



ЖАСАНДЫ ИНТЕЛЛЕКТ
ИСКУССТВЕННЫЙ ИНТЕЛЛЕКТ
ARTIFICIAL INTELLIGENCE

DOI 10.51885/1561-4212_2023_4_68
MPHTI 28.23.29

Ш.Е. Ахметжанова¹, А.А. Хасенова², Ж.О. Оралбекова³, Л.Т. Курмангазиева⁴

¹Таразский региональный университет имени М.Х. Дулати, г. Тараз, Казахстан
E-mail: shina_70@mail.ru

²Карагандинский университет имени академика Е.А. Букетова, г. Караганда, Казахстан
E-mail: aigera_0089@mail.ru*

³Евразийский национальный университет имени Л.Н. Гумилева, г. Астана, Казахстан
E-mail: oralbekova@bk.ru

⁴Атырауский университет имени Х. Досмухамедова, г. Атырау, Казахстан
E-mail: kurmangazieval@mail.ru

ИНТЕЛЛЕКТУАЛЬНЫЙ АНАЛИЗ КЛИМАТИЧЕСКИХ ПАРАМЕТРОВ ПОЧВЫ ТОПЫРАҚТЫҢ КЛИМАТТЫҚ ПАРАМЕТРЛЕРІН ИНТЕЛЛЕКТУАЛДЫ ТАЛДАУ INTELLIGENT ANALYSIS OF SOIL CLIMATIC PARAMETERS

Аннотация. В данной статье представлен интеллектуальный подход к анализу параметров почвы с использованием методов машинного обучения и статистического анализа данных. Были проанализированы сезонные тенденции изменения показателей почвы. Результаты кластеризации выявили регионы со схожими параметрами влажности и температуры воздуха. Для прогнозирования временных рядов параметров влажности почвы, освещенности и температуры воздуха использовались методы ARIMA (авторегрессионная интегрированная скользящая средняя) и LSTM (долгая краткосрочная память). Интеллектуальный подход к анализу параметров почвы демонстрирует эффективность и перспективность применения методов машинного обучения и анализа данных в сельском хозяйстве. Такой подход может быть полезен для улучшения управления земельными ресурсами, повышения урожайности и устойчивого развития сельского хозяйства. В ходе исследования использовался язык программирования Python, который обладает богатым набором библиотек и модулей для анализа данных, что позволяет гибко подходить к решению сложных задач и создавать настраиваемые решения, наилучшим образом соответствующие требованиям проекта.

Ключевые слова: интеллектуальный анализ данных, машинное обучение, кластеризация, визуализация данных, параметры почвы, точное земледелие, прогнозирование.

Аңдатпа. Бұл мақалада машиналық оқыту әдістерін және статистикалық деректерді талдауды пайдалана отырып, топырақ параметрлерін талдаудың интеллектуалды тәсілі ұсынылған. Топырақ көрсеткіштерінің өзгеруінің маусымдық тенденциялары талданды. Кластерлеу нәтижелері бойынша ылғалдылық пен ауа температурасы бойынша ұқсас аймақтар анықталды. Топырақ ылғалдылығы, жарық және ауа температурасы параметрлерінің уақыт қатарын болжау үшін ARIMA және LSTM әдістері қолданылды. Топырақ параметрлерін талдаудың интеллектуалды тәсілі ауыл шаруашылығында машиналық оқыту әдістерін және деректерді талдауды қолданудың тиімділігі мен болашағын көрсетеді. Мұндай тәсіл жерді басқаруды жақсарту, ауыл шаруашылығы дақылдарының өнімділігін арттыру және ауыл шаруашылығының тұрақты дамуы үшін пайдалы болуы мүмкін. Зерттеу барысында күрделі мәселелерді шешуге икемді түрде келуге және жобаның талаптарына барынша сәйкес келетін реттелетін шешімдерді жасауға мүмкіндік беретін деректерді талдауға арналған кітапханалар

мен модульдердің бай жиынтығы бар Python бағдарламалау тілі пайдаланылды.

Түйін сөздер: деректерді интеллектуалды талдау, машиналық оқыту, кластерлеу, деректерді визуализациялау, топырақ параметрлері, дәл ауыл шаруашылығы, болжау.

Abstract. This article presents an intelligent approach to the analysis of soil parameters using machine learning methods and statistical data analysis. Seasonal trends in changes in soil indicators were analyzed. The results of clustering revealed regions with similar parameters of humidity and air temperature. ARIMA and LSTM methods were used to predict the time series of soil moisture, light and air temperature parameters. An intelligent approach to the analysis of soil parameters demonstrates the effectiveness and prospects of using machine learning methods and data analysis in agriculture. Such an approach can be useful for improving land management, increasing crop yields and sustainable agricultural development. The study used the Python programming language, which has a rich set of libraries and modules for data analysis, which allows you to flexibly approach solving complex problems and create custom solutions that best suit the requirements of the project.

