

Закирьянова Г.Т., Трегубова А.Х.

ПРОГРАММА LOGINOM: ОБРАБОТКА И АНАЛИЗ МЕДИЦИНСКИХ ДАННЫХ

Башкирский государственный медицинский университет, г. Уфа

Медицинская сфера является одной из областей, где объем данных постоянно растет, и где анализ этих данных играет ключевую роль в улучшении качества здравоохранения. Программа Loginom представляет собой мощный инструмент для анализа медицинских данных, который позволяет эффективно «очищать» и обрабатывать информацию, выявлять закономерности и принимать обоснованные решения на основе фактических данных.

Ключевые слова: аналитические платформы, интеллектуальный анализ данных, оптимизация данных, визуализация данных, структурное проектирование, Loginom.

Loginom представляет собой инструмент для разработки прикладных решений в области анализа данных. Технологии, воплощенные в Loginom, позволяют провести все этапы создания аналитической системы на основе единой архитектуры: от объединения данных до построения моделей и визуализации результатов. Платформа может быть применена в различных областях и способна обрабатывать разнообразные структурированные табличные данные.

Программа Loginom обладает широкими возможностями для работы с медицинскими данными. С ее помощью можно проводить анализ различных типов данных, включая результаты клинических исследований, данные об эпидемиологических показателях, информацию о пациентах и многое другое. Графический интерфейс программы позволяет легко создавать сложные цепочки обработки данных, включая этапы очистки, преобразования и анализа.

Одним из ключевых преимуществ программы Loginom при работе с медицинскими данными является возможность создания моделей прогнозирования и классификации. С ее помощью можно строить прогностические модели для прогнозирования развития заболеваний, определения рисков и принятия предиктивных мер по улучшению здоровья пациентов. Благодаря возможности автоматизации процессов обработки данных пользователи могут быстро и эффективно анализировать большие объемы информации.

Программа Loginom также обладает возможностью интеграции с различными медицинскими информационными системами, что позволяет создавать комплексные решения для анализа и управления медицинскими данными. Это открывает новые возможности для исследований в области медицины, помогает улучшить диагностику и лечение пациентов, а также оптимизировать процессы управления здравоохранением.

Нами для примера работы в Loginom был проведен анализ массива данных по прогнозированию сердечно-сосудистых заболеваний. Сформирован сценарий, выполняющий импорт и экспорт данных, «очистку» и преобразование данных (удаление лишних символов, удаление дубликатов, пустых строк, проверка основного языка описания и т.д.), а также

сортировка данных, которая позволяет быстро отфильтровать необходимую информацию. Выполнено статистическое исследование, проанализирована и составлена прогностическая модель логистической регрессии. Часть действий отображены в окне программы Loginom в сценариях, действия - в узлах (рис.1).

