**Course Title:** Machine Learning

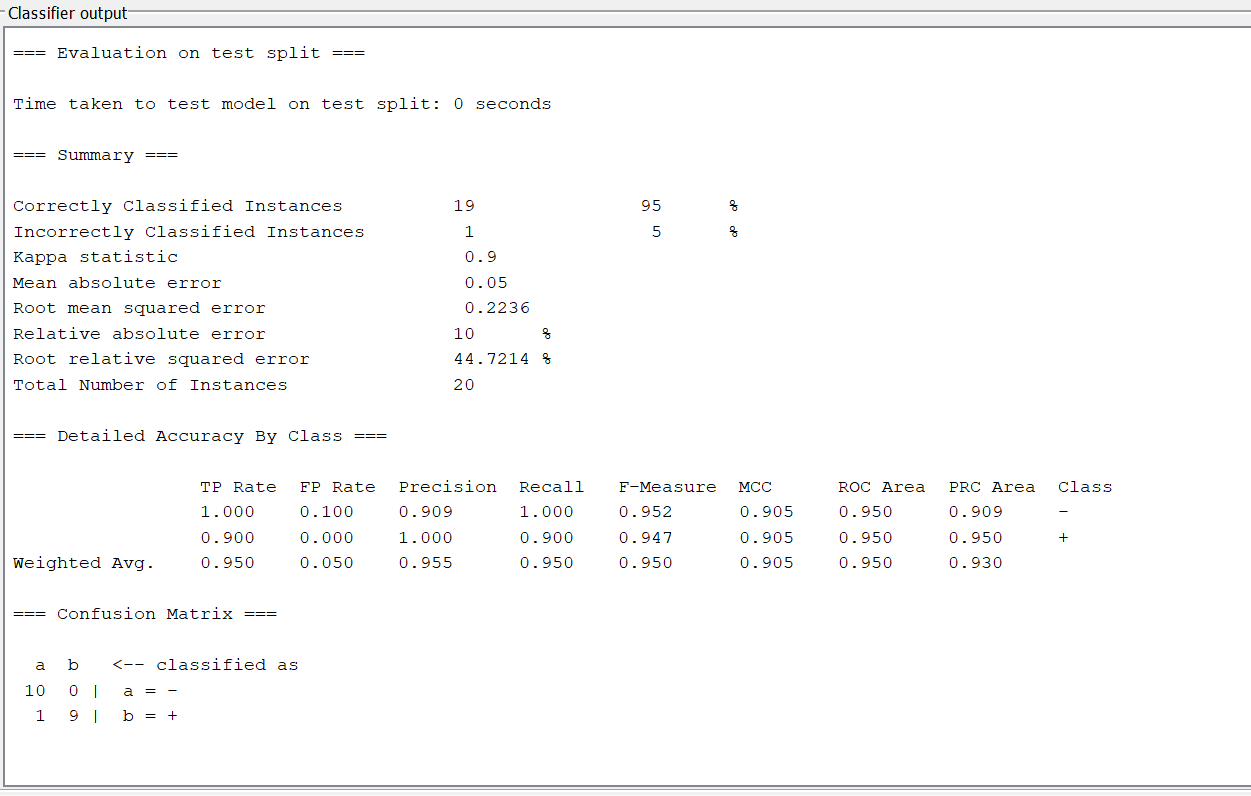
**Assignment** #2

**Course Code:** CSC668 - PCS 716

**Student Name:** Muhammad Naeem

***Report File***

In my experience working with the standard ML pipeline, I learned how important it is to go through each step carefully, from choosing the right features to running the classifier. Using WEKA, I started by manually picking out key features from the dataset, which helped the model perform better. Each feature plays a role in how well the algorithm can separate different classes. When I loaded the dataset into WEKA, it helped me visualize the data and understand how it's spread out. Each feature plays a role in how well the algorithm can separate different classes. I used 80% of the data for training and 20% for testing, which gave me a balanced way to evaluate the model. Running the J48 classification algorithm was easy, and WEKA made it simple to split the data and get evaluation metrics. The results, especially in terms of accuracy and precision showed me that when you select the right features and use the right algorithm, you can achieve good classification outcomes. Overall, I found that the pipeline is a useful way to turn raw data into meaningful results, but it depends a lot on choosing the right features and algorithms.



During this assignment, I added a feature to the dataset that identifies whether the second letter of a name is a vowel ("second word vowel"). This feature helps in distinguishing names based on sound patterns, as names with vowels in certain positions might be more common for specific genders or cultural groups. By adding this feature, I improved the classifier's ability to catch subtle differences that might not be obvious with just basic alphabetic features.

I also included a feature based on the last letter or sound in a name ("last feature"), which can help spot linguistic trends. For example, names that end in certain letters or sounds could be more frequent for one gender over another.

Incorporating these features made me realize how important it is to recognize real-world patterns when building a machine learning model. Customizing features like this not only boosted the classifier’s performance but also deepened my understanding of how even small details can have a significant impact on the outcome of classification tasks.

Tree View:

