, about 805,000 Americans have a heart attack. Of these, 605,000 are a first heart attack and 200,000 happen to people who have already had a heart attack.

Reference:

Centers for Disease Control and Prevention (CDC), "Heart Disease Facts."

Available: https://www.cdc.gov/heartdisease/facts.htm

Data Sets

Heart Attack

Records: 8763

Features: 26

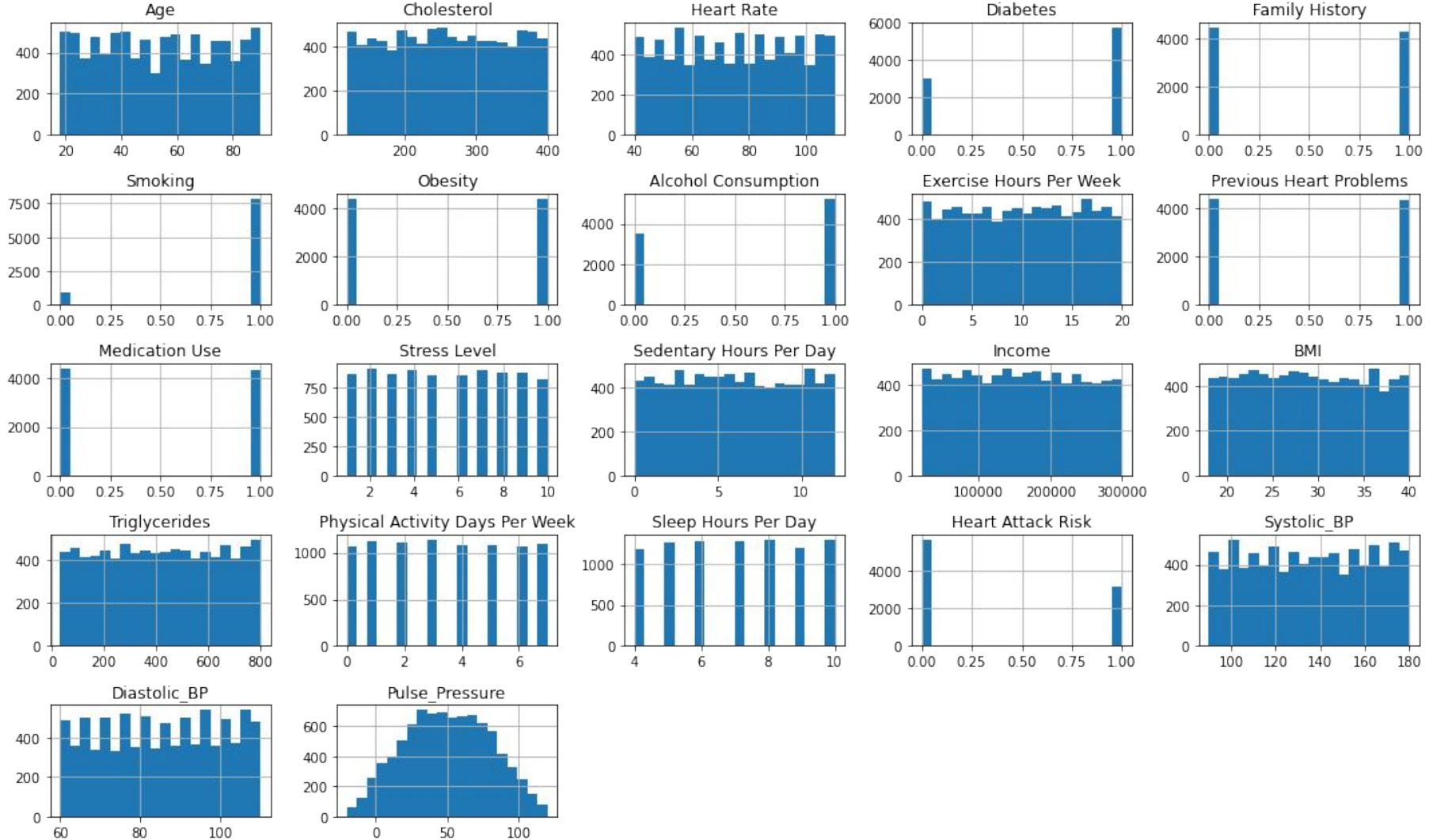
Age
Sex
Cholesterol
Blood Pressure
Heart Rate
Diabetes
Family History
Smoking
Obesity
Alcohol Consumption
Exercise Hours Per Week
Diet
Previous Heart Problems
Medication Use
Stress Level
Sedentary Hours Per Day
Income
BMI
Triglycerides
Physical Activity Days Per Week
Sleep Hours Per Day
Country
Continent
Hemisphere
Heart Attack Risk

