

Name: Kshitij Jadhav  
Class: TY-3 Batch: B  
RollNo.: 24  
DWM Expt 8

## Demonstrate Classification, Clustering, Association using weka

### Data Pre-processing and Classification:

The screenshot displays the Weka Explorer application window. The 'Preprocess' tab is active, showing a list of attributes on the left and a summary of the selected attribute 'checking\_status' on the right. The 'checking\_status' attribute is selected, and its summary shows 4 distinct values with counts and weights. A bar chart at the bottom visualizes the distribution of the 'checking\_status' attribute, with bars for each value: 1 (red), 2 (red), 3 (red), and 4 (blue). The status bar at the bottom indicates 'OK'.

Weka Explorer

Preprocess | Classify | Cluster | Associate | Select attributes | Visualize

Open file... | Open URL... | Open DB... | Generate... | Undo | Edit... | Save...

Filter: Choose **NumericToBinary** -R first-last | Apply | Stop

Current relation: german\_credit-weka.filters.unsupervised.attribute.NumericToBinary-Rfirst-last  
Instances: 1000 | Attributes: 21 | Sum of weights: 1000

Attributes: All | None | Invert | Pattern

No.	Name
1	<input checked="" type="checkbox"/> checking_status
2	<input type="checkbox"/> duration_binarized
3	<input type="checkbox"/> credit_history
4	<input type="checkbox"/> purpose
5	<input type="checkbox"/> credit_amount_binarized
6	<input type="checkbox"/> savings_status
7	<input type="checkbox"/> employment
8	<input type="checkbox"/> installment_commitment_binarized
9	<input type="checkbox"/> personal_status
10	<input type="checkbox"/> other_parties
11	<input type="checkbox"/> residence_since_binarized
12	<input type="checkbox"/> property_magnitude
13	<input type="checkbox"/> age_binarized
14	<input type="checkbox"/> other_payment_plans
15	<input type="checkbox"/> housing
16	<input type="checkbox"/> existing_credits_binarized
17	<input type="checkbox"/> job
18	<input type="checkbox"/> num_dependents_binarized
19	<input type="checkbox"/> own_telephone
20	<input type="checkbox"/> foreign_worker
21	<input type="checkbox"/> class

Remove

Selected attribute: Name: checking\_status  
Missing: 0 (0%) | Distinct: 4 | Type: Nominal | Unique: 0 (0%)

No.	Label	Count	Weight
1	<0	274	274
2	0<=X<200	269	269
3	>=200	63	63
4	no checking	394	394

Class: class (Nom) | Visualize All

Status: OK | Log | x 0

**Weka Explorer**

Preprocess   **Classify**   Cluster   Associate   Select attributes   Visualize

Classifier: Choose **J48 - C 0.25 - M 2**

Test options:  
☐ Use training set  
☐ Supplied test set  
☒ Cross-validation   Folds: **10**  
☐ Percentage split   %: **66**  
 More options...

(Nom) class: **Start**   **Stop**

Result list (right-click for options):  
 22:49:50 - trees.J48

Classifier output:

```

| | credit_history = critical/other existing credit: good (5.0/1.0)
| | savings_status = 500<=<1000: good (11.0/3.0)
| | savings_status = >=1000: good (14.0/4.0)
| | savings_status = no known savings: good (45.0/7.0)
| checking_status = >=200: good (63.0/14.0)
| checking_status = no checking: good (394.0/46.0)

```

Number of Leaves : 62  
 Size of the tree : 81  
 Time taken to build model: 0.08 seconds

==== Stratified cross-validation ====

==== Summary ====

	Correctly Classified Instances	Incorrectly Classified Instances	Kappa statistic	Mean absolute error	Root mean squared error	Relative absolute error	Root relative squared error	Total Number of Instances
	700	300	0.2485	0.354	0.4584	84.2521 %	100.025 %	1000

==== Detailed Accuracy By Class ====

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
Weighted Avg.	0.823	0.587	0.766	0.823	0.793	0.251	0.680	0.784	good
	0.413	0.177	0.500	0.413	0.453	0.251	0.680	0.449	bad

