PARALLELIZED GRID SEARCH OPTIMIZATION FOR WATER QUALITY CLASSIFICATION

Overview

Presenter: Haya

Introduction

Water quality is an important indicator of overall health. Water pollution can alter the chemistry of water which leads to low water quality, and the consumption of poor quality water can harm both human health and the surrounding environment.

In our project, we applied Machine Learning techniques to build a water quality prediction model, that is time and cost effective to predict the quality of water based on several parameters to detect if the water source is safe for consumption or not.

Presenter: Haya

Proposed Techniques

01 Random Forest

02 Support Vector Machine

03 Gradient Boosting Machine

Presenter: Joury

Dataset

pH 01

Hardness 02

Solids (total dissolved solids) 03

Chloramines 04

Sulfate 05

Conductivity 06

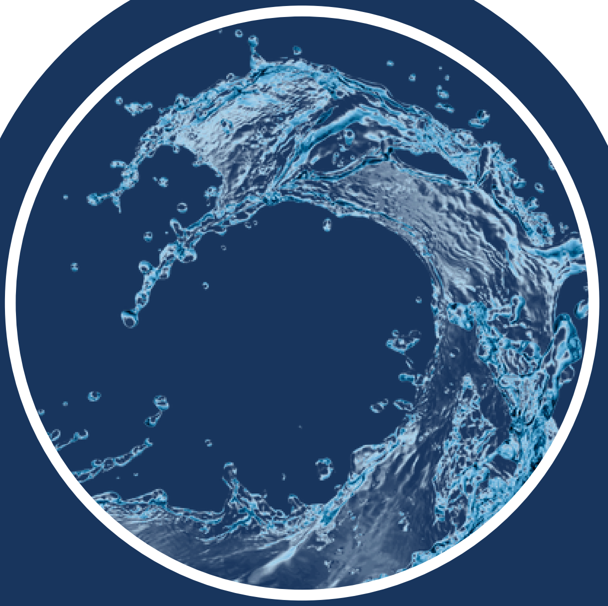
Organic carbon 07

Trihalomethanes 08

Turbidity 09

Presenter: Joury

Experimental Setup



Filling mising data with median

SMOT

Training model using:
-SVM
-RF
-GB

Feature Selection

Presenter: Abrar

Performance Measures

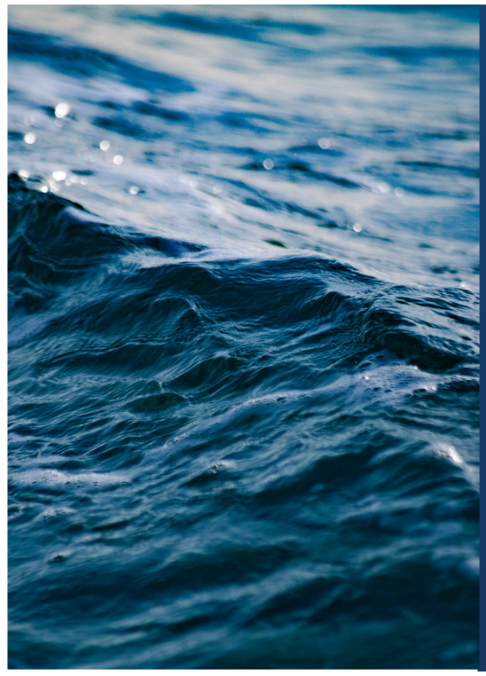
01 Accuracy

02 Precision

03 Recall

Presenter: Abrar

Optimization Strategy



01

Random Forest

- max_depth: 10
- max_features: ‘auto ‘
- min_samples_leaf: 4
- n_estimators: 5

03

Support Vector Machine

- max_depth: 3
- Max_features: ‘Auto ‘
- min_samples_leaf: 2
- min_samples_split: 5
- n_estimators: 100

