

Методы интеллектуального анализа данных в виртуальном практикуме для целей цифровизации образования

Е. В. Тимошенко, кандидат физ.-мат. наук, доцент, профессор кафедры программного обеспечения информационных технологий

E-mail: : timoshchenko@msu.by

ORCID ID: 0000-0003-1373-5113

УО «Могилевский государственный университет имени А.А.Кулешова», ул. Космонавтов, д.1, 212022, г. Могилёв, Республика Беларусь

А. Ф. Разжков, магистр педагогических наук, аспирант

E-mail: razhkov.a@mail.ru

ORCID ID: 0000-0001-5200-4167

ГНУ «Объединенный институт проблем информатики Национальной академии наук Беларуси», ул. Сурганова, д.6, 220012, г. Минск, Республика Беларусь

Аннотация. Разработан программный модуль, лежащий в основе виртуального лабораторного практикума и состоящий из двух компьютерных приложений, использующих методы интеллектуального анализа биомедицинских данных в качестве примера возможности прогнозирования заболевания по предложенным симптомам пациента и определения вероятности наличия сердечно-сосудистого заболевания. Разработанный виртуальный лабораторный практикум может быть использован при преподавании блока медико-биологических дисциплин и включает в себя перечень лабораторных работ, которые помогают приобрести практические навыки биоаналитической работы и навыки программирования на Python.

Ключевые слова: цифровизация образования, практико-ориентированное обучение, виртуальный практикум, интеллектуальный анализ данных, прогнозирование заболеваний, ИТ в образовании, язык программирования Python.

Для цитирования: Тимошенко, Е. В. Методы интеллектуального анализа данных в виртуальном практикуме для целей цифровизации образования / Е. В. Тимошенко, А. Ф. Разжков//Цифровая трансформация. – 2021. – № 4 (17). – С. 52–62.



© Цифровая трансформация, 2021

Data Mining Techniques in a Virtual Workshop for the Digitalization of Education

E. V. Timoschenko, PhD in Physico-mathematical sciences, Associate Professor, Professor of the Department of Information Technology Software

E-mail: : timoshchenko@msu.by

ORCID ID: 0000-0003-1373-5113

"Mogilev State A. Kuleshov University", 1, Kosmonavtov str., 212022, Mogilev, The Republic of Belarus

A.F. Razhkov, Master of Pedagogic sciences, postgraduate student

E-mail: razhkov.a@mail.ru

ORCID ID: 0000-0001-5200-4167

The State Scientific Institution «The United Institute of Informatics Problems of the National Academy of Sciences of Belarus», 6, Surganova str., 220012, Minsk, The Republic of Belarus

Abstract. A software module has been developed that forms the basis of a virtual laboratory workshop and consists of two computer applications that use the methods of intelligent analysis of biomedical data for an example of the possibility

of predicting a disease based on the proposed symptoms of a patient and determining the likelihood of cardiovascular disease. The developed virtual laboratory practice can be used in teaching the block of biomedical disciplines and includes a list of laboratory works that help to acquire practical skills in bioanalytical work and programming skills in Python.

Key words: digitalization of education, practice-oriented learning, virtual workshop, data mining, disease prediction, IT in education, programming language Python.

For citation: Timoschenko E. V., Razhkov A. F. Data Mining Techniques in a Virtual Workshop for the digitalization of education. *Cifrovaja transformacija* [Digital transformation], 2021, 4 (17), pp. 52–62(in Russian).

© Digital Transformation, 2021

Введение. В настоящее время в Республике Беларусь в рамках реализации концепции модели «Университет 3.0» большое внимание уделяется развитию системы практико-ориентированного обучения в целях формирования личностных и профессиональных компетенций выпускников, как будущих специалистов. При этом в учебно-методических планах учреждений высшего образования значительное место уделяется внедрению в учебный процесс инновационных обучающих технологий, а также учебных дисциплин, связанных с изучением современных компьютерных технологий. Поэтому, использование методов, которые ориентированы на формирование у обучающихся профессиональных компетенций, востребованных сегодня работодателями и максимально приближенных к потребностям реального сектора экономики, становится все более актуальным.

Практико-ориентированный подход, в большей степени, направлен на приближение обучающей деятельности учреждения образования к потребностям реальной жизни. Такой подход позволяет создать условия для целенаправленной подготовки выпускников ВУЗов, как будущих молодых специалистов к конкурентной борьбе на рынке труда. Основной задачей преподавателя при этом является не только передача определенного объема знаний и навыков в рамках учебных программ преподаваемых дисциплин, но и подготовка студента таким образом, чтобы он мог применить полученные теоретические знания на практике и быть готовым к определенным видам профессиональной деятельности. Этому, несомненно, благоприятствует внедрение информационных технологий в образовательный процесс, создавая предпосылки для кардинального обновления, как содержательно-целевых, так и технологических сторон обучения. Это проявляется в существенном обогащении системы дидактических средств и формировании нетрадиционных информационно-коммуникационных технологий

(ИКТ) и методических приемов обучения.

Практико-ориентированное обучение предполагает реализацию следующих основных концептуальных положений:

- активные формы приобретения и усвоения знаний;
- гарантия исследовательской свободы;
- самоанализ собственного опыта;
- расширение возможностей социализации обучения.

