Detecting Potability By Water Quality

ADS - 503 Team 2
Applied Predictive Modeling

The Team

Jimmy Nguyen

MADS Student

Sarah Alqaysi

MADS Student

Sai Thiha

MADS Student

The Utopia

Top 3 problems our project is addressing are:

Problem 1

One of the most basic human rights in life is to have access to a drinkable water that adheres to all safety criteria and regulations.

Problem 2

Water potability is essential to human survival and should be prioritized for detecting safe drinking water.

Problem 3

National and local authorities as well as private companies, are all obliged to regularly monitoring and enforcing the right procedures when it comes to water safety regulations.

Main Water Quality Metrics

Quality 1

The pH value evaluates the acid-base range of water.

Quality 2

Hardness is an observed metric of how much hardness is in raw water when meeting calcium and magnesium salts

Quality 3

Total dissolved solids produce an unpleasant taste and diluted color in water.

Quality 4

The amount of chlorine and chloramine in public water systems determines the safety for potable water.

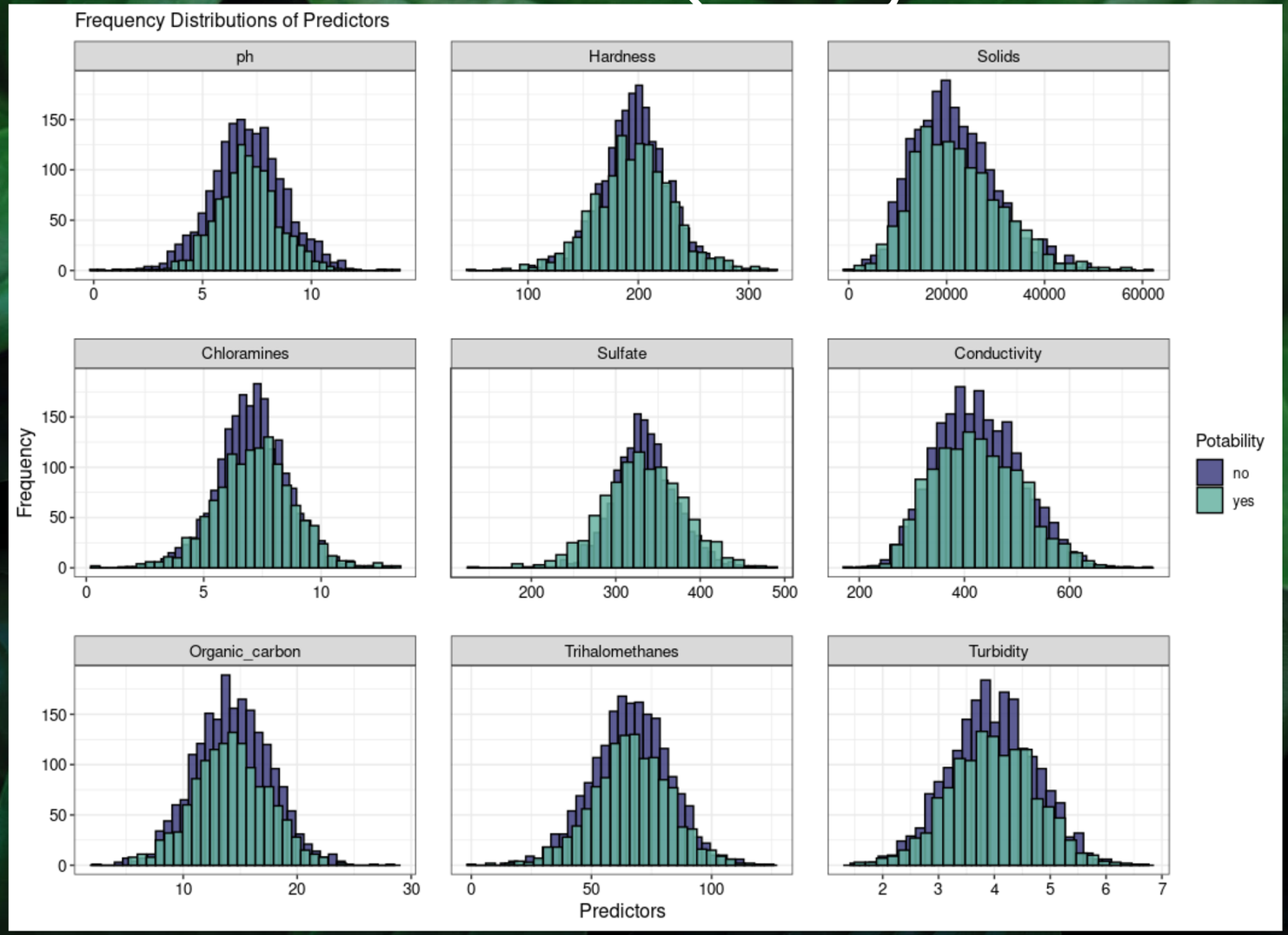
Quality 5

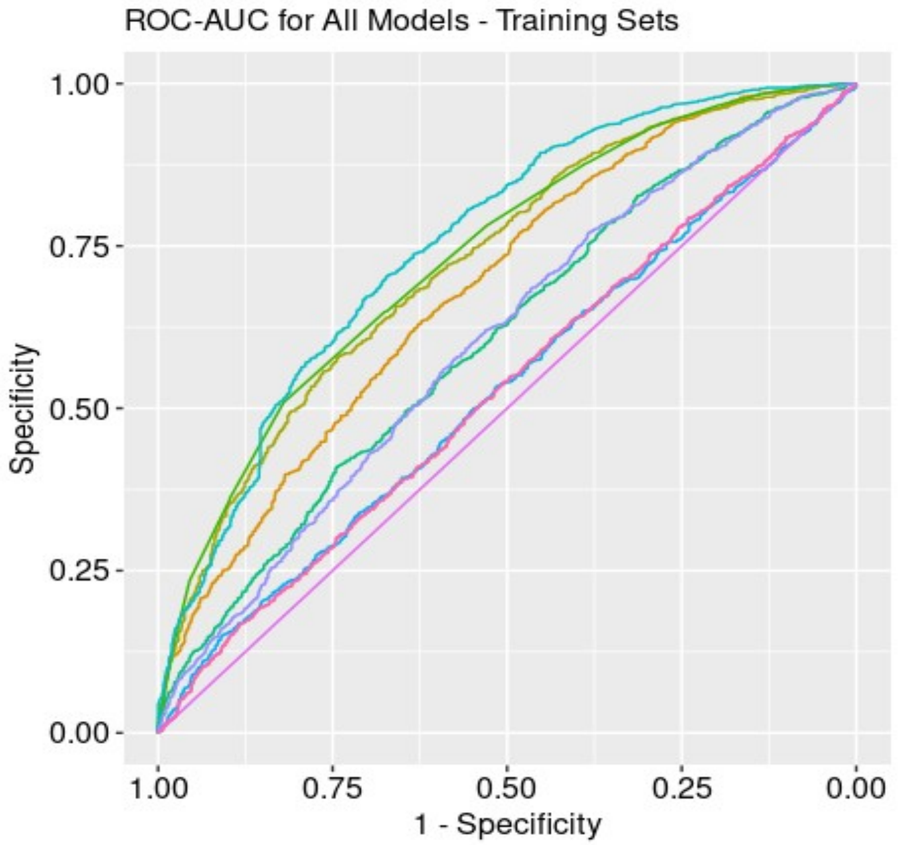
total amount of organic carbon compounds found in water sources which can determine the potable water quality.

