



Министерство науки и высшего образования Российской Федерации
Калужский филиал
федерального государственного бюджетного
образовательного учреждения высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(КФ МГТУ им. Н.Э. Баумана)

ФАКУЛЬТЕТ ИУК «Информатика и управление»

КАФЕДРА ИУК4 «Программная инженерия»

Лабораторная работа №4

«Классификация и кластеризация в среде WEKA»

**ДИСЦИПЛИНА: «Интеллектуальные информационные системы
анализа данных»**

Выполнил: студент гр. ИУК4-21М _____ (Сафронов Н.С.)
(подпись) (Ф.И.О.)

Проверил: _____ (Белов Ю.С.)
(подпись) (Ф.И.О.)

Дата сдачи (защиты):

Результаты сдачи (защиты):

- Балльная оценка:

- Оценка:

Калуга, 2025

Цель работы: приобретение практических навыков по обучению искусственной нейронной сети с использованием различных библиотек.

Задание:

Используя готовые коллекции наборов данных произвести классификацию и кластеризацию данных. Для классификации предварительно разбить исходный датасет на 2 части. На основе полученных данных сделать выводы. Результаты визуализировать с помощью встроенного инструментария WEKA

Вариант 7

Разработать и произвести классификацию, кластеризацию датасета больных гепатитом. Использовать файл hepatitis.arff.

Результат выполнения работы

The screenshot shows the Weka Explorer interface with the 'Classify' tab selected. The classifier chosen is 'J48 - C 0.25 - M 2'. The test options are set to 'Cross-validation' with 10 folds. The classifier output is displayed, showing a summary of performance metrics and a detailed accuracy table.

Classifier output

Number of Leaves : 12
Size of the tree : 23
Time taken to build model: 0.01 seconds

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

Metric	Value	Percentage
Correctly Classified Instances	125	80.6452 %
Incorrectly Classified Instances	30	19.3548 %
Kappa statistic	0.3807	
Mean absolute error	0.218	
Root mean squared error	0.4203	
Relative absolute error	66.0075 %	
Root relative squared error	103.8002 %	
Total Number of Instances	155	

=== Detailed Accuracy By Class ===

	TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
	0,469	0,106	0,536	0,469	0,500	0,382	0,692	0,433	DIE
	0,894	0,531	0,866	0,894	0,880	0,382	0,692	0,846	LIVE
Weighted Avg.	0,806	0,443	0,798	0,806	0,802	0,382	0,692	0,760	

=== Confusion Matrix ===

```
a  b  <-- classified as
15  17 |  a = DIE
13  110 |  b = LIVE
```

Рисунок 1 – Классификация с использованием классификатора J48

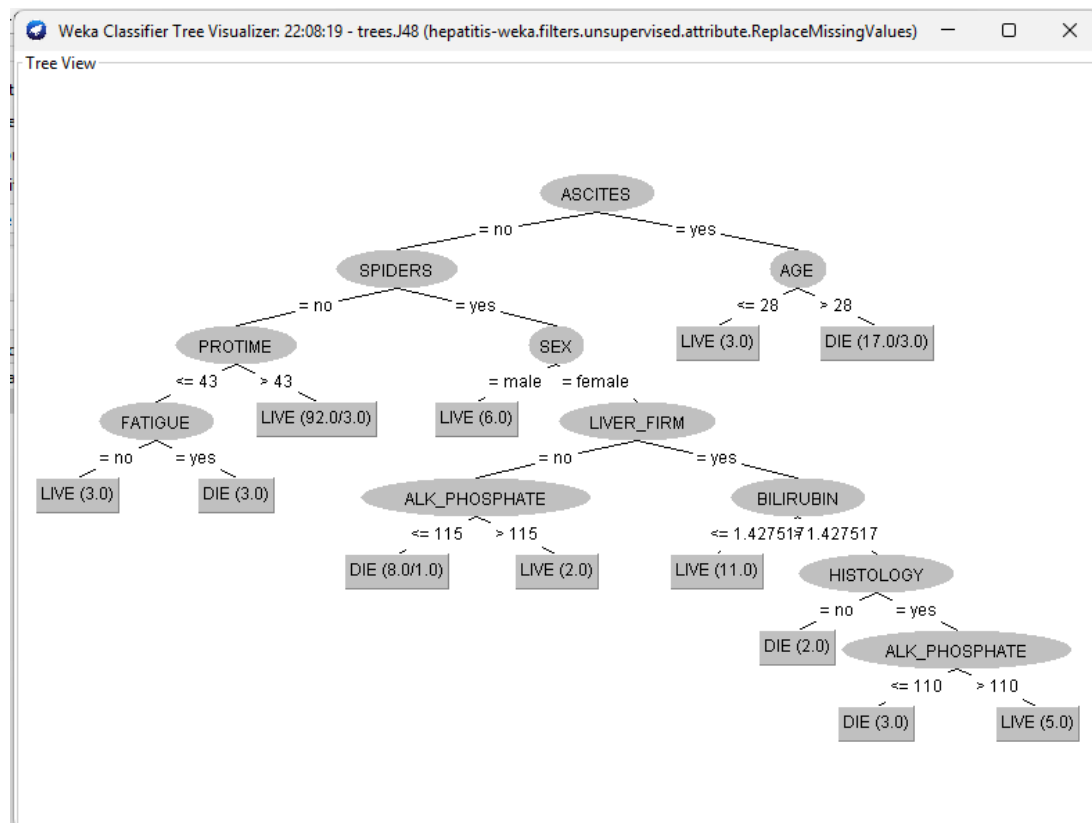


Рисунок 2 – Визуализация дерева признаков

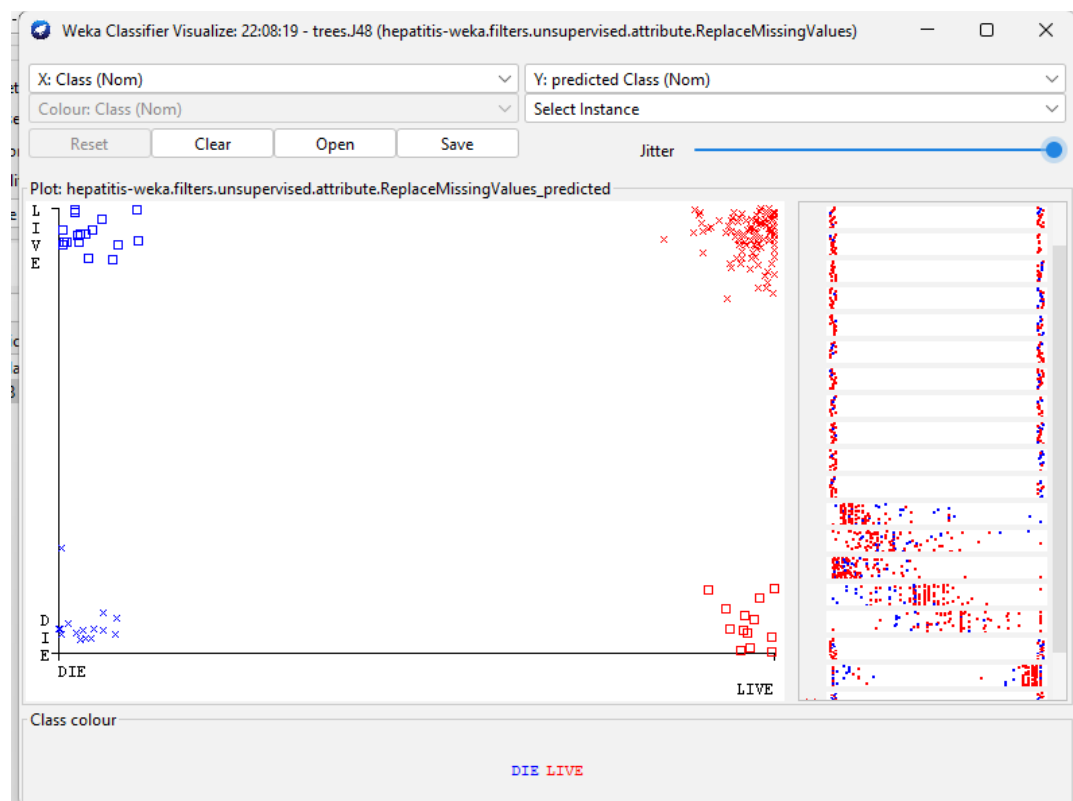


Рисунок 3 – Визуализация ошибки классификации

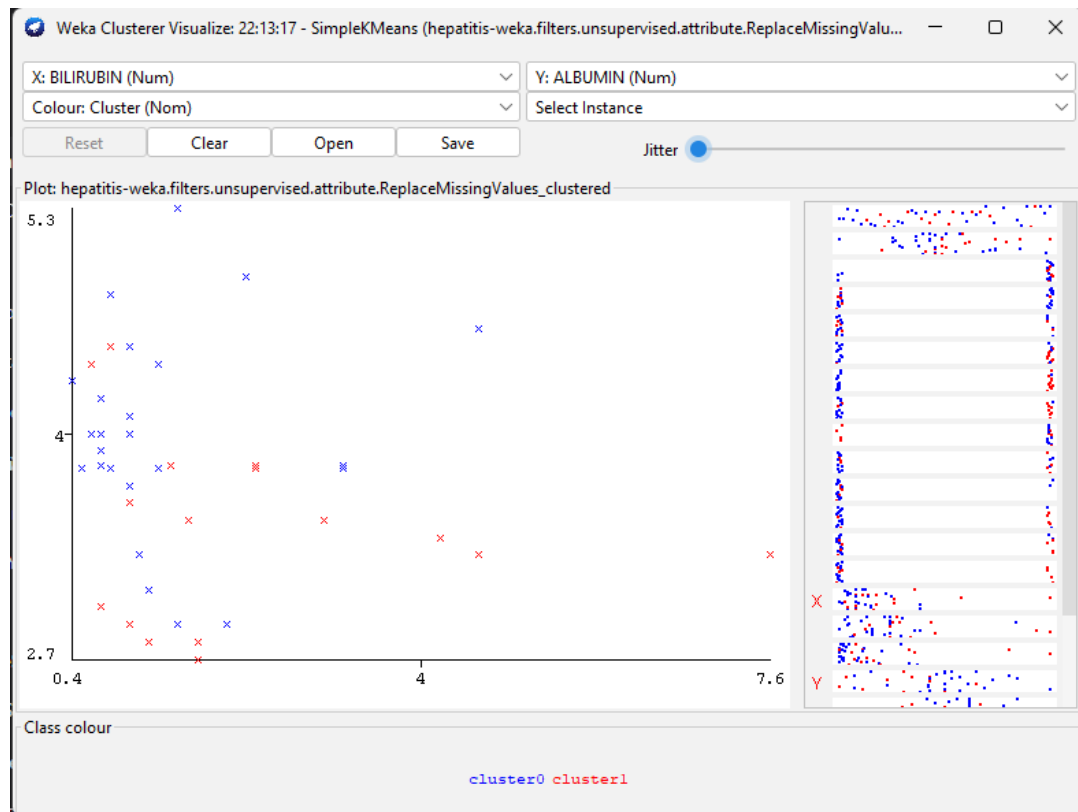


Рисунок 4 – Визуализация кластеризованных меток

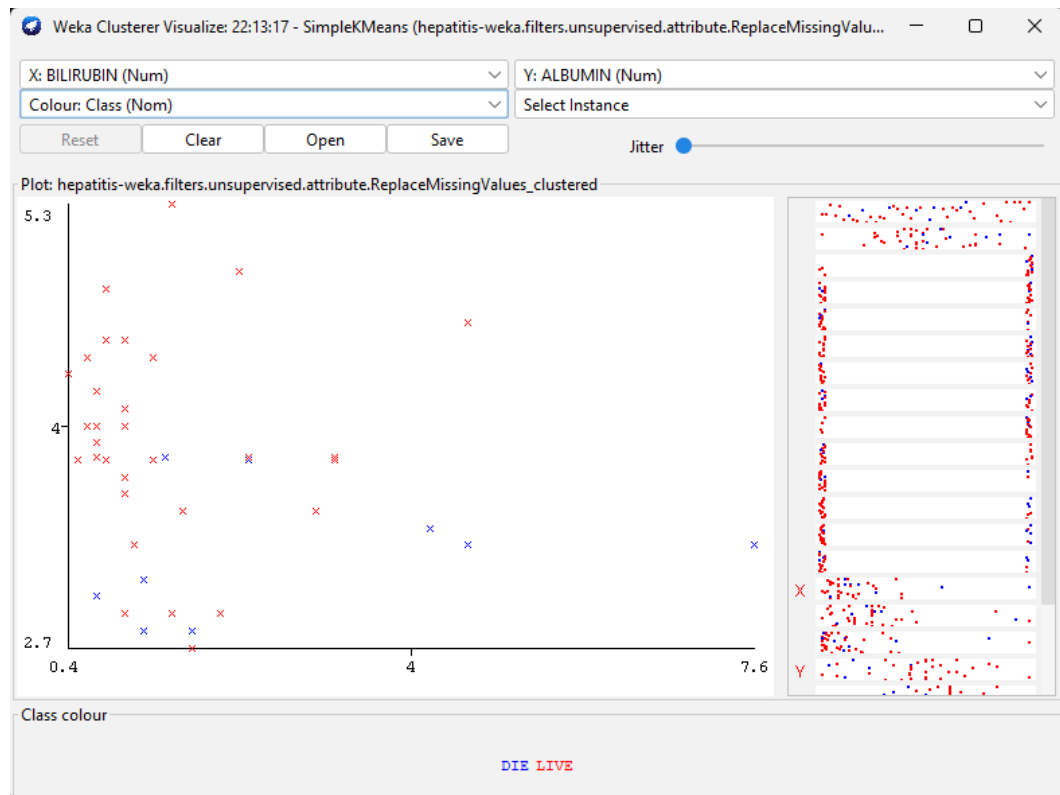


Рисунок 5 – Визуализация действительных меток

Выводы: в ходе выполнения работы были сформированы практические навыки по обучению искусственной нейронной сети с использованием различных библиотек.