Для реализации этих принципов может применяться такая форма учебных занятий, как виртуальный лабораторный практикум, который является одним из инновационных методов обучения, предполагающих использование ИКТ. Он позволяет перевести обучение в целом из реального пространства в виртуальное, что избавляет от необходимости в учебных помещениях, физическом присутствии обучающихся и преподавателей, а также позволяет выполнять задания удаленно. Возможность удаленной работы позволяет, кроме того, расширить географическую составляющую обучающихся, что может положительным образом сказаться на престиже ВУЗа, его известности в информационном пространстве и привлекательности для иностранных абитуриентов.

Виртуальный лабораторный практикум представляет собой программный модуль, состоящий из одного или нескольких программных средств, позволяющих моделировать или правдоподобно имитировать процессы, явления и закономерности реальной окружающей среды. Виртуальная лабораторная работа при этом иногда способна полностью заменить реальный объект исследования, что позволяет гарантированно получить результаты, основанные на данных статистического анализа определенных закономерностей, сфокусировать внимание на ключевых сторонах исследуемого явления, а также, в некоторых случаях, сократить время проведения экс-

перимента.

Однако при проведении таких лабораторных работ необходимо помнить, что виртуальная модель отображает реальные процессы и явления в усредненном, в целом – упрощённом виде, содержащем лишь существенные свойства моделируемого объекта или процесса. Поэтому вопрос о том, какие ключевые моменты на самом деле подчёркивает модель, а какие можно в данных условиях считать несущественными, может быть сформулирован в качестве одной из форм задания. Такие виртуальные лабораторные работы можно выполнять целиком, используя предложенное программное обеспечение, а можно расширить область исследования и предложить задачи, где использование программного продукта будет являться одним из этапов моделирования более широкой работы, которая предполагает также, например, сравнительный анализ с реальными данными и лабораторным оборудованием.

Согласно образовательному стандарту высшего образования «Высшее образование. Первая ступень. Специальность 1-80 02 01 Медико-биологическое дело. Квалификация «Биолог-аналитик. Преподаватель биологии» [1], специалисты данного профиля должны быть подготовлены к решению ряда профессиональных задач. Из них можно выделить следующие:

- сбор и подготовка научных материалов, квалифицированная постановка экспериментов, исследование молекулярно-биологических и молекулярно-генетических характеристик клеток в норме и их изменения после неблагоприятных экологических воздействий или других патологий; обработка результатов экспериментальных исследований;

- морфологические, физиолого-биохимические и иммунологические исследования на клеточном, тканевом и организменном уровне с целью оценки степени и уровня их патологических изменений;

- подготовка результатов лабораторных исследований для медико-диагностических целей;

- получение материалов для лабораторных анализов, квалифицированное проведение экспериментов, заключение по результатам экспериментов и анализов.

Учебный план, разработанный для специальности 1-80 02 01 Медико-биологическое дело в рамках учебной дисциплины «Информацион-

ные технологии», предполагает необходимость приобретения студентами навыков решения профессиональных задач с использованием информационно-технических средств [2]. В частности, студенты должны иметь представление об алгоритмах машинного обучения искусственного интеллекта и использовать в своей работе методы интеллектуального анализа данных.

Осознав актуальность и необходимость внедрения новых прогрессивных методов для осуществления процессов цифровизации образования, коллектив авторов приступил к реализации идеи создания виртуального лабораторного практикума, который можно будет использовать также в целях повышения практико-ориентированности обучения при преподавании дисциплин, предполагающих анализ большого объема биохимических и медико-биологических данных, для студентов специальностей медико-биологического профиля.

В основу виртуального лабораторного практикума положен программный модуль, включающий в себя разработанное специально для этих целей программное обеспечение, которое использует алгоритмы и методы интеллектуального анализа данных [3], так как использование других методов, по объективным причинам весьма затруднительно, в силу огромного количества обрабатываемой информации.

Методы и технологии разработки ПО. Замечено, что из-за интенсивного развития искусственного интеллекта и растущего количества баз данных различного назначения, почти каждый человек в быту и в профессиональной деятельности постоянно производит большое количество данных. В связи с неуклонно растущим объемом данных традиционным технологиям анализа данных становится все труднее эффективно работать с информацией – принимать её, обрабатывать, анализировать и изменять. Для решения этой проблемы постепенно развивается и совершенствуется технология интеллектуального анализа данных таким образом, чтобы полностью соответствовать требованиям различных сфер деятельности человека. Целью технологии интеллектуального анализа данных является обнаружение неявных закономерностей в объемах данных. Иными словами, интеллектуальный анализ данных – это процесс извлечения определенных знаний из совокупности данных [4].

Если говорить о медико-биологической об-

ласти исследований, то в качестве иллюстрации обоснованности использования интеллектуального анализа можно привести следующий пример: имеется информация о пациенте, которая содержит возраст, пол, артериальное давление, данные биохимического анализа крови. Требуется обнаружить закономерность между этими данными, наличием патологии или обнаружением её вероятности. В этом случае, технология интеллектуального анализа данных может помочь обнаружить закономерности и спрогнозировать различные сценарии – в данном случае наличие болезни у данного пациента.

После анализа существующих технологий для разработки обучающего программного обеспечения, которое ляжет в основу виртуального лабораторного практикума, было решено использовать следующие методы машинного обучения искусственного интеллекта.

1. *Logistic Regression (Логистическая регрессия)*.

Логистическая регрессия [5] используется в виде модели бинарной логистической регрессии, которая используется в том случае, когда зависимая переменная имеет два значения, например, 0 и 1, используется для моделирования вероятности существования определенного класса или события, позволяет оценивать апостериорные вероятности принадлежности объектов классам.

В модели логистической регрессии применяется сигмовидная функция, приведенная ниже.

Эта функция успешно отображает любое

$$\sigma(z) = \frac{1}{1 + e^{-z}} \quad (1)$$

число в значение от 0 до 1, которое рассматривается как вероятность предсказания классов. Например, имеется два класса: наличие и отсутствие сердечно-сосудистого заболевания. Если установлен порог 0.5 и значение сигмовидной функции равно 0.7, это означает, что у пациента высокая вероятность наличия сердечно-сосудистого заболевания – 70%.

2. *Naive Bayes Classifier Algorithm (Наивный байесовский классификационный алгоритм)*.

Алгоритм основан на теореме Байеса [6] и является одним из простых методов классификации объектов [7]. В его основе лежит предположение о статистической независимости призна-

ков, что ограничивает его применимость. Тем не менее, он иногда используется на практике и в некоторых задачах показывает приемлемые результаты.

В задаче обучения нейронной сети можно применить теорему Байеса следующим образом:

Здесь $P(y)$ – вероятность события A, опреде-

$$P(y|X) = \frac{P(X|y)P(y)}{P(X)} \quad (2)$$

ляемого переменной класса y , $P(X)$ – вероятность события B, определяемого вектором зависимых объектов $X=(x_1, x_2, \dots, x_n)$, $P(X|y)$ – вероятность события A при наступлении события B, где:

Например, $X = (\text{Пол}=1, \text{Возраст}=63, \text{Артериальное давление}=145, \dots)$ и $y = (\text{Наличие болезни}=1)$. Таким образом, $P(X|y)$ здесь означает высокую вероятность наличия болезни, учитывая биомедицинские данные пациента.

3. *k-Nearest Neighbor (Метод ближайших соседей)*

Принцип метода ближайших соседей [8] состоит в том, чтобы оценить относится ли классифицируемый объект к тому классу, к которому принадлежат ближайшие к нему объекты обучающей выборки. В области прогнозирования сердечно-сосудистых заболеваний объект является вектором в n-мерном пространстве, каждое измерение в котором представляет собой описание одного из признаков объекта (атрибуты пациента, изложенные выше).

4. *Decision Tree (Дерево принятия решений)*

Дерево принятия решений – это алгоритм обучения с учителем, применимый к задачам классификации, в которых в основном используются категориальные переменные [9]. Дерево состоит из узлов-листьев, которые являются классами, и узлов принятия решений, представляющих правила, которые делят заданное множество объектов на подмножества. Например, правило, которое касается максимальной достигнутой частоты сердечных сокращений: если у пациента она больше 120 ударов в минуту, то вероятность наличия сердечно-сосудистого заболевания значительно высокая.

5. *Random Forest (Метод случайного леса)*

Данный алгоритм [10] представляет собой ансамбль многочисленных деревьев решений. На

основе каждого из случайно выбранных подмножеств обучающего набора строится свое дерево принятия решений. Отличием метода, случайного леса построения деревьев решений от классического метода является использование только фиксированного числа случайно отбираемых признаков обучающего набора и построение полного дерева.

6. Support Vector Machine (Метод опорных векторов)

Метод опорных векторов [11] является одним из популярных алгоритмов машинного обучения с учителем, применяемых для задач классификации. Принцип алгоритма состоит в представлении атрибутов объекта в виде векторов, где значение атрибута – координаты вектора. Объект, имеющий n атрибутов, представляет собой вектор n -мерного пространства. Пространством, разделяющим точки из разных классов наилучшим образом, будет являться гиперплоскость.

7. LightGBM (Light Gradient Boosted Machine)

LightGBM – это алгоритм машинного обучения, разработанный командой Google на основе градиентного бустинга [12]. Принцип алгоритма заключается в вертикальном продолжении дерева, в отличие от других алгоритмов на основе бустинга деревьев – горизонтальном «выращивании» деревьев.

8. XGBoost (eXtreme Gradient Boosting)

XGBoost – алгоритм машинного обучения, принцип которого состоит в построении комбинации простых алгоритмов путем изменения веса входных данных [13]. XGBoost представляет собой идеальное сочетание программных и аппаратных возможностей, разработанных для повышения точности существующих методов повышения мощности в кратчайшие сроки.

Разработанное программное обеспечение предназначено для работы в операционных системах Windows, Linux и других, поддерживающих работу Python – высокоуровневого языка программирования с динамической типизацией данных и поддерживающий объектно-ориентированный, функциональный и императивный стили программирования. В качестве среды разработки языка Python было решено использовать PyCharm [14], представляющую собой кроссплатформенную интегрированную среду разработки, которая предоставляет средства для анализа кода, графический отладчик, а также большое

количество плагинов, облегчающих разработку программ.

Программный модуль виртуального лабораторного практикума. Специалисты медико-биологического профиля, как будущие биологи-аналитики, должны уметь обрабатывать большие объемы биохимических и медико-биологических данных, а также владеть навыками применения технологий машинного обучения искусственного интеллекта на практике. Поэтому было решено разработать программное обеспечение в виде компьютерных приложений с простым и удобным интерфейсом, которые позволяют студентам не только закрепить навыки аналитической работы для прогнозирования заболевания по предложенным симптомам пациента и определения степени его вероятности, но и приобрести знания возможностей машинного обучения нейронных сетей при решении ряда профессиональных задач с использованием информационно-технических средств.

Используя методы интеллектуального анализа данных, разработано программное обеспечение, включающее в себя два программных средства, одно из которых позволяет прогнозировать наличие заболеваний сердечно-сосудистой системы по перечню биомедицинских данных пациента, а второе – определять наличие характерного заболевания по существующим у пациента симптомам. Разработанные компьютерные приложения легли в основу программного модуля виртуального лабораторного практикума [15], апробированного и внедренного в учебный процесс кафедрой естествознания Могилевского государственного университета имени А.А. Кулешова.

Программное обеспечение прогнозирования заболеваний сердечно-сосудистой системы.

При разработке программного обеспечения для определения вероятности возникновения сердечных заболеваний у пациента по его биомедицинским данным, использовалась информация, включающая в себя в общей сложности 303 наблюдения, в том числе, 13 описательных признаков и 1 целевой. Информация о признаках представлена в таблице 1.

В процессе разработки, алгоритмы интеллектуального анализа данных были реализованы с помощью библиотеки Scikit-Learn.

Интерфейс разработанного приложения имеет вид классического диалогового окна (ри-

Таблица 1. Информация о пациентах
Table 1. Patient information


Атрибут	Описание и тип данных
Возраст	Год рождения. Числовой.
Пол	1 = мужчина, 0 = женщина.
Боль в груди	Тип боли в груди. 0: бессимптомная, 1: атипичная стенокардия, 2: неангинальная, 3: типичная стенокардия
Артериальное давление в состоянии покоя	Давление в мм рт.ст. Числовой.
Холестерин сыворотки	В мг/дл. Числовой.
Уровень сахара в крови натощак	0: менее 120 мг / дл, 1: более 120 мг / дл
Результаты электрокардиографии в покое	0: норма, 1: наличие аномалии ST-T, 2: наличие вероятной или определенной гипертрофии левого желудочка по критериям Эстеса
Максимальная достигнутая частота сердечных сокращений	Удар/минута. Числовой.
Стенокардия, вызванная физической нагрузкой	0: нет, 1: да
Депрессия ST, вызванная упражнениями по сравнению с отдыхом	Числовой.
Наклон пикового сегмента ST при физической нагрузке	0: нисходящий, 1: плоский, 2: восходящий
Количество крупных сосудов (0–3), окрашенных флуороскопией.	0: «0», 1: «1», 2: «2», 3: «3», 4: «4»
Талассемия	0: отсутствует, 1: исправленный дефект, 2: нормальный, 3: обратимый дефект
Целевой признак	0 – отсутствие заболевания, 1 – наличие заболевания.

сунк 1). На главной форме приложения расположено 13 текстовых полей для ввода соответствующих атрибутов (биомедицинских данных) пациента.

При нажатии на кнопку «Выполнение» происходит реализация программного кода, в завершение которого выводится подробная информация о вероятности наличия заболевания пациента (рисунок 2). Здесь можно увидеть не только сте-

пень вероятности диагностирования заболевания – высокая, средняя, низкая – но и количественное отображение этой вероятности в процентном соотношении. Результат представлен достаточно подробно – вычислена вероятность наличия заболевания всеми используемыми здесь аналитическими методами, что даёт преподавателю вариативность в подборе заданий для студентов.

Разработанное программное обеспечение

 Prediction Heart Disease

Удалить модели

Возраст пациента в годах	41
Пол (1 = мужчина; 0 = женщина)	0
Тип боли в груди (0, 1, 2, 3, 4)	1
Артериальное давление в состоянии покоя (в мм рт. ст. при поступлении в стационар)	130
Сывороточный холестерин в мг/дл	204
(Уровень сахара в крови натощак > 120 мг/дл) (1 = да; 0 = нет)	0
Результаты электрокардиографии в покое (0, 1, 2)	0
Максимальная достигнутая частота сердечных сокращений	172
Вызванная физическими упражнениями стенокардия (1 = да; 0 = нет)	0
Депрессия сегмента ST, вызванная физическими упражнениями относительно отдыха	1.4
Наклон пика ST сегмента	2
Количество крупных сосудов (0-3), окрашенных по цвету	0
Заблевание крови, называемое талассемией 3 = нормальный; 6 = исправленный дефект; 7 = обратимый дефект	3

Выполнение

Рис. 1. Интерфейс программного обеспечения прогнозирования заболеваний сердечно-сосудистой системы
Fig. 1. Cardiovascular disease prediction software interface

```

Результат NB: Вероятность наличия заболевания высока
(Точность = 82.18279569892474% )

Результат RF: Вероятность наличия заболевания высока
(Точность = 82.48387096774194% )

Результат SVM: Вероятность наличия заболевания высока
(Точность = 83.18279569892472% )

Результат LR: Вероятность наличия заболевания высока
(Точность = 84.13978494623656% )

Результат DT: Вероятность наличия заболевания высока
(Точность = 76.9247311827957% )

Результат LGBM: Вероятность наличия заболевания высока
(Точность = 81.49462365591398% )

Результат XGB: Вероятность наличия заболевания высока
(Точность = 80.82795698924731% )

Результат KNN: Вероятность наличия заболевания высока
(Точность = 65.31182795698925% )

```

Рис. 2. Результат работы приложения (вычисление вероятности наличия сердечно-сосудистого заболевания)
NB – наивный байесовский алгоритм, RF - метод случайного леса, SVM – Метод опорных векторов, LR – Логистическая регрессия, DT – Дерево принятия решений, LGBM – Light Gradient Boosted Machine, XGB – eXtreme Gradient Boosting, KNN – Метод ближайших соседей.