Stroke

Records: 5110

Features: 11

Gender
Age
Hypertension
Ever Married
Work Type
Residence Type
Glucose Level
BMI
Smoking Status
History of stroke
Heart Disease


Age
Cholesterol
Heart Rate
Diabetes
Family History
Smoking
Obesity
Alcohol Consumption
Exercise Hours Per Week
Previous Heart Problems
Medication Use
Stress Level
Sedentary Hours Per Day
Income
BMI
Triglycerides
Physical Activity Days Per Week
Sleep Hours Per Day
Heart Attack Risk
Systolic_BP
Diastolic_BP
Pulse_Pressure

High cardinality

Blood Pressure	8763
Exercise Hours Per Week	8763
Sedentary Hours Per Day	8763
Income	8615
BMI	8763
Triglycerides	771
Cholesterol	281

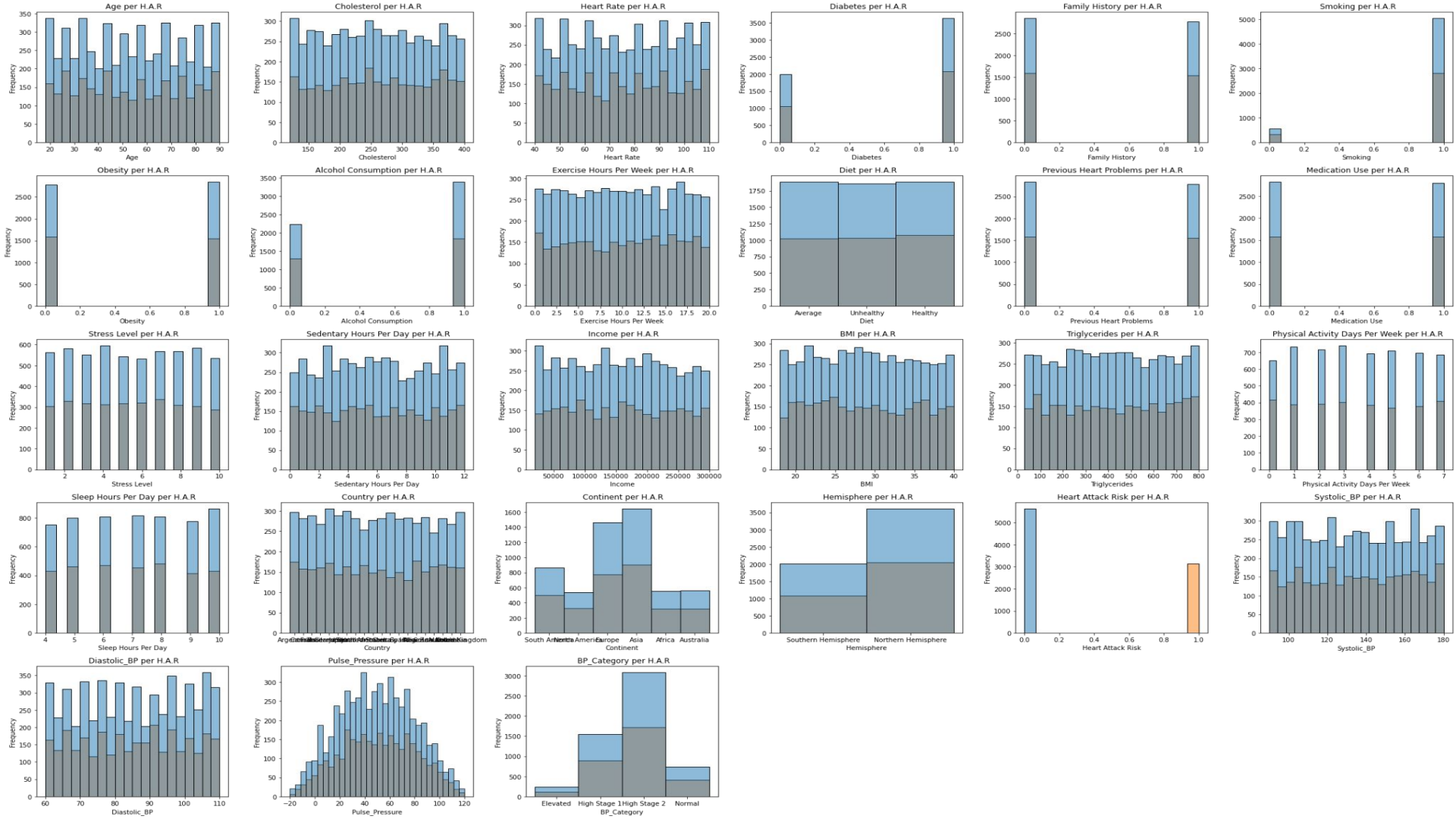
create categories based on medical standards

