

СЕМЕРИКОВ А. В., ГЛАЗЫРИН М. А.
ПЛАТФОРМА LOGINOM ДЛЯ ПОСТРОЕНИЯ НЕЙРОННОЙ СЕТИ ПО
ПРОГНОЗИРОВАНИЮ ИТОГОВОГО РЕЗУЛЬТАТА ОБУЧЕНИЯ
СТУДЕНТА УНИВЕРСИТЕТА

УДК 378.141.21:330.47, ВАК 1.2.2, ГРНТИ 28.17.31

Платформа Loginom для построения нейронной сети по прогнозированию итогового результата обучения студента университета

Loginom platform for building a neural network to predict the final learning outcome of a university student

А. В. Семериков¹, М. А. Глазырин²

A.V. Semerikov¹, M.A. Glazyrin²

¹Ухтинский государственный технический университет, г. Ухта

¹Ukhta State Technical University, Ukhta

²Вятский государственный университет, г. Киров

²Vyatka State University, Kirov

В статье описано построение нейронной сети для прогнозирования результата успешности завершения обучения потенциальным студентом университета. При построении, обучении и анализе нейронной сети использовалась система Loginom. В качестве исходных данных использовались результаты ЕГЭ, представленные в формате MS Excel. Количество примеров составляет 10311 с пятью параметрами. Целевая функция принимает бинарные значения. Представлен пример расчета вероятности итогового результата обучения абитуриента.

The article describes the construction of a neural network to predict the result of the successful completion of training by a potential university student. When building, training and analyzing the neural network, the Loginom system was used. The results of the USE presented in MS Excel format were used as initial data. The number of examples is 10311 with five parameters. The objective function accepts binary values. An example of calculating the probability of the final result of an applicant's education is presented.

Ключевые слова: нейронная сеть, сетевая модель, нейрон, Loginom, Low-cod

Keywords: neural network, network model, neuron, Loginom, Low-cod

Введение

В настоящее время в процессе электронного учета хозяйственной деятельности предприятий накопилось большое количество данных Big Data, при обработке которых можно извлекать из них нетривиальные ценные знания.

Для поиска новых знаний (Data Mining) о изучаемой системе используются алгоритмы машинного обучения, базирующиеся на математическом анализе и статистике. Вполне понятно, что в этом случае необходимы знания по анализу

данных, теории вероятности, базам данных. Вместе с тем, аналитик должен понимать область применения методов для решения конкретной проблемы и уметь их использовать на практике.

Для решения бизнес-задач, связанных с анализом больших данных (Big Data), существует большое количество различных инструментов. Все они с разной степенью успешности обрабатывают, моделируют и визуализируют данные. Чтобы понять, насколько тот или иной инструмент соответствует бизнес-задачам, необходимо провести сравнение этих методов, а не полагаться на имеющейся информации об продукте.

Для создания приложений по анализу данных и их использованию требуются высококлассные программисты. На рынке наблюдается дефицит этих специалистов. Поэтому естественно сформировалась концепция по созданию, настройке и модификации систем и приложений, которая практически не требует написания программного кода (Low-code).

Low-code платформа Loginom предоставляет аналитику возможность провести анализ больших данных без создания сложного программного кода. Платформа Loginom имеет удобный визуальный конструктор, позволяющий аналитику настроить все процессы анализа больших данных: интеграцию, подготовку данных, моделирование, визуализацию. При этом сокращается время от тестирования модели до создания работающего бизнес-процесса.

Чтобы проиллюстрировать и оценить возможности платформы [1], в настоящей статье представлен процесс создания нейронной сети (НС) с использованием Loginom. В качестве исходных данных для анализа использовалась информация об итоговых результатах обучения студентов.

Экспериментальная часть

При поступлении в ВУЗ абитуриент предоставляет результаты ЕГЭ, которые представляют собой набор оценок: «Русский язык», «Математика», «Физика», «Информатика» и т.д.