Fig. 2. The result of the application (calculating the probability of having a cardiovascular disease)
NB - Naive Bayes Classifier Algorithm, RF - Random Forest, SVM - Support Vector Machine, LR - Logistic Regression, DT - Decision Tree, LGBM - Light Gradient Boosted Machine, XGB - eXtreme Gradient Boosting, KNN – k-Nearest Neighbor.

позволяет не только осуществить прогноз заболеваний сердечно-сосудистой системы по введенной информации о пациенте, но и предоставляет возможность демонстрации применения разных методов интеллектуального анализа к одному и тому же набору исходных данных.

Программное обеспечение прогнозирования диагноза по существующим симптомам у пациента.

При разработке программного обеспечения для определения наличия определенного заболевания у пациента по имеющимся у него симптомам использовался набор данных, включающий в себя в общей сложности более 5000 наблюдений, в том числе, 132 описательных признака (симптомы) и 1 целевой – одно из 41 имеющегося заболевания в базе данных.

Интерфейс программы состоит из диалогового окна классического вида, представленного на рисунке 3.

На главной форме приложения (рисунок 4) предлагается из списка выбрать симптомы пациента. После чего требуется нажатие на кнопку «Выполнить» и в текстовом поле справа появится результат прогнозирования, определенный разными используемыми методами (рисунок 5).

Программное обеспечение позволяет не только прогнозировать диагноз по существующим симптомам у пациента, но и сравнивать эффективность алгоритмов машинного обучения при постановке диагноза и степень полноты пе-

речня симптомов для достоверной диагностики.

Содержание виртуального лабораторного практикума. Освоение содержания представленного виртуального практикума позволяет студентам приобрести навыки биоаналитической работы, а также элементарные навыки программирования на Python с целью возможности самостоятельной модификации программного обеспечения под любой набор анализируемых данных и любое направление исследований. Например, студенты медико-биологического профиля, а в частности, студенты специальности 1-80 02 01 «Медико-биологическое дело» в ходе обучения приобретают навыки биоаналитической работы для проведения биохимических, иммунологических и других анализов и исследований биологических проб в медицинских целях.

Содержательная часть виртуального практикума включает следующие лабораторные работы.

Лабораторная работа 1. Введение в язык программирования Python. Целочисленная арифметика.

Цель работы: приобрести начальные навыки работы с языком программирования Python, на примерах программного кода освоить технологию разработки простейших программ на Python.

Лабораторная работа 2. Искусственный интеллект для целей прогнозирования сердечно-сосудистых заболеваний.

Цель: ознакомиться с работой искусствен-

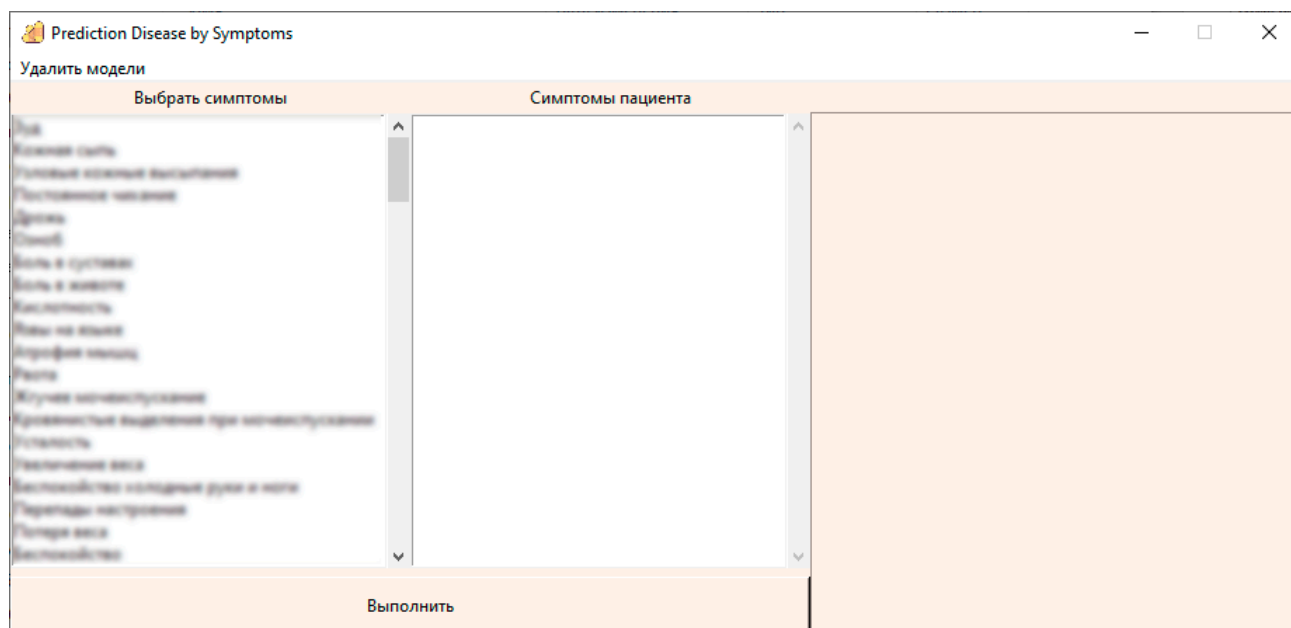


Рис. 3. Интерфейс программного обеспечения прогнозирования диагноза по существующим симптомам у пациента
Fig. 3. Diagnosis prediction software interface based on existing patient symptoms

