COMSATS UNIVERSITY ISLAMABAD, LAHORE CAMPUS
MACHINE LEARNING

ASSIGNMENT NO. 2

SUBMITTD TO:

DR. MUHAMMAD SHARJEEL

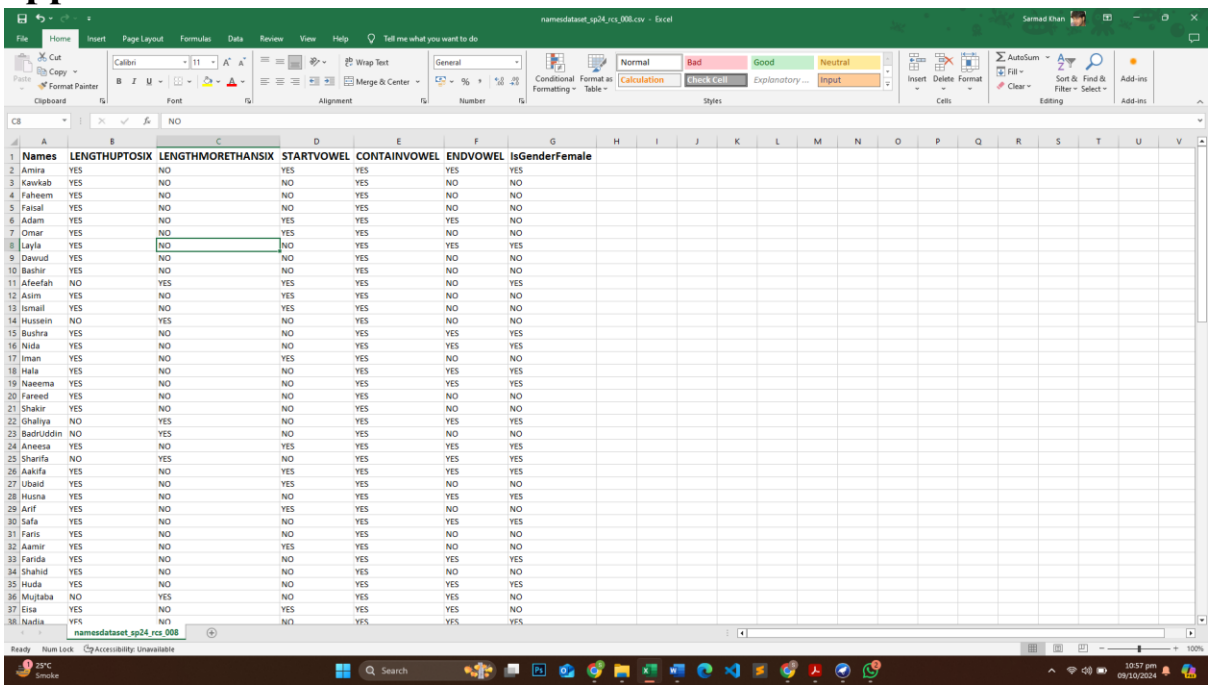
SUBMITTED BY:

SARMAD ZAMIR KHAN

SP24 – RCS – 008

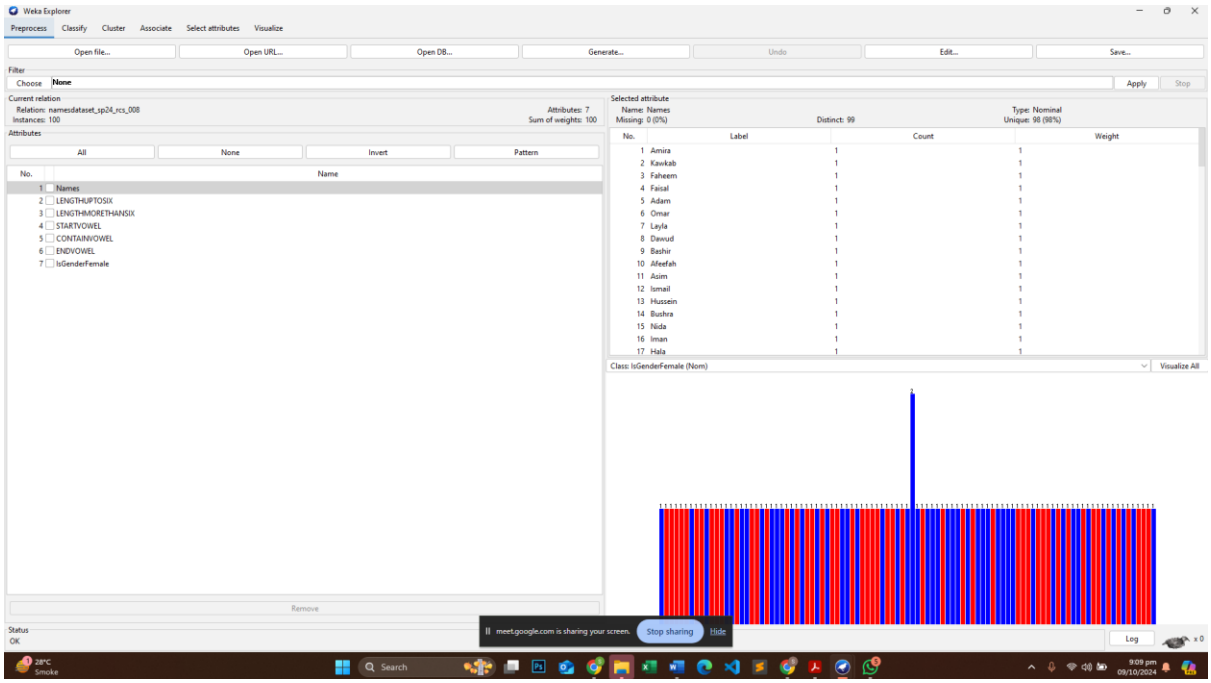
COMSATS UNIVERSITY ISLAMABAD, LAHORE CAMPUS
MACHINE LEARNING

The name dataset was explored for various attributes to be applied.