После завершения обучения фиксируется результат обучения студента. Так, если студент по каким-либо причинам не закончил обучение с получением диплома в журнале с оценками по единому государственному экзамену (ЕГЭ), с которыми студент был принят в ВУЗ, ставится отметка в виде числа 0. В случае успешного окончания ВУЗа (получение диплома) в этом же журнале ставится число 1. Кроме того, в этом наборе данных отмечены также пол студента и вид учебного заведения, которое закончил студент при поступлении в ВУЗ.

Таким образом в наличии имеется журнал в котором фиксируются данные по студентам. Графу «Диплом» можно представить в виде значений целевой функции, которая принимает значения в зависимости от значения вступительных параметров (оценок), перечисленных выше.

Имея большое количество значений целевой функции, представляется возможным, используя методы машинного обучения, построить нейронную сеть НС [1] используя платформу Loginom.

Для обучения и тестирования нейронной сети использовались данные в виде таблицы MS Excel в формате «xlsx», в которых строки отражают набор значений

признаков для описания отдельного обезличенного студента, а столбцы соответствуют этим признакам (таблица 1). Последний столбец «Диплом» представляет собой целевой признак, который принимает значения 0 или 1. Значение этого признака равного 1 означает, что студент закончил обучение с получением диплома. В противном случае студент закончил образование без получения диплома.

Таблица 1. Исходный набор данных по каждому абитуриенту

Студент	Пол	Обр. учрежд.	Математика	Русский язык	Физика	Диплом
15056	0	1	72	60	58	1
15060	1	3	49	43	42	0
15061	1	2	29	45	37	0
15062	0	3	41	70	30	1
15068	0	3	35	65	32	0

Как видно из Таблицы 1 в НС должно быть 5 элементов на входе. НС в этом случае будет выглядеть следующим образом (Рисунок 1).

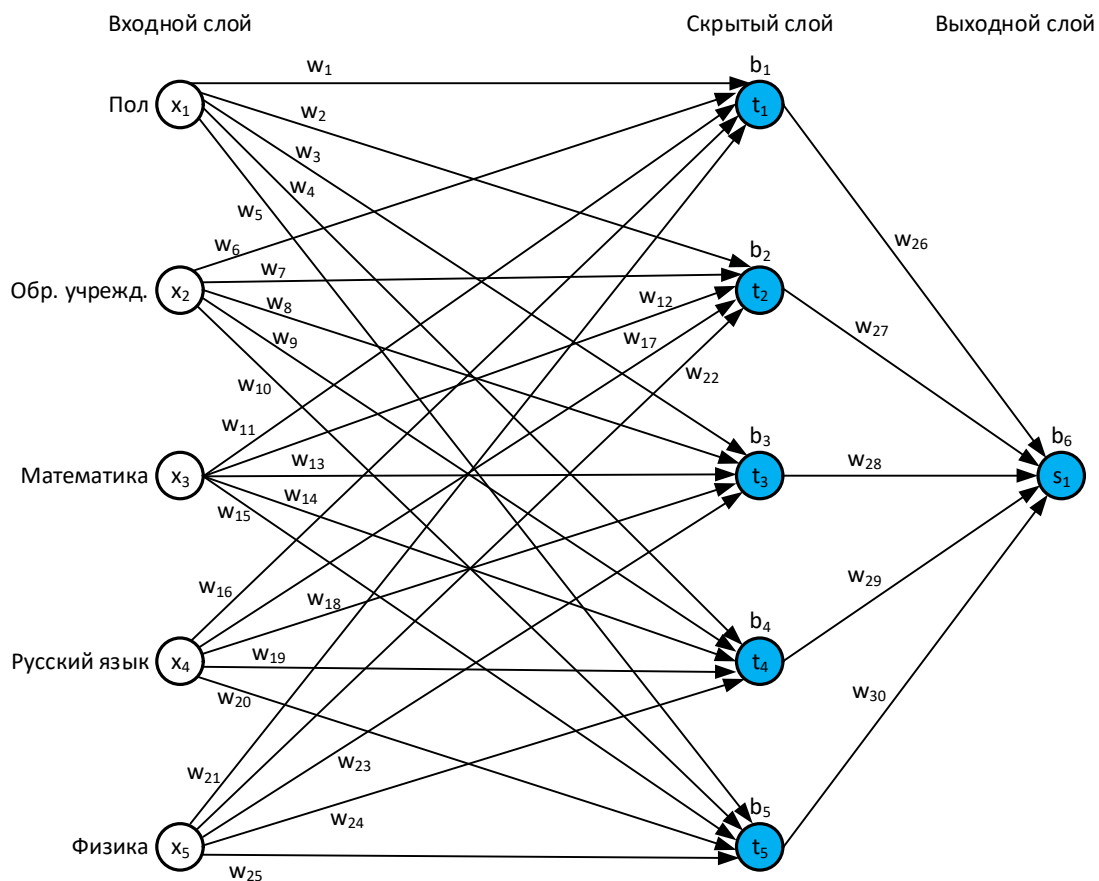


Рисунок 1. Нейронная сеть с пятью входами и пятью нейронами скрытом слое

На Рисунке 1 изображена НС, состоящая из 6 нейронов. Они представлены в виде кругов синего цвета. Внутри кругов символами t_1 , t_2 , t_3 , t_4 , t_5 , s_1 представлены параметры, значения которых равны значениям активационной функции в этих нейронах.

Во входном слое находится информация в виде чисел x_1, x_2, x_3, x_4, x_5 , которые подаются в НС.

Первые пять нейронов находятся в так называемом скрытом слое. В данном случае представлен один скрытый слой. В самом общем случае количество скрытых слоев может быть больше единицы. Их количество определяется по результатам проведенного анализа НС.

Выходной нейрон располагается в выходном слое. На рис.1 изображен один выходной нейрон, в котором содержится итоговая информация, ради которой и строится НС.

Для создания НС можно воспользоваться одним из имеющихся пакетов [2-4]. В настоящее время лидером по использованию для разработки нейросетей является платформа TensorFlow, разработанная компанией Google специально для создания нейросетей самой разной структуры. При этом разработчику требуется знания по программированию и умение использовать библиотеки для анализа работы НС.

Представленную НС можно построить также с использованием платформы Logitom. При этом не требуется написания программного кода. Реализация НС осуществляется путем создания сценария из последовательно соединенных узлов, каждый из которых представляет собой объект с определенной функциональностью и свойством (Рисунок 2).

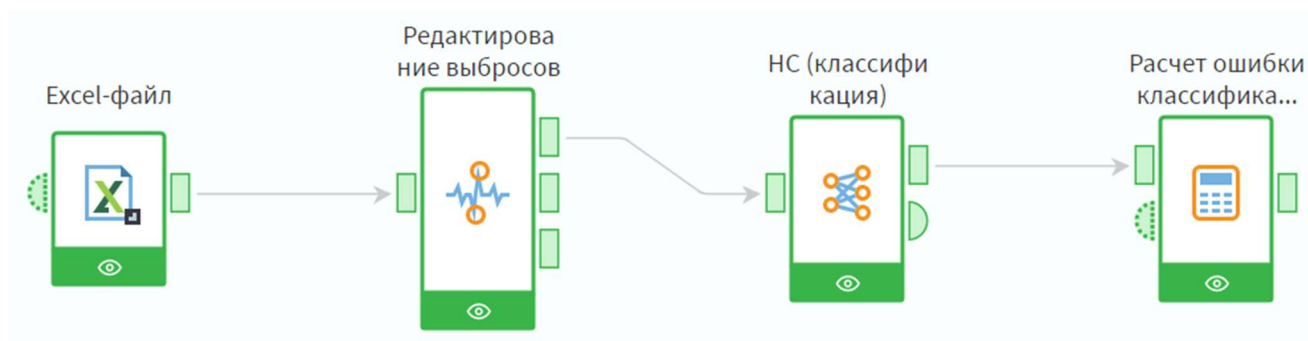


Рисунок 2. Сценарий приложения НС платформе Logitom

Для построения рассматриваемой НС потребуется 4 компонента Logitom. Последовательность шагов для реализации рассматриваемой НС представляется следующим образом:

1. Вызываем предварительно установленную платформу Logitom [1].
2. Создаем пакет со сценарием приложения НС (рис. 2), представляющий собой последовательный набор 4 объектов (узлов).