Blood Pressure

158/88

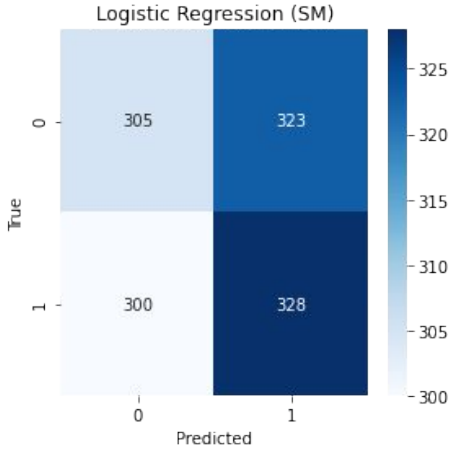
1. Systolic blood pressure (SBP)
2. Diastolic blood pressure (DBP)

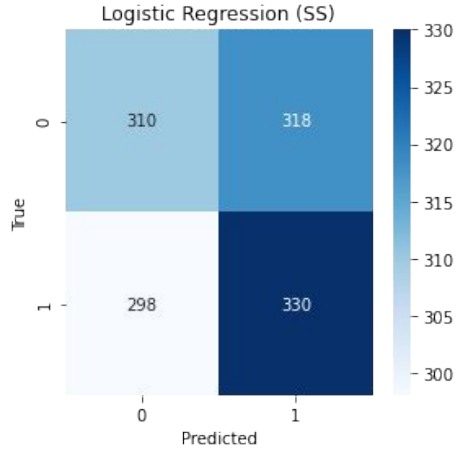
3.Pulse pressure = (SBP) - (DBP)