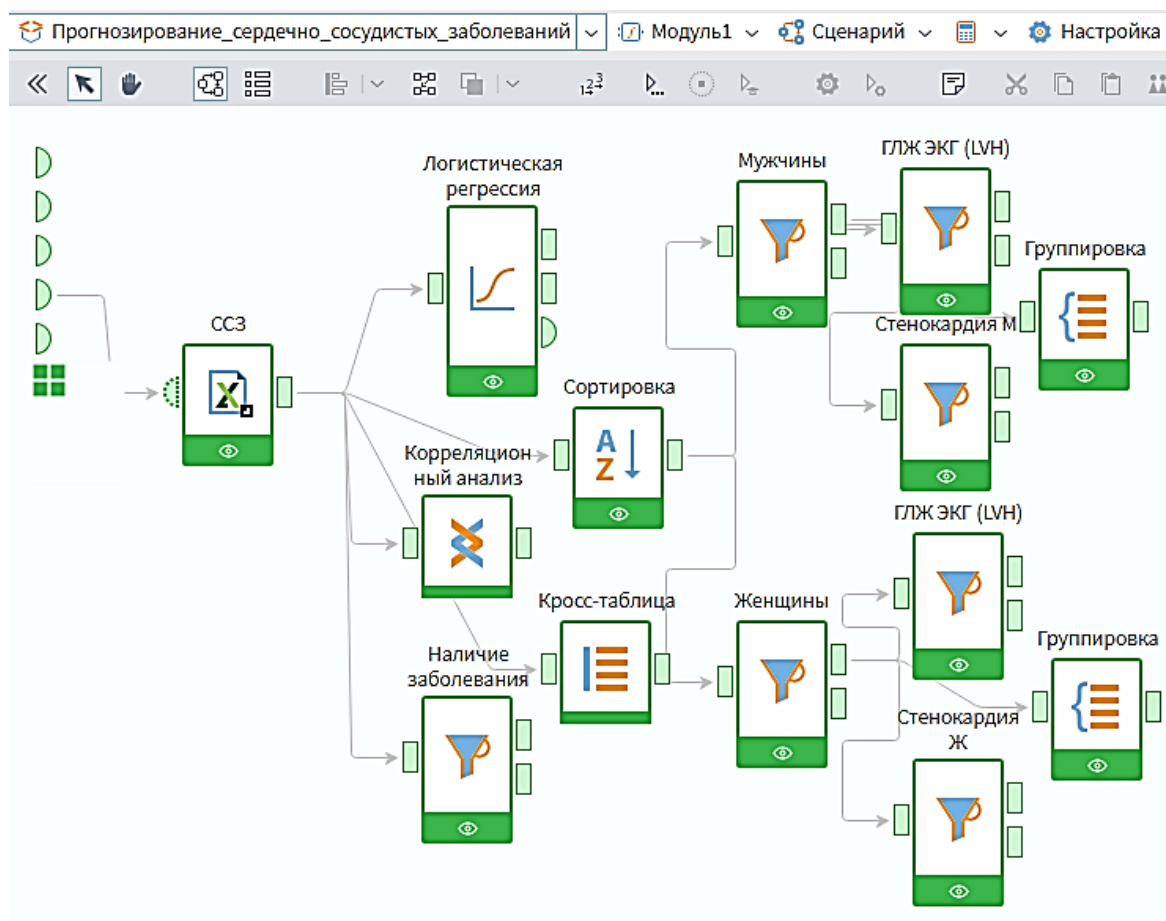


Рис.1. Окно сценария с выполненными узлами в Loginom

Предоставленный объем данных представляет собой сгенерированный набор информации, который был разработан для моделирования сценария прогнозирования сердечно-сосудистых заболеваний. Главная цель этого исследовательского набора данных заключается в создании инструмента, который позволяет исследователям изучать и экспериментировать с различными прогностическими моделями обнаружения сердечно-сосудистых заболеваний.

Набор данных включает в себя информацию о 918 пациентах. Каждый пациент характеризуется 12 различными параметрами: пол, возраст, тип боли в груди, холестерин в сыворотке крови, артериальное давление, уровень сахара в крови натощак, ЭКГ, максимально достигнутая частота сердечных сокращений, депрессия, стенокардия. Все эти параметры играют важную роль в прогнозировании вероятности развития сердечно-сосудистых заболеваний.

Обработанные данные отображаются в табличном виде. Источниками данных являются пациенты, а значения атрибутов – это результаты клинико-диагностических исследований (рис.2).

Ц3 ССЗ - Быстрый просмотр

Набор данных

#	12 Age	ab Sex	ab ChestPainType	12 RestingBP	12 Cholesterol	12 FastingBS	ab RestingECG	12 MaxHR	ab ExerciseAngina	9.0 Oldpeak	ab ST_Slope	12 HeartDisease
1	40	M	ATA	140	289	0	Normal	172	N	0,00	Up	0
2	49	F	NAP	160	180	0	Normal	156	N	1,00	Flat	1
3	37	M	ATA	130	283	0	ST	98	N	0,00	Up	0
4	48	F	ASY	138	214	0	Normal	108	Y	1,50	Flat	1
5	54	M	NAP	150	195	0	Normal	122	N	0,00	Up	0
6	39	M	NAP	120	339	0	Normal	170	N	0,00	Up	0
7	45	F	ATA	130	237	0	Normal	170	N	0,00	Up	0
8	54	M	ATA	110	208	0	Normal	142	N	0,00	Up	0
9	37	M	ASY	140	207	0	Normal	130	Y	1,50	Flat	1
10	48	F	ATA	120	284	0	Normal	120	N	0,00	Up	0
11	37	F	NAP	130	211	0	Normal	142	N	0,00	Up	0
12	58	M	ATA	136	164	0	ST	99	Y	2,00	Flat	1
13	39	M	ATA	120	204	0	Normal	145	N	0,00	Up	0
14	49	M	ASY	140	234	0	Normal	140	Y	1,00	Flat	1
15	42	F	NAP	115	211	0	ST	137	N	0,00	Up	0
16	54	F	ATA	120	273	0	Normal	150	N	1,50	Flat	0
17	38	M	ASY	110	196	0	Normal	166	N	0,00	Flat	1
18	43	F	ATA	120	201	0	Normal	165	N	0,00	Up	0
19	60	M	ASY	100	248	0	Normal	125	N	1,00	Flat	1
20	36	M	ATA	120	267	0	Normal	160	N	3,00	Flat	1
21	43	F	TA	100	223	0	Normal	142	N	0,00	Up	0
22	44	M	ATA	120	184	0	Normal	142	N	1,00	Flat	0
23	40	F	ATA	124	201	0	Normal	164	N	0,00	Up	0

Рис.2. Табличное представление массива данных сердечно-сосудистых заболеваний

В программе Logiном проведен статистический анализ массива данных, наглядно отображена описательная статистика всех переменных

№	Метка	Вид	Гистограмма	Диаграмма размаха	Минимум	Максимум	Среднее	Стандарт...	Несмеще...	Медиана	Верхний ...	Пропуски	Ун
1	12 Age				28	77	54	9,4326165...	88,974254...	54	60	0	
2	ab Sex			Недоступно	1	1	1	0	0			0	2
3	ab ChestPainType			Недоступно	2	3	2,9498910...	0,2182886...	0,0476499...			0	4
4	12 RestingBP				0	200	132	18,514154...	342,77390...	130	140	0	
5	12 Cholesterol				0	603	199	109,38414...	11964,891...	223	267	0	
6	12 FastingBS				0	1	0	0,4230456...	0,1789676...	0	0	0	

Рис.3. Описательная статистика массива данных сердечно-сосудистых заболеваний

Возможно визуализировать исходные данные графически, например (рис.4).:

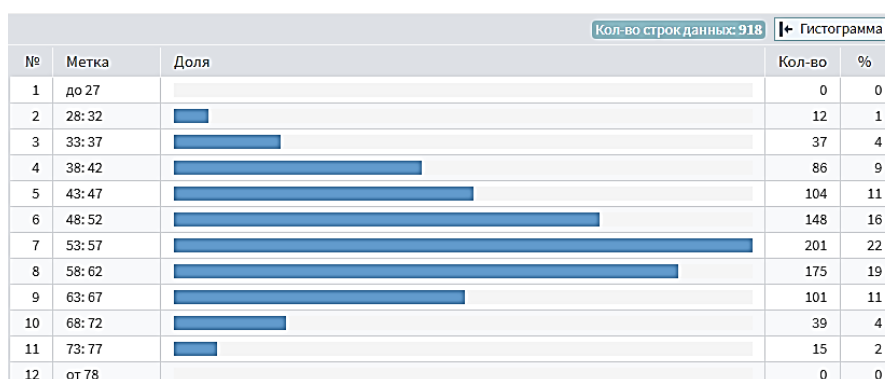


Рис.4. Гистограмма частот переменной Возраст.

По рассматриваемому массиву данных провели сортировку, группировку и фильтрацию данных, среди 918 пациентов: мужчин – 725, женщин – 193. Средний возраст мужчин – 55 лет, женщин – 53 года. Больных сердечно-сосудистого заболевания, вызванным критерием ГЛЖ ЭКГ: женщин – 47, мужчин – 141. Гипертрофия левого желудочка – это выявляемое при инструментальном исследовании утолщение стенок левого желудочка в совокупности или без увеличения его полости. Существует большое количество заболеваний, способных приводить к такому отклонению от нормы. Данное патологическое состояние опасно тем, что при отсутствии необходимого медицинского контроля оно рано или поздно приведет к развитию хронической сердечной недостаточности.

Анализируя данные, получим, что у мужчин по сравнению с женщинами риск заболеваний сердца выше. Это связано с половым различием. В организме женщины детородного возраста количество женского гормона – эстрогена – в разы выше, чем у мужчин.

Синтетический характер этого набора данных позволяет проводить контролируемые эксперименты и служит ценным ресурсом для предварительных исследований в области прогнозирования сердечно-сосудистых заболеваний.