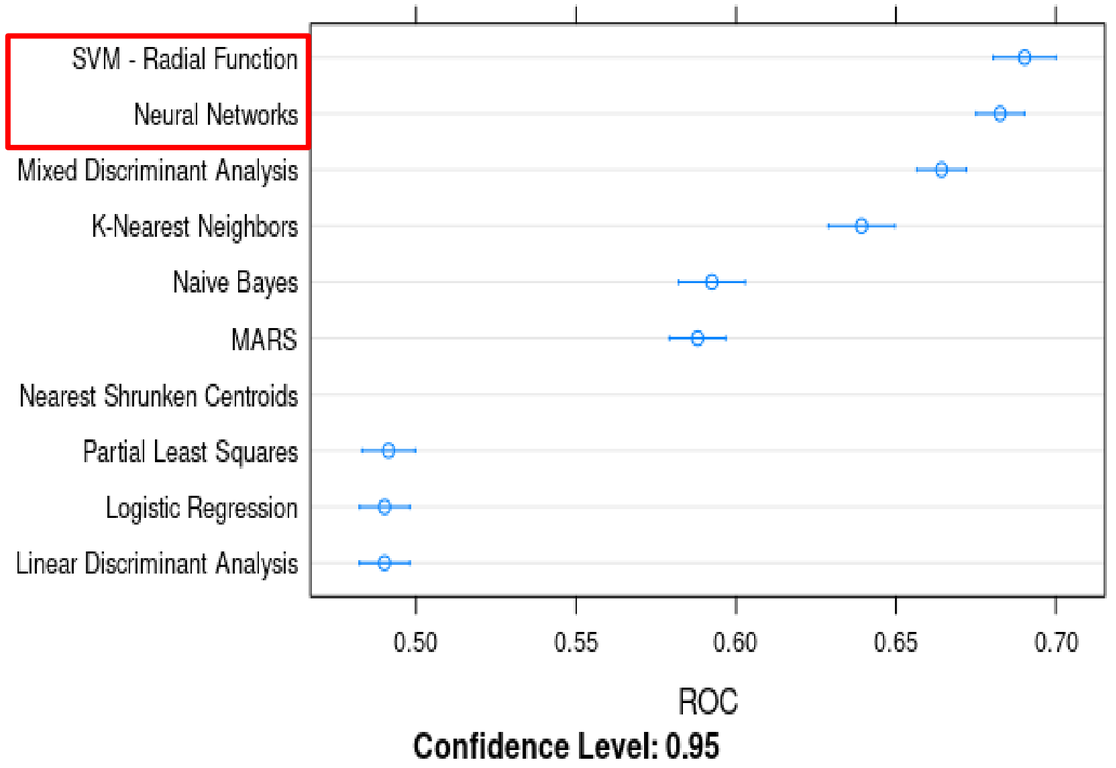
Quality 6

turbidity describes the amount of solid matter present in the suspended state of water.