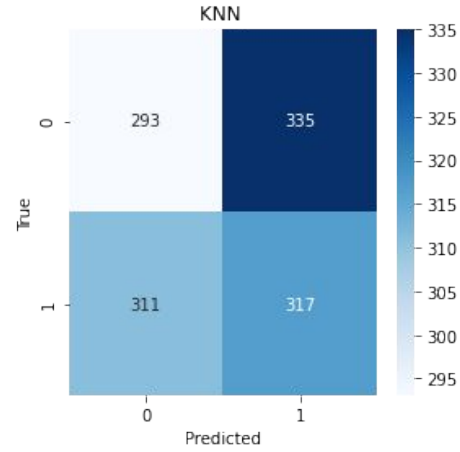

Age per H.A.R
Cholesterol per H.A.R
Heart Rate per H.A.R
Diabetes per H.A.R
Family History per H.A.R
Smoking per H.A.R
Obesity per H.A.R
Alcohol Consumption per H.A.R
Exercise Hours Per Week per H.A.R
Diet per H.A.R
Previous Heart Problems per H.A.R
Medication Use per H.A.R
Stress Level per H.A.R
Sedentary Hours Per Day per H.A.R
Income per H.A.R
BMI per H.A.R
Triglycerides per H.A.R
Physical Activity Days Per Week per H.A.R
Sleep Hours Per Day per H.A.R
Country per H.A.R
Continent per H.A.R
Hemisphere per H.A.R
Heart Attack Risk per H.A.R
Systolic_BP per H.A.R
Diastolic_BP per H.A.R
Pulse_Pressure per H.A.R
BP_Category per H.A.R
Frequency
Age
Cholesterol
Heart Rate
Diabetes
Family History
Smoking
Obesity
Alcohol Consumption
Exercise Hours Per Week
Diet
Average
Unhealthy
Healthy
Previous Heart Problems
Medication Use
Stress Level
Sedentary Hours Per Day
Income
BMI
Triglycerides
Physical Activity Days Per Week
Sleep Hours Per Day
Country
Continent
South America
North America
Europe
Asia
Africa
Australia
Hemisphere
Southern Hemisphere
Northern Hemisphere
Heart Attack Risk
Systolic_BP
Diastolic_BP
Pulse_Pressure
BP_Category
Elevated
High Stage 1
High Stage 2
Normal

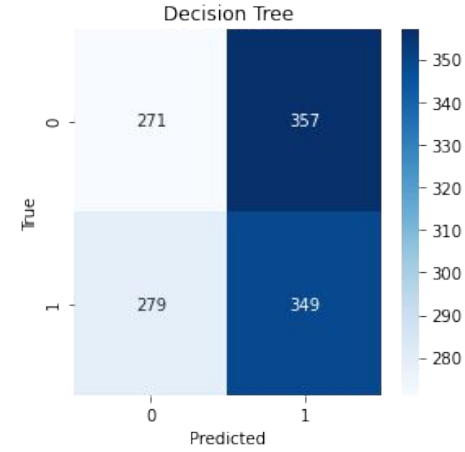
Uniform distribution across the various categories

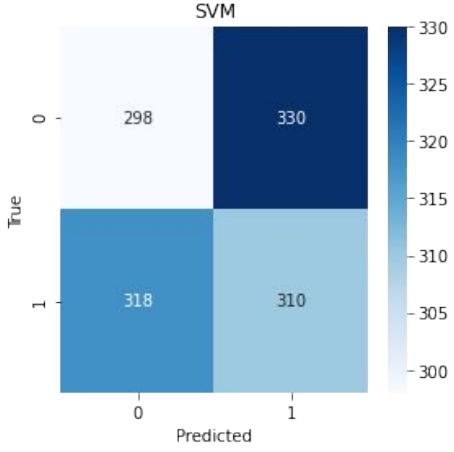
- Balanced Category Representation
- Enhanced Model Generalizability
- Uniform Contribution to Predictive Power
- Simplified Feature Engineering
- Applicability of Standard Evaluation Metrics
- Opportunities for Further Exploration

