|  |  |
| --- | --- |
| **Student Name:** | Xihai Ren |
| **Student No.:** | 041127486 |
| **Due Date:** | June 7, 2024 |
| **Course:** | CST8390 BI |
| **Professor:** | Anu Thomas |

**CST8390 Assignment 1 Report**

**Classification using KNN and Decision Trees**

**Table of Contents**

[1 Introduction 3](#_Toc168605888)

[2 Business Understanding 3](#_Toc168605889)

[3 Data Understanding 3](#_Toc168605890)

[3.1 Collect Initial Data 3](#_Toc168605891)

[3.2 Describe Data 3](#_Toc168605892)

[3.3 Explore Data 4](#_Toc168605893)

[3.4 Verify Data Quality 7](#_Toc168605894)

[4 Data Preparation 9](#_Toc168605895)

[4.1 Select Data 9](#_Toc168605896)

[4.2 Clean Data 9](#_Toc168605897)

[4.3 Construct Data 10](#_Toc168605898)

[4.4 Integrate Data 10](#_Toc168605899)

[4.5 Format Data 10](#_Toc168605900)

[5 Modeling 10](#_Toc168605901)

[5.1 KNN 10](#_Toc168605902)

[5.2 Decision Tree 12](#_Toc168605903)

[6 Comparison of Results 18](#_Toc168605904)

[7 Conclusion 19](#_Toc168605905)

[8 Reference 19](#_Toc168605906)

Tables

[Table 1 Attributes Table 4](#_Toc168605748)

[Table 2 Attributes and Instances statistic 4](#_Toc168605749)

[Table 3 Attributes Data Distribution 4](#_Toc168605750)

[Table 4 Gender table statistic 5](#_Toc168605751)

[Table 5 Selector Distribution 5](#_Toc168605752)

[Table 6 A/G Missing Values 7](#_Toc168605753)

[Table 7 Outliers Table 9](#_Toc168605754)

[Table 8 KNN Classify Table 11](#_Toc168605755)

[Table 9 J48 Decision Tree Train Table 15](#_Toc168605756)

[Table 10 Metric Compare 19](#_Toc168605757)

Figures

[Figure 1 the relationship between Total Proteins (TP) and Albumin (ALB) 5](#_Toc168605758)

[Figure 2 the relationship between Total Bilirubin (TB) and Direct Bilirubin (DB) 6](#_Toc168605759)

[Figure 3 the relationship between the Selector (class label) and Total Bilirubin (TB) 7](#_Toc168605760)

[Figure 4 Class Distribution 9](#_Toc168605761)

[Figure 5 KNN Statistic 12](#_Toc168605762)

[Figure 6 J48 Train Result Statistic 16](#_Toc168605763)

[Figure 7 J48 pruned tree -C 0.25 -M 33, Test mode: 10-fold cross-validation, Seed = 10. 17](#_Toc168605764)

[Figure 8 J48 unpruned tree, 10-fold cross-validation, Seed = 10, -M=33 18](#_Toc168605765)

# Introduction

Liver disease rates are rising due to excessive alcohol consumption, exposure to harmful gases, and the intake of contaminated food and drugs. Early detection of liver disease is crucial for improving patient outcomes and reducing the burden on healthcare professionals.

This dataset, containing 583 patient records from Northeast Andhra Pradesh, India, includes various biochemical markers and demographic information. The objective is to use machine learning algorithms (KNN and Decision Trees) to predict whether a patient has liver disease based on various biochemical markers.

This report will detail each step from business understanding, data understanding, and preparation to modeling, evaluation, and conclusion.

# Business Understanding

* The ILPD dataset contains a total of 583 records, including 416 liver patient records and 167 non-liver patient records.
* The dataset includes 11 features, such as age, gender, total bilirubin, direct bilirubin, alkaline phosphatase, alanine Aminotransferase, aspartate Aminotransferase, total proteins, albumin, albumin/globulin ratio, and Selector.
* Exploratory data analysis could be performed to understand the distribution and correlation of the features.
* Exploratory data analysis shows that liver patients tend to have higher levels of total bilirubin, direct bilirubin, alkaline phosphatase, and aspartate transaminase.
* Correlation analysis indicates a significant relationship between total bilirubin and direct bilirubin.

# Data Understanding

## Collect Initial Data

This dataset was downloaded from the UCI ML Repository:

Ramana, Bendi and Venkateswarlu, N.. (2012). ILPD (Indian Liver Patient Dataset). UCI Machine Learning Repository. <https://doi.org/10.24432/C5D02C>.

The ILPD dataset contains liver patient records collected from Northeast Andhra Pradesh, India, with a total of 583 records, including 416 liver patient records and 167 non-liver patient records, the dataset provides valuable insights into the factors that impact liver disease.

## Describe Data

|  |  |  |
| --- | --- | --- |
| Attribute Name | Description | Data Type |
| Age | Age of the patient. | Numeric |
| Gender | Gender of the patient | Nominal |
| TB | Total Bilirubin | Numeric |
| DB | Direct Bilirubin | Numeric |
| Alkphos | Alkaline Phosphatase | Numeric |
| Sgpt | Alamine Aminotransferase | Numeric |
| Sgot | Aspartate Aminotransferase | Numeric |
| TP | Total Proteins | Numeric |
| ALB | Albumin | Numeric |
| A/G Ratio | Albumin and Globulin Ratio | Numeric |
| Selector | Class label  1 - Liver disease  2 - not Liver disease | Nominal |

Table 1 Attributes Table

## Explore Data

|  |  |
| --- | --- |
| **Attributes** | 11 |
| **Instances** | 583 |

Table 2 Attributes and Instances statistic

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Attribute | Missing | Minimum | Maximum | Mean | StdDev | Distribution |
| Age | 0 | 4 | 90 | 44.746 | 16.19 | Normal Distribution |
| TB | 0 | 0.4 | 75 | 3.299 | 6.21 | Skewed Distribution |
| DB | 0 | 0.1 | 19.7 | 1.486 | 2.808 | Skewed Distribution |
| Alkphos | 0 | 63 | 2110 | 290.576 | 242.938 | Skewed Distribution |
| Sgpt | 0 | 10 | 2000 | 80.714 | 182.62 | Skewed Distribution |
| Sgot | 0 | 10 | 4929 | 109.911 | 288.919 | Skewed Distribution |
| TP | 0 | 2.7 | 9.6 | 6.483 | 1.085 | Normal Distribution |
| ALB | 0 | 0.9 | 5.5 | 3.142 | 0.796 | Normal Distribution |
| A/G Ratio | 4 | 0.3 | 2.8 | 0.947 | 0.32 | Skewed Distribution |

Table 3 Attributes Data Distribution

|  |  |  |
| --- | --- | --- |
|  | Female | Male |
| Gender | 142 | 441 |

Table 4 Gender table statistic

|  |  |  |
| --- | --- | --- |
|  | 1 - Liver disease | 2 - not Liver disease |
| Selector | 416 | 167 |

Table 5 Selector Distribution

A screenshot of a computer

Description automatically generated

Figure 1 the relationship between Total Proteins (TP) and Albumin (ALB)

The scatter plot visualizes the relationship between Total Proteins (TP) and Albumin (ALB) in the Indian Liver Patient Dataset. The scatter plot shows a clear positive correlation between total proteins and albumin, indicating that as the total proteins increase, the albumin levels also increase.

A screenshot of a computer

Description automatically generated

Figure 2 the relationship between Total Bilirubin (TB) and Direct Bilirubin (DB)

The scatter plot visualizes the relationship between Total Bilirubin (TB) and Direct Bilirubin (DB) in the Indian Liver Patient Dataset. The scatter plot shows a strong positive correlation between total bilirubin and direct bilirubin, indicating that as the total bilirubin increases, the direct bilirubin levels also increase.

A screenshot of a computer

Description automatically generated

Figure 3 the relationship between the Selector (class label) and Total Bilirubin (TB)

The scatter plot visualizes the relationship between the Selector (class label) and Total Bilirubin (TB) in the Indian Liver Patient Dataset. When total bilirubin values exceed 18, the proportion of liver patients is significantly higher.

## Verify Data Quality

**Missing value:**

A/G Ratio has four missing values, 1%. Populate with the mean value.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age | Gender | TB | DB | Alkphos | Sgpt | Sgot | TP | ALB | A/G Ratio | Selector |
| 45 | Female | 0.9 | 0.3 | 189 | 23 | 33 | 6.6 | 3.9 |  | 1 |
| 51 | Male | 0.8 | 0.2 | 230 | 24 | 46 | 6.5 | 3.1 |  | 1 |
| 35 | Female | 0.6 | 0.2 | 180 | 12 | 15 | 5.2 | 2.7 |  | 2 |
| 27 | Male | 1.3 | 0.6 | 106 | 25 | 54 | 8.5 | 4.8 |  | 2 |

Table 6 A/G Missing Values

**Data type:**

All Data types are right.

**Outlier:**

There are many outliers.

|  |  |  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- | --- | --- |
| Age | Gender | TB | DB | Alkphos | Sgpt | Sgot | TP | ALB | A/G Ratio | Selector |
| 34 | Male | 4.1 | 2 | 289 | 875 | 731 | 5 | 2.7 | 1.1 | 1 |
| 34 | Male | 4.1 | 2 | 289 | 875 | 731 | 5 | 2.7 | 1.1 | 1 |
| 68 | Female | 0.6 | 0.1 | 1620 | 95 | 127 | 4.6 | 2.1 | 0.8 | 1 |
| 50 | Male | 7.3 | 3.6 | 1580 | 88 | 64 | 5.6 | 2.3 | 0.6 | 2 |
| 48 | Male | 0.7 | 0.1 | 1630 | 74 | 149 | 5.3 | 2 | 0.6 | 1 |
| 32 | Male | 12.7 | 6.2 | 194 | 2000 | 2946 | 5.7 | 3.3 | 1.3 | 1 |
| 32 | Male | 15.9 | 7 | 280 | 1350 | 1600 | 5.6 | 2.8 | 1 | 1 |
| 32 | Male | 18 | 8.2 | 298 | 1250 | 1050 | 5.4 | 2.6 | 0.9 | 1 |
| 58 | Female | 1.7 | 0.8 | 1896 | 61 | 83 | 8 | 3.9 | 0.95 | 1 |
| 66 | Male | 11.3 | 5.6 | 1110 | 1250 | 4929 | 7 | 2.4 | 0.5 | 1 |
| 60 | Male | 11.5 | 5 | 1050 | 99 | 187 | 6.2 | 2.8 | 0.8 | 1 |
| 55 | Male | 75 | 3.6 | 332 | 40 | 66 | 6.2 | 2.5 | 0.6 | 1 |
| 75 | Male | 14.8 | 9 | 1020 | 71 | 42 | 5.3 | 2.2 | 0.7 | 1 |
| 60 | Male | 2.4 | 1 | 1124 | 30 | 54 | 5.2 | 1.9 | 0.5 | 1 |
| 7 | Female | 27.2 | 11.8 | 1420 | 790 | 1050 | 6.1 | 2 | 0.4 | 1 |
| 40 | Male | 3.9 | 1.7 | 350 | 950 | 1500 | 6.7 | 3.8 | 1.3 | 1 |
| 45 | Female | 0.9 | 0.3 | 189 | 23 | 33 | 6.6 | 3.9 |  | 1 |
| 33 | Male | 2 | 1.4 | 2110 | 48 | 89 | 6.2 | 3 | 0.9 | 1 |
| 51 | Male | 0.8 | 0.2 | 230 | 24 | 46 | 6.5 | 3.1 |  | 1 |
| 35 | Female | 0.6 | 0.2 | 180 | 12 | 15 | 5.2 | 2.7 |  | 2 |
| 27 | Male | 1.3 | 0.6 | 106 | 25 | 54 | 8.5 | 4.8 |  | 2 |
| 55 | Female | 8.2 | 3.9 | 1350 | 52 | 65 | 6.7 | 2.9 | 0.7 | 1 |
| 55 | Female | 10.9 | 5.1 | 1350 | 48 | 57 | 6.4 | 2.3 | 0.5 | 1 |
| 73 | Male | 1.9 | 0.7 | 1750 | 102 | 141 | 5.5 | 2 | 0.5 | 1 |
| 45 | Female | 23.3 | 12.8 | 1550 | 425 | 511 | 7.7 | 3.5 | 0.8 | 1 |
| 58 | Male | 0.9 | 0.2 | 1100 | 25 | 36 | 7.1 | 3.5 | 0.9 | 1 |
| 33 | Male | 7.1 | 3.7 | 196 | 622 | 497 | 6.9 | 3.6 | 1.09 | 1 |
| 33 | Male | 3.4 | 1.6 | 186 | 779 | 844 | 7.3 | 3.2 | 0.7 | 1 |
| 32 | Male | 30.5 | 17.1 | 218 | 39 | 79 | 5.5 | 2.7 | 0.9 | 1 |
| 40 | Male | 30.8 | 18.3 | 285 | 110 | 186 | 7.9 | 2.7 | 0.5 | 1 |
| 46 | Female | 1.4 | 0.4 | 298 | 509 | 623 | 3.6 | 1 | 0.3 | 1 |
| 42 | Male | 11.1 | 6.1 | 214 | 60 | 186 | 6.9 | 2.8 | 2.8 | 1 |
| 26 | Male | 42.8 | 19.7 | 390 | 75 | 138 | 7.5 | 2.6 | 0.5 | 1 |
| 32 | Male | 15.6 | 9.5 | 134 | 54 | 125 | 5.6 | 4 | 2.5 | 1 |
| 32 | Male | 25 | 13.7 | 560 | 41 | 88 | 7.9 | 2.5 | 2.5 | 1 |

Table 7 Outliers Table

# Data Preparation

## Select Data

* Choose the ILPD dataset.
* The ILPD dataset contains liver patient records collected from the Northeast of Andhra Pradesh, India.
* The ILPD dataset contains a total of 583 records, including 416 liver patient records and 167 non-liver patient records.

A screenshot of a computer

Description automatically generated

Figure 4 Class Distribution

## Clean Data

* Replace missing values: The attribute “**A/G Ratio**” has four missing values (1%). Populating the missing value with the mean value.
* Remove duplicates data.
* Normalize data.

## Construct Data

* This dataset is well-formatted, this step is omitted.

## Integrate Data

* This dataset is the unique data source, this step is omitted.

## Format Data

* Change the attribute “**Selector”** from Numeric to Nominal.
* Change the attribute “**Gender**” from Nominal to Binary.
* Any patient whose “**Age”** exceeded 89 is listed as being of age "90".

# Modeling

## KNN

|  |  |  |
| --- | --- | --- |
| K | Percentage of correctly | Confusion Matrix |
| 3 | 64.56% | a b <-- classified as  307 99 | a = 1  103 61 | b = 2 |
| 5 | 64.39% | a b <-- classified as  320 86 | a = 1  117 47 | b = 2 |
| 7 | 65.09% | a b <-- classified as  322 84 | a = 1  115 49 | b = 2 |
| 9 | 66.32% | a b <-- classified as  333 73 | a = 1  119 45 | b = 2 |
| 11 | 66.14% | a b <-- classified as  338 68 | a = 1  125 39 | b = 2 |
| 13 | 65.96% | a b <-- classified as  341 65 | a = 1  129 35 | b = 2 |
| 15 | 66.84% | a b <-- classified as  349 57 | a = 1  132 32 | b = 2 |
| 17 | 68.42% | a b <-- classified as  357 49 | a = 1  131 33 | b = 2 |
| 19 | 68.77% | a b <-- classified as  359 47 | a = 1  131 33 | b = 2 |
| 21 | 69.12% | a b <-- classified as  363 43 | a = 1  133 31 | b = 2 |
| 23 | 68.07% | a b <-- classified as  362 44 | a = 1  138 26 | b = 2 |
| 25 | 68.42% | a b <-- classified as  369 37 | a = 1  143 21 | b = 2 |
| 27 | 68.77% | a b <-- classified as  374 32 | a = 1  146 18 | b = 2 |
| 29 | 69.30% | a b <-- classified as  380 26 | a = 1  149 15 | b = 2 |
| 31 | 70.00% | a b <-- classified as  381 25 | a = 1  146 18 | b = 2 |
| 33 | 70.70% | a b <-- classified as  384 22 | a = 1  145 19 | b = 2 |
| 35 | 70.53% | a b <-- classified as  384 22 | a = 1  146 18 | b = 2 |
| 37 | 69.65% | a b <-- classified as  383 23 | a = 1  150 14 | b = 2 |
| 39 | 70.53% | a b <-- classified as  388 18 | a = 1  150 14 | b = 2 |
| 41 | 70.18% | a b <-- classified as  389 17 | a = 1  153 11 | b = 2 |

Table 8 KNN Classify Table

A graph with a line going up

Description automatically generated

Figure 5 KNN Statistic

Different values of k were tested to find the best k value for classification. The best performance was achieved with **k=33**, achieving an accuracy of **70.70%.**

## Decision Tree

**Cross-Validation = 10, Seed = 10, confidence Factor = 0.25**

|  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- |
| Parameters | | Result | | Parameter | Result | |
| minNumObj | **unpruned** | **Rate of Correctly** | **Confusion Matrix** | **unpruned** | **Rate of Correctly** | **Confusion Matrix** |
| 2 | FALSE | 69.12% | a b <-- classified as  349 57 | a = 1  119 45 | b = 2 | TRUE | 68.77% | a b <-- classified as  346 60 | a = 1  118 46 | b = 2 |
| 3 | FALSE | 69.12% | a b <-- classified as  350 56 | a = 1  120 44 | b = 2 | TRUE | 68.60% | a b <-- classified as  344 62 | a = 1  117 47 | b = 2 |
| 5 | FALSE | 68.77% | a b <-- classified as  354 52 | a = 1  126 38 | b = 2 | TRUE | 68.42% | a b <-- classified as  344 62 | a = 1  118 46 | b = 2 |
| 7 | FALSE | 68.95% | a b <-- classified as  346 60 | a = 1  117 47 | b = 2 | TRUE | 68.77% | a b <-- classified as  337 69 | a = 1  109 55 | b = 2 |
| 9 | FALSE | 68.77% | a b <-- classified as  351 55 | a = 1  123 41 | b = 2 | TRUE | 69.12% | a b <-- classified as  343 63 | a = 1  113 51 | b = 2 |
| 11 | FALSE | 67.72% | a b <-- classified as  338 68 | a = 1  116 48 | b = 2 | TRUE | 67.19% | a b <-- classified as  332 74 | a = 1  113 51 | b = 2 |
| 13 | FALSE | 68.95% | a b <-- classified as  342 64 | a = 1  113 51 | b = 2 | TRUE | 69.12% | a b <-- classified as  341 65 | a = 1  111 53 | b = 2 |
| 15 | FALSE | 68.25% | a b <-- classified as  332 74 | a = 1  107 57 | b = 2 | TRUE | 68.07% | a b <-- classified as  331 75 | a = 1  107 57 | b = 2 |
| 17 | FALSE | 70.35% | a b <-- classified as  347 59 | a = 1  110 54 | b = 2 | TRUE | 69.47% | a b <-- classified as  339 67 | a = 1  107 57 | b = 2 |
| 19 | FALSE | 69.65% | a b <-- classified as  358 48 | a = 1  125 39 | b = 2 | TRUE | 68.60% | a b <-- classified as  347 59 | a = 1  120 44 | b = 2 |
| 21 | FALSE | 67.89% | a b <-- classified as  342 64 | a = 1  119 45 | b = 2 | TRUE | 68.07% | a b <-- classified as  338 68 | a = 1  114 50 | b = 2 |
| 23 | FALSE | 69.82% | a b <-- classified as  346 60 | a = 1  112 52 | b = 2 | TRUE | 70% | a b <-- classified as  343 63 | a = 1  108 56 | b = 2 |
| 25 | FALSE | 71.23% | a b <-- classified as  344 62 | a = 1  102 62 | b = 2 | TRUE | 71.05% | a b <-- classified as  339 67 | a = 1  98 66 | b = 2 |
| 27 | FALSE | 69.82% | a b <-- classified as  340 66 | a = 1  106 58 | b = 2 | TRUE | 70.35% | a b <-- classified as  342 64 | a = 1  105 59 | b = 2 |
| 29 | FALSE | 70.53% | a b <-- classified as  349 57 | a = 1  111 53 | b = 2 | TRUE | 70.18% | a b <-- classified as  343 63 | a = 1  107 57 | b = 2 |
| 31 | FALSE | 69.65% | a b <-- classified as  337 69 | a = 1  104 60 | b = 2 | TRUE | 69.82% | a b <-- classified as  340 66 | a = 1  106 58 | b = 2 |
| 33 | FALSE | 71.58% | a b <-- classified as  348 58 | a = 1  104 60 | b = 2 | TRUE | 71.23% | a b <-- classified as  345 61 | a = 1  103 61 | b = 2 |
| 35 | FALSE | 69.47% | a b <-- classified as  350 56 | a = 1  118 46 | b = 2 | TRUE | 69.47% | a b <-- classified as  348 58 | a = 1  116 48 | b = 2 |
| 37 | FALSE | 71.40% | a b <-- classified as  366 40 | a = 1  123 41 | b = 2 | TRUE | 71.05% | a b <-- classified as  357 49 | a = 1  116 48 | b = 2 |
| 39 | FALSE | 70.88% | a b <-- classified as  370 36 | a = 1  130 34 | b = 2 | TRUE | 70.88% | a b <-- classified as  354 52 | a = 1  114 50 | b = 2 |
| 41 | FALSE | 70.88% | a b <-- classified as  368 38 | a = 1  128 36 | b = 2 | TRUE | 70.88% | a b <-- classified as  352 54 | a = 1  112 52 | b = 2 |

