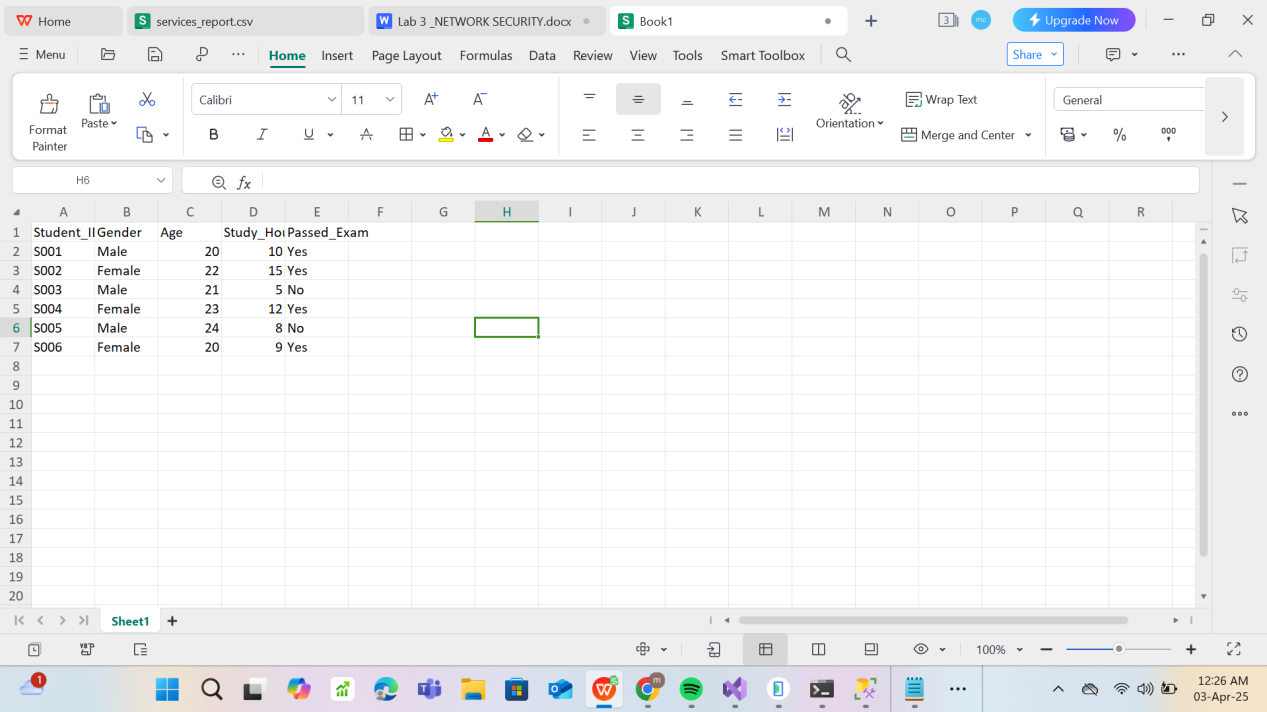
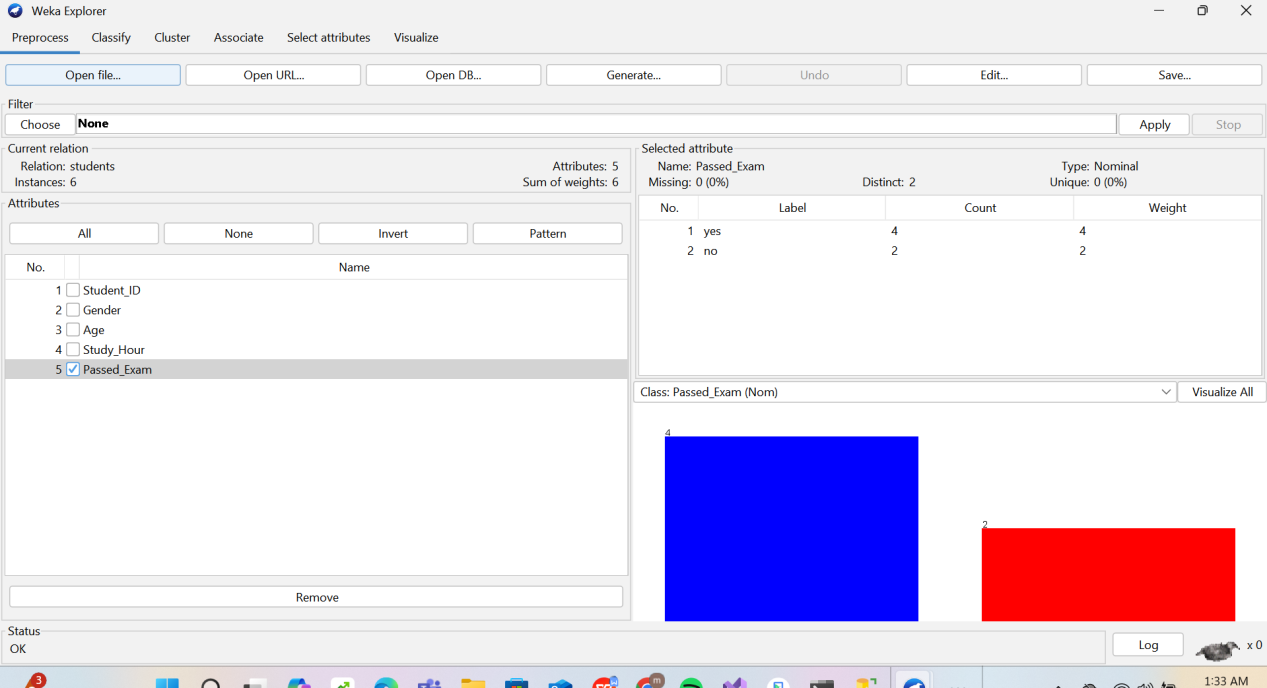
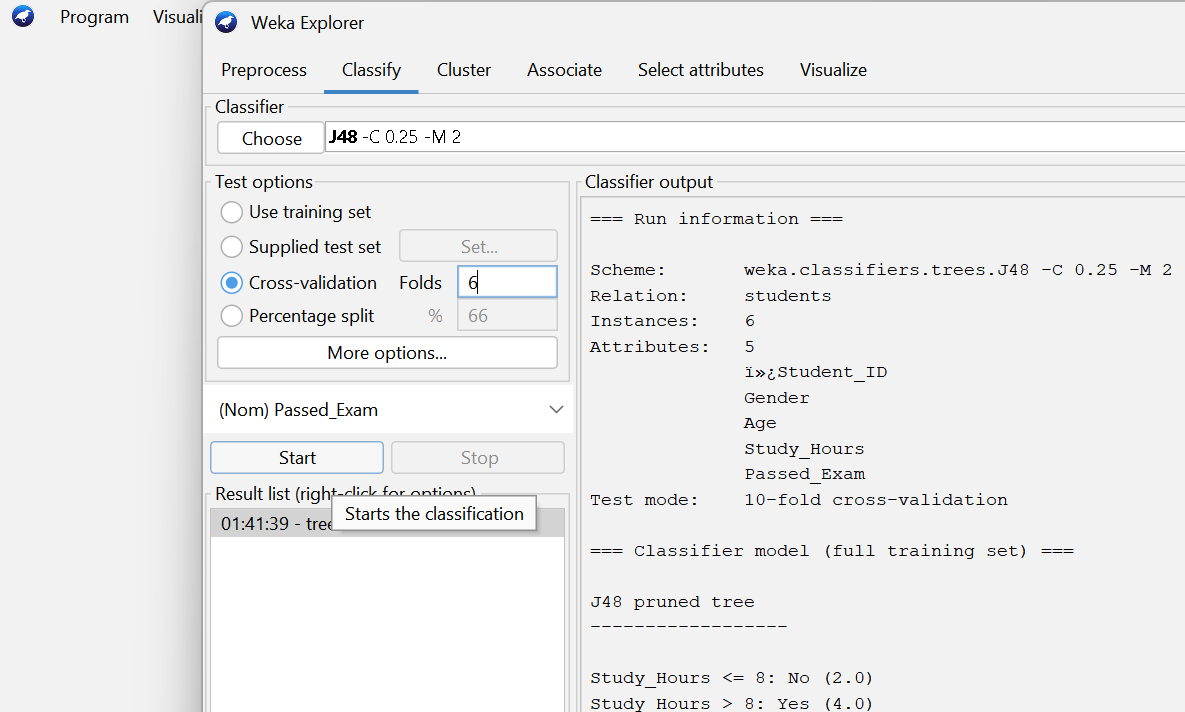
