Journal of Trends in Computer Science and Smart Technology (ISSN: 2582-4104)
www.irojournals.com/tcsst/

Analysis of Classification Algorithms in Drug Classification Using Weka Data Mining Tool

B. Deepthi[1], K. V. Siva Prasad Reddy[2], B. S. Jubedha[3]

[1,3]Student, CSE Department, JNTUACEP, Pulivendula, AP, India
[2]Assistant Professor, CSE Department, JNTUACEP, Pulivendula, AP, India.

E-mail: [1]banala.deepthi2002@gmail.com, [2]sivajntuacep@gmail.com, [3]beejapurshaikjubedha@gmail.com

Abstract

Classification algorithms have been found to produce better results in terms of performance and accuracy, when used with drug observation dataset. Three machine learning algorithms such as J48, Naive Bayes, and K-Nearest Neighbor are compared using Waikato Environment for Knowledge Analysis (WEKA) in this paper. In addition, these three well-known classification methods are evaluated based on different Quality of Service parameters to find the best fit classifier for the design of the model. The analysis procedure of dataset and the performance indicators are discussed. These results help to draw a conclusion about which of the three algorithms is the best for drug classification.

Keywords: Classification, J48, Naïve Bayes, K-Nearest Neighbour, cross validation, accuracy, correctly classified and incorrectly classified instances.

1. Introduction

Data mining is advancing significantly across several industries. Data mining is used in a variety of applications and is created for various databases. Assigning the objects in a collection to specific categories or classes is the function of classification in data mining. To accurately determine the target class for each case in the data is the aim of classification [1]. Classification, clustering, association rule extraction, regression, and visualisation are some of the words used in data mining functions [4].

The classification is supervised learning, where grade levels or prophetic objectives are previously known. As a result, the rule structure that represents the classification form is created during the classification process. In this instance, the model developed reflects

DOI: https://doi.org/10.36548/jtcsst.2022.4.003
Received: 07.09.2022, received in revised form: 12.10.2022, accepted: 27.11.2022, published: 08.12.2022

important data and is applied to future planning. Analytical modelling is a sort of classification. A concept for transforming the current object into a known type or class is classification. From a group of tagged records, a decision tree-like prototype is created and labels are assigned to the upcoming untagged records [2].

Database technology, statistical analysis, and artificial intelligence are all used extensively in data mining technology, which has proven to have significant commercial value, and widely used in retail, insurance, telecommunications, and in power industries have been gradually adopted by other professions.

A thorough evaluation of several classification methods in data mining is provided through the study of various classification strategies, including ID3, Decision Tree (DT), C4.5, Bayesian classification, and KNN classifier [3]. The differences between the K-Nearest Neighbour (KNN), Naïve Bayes and Decision Tree classification algorithms are summarized.

Table 1. Comparison of different parameters between different classification techniques

Parameter	**KNN**	**Naïve Bayes**	**Decision Tree**
Deterministic/ Nondeterministic	Non-deterministic	Non-deterministic	Deterministic
Effectiveness on	Small data	Huge data	Large data
Speed	Slow	Faster than KNN	Faster
Dataset	It is unable to handle noisy data	It can handle noisy data	It can handle noisy data.
Accuracy	High	It needs a lot of records	High

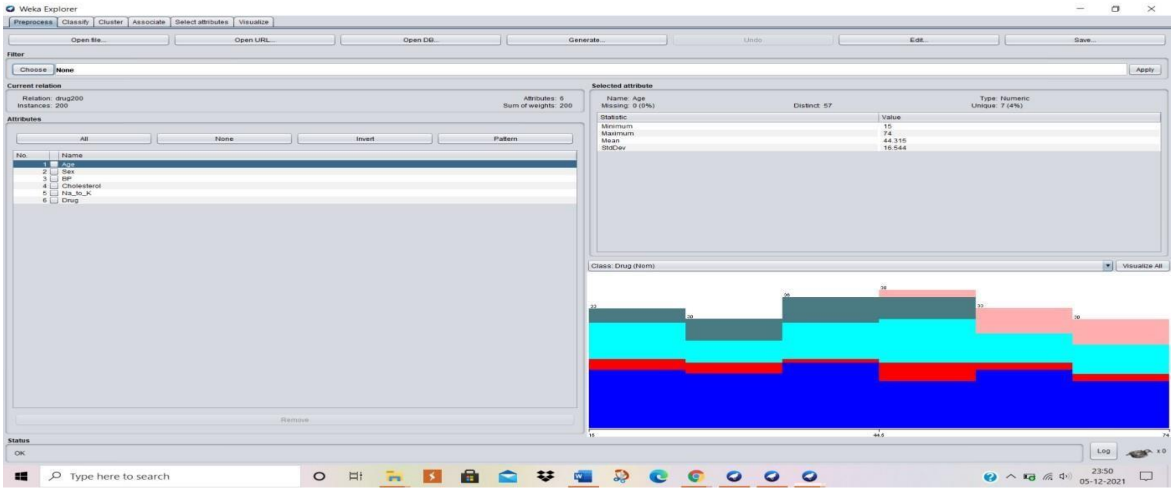
The above table explains the comparison of different classification techniques (KNN, Naïve Bayes and Decision Tree) based on different parameters like deterministic/non-deterministic, effectiveness, speed, data handling and accuracy.

