Problem Definition Description:

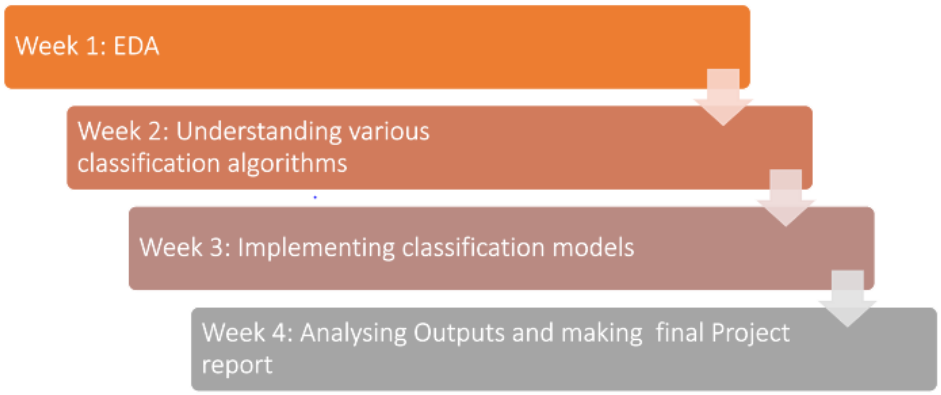
The presence of contaminants such as sewage waste, heavy metal, fertilizers, pesticides can be the cause of many illnesses in humans such as gastrointestinal illness, reproductive problems, and neurological disorders. Individuals with weakened immune systems may be at greater risk. Water testing is important for monitoring the safety of drinking water, investigation of disease outbreaks, validation process and preventive measures, especially in a country like India where 40% of the population is not expected to have clean water by 2030

The attempt here is to classify whether water is potable for drinking based on pH value, hardness of water, TDS (Total dissolved solids)value, Chloramines and sulphates content, conductivity, Total Organic Carbon (TOC), Trihalomethanes and turbidity of water and further to perform intense EDA and compare various classification models based on their prediction accuracy and effectiveness. This aims at the tapping of the cumulative effect of contaminants on the potability of water which is often overlooked

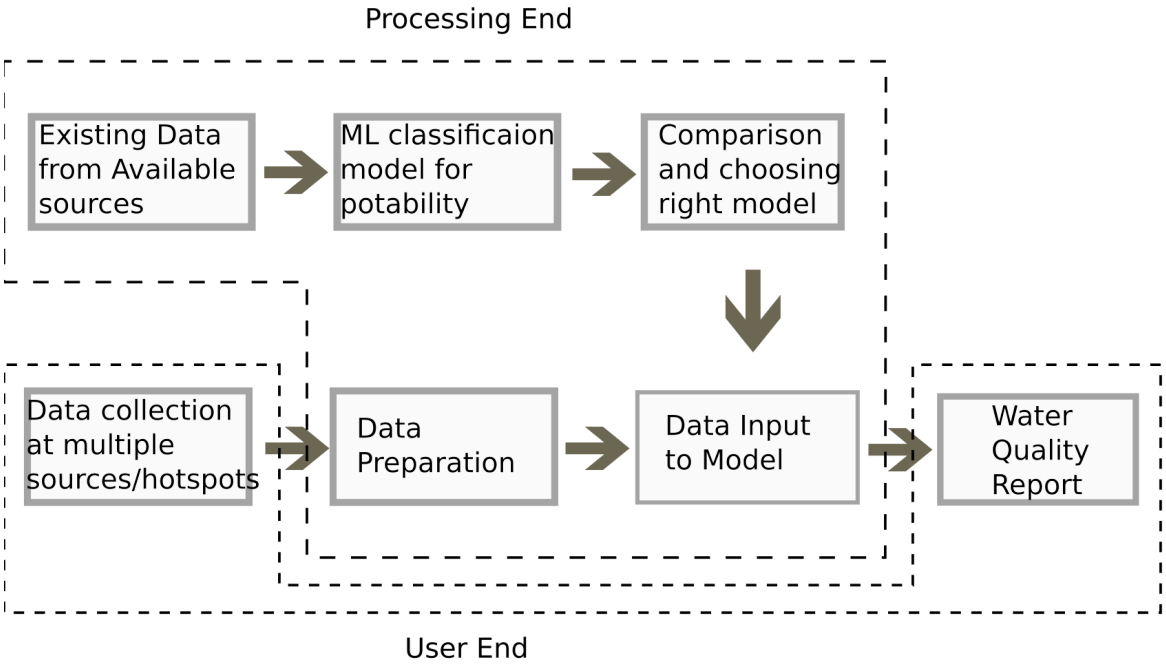
Technology landscape assessment:

In some regions, it has been shown that investments in water supply and sanitation can yield a net economic benefit, since the reductions in adverse health effects and health care costs outweigh the costs of undertaking the interventions. The general procedure for water quality assessment pivots around the investigative analysis followed by comparison to defined levels of each parameter, which neglects the obvious cumulative effect of chemicals/contaminants. Replacing Manual interventions with the deployed model could reduce overall cost of the project.

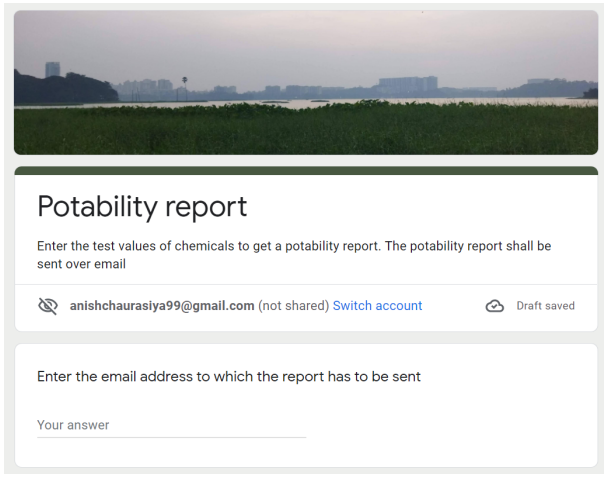
Project Planning:



Conceptual Design



Screenshots of User interface:



Potability report

Enter the test values of chemicals to get a potability report. The potability report shall be sent over email

