6th October,2024

Machine-Learning

Analysis of “Names” Characteristics Using Different Classifiers

Submitted To: Dr. Muhammad Sharjeel

Submitted By: Shehar Zaad

Registratin No: SP24-RCS-010

## Introduction:

The assignment involves applying different machine learning classifiers to a dataset of “Names” to understand how various characteristics of names impact classification outcomes. The dataset includes names labeled with symbols like “Y” for yes and “N” for no, and the main goal is to identify patterns in the names that can be useful for classification. For this purpose, I manually extracted several features from each name, such as whether the second letter is a vowel, if the name starts or ends with a vowel, and whether the name’s length is even or odd. For analysis, I primarily used classifiers like the J48 decision tree and LMT.

**Dataset:**

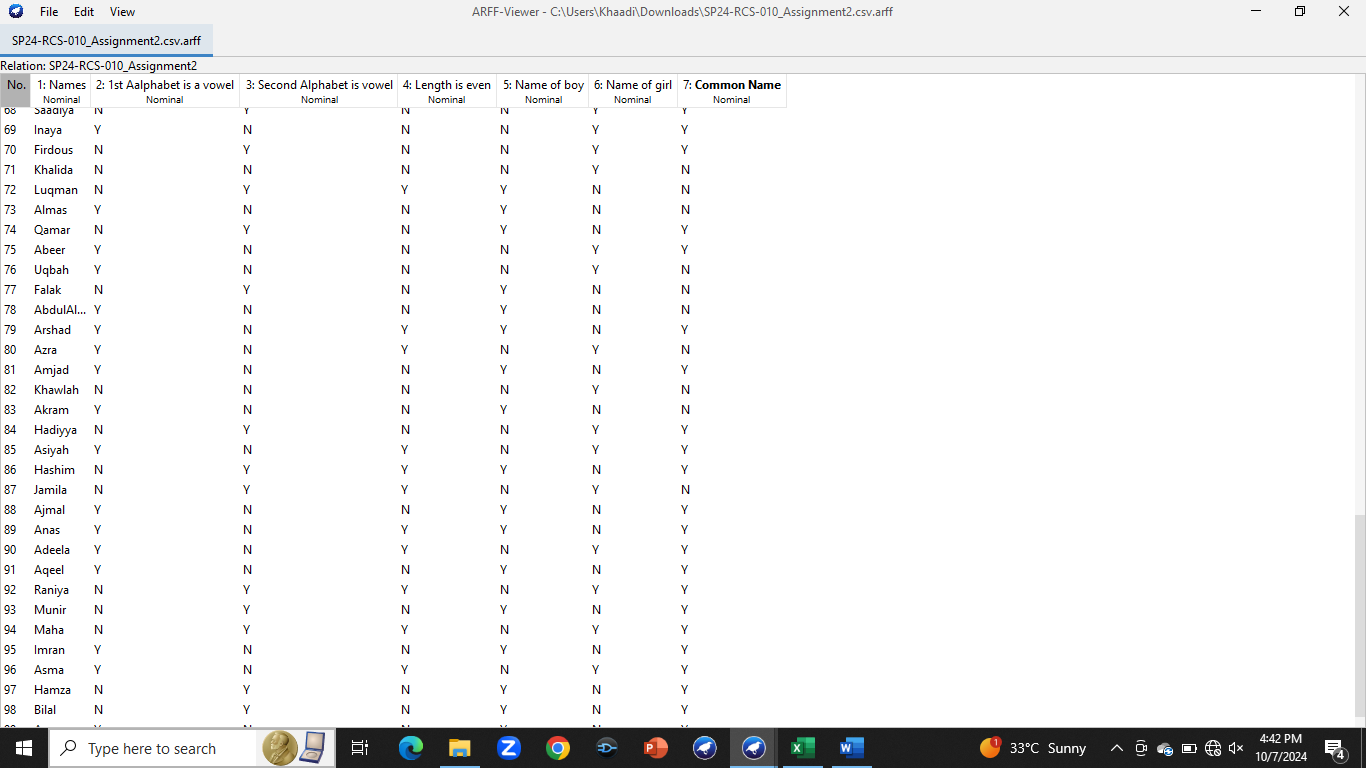
The manually extracted features are as follows:

1. 1st Alphabet is a vowel.
2. Second Alphabet is a vowel.
3. Length is even.
4. Name of boy.
5. Name of girl.
6. Common Name.

After extracting these features, I converted the dataset into ARFF format to ensure compatibility with the WEKA machine-learning environment. The dataset was formatted into ARFF, making it compatible with WEKA, a popular machine-learning tool. I applied different classifiers like the J48 decision tree and the Logistic Model Tree (LMT).

Following are the screenshots of the outputs of the decision tree and their output.

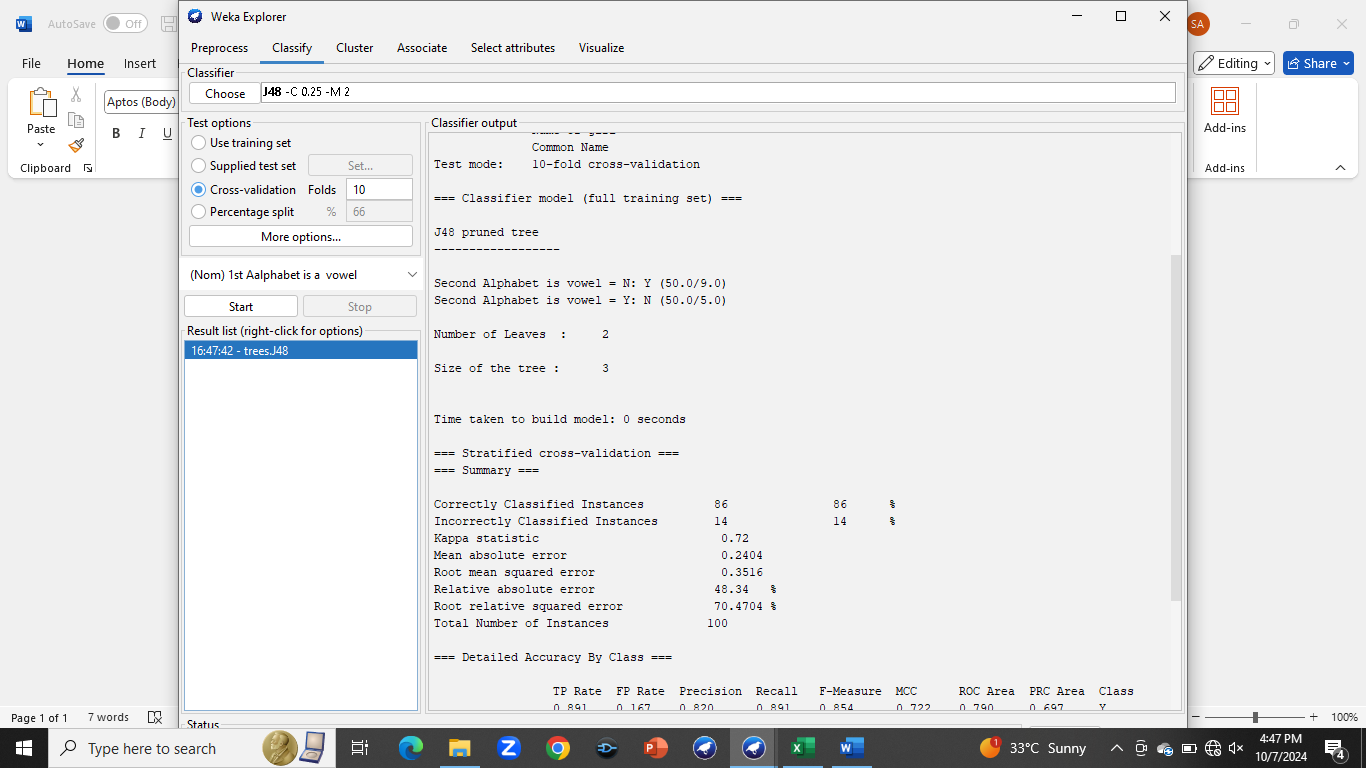
## Converting CSV file into ARFF file format.