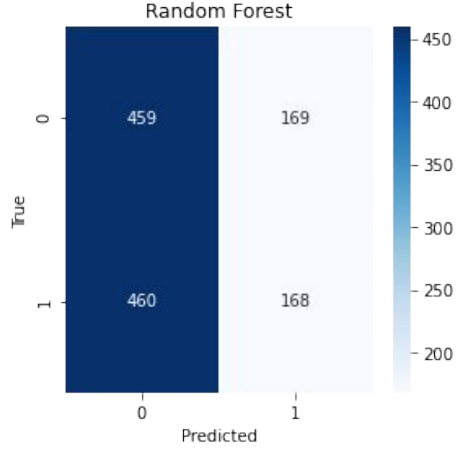

Logistic Regression (SM)
305
323
300
328
True
Predicted
0
1
325
320
315
310
305
300

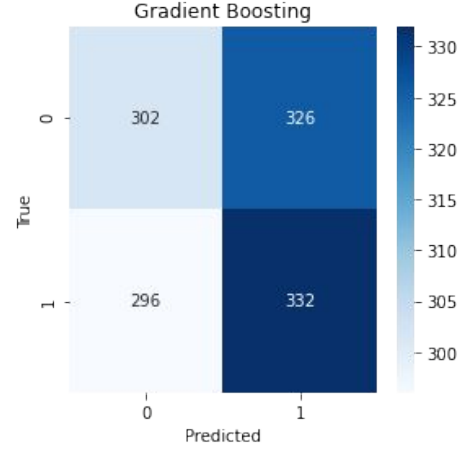

Logistic Regression (SS)
310
318
298
330
True
Predicted
0
1
330
325
320
315
310
305
300

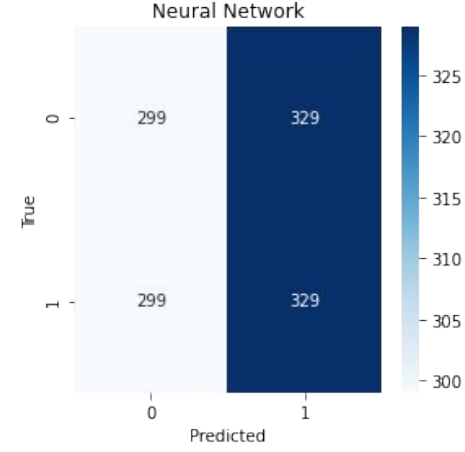

KNN
293
335
311
317
True
Predicted
0
1
335
330
325
320
315
310
305
300
295

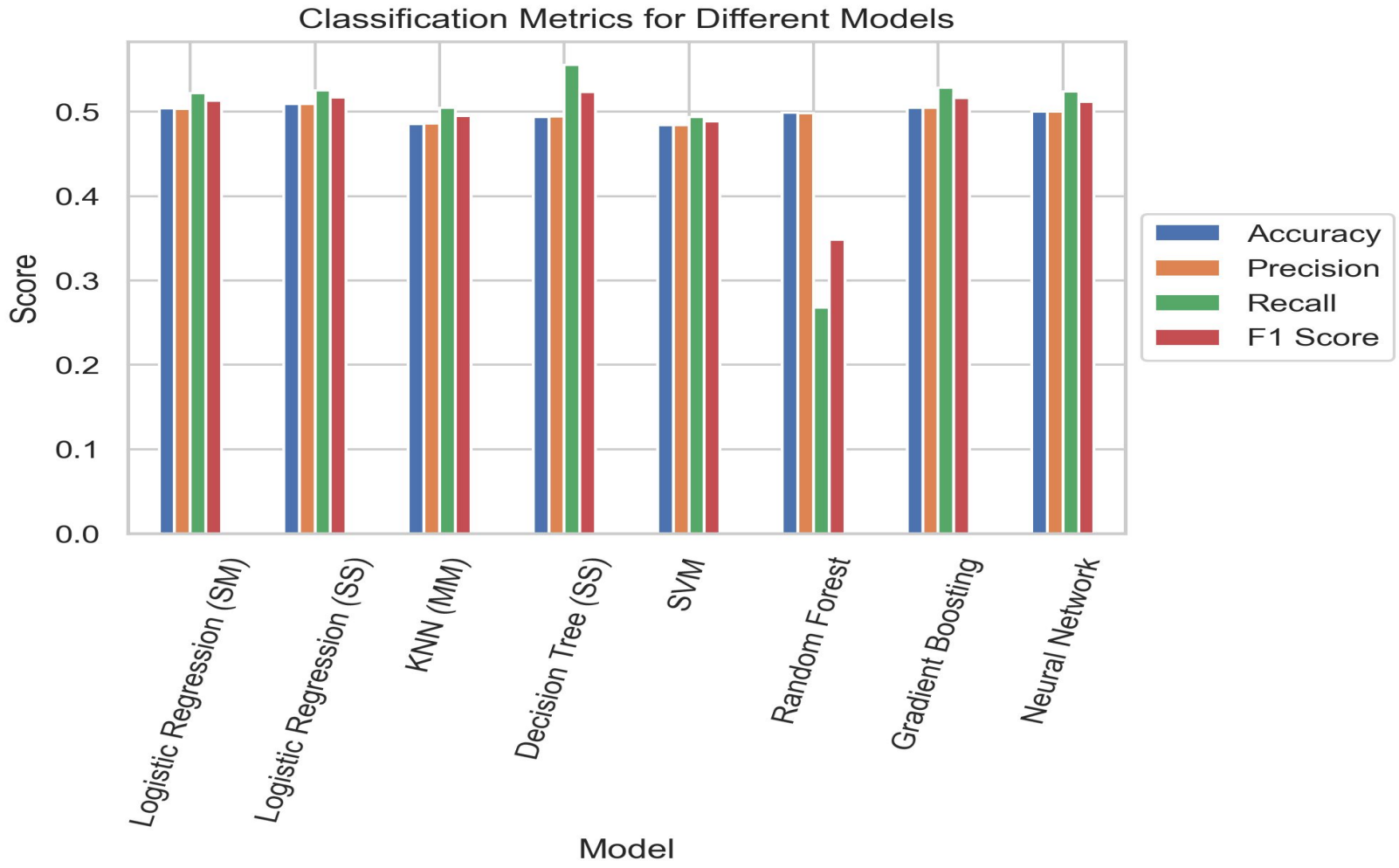

Decision Tree
271
357
279
349
True
Predicted
0
1
350
340
330
320
310
300
290
280

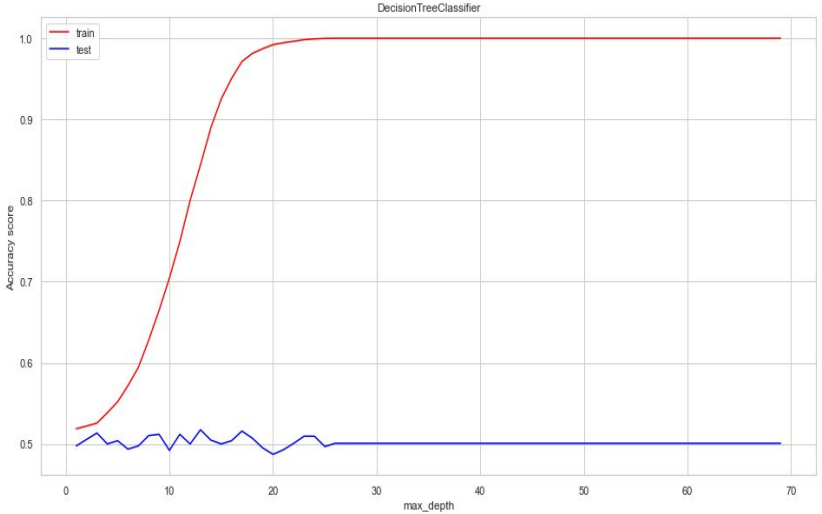

SVM
298
330
318
310
True
Predicted
0
1
330
325
320
315
310
305
300

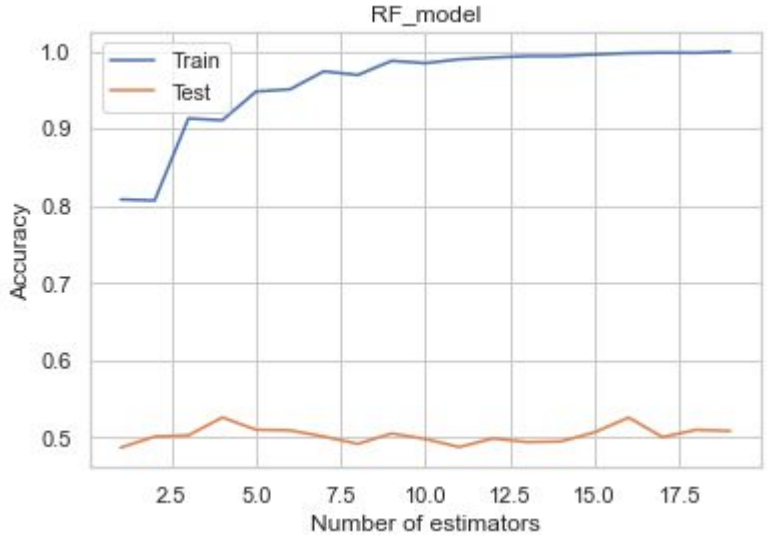

Random Forest
459
169
460
168
True
Predicted
0
1
450
400
350
300
250
200