С помощью первого узла «Excel-файл» осуществляется загрузка данных в формате MS Excel. В рассматриваемом случае количество записей составляет 10311 человек, которые обучались в период 2007 по 2017 годы. В записях представлены значения вступительных оценок ЕГЭ и значения результата обучения в ВУЗе (1 – есть диплом, 0 – нет диплома). Статистика входных данных для построения НС имеет следующие показатели.

Таблица 2. Статистика входных данных

Статистический показатель	Пол	Обр. учрежд.	Математика	Русский язык	Физика	Диплом
минимум	0	1	17	11	25	0
максимум	1	3	100	100	100	1
среднее	1	2	51	59	48	0
отклонение	0,42	0,72	17,00	14,70	12,36	0,50

Гистограмма входных данных по математике, русскому языку, физике имеет вид нормального распределения, что указывает на возможность построить качественную модель для прогнозирования результата обучения студента.

Второй узел «Редактирование выбросов» предназначен для очистки некорректно записанных примеров, содержащих выбросы. В представленном наборе экстремальных выбросов не наблюдается.

Третий узел «НС (классификация)» используется для создания и обучения НС (рис. 1), которая включает в себя 5 элементов на входе, 5 нейронов в скрытом слое и 1 нейрон на выходе.

Для обучения НС принято 90 % примеров, 3000 эпох, k-fold кросс-валидация (начальные веса параметров изменяются 5 раз), порог минимального изменения весов 0,005. После обучения и тестирования НС формируется таблица со значениями параметров, один из которых представляет собой прогноз получения диплома (Диплом-прогноз) в соответствии с оценками ЕГЭ (таблица 3).

Таблица 3. Выходные данные нейронной сети

Пол	Обр. учрежд.	Математика	Русский язык	Физика	Диплом	Дипломпрогноз
0	1	72	60	58	1	1
1	3	49	43	42	0	0
1	2	29	45	37	0	1
0	3	41	70	30	1	0
0	3	35	65	32	0	1
..

Четвертый компонент «Расчет ошибки классификации» позволяет определить ошибку классификации, которая рассчитывается как разность параметров «Диплом-прогноз» и «Диплом». Для выполнения этой операции в системе LogInot имеется встроенный язык, который может быть использован в режиме настройки этого компонента. Выражение для расчета ошибки имеет вид: $\text{abs}(\text{Diplom} - \text{Diplom_predicted})$.

В компоненте «Расчет ошибки классификации» имеется ряд визуализаторов, которые позволяют провести просмотр и анализ данных.

Так с помощью визуализатора «Таблица» можно посмотреть весь выходной набор данных (Таблица 4).

Таблица 4. Выходные данные компонента «Расчет ошибки классификации»

Пол	Обр. учрежд.	Математика	Русский язык	Физика	Диплом	Дипломпрогноз	Ошибка прогноза
0	1	72	60	58	1	1	0
1	3	49	43	42	0	0	0
1	2	29	45	37	0	1	1
0	3	41	70	30	1	0	1
0	3	35	65	32	0	1	1
..

В частности, представляется возможным визуально оценить качество работы НС по значению параметра «Ошибка прогноза», который принимает значение 0 или 1. Значение ошибки 0 соответствует правильному решению НС. В противном случае НС показывает неверный прогноз.

В системе Loginom имеется визуализатор «Куб», с помощью которого можно провести анализ как отдельно выделенных параметров, так и в разрезе их различных сочетаний.

Так, например, выбрав по отдельности параметры «Диплом-прогноз», «Диплом», «Ошибка прогноза» (таблица 5) визуализатор позволяет определить количество фактически выданных и прогнозируемых дипломов об окончании обучения в ВУЗе и количество ошибочных решений НС.

