

SPRINT 2

Project Deliverables (Model Building Code & Evaluation)

Team ID	PNT2022TMID37191 Type your text
Project Name	Efficient Water Quality Analysis & Prediction using Machine Learning

LOGISTIC REGRESSION EVALUATION:

colab.research.google.com/drive/1n-8a3LcpLFWWCgFkNYEVVRdCGCo3pGPLI#scrollTo=XH3BxirdFnFc

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

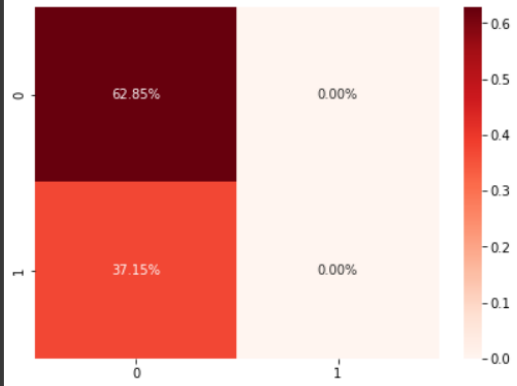
Files

- sample_data
- water(1).csv

```
[46] /usr/local/lib/python3.7/dist-packages/sklearn/metrics/_classification.py:1318: UndefinedMetricWarning: Precision and F-score are ill-defined and b
_warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.7/dist-packages/sklearn/metrics/_classification.py:1318: UndefinedMetricWarning: Precision and F-score are ill-defined and b
_warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.7/dist-packages/sklearn/metrics/_classification.py:1318: UndefinedMetricWarning: Precision and F-score are ill-defined and b
_warn_prf(average, modifier, msg_start, len(result))
```

```
[48] # confusion Maxtrix
cm1 = confusion_matrix(y_test, pred_lg)
sns.heatmap(cm1/np.sum(cm1), annot = True, fmt= '0.2%', cmap = 'Reds')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f5e9a6a2090>



	0	1
0	62.85%	0.00%
1	37.15%	0.00%

0s completed at 10:59

model evaluation.pdf

29°C Haze

Search

ENG IN

11:26 18-11-2022

DECISION TREE EVALUATION:

colab.research.google.com/drive/1n-8a3LcpLFWWCgFkNYEVrdCGCo3pGPLI#scrollTo=XH3BxirdFnFc

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- water(1).csv

+ Code + Text

Insert code cell below Ctrl+M B

	precision	recall	f1-score	support
0	0.66	0.90	0.76	680
1	0.56	0.22	0.32	402
accuracy			0.65	1082
macro avg	0.61	0.56	0.54	1082
weighted avg	0.62	0.65	0.60	1082

```
[55] # confusion Maxtrix
cm2 = confusion_matrix(y_test, pred_dt)
sns.heatmap(cm2/np.sum(cm2), annot = True, fmt= '0.2%', cmap = 'Reds')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f5e9984d3d0>

Confusion matrix values:

- Actual 0, Predicted 0: 56.28%
- Actual 0, Predicted 1: 6.56%
- Actual 1, Predicted 0: 28.93%
- Actual 1, Predicted 1: 8.23%

Disk 85.16 GB available

29°C Haze

Search

ENG IN

11:29 18-11-2022

RANDOM FOREST EVALUATION:

colab.research.google.com/drive/1n-8a3LpLFWWCgFkNYEVrdCGCo3pGPLI#scrollTo=XH3BxirdFnFc