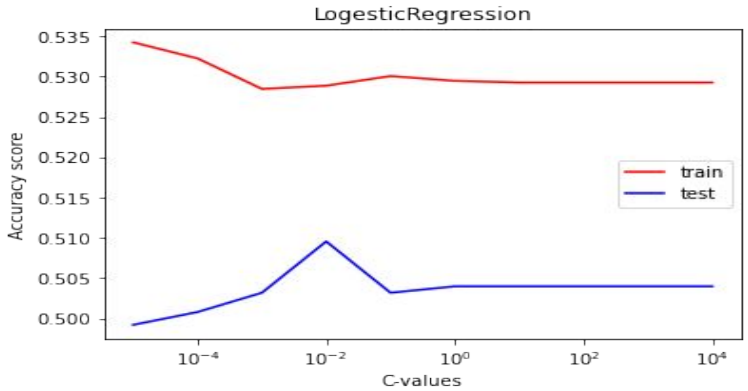

Gradient Boosting
302
326
296
332
True
Predicted
0
1
330
325
320
315
310
305
300

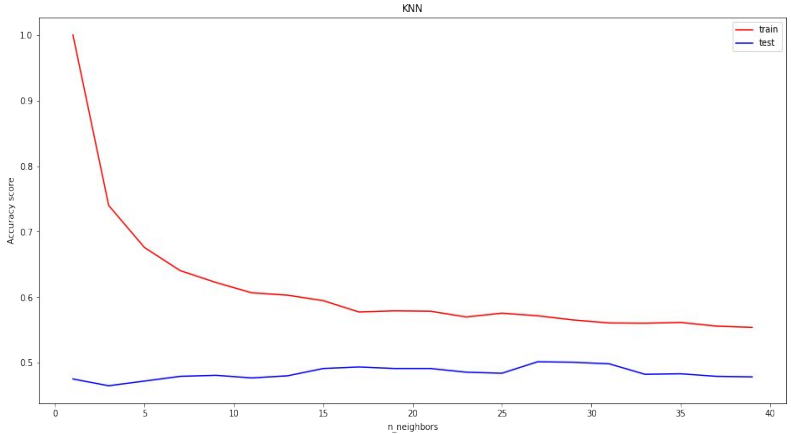

Neural Network
299
329
299
329
True
Predicted
0
1
325
320
315
310
305
300


Classification Metrics for Different Models
Score
0.0
0.1
0.2
0.3
0.4
0.5
Accuracy
Precision
Recall
F1 Score
Logistic Regression (SM)
Logistic Regression (SS)
KNN (MM)
Decision Tree (SS)
SVM
Random Forest
Gradient Boosting
Neural Network
Model