## J48 classifier for attribute “1st Alphabet is a vowel”

Correctly classified instances 86 %

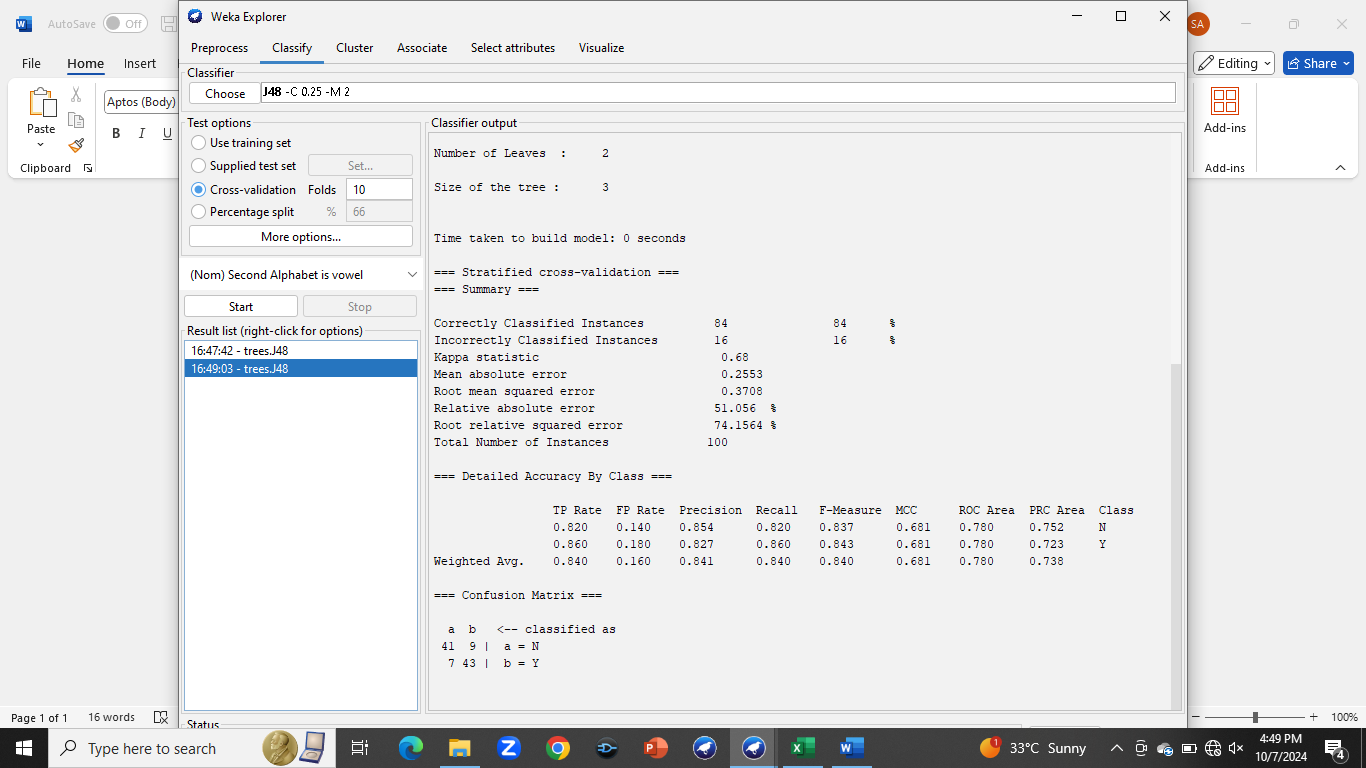
Incorrectly classified instances 14%



## J48 classifier for attribute "2nd Alphabet is a vowel”

Correctly classified instances 84 %

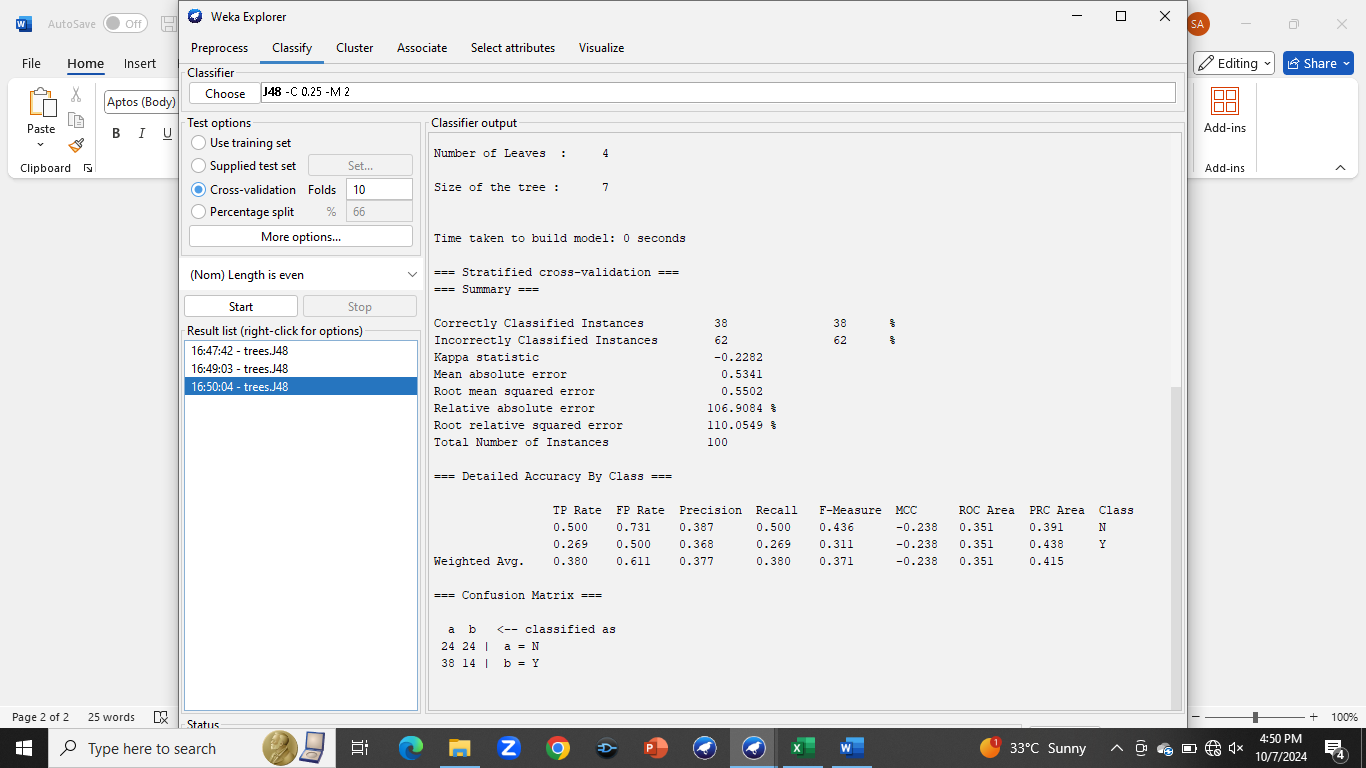
Incorrectly classified instances 16%



## J48 classifier for attribute "Length is even”

Correctly classified instances 38 %