2. Problem Statement

WEKA's (Waikato Environment for Knowledge Analysis) data mining tool was used to compare the classification algorithms (Decision Trees, Naive Bayes, and KNN) on the drug dataset to determine which algorithm works efficiently and which one is the best for a given dataset.

3. Drug Dataset

The UCI ML repository is used to choose drug-related files. Analysis is done on the effectiveness of a full set of classification algorithms. 200 instances and 6 attributes are included in the collection. The major purpose of this dataset is to differentiate between test outcomes, i.e., the proportion of subjects that tested positively and negatively. The information gathered allows one to determine whether or not a person is a drug addict. Attributes used include age (a person's age), gender (a person's sex), BP (Blood Pressure readings such as high, normal, and low), cholesterol (a person's cholesterol readings), Na_to_K, Drug (any kind), etc.



Filter: None

Relation: drug200
Instances: 200
Attributes: 6
Sum of weights: 200

No.	Name
1	Age
2	Sex
3	BP
4	Cholesterol
5	Na_to_K
6	Drug

Name: Age
Missing: 0 (0%)
Distinct: 57
Type: Numeric
Unique: 7 (4%)

Statistic	Value
Minimum	15
Maximum	74
Mean	44.315
StdDev	16.544

Status: OK

Figure 1. Weka Explorer interface

The above picture shows the drug dataset open in WEKA and the number of instances present in the attributes of the dataset. Information is gathered from a large number of patients with the same condition. Each patient responded to one of the five medications—drugs A, B, C, X, and Y, during the course of treatment. Building a model to determine which medications are appropriate for future patients with the same problem is part of the job description.

Table 2. Brief description about the dataset (Drug dataset)

Age	Sex	BP	Cholesterol	Na_to_K	Drug
47	M	LOW	HIGH	13.093	Drug C
28	F	NORMAL	HIGH	7.798	Drug X
22	F	NORMAL	HIGH	8.607	Drug X
23	F	HIGH	HIGH	25.355	Drug Y

47	M	LOW	HIGH	10.114	Drug C
61	F	LOW	HIGH	18.043	Drug Y
49	F	NORMAL	HIGH	16.275	Drug Y

The above table is an illustration of a multiclass classifier, where the dataset's training component is utilised to create a decision tree that can be used to predict unknown patient classes or recommend medications to new patients.

4. Literature Survey

Table 3. Tabulation of Related Works

Ref. No.	Authors	Title	Description
[2]	Swathi Agarwal, G.L.Anand Babu, and Dr.K.S.Reddy	Classification Techniques in Data Mining -Case Study	This paper focused on the data mining techniques that have been used to categorise medical data. The data mining methods utilised for medical data used to identify and diagnose various disorders affecting human health were discussed in this study.
[10]	Safae Sossi Alaoui, Yousef Farhaoui and Brahim Aksasse.	Classification algorithms in Data Mining	The use of data mining techniques to manage healthcare datasets has grown significantly. According to the stage of each condition, they classify patients who may have diabetes, hypothyroidism, or breast cancer.
[11]	Dr.S.Vijayarani, Mr.S.Dhayanand	Data mining classification algorithms for kidney disease prediction.	The primary goal of this project was to use classification techniques like Naive Bayes and Support Vector Machine to predict kidney disorders.
[12]	N. Chandra Sekhar Reddy, K. Sai Prasad and A. Mounika.	Classification of Algorithms on Data mining: A Study	This paper provided a detailed analysis of the benefits and drawbacks of several classification algorithms.
[13]	Archit Verma	Study and evaluation of Classification Algorithms in data mining	The classification method is also known as the supervised learning method. In order to place a data instance into one of the specified classes, classification was used to predict the categorical class label of the data instance.
[14]	Parneet Kaura, Manpreet Singhb, and Gurpreet Singh Josanc	Classification and prediction based data mining algorithms to	This paper focused on identifying students who are slow learners and showing them so, using a predictive data mining model with classification-based algorithms. Using WEKA,

		predict slow learners in education sector	an open source tool, real-world data from a high school was filtered to exclude the desired possible variables.
[15]	Saima Anwar Lashari , Rosziati Ibrahim, Norhalina Senan and N. S. A. M. Taujuddin	Application of Data Mining Techniques for Medical Data Classification: A Review	It is clear from this paper that data mining techniques have been used to categorize medical data. The data mining methods utilised for medical data used to identify and diagnose various disorders affecting human health were discussed in this study.

5. Methodlogy

The approach used to achieve the goals in comparing a number of data mining classification algorithms is outlined in this section.

1) To make the training datasets precise.
2) To choose the appropriate open source application for the task.
3) To assess the models produced by a few methodologies.
4) To evaluate the effectiveness of each strategy.

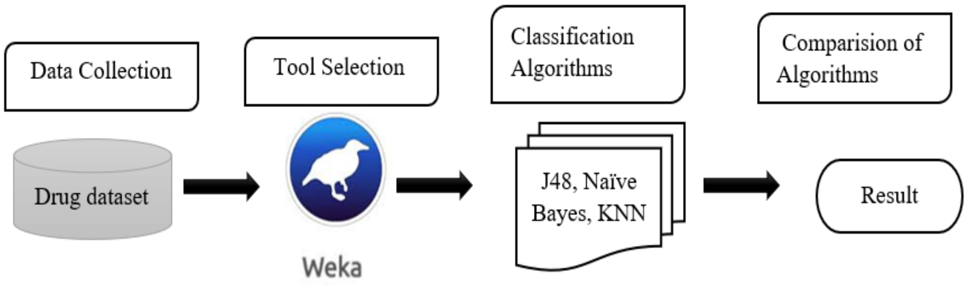


Figure 2. Methodology of Research

Although the WEKA tool offers a wide variety of categorization methods, only three have been chosen for this article. In data mining research fields, classification algorithms received a lot of attention [5]. To group data into distinct classifications, classification procedures must be analysed.

5.1 J48

The J48 algorithm is used to classify various applications and get accurate classification results. The J48 algorithm is one of the best machine learning algorithms for categorically and continuously examining data. For example, when used for that purpose, it takes up more storage space and reduces the performance and accuracy of medical data classification. The proposed method is to measure the improved performance and yield higher accuracy rates.

5.1.1 Advantages of J48

- J48 is easy to train.
- J48 is efficient.
- J48 has been included in many winning packages.

Some specific uses included are classification in soybean disease and web search.

5.2 Naive Bayes

It is an easy probabilistic classifier that determines a set of possibilities by adding values from the given different datasets and counting invariants [7]. This approach can be used in machine learning to infer new data or test data based on "Bayesian" theory [8].

The WEKA tool uses this algorithm and applies it. Using supervised discretization to change numeric characteristics into regular attributes, this tool employs an estimator for numeric attributes and offers the option to perform the aforementioned algorithm [9].

5.2.1 Advantages

- Fast training (simple scan)
- Sort quickly
- Insensitive to irrelevant features
- Handle real and discrete data
- Properly handle streaming data

5.2.2 Disadvantages

- It assumes functional independence

5.3 K-Nearest-Neighbor algorithm

All existing cases and categories simply provide a measure of similarity. It is useful for statistical analysis and pattern recognition. KNN is a non-parametric lagged learning algorithm. If the method is non-parametric, it means that it makes no assumptions about the underlying data distribution [6]. KNN is a lagging algorithm. This indicates that the training dataset is not used to perform the generalization.

The following are some benefits of KNN

- The algorithm's accuracy won't be impacted by the smooth addition of the new data.
- Implementation is quite simple.

KNN has a number of drawbacks, including the following

- Features must be scaled before being applied to any dataset; else, incorrect predictions will be generated.
- Does not perform well with large datasets since it is very expensive to calculate the distance between each new point and each old point in large datasets, which degrades the algorithm's speed.
- Sensitive to missing numbers, noisy data, and outliers.

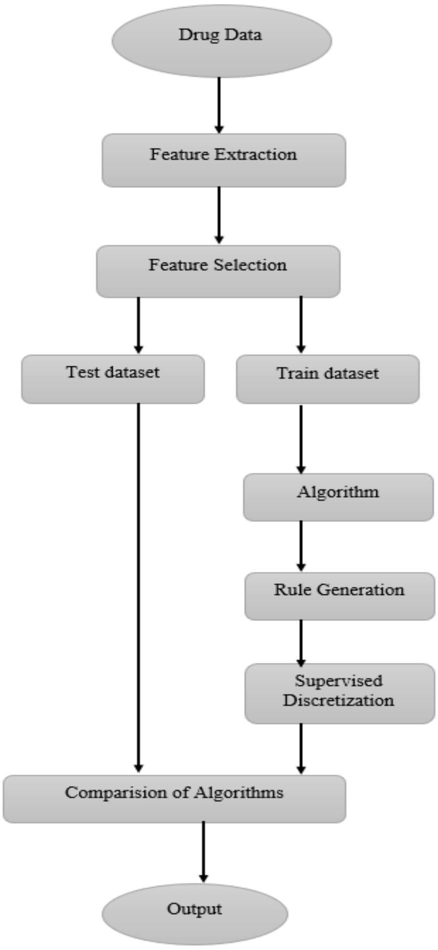


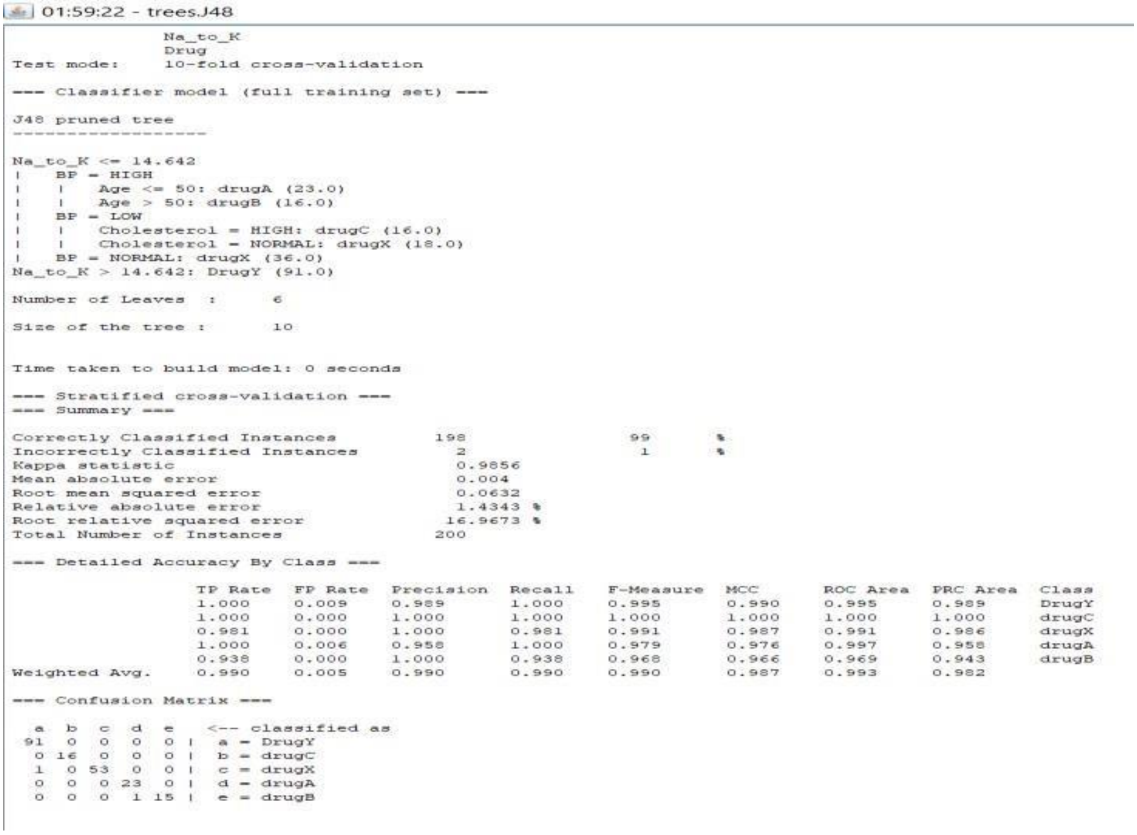
Figure 3. Flowchart of the Proposed Methodology

The above flow chart explains the applying of different classification algorithms (KNN, Naïve Bayes, and J48) on the drug dataset and then under the rule generation, the Supervised Discretization must be selected.

6. Experimental Results

- This investigation examines the effectiveness of several classification algorithms on drug dataset.
- Drug record classifiers are compared based on the following criteria: accuracy, correctly classified instances, incorrectly classified instances, error rate, and execution time.
- Each algorithm's models are created using one of two primary techniques.
- 10-fold cross-validation (66% of the dataset is utilised for training, 34% for testing, and 9 folds and 1 fold on the training set are used for testing).
- Using a 10-fold cross-validation test bed, the figures are shown to compare several drug record classifiers.

6.1 J48



01:59:22 - trees.J48

```
               Na_to_K
               Drug
Test mode:     10-fold cross-validation

=== Classifier model (full training set) ===

J48 pruned tree
------------------

Na_to_K <= 14.642
|   BP = HIGH
|   |   Age <= 50: drugA (23.0)
|   |   Age > 50: drugB (16.0)
|   BP = LOW
|   |   Cholesterol = HIGH: drugC (16.0)
|   |   Cholesterol = NORMAL: drugX (18.0)
|   BP = NORMAL: drugX (36.0)
Na_to_K > 14.642: DrugY (91.0)

Number of Leaves  :      6

Size of the tree :       10


Time taken to build model: 0 seconds

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

Correctly Classified Instances         198               99      %
Incorrectly Classified Instances         2                1      %
Kappa statistic                          0.9856
Mean absolute error                      0.004 
Root mean squared error                  0.0632
Relative absolute error                  1.4343 %
Root relative squared error             16.9673 %
Total Number of Instances              200     

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.009    0.989      1.000    0.995      0.990    0.995     0.989     DrugY
                 1.000    0.000    1.000      1.000    1.000      1.000    1.000     1.000     drugC
                 0.981    0.000    1.000      0.981    0.991      0.987    0.991     0.986     drugX
                 1.000    0.006    0.958      1.000    0.979      0.976    0.997     0.958     drugA
                 0.938    0.000    1.000      0.938    0.968      0.966    0.969     0.943     drugB
Weighted Avg.    0.990    0.005    0.990      0.990    0.990      0.987    0.993     0.982     

=== Confusion Matrix ===

  a  b  c  d  e   <-- classified as
 91  0  0  0  0 |  a = DrugY
  0 16  0  0  0 |  b = drugC
  1  0 53  0  0 |  c = drugX
  0  0  0 23  0 |  d = drugA
  0  0  0  1 15 |  e = drugB
```

Figure 4. Outcome of J48 algorithm

6.1.1 Analysis using J48 Algorithm

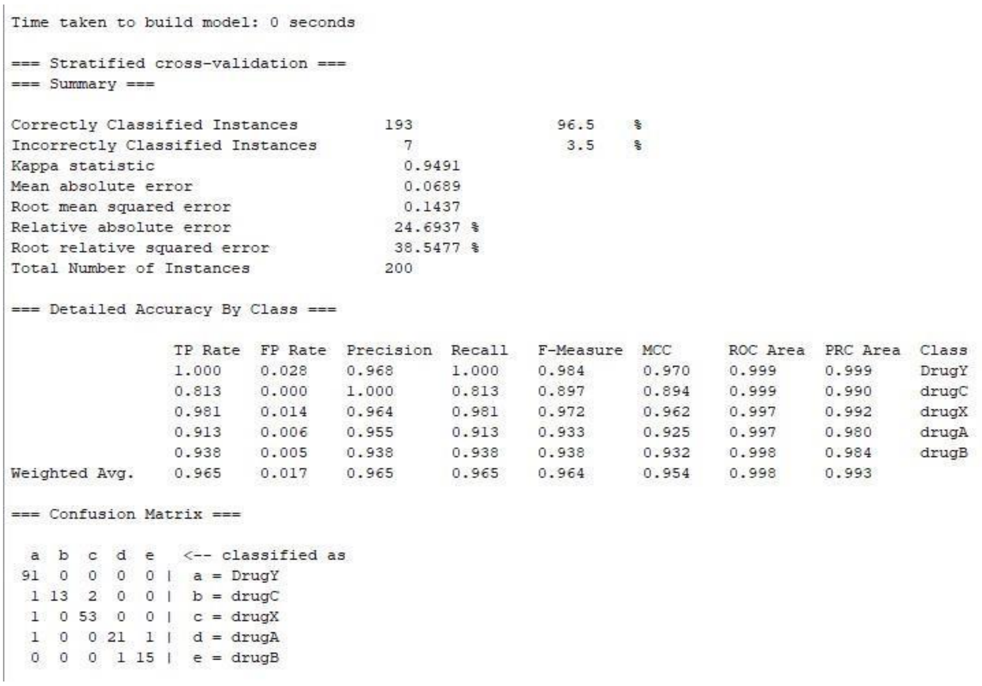
WEKA's standard settings are used for the initial analysis of this dataset. Choosing "Use training" set with a 66% training set and various test options allows for classification, with 80% of the data being used for training and the remaining 20% being utilised for testing. The J48 method is shown in this table for two different parameters.

Table 4. J48 Algorithm table

	Default parameter		Supervised descretize	
Correct instance	193	96.5%	198	99%
Wrong instance	7	3.5%	2	1%
Total instance	200	100%	200	100%

According to the above table, 193 of the 200 examples are correctly classified, while 7 are wrongly labelled. By switching the parameters to "supervised discretization," the final analysis is carried out. Instances that are correctly classified are changed to 198, while instances that are mistakenly categorised are changed to 2.

6.2 Naïve Bayes



```
Time taken to build model: 0 seconds

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

Correctly Classified Instances         193               96.5    %
Incorrectly Classified Instances         7                3.5    %
Kappa statistic                          0.9491
Mean absolute error                      0.0689
Root mean squared error                  0.1437
Relative absolute error                 24.6937 %
Root relative squared error             38.5477 %
Total Number of Instances              200

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 1.000    0.028    0.968      1.000    0.984      0.970    0.999     0.999     DrugY
                 0.813    0.000    1.000      0.813    0.897      0.894    0.999     0.990     drugC
                 0.981    0.014    0.964      0.981    0.972      0.962    0.997     0.992     drugX
                 0.913    0.006    0.955      0.913    0.933      0.925    0.997     0.980     drugA
                 0.938    0.005    0.938      0.938    0.938      0.932    0.998     0.984     drugB
Weighted Avg.    0.965    0.017    0.965      0.965    0.964      0.954    0.998     0.993

=== Confusion Matrix ===

  a  b  c  d  e   <-- classified as
 91  0  0  0  0 |  a = DrugY
  1 13  2  0  0 |  b = drugC
  1  0 53  0  0 |  c = drugX
  1  0  0 21  1 |  d = drugA
  0  0  0  1 15 |  e = drugB
```

Figure 5. Outcome of Naïve Bayes algorithm

6.2.1 Analysis using Naive Bayes Algorithm

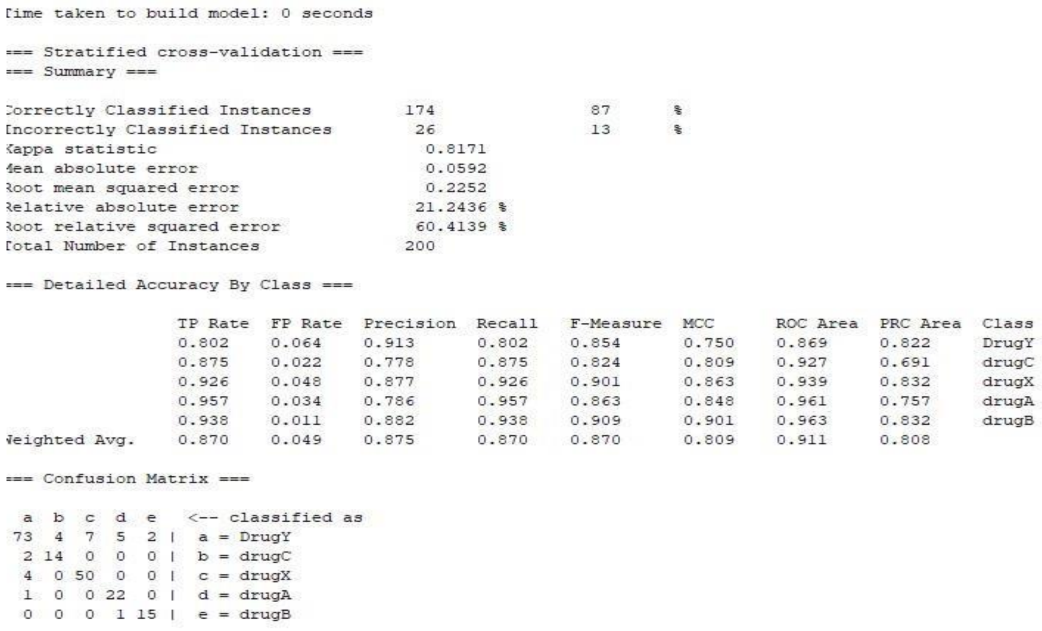
WEKA's standard settings are used for the initial analysis of this dataset. Choosing "Use training" set with a 66% training set and various test options allows for classification, with 80% of the data being used for training and the remaining 20% being utilised for testing. The Naïve Bayes method is shown in this table for two different parameters.

Table 5. Naïve Bayes Algorithm table

	Default parameter		Supervised descretize	
Correct instance	198	96.5%	200	99%
Wrong instance	2	3.5%	0	1%
Total instance	200	100%	200	100%

According to the above table, 198 of the 200 examples are correctly classified, while 2 are wrongly labelled. By switching the parameters to "supervised discretization," the final analysis is carried out. Instances that are correctly classified are changed to 200, while instances that are mistakenly categorised are changed to zero.

6.3 K-Nearest Neighbor



```
Time taken to build model: 0 seconds

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

Correctly Classified Instances         174               87      %
Incorrectly Classified Instances        26               13      %
Kappa statistic                          0.8171
Mean absolute error                      0.0592
Root mean squared error                  0.2252
Relative absolute error                 21.2436 %
Root relative squared error             60.4139 %
Total Number of Instances              200

=== Detailed Accuracy By Class ===

                 TP Rate  FP Rate  Precision  Recall   F-Measure  MCC      ROC Area  PRC Area  Class
                 0.802    0.064    0.913      0.802    0.854      0.750    0.869     0.822     DrugY
                 0.875    0.022    0.778      0.875    0.824      0.809    0.927     0.691     drugC
                 0.926    0.048    0.877      0.926    0.901      0.863    0.939     0.832     drugX
                 0.957    0.034    0.786      0.957    0.863      0.848    0.961     0.757     drugA
                 0.938    0.011    0.882      0.938    0.909      0.901    0.963     0.832     drugB
Weighted Avg.    0.870    0.049    0.875      0.870    0.870      0.809    0.911     0.808

=== Confusion Matrix ===

  a  b  c  d  e   <-- classified as
 73  4  7  5  2 |  a = DrugY
  2 14  0  0  0 |  b = drugC
  4  0 50  0  0 |  c = drugX
  1  0  0 22  0 |  d = drugA
  0  0  0  1 15 |  e = drugB
```

Figure 6. Outcome of KNN algorithm

6.3.1 Analysis using KNN Algorithm

WEKA's standard settings are used for the initial analysis of this dataset. Choosing "Use training" set with a 66% training set and various test options allows for classification, with 80% of the data being used for training and the remaining 20% being utilised for testing. The KNN method is shown in this table for two different parameters.

Table 6. KNN Algorithm Table

	Default parameter		Supervised descretize	
Correct instance	174	96.5%	200	99%
Wrong instance	26	3.5%	0	1%
Total instance	200	100%	200	100%

According to the above table, 174 of the 200 examples are correctly classified, while 26 are wrongly labelled. By switching the parameters to "supervised discretization," the final analysis is carried out. Instances that are correctly classified are changed to 200, while instances that are mistakenly categorised are changed to zero.

7. Evaluation of Errors of the Methods

The following errors are evaluated:

- **Kappa statistics:** It estimates the degree of agreement between observers when categorising qualitatively.
- **Mean Absolute Error**: The mean of the absolute differences between predictions and actual observations is known as the mean absolute error.
- **Mean Squared Error:** The mean squared deviation between the forecast and the actual observation is the mean squared error.
- **Relative square root error.**
- **Absolute relative error.**

Table 7. Evaluation of errors of the methods

Method	**Kappa statistic**	**Mean absolute error**	**Mean squared error**	**Relative absolute error**	**Relative squared error**
J48	0.9856	0.004	0.0632	1.4343	16.9673
Naïve bayes	0.9491	0.0689	0.1437	24.6937	38.5477
KNN	0.9491	0.0689	0.1437	24.6937	38.5477

The above table explains the evaluation of error of the methods like Kappa statistics, Mean Absolute Error, Mean Squared Error, Relative absolute error and Relative squared error for J48, Naïve Bayes and KNN.

8. Results

The test option is 10-fold cross-validation. Comparing all the three algorithms shows that the J48 algorithm is more accurate.

Table 8. Comparison of classifiers for Drug dataset using cross validation testing bed

Classifier	**J48**	**Naïve Bayes**	**KNN**
Testing Bed	Cross Validation	Cross Validation	Cross Validation
Applications	Emotion recognition, Verbal column pathologies.	Text classification, Spam filtering, Online Application, Hybrid recommender system.	Agriculture, finance, medicine, face identification, Recommendation system like Amazon, Netflix, etc.
Accuracy	99%	96.5%	87%

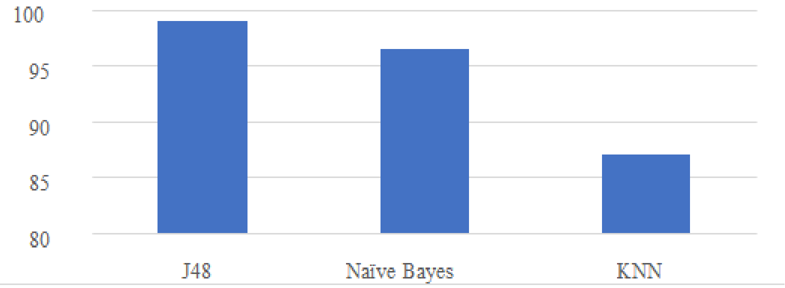


Figure 7. Graphical representation of different classification algorithms' accuracy using cross validation method

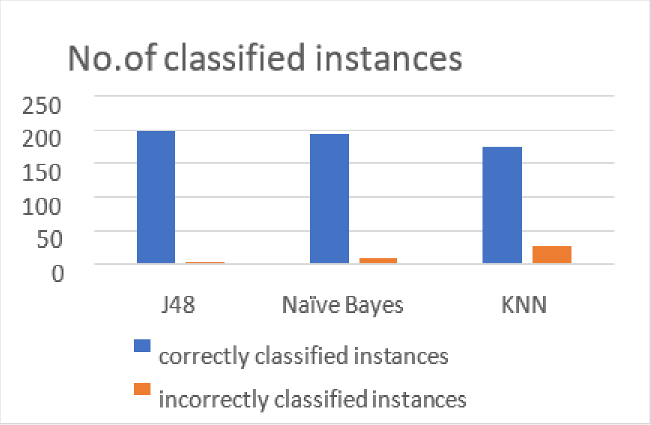


Figure 8. Correctly and Incorrectly classified instances in case of Cross Validation

The above table shows the comparison of classifiers for the drug dataset using cross validation testing bed, which explains some applications related to that algorithms, and it shows the accuracy results for the J48, Naive Bayes, and KNN algorithms. The above graphical representations tells about the different classification algorithms' accuracy, correctly and incorrectly classified instances by using the cross validation method.