DecisionTreeClassifier
train
test
Accuracy score
max_depth


RF_model
Train
Test
Accuracy
Number of estimators


LogesticRegression
train
test
Accuracy score
C-values


KNN
train
test
Accuracy score
n_neighbors

Insights

🌐 **Balanced Accuracy Across Models:** All models, from Logistic Regression to Neural Networks, consistently show around 50% accuracy, indicating a balanced but challenging dataset.

🎯 **Precision and Recall Trade-Off:** Each model demonstrates a similar precision-recall balance, suggesting equal strengths and weaknesses in identifying true positives and negatives.

⚖️ **Moderate F1 Scores:** The F1 scores across models are moderate, reflecting the ongoing challenge of achieving a perfect balance between precision and recall.

🚀 **Model Selection Tip:** Choose Logistic Regression or Decision Trees for higher sensitivity, or opt for SVM or Random Forest for better precision.

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
const	-0.0000	0.0283	-0.0004	0.9997	-0.0555	0.0555
x1	0.0370	0.0342	1.0827	0.2789	-0.0300	0.1041
x2	0.0436	0.0284	1.5351	0.1248	-0.0121	0.0992
x3	0.0015	0.0284	0.0519	0.9586	-0.0541	0.0570
x4	-0.0494	0.0373	-1.3251	0.1851	-0.1225	0.0237
x5	-0.0217	0.0284	-0.7649	0.4444	-0.0773	0.0339
x6	0.0052	0.0284	0.1839	0.8541	-0.0504	0.0609
x7	-0.0076	0.0284	-0.2684	0.7884	-0.0632	0.0480
x8	0.0139	0.0284	0.4900	0.6241	-0.0417	0.0695
x9	-0.0091	0.0284	-0.3196	0.7493	-0.0647	0.0465
x10	-0.0238	0.0284	-0.8391	0.4014	-0.0794	0.0318
x11	0.0531	0.0393	1.3534	0.1759	-0.0238	0.1301
x12	0.0730	0.0288	2.5334	0.0113	0.0165	0.1295
x13	-0.0102	0.0288	-0.3536	0.7236	-0.0665	0.0462
x14	0.0444	0.0284	1.5612	0.1185	-0.0113	0.1001

	Coef.	Std.Err.	z	P>|z|	[0.025	0.975]
const	-0.0000	0.0283	-0.0004	0.9997	-0.0555	0.0555
x15	-0.0146	0.0284	-0.5130	0.6079	-0.0702	0.0411
x16	-0.0173	0.0284	-0.6097	0.5421	-0.0731	0.0384
x17	0.0480	0.0284	1.6901	0.0910	-0.0077	0.1036
x18	0.0086	0.0283	0.3051	0.7603	-0.0469	0.0642
x19	0.0598	0.0284	2.1065	0.0352	0.0042	0.1155
x20	0.0071	0.0319	0.2215	0.8247	-0.0554	0.0695
x21	-0.0216	0.0304	-0.7097	0.4779	-0.0812	0.0380
x22	-0.0101	0.0405	-0.2496	0.8029	-0.0895	0.0692
x23	-0.0695	0.0611	-1.1371	0.2555	-0.1894	0.0503
x24	-0.0308	0.0508	-0.6066	0.5441	-0.1303	0.0687
x25	0.0007	0.0453	0.0152	0.9879	-0.0881	0.0895
x26	-0.0128	0.0428	-0.3001	0.7641	-0.0967	0.0710
x27	-0.0224	0.0328	-0.6823	0.4951	-0.0867	0.0419
x28	0.0094	0.0328	0.2875	0.7737	-0.0549	0.0737