

ГУАП

КАФЕДРА № 42

ОТЧЕТ
ЗАЩИЩЕН С ОЦЕНКОЙ _____

ПРЕПОДАВАТЕЛЬ

профессор, д-р.т.н., профессор				В. В. Фомин
должность, уч. степень, звание		подпись, дата		инициалы, фамилия

ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ №4

МЕТОД НЕЙРОННЫХ СЕТЕЙ

Вариант 5

по курсу: МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА

РАБОТУ ВЫПОЛНИЛ

СТУДЕНТ ГР. №	4128			Воробьев В. А.
			подпись, дата	инициалы, фамилия

Санкт-Петербург 2024

СОДЕРЖАНИЕ

1	Введение	3
1.1	Цель лабораторной работы	3
1.2	Задание	3
2	Выполнение работы	4
2.1	Набор данных	4
2.2	Рабочий процесс	4
3	Вывод	10

1 Введение

1.1 Цель лабораторной работы

Изучение основ организация работы с технологической платформой для создания законченных аналитических решений использованием метода нейронных сетей.

1.2 Задание

1. Для набора данных выполнить классификацию с помощью метода нейронных сетей.
2. Выполнить оценку качества классификации.

2 Выполнение работы

2.1 Набор данных

Набор данных взят с Kaggle (URI - <https://www.kaggle.com/datasets/sudhanshu2198/wheat-variety-classification>).

Набор данных включает зерна пшеницы, принадлежащие к трем различным сортам пшеницы: **Кама**, **Роза** и **Канадская**, по 70 элементов каждый.

Для построения данных были измерены семь геометрических параметров зерен пшеницы:

- 1) Область — размер поверхности зерна пшеницы.
- 2) Периметр — общая длина внешней границы зерна.
- 3) Компактность — насколько форма зерна близка к идеальной круговой.
- 4) Длина ядра — измерение самой длинной оси внутренней части зерна пшеницы.
- 5) Ширина ядра — поперечное измерение внутренней части зерна.
- 6) Коэффициент асимметрии — отклонение формы зерна от симметричной.
- 7) Длина бороздки ядра — протяженность центральной линии или углубления в зерне.

Для каждого этого параметра был сопоставлен сорт пшеницы:

- **Кама** — сорт пшеницы, известный своей устойчивостью к болезням и приспособленностью к различным климатическим условиям.
- **Роза** — сорт пшеницы, который ценится за качество зерна и применяется для муки высшего сорта.
- **Канадская** — сорт пшеницы с высоким содержанием белка, используемый для производства высококачественной муки.

2.2 Рабочий процесс

Целью создания данной системы является проверка гипотезы, что вышеуказанных 7 параметров достаточно для определения сорта пшеницы. Гипотезу будем считать доказанной, если точность составит 95%.

Для создания модели в программе KNIME создаём следующие узлы:

- Excel Reader для считывания файла;

- Number to String для преобразования номера сорта пшеницы в строку.
- String Manipulation для сопоставления номера сорта с его названием.
- Color Manager для цветового разделения на графике;
- Partitioning для разделения данных на обучающие и тестовые(50/50). Дополнительно выбран Linear Sampling, так как набор данных отсортирован по сорту пшеницы;
- RProp MLP Learner для обучения модели;
- MultiLayerPerceptron Predictor непосредственно для предсказания;
- Scorer для вычисления статистики;

На рисунке 1 представлена схема рабочего процесса.

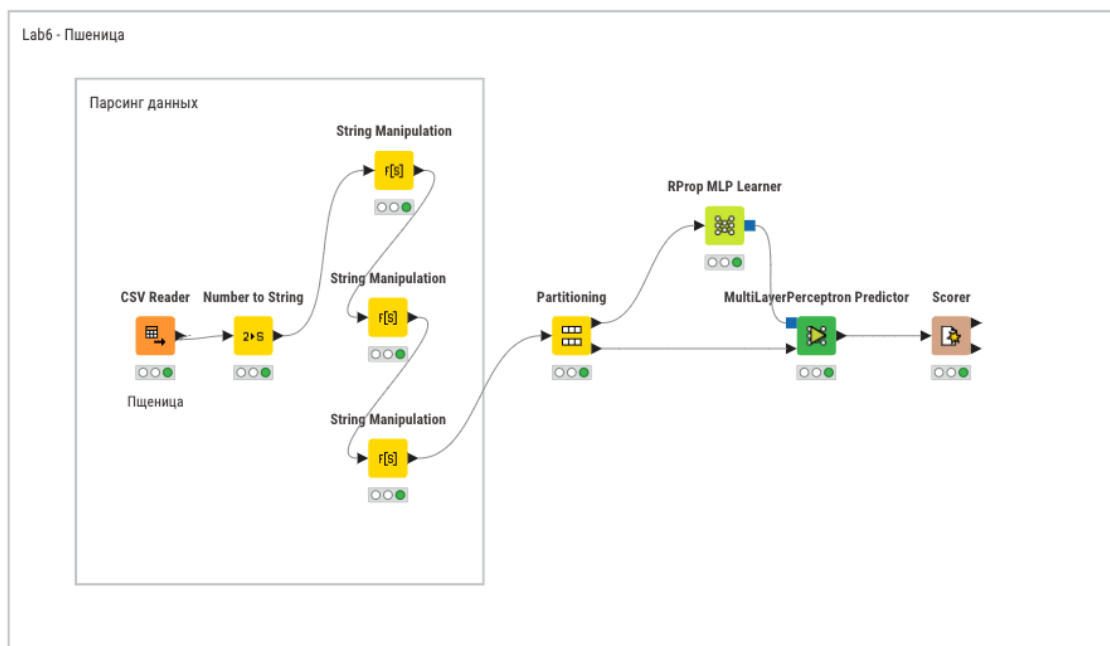


Рисунок 2.1 - Схема в KNIME

Так как количество параметров равно 7, было принято решение увеличить число слоев до 2, а количество нейронов на слое до 20.

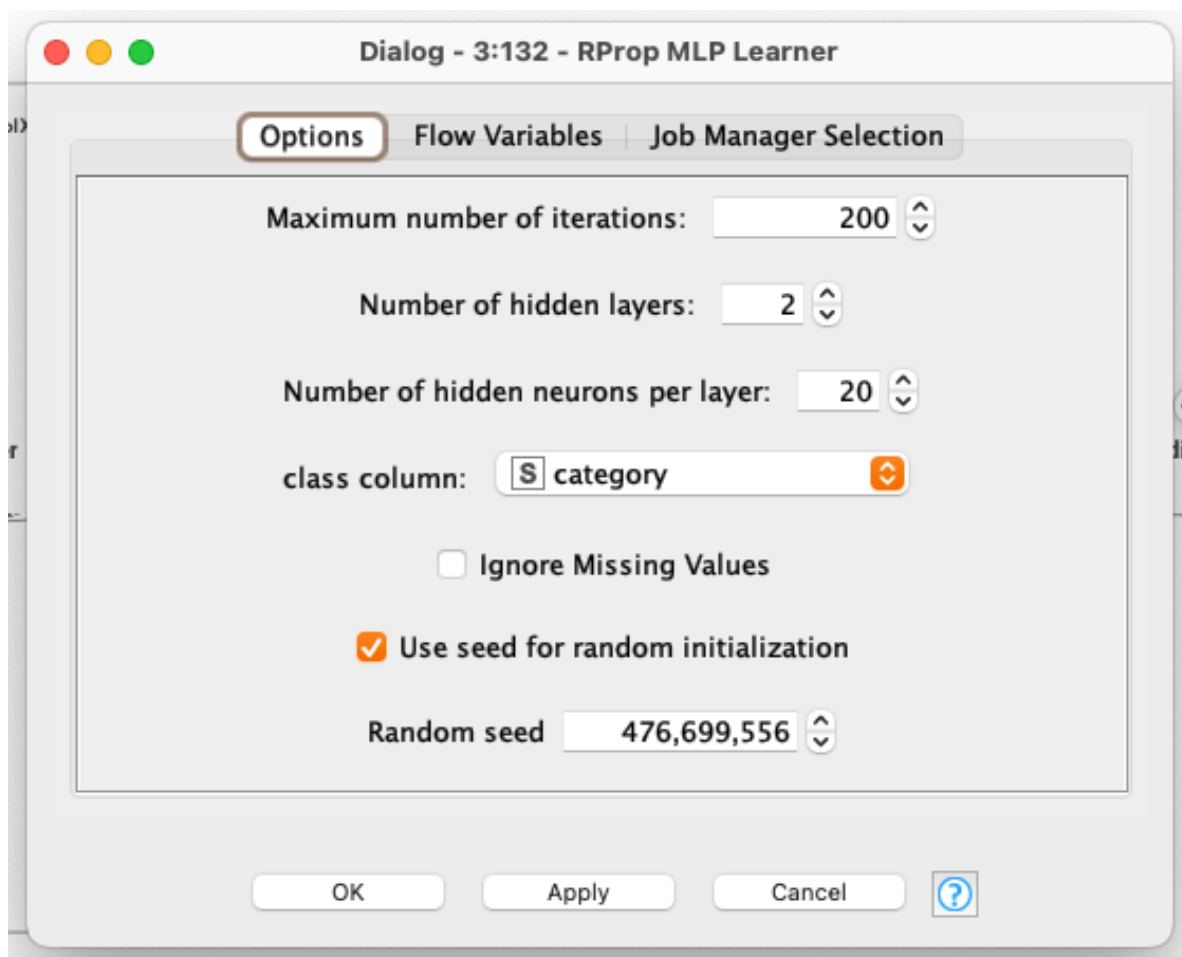


Рисунок 2.2 - Настройки узла

На рисунке 3 представлен фрагмент набора данных с вычисленными вероятностями появления классов на основе обученной модели.

Classified Data (Table)

Rows: 105

Columns: 12

<

Рисунок 2.3 - Вероятность появления классов

В результате обучения нейронной сети построен график распределения ошибки на итерациях, демонстрирующий постепенное обучение модели, минимизирующее ошибку. Этот график представлен на рисунке 4.

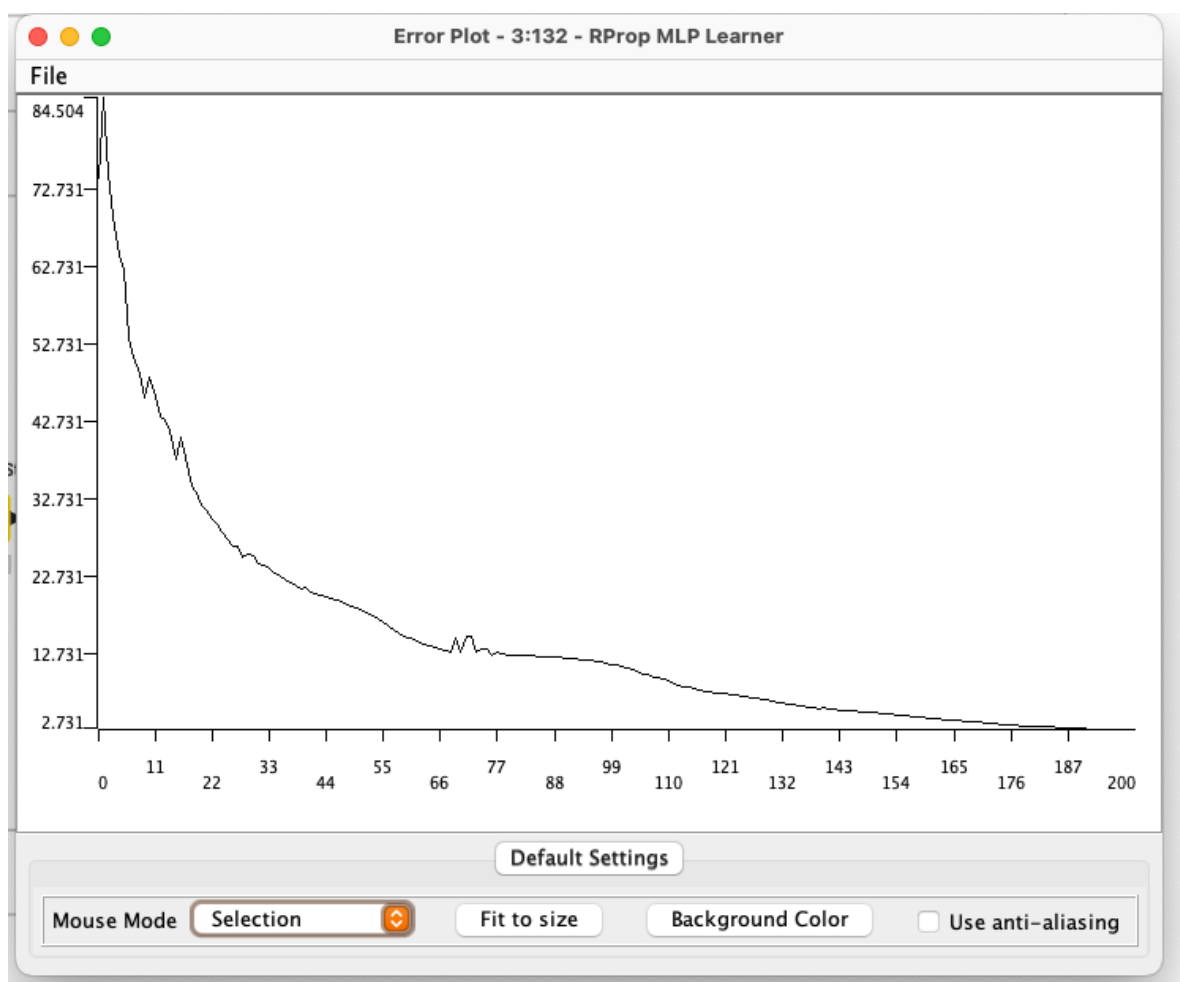


Рисунок 2.4 - Процент ошибки на N итерации

Также были получены метрики оценки качества.

Confusion Matrix - 3:127 - Scorer				
File	Hilite			
category \ ...	Kama	Rosa	Canadian	
Kama	34	0	0	
Rosa	1	34	0	
Canadian	0	0	36	

Correct classified: 104	Wrong classified: 1
Accuracy: 99.048%	Error: 0.952%
Cohen's kappa (κ): 0.986%	

Рисунок 2.5 - Матрица сопряженности

Rows: 4 | Columns: 11

Table Statistics

#	RowID	TruePositives Number (integer)	FalsePositives Number (integer)	TrueNegatives Number (integer)	FalseNegatives Number (integer)	Recall Number (double)	Precision Number (double)	Sensitivity Number (double)	Specificity Number (double)	F-measure Number (double)	Accuracy Number (double)	Cohen's kappa Number (double)
1	Kama	34	1	70	0	1	0.971	1	0.986	0.986	0	0
2	Rosa	34	0	70	1	0.971	1	0.971	1	0.986	0	0
3	Cana...	36	0	69	0	1	1	1	1	1	0	0
4	Overall	0	0	0	0	0	0	0	0	0	0.99	0.986

Рисунок 2.6 - Метрики оценки качества

Из метрик оценки качества видно, что модель работает безошибочно, за исключением очередного ложноположительного срабатывания сорта Камы на сорте Роза.

Из 104 записей точность составила 99.048%.

3 Вывод

Полученная точность составляет 99.048%. Что является самым высоким среди всех методов. Одно ложноположительное срабатывание можно оправдать маленьким размером выборки и схожестью сорта Камы и Роза. Полученная модель позволяет доказать гипотезу о возможности однозначно определить сорт пшеницы на основе 7 параметров.

В результате выполнения всех лабораторных работ также было выявлено, что для данного набора данных оптимальным решением является метод нейронных сетей и дерево решений.