anishchaurasiya99@gmail.com (not shared) Switch account

Draft saved

Enter the email address to which the report has to be sent

Your answer

Select the type of model you want to go ahead with

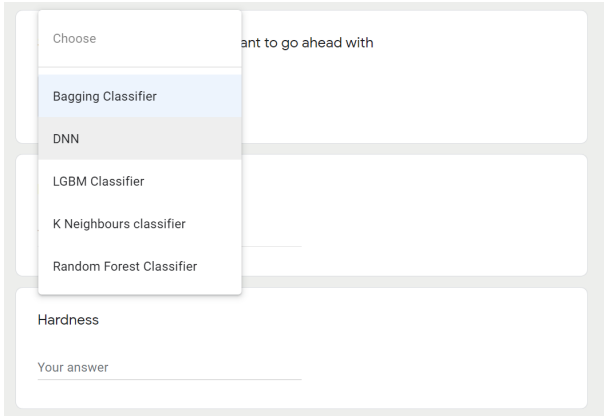
Bagging Classifier

pH

Your answer

Hardness

Your answer



Choose

Bagging Classifier

DNN

LGBM Classifier

K Neighbours classifier

Random Forest Classifier

ant to go ahead with

Hardness

Your answer

Solids

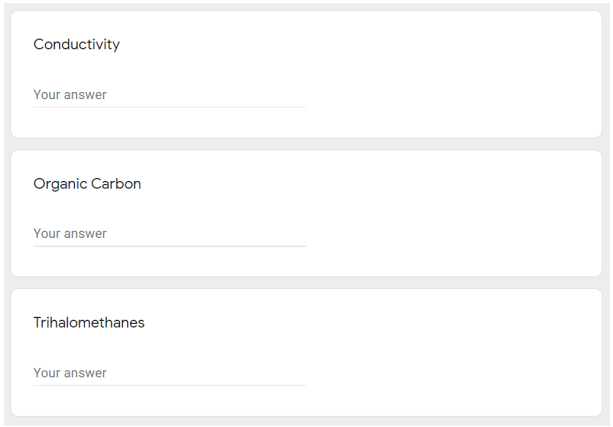
Your answer

Chloramines

Your answer

Sulphates

Your answer



Conductivity

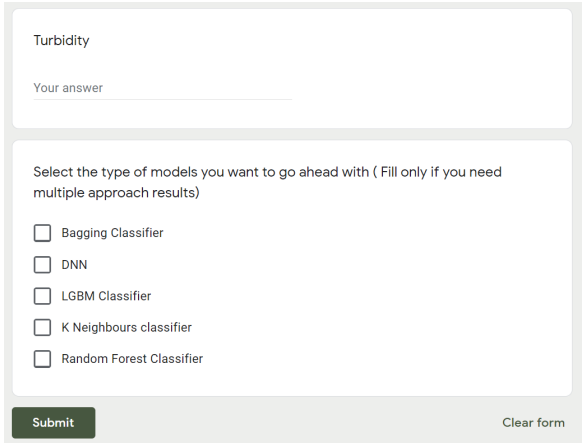
Your answer

Organic Carbon

Your answer

Trihalomethanes

Your answer



Turbidity

Your answer

Select the type of models you want to go ahead with (Fill only if you need multiple approach results)

- [ ] Bagging Classifier
- [ ] DNN
- [ ] LGBM Classifier
- [ ] K Neighbours classifier
- [ ] Random Forest Classifier

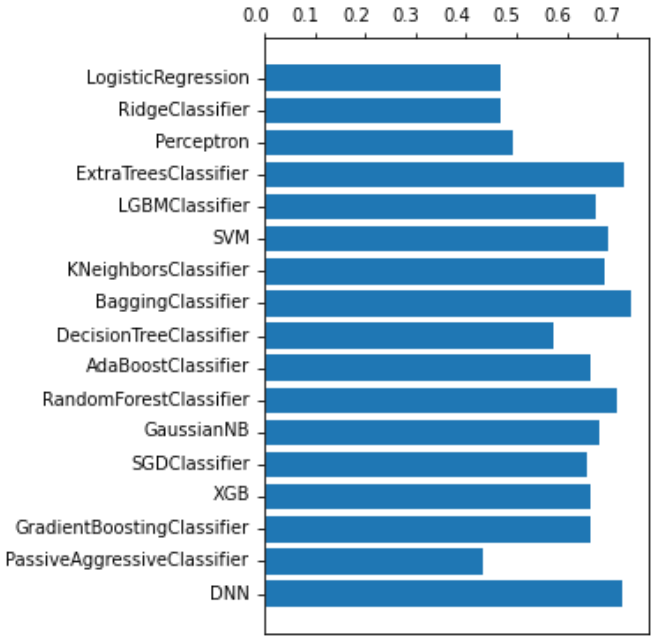
Submit

Clear form

Unit testing report:

0	LogisticRegression	0
1	RidgeClassifier	0
2	Perceptron	0
3	ExtraTreesClassifier	0
4	LGBMClassifier	0
5	SVM	0
6	KNeighborsClassifier	0
7	BaggingClassifier	0
8	DecisionTreeClassifier	0
9	AdaBoostClassifier	0
10	RandomForestClassifier	0
11	GaussianNB	0
12	SGDClassifier	0
13	XGB	0

Model training and testing report:



	Model_Name	Precission	Recall	Accuracy
0	LogisticRegression	0.284211	0.409091	0.465
1	RidgeClassifier	0.284211	0.409091	0.465
2	Perceptron	0.336364	0.560606	0.490
3	ExtraTreesClassifier	0.586957	0.409091	0.710
4	LGBMClassifier	0.480519	0.560606	0.655
5	SVM	0.526316	0.303030	0.680
6	KNeighborsClassifier	0.530303	0.500000	0.670
7	BaggingClassifier	0.622222	0.424242	0.725
8	DecisionTreeClassifier	0.383721	0.500000	0.570
9	AdaBoostClassifier	0.454545	0.378788	0.645
10	RandomForestClassifier	0.543860	0.469697	0.695
11	GaussianNB	0.475000	0.287879	0.660
12	SGDClassifier	0.436364	0.363636	0.635
13	XGB	0.333333	0.075758	0.645
14	GradientBoostingClassifier	0.452830	0.363636	0.645
15	PassiveAggressiveClassifier	0.293103	0.515152	0.430
16	DNN	0.559322	0.500000	0.705

User Manual:

User Inputs through Google-form and Model Outputs through Email.

1st Method:
Manually entering all the 9 values from the lab report
Selecting Models
Entering Email-ID for report/models output

2nd Method: (Optional)
Clicking photo of the Lab report and uploading them
Selecting Models
Entering Email-ID for report/models output
Using OCR Technique, table/values are extracted from the image for model Input

Link to code:
https://colab.research.google.com/drive/1pg3axtCLtNTRYsh3lKoj-S_UTEJMCvUS?usp=sharing