Таблица 5. Статистика отдельных параметров

Параметр	Количество	Общее количество студентов	Среднее значение
Диплом-прогноз	5119	10311	0,50
Диплом	5340	10311	0,52
Ошибка НС	3959	10311	0,38

Представим значения параметров в разрезе по половому признаку (таблица 6).

Таблица 6. Статистика отдельных параметров в разрезе по половому признаку

Параметр	Количество	Общее количество студентов	Среднее значение
женщины			
Диплом-прогноз	1573	2393	0,66
Диплом	1429	2393	0,60
Ошибка НС	846	2393	0,35
мужчины			
Диплом-прогноз	3346	7918	0,45
Диплом	3911	7918	0,49
Ошибка НС	3959	10311	0,34

Представим значения параметров в разрезе по виду образовательного учреждения, которое закончил абитуриент перед зачислением в ВУЗ (таблица 7).

Таблица 7. Статистика параметров в разрезе образовательного учреждения

Параметр	Количество	Общее количество студентов	Среднее значение
образовательное учреждение 0			
Диплом-прогноз	11	504	0,02
Диплом	171	504	0,34
Ошибка НС	174	504	0,35
образовательное учреждение 1			
Диплом-прогноз	387	876	0,44
Диплом	207	876	0,24
Ошибка НС	339	876	0,39
образовательное учреждение 2			
Диплом-прогноз	2424	6487	0,38
Диплом	4847	6487	0,75
Ошибка НС	3770	6487	0,58
образовательное учреждение 3			
Диплом-прогноз	1012	2444	0,38
Диплом	54	2444	0,02
Ошибка НС	1000	2444	0,41

Построим матрицу соответствия правильных и неправильных предсказаний НС с фактическими результатами получения диплома после окончания ВУЗа (таблица 8).

Таблица 8. Матрица соответствия

Диплом	Диплом-прогноз		
	0	1	Итого
0	3102	1869	4971
1	2090	3250	5340
Итого	5192	5119	10311

Согласно матрице соответствия получим следующие данные:

1) общая точность (ассигасу) правильных решений НС составляет

$$(3102+3250)/10311=0,616;$$

2) точность (precision) по результату прогноза 0 составляет:

$$3102/4971=0,624;$$

3) точность (precision) по результату прогноза 1 составляет:

$$3250/5340=0,609;$$

4) полнота (recall) правильных решений по результату прогноза 0 составляет:

$$3102/5192=0,597;$$

5) полнота (recall) правильных решений по результату прогноза 1 составляет:

$$3250/5119=0,635.$$

Таким образом, на основании вышеизложенного можно утверждать, что платформа Loginom позволяет построить НС, не обладая глубокими знаниями в программировании. Для получения результатов ее работы необходимы общие навыки по эксплуатации программного обеспечения и знания по статистическому анализу данных.

Результаты

1. Построена нейронная сеть для прогнозирования успешности окончания университета потенциальны студентом с помощью платформы Loginom.
2. Выполнено обучение нейронной сети. В качестве учителя использованы данные о 10311 студентах.
3. Определено качество нейронной сети и процент предсказаний.

Список использованных источников и литературы

1. Платформа для продвинутого анализа данных. Бесплатно и без ограничения по времени [Электронный ресурс]. – Режим доступа: <https://loginom.ru/blog/about-community-edition> (дата обращения: 21.09.2022).
2. Нейронные сети. Электронный учебник по статистике [Электронный ресурс]. – Режим доступа: <http://statsoft.ru/home/textbook/modules/stneunet.html> (дата обращения: 14.09.2022).
3. Пишем нейросеть на Python с нуля [Электронный ресурс]. – Режим доступа: <https://proglib.io> (дата обращения: 20.09.2022).

List of references

1. Platform for advanced data analysis. Free and without time limit, <https://loginom.ru/blog/about-community-edition>, (date of access: 09/21/2022).
2. Neural networks. Electronic textbook on statistics, <http://statsoft.ru/home/textbook/modules/stneunet.html>, (date of access: 09/14/2022).
3. We write a neural network in Python from scratch, <https://proglib.io> (date of access: 09/20/2022).