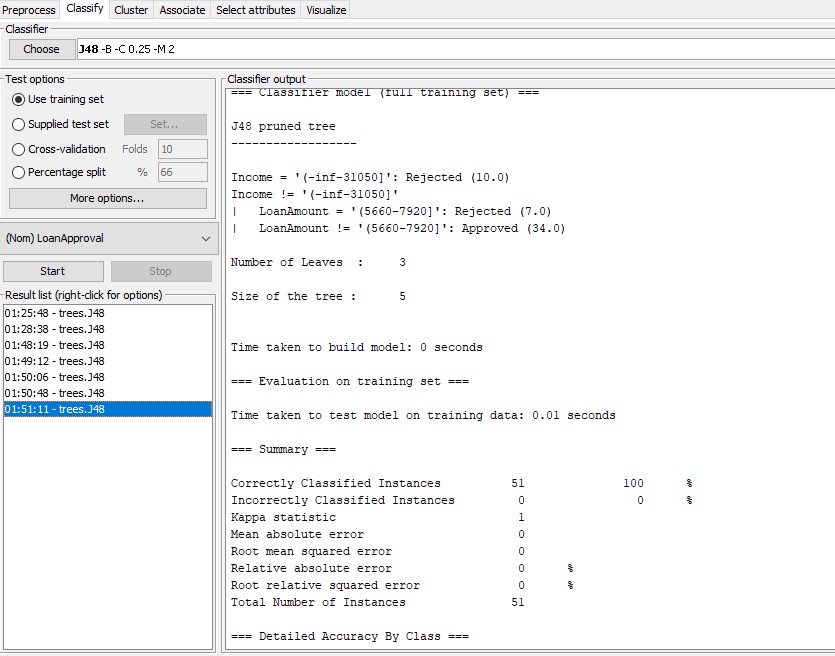
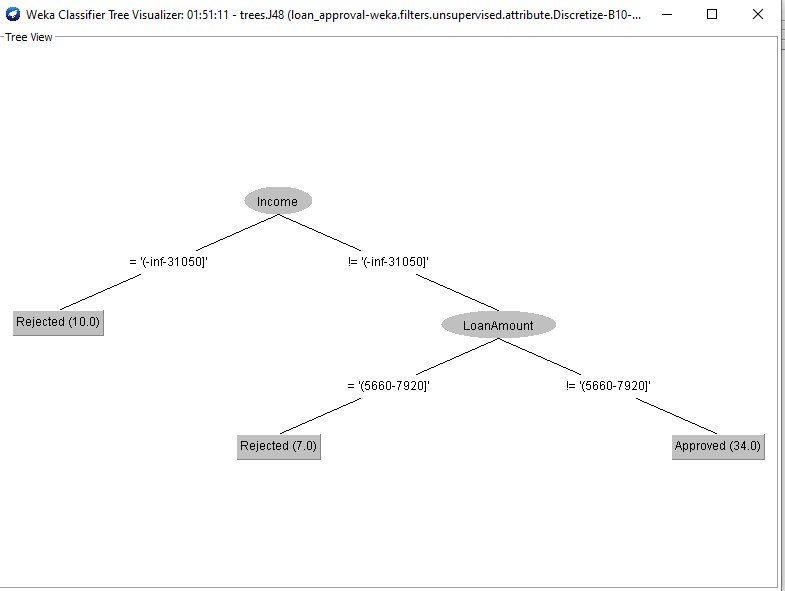
**Bank Loan Approval Analysis**

**Decision Tree Agorithm**

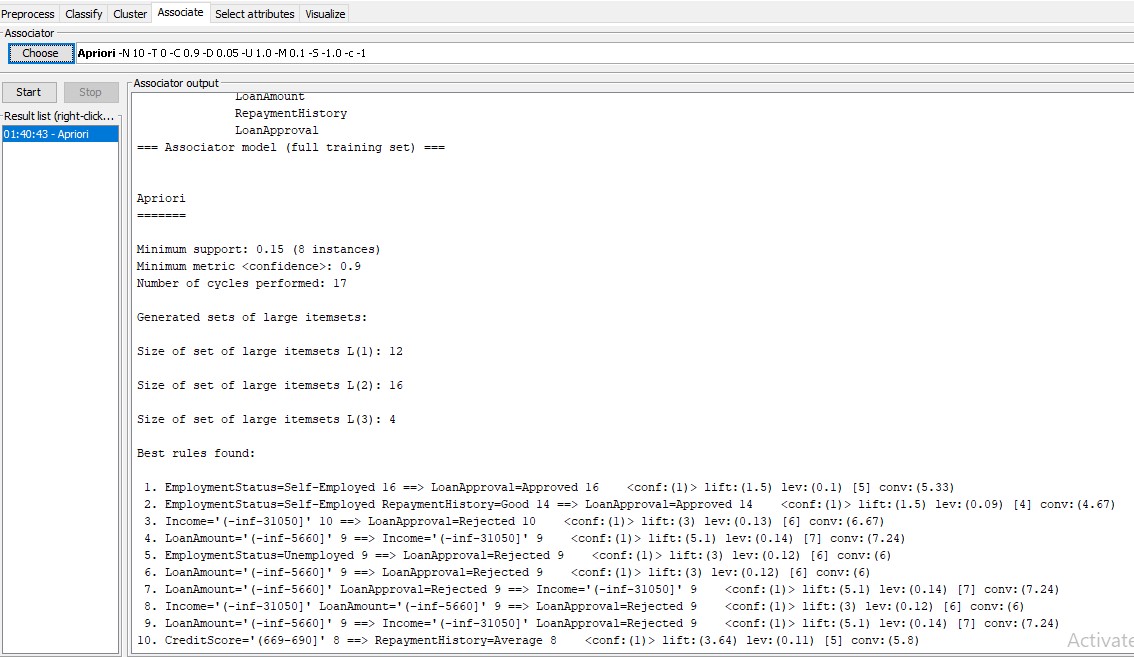
****

****

**Key Findings**

Income is the key factor influencing loan approvals. In our case, if the income amount is equal to 31,050, the loan is rejected. As for a customer whose income is not equal to this amount, the bank will have to first consider their loan amounts. If the loan amounts between 5,660 to 7,920 shillings they are rejected and for those whose amount is not between this range, they are approved of taking a loan.

**Apriori Algorithm**



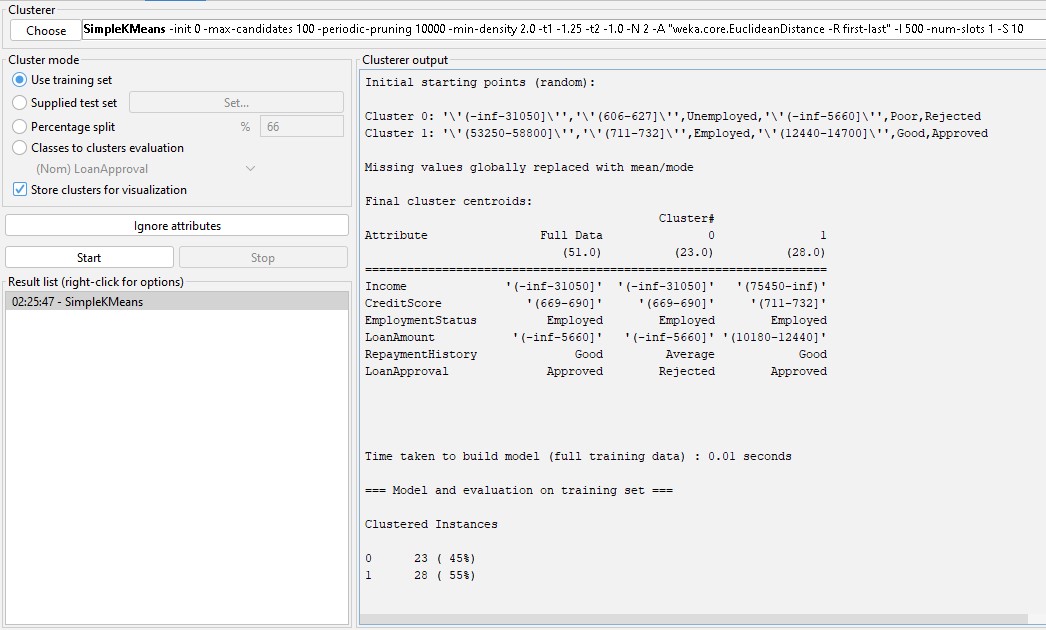
1. **Self-employed applicants with good repayment history have a strong chance of approval.**

B) **Applicants with low income (≤31,050) or who are unemployed are automatically rejected.**

C) **Small loan requests are associated with lower-income applicants, leading to a higher rejection rate.**

**D) Credit score influences repayment behavior but is not directly linked to approval.**

1. ****Means Algorithm****



Weke has grouped the data into two clusters.

**Cluster 0**

* The first cluster contains **low-income individuals** who are **more likely to be rejected**.
* Lower **credit scores** (669-690) are associated with **Cluster 0 (more rejections)**
* Cluster 1 contains **mostly employed individuals**, aligning with more **loan approvals**.
* Cluster 0 mostly applies for **smaller loans** (< 5,660)

**Cluster 1**

* The second cluster consists of **higher-income applicants** who are **mostly approved**
* Higher **credit scores** (711-732) and **good repayment history** appear in **Cluster 1 (more approvals)**
* Cluster 0 has **many unemployed individuals**, leading to **more loan rejections**
* Cluster 1 applies for **higher loan amounts** (10,180-12,440).

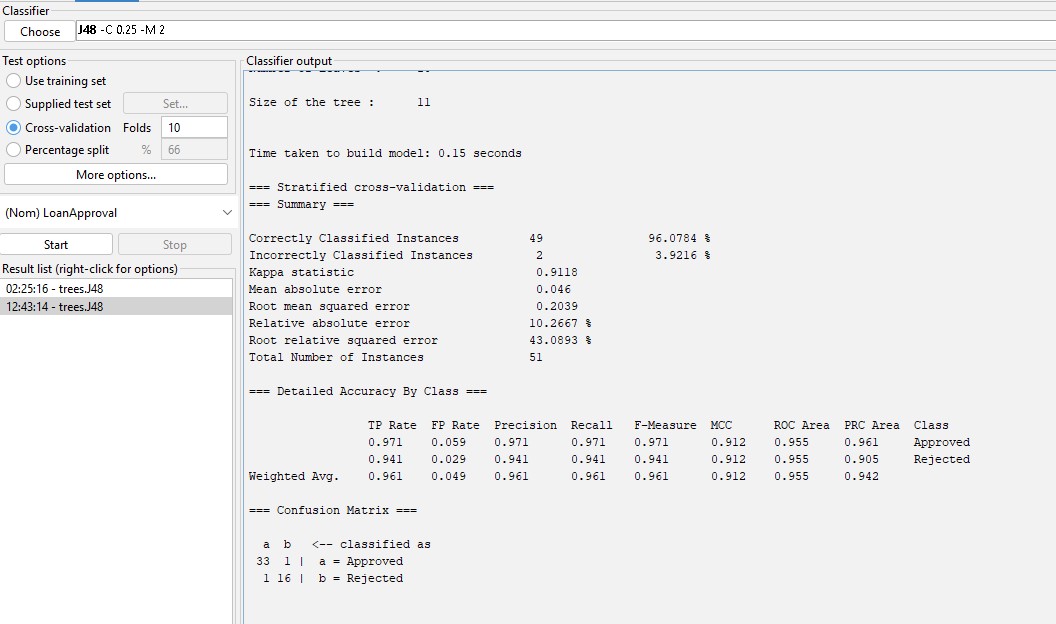
**For Loan Decision Policies -** The bank can **target high-income, good credit score customers for approvals**

**For Risk Management -** Cluster 0 customers pose a **higher risk**—banks should consider **alternative verification steps** or **guarantees.** Cluster 1 represents **more reliable borrowers**—they can be offered **better interest rates.**

****Evaluation and optimization****

**Results obtained from the decision tree above has 100% correctly classified instances which rises concerns of overfitting.**

**To solve this issue I optimized the model by appling cross-validation of 10 folds leading to a more realistic accuracy.**

****

* ****Accuracy** - 96.08%**, meaning the model correctly classified 49 out of 51 instances.
* ****Precision, Recall, and F1-Score:****

For "Approved" class: **Precision = 0.971, Recall = 0.971, F1 = 0.971**

For "Rejected" class: **Precision = 0.941, Recall = 0.941, F1 = 0.941**

The model is still highly effective at predicting approvals and rejections.

* ****Confusion Matrix Analysis:****

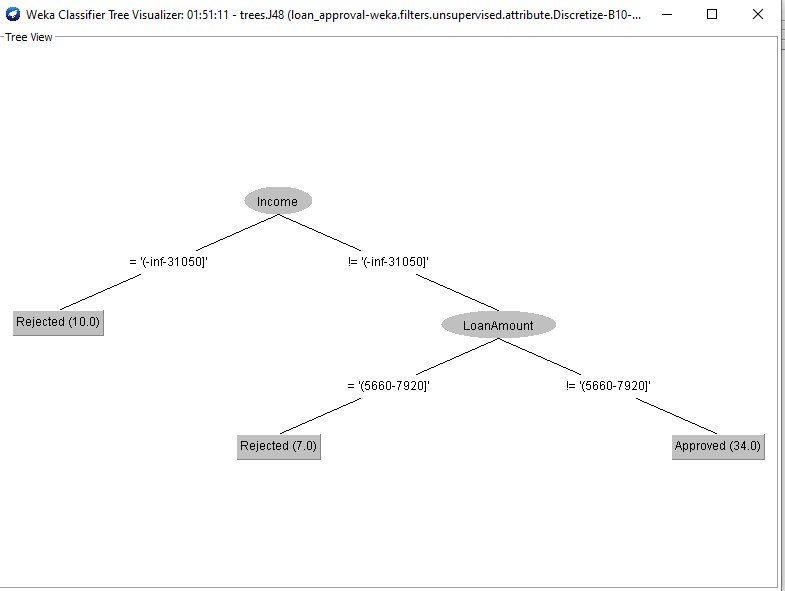
**1 false positive** (an approved loan was incorrectly rejected).

**1 false negative** (a rejected loan was incorrectly approved).

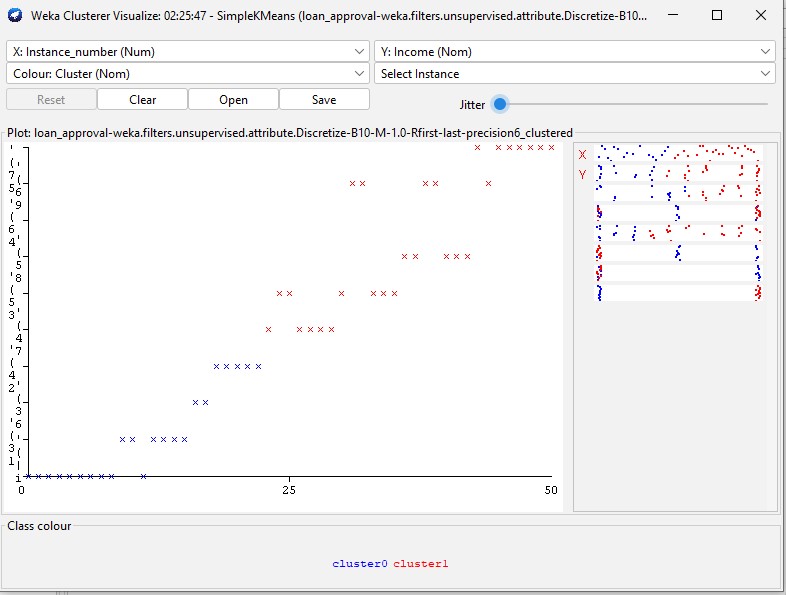
Shows that the model now allows for some misclassification, improving generalizability.

**Visualisation**

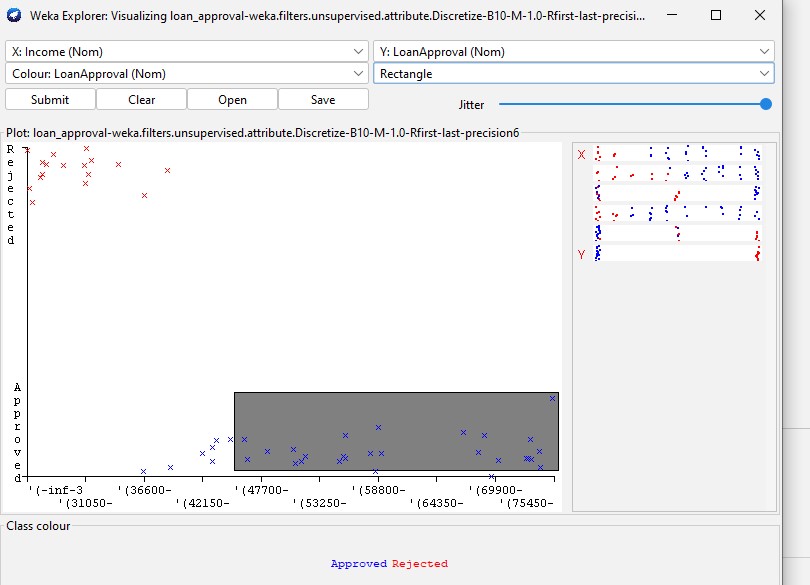
**Decision Tree**

****

****K-Means Visualization****

********

****Apriori Visual****



### ****Risk-Based Loan Approval Policies****

### **High-risk applicants (low-income, poor repayment history)** → Offer **secured loans with stricter terms**.

* **Medium-risk applicants (average repayment history)** → Offer **loans with moderate interest rates**.
* **Low-risk applicants (high income, good credit score)** → Offer **premium loans with better terms**.

### **B. **Tailored Loan Products****

* Create **custom loan products** for different employment types (e.g., flexible loans for self-employed individuals).
* Offer **higher credit limits to customers with strong repayment histories.**

### **C. **Targeted Customer Communication****

* **Send automated loan approval/rejection explanations** based on key factors.
* Provide **personalized financial guidance** to high-risk customers.