Keywords: data mining, Machine learning, Clustering, Data visualization, Soil parameters, Precision Agriculture, Forecasting

Введение. В сельском хозяйстве все больше используются технологии интернета вещей (IoT), датчики и автоматизированные системы. Мониторинг состояния почвы на основе анализа данных таких систем становится неотъемлемой частью процесса принятия решений в сельском хозяйстве и помогает сельскохозяйственным предприятиям перейти к более эффективным и точным методам управления [1]. Продуктивность сельского хозяйства в значительной мере зависит от состояния почвы, характеризующейся как химико-биологическим составом, так и климатическими показателями, такими как влажность, уровень освещенности и температура. Использование недорогих датчиков для получения и анализа этих данных может повысить производительность сельского хозяйства, снизить затраты и способствовать внедрению устойчивых методов ведения сельского хозяйства, которые могут помочь фермерам принимать обоснованные решения об ирригации, внесении удобрений и других методах управления [2].

Применение методов интеллектуального анализа данных на основе анализа климатических параметров состояния почвы имеет несколько ключевых актуальных применений:

– *оптимизация урожайности:* зная точные параметры состояния почвы, сельскохозяйственные производители могут оптимизировать полив, выбирать подходящие культуры и контролировать условия роста;

– *снижение затрат и рисков:* сельскохозяйственные предприятия могут применять удобрения и воду более точно, исходя из реальных потребностей растений;

– *сельскохозяйственное управление:* сельскохозяйственные производители могут принимать своевременные меры и управлять производством, чтобы минимизировать потери;

– *экологическая устойчивость:* аккуратное использование воды и удобрений, что способствует устойчивости экосистем и окружающей среды;

– *принятие информированных решений:* анализ данных о состоянии почвы предоставляет сельскохозяйственным предприятиям ценную информацию для принятия решений. Они могут использовать эти данные для планирования будущих сезонов, выбора наилучших сортов культур и прогнозирования возможных проблем.

Применению методов интеллектуального анализа данных в сельском хозяйстве уделяется значительное внимание [3-25]. Согласно этим исследованиям, наиболее широко распространенными методами и техниками анализа являются нейронные сети, Decision tree (DT), градиентный бустинг (GBT), K-Nearest Neighbor, Support vector machine, K-means, и гибридные методы решающих деревьев и искусственных нейронных

сетей. Большая часть исследований посвящена применению Data Mining в прогнозировании благоприятных условий для выращивания определенных агрокультур. С использованием этих методов были решены такие задачи в области сельского хозяйства, как обнаружение вредителей, определение, классификация и прогнозирование болезней сельскохозяйственных культур, прогнозирование урожайности, управление затратами (планирование орошения и пестицидов), рекомендации по удобрениям [11], а также прогнозирование влажности почвы в режиме реального времени [12].

В работе [13] автор разработал эмпирические модели, прогнозирующие урожай кукурузы и сои в центральной части Соединенных Штатов, учитывая географические и межгодовые вариации фенологии сельскохозяйственных культур. Автор [15] применил регрессионный анализ, основанный на ансамблевых деревьях решений и искусственной нейронной сети, для определения взаимосвязи образцов листьев в зависимости от различных азотных удобрений с параметрами урожайности. В работе [14] была разработана модель на основе R2U-Net, более устойчивая и надежная для разных видов агрокультур (точность 97,59-99,96 %).

Материалы и методы исследования. Рассматриваемая работа выполнялась в университете Данди (Великобритания) в рамках научной стажировки по международной программе «Болашак». Цель исследования заключалась в изучении и анализе данных почвенных сенсоров, установленных в различных сельскохозяйственных угодьях.

Обсерватория GROW собирала данные [10] со всей Европы, используя недорогие почвенные датчики для сбора данных о почве и условиях выращивания. В 22 сельскохозяйственных угодьях установлено 6500 датчиков Flower Power. Для этого были привлечены различные общественные группы, имеющие какое-либо отношение к угодьям: управляющие земельных участков, малых и семейных ферм, научные и образовательные учреждения, имеющие доступ к участкам земель, фермеры виноградников, агролесоводств и т.д. В наборе представлены данные с ноября 2017 г. по октябрь 2019 г. Данные собирались датчиками каждые 15 минут и загружались участниками на серверы GROW с помощью мобильных телефонов. В наборе представлены следующие показатели:

- Влажность почвы (%);
- Температура (C);
- Уровень освещенности (люмены);
- Широта и долгота установки датчика;
- Название датчика;
- Дата наблюдения.

Применение интеллектуальных методов анализа и машинного обучения позволило провести анализ данных по сезонам, по пространственному положению, а также кластерный анализ.

1. *Сезонный анализ.* Набор данных охватывает период с ноября 2017 г. по октябрь 2019 г., что позволяет проводить сезонный анализ данных о влажности почвы, температуре и уровне освещенности. Путем агрегирования данных по сезонам (зима, весна, лето, осень) можно определить закономерности и тенденции в условиях для выращивания агрокультур.

2. *Пространственные закономерности.* Пространственная информация (широта, долгота) датчиков может использоваться для определения пространственных закономерностей влажности почвы, температуры и уровней освещенности в разных местах. Это может помочь определить области с благоприятными или неблагоприятными условиями для выращивания, а также получить представление о факторах, влияющих на

влажность почвы, температуру и уровень освещенности.

3. *Кластерный анализ.* K-Means, иерархическая кