УДК 00.004.89
ГРНТИ 50.53.17

Кубегенова А.Д., магистр технических наук, старший преподаватель, https://orcid.org/0000-0003-0156-7757
НАО «Западно-Казахстанский аграрно-технический университет имени Жангир хана», г. Уральск, ул. Жангир хана 51, 090009, Казахстан, aigul-03@mail.ru
Зайцева Е., доктор PhD, координатор проекта ACeS RU ERASMUS+, Depth. факультет информатики, Университет Жилина, 01026, Словакия, г.Жилина Universitna 8215/1,

Kubegenova A.D., Master of Technical Sciences, Senior Lecturerhttps://orcid.org/0000-0003-0156-7757
NJSC «West Kazakhstan Agrarian and Technical University named after Zhangir khan», Uralsk, st. Zhangir khan 51, 090009, Kazakhstan, aigul-03@mail.ru
Zaitseva E., PhD, Coordinator of the ACeS RU ERASMUS+ project, Depth. Faculty of Computer Science, University of Zilina, 01026, Slovakia, Zilina Universitna 8215/1,

ПРИМЕНЕНИЕ ИНТЕЛЛЕКТУАЛЬНОГО АНАЛИЗА ДАННЫХ В МЕДИЦИНЕ
APPLICATION OF DATA MINING IN MEDICINE

АННОТАЦИЯ

В статье рассматриваются аспекты исследования данных, глубокого анализа данных, получения знаний, способов обработки данных в базе знаний, методов интеллектуального анализа и применения технологии Data Mining в области медицины. Выявлена группа ВИЧ-инфицированных больных, проведен анализ с историей заболеваний, разработаны модели и разработан алгоритм действий (входные данные), проведен анализ и эксперименты с методами поиска данных. Все болезни были представлены в виде набора числовых векторов и были сгруппированы в кластеры, согласно описанной методологии, с помощью этого распределения было рассчитано значение статистики Хопкинса. Сама кластеризация осуществлялась с использованием обычных инструментов библиотеки sklearn. Предложены различные методы представления многомерных данных в двумерной плоскости метод основных компонентов, линия Кохонена и др.

Были рассмотрены два разных способа кластеризации: метод k-Medium (с использованием функции Kmeans из библиотеки sklearn языка Python), методы кластеризации на основе плотности с автоконфигурацией (из функции HDBSCAN из библиотеки Hdbscan языка Python). В случае сравнения оценивают структуру кластеров путем изменения различных параметров одного алгоритма (например, количество групп k); на полученных и подготовленных объектах строится модель (или несколько) и корректируются ее параметры. Затем было проведено тестирование и анализ результатов.

ANNOTATION

The article discusses aspects of data research, in-depth data analysis, knowledge acquisition, methods of data processing in the knowledge base, methods of intellectual analysis and application of Data Mining technology in the field of medicine. A group of HIV-infected patients was identified, an analysis with a history of diseases was carried out, models were developed and an algorithm of actions (input data) was developed, analysis and experiments with data search methods were carried out. All diseases were presented as a set of numerical vectors and were grouped into clusters, according to the described methodology, with the help of this distribution, the value of Hopkins statistics was calculated. Clustering itself was carried out using the usual tools of the sklearn library. Various

methods of representation of multidimensional data in a two-dimensional plane are proposed, the method of basic components, the Kohonen line, etc.

Two different clustering methods were considered: the k-Medium method (using the Kmeans function from the Python sklearn library), density-based clustering methods with autoconfiguration (from the HDBSCAN function from the Python Hdbscan library). In the case of comparison, the cluster structure is evaluated by changing various parameters of one algorithm (for example, the number of groups k); a model (or several) is built on the received and prepared objects and its parameters are adjusted. Then testing and analysis of the results were carried out.

Ключевые слова: *кластеризация, векторизация, корреляция, sklearn, манипуляция.*
Key words: *clustering, vectorization, correlation, sklearn, manipulation.*

Введение. Развитие современных методов хранения и обработки данных приводит к быстрому росту накопленной, требующей анализа информации. Такое большое количество накопленных данных не позволяет обрабатывать их силами человека, и очевидно, что среди этих необработанных данных есть информация, необходимая для принятия важных решений.

Поэтому для автоматического анализа данных необходимо будет использовать технологию Data Mining.

Интеллектуального анализа данных Data Mining в переводе с английского означает «добыча данных», «раскопка данных» является обнаружение неявных закономерностей в наборах данных, а у нас принято переводить как «интеллектуальный анализ данных».

Как научное направление он стал активно развиваться в 90-х годах XX века, что было вызвано широким распространением технологий автоматизированной обработки информации и накоплением в компьютерных системах больших объемов данных [1,2].

Мы знаем, что Data Mining является многопрофильной областью, которая возникла и развивается на основе таких наук, как прикладная статистика, распознавание знаний, искусственный интеллект, теория баз данных и т. д.

DataMining-это процесс поддержки принятия решений, основанный на поиске скрытых шаблонов (шаблонов информации) из данных.

Data Mining можно охарактеризовать как технологию, предназначенную для поиска большого объема нечетких, объективных и полезных на практике закономерностей:

- нечетко, потому что обнаруженные закономерности не могут быть определены стандартными методами обработки информации или экспертным путем;
- объективные, так как выявленные закономерности полностью соответствуют действительности, в отличие от экспертного знания, которое всегда субъективно;
- на самом деле полезно, потому что выводы имеют реальный смысл, который можно применить на практике.

Также очень успешно анализ данных применяется в медицине. Примерами этого являются анализ результатов обследования, диагностика, сравнение эффективности методов лечения и лекарств, анализ заболеваний и их распространенность, выявление побочных эффектов. Такие технологии Data Mining, как ассоциативные правила и цепные шаблоны, успешно используются для определения связи между приемом лекарств и их побочными эффектами.

При визуализации и идентификации, скрытых сложных взаимосвязей между диагностическими особенностями разных групп пациентов применяют различные виды алгоритмов связанные с интелектуальным анализом данных. [3]

Технология Data Mining позваляют обнаруживать в медицинских данных шаблоны, состовляющие основу указанных правил, исходя из этого, можно узнать диагноз пациента и метод его лечения при описании, сочетании различных симптомов заболевании.

В основе Data Mining лежат методы классификации и кластеризации, моделирования и прогнозирования, генетические и эволюционные алгоритмы, методы «мягких вычислений».

Также можно отметить начинающего в Data Mining методы прикладной статистики, анализы регрессионный и корреляционный, дискриминантный и факторный анализ [4]. Для обработки медицинских задач используют следующие алгоритмы алгоритм C4.5; метод к-средних; алгоритм Apriori; алгоритм PageRank; алгоритм Ada Boost; алгоритм k-ближайших

соседей (kNN); наивный байесовский классификатор и др. Эти алгоритмы входят в качестве инструментарии Data Mining и используются для обработки больших данных.

В сатьях С. Кабанихина, О.Криворотько был произведен мониторинг и анализ прогнозирование распространения эпидемии в регионе и построен математическая модель, учитывая численность населения и визуализация, обработки больших данных. Анализировали численный метод решения обратной задачи эпидемиологии, основанный на генетическом алгоритме и традиционных идеях оптимизации.

После того, как все эти факторы были учтены, произведены в модели и сделали прогноз относительно числа людей, которые, как ожидается, будут инфицированы во время эпидемии и продолжительности эпидемии и максимального уровня заболеваемости. [5]

В исследовании А. Бушница, Г. Бочарова сформулировано многомасштабная модель острой ВИЧ-инфекции, которая объединяет процессы распространения инфекции и иммунные реакции в лимфатических узлах и связывает с динамикой ВИЧ, наблюдаемой в крови.

Многомасштабная модель ВИЧ-инфекции, с формулировалась в исследовании, основанной в ряде упрощающих предположений, в которых можно выделить следующие:

1. пространственная динамика клеток и цитокинов в ЛУ рассматривается в двумерной регулярной области;
2. модель ограничена первичной острой ВИЧ-инфекцией и сопутствующими реакциями цитотоксических Т-клеток;
3. внутриклеточная регуляция клеточной судьбы с помощью множественной передачи сигналов цитокинов описывается через иерархию порогов активации:

Элементарные механистическиеие модули высокого разрешения и их интеграция в разработанной многомасштабнойой модельной основы позволил изучить с помощью анализа чувствительности эффективностии мультимодальных подходов к лечению ВИЧ-инфекции, сочетающих ВРТ, антифиброзные и иммунологические методы лечения, модулирующие препараты. Это исследование помагло клиницистам в продвижении к амбициозной цели идеального долгосрочного контроля для лечения инфекции с минимальными побочными эффектами.[6]

В статье Численный алгоритм построения индивидуальной математической модели динамики ВИЧ на клеточном уровне, исследуется проблема определения параметров ВИЧ-инфекции и иммунного ответа с использованием дополнительных измерений концентраций Т-лимфоцитов, свободного вируса и иммунных эффекторов в фиксированные моменты времени для математической модели динамики ВИЧ. Задача задания параметров математической модели (обратная задача) сводится к задаче минимизации целевой функции, описывающей отклонение результатов моделирования от экспериментальных данных. Реализовано и исследовано генетический алгоритм решения задачи, минимизации функции наименьших квадратов. Анализировалось результаты численного решения обратной задачи. [7]

Э.О.Омонди в своей статье анализировав, формулировал математические компартментные модели передачи ВИЧ внутри и между двумя возрастными группами в Кении. С помощью метода Монте-Карло подогнал модель к данным и вывел параметры, оценив цифры размножения в приделах передачи в возрастных группах и между возрастными группами передачи основных чисел размножения. В анализе данных получили результат, что существует значительная разница в среднем числе новых случаев ВИЧ-инфекции между мужчинами и женщинами в двух возрастных группах.

В большей степени результат показывал, что большинство случаев инфицированы ВИЧ женщины, и уровень передачи ВИЧ на душу населения был самым высоким, так как существует взаимодействие между молодыми людьми и взрослыми. Анализ чувствительности показывал, что показатели размножения зависят главным образом от вероятности инфекции.[8]

Значимость моего исследования является выявление групп пациентов ВИЧ инфицированных с помощью специализированного программного обеспечения для анализа данных. Важность исследования этой проблемы заключается в своевременной постановке

правильного диагноза и проведении необходимого лечения, соответствующего одной из их клинических форм. Отсутствие лечения приводит к ухудшению здоровья, и могут возникнуть осложнения и рост заболевания. Для этого нужно изучить методы обработки медицинских данных, создать взаимосвязь и закономерность между симптомами, оценить и построить различные прогностические модели. Полученные результаты могут быть использованы специалистами для принятия решений при постановке диагноза.

Материалы и методы исследования. Архив медицинских данных содержит много информации о различных случаях конкретных заболеваний, методах их диагностики. Поиск образцов является одной из задач многих медицинских исследований. Для решения таких задач часто используются методы автоматического анализа данных.

Концепция Data Mining используется для обозначения совокупности методов определения практически полезных знаний в данных, необходимых для принятия решений в различных областях деятельности, включая медицину. Поиск данных-это обширная область, которая возникла и развивается в поколении таких наук, как статистика, машинное обучение, искусственный интеллект.

Рассмотрим краткое описание. Статистика-это наука, которая собирает информацию и тщательно изучает объекты для их дальнейшего анализа и обработки. Машинное обучение можно охарактеризовать как процесс, изучающий методы построения алгоритмов, способных к обучению. Искусственный интеллект-это научная область, которая разрабатывает методы, позволяет решать интеллектуальные проблемы на электронном компьютере, если эти проблемы решаются людьми.

Data Mining объединяет методы и алгоритмы, такие как искусственные нейронные сети, деревья решений, корреляция, кластерный анализ, линейная регрессия, байесовские сети и многое другое. Решаются такие задачи, как классификация, кластеризация, прогнозирование.[9]

Статистические методы часто сводятся к решению уравнений линейной регрессии. Однако при таком подходе не всегда удается найти контакт. В таких случаях применяются методы машинного обучения. В медицине существует множество экспертных систем для диагностики, основанных на закономерностях и правилах, описывающих сочетание симптомов различных заболеваний.

Эти правила помогут определить, как болел больной, какое лечение назначать, предсказать результат назначенного лечения, изучить причины различных патологий. Технологии поиска данных позволяют находить медицинские данные, такие как правила и схемы. Разработка методов диагностики является актуальной задачей медицины, которая, в свою очередь, относится к задачам классификации.

Важность этого анализа заключается в своевременной постановке правильного диагноза и проведении необходимого лечения, соответствующего одной из их клинических форм.Несвоевременное лечение приводит к ухудшению здоровья и может стать осложнением заболевания.

Результаты исследования. Объект анализа-сбор данных больных с различными клиническими формами ВИЧ-инфекции, построение прогнозной модели и проведение экспериментов с использованием методов поиска данных.

Целью анализа является определение группы пациентов с ВИЧ-инфекцией с помощью специализированного программного обеспечения для анализа данных.

Полученные результаты могут быть использованы специалистами для принятия решения, при постановке диагноза.В процессе разработки моделей выстраивается алгоритм действий и вводятся входные данные.В первую очередь в качестве исходных данных был проанализирован анамнез данных больных ВИЧ-инфекцией, получающих лечение.Обязательный анализ для каждого пациента на рисунке #1 представлена выписка из одного из документов

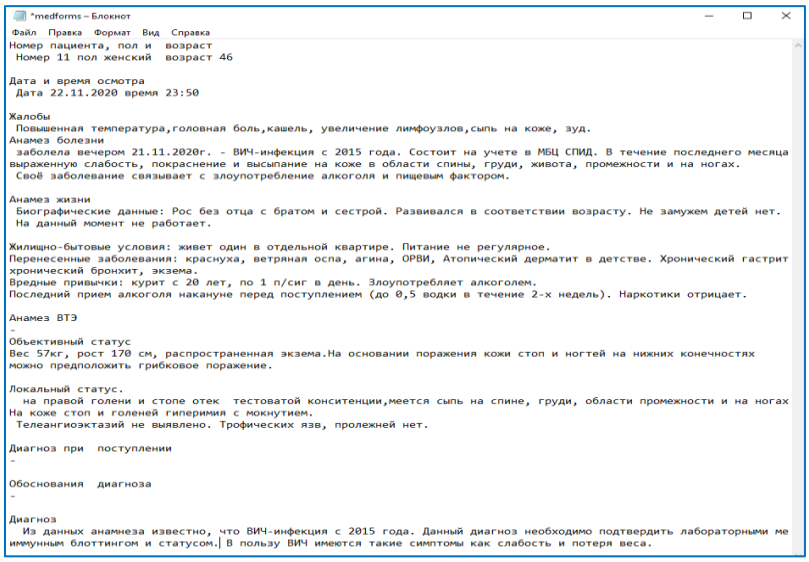

Рисунок 1 – Входные данные

Как мы видим, в документе есть неструктурированные записи (за исключением явных признаков, их можно просто разбить на шаблоны и проанализировать).Данные добровольной передачи больного заносятся в базу данных.Кроме того, некоторые разделы могут отсутствовать в базе данных пациента. Например, не во всех документах указаны подробные сведения о диагнозе и жалобах пациента. Если блоки результатов теста представлены в виде нескольких различных шаблонов, из которых можно получить параметры, то такие разделы, как жалобы и история болезни, заполняются в совершенно произвольной форме. Кроме того, прежде чем приступить к построению кластерной модели, необходимо изучить вероятностное решение этой задачи.

Векторизация

Представление текстов в формате, которое можно было бы выполнить с помощью таких манипуляций, осуществлялось с помощью функции TfidfVectorizer из библиотеки sklearn языка Python.Статистический критерий TF-IDF, лежащий в основе этого метода, используется для оценки значимости слова. Все болезни представлены в виде набора числовых векторов, с которыми можно проводить дальнейшие манипуляции. [10]

Перед началом кластеризации необходимо определить предрасположенность данных, сгруппированных в первые кластеры. Для этого выбирается статистика Хопкинса.На самом деле он основан на нулевой гипотезе о том, что данные не склонны к группировке. Для расчета его значения на основе распределения с тем же стандартным отклонением, что и исходный набор данных, создается несколько случайно генерируемых ложных наборов данных.

Для каждого i наблюдения рассчитывается среднее расстояние от n до k: ω_i между конкретными объектами q_i между искусственными объектами и их ближайшими фактическими соседями. (1)

$$H_{ind} = \frac{\Sigma_n \omega_i}{\Sigma_n q_i + \Sigma_n \omega_i} \qquad (1)$$

Затем статистика Хопкинса показывает, что q больше 0,5 аналогично, а сгруппированные объекты делятся на случайные и однородные и соответствуют нулевой гипотезе.

Значение H .25 ind < 0 отражает тенденцию к группировке данных на уровне надежности 90%. Если эта статистика показывает, что нулевая гипотеза неверна, и наши доходы имеют тенденцию к кластеризации, тогда мы переходим к кластеризации.[11]

Алгоритмы кластеризации

Были рассмотрены два разных способа кластеризации: метод k-Medium (с использованием функции Kmeans из библиотеки sklearn языка Python), методы кластеризации на основе плотности с Авто-конфигурацией (с использованием функции HDBSCAN из

библиотеки Hdbscan языка Python).На полученных и подготовленны векторизация текстовых данныхх объектах необходимо построить модель (или несколько) и отрегулировать ее параметры. Затем проводится тестирование и анализ результатов. На рисунке #2 показана последовательность работ по анализу.

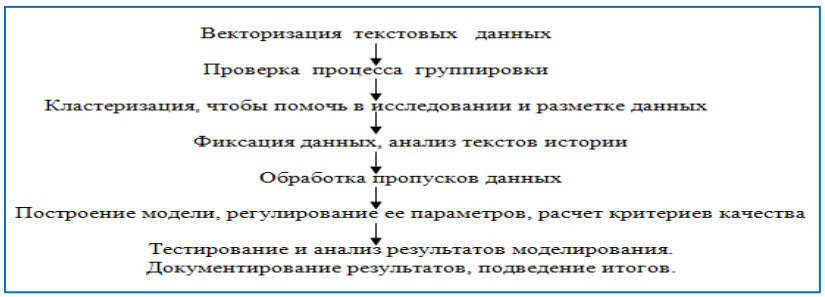

Рисунок 2 – Порядок проведения анализа

На разных этапах работы над задачей требовалась визуализация, использовались различные функции на основе модуля pyplot в библиотеке Python matplotlib.[12]

В частности:

1. NumPy-Фундаментальная библиотека, необходимая для проведения научных вычислений в Python.
2. Matplotlib-библиотека для работы с двумерными графиками
3. Pandas-инструмент анализа структурных данных и временных рядов.
4. Scikit-learn-интегратор классических алгоритмов машинного обучения.
5. SciPy-библиотека, используемая в области математики, науки и инженерии
6. Jupyter-интерактивная вычислительная среда.

Прежде всего, мы должны кластеризовать данные для их детализации. Чтобы данные были простыми, все ненужные знаки препинания были удалены, и библиотека регулярных выражений Python используется для выполнения этой задачи. Разделение очищенного документа на пустые символы и отдельные слова-метод разделения документов на лексемы.

Например:

```
def correct_known_words(story):
                    dict_ = {'сопут диагноз': 'сопутствующий диагноз',
                    'табл ':'таблеток ',
                     ' ст. вторичных': ' стадии вторичных',
                     'голов. боль': 'головная боль'
```

После манипуляций, проведенных на предыдущем шаге, достаточно загрузить полнотекстовый модифицированный и очищенный фрейм данных и обработать все события, как и значения ячеек фрейма данных, индексы являются именами файлов истории. Функция NearestNeighbors из библиотеки Sklearn, как правило, используется для создания неконтролируемых и управляемых моделей, что позволяет нам создавать псевдообъект, аналогичный нашему векторизированному тексту.[13]

С помощью этого распределения по описанной выше методологии мы можем вычислить значение статистики Хопкинса. Сама кластеризация осуществляется с помощью обычных инструментов библиотеки sklearn. Существуют различные методы представления многомерных данных на двумерной плоскости: метод основных компонентов, линия Кохонена и др.

PCA tSNE поскольку эти методы основаны на различных принципах, если один не показывает визуальных различий между кластерами, другой может работать лучше. Оба метода используются в библиотеке sklearn. В реализации Sklearn PCA и tSNE возвращают набор векторов, соответствующих количеству осей, которые мы указали при работе.

В нашем случае работа с плоскостью, это будут два вектора, каждый из которых имеет длину, равную числу строк в рамках данных, содержащих событие. Эти векторы должны быть переданы функции визуализации, определяя конкретные признаки класса и номера кластера.

В результате был создан график, в котором точки в разных кластерах будут иметь разные цвета, и если мы зададим им необходимые параметры, кресты разных цветов будут над ними.

После создания кластерного решения возникает вопрос о том, насколько оно стабильно и статистически важно. Устойчивая группировка должна сохраняться при изменении методов кластеризации: например, если в результатах иерархического кластерного анализа при группировке с использованием метода k-среднего доля совпадений составляет более 70%, то принимается допуск к устойчивости.[14]

Сравнительная проверка оценивает структуру кластеров путем изменения различных параметров одного алгоритма(например, количество групп k); это делается обычными средствами библиотеки sklearn.

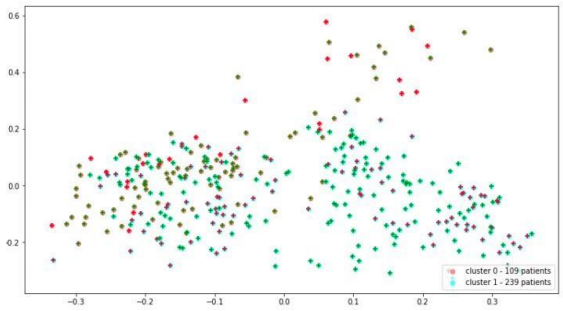

Рисунок 3 – Пример кластеризации

Расчеты статистики Хопкинса по исходному набору неструктурированных текстовых данных показали, что в любом эксперименте значение H не превышало 0,150, что свидетельствует о правильности наших предположений о наличии в данных кластерной структуры. Хотя визуализация методом основных компонентов не может быть визуально построена линия распределения между кластерами, мы видим, что существует определенная тенденция визуального разделения.[15-16]

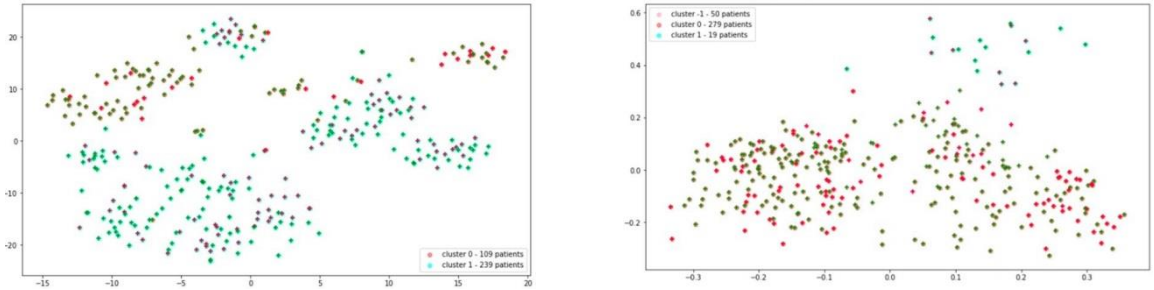

Рисунок 4 – Кластерный эксперимент

Конкретных закономерностей не выявлено: A) и B) кластеры во многих точках сходны и с совершенно разными, но при этом расположены в одном и разных кластерах. На рисунке # 4. попытка B) кластерного эксперимента по визуализации и поиску распределения для извлечения информации из данных была осуществлена с использованием метода tSNE.

Этот метод визуализации показывает выраженную выраженность структуры кластера, однако определить его по методу соседей не представляется возможным. В этом эксперименте можно определить количество кластеров, равное количеству областей, четко выделенных для этой визуализации.Однако добиться соответствия с визуальными кластерами невозможно. В то время применялся принципиально иной подход к кластеризации-на основе плотности, которая автоматически регулирует расстояние. (в частности, реализация hdbscan)[17]

Заключение. Использование технологии Data Mining в области медицины как аспектов применения и методов интеллектуального анализа становится все более эффективным. Задача определения групп ВИЧ-инфицированных пациентов, были зафиксированы все основные моменты реализации, для оценки статистической критерия TF-IDF была использована функция Tfidf Vectorizer из библиотеки sklearn языка Python. Была выбрана статистика Хопкинса, определены склонности сгруппированных данных, проведена кластеризация пациента по

истории болезни. Рассматривались два вида кластеризации. Из подготовленных объектов построены модели и составлен график в соответствии с методологией.

В результате была выявлена определенная тенденция к визуальному делению методом основных компонентов. Все выводы и полученные критерии были записаны и проанализированы.

Все решения, использованные и описанные в работе, применяются для поиска и реализации групп пациентов с другими заболеваниями, которые могут быть как масштабными, так и позволяющими применять их в экспериментах.

СПИСОК ЛИТЕРАТУРЫ

1. Макленнен, Джеми, Чжаохуэй Танг, Богдан Криват,2008. Microsoft SQL Server: Data mining – интеллектуальный анализ данных:– СПб.: БХВ-Петербург, 2009. – 720с.
2. Барсегян А. А. С. Куприянов, И. И. Холод, М. Д. Тесс, С. И. Елизаров, 2009. Анализ данных и процессов: учеб. пособие /– 3-е изд., перераб. и доп. – СПб.: БХВ-Петербург, – 512 с.
3. U. Rajendra Acharya, 2010.Wenwei Yu Data Mining Techniques in Medical Informatics // The Open Medical Informatics Journal. — № 4, p. 21–22
4. Шумейко А. А., Сотник С. Л, Белая Е. А, 2012. Интеллектуальный анализ данных (Введение в Data Mining): учеб. пособ. / Днепропетровск, – 212 с.
5. Sergey Kabanikhin, Olga Krivorotko, Aliya Takuadina Geo-information system of spread of tuberculosis based on inversion and prediction.//Journal of Inverse and III – posed Problems, 2021.
6. Anass Bouchnita, Gennady Bocharov , Andreas Meyerhans and Vitaly Volpert Towards a multiscale model of acute hiv infection. Computation-2017
7. H.Th. Banks, S.I.Kabanikhin, O.I.Krivorotko, D.V.Yermolenko. A numerical algorithm for constructing an individual mathematical model of HIV dynamics at cellular level. Journal of Inverse and ILL-Posed Problems. 2018 V.26(6).P.859-873.
8. E.O.Omondi. A mathematical modelling study of HIV infection in two heterosexual age groups in Kenya. Infectious Disease Modelling. V.4, 2019, P. 83-98.
9. Лисинин А.В.,Файзулин Р.Т., 2015.Применение метаэвристических алгоритмов к решению задач кластеризации методом k-средних // Компьютерная оптика.Т.39, №. 3. — С. 406–412.
10. Андреас Мюллер, 2017. Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными.- М.: Альфа-книга, С.153-170.
11. Нейский И.М., 2006. Классификация и сравнение методов Кластеризации. Интеллектуальные технологии и системы. Сборник учебно-методических работ и статей аспирантов и студентов. — М.: НОК «CLAIM»,. —Выпуск 8. — С. 130–142.
12. Барсегян А.А., Куприянов М.С., Степаненко В.В., Холод И.И., 2008.Методы и модели анализа данных // OLAP и DataMining. СПб.: БХВ-Петрбург,-336с.
13. Петрунин Ю. Ю., 2010. Информационные технологии анализа данных.// Анализ данных. – М: КДУ, -С.180-200.
14. Плас, Джейк Вандер, 2018. Python для сложных задач. Наука о данных и машинное обучение. Руководство. // - М.: Питер,
15. Charikar, M. E. 2004.Incremental clustering and dynamic information retrieval.// SIAM Journal on Computing. — Vol. 33, №. 6. — P. 1417–1440.
16. Мокина Е.Е., Марухина О.В., Шагарова М.Д., Дубинина И.А., 2016. Использование методов Data Mining при принятии медицинских диагностических решений.// фундаментальные исследования. — № 5-2. – с. 269-274;
17. Kubegenova, A.D., Zhakhiena, A.G., Baigubenova, S.K., Utyasheva, G.S., Omarov, A.N. Clustering and data mining on the example of hiv-infected people data. (2022) Journal of Theoretical and Applied Information Technology, 100 (13), pp. 5010-5018.
18. Roeger L.V., Feng Z., Castillo-Chaves.C. Modelling TB and HIV Co-infections.Mathematical biosciences and engineering, 2009, vol.6, no.4, pp.815-837, doi:10.3934/mbe.2009.6.815
19. Kubegenova, A.D., Kubegenov, E.S., Gumarova, Z.M., Kamalova, G.A., Zhazykbaeva,

G.M. Using Data Mining Technology in Monitoring and Modeling the Epidemiological Situation of the Human Immunodeficiency Virus in Kazakhstan (2022) Communications in Computer and Information Science, 1703 CCIS, pp. 57-65.DOI: 10.1007/978-3-031-21340-3_6

20. А.Кубегенова, К. Искаков, Е. Кубегенов, О. Криворотько. Мониторинг и моделирование эпидемиологической ситуации с помощью интеллектуального анализа данных.№4 (2022) Известия НАН РК.Серия информатика.С.44-55.

ТҮЙІН

Мақалада деректерді зерттеу, деректерді терең талдау, білім алу, білім қорындағы деректерді өңдеу әдістері, интеллектуалды талдау әдістері және медицина саласындағы Data Mining технологиясын қолдану аспектілері қарастырылады. АИТВ жұқтырған пациенттер тобы бөлінді, ауру тарихына талдау жүргізілді, іс-қимыл модельдері мен алгоритмі (кіріс деректері) әзірленді, деректерді іздеу әдістерімен талдау және эксперименттер жүргізілді. Барлық аурулар сандық векторлар жиынтығы ретінде ұсынылды және кластерлерге топтастырылды, сипатталған техникаға сәйкес, Хопкинс статистикасының мәні осы үлестіру арқылы есептелді. Кластерлеудің өзі әдеттегі sklearn кітапхана құралдарының көмегімен жүзеге асырылды. Екі өлшемді жазықтықта көп өлшемді деректерді ұсынудың әртүрлі әдістері, негізгі компоненттер әдісі, кохонен сызығы және т. б. ұсынылған.

Кластерлеудің екі түрлі әдісі қарастырылды: k-орташа әдісі (Python sklearn кітапханасының Kmeans функциясын қолдана отырып), АВТО конфигурациясы бар тығыздыққа негізделген кластерлеу әдістері (Python hdbscan кітапханасының HDBSCAN функциясын қолдана отырып). Салыстыру жағдайында кластерлік құрылым бір алгоритмнің әртүрлі параметрлерін өзгерту арқылы бағаланады (мысалы, k топтарының саны); алынған және дайындалған объектілерде модель (немесе бірнеше) салынып, оның параметрлері реттеледі. Содан кейін тестілеу және нәтижелерді талдау жүргізіледі.

АЛҒЫС

Бұл жариялым Erasmus+ ACeSYRI «Компьютерлік ғылымдар саласындағы докторанттар мен жас зерттеушілерге арналған озық орталық» жобасын іске асыру нәтижесі, тіркеу нөмері 610166-EPP-1-2019-1-SK-EPPKA2-CBHE-JP.

UDK 004.5
IRSTI 50.41.29

Levashenko V., Ph.D., head of the Dept.of Informatics, Faculty of Management Science and Informatics, University of Zilina, vitally.levashenko@fri.uniza.sk
Ainura G.A., master of Engineering sciences, https://orcid.org/ 0000-0002-3702-3260, teacher of the high school of Information technologies, Zhangir Khan West Kazakhstan Agrarian-Technical University, str. Zhangir Khan, 51, 090009, Kazakhstan, g_ainura_91@mail.ru
Kamalova G.A., candidate of Physical and Mathematical Sciences, https://orcid.org/0000-0002-5252-4573, Associate Professor of Higher School of Information Technologies, Zhangir Khan West Kazakhstan Agrarian-Technical University, str. Zhangir Khan, 51, 090009, Kazakhstan, gokhakam@gmail.com

FEATURES OF USING THE BOOTSTRAP 4 FRAMEWORK IN THE DESIGN OF AN ADAPTIVE WEBSITE

ANNOTATION

This article discusses the concept of adaptive design in the process of creating a website. The essence of the frontend frameworks of their varieties is revealed, as well as a detailed analysis of the Bootstrap framework is described. It is concluded that Bootstrap frameworks and the like make the web development process simpler and facilitate the development of an adaptive website design. The world of web programmers has already seen Bootstrap 5, but many are still interested in what Bootstrap 4 has brought in itself. Because it was this version that brought a new round of technologies