Prediction Disease by Symptoms

Удалить модели

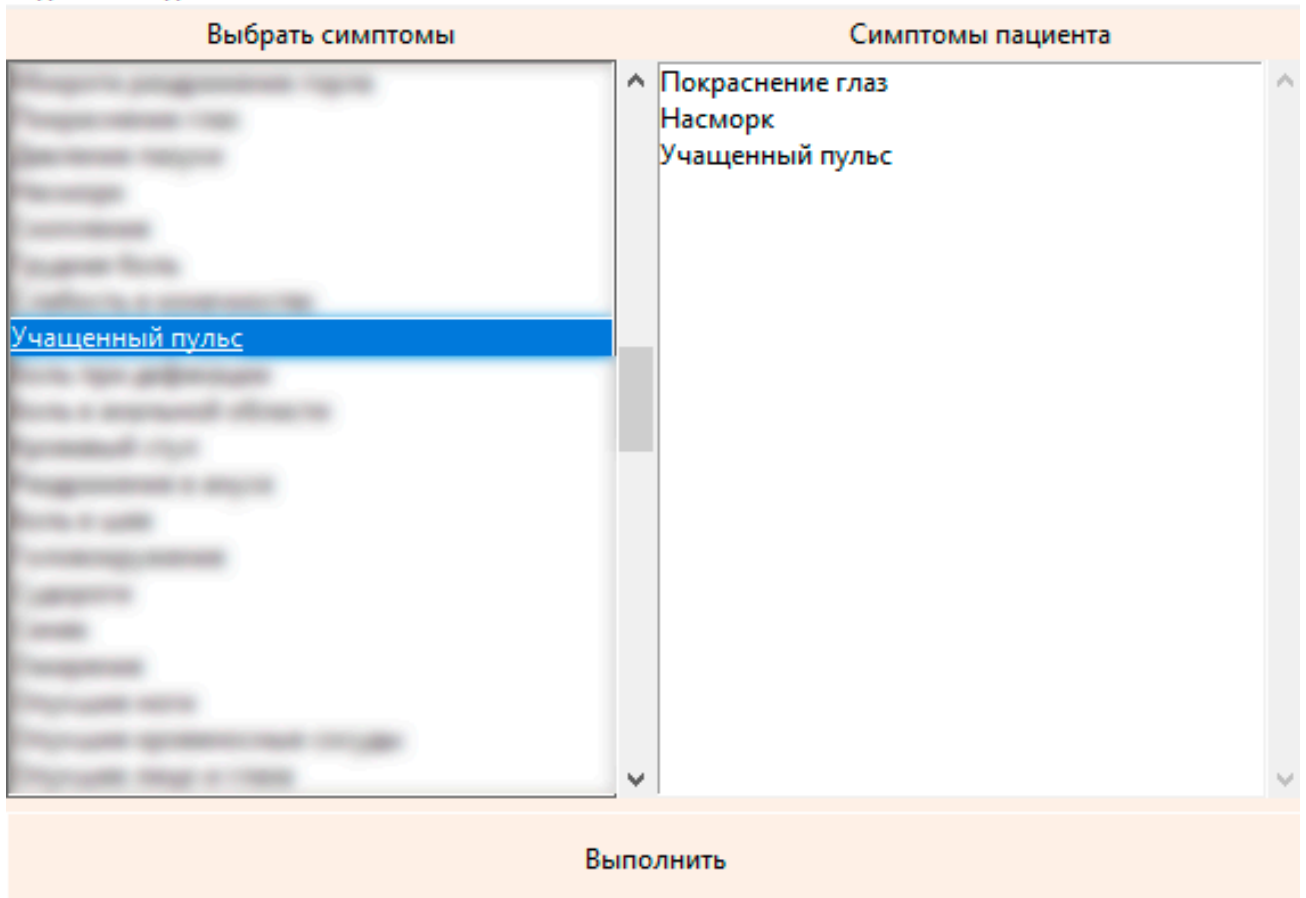


Рис. 4. Выбор симптомов пациента из предлагаемого списка
Fig. 4. Selecting the patient's symptoms from the suggested list