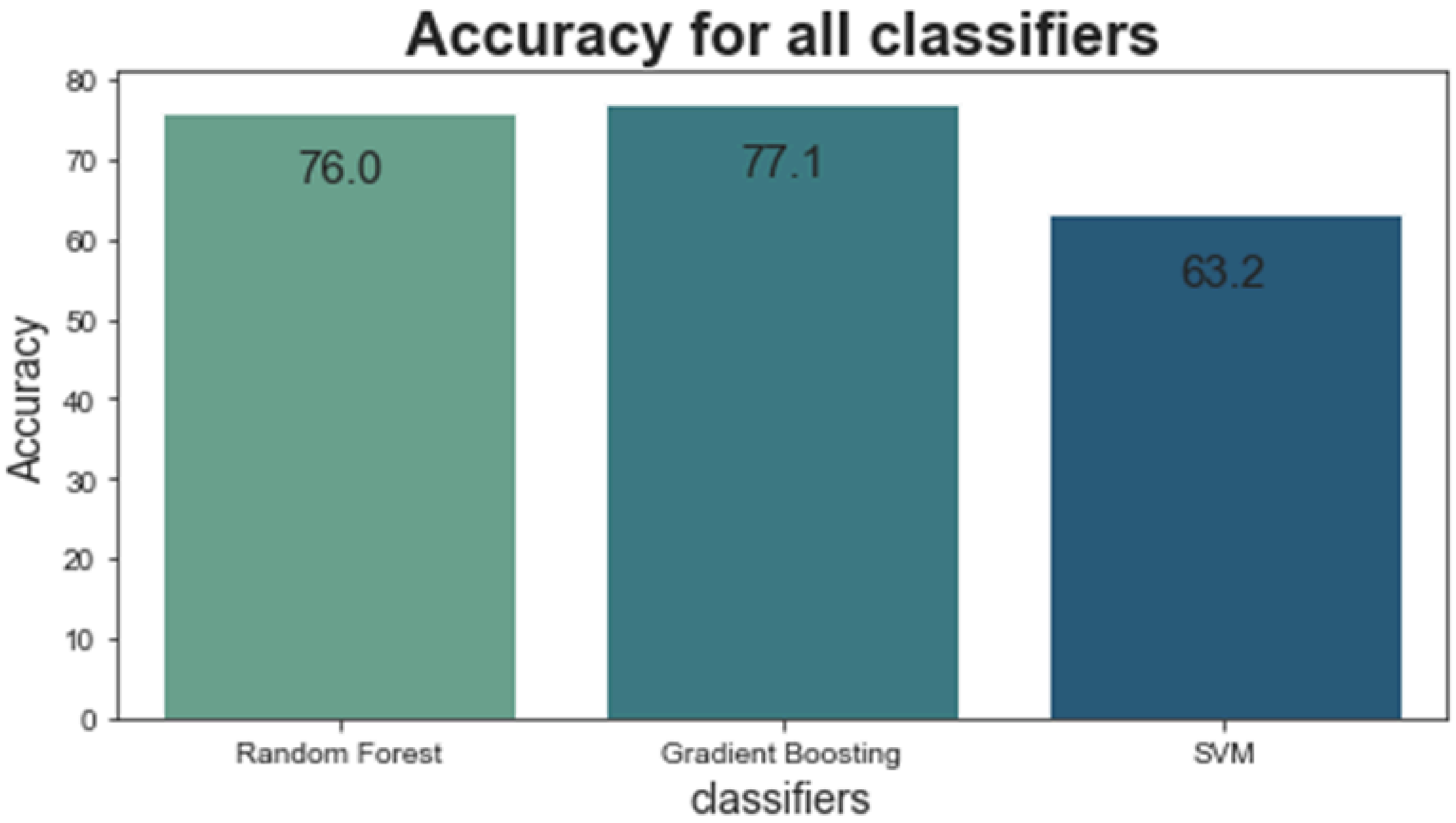
02

Gradient Boosting

- C: 1
- coef0: 0.5
- degree: 1
- gamma: 0.1
- kernel: ‘rbf ’

Presenter: Manar

Results Using All Features

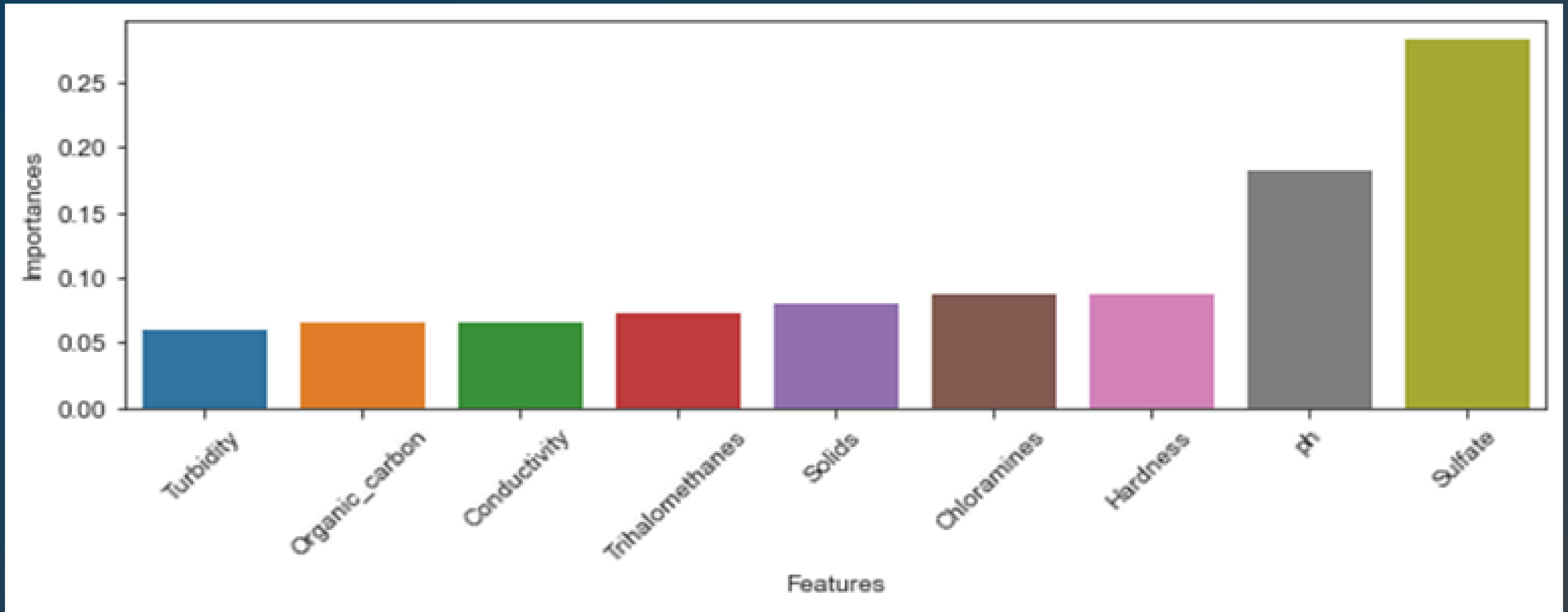


	Accuracy	Recall	Precision	F1-score
Random Forest	76%	73%	67%	70%
Gradient Boosting Machine	77.1%	79%	67%	73%
Support Vector Machine	63%	61%	52%	56%

Presenter: Manar

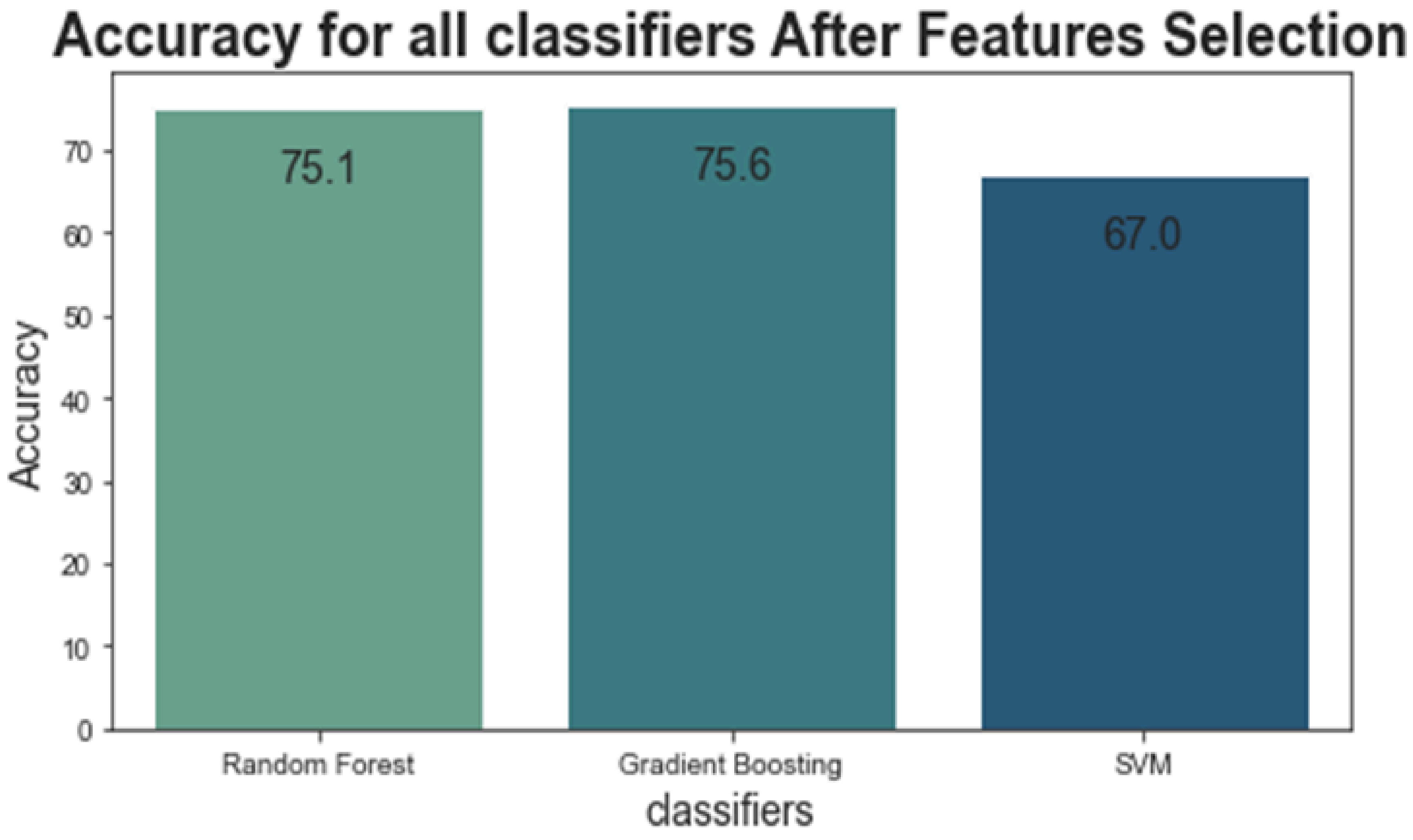
Feature Selection

Output of Embedded method



Presenter: Manar

Results After Feature Selection



	Accuracy	Recall	Precision	F1-score
Random Forest	75.1%	76%	65%	70%
Gradient Boosting Machine	75.6%	77%	66%	71%
Support Vector Machine	67%	64%	56%	60%

Presenter: Manar

Our Result Vs. Previous Studies

Previous studies

73% Svm

Percentage difference

4% More

Our highest result

77.1%

Gradient Boosting

by Kaddoura reference 13 in the report

Presenter:
Wahbia

Tools

Python using Jupyter Notebook

RF, SVM, GBM

Lenovo yoga core i7 11th gen

Hp core i7 11th gen

Hp core i7 10th gen

Numpy, Pandas, MATLAB

Presenter:
Wahbia

Conclusion

- Analyzing water quality is crucial, and machine learning algorithms aid in that by making it faster and cost-effective.
- A dataset of 3276 water samples with 9 attributes was used in the model.
- The algorithms that were used were Random Forest, Support Vector Machine and Gradient Boosting Machine.
- GBM outperformed the other algorithms with an accuracy of 77.1%, 79% recall, 67% precision and 73% F1-score.
- The proposed algorithms helped with classifying water into potable and non-potable.