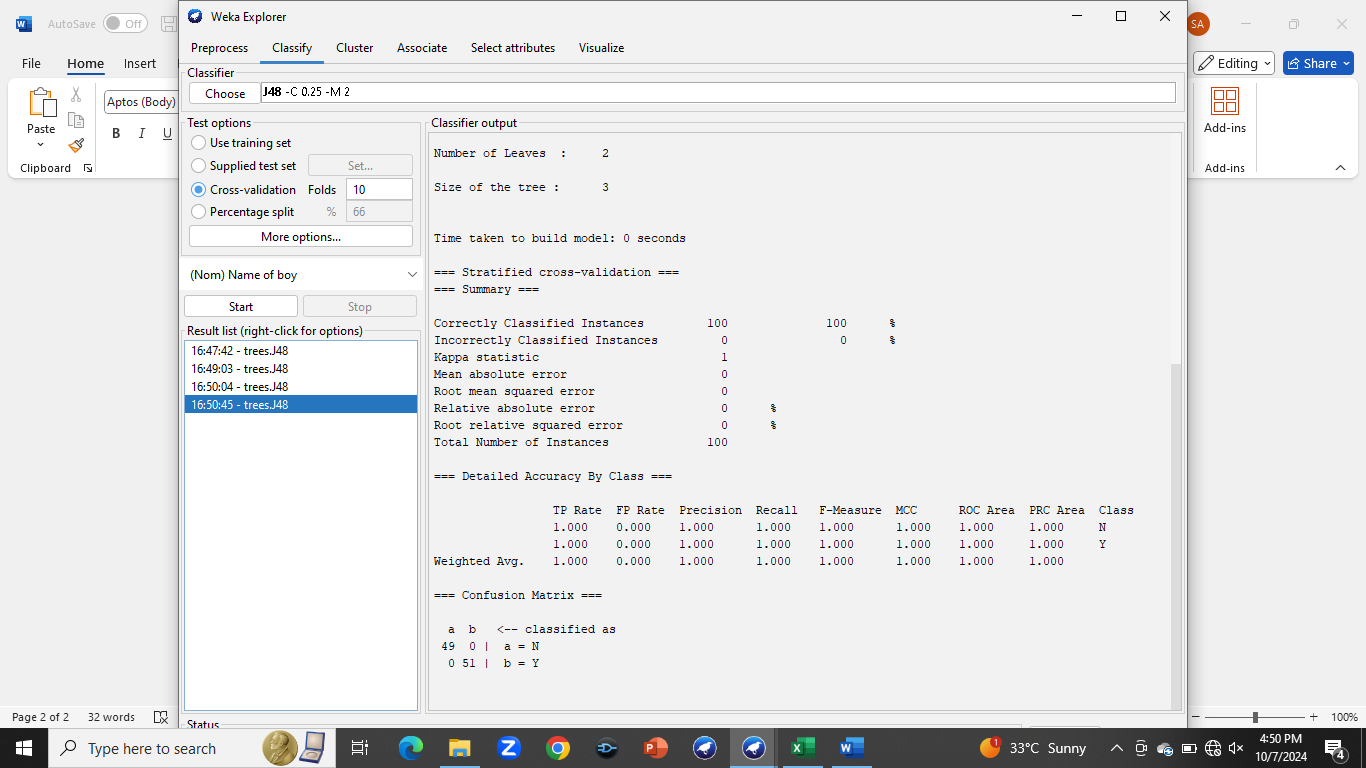
Incorrectly classified instances 62%



## J48 classifier for attribute "Name of a boy”

Correctly classified instances 100 %

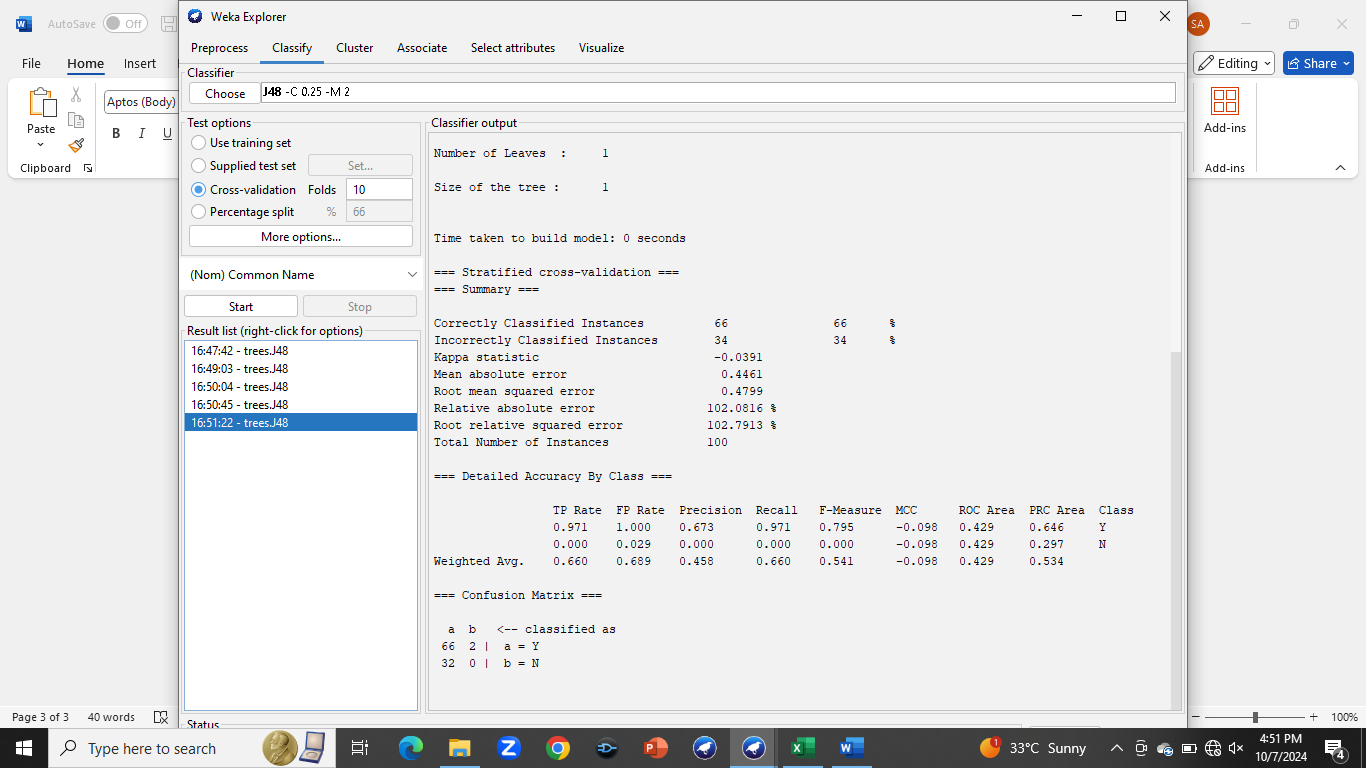
Incorrectly classified instances 0%



## J48 classifier for attribute "Common Name”

Correctly classified instances 66 %

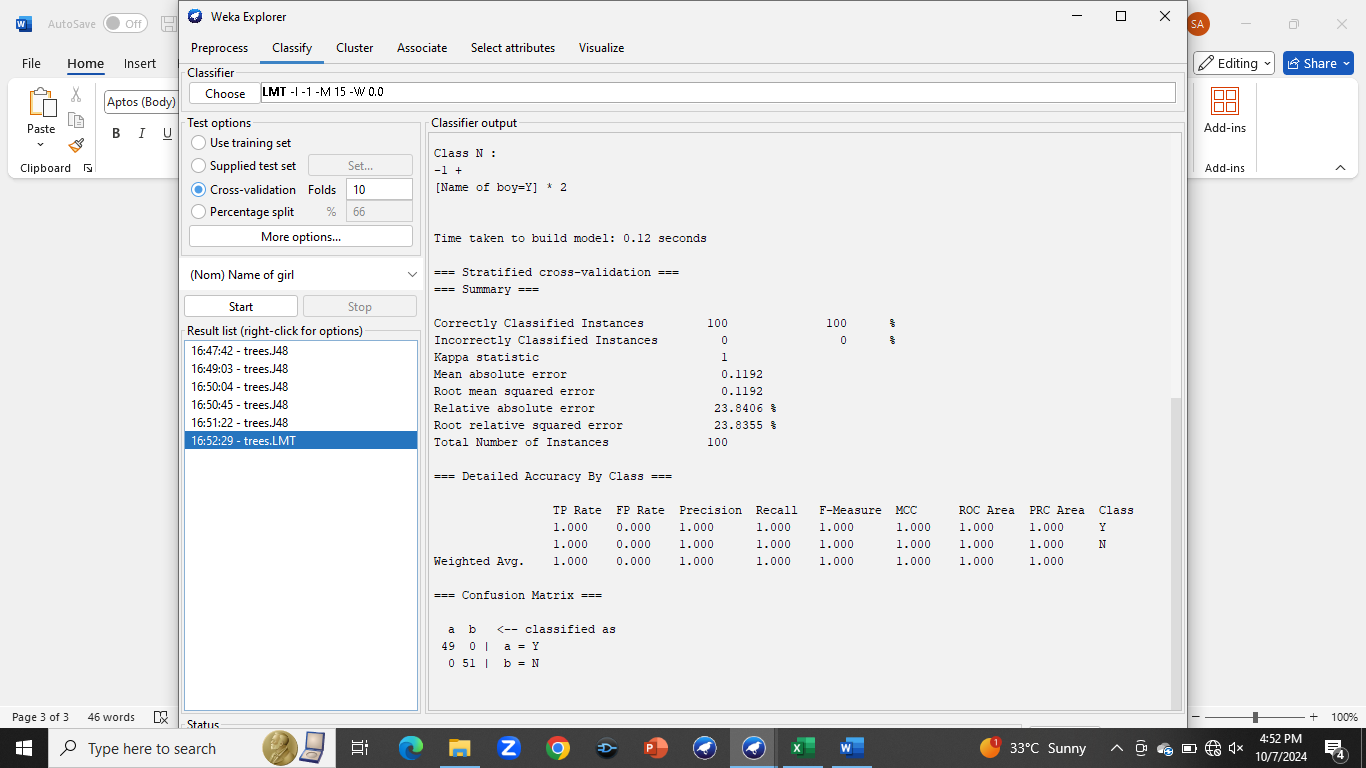
Incorrectly classified instances 34%



## LMT classifier for attribute "Name of Girl”

Correctly classified instances 100 %

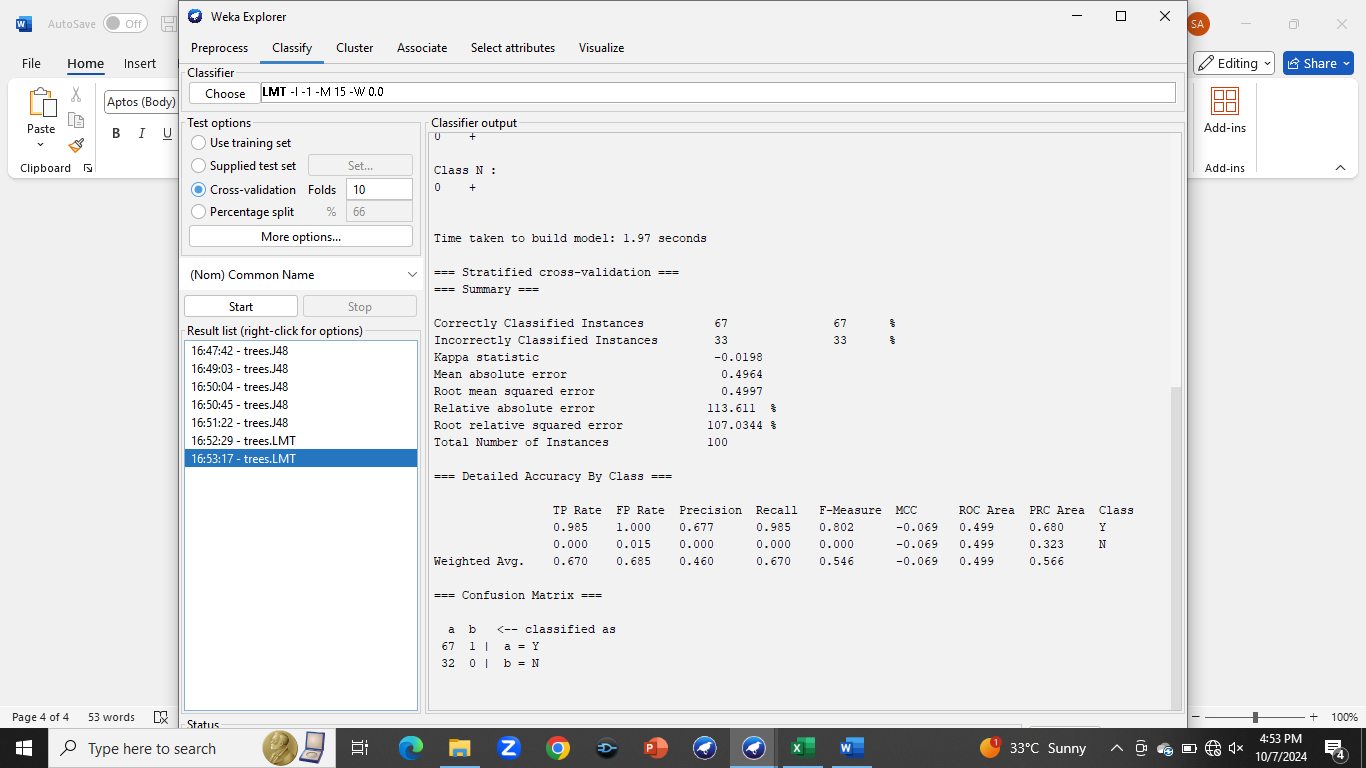
Incorrectly classified instances 0%



## LMT classifier for attribute "Common Name”

Correctly classified instances 67 %

Incorrectly classified instances 33%



## Conclusion:

I also applied the Logistic Model Tree (LMT) classifier to the same dataset, focusing on the attribute of whether the length of the name is even or odd. For the attribute “Name of Girl/Boy”, both J48 and LMT classifiers gave the100% results. For this common feature, J48 correctly classified 66% of the instances, while LMT achieved a slightly higher 67% accuracy on the same dataset. This comparison underscores the differences in how each model handles data variability. The results highlight the sensitivity of decision trees, like J48, to minor data adjustments, which can lead to overfitting. In contrast, LMT demonstrated more consistent performance, suggesting it may be better suited for datasets prone to bias or subtle variations in features. Thus, selecting robust classifiers is crucial for achieving reliable results.