Distribution of Water Quality Metrics

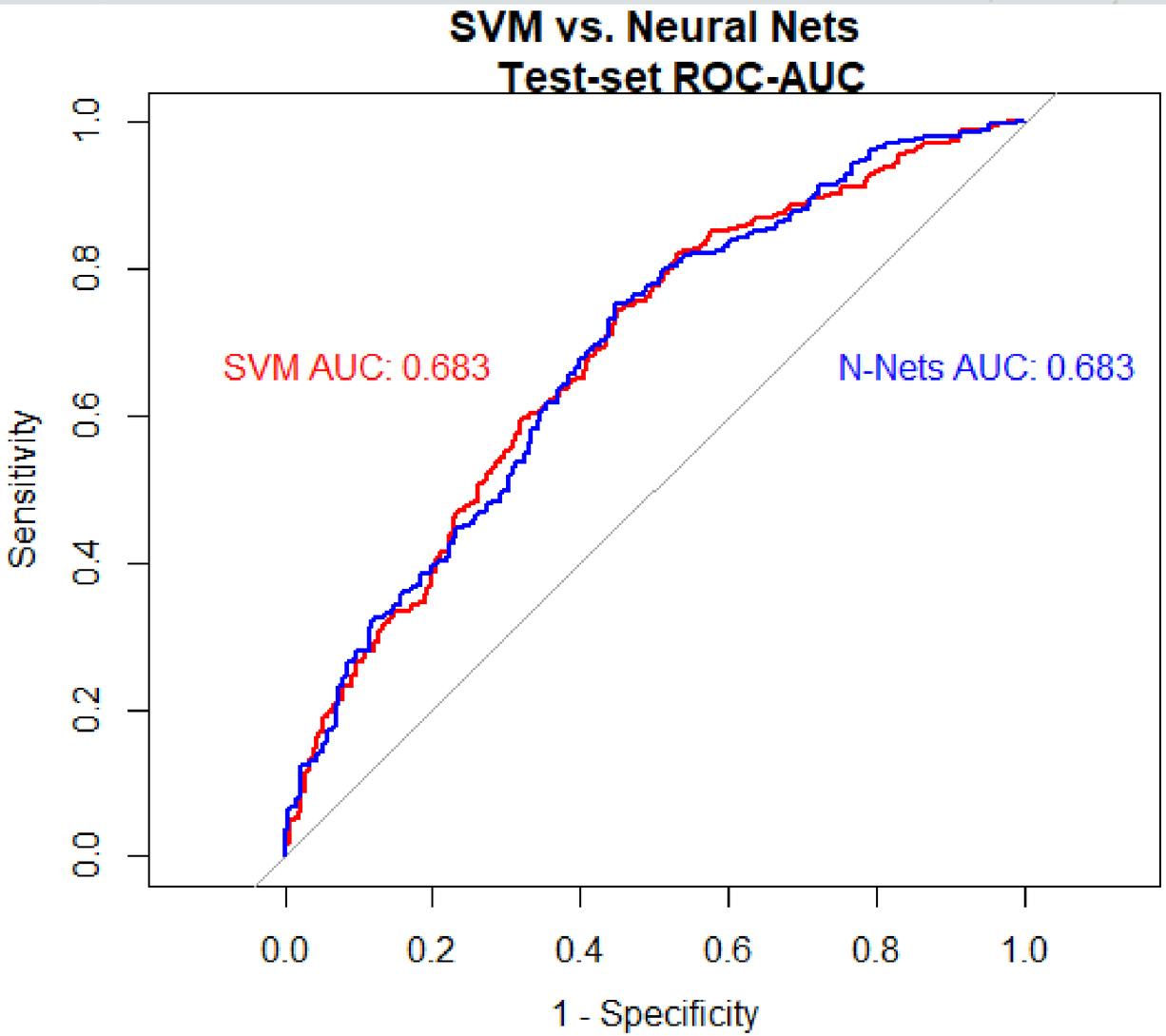


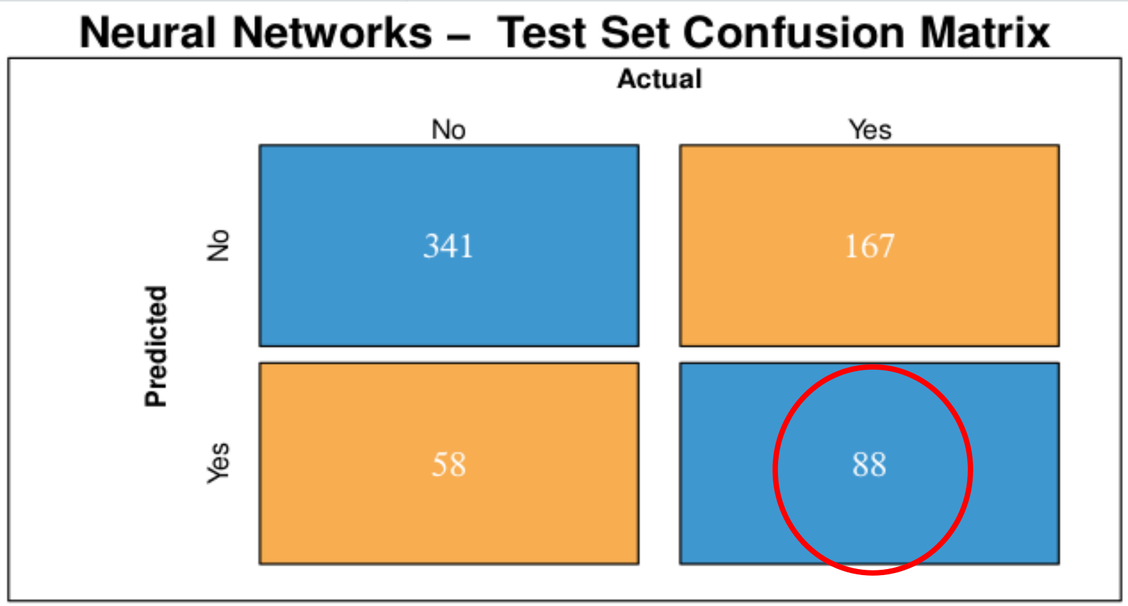


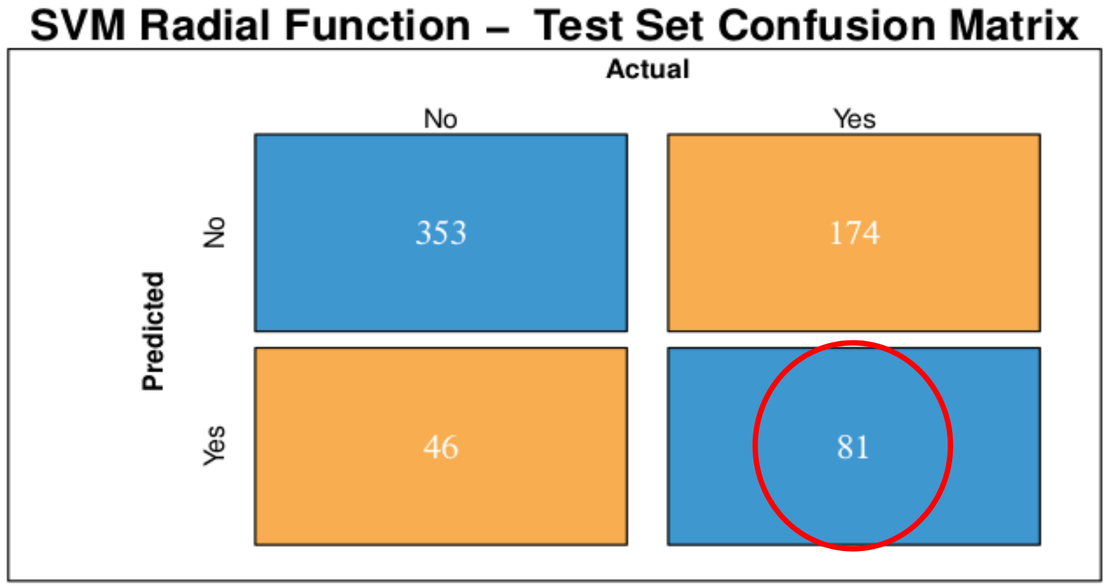


Models	AUC	Sensitivity	Specificity
Linear Discriminant Analysis	0.5297	0	1
Mixed Discriminant Analysis	0.6841	0.305	0.9012
Neural Networks	0.7241	0.4223	0.8618
K-Nearest Neighbors	0.7335	0.2884	0.935
Naive Bayes	0.6072	0.2336	0.8768
SVM - Radial Function	0.7545	0.3744	0.9293
Partial Least Squares	0.5292	0	1
MARS	0.6032	0.1672	0.9199
Nearest Shrunken Centroids	0.5	0	1
Logistic Regression	0.5297	0	1

Modeling







Final Results

Conclusion & Suggestions

01

First Conclusion

After resample of all ROC models, Neural Network outperformed K-nearest neighbors.

02

Second Conclusion

Neural Network is the preferred model for this Dataset.

03

First suggestion...

04

Second suggestion…