В программе Logitom рассчитаны коэффициенты логистической регрессии и построены графики визуализации результатов, (рис.5)

Атрибут	Коэффициент	Стандартная ошиб...	Статистика В...	P-значение	Отношение ша...	Нижняя граница...	Верхняя границ...
12 Age	-0,023059	0,011640	3,924282	0,047593	0,977205	0,955163	0,999755
ab ChestPainType							
ab ATA	1,857270	0,322208	33,225925	8,204931e-9	6,406223	3,406732	12,046645
ab NAP	1,718681	0,261938	43,051877	5,330747e-11	5,577169	3,337731	9,319148
ab TA	1,491458	0,427361	12,179596	0,000483	4,443568	1,922923	10,268377
12 Cholesterol	0,003981	0,001023	15,149347	0,000099	1,003989	1,001979	1,006004

Рис 5. Результаты расчетов программы Logitom логистической регрессии

Была разработана прогностическая модель для определения вероятности наличия сердечных заболеваний.

$$P = 1 / (1 + e^{-z}) \times 100\%,$$

$$z = 1,466X_M + 1,86X_{ATA} + 1,72X_{NAP} + 1,49X_{TA} + 0,004X_{Холестерин} + 1,193X_{Наличие} + 0,991X_Y + 0,410X_{Пиковый\ низкий\ ST} - 1,060X_{Up} + 1,444X_{Flat}$$

Полученная регрессионная модель является статистически значимой ($p < 0,001$)

Визуализация результатов представлена графиками, где площадь под ROC-кривой составила 0,9. Пороговое значение логистической функции P в точке cut-off, составило 0,5.

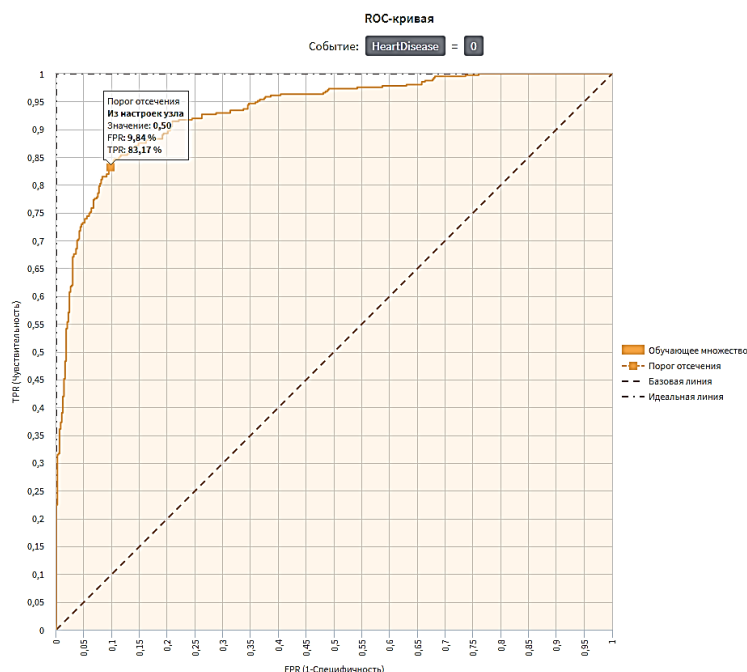


Рис 6. ROC-кривая, характеризующая зависимость вероятности наличия сердечной заболеваний от значения логистической функции P

Выполнив тестирование алгоритмов и оценку прогноза модели, можно заключить, что у логистической регрессии показатели коэффициента точности алгоритма составили 84% при прогнозировании отсутствия сердечной недостаточности и 89% при диагностировании сердечной недостаточности. Прогнозирование сердечной недостаточности на примере тестового набора данных по средствам данным алгоритмов, показали высокую погрешность прогнозов.

Прогностические модели позволяют увидеть незначительные закономерности при диагностике заболеваний, что позволяет снизить затраты на дополнительную диагностику и увеличить эффективность работы медицинского персонала.

Инструменты Logipom имеют большой потенциал для исследований и проверки прогностических моделей. Исследователи могут использовать его для проведения различных экспериментов и оценки эффективности различных подходов прогнозирования. Особенно ценно его использование для разработки новых методов, что является одной из основных задач в медицинской науке.

Заключение и выводы

Программа Loginom представляет собой ценный инструмент для анализа медицинских данных, который помогает улучшить качество здравоохранения, оптимизировать процессы управления и принимать обоснованные решения на основе данных. Благодаря программе Loginom, исследователи и врачи могут эффективно анализировать медицинские данные, выявлять закономерности и принимать обоснованные решения для улучшения здоровья пациентов.

ЛИТЕРАТУРА

1. Loginom Official Website. URL: <https://www.loginom.com/>
2. Kaggle: Your Home for Data Science. <https://www.kaggle.com/>