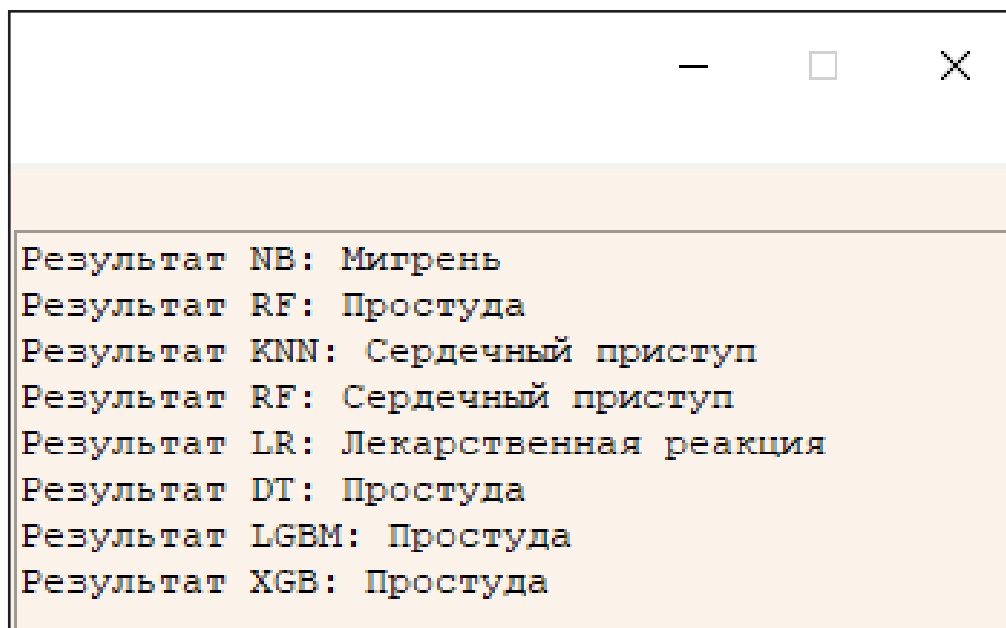


Рис. 5. Результат работы приложения (диагноз)
Fig. 5. The result of the software (diagnosis)

ного интеллекта и алгоритмами машинного обучения, закрепить знание основ проектирования программного обеспечения с учетом специфики диагностики сердечно-сосудистых заболеваний.

Лабораторная работа 3. Программное обеспечение для определения вероятности сердечно-сосудистых заболеваний у пациента.

Цель: ознакомиться с предложенным программным обеспечением для прогнозирования сердечно-сосудистых заболеваний по набору биохимических, медицинских и антропологических данных пациента, проанализировать работу приложения.

Лабораторная работа 4. Программное обеспечение для прогнозирования диагноза по имеющимся симптомам пациента.

Цель: ознакомиться с предложенным программным обеспечением для прогнозирования диагноза по имеющимся у пациента симптомам, проанализировать работу приложения.

Заключение. Представленный виртуальный лабораторный практикум помогает в решении задачи цифровизации образования и повышения практико-ориентированности процесса обучения при преподавании учебных дисциплин у студентов специальностей медико-биологического профиля. Так, студенты смогут приобрести не только навыки аналитической работы, связанной с их профессиональной деятельностью, но и знание алгоритмов машинного обучения искусственного интеллекта для решения профессиональных задач, основанных на необходимости

анализа большого объема исходных биохимических и медико-биологических данных с использованием информационно-технических средств. Используя разработанное программное обеспечение студенты смогут также сосредоточиться на приобретении практических навыков программирования на Python для возможности модификации приложений, беря за основу предложенный открытый программный код и имея при этом минимальные начальные теоретические знания, необходимые для поддержки этих навыков.

Разработанный программный модуль может быть использован при преподавании дисциплин медико-биологического профиля не только в учреждениях образования, выпускающих специалистов, не осуществляющих врачебную деятельность, таких как биолог-аналитик, но и при преподавании блока медико-биологических дисциплин в профильных учреждениях, осуществляющих подготовку медицинских кадров для системы здравоохранения.

Благодаря программному модулю, допускающему возможность модификации программного обеспечения под определенную специфику и исследовательские цели, практикум универсален и может быть дополнен лабораторными работами и дополнительными заданиями в зависимости от целей использования и преподаваемых учебных дисциплин.

Список литературы

1. Образовательный стандарт высшего образования «Высшее образование. Первая ступень. Специальность 1-80 02 01 Медико-биологическое дело. Квалификация «Биолог-аналитик. Преподаватель биологии»: постановление Министерства Респ. Беларусь, 14 мая 2019 г., № 51 // Нац. правовой Интернет-портал Респ. Беларусь. – 31.07.2019, 8/34277.
2. Борковский, Н. Б. Информационные технологии. Учебная программа учреждения высшего образования по учебной дисциплине для специальности: 1- 80 02 01 Медико-биологическое дело № УД-379-14/р. / Н. Б. Борковский. – Минск: Международный государственный экологический институт имени А. Д. Сахарова Белорусского государственного университета, 2014.
3. Тимошенко, Е. В. Методы интеллектуального анализа биомедицинских данных / Тимошенко Е. В., Ражков А. Ф. // Итоги научных исследований ученых МГУ имени А. А. Кулешова 2019 г.: материалы научно-методической конференции, 29 января – 10 февраля. 2020 г. / под ред. Н. В. Маковской, Е. К. Сычовой. – Могилев: МГУ имени А. А. Кулешова, 2020. – С. 106-107.
4. Han, J. Data Mining: Concepts and Techniques / J. Han, M. Kamber, J. Pei. – UK: Morgan Kaufmann: Oxford, 2012. – pp. 12-18.
5. Schwender, H. Logic regression and its extensions / H. Schwender, I. Ruczinski.: Advanced in Genetics, 2010. – pp. 25-45.
6. Bayes, T. An essay towards solving a problem in the doctrine of chances / T. Bayes.: Resonance, 2003. – pp. 80-88.
7. Karthika, S. A Naïve Bayesian Classifier. For Educational Qualification / S. Karthika, N. Sairam.: Indian Journal of Science and Technology, Vol. 8, 2015. – pp. 1-5.
8. Imandoust, S.B. Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical

- Background / S.B. Imandoust.: Int. Journal of Engineering Research and Applications Vol. 3, Issue 5, Sep-Oct 2013. – pp. 605- 610.
9. Brijain, R Patel A Survey on Decision Tree Algorithm for Classification / R Patel Brijain, K Rana Kushik.: IJEDR, Volume 2, Issue 1, 2014.
10. Biau, G. Analysis of a Random Forests Model / G. Biau.: Journal of Machine Learning Research, 13, 2012. – pp. 1063-1095.
11. Osuna, E. Training support vector machines: an application to face detection / E. Osuna, R. Freund, F. Girosi.: IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997. – pp. 130-136.
12. Guolin Ke. LightGBM: A Highly Efficient Gradient Boosting Decision Tree./ Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu. – Advances in Neural Information Processing Systems 30 (NIPS 2017), pp. 3149-3157.
13. Natekin, A. Gradient boosting machines, a tutorial / A. Natekin, A. Knoll.: Frontiers in Neuroinformatics, Vol. 7, Art. 21., 2013.
14. Get started [Электронный ресурс]. – Режим доступа: <https://www.jetbrains.com/help/pycharm/quick-start-guide.html>. – Дата доступа: 01.05.2021.
15. Разжков, А.Ф. Виртуальный лабораторный практикум «Интеллектуальный анализ данных для прогнозирования заболеваний» / Разжков А.Ф., Тимошенко Е.В. // V международная научно-методическая конференция «Качество подготовки специалистов в техническом университете: проблемы, перспективы, инновационные подходы», 19-20 ноября 2020. – Могилев: МГУП, 2020. – С. 172-175.

References

1. Educational standard of higher education “Higher education. First stage. Specialty 1-80 02 01 Biomedical business. Qualification “Analyst Biologist. Biology teacher ”: Resolution of the Ministry of the Republic. Belarus, May 14, 2019, No. 51 // Nat. legal Internet portal Resp. Belarus. - 07/31/2019, 8/34277.
2. Borkovsky, NB Information technologies. The curriculum of the institution of higher education in the discipline for the specialty: 1- 80 02 01 Biomedical business № UD-379-14 / r. / N.B. Borkovsky. - Minsk: International State Ecological Institute named after A. D. Sakharov of the Belarusian State University, 2014.
3. Timoshchenko, E.V. Methods of intellectual analysis of biomedical data / Timoshchenko E.V., Razhkov A.F. // Results of scientific research of scientists of the Mogilev State University named after A.A. Kuleshov 2019: materials of the scientific and methodological conference, January 29 - February 10. 2020 / ed. N. V. Makovskaya, E. K. Sychova. - Mogilev: Mogilev State University named after A.A.Kuleshov, 2020. -- P. 106-107.
4. Han, J. Data Mining: Concepts and Techniques / J. Han, M. Kamber, J. Pei. - UK: Morgan Kaufmann: Oxford, 2012. - pp. 12-18.
5. Schwender, H. Logic regression and its extensions / H. Schwender, I. Ruczinski. : Advances in Genetics, 2010. - pp. 25-45.
6. Bayes, T. An essay towards solving a problem in the doctrine of chances / T. Bayes.: Resonance, 2003. – pp. 80-88.
7. Karthika, S. A Naïve Bayesian Classifier. For Educational Qualification / S. Karthika, N. Sairam.: Indian Journal of Science and Technology, Vol. 8, 2015. – pp. 1-5.
8. Imandoust, S.B. Application of K-Nearest Neighbor (KNN) Approach for Predicting Economic Events: Theoretical Background / S.B. Imandoust.: Int. Journal of Engineering Research and Applications Vol. 3, Issue 5, Sep-Oct 2013. - pp. 605-610.
9. Brijain, R Patel A Survey on Decision Tree Algorithm for Classification / R Patel Brijain, K Rana Kushik. : IJEDR, Volume 2, Issue 1, 2014.
10. Biau, G. Analysis of a Random Forests Model / G. Biau.: Journal of Machine Learning Research, 13, 2012. - pp. 1063-1095.
11. Osuna, E. Training support vector machines: an application to face detection / E. Osuna, R. Freund, F. Girosi. : IEEE Computer Society Conference on Computer Vision and Pattern Recognition, 1997. - pp. 130-136.
12. Guolin Ke. LightGBM: A Highly Efficient Gradient Boosting Decision Tree ./ Guolin Ke, Qi Meng, Thomas Finley, Taifeng Wang, Wei Chen, Weidong Ma, Qiwei Ye, Tie-Yan Liu. - Advances in Neural Information Processing Systems 30 (NIPS 2017), pp. 3149-3157.
13. Natekin, A. Gradient boosting machines, a tutorial / A. Natekin, A. Knoll. : Frontiers in Neuroinformatics, Vol. 7, Art. 21., 2013.
14. Get started [Electronic resource]. - Access mode: <https://www.jetbrains.com/help/pycharm/quick-start-guide.html>. - Date of access: 01/05/2021.
15. Razhkov, A.F. Virtual laboratory workshop "Data mining for disease prediction" / Razhkov A.F., Timoshchenko E.V. // V international scientific and methodological conference "The quality of training of specialists at a technical university: problems, prospects, innovative approaches", November 19-20, 2020. - Mogilev: MGUP, 2020. - pp. 172-175.

Received: 17.05.2021

Поступила: 17.05.2021