ГУАП

КАФЕДРА № 42

ОТЧЕТ ЗАЩИЩЕН С ОЦЕНКОЙ									
ПРЕПОДАВАТЕЛЬ									
профессор, д-р.т.н., профессор должность, уч. степень, звание	подпись, дата	В. В. Фомин инициалы, фамилия							
ОТЧЕТ О ЛАБОРАТОРНОЙ РАБОТЕ №6									
МЕТОД МАШИННЫХ ОПОРНЫХ ВЕКТОРОВ (SVM)									
Вариант 5									
по курсу: МЕТОД	Ы ИСКУССТВЕННОГО	ИНТЕЛЛЕКТА							

РАБОТУ ВЫПОЛНИЛ

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

СОДЕРЖАНИЕ

1	Вве	дение	3
	1.1	Цель лабораторной работы	3
	1.2	Задание	3
2	Выі	полнение работы	4
	2.1	Набор данных	4
	2.2	Рабочий процесс	4
3	Вы	вод	8

1 Введение

1.1 Цель лабораторной работы

Изучение основ организация работы с технологической платформой для создания законченных аналитических решений использованием метода машинных опорных векторов.

1.2 Задание

- 1. Для набора данных выполнить классификацию с помощью метода машинных опорных векторов.
- 2. Выполнить оценку качества классификации.

2 Выполнение работы

2.1 Набор данных

Hабор 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;
- SVM Learner для обучения модели;
- SVM Predictor непосредственно для предсказания;
- Scorer для вычисления статистики;

На рисунке 2.1 представлена схема рабочего процесса.

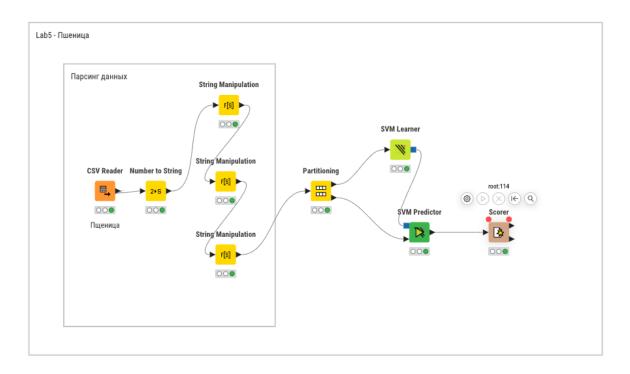


Рисунок 2.1 - Схема в KNIME

После выполнения процесса были получены: фрагмент опорных векторов, матрица смежности и метрики оценки качества.

```
SVM View - 3:121 - SVM Learner
File

SVM O Class: Kama
Support Vectors:
15.26, 14.84, 0.871, 5.763, 3.312, 2.221, 5.22, Kama
14.88, 14.57, 0.8811, 5.554, 3.333, 1.018, 4.956, Kama
13.84, 13.94, 0.8951, 5.386, 3.312, 2.462, 4.956, Kama
14.38, 14.21, 0.8951, 5.386, 3.312, 2.462, 4.956, Kama
14.38, 14.21, 0.8951, 5.386, 3.312, 2.462, 4.956, Kama
13.78, 14.06, 0.8759, 5.479, 3.156, 3.136, 4.872, Kama
14.59, 14.28, 0.8993, 5.351, 3.333, 1.818, 4.781, Kama
15.69, 14.75, 0.9058, 5.527, 3.514, 1.599, 5.046, Kama
15.72, 13.57, 0.8666, 5.226, 3.049, 4.102, 4.914, Kama
12.08, 13.23, 0.8664, 5.099, 2.936, 1.415, 4.961, Kama
12.08, 13.23, 0.8664, 5.099, 2.936, 1.415, 4.961, Kama
12.74, 13.76, 0.8564, 5.395, 2.956, 2.504, 4.869, Kama
13.45, 14.02, 0.8604, 5.516, 3.065, 3.531, 5.097, Kama
13.45, 14.02, 0.8604, 5.516, 3.065, 3.531, 5.097, Kama
14.28, 14.17, 0.8944, 5.397, 3.298, 6.685, 5.001, Kama
14.28, 14.17, 0.8944, 5.397, 3.298, 6.685, 5.001, Kama
13.5, 13.86, 0.8852, 5.351, 3.158, 2.249, 5.176, Kama
13.5, 13.86, 0.8852, 5.351, 3.158, 2.249, 5.176, Kama
13.8, 14.04, 0.8794, 5.376, 3.155, 1.56, 4.961, Kama
14.99, 14.56, 0.8883, 5.877, 3.376, 4.711, 5.528, Kama
13.8, 14.04, 0.8794, 5.376, 3.155, 1.56, 4.961, Kama
14.93, 14.43, 0.9006, 5.5678, 3.258, 2.129, 5.351, Kama
14.86, 14.67, 0.8676, 5.678, 3.258, 2.129, 5.351, Kama
14.92, 14.43, 0.9006, 5.384, 3.412, 1.142, 5.088, Kama
12.11, 13.47, 0.8392, 5.159, 3.032, 1.502, 4.519, Kama
14.92, 14.43, 0.9006, 5.384, 3.412, 1.142, 5.088, Kama
12.11, 13.47, 0.8392, 5.159, 3.032, 1.502, 4.519, Kama
14.92, 14.44, 4.0879, 5.609, 3.158, 2.217, 5.132, Kama
14.92, 14.44, 0.879, 6.384, 3.412, 1.425, 0.88, Kama
12.11, 13.67, 0.868, 5.395, 3.07, 4.157, 5.088, Kama
12.12, 15.50, 6.889, 5.879, 3.3032, 1.502, 4.519, Kama
18.22, 13.84, 0.868, 5.395, 3.07, 4.157, 5.088, Kama
12.13, 12.66, 0.889, 5.879, 3.694, 2.268, 5.795, Rosa
10.71, 17.23, 0.8763, 6.579, 3.814, 4.457, 5.879, Rosa
10.77, 15.62, 0.8648, 6.139, 3.694, 2.268, 5.837, Rosa
18.87, 16.61, 0.8871, 6.603, 3.738, 6.6001, 5.877, Rosa
18.89, 16.66, 0.859, 6.549, 3.677, 3.6
                         SVM 0 Class: Kama
```

Рисунок 2.2 - Фрагмент опорных векторов

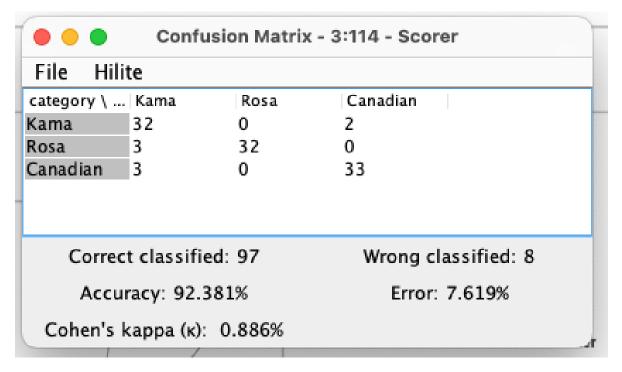


Рисунок 2.3 - Матрица смежности



Рисунок 2.4 - Метрики оценки качества

Из метрик оценки качества следует, что определение сорта Розы на 100% верное, тем не менее сама полнота определения сорта ~ 0.941. Доля ошибок в определении всех сортов составляет 7.619%. Сорт Кама в очередной раз показывает наибольшее количество ложноположительных срабатываний при исходном сорте Роза. К тому же сорт Кама обладает самой низкой точностью равной 0.842.

3 Вывод

Полученная точность составляет 92.381%, что чуть хуже, чем у наивного подхода Байеса и приближено к точности К-ближайших соседей. Тем не менее точность недостаточна для подтверждения гипотезы.