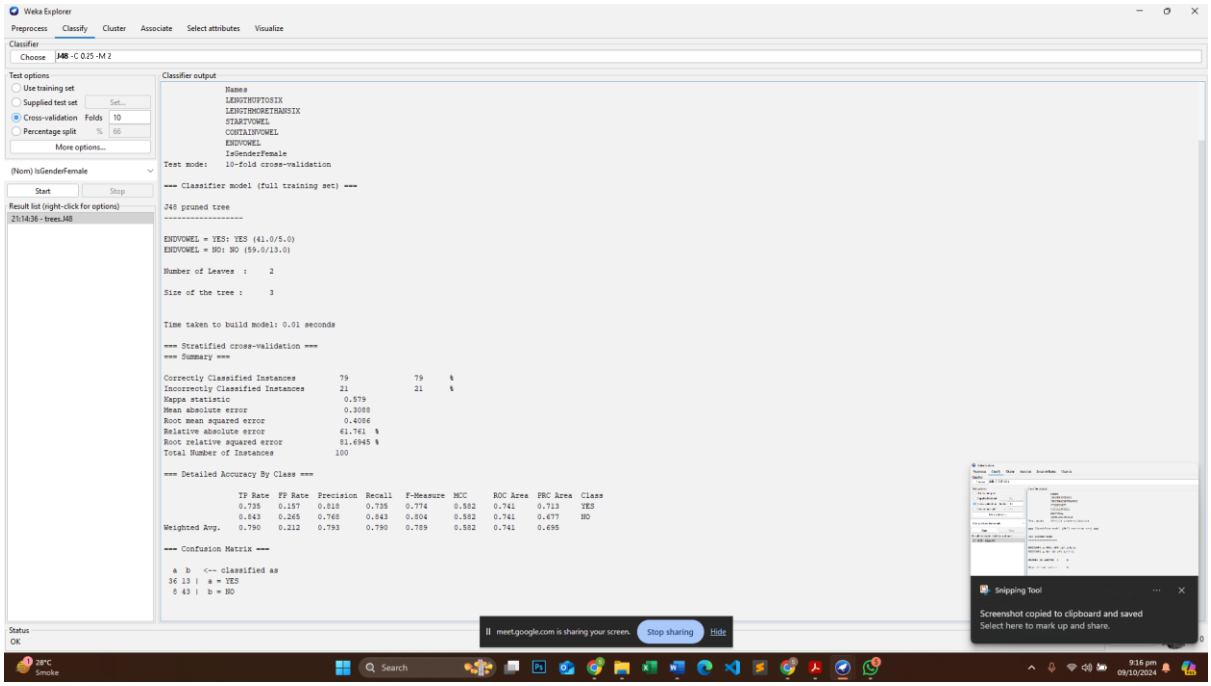


Names	LENGTHUPTOSIX	LENGTHMORETHANSIX	STARTVOWEL	CONTAINVOWEL	ENDVOWEL	IsGenderFemale
Amira	YES	NO	YES	YES	YES	YES
Kawkab	YES	NO	NO	YES	NO	NO
Faheem	YES	NO	NO	YES	NO	NO
Faisal	YES	NO	NO	YES	NO	NO
Adam	YES	NO	YES	YES	YES	NO
Omar	YES	NO	YES	YES	NO	NO
Layla	YES	NO	NO	YES	YES	YES
Dawud	YES	NO	NO	YES	NO	NO
Bashir	YES	NO	NO	YES	NO	NO
Afeefah	NO	YES	YES	YES	NO	YES
Asim	YES	NO	YES	YES	NO	NO
Ismail	YES	NO	YES	YES	NO	NO
Hussein	NO	YES	NO	YES	NO	NO
Bushra	YES	NO	NO	YES	YES	YES
Nida	YES	NO	NO	YES	YES	YES
Iman	YES	NO	YES	YES	NO	NO
Hala	YES	NO	NO	YES	YES	YES
Naeema	YES	NO	NO	YES	YES	YES
Fareed	YES	NO	NO	YES	NO	NO
Shakir	YES	NO	NO	YES	NO	NO
Ghaliya	NO	YES	NO	YES	YES	YES
BadrUddin	NO	YES	NO	YES	NO	NO
Aneesa	YES	NO	YES	YES	YES	YES
Sharifa	NO	YES	NO	YES	YES	YES
Aakifa	YES	NO	YES	YES	YES	YES
Ubaid	YES	NO	YES	YES	NO	NO
Husna	YES	NO	NO	YES	YES	YES
Arif	YES	NO	YES	YES	NO	NO
Safa	YES	NO	NO	YES	YES	YES
Faris	YES	NO	NO	YES	NO	NO
Aamir	YES	NO	YES	YES	NO	NO
Farida	YES	NO	NO	YES	YES	YES
Shahid	YES	NO	NO	YES	NO	NO
Huda	YES	NO	NO	YES	YES	YES
Mujtaba	NO	YES	NO	YES	YES	NO
Eisa	YES	NO	YES	YES	YES	NO
Nadia	YES	NO	NO	YES	YES	YES

This dataset was then fed to weka for application of Machine Learning Algorithm. Before that ARFF file was created using weka.



After that the ARFF file was fed to WEKA and J48 machine learning algorithm was applied.



Conclusion:

The ML Pipeline explains that outcome obtained is a classification model evaluation report generated by WEKA. A breakdown of the key components and their significance suggest that J48 (C4.5) decision tree with a confidence factor of 0.25 and minimum number of instances per leaf of 2.

Dataset: namesdataset_sp24_rcs_008 with 100 instances and 7 attributes.

Evaluation: 10-fold cross-validation.

Model Performance:

The model performance was almost satisfactory,

Overall Accuracy: 79% (correctly classified instances).

Detailed Accuracy By Class: TP Rate, FP Rate, Precision, Recall, F-Measure, MCC, ROC Area, PRC Area for each class (YES and NO).

Confusion Matrix: Shows the distribution of predicted and actual classes.

Interpretation:

The J48 decision tree model achieved a relatively high accuracy of 79% on the given dataset.

The detailed accuracy by class metrics provide insights into the model's performance for each class. For example, the TP Rate for class YES is 0.735, indicating that 73.5% of actual YES instances were correctly classified.

The confusion matrix helps visualize the model's classification errors. In this case, 13 instances were misclassified as YES, and 8 instances were misclassified as NO.

Overall, the model appears to perform reasonably well on the dataset, but further analysis and domain knowledge would be required to assess its suitability for specific applications. However, in my opinion, no decisive results have been produced by the model.