

SUBJECT: CA-605-MJP
Lab Course on CA-602-MJ (Machine Learning)

1. Using any open source software such as WEKA and its datasets, perform classification using Naïve Bayes classifier, note accuracy. Describe the output of confusion matrix & state the formula required for calculation of accuracy.
2. Using any open source software such as WEKA and its datasets, perform classification using Neural network classifier. Using any available attribute selection algorithm in WEKA note the accuracy and compare with it. Elaborate the working of NN classifier.
3. Using any open source software such as WEKA and its datasets, perform clustering using 'EM' algorithm. State about the generation of confusion matrix & its accuracy.
4. Using any open source software such as WEKA and its datasets, perform classification using Neural network classifier. Using any available attribute selection algorithm in WEKA note the accuracy and compare with it. Elaborate the working of NN classifier.
5. Implement K-Means clustering/ hierarchical clustering on sales_data_sample.csv dataset. Determine the number of clusters using the elbow method. Elaborate the output with respective clustering method.
Dataset link : <https://www.kaggle.com/datasets/kyanyoga/sample-sales-data>
6. Using any open source software such as WEKA and its datasets, perform clustering using 'EM' algorithm. State about the generation of confusion matrix & its accuracy.
7. Using any open source software such as WEKA and its datasets, perform classification using Neural network classifier. Using any available attribute selection algorithm in WEKA note the accuracy and compare with it. Elaborate the working of NN classifier.
8. Using any open source software such as WEKA and its datasets, perform classification using C4.5 – the decision tree classifier. Describe the output of confusion matrix & state the formula required for calculation of accuracy.
9. Using any open source software such as WEKA and its datasets, perform classification using Naïve Bayes classifier, note accuracy. State about the generation of confusion matrix & its accuracy.
10. Using any open source software such as WEKA and its datasets, perform classification using Naïve Bayes classifier, note accuracy. Using any available attribute selection algorithm in WEKA and note the accuracy and compare with it. Describe the output of confusion matrix & state the formula required for calculation of accuracy.
11. Using any open source software such as WEKA and its datasets, perform classification using Neural network classifier. Describe the output of confusion matrix & state the formula required for calculation of accuracy.
12. Implement K-Means clustering/ hierarchical clustering on sales_data_sample.csv dataset. Determine the number of clusters using the elbow method. Elaborate the output with respective clustering method.
Dataset link : <https://www.kaggle.com/datasets/kyanyoga/sample-sales-data>
13. Using any open source software such as WEKA and its datasets, perform classification using Naïve Bayes classifier, note accuracy. Using any available attribute selection algorithm in WEKA and note the accuracy and compare with it. Describe the output of confusion matrix & state the formula required for calculation of accuracy.
14. Using any open source software such as WEKA and its datasets, perform classification using C4.5 – the decision tree classifier. Using both attribute and instance selection

algorithm in WEKA and note the accuracy and compare with it. Elaborate the working of Decision tree classifier.

15. Using any open source software such as WEKA and its datasets, perform classification using Neural network classifier. Elaborate the working of NN classifier. State about the generation of confusion matrix & its accuracy.
16. Using any open source software such as WEKA and its datasets, perform classification using C4.5 – the decision tree classifier. Using any available attribute selection algorithm in WEKA and note the accuracy and compare with it. Describe the output of confusion matrix & state the formula required for calculation of accuracy.
17. Using any open source software such as WEKA and its datasets, perform classification using Neural network classifier. Using any available attribute selection algorithm in WEKA note the accuracy and compare with it. Elaborate the working of NN classifier.
18. Using any open source software such as WEKA and its datasets, perform classification using Naïve Bayes classifier, note accuracy. Using any available attribute selection algorithm in WEKA and note the accuracy and compare with it. Describe the output of confusion matrix & state the formula required for calculation of accuracy.
19. Find the missing values from “*weather.arff*” dataset and replace that values using WEKA. Analyze how it will affect the dataset.
20. Scale all features between 0 and 1 of “*iris.arff*” dataset using WEKA. Compare the statistics of one feature before and after normalization.
21. Using any open-source software such as WEKA and its datasets, perform classification using Naïve Bayes classifier, note accuracy.
22. Apply the J48 algorithm on “*iris.arff*” dataset and classify the data using WEKA. Report the accuracy and visualize the resulting decision tree.
23. Use the Naive Bayes classifier on *weather.nominal.arff* dataset and classify the data using WEKA. Note the accuracy of the model.
24. Apply the KNN classifier with $k=3$ on “*iris.arff*” dataset. Compare the performance with $k=5$ and explain which value of k performs better and why using WEKA.
25. Use the J48 classifier and perform 10-fold cross-validation on the “*cancer.arff*” Dataset using WEKA. Note the accuracy, and analyze how does it compare to training on the full dataset.
26. Train a Random Forest classifier and J48 classifier on the “*iris.arff*” dataset using WEKA. Compare performance of the both the classifier.
27. Apply k-Means clustering with $k=3$ on *iris.arff* dataset using WEKA. Find out centroids of the clusters.
28. Apply the hierarchical clustering algorithm on *iris.arff* dataset using WEKA. Interpret the pattern and formation of the cluster using dendrogram.
29. Identify the top 3 most significant attributes from “*Cancer.arff*” dataset using WEKA. Find which attributes were selected, and why they are important?
30. Using any open-source software such as WEKA and its datasets, perform classification using C4.5 the decision tree classifier.
31. Using any open-source software such as WEKA and its datasets, perform classification using C4.5 the decision tree classifier.
32. Apply the Naive Bayes classifier on *cancer.arff* dataset and generate the confusion Matrix using WEKA. Interpret the precision and recall.

