

**ИСПОЛЬЗОВАНИЕ ВОЗМОЖНОСТЕЙ ЯЗЫКА R ДЛЯ РЕАЛИЗАЦИИ
АЛГОРИТМОВ МАШИННОГО ОБУЧЕНИЯ В СРЕДЕ MS SQL SERVER
2019**

Крумкаченко Артём Андреевич

студент

Мичуринский государственный аграрный университет
г. Мичуринск, Россия

Косенков Дмитрий Владимирович

студент

Мичуринский государственный аграрный университет
г. Мичуринск, Россия

Гавриков Владимир Витальевич

студент

Мичуринский государственный аграрный университет
г. Мичуринск, Россия

Абалуев Роман Николаевич

кандидат педагогических наук, доцент

Мичуринский государственный аграрный университет
г. Мичуринск, Россия

Аннотация: В статье проводится обзор современных методов анализа данных средствами языка программирования R интегрированного в систему управления базами данных MS SQL Server 2019.

Ключевые слова: Машинное обучения, системы управления базами данных, информационные системы, хранимая процедура, язык R.

Модернизация аграрной отрасли экономики и переход к инновационной модели развития требуют внедрения прогрессивных цифровых технологий, масштабных управленческих решений и корпоративных стратегий. Современное сельскохозяйственное предприятие функционирует в постоянно

изменяющихся условиях, при планировании и организации его деятельности необходимо учитывать неопределенность и динамичные изменения, происходящие во внешней среде. Для предприятий важно иметь возможность прогнозировать результаты своей деятельности используя накопленную информацию. Для решения такого рода задач могут использоваться службы машинного обучения на базе MS SQL Server 2019.

На сегодняшний день одной из наиболее мощных систем работы с базами данных на основе архитектуры "клиент-сервер" является СУБД MS SQL Server, которая в своем составе имеет средства создания баз данных, обработки информации, импорта и экспорта данных, разграничения доступа к информации, резервного копирования и восстановления информации, оптимизации и выполнения запросов к данным, которые сейчас активно используются для информатизации различных отраслей в том числе и сельскохозяйственными предприятиями [1].

Под машинным обучением понимают класс методов искусственного интеллекта, характерной чертой которых является не прямое решение задачи, а обучение в процессе применения решений множества сходных задач. Для построения таких методов используются средства математической статистики, численных методов, методов оптимизации, теории вероятностей, теории графов, различные техники работы с данными в цифровой форме.

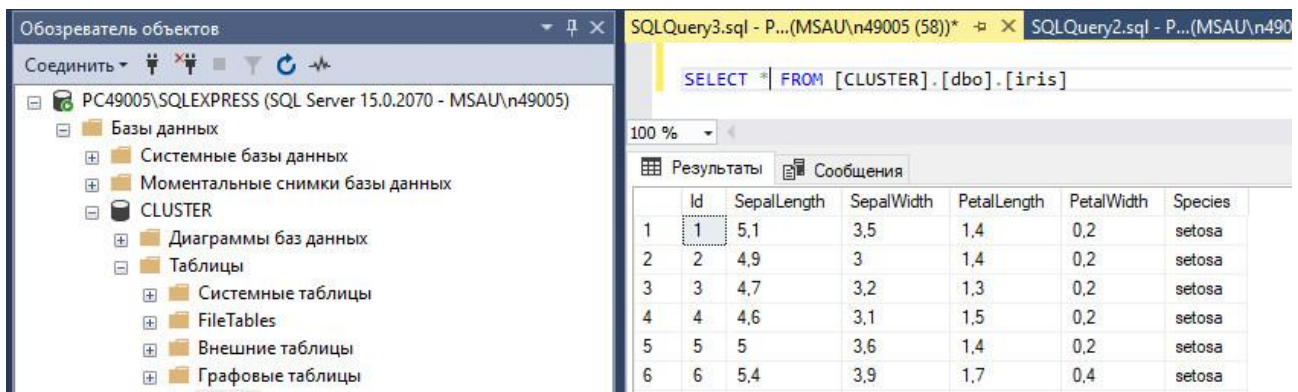
Службы машинного обучения MS SQL Server 2019 позволяют выполнять скрипты языков Python и R непосредственно в системе управления базой данных, что позволяет генерировать новые данные и аналитику без увеличения транзакционной рабочей нагрузки.

В данной статье рассмотрим использование возможностей языка R для классификации данных, последовательно выполнив загрузку исходных данных из таблицы базы данных MS SQL Server 2019 с последующей их обработкой и визуализацией.

В качестве исходных данных был взят набор данных [2, 4], который содержит описание трех видов ирисов (лат. *Íris*) — род многолетних

корневищных растений семейства Ирисовые, или Касатиковые (Iridaceae) и состоит из 5 переменных, первые четыре – геометрические размеры цветка и последняя – название вида.

Исходные данные (150 строк), находятся в таблице iris базы данных CLUSTER (Рис. 1). В столбце Id содержится идентификационный номер объекта наблюдения, в столбцах SepalLength, SepalWidth, PetalLength, PetalWidth содержатся геометрические размеры цветка для каждого объекта наблюдения, в столбце Species название вида для каждого объекта.



	Id	SepalLength	SepalWidth	PetalLength	PetalWidth	Species
1	1	5,1	3,5	1,4	0,2	setosa
2	2	4,9	3	1,4	0,2	setosa
3	3	4,7	3,2	1,3	0,2	setosa
4	4	4,6	3,1	1,5	0,2	setosa
5	5	5	3,6	1,4	0,2	setosa
6	6	5,4	3,9	1,7	0,4	setosa

Рисунок 1. Фрагмент таблицы массива исходных данных в базе данных MS SQL Server.

Для классификации будем использовать функцию fanny из библиотеки cluster языка R. Данная функция выполняет нечеткую кластеризацию массива данных x на заданное количество кластеров k [3, 4]:

```
fanny(x, k, diss = inherits(x, "dist"), memb.exp = 2,  
      metric = c("euclidean", "manhattan", "SqEuclidean"),  
      stand = FALSE, iniMem.p = NULL, cluster.only = FALSE,  
      keep.diss = !diss && !cluster.only && n < 100,  
      keep.data = !diss && !cluster.only,  
      maxit = 500, tol = 1e-15, trace.lev = 0)
```

В среде SQL Server Management Studio был создан скрипт, который используя в качестве входных данных выборку из таблицы iris, производит классификацию объектов на три кластера. Результатом работы скрипта является набор данных, где для каждого объекта вычисляется степень принадлежности к каждому кластеру [5]. Скрипт и фрагменты массива выходных данных приведены на Рис. 2.

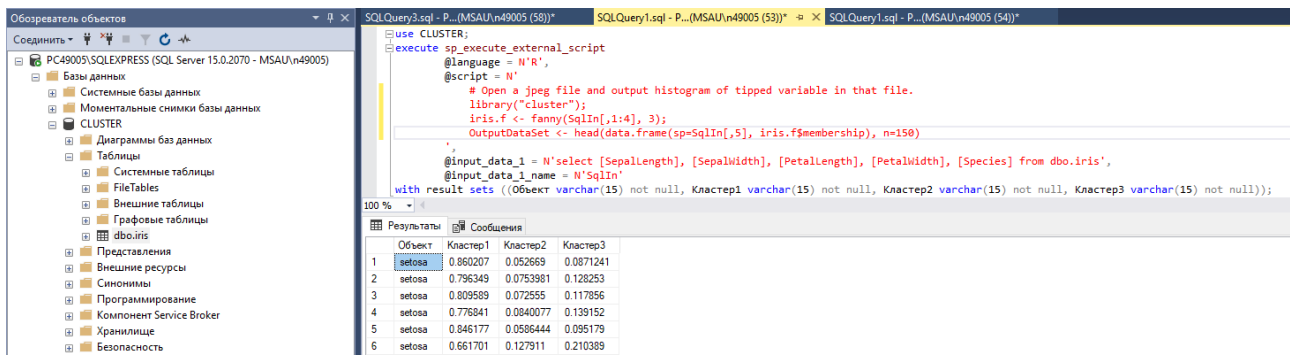


Рисунок 2. Скрипт реализующий нечеткую кластеризацию в таблице БД с результатами выполнения.

Язык R обладает развитыми средствами визуализации, которые можно использовать также и в среде MS SQL Server [6, 7]. Дополнив скрипт строками, где указывается функция визуализации, тип графического файла и путь его сохранения, можно получить графическую интерпретацию результатов кластеризации (Рис.3):

```
library("cluster");
iris.f <- fanny(SqlIn[,1:4], 3);
dest_filename = tempfile(pattern = "rGrapfFanny_", tmpdir = "D:\\tmp\\plot");
dest_filename = paste(dest_filename, '.jpg', sep="");
jpeg(filename=dest_filename);
plot(iris.f, which=1);
dev.off();
```

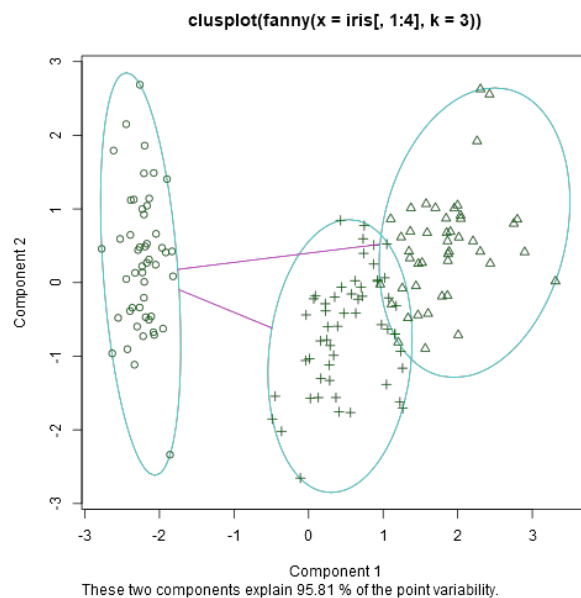


Рисунок 3. Графическая интерпретация результатов кластеризации.

В данной работе мы рассмотрели возможности новой версии СУБД MS SQL Server 2019 с интегрированными службами машинного обучения, создали и апробировали скрипт на языке R для нечеткой классификации информации

базы данных. В дальнейшем данный скрипт можно использовать для классификации данных хранящихся в различных СУБД реализованных на базе платформы MS SQL Server.

Список литературы

1. Абалуев Р.Н. Машинное обучение в среде СУБД MS SQL SERVER / Р.Н. Абалуев, А.А. Крумкаченко // Наука и Образование – 2019. – №4. – С. 52.
2. Fisher, R. A. The use of multiple measurements in taxonomic problems. *Annals of Eugenics*, 1936-7, Part II, 179–188.
3. Иерархический анализ экспериментальных данных / Л.В. Бобрович, Н.В. Картечина, Н.В. Андреева, С.О Чиркин. // Наука и Образование (Мичуринск). – 2019. – № 3. – С. 2.
4. Некоторые возможности применения Mathcad для решения инженерных задач в АПК / О.С. Дьячкова, С.В. Дьячков, О.С. Картечина, Н.В. Картечина // Наука и Образование (Мичуринск). – 2019. – № 4. – С. 203.
5. Абалуев Р.Н. Анализ и оценка материалов для 3d-печати с использованием технологии лазерной стереолитографии / С.О. Чиркин, Р.Н. Абалуев //Наука и Образование (Мичуринск) . – 2019. – № 4. – С. 131.
6. Абалуев Р.Н. Обзор современных подходов к обеспечению информационной безопасности при создании инфраструктуры интернета вещей в агропромышленном комплексе / Р.Н. Абалуев, А.А. Крумкаченко // Наука и Образование (Мичуринск). - 2019. – № 2. – С. 289.
7. Абалуев Р.Н. Информационное обеспечение сельского хозяйства / Р.Н. Абалуев, Д.В. Косенков // Наука и Образование (Мичуринск). - 2019. – № 2. – С. 290.

USING THE CAPABILITIES OF THE R LANGUAGE TO IMPLEMENT MACHINE LEARNING ALGORITHMS IN MS SQL SERVER 2019

Krumkachenko Artem Andreevich,

student

Michurinsk state agrarian University

Michurinsk, Russia

krumkachenko@mail.ru

Kosenkov Dmitry Vladimirovich

student

Michurinsk state agrarian University

Michurinsk, Russia

mr.wow33@mail.ru

Gavrikov Vladimir Vitalievich,

student

Michurinsk state agrarian University

Michurinsk, Russia

gvv_kk1998@gmail.com

Abaluev Roman Nikolaevich,

candidate of pedagogics, Associate Professor

Michurinsk state agrarian University

Michurinsk, Russia

abaluevrn@mgau.ru

Abstract: The article provides an overview of modern methods of data analysis using the programming language R integrated into the MS SQL Server 2019 database management system.

Keywords: machine learning, database management systems, information systems, stored procedure, language R.