

**ПРОГНОЗ ИСПАРЕНИЯ ПОДЗЕМНЫХ ВОД С ИСПОЛЬЗОВАНИЕМ
МНОЖЕСТВЕННОЙ ЛИНЕЙНОЙ РЕГРЕССИИ**

Яхшибоев Р.Э., Хусанов У.А., Довлетова С.Б., Сиддиқов Б.Н.

Аннотация - Модели регрессионного анализа и классификации данных временных рядов на основе алгоритмов машинного обучения позволяют решать задачи прогнозирования состояния региона в различных сферах, в том числе в сельском хозяйстве. Одной из проблем в этой сфере является засоление почв, одна из основных причин засоления которых связана с повышением уровня грунтовых вод. Эта статья посвящена определению модели для прогнозирования испарения подземных вод с использованием метода линейной регрессии с несколькими переменными с использованием географических данных из региона. В качестве исходных данных для построения модели использовались данные Хорезмской области за период с 1980 по 2010 годы, и на основе этих данных была сформирована обучающая выборка. Был проведен корреляционный анализ для изучения взаимосвязи между переменными выборки и моделью линейной регрессии с тремя переменными, состоящей из осадков, испарение воды и температура воздуха использовались для прогнозирования уровня грунтовых вод и повышения точности модели. В статье также представлен метод очистки данных в обучающей выборке от помех.

Ключевые слова – анализ, испарение воды, прогноз, линейная регрессия, геоинформационные данные, МНК

ВВЕДЕНИЕ

Обработка информации о подземных водах является сложной задачей. Оценка состояния подземных вод неразрывно связана со многими экосистемами, такими как зарядка, засоление, уровень мелководья и явления поверхностных вод [1]. Поскольку ресурс находится под землей, его отслеживание или измерение — сложный процесс. Самый распространенный способ оценки ресурса — посмотреть на уровень грунтовых вод, измеренный в этих колодцах. Для водных ресурсов использование и практика искусственного интеллекта может дать отличные результаты.

Сегодня в мире растет спрос на питьевую воду и в связи с этим возникает проблема нехватки воды. Более высокие температуры и небольшое количество осадков приводят к засухе. Для решения этой задачи разрабатываются современные технологии, программы, математические модели и алгоритмы. В данной работе представлены результаты исследований по построению модели

прогнозирования подъема грунтовых вод методом линейной регрессии с множественными переменными по геопространственным данным региона.

МАТЕРИАЛЫ И МЕТОДЫ

Пространственные данные составляют основу информационного обеспечения геоинформационных систем. Современный анализ геопространственных данных позволяет совместить геоинформационную систему с бизнес-аналитикой, что приводит к качественному, быстрому принятию решений за счет сокращения времени на поиск и анализ необходимой информации. Пространственный анализ позволяет использовать карту в качестве одного из стандартных измерений, таких как время. [1,9]

Линейная регрессия является одним из наиболее важных и широко используемых методов регрессии. Это самый простой метод регрессии. Одним из его преимуществ является простота интерпретации результатов.

Линейная регрессия некоторой зависимой переменной y по набору независимых переменных $x = (x_1, \dots, x_r)$, где r — количество предикторов, предполагает линейную зависимость между y и x : $y = \beta_0 + \beta_1 x_1 + \dots + \beta_r x_r + \varepsilon$. Это уравнение регрессии. [2,10]

β_0, β_1, \dots , — коэффициенты регрессии, а ε является случайной ошибкой.

Линейная регрессия вычисляет оценочные функции коэффициентов регрессии или просто предсказанные веса измерения, обозначаемые как b_0, b_1, \dots, b_r .

Они определяют оценочную функцию регрессии $f(x) = b_0 + b_1 x_1 + \dots + b_r x_r$. Эта функция достаточно хорошо фиксирует зависимости между входными и выходными данными. [3,11]

Для каждого результата наблюдения $i = 1, \dots, n$ предполагаемый или прогнозируемый ответ $f(x_i)$ должен быть как можно ближе к соответствующему фактическому ответу y_i .

Разности $y_i - f(x_i)$ для всех наблюдений называются остатками. Регрессия определяет наилучшие предсказанные веса измерений, которые соответствуют наименьшим остаткам. [4,12]

Чтобы получить наилучшие веса, вам нужно минимизировать сумму остаточных квадратов (SSR) для всех наблюдений:

$$SSR = \sum_i (y_i - f(x_i))^2.$$

Этот подход называется методом наименьших квадратов.

Пакеты Python для линейной регрессии:

NumPy — это фундаментальный научный пакет для быстрых операций с одномерными и многомерными массивами. NumPy упрощает математическую рутину и, конечно же, имеет открытый исходный код. [5,13]

Пакет `scikit-learn` — это библиотека, широко используемая в машинном обучении. `scikit-learn` предоставляет значения для предварительной обработки данных, уменьшения размера, реализует регрессию, классификацию, кластеризацию и т. д. Он с открытым исходным кодом, как и `NumPy`.

Вы можете указать несколько необязательных параметров для класса `LinearRegression`:

- `fit_intercept` — логический (по умолчанию `True`) параметр, который определяет, следует ли вычислять сегмент b_0 (`True`) или считать его равным нулю (`False`).
- `normalize` — логический (по умолчанию `False`) параметр, который определяет, следует ли нормализовать входные переменные (`True`) или нет (`False`).
- `copy_X` — логический (по умолчанию `True`) параметр, который определяет, следует ли копировать (`True`) или перезаписывать входные переменные (`False`).
- `n_jobs` — это целое число или `None` (по умолчанию), представляющее количество процессов, задействованных в параллельных вычислениях. `None` означает отсутствие процессов, `-1` использует все доступные процессоры. [6]

РЕЗУЛЬТАТЫ

Была выбрана географическая карта Республики Узбекистан (рис. 1.).

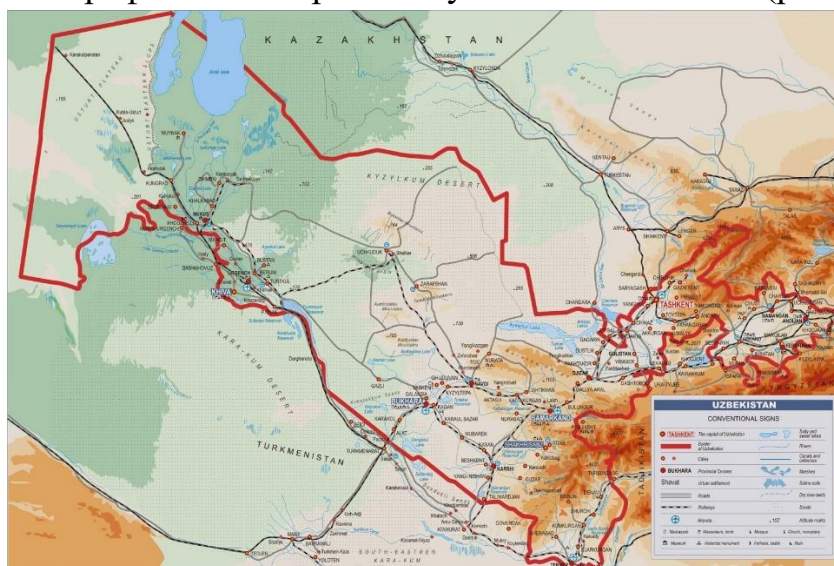


Рис. 1. Географическая карта Узбекистана.

На основе этой карты был подготовлен Набор данных о характере, температуре воздуха и относительной влажности выбранного объекта. На основе созданной математической модели, матрицы и алгоритма выполняются прогнозы на языке программирования `Python`. Все данные в наборе данных анализируются и сравниваются друг с другом. [15,16,17]

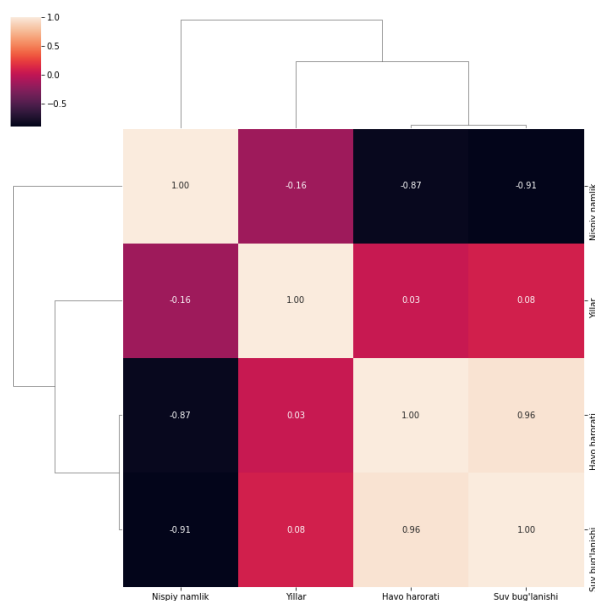


Рис.2. Корреляционная матрица

Данные анализируются с использованием набора данных для выбранной математической модели и корреляционной матрицы. Затем результаты анализа (рис. 2) проверяются на наличие ошибок. Набор данных указан в таблице №1. По умолчанию используются годы, температура воздуха, испарение воды и относительная влажность. [7,15,18]

Таблица 1

	Yillar	Havo harorati	Suv bug'lanishi	Nispiy namlik
0	1980	-5.1	14.26	80
1	1980	-4.8	20.57	72
2	1980	3.6	61.84	58
3	1980	16.9	154.84	51
4	1980	23.3	243.55	42

Анализ на наличие шума производится по отношению к таблице. Шумы — это числа, которые внезапно падают во время анализа. Кроме того, данные, которые значительно отличаются от других точек данных. По этой причине каждый столбец проверяется, чтобы убедиться, что он работает правильно. (Рисунок 3.) Все сравнительные результаты представлены ниже. [8,16,17]

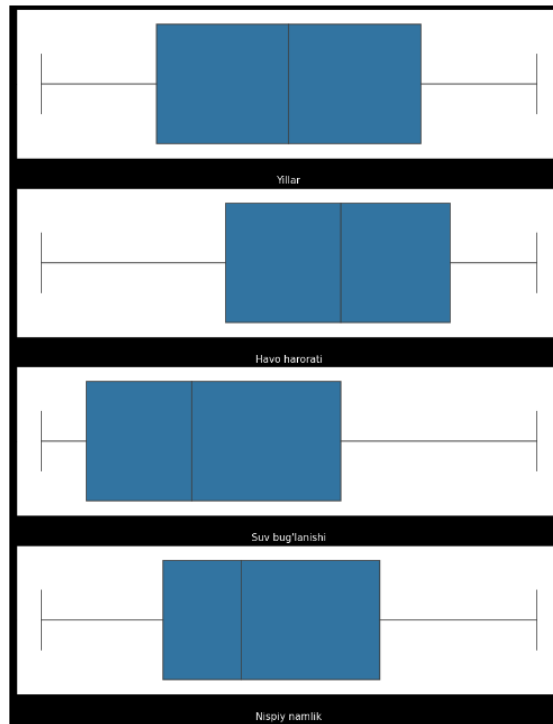


Рис. 3. Процесс анализа шума

После удаления шумов относительная влажность и температура воздуха сравниваются между собой по осям x и y на основе значений, приведенных в таблице. В результате видно, что температура и относительная влажность по осям x и y увеличиваются из года в год. (Рис.4.5)

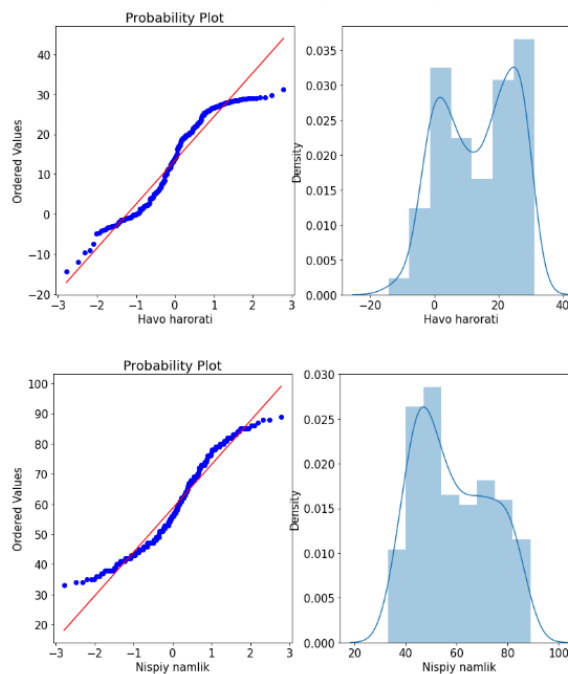


Рис.4. График распределения данных по столбцам.

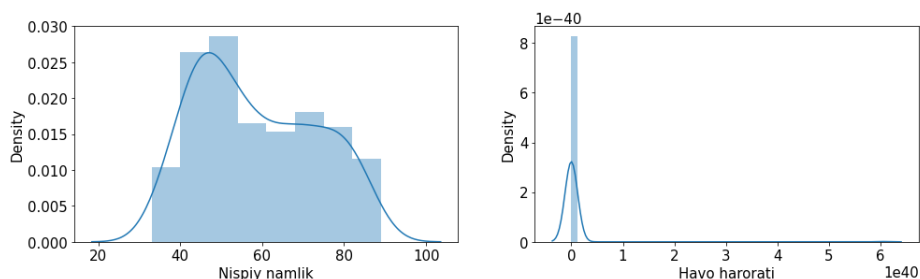


Рис.5. График распределения данных по столбцам.

В результате вы можете увидеть результат сравнения фактических и прогнозируемых результатов. (Рис. 6.). С развитием множественной линейной регрессии можно предсказать год за годом, сколько литров или тонн воды будет испаряться и как изменится баланс питьевой воды.

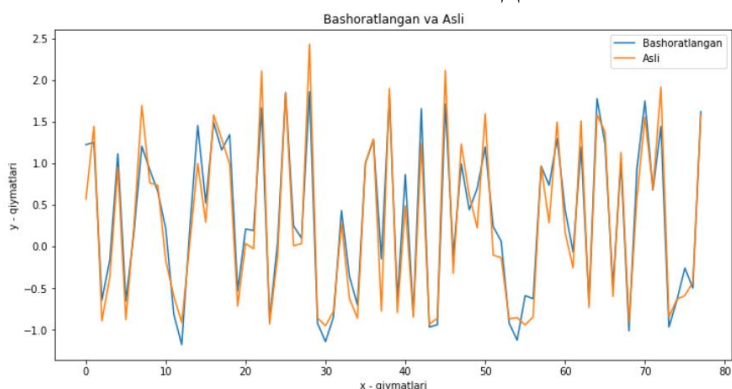


Рис. 6. Результат линейной регрессии

ОБСУЖДЕНИЕ

Модели регрессии описывают линейную зависимость между выбранной переменной и зависимыми переменными. В моделях линейной регрессии используется прямая линия, а в моделях логистической и нелинейной регрессии — кривая. Регрессия предсказывает, как произвольная переменная изменится с изменением свободных переменных. Ниже приведена формула линейной регрессии с несколькими переменными.

$$Y = \beta_0 + \beta_1 X_1 + \beta_2 X_2 + \dots + \beta_n X_n \quad (1)$$

Y - произвольная переменная, β_0 - свободный предел, X - угловой коэффициент
 X_1, X_2, \dots, X_n - произвольная переменная.

В случае линейной регрессии угловой коэффициент определяет влияние этих произвольных переменных на произвольную переменную. Отсюда следует, что чем больше угловой коэффициент произвольной переменной, тем больше степень его влияния на произвольную переменную. Целью модели линейной регрессии является $\beta_0, \beta_1, \beta_2, \dots$. Нахождение β_k есть. Мы использовали метод обычных наименьших квадратов (OLS) для определения оптимальных коэффициентов.

Модели линейной регрессии показывают более точные результаты, когда обучающая выборка свободна от шума и выбраны наиболее релевантные столбцы. Поэтому приведенная выше модель линейной регрессии была сначала обработана перед построением.

В используемой модели линейной регрессии с несколькими переменными в качестве вынужденной переменной выбрано повышение уровня грунтовых вод, а в качестве независимых переменных – температура воздуха, осадки и испарение воды.

На основе приведенной выше формулы была построена структура модели многомерной линейной регрессии.

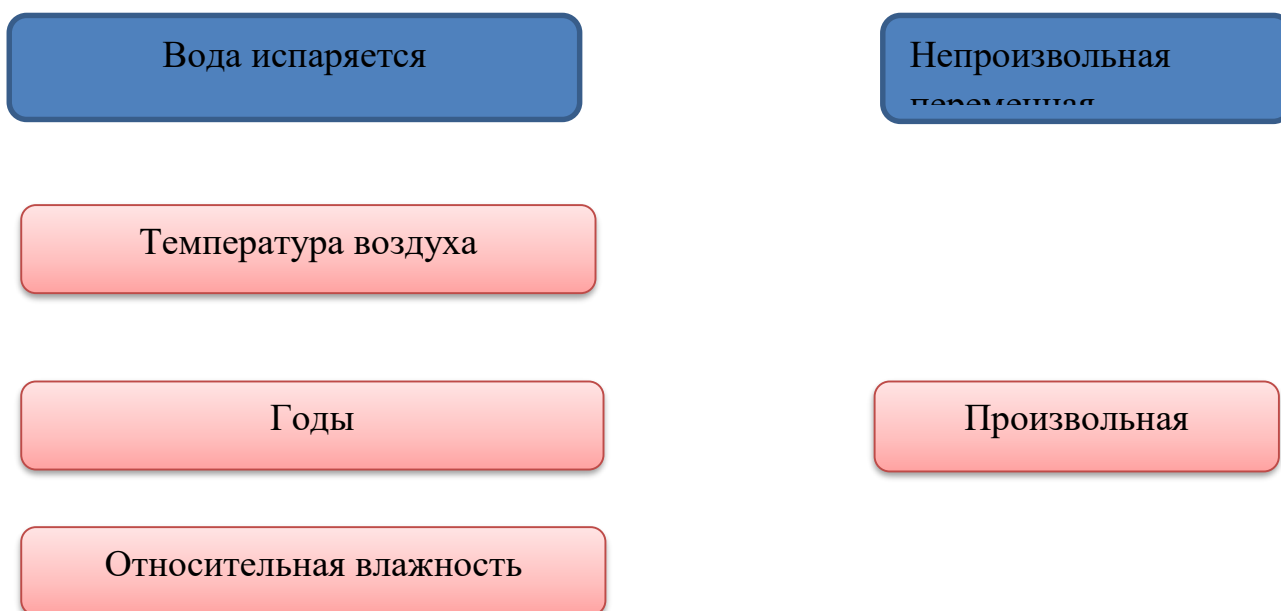


Рис.7. Структурное моделирование

ЗАКЛЮЧЕНИЕ

В заключение можно отметить, что без учета человеческого фактора можно предсказать, какие изменения произойдут в природе в будущем. Это видно из результатов вышеприведенного анализа. Если наука и техника будут продолжать развиваться, можно будет предотвратить стихийные бедствия и засухи. Благодаря развитию искусственного интеллекта можно добиться хороших результатов в различных сферах.

Сегодня сельское хозяйство является одной из важнейших отраслей экономики. Эффективность сельскохозяйственного производства также напрямую связана с влажностью почвы и воздуха. Данное исследование направлено на прогнозирование уровня испарения воды, благодаря чему можно добиться ряда успехов в повышении продуктивности за счет прогнозирования состояния влажности почвы и воздуха.

В заключение можно спрогнозировать, какие изменения произойдут в природе в будущем без человеческого фактора. Это видно и из результатов вышеприведенного анализа. Благодаря развитию искусственного интеллекта можно добиться хороших результатов в различных сферах.

БЛАГОДАРНОСТЬ

Работа выполнена при поддержке Ташкентского университета информационных технологий и лаборатории геоинформационных технологий.

ИСПОЛЬЗОВАННАЯ ЛИТЕРАТУРА

1. Рахмани, Ф., Лоусон, К., Оуян, 545 В., Аплинг, А., Оливер, С. и Шен, К.: Изучение исключительной производительности модели температуры потока с глубоким обучением и ценности потока потока Данные, Окружающая среда. Рез. Lett., <https://doi.org/10/ghsw9p>, 2020.
2. Раджаи Т., Эбрахими Х. и Нурани В.: Обзор методов искусственного интеллекта в моделировании уровня подземных вод, Журнал гидрологии, <https://doi.org/10/gfvfg3>, 2019.
3. Раут М., Штайнер Х., Ридигер У., Мазуркевич А. и Грацки А. Климатология осадков в Центральной Европе. (HYRAS), Метеорол. З., с. 22, <https://doi.org/10/f5gf49>, 2013 г.
4. Ребак Дж., МакКинни В., Джброкмендель, Босше ДВД, Аугспургер Т., Клауд П., Гфьюнг, Синхркс, Клейн А., Решке М., Хокинс С., Тратнер Дж., Она, К., Айд, В., Петерсен, Т., Гарсия, М., Шендель, Дж., Хейден, А., MomIsBestFriend, Янкаускас, В., Баттистон, П., Сиболд, С., Крис -В1, Н-Vetinari, Hoyer, S., Overmeire, W., Alimcmaster1, Dong, K., Whelan, C. и Mehvar, M.: Pandas-Dev / Pandas: Pandas 1.0.3, Zenodo, <https://doi.org/10.5281/ZENODO.3509134>, 2020.
5. Регион Эльзас - Страсбург: Bestandsaufnahme Der Grundwasserqualität Im Oberrheingraben / Inventaire de La Qualité Des Eaux Souterraines Dans La Vallée Du Rhin Supérieur, 1999.
6. Шен, К.: Междисциплинарный обзор исследований глубокого обучения и его актуальность для ученых-водников, Water Resour. рез., 54, 8558-8593, <https://doi.org/10/gd8cqb>, 2018.
7. Судхир, К.П., Наяк, П.С., и Рамасастри, К.С.: Улучшение оценок пикового расхода в моделях речного стока с искусственной нейронной сетью, Гидрологические процессы, 17, 677-686, <https://doi.org/10/b39k4k>, 2003.
8. Закон Республики Узбекистан «Об информатизации», 12.11.2003 г., № 560-П, Ташкент Собрание законодательства Республики Узбекистан, 2014 г., № 36.
9. Постановление Президента Республики Узбекистан №ПП-4642 от 17.03.2020 г. «О мерах по широкому внедрению цифровых технологий в городе Ташкенте» Национальная база данных законодательства, 18.03.2020 г., №07/ 20/4642/0328

10. Безручко Б.П., Смирнов Д.А. Математическое моделирование и хаотические временные ряды. — Саратов: ГосУНЦ «Колледж», 2005. — ISBN 5-94409-045-6.
11. Khamzaev J. et al. DRIVER SLEEPINESS DETECTION USING CONVOLUTION NEURAL NETWORK //CENTRAL ASIAN JOURNAL OF EDUCATION AND COMPUTER SCIENCES (CAJECS). – 2022. – Т. 1. – №. 4. – С. 31-35.
12. Яхшибоев Р. и др. РАЗРАБОТКА МОДЕЛИ РАСПОЗНАВАНИЯ ГРАФИЧЕСКИХ ОБЪЕКТОВ НА ОСНОВЕ МЕТОДА «TRANSFER LEARNING» ДЛЯ ДИАГНОСТИКИ В СФЕРЕ ЗДРАВООХРАНЕНИЯ //Conferences. – 2022. – С. 156-164.
13. Yaxshiboyev R. DEVELOPMENT OF A SOFTWARE AND HARDWARE COMPLEX FOR PRIMARY DIAGNOSTICS BASED ON DEEP MACHINE LEARNING //CENTRAL ASIAN JOURNAL OF EDUCATION AND COMPUTER SCIENCES (CAJECS). – 2022. – Т. 1. – №. 4. – С. 20-24.
14. Яхшибоев Р., Базарбаев М., Эрметов Э. ПРИМЕНЕНИЕ НЕЙРОННЫХ СЕТЕЙ ДЛЯ ДИАГНОСТИКИ МЕДИЦИНСКИХ ЗАБОЛЕВАНИЙ //Conferences. – 2022. – С. 148-156.
15. Yaxshiboyev R. DEVELOPMENT OF A MODEL OF OBJECT RECOGNITION IN IMAGES BASED ON THE «TRANSFER LEARNING» METHOD //CENTRAL ASIAN JOURNAL OF EDUCATION AND COMPUTER SCIENCES (CAJECS). – 2022. – Т. 1. – №. 4. – С. 36-41.
16. Яхшибоев Р. Э., Очилов Т. Д., Сиддигов Б. Н. РАЗРАБОТКА ПРОГРАММНОГО СРЕДСТВА ДЛЯ ИДЕНТИФИКАЦИИ НОМЕРНЫХ ЗНАКОВ ТРАНСПОРТНЫХ СРЕДСТВ НА ОСНОВЕ МЕТОДОВ КОМПЬЮТЕРНОГО ЗРЕНИЯ //Journal of new century innovations. – 2022. – Т. 15. – №. 1. – С. 81-93.
17. Яхшибоев Р., Сиддигов Б. ЦИФРОВЫЕ ТЕХНОЛОГИИ ДЛЯ ПЕРВИЧНОЙ ДИАГНОСТИКИ РАЗНЫХ МЕДИЦИНСКИХ ЗАБОЛЕВАНИЙ //Innovations in Technology and Science Education. – 2022. – Т. 1. – №. 4. – С. 94-105.
18. Yaxshiboyev R., Yaxshiboyeva D. ANALYSIS OF ALGORITHMS FOR PREDICTION AND PRELIMINARY DIAGNOSTICS OF GASTROENTEROLOGICAL DISEASES //CENTRAL ASIAN JOURNAL OF EDUCATION AND COMPUTER SCIENCES (CAJECS). – 2022. – Т. 1. – №. 2. – С. 49-56.
19. Djumanov J. et al. Mathematical model and software package for calculating the balance of information flow //2021 International Conference on Information Science and Communications Technologies (ICISCT). – IEEE, 2021. – С. 1-6.