Table 9 J48 Decision Tree Train Table

A graph with lines and dots

Description automatically generated

Figure 6 J48 Train Result Statistic

Various parameters were tested to find the best configuration for the decision tree. The best performance was achieved with a minimum number of objects per leaf (minNumObj) of **33**, achieving an accuracy of **71.58%.**

A screenshot of a computer

Description automatically generated

Figure 7 J48 pruned tree -C 0.25 -M 33, Test mode: 10-fold cross-validation, Seed = 10 ,

The root of the tree is Sgpt. If Sgpt is greater than 0.027638, the sample is classified as 1.

Otherwise, evaluate TB.

If TB is greater than 0.016086, the sample is classified as 1.

Otherwise, evaluate Alkphos.

If Alkphos is greater than 0.072301, the sample is classified as 1.

Otherwise, evaluate DB.

If DB is less than or equal to 0, the sample is classified as 1.

Otherwise, evaluate TB.

If TB is less than or equal to 0.004021, the sample is classified as 2.

Otherwise, evaluate TB again.

If TB is less than or equal to 0.006702, evaluate Age.

If Age is less than or equal to 0.465116, the sample is classified as 2.

Otherwise, the sample is classified as 1.

If TB is greater than 0.006702, the sample is classified as 2.

**A diagram of a tree

Description automatically generated**

Figure 8 J48 unpruned tree, 10-fold cross-validation, Seed = 10, -M=33

The root of the tree is DB. If DB is greater than 0.045918, the sample is classified as 1.

Otherwise, evaluate Sgpt.

If Sgpt is greater than 0.027638, the sample is classified as 1.

Otherwise, evaluate TB.

If TB is greater than 0.016086, the sample is classified as 1.

Otherwise, evaluate Alkphos.

If Alkphos is greater than 0.072301, the sample is classified as 1.

Otherwise, evaluate DB.

If DB is less than or equal to 0, the sample is classified as 1.

Otherwise, evaluate TB.

If TB is less than or equal to 0.004021, the sample is classified as 2.

Otherwise, evaluate TB again.

If TB is less than or equal to 0.006702, evaluate Age.

If Age is less than or equal to 0.465116, the sample is classified as 2.

Otherwise, the sample is classified as 1.

If TB is greater than 0.006702, the sample is classified as 2.

# Comparison of Results

KNN: K = 33

Decision Tree: Cross-Validation = 10, Seed = 10, confidence Factor = 0.25, minNumObj =33, unpruned = false

|  |  |  |
| --- | --- | --- |
| Metric | KNN | Decision Tree |
| Accuracy | 0.707 | 0.716 |
| Precision | 0.650 | 0.695 |
| Sensitivity | 0.707 | 0.716 |
| Specificity | 0.355 | 0.507 |
| F1 Measure | 0.638 | 0.700 |

Table 10 Metric Compare

**Accuracy:** The Decision Tree (J48) classifier has a slightly higher accuracy (0.716) compared to the KNN accuracy (0.707).

**Precision:** The Decision Tree (J48) classifier also has a higher precision (0.695) compared to the KNN precision (0.650).

**Sensitivity**: The Decision Tree (J48) classifier has a higher sensitivity (0.716) compared to the KNN sensitivity (0.707).

**Specificity**: The Decision Tree (J48) classifier has a higher specificity (0.507) compared to the KNN specificity (0.355).

**F1 Measure**: The Decision Tree (J48) classifier has a higher F1 measure (0.700) compared to the KNN classifier (0.638).

# Conclusion

Based on the above metrics, the Decision Tree (J48) classifier generally outperforms the KNN classifier in terms of accuracy, precision, sensitivity, specificity, and F1 measure. This indicates that the Decision Tree (J48) classifier is better.

# Reference

[1] Ramana,Bendi and Venkateswarlu,N.. (2012). ILPD (Indian Liver Patient Dataset). UCI Machine Learning Repository. https://doi.org/10.24432/C5D02C.