9. Conclusion

The dataset is analysed by utilizing the WEKA tool's standard parameters. A training set of 66% is used for classification, followed by a variety of testing choices employing 80% of the data for training and 20% for testing. It displays both correctly and incorrectly classified examples. Changing the parameters to Supervised Discretization completes the final analysis. The accuracy has increased by this adjustment. This establishes J48 as a simple classification technique for creating decision trees. In experiments, the WEKA tool produced effective results using drug datasets. The results from KNN and Naive Bayesian classifiers are similarly promising. The study's experimental findings reveal that J48 has a better level of classification accuracy than the other two methods.

References

[1] Ghopi Gandi, Rohith Shreevastav "Modified k-methods algorithms for analysis and application to increase scalabilities and efficiencies for larger datasets" International Journal of Research in Engineering and Technology, no. 3(2014):150-153.

[2] Swathi Agarwal , G.L.Anand Babu , Dr.K.S.Reddy "Classification Techniques in Data Mining-Case Study" - IOSR Journal of Computer Engineering, Volume 18 (2016): 30-33.

[3] KS Reddy, GPS Varma, MK Reddy, An Effective Methodology for Pattern Discovery in Web Usage Mining - International Journal of Computer Science and Information Technologies, International Journal of Computer Science and Information Technologies, Vol. 3 (2) , 2012: 3664-3667.

[4] KS Reddy, GPS Varma, SSS Reddy, Understanding the scope of web usage mining & applications of web data usage patterns, International Conference on Computing, Communication and Applications, 2012: 1-5.

[5] L. Breiman, L. Friedman, R. Olshen, and C. Stone. Classification and Regression Trees. Wadsworth and Brooks, 1984.

[6] Boser, B. E., I. Guyon, and V. Vapnik (1992). A training algorithm for optimal margin classifiers. In Proceedings of the Fifth Annual Workshop on Computational Learning Theory, ACM Press (1992): 144 -152.

[7] George Dimytoglou, James Adam, and Carol M. Jhim, “Comparisons of C4.5 and Naive Bayes Classification for the Analysis of Lung Cancer Survivabilities” Journal of Computing, Volume 4, Issue 8 (2012): 1-9.

[8] Olivier C. Fhran, kois and Philip Leray, “Study of the Tree Augmented Naive Bayes Classification from deficient datasets”, vol 06, issue no. 15 (2018): 1-4

[9] Tina R. Patil, Mrs. S. S. Sherekar, “ Performance Analysis of Naive Bayes and J48 Classification Algorithm for Data Classification”, International Journal Of Computer Science And Applications Vol. 6, No.2 (2013) :256-261.

[10] Safae Sossi Alaoui , Yousef Farhaoui, Brahim Aksasse, “Classification algorithms in Data Mining”, International Journal of Tomography and Simulation(2018): 1-12.

[11] Dr. S. Vijayarani , Mr.S.Dhayanand, “DATA MINING CLASSIFICATION ALGORITHMS FOR KIDNEY DISEASE PREDICTION” - International Journal on Cybernetics & Informatics(2015): 13-25.

[12] N. Chandra Sekhar Reddy, K. Sai Prasad and A. Mounika, “Classification Algorithms on Datamining: A Study” - International Journal of Computational Intelligence Research Volume 13, Number 8 (2017): 2135-2142.

[13] Archit Verma, “Study and evaluation of Classification Algorithms in data mining” - International Research Journal of Engineering and Technology, Volume 05 Issue 08 (2018): 1297 - 1308.

[14] Parneet Kaura ,Manpreet Singhb ,Gurpreet Singh Josanc, “ Classification and prediction based data mining algorithms to predict slow learners in education sector” – ICRTC, Procedia Computer Science 57 (2015): 500-508.

[15] Saima Anwar Lashari, Rosziati Ibrahim , Norhalina Senan and N. S. A. M. Taujuddin, “Application of Data Mining Techniques for Medical Data Classification: A Review”(2017): 1-6.

Author’s biography

B. Deepthi is pursuing B.Tech, Department of Computer Science and Engineering in Jawaharlal Nehru Technological University, Pulivendula, Kadapa District, Andhra Pradesh, India. Her main areas of Interest is in Data mining, Machine Learning, Deep Learning, Data Science and Artificial Intelligence.

K. V. Siva Prasad Reddy, Assistant Professor (Adhoc), Department of Computer Science and Engineering in Jawaharlal Nehru Technological University, Pulivendula, Kadapa

District, Andhra Pradesh, India. His main areas of Interest is in Big Data, Machine Learning, Artificial Intelligence, Cloud Computing and Data Mining.

B .S. Jubedha is pursuing B.Tech, Department of Computer Science and Engineering in Jawaharlal Nehru Technological University, Pulivendula, Kadapa District, Andhra Pradesh, India. Her main areas of Interest is in Data Mining, Machine Learning, Database Management System(DBMS).