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

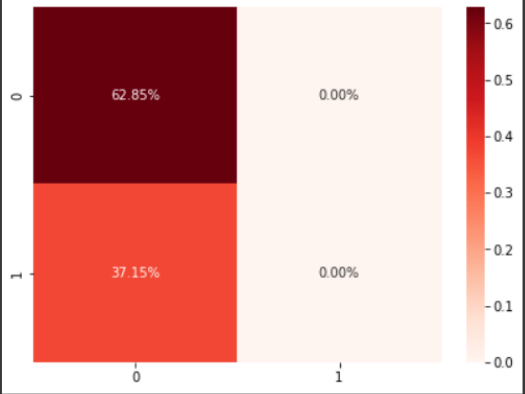
Files

- sample_data
- water(1).csv

```
[61] /usr/local/lib/python3.7/dist-packages/sklearn/metrics/_classification.py:1318: UndefinedMetricWarning: Precision and F-score are ill-defined and b
_warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.7/dist-packages/sklearn/metrics/_classification.py:1318: UndefinedMetricWarning: Precision and F-score are ill-defined and b
_warn_prf(average, modifier, msg_start, len(result))
/usr/local/lib/python3.7/dist-packages/sklearn/metrics/_classification.py:1318: UndefinedMetricWarning: Precision and F-score are ill-defined and b
_warn_prf(average, modifier, msg_start, len(result))
```

```
# confusion Maxtrix
cm3 = confusion_matrix(y_test, pred_rf)
sns.heatmap(cm3/np.sum(cm3), annot = True, fmt= '0.2%', cmap = 'Reds')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f5e98168190>



	0	1
0	62.85%	0.00%
1	37.15%	0.00%

Using XGBoost Classifier

Disk 85.16 GB available

29°C Haze

Search

ENG IN 11:31 18-11-2022

XGBOOST CLASSIFIER EVALUATION:

colab.research.google.com/drive/1n-8a3LcpLFWWCgFkNYEVRdCGCo3pGPLI#scrollTo=XH3BxirdFnFc

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- water(1).csv

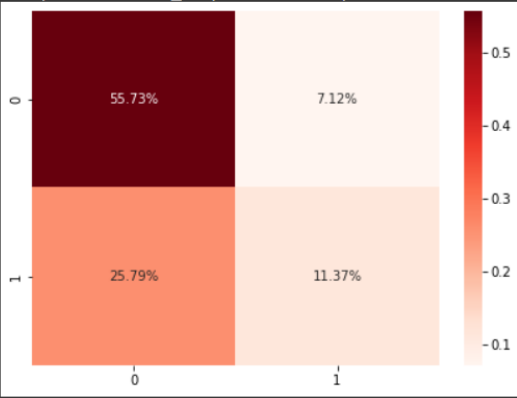
+ Code + Text

precision recall f1-score support

0	0.68	0.89	0.77	680
1	0.61	0.31	0.41	402
accuracy			0.67	1082
macro avg	0.65	0.60	0.59	1082
weighted avg	0.66	0.67	0.64	1082

```
# confusion Maxtrix
cm4 = confusion_matrix(y_test, pred_xgb)
sns.heatmap(cm4/np.sum(cm4), annot = True, fmt= '0.2%', cmap = 'Reds')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f5e8f2fc110>



Using KNeighbours

Disk 85.16 GB available

29°C Haze

Search

ENG IN

11:32 18-11-2022

KNN EVALUATION:

colab.research.google.com/drive/1n-8a3LcpLFWWCgFkNYEVrdCGCo3pGPLI#scrollTo=XH3BxirdFnFc

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- ..
- sample_data
- water(1).csv

```
+ Code + Text
[75] print(classification_report(y_test, pred_kn))

              precision    recall  f1-score   support

      0       0.69       0.82       0.75        680
      1       0.55       0.37       0.44        402

 accuracy          0.65          1082
 macro avg       0.62       0.60       0.59          1082
 weighted avg    0.64       0.65       0.63          1082
```

```
# confusion Maxtrix
cm5 = confusion_matrix(y_test, pred_kn)
sns.heatmap(cm5/np.sum(cm5), annot = True, fmt = '0.2%', cmap = 'Reds')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f5e8d230510>

	Actual 0	Actual 1
Predicted 0	51.66%	11.18%
Predicted 1	23.48%	13.68%

Disk 85.16 GB available

29°C Haze

Search

ENG IN 11:33 18-11-2022

SVM EVALUATION:

colab.research.google.com/drive/1n-8a3LcplfWWCgFkNYEVrdCGCo3pGPLI#scrollTo=XH3BxirdFnFc

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

RAM Disk

Editing

Files

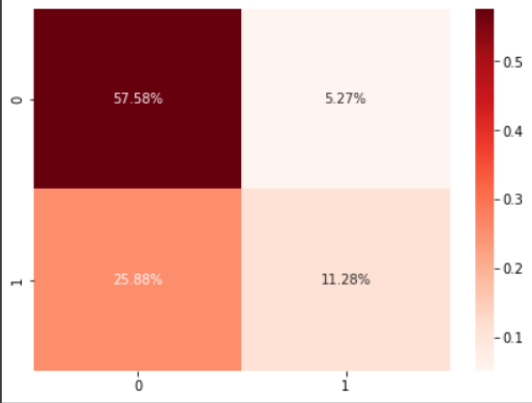
sample_data

water(1).csv

```
[82] accuracy      0.65      1082
      macro avg   0.62      1082
      weighted avg 0.64      1082
```

```
# confusion Maxtrix
cm6 = confusion_matrix(y_test, pred_svm)
sns.heatmap(cm6/np.sum(cm6), annot = True, fmt= '0.2%', cmap = 'Reds')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f5e8d164ad0>



```
## Using AdaBoost Classifier
from sklearn.ensemble import AdaBoostClassifier
```

```
model_ada = AdaBoostClassifier(learning_rate= 0.002,n_estimators= 205,random_state=42)
```

29°C Haze

Search

ENG IN

11:33 18-11-2022

ADABOOST CLASSIFIER:

colab.research.google.com/drive/1n-8a3LcpLFWWCgFkNYEVrdCGCo3pGPLI#scrollTo=XH3BxirdFnFc

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

RAM Disk

Editing

Files

sample_data

water(1).csv

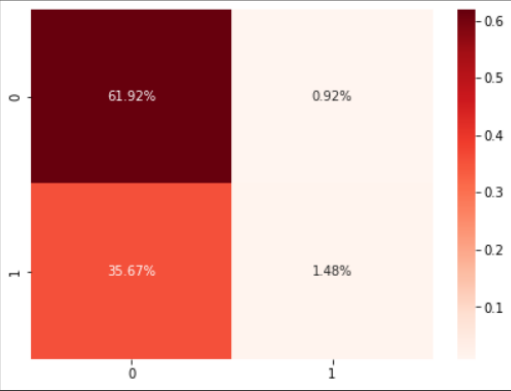
```
[89]
      precision    recall  f1-score   support

     0       0.63      0.99      0.77       680
     1       0.62      0.04      0.07       402

 accuracy          0.63       1082
 macro avg          0.62       1082
 weighted avg       0.63       1082
```

```
# confusion Maxtrix
cm7 = confusion_matrix(y_test, pred_ada)
sns.heatmap(cm7/np.sum(cm7), annot = True, fmt= '0.2%', cmap = 'Reds')
```

<matplotlib.axes._subplots.AxesSubplot at 0x7f5e8d0e4950>



	Actual 0	Actual 1
Predicted 0	61.92%	0.92%
Predicted 1	35.67%	1.48%

Disk 85.16 GB available

29°C Haze

Search

ENG IN

11:34 18-11-2022

COMPARISSON:

colab.research.google.com/drive/1n-8a3LcpLFWWCgFkNYEVrdCGCo3pGPLI#scrollTo=XH3BxirdFnFc

Untitled2.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Files

- sample_data
- water(1).csv

```
models
sns.barplot(x='Accuracy_score', y='Model', data=models)

models.sort_values(by='Accuracy_score', ascending=False)
```

	Model	Accuracy_score
5	SVM	0.688540
3	XGBoost	0.670980
4	KNeighbours	0.653420
1	Decision Tree	0.645102
6	AdaBoost	0.634011
0	Logistic Regression	0.628466
2	Random Forest	0.628466

Model

Logistic Regression

Decision Tree

Random Forest

XGBoost

KNeighbours

SVM

AdaBoost

29°C Haze

Search

ENG IN

11:37 18-11-2022

CONCLUSION:

Out of all we found that *SVM Classifier* has the highest accuracy among all , so we are going to use *SVM Classifier* to train **our Water Prediction machine learning model**