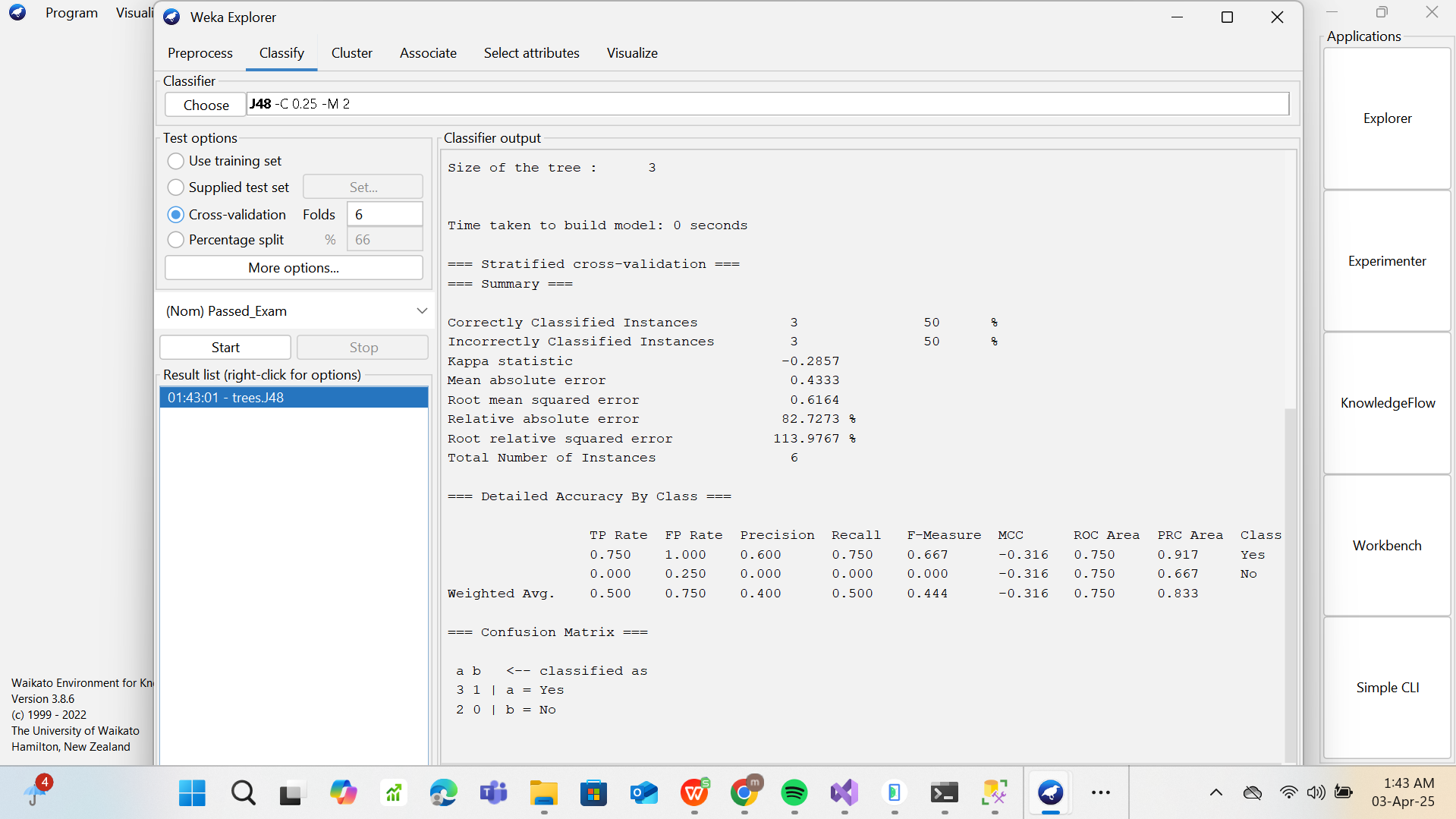
1. Open the "students.xlsx" file in Excel.









1. **What are the key features used in the decision tree to classify whether a student passes the exam or not?**

* Key features are determined by the attributes that most effectively split the data based on the target class (Passed\_Exam).
* Study\_Hour: This is often the strongest predictor. Students who study more hours tend to pass.
* Age: Depending on variation, it might influence the outcome slightly.

**2. Explain the logic behind the decision tree generated by the J48 algorithm based on your dataset.**

**The J48 algorithm (C4.5) builds the tree like this:**

* It looks at all attributes and picks the one that best splits the data (based on information gain).
* It creates a decision node using this attribute.
* It repeats the process for each branch using the remaining attributes.

**3. If you were to add more attributes (e.g., "Study Method"), how do you think it would affect the decision tree?**

**Adding more attributes can:**

* Improve prediction accuracy if the new feature is informative (e.g., "Study Method" like group, solo, video, etc.).
* Change the structure of the tree, because J48 might choose the new attribute for splitting if it has higher information gain.
* Risk overfitting if too many features are added without enough data, especially if the new features are noisy or irrelevant.

**4. What is the significance of the "Age" and "Study\_Hours" attributes in the classification model?**

**Study\_Hour:**

Most likely the primary predictor in your dataset.

Shows a direct correlation more study time higher chance of passing.

**Age:**

Might have weaker correlation, unless certain age groups show different study habits or performance.

Weka might not choose it at all unless it adds value to the split.