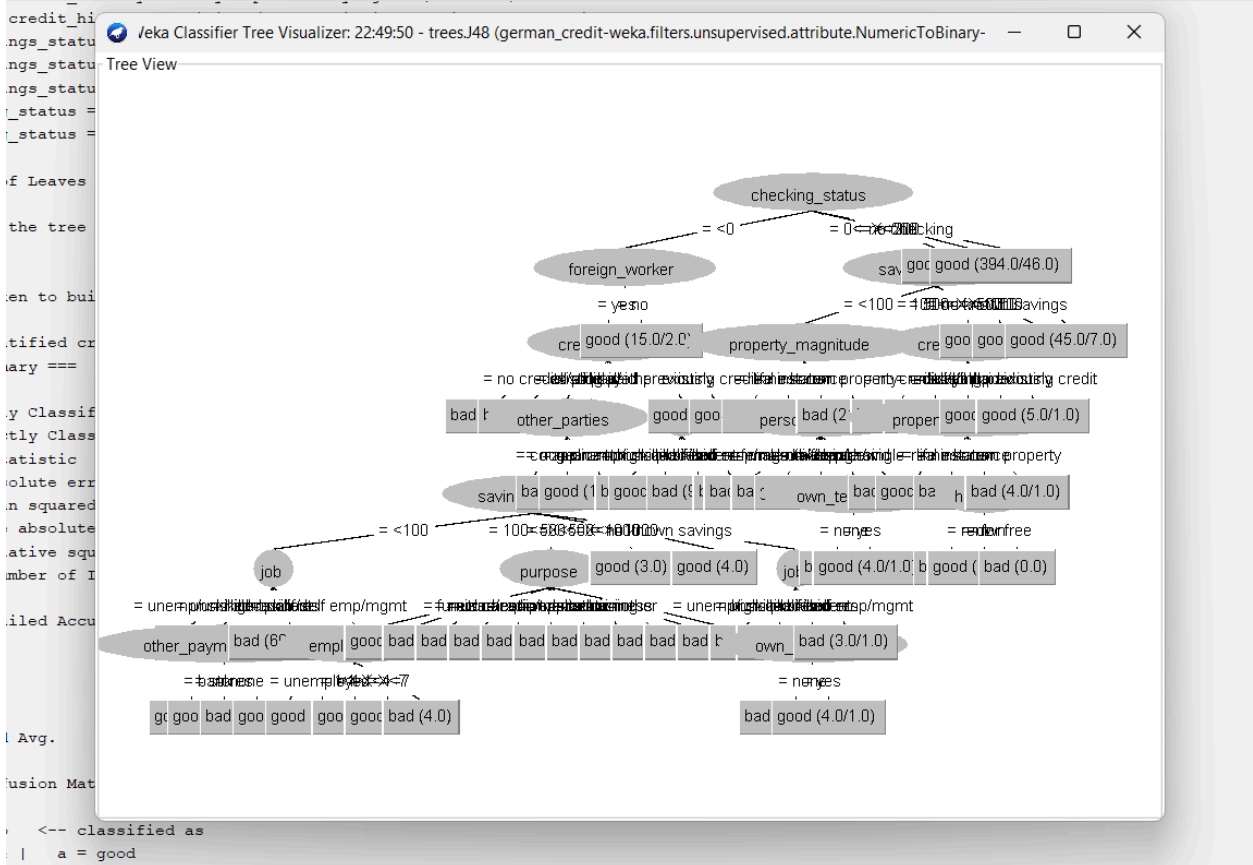
==== Confusion Matrix ====

```

a b <-- classified as
576 124 | a = good
176 124 | b = bad

```

Status: OK



# Clustering:

The screenshot shows the Weka Explorer interface with the SimpleKMeans algorithm selected. The 'Clusterer' tab is active, and the 'Cluster mode' is set to 'Use training set'. The 'Result list' on the left shows '22:52:04 - SimpleKMeans'. The 'Cluster output' pane displays the following information:

Final cluster centroids:

Attribute	Full Data (1000.0)	Cluster# 0 (725.0)	Cluster# 1 (275.0)
checking_status	no checking	no checking	<0
duration_binarized	1	1	1
credit_history	existing paid	existing paid	existing paid
purpose	radio/tv	radio/tv	new car
credit_amount_binarized	1	1	1
savings_status	<100	<100	<100
employment	1<=X<4	1<=X<4	>=7
installment_commitment_binarized	1	1	1
personal_status	male single	male single	male single
other_parties	none	none	none
residence_since_binarized	1	1	1
property_magnitude	car	real estate	no known property
age_binarized	1	1	1
other_payment_plans	none	none	none
housing	own	own	own
existing_credits_binarized	1	1	1
job	skilled	skilled	skilled
num_dependents_binarized	1	1	1
own_telephone	none	none	yes
foreign_worker	yes	yes	yes
class	good	good	good

Time taken to build model (full training data) : 0.06 seconds

=== Model and evaluation on training set ===

Clustered Instances

Cluster	Instances
0	725 ( 73%)
1	275 ( 28%)

# Association Algorithm (Apriori):

The screenshot shows the Weka Explorer interface with the Apriori algorithm selected. The 'Associate' tab is active, and the 'Associator' is set to 'Apriori'. The 'Result list' on the left shows '22:49:28 - Apriori'. The 'Associator output' pane displays the following information:

Apriori

Minimum support: 0.95 (950 instances)  
Minimum metric <confidence>: 0.9  
Number of cycles performed: 1

Generated sets of large itemsets:

Size of set of large itemsets L(1): 8  
Size of set of large itemsets L(2): 28  
Size of set of large itemsets L(3): 56  
Size of set of large itemsets L(4): 70  
Size of set of large itemsets L(5): 56  
Size of set of large itemsets L(6): 28  
Size of set of large itemsets L(7): 8  
Size of set of large itemsets L(8): 1

Best rules found:

- credit\_amount\_binarized=1 1000 ==> duration\_binarized=1 1000 <conf: (1)> lift: (1) lev: (0) [0] conv: (0)
- duration\_binarized=1 1000 ==> credit\_amount\_binarized=1 1000 <conf: (1)> lift: (1) lev: (0) [0] conv: (0)
- installment\_commitment\_binarized=1 1000 ==> duration\_binarized=1 1000 <conf: (1)> lift: (1) lev: (0) [0] conv: (0)
- duration\_binarized=1 1000 ==> installment\_commitment\_binarized=1 1000 <conf: (1)> lift: (1) lev: (0) [0] conv: (0)
- residence\_since\_binarized=1 1000 ==> duration\_binarized=1 1000 <conf: (1)> lift: (1) lev: (0) [0] conv: (0)
- duration\_binarized=1 1000 ==> residence\_since\_binarized=1 1000 <conf: (1)> lift: (1) lev: (0) [0] conv: (0)
- age\_binarized=1 1000 ==> duration\_binarized=1 1000 <conf: (1)> lift: (1) lev: (0) [0] conv: (0)
- duration\_binarized=1 1000 ==> age\_binarized=1 1000 <conf: (1)> lift: (1) lev: (0) [0] conv: (0)
- existing\_credits\_binarized=1 1000 ==> duration\_binarized=1 1000 <conf: (1)> lift: (1) lev: (0) [0] conv: (0)
- duration\_binarized=1 1000 ==> existing\_credits\_binarized=1 1000 <conf: (1)> lift: (1) lev: (0) [0] conv: (0)