33. Apply the J48 classifier and generate the ROC curve on *weather.nominal.arff* dataset using WEKA. Analyze ROC curve indicate about the model's performance in terms of sensitivity and specificity?
34. Using WEKA compare the accuracy and ROC area of the J48, Naive Bayes, and Random Forest classifiers when applied to the *iris.arff* dataset. Analyze which classifier performs best.
35. Create a scatter plot for the *iris.arff* dataset using WEKA's visualization tool. Analyze patterns observed in the plot, and the classes distributed.
36. Train a model using the J48 algorithm on the *cancer.arff* dataset using WEKA, save the model to disk. Analyze how this model can be reloaded in WEKA for future predictions.
37. Train a Random Forest classifier and J48 classifier on the "iris.arff" dataset using WEKA. Compare performance of the both the classifier.
38. Using any open-source software such as WEKA and its datasets, perform regression using logistic regression, note accuracy.
39. Implement K-Means clustering/ hierarchical clustering (Use Credit Card Dataset GENERAL.csv Download from kaggle.com). Determine the number of clusters using the elbow method.
40. Using any open-source software such as WEKA and its datasets, perform regression using linear regression, note accuracy.(use airline dataset).
41. Use the Naive Bayes classifier on *weather.nominal.arff* dataset and classify the Data using WEKA. Note the accuracy of the model.
42. Perform classification using KNN on heart disease dataset.
43. Using any open-source software such as WEKA and its datasets, perform classification using Random Forest classifier.
44. Using any open-source software such as WEKA and its datasets, perform classification using AdaBoost classifier.
45. Perform classification using SVM on heart disease dataset.
46. Using any open-source software such as WEKA and its datasets, perform clustering using EM clustering, note accuracy.
47. Evaluate the impact of instance selection on the accuracy of classification algorithms using WEKA.(use Iris Dataset)
48. Implement a decision tree classifier using WEKA and evaluate its performance on a sample dataset.
49. Using any open-source software such as WEKA and its datasets, perform clustering using Hierarchical clustering, note accuracy.
50. Using any open-source software such as WEKA and its datasets, perform clustering using Density-Based clustering, note accuracy.
51. Develop a classification model using a combination of attribute and instance selection algorithms in WEKA.
52. Implement K-Means clustering/ hierarchical clustering on *sales_data_sample.csv* dataset. Determine the number of clusters using the elbow method.
Dataset link : <https://www.kaggle.com/datasets/kyanyoga/sample-sales-data>
53. Investigate the effect of attribute selection on the accuracy of classification algorithms using WEKA.(use Adult Data set).
54. Compare the accuracy of Naïve Bayes, C4.5, and Neural Network classifiers on a common dataset using WEKA.

55. Implement hierarchical clustering on sales_data_sample.csv dataset. Determine the number of clusters using the elbow method.
56. Using any open-source software such as WEKA and its datasets, perform classification using C4.5 the decision tree classifier.
57. Use the J48 classifier and perform 10-fold cross-validation on the “cancer.arff” Dataset using WEKA. Note the accuracy, and analyze how does it compare to training on the full dataset.
58. Use the J48 classifier and perform 10-fold cross-validation on the “cancer.arff” Dataset using WEKA. Note the accuracy, and analyze how does it compare to training on the full dataset.
59. Using any open-source software such as WEKA and its datasets, perform classification using C4.5 the decision tree classifier.
60. Using any open-source software such as WEKA and its datasets, perform classification using SVM Classifier.
61. Apply the J48 algorithm on “iris.arff” dataset and classify the data using WEKA. Report the accuracy and visualize the resulting decision tree.
62. Use the Naive Bayes classifier on weather.nominal.arff dataset and classify the data using WEKA. Note the accuracy of the model.
63. Compare the accuracy of Naïve Bayes, C4.5, and Neural Network classifiers on a common dataset using WEKA.
64. Apply the KNN classifier with $k=3$ on “iris.arff” dataset. Compare the performance with $k=5$ and explain which value of k performs better and why using WEKA.
65. Using open source software such as WEKA and its datasets, perform classification using C4.5 – the decision tree classifier.
66. Use the J48 classifier and perform 10-fold cross-validation on the “cancer.arff”. Dataset using WEKA. Note the accuracy, and analyze how does it compare to training on the full dataset.
67. Use the Naive Bayes classifier on weather.nominal.arff dataset and classify the data using WEKA. Note the accuracy of the model.
68. Train a Random Forest classifier and J48 classifier on the “iris.arff” dataset using WEKA. Compare performance of the both the classifier.
69. Using any open-source software such as WEKA and its datasets, perform classification using SVM Classifier.
70. Train a Random Forest classifier and J48 classifier on the “iris.arff” dataset using WEKA. Compare performance of the both the classifier.
71. Apply k-Means clustering with $k=3$ on iris.arff dataset using WEKA. Find out centroids of the clusters.
72. Compare the accuracy of Naïve Bayes, C4.5, and Neural Network classifiers on a common dataset Using WEKA.
73. Using any open-source software such as WEKA and its datasets, perform regression using linear regression, note accuracy.(use airline dataset).
74. Using any open-source software such as WEKA and its datasets, perform clustering using Density-Based clustering, note accuracy.
75. Implement a decision tree classifier using WEKA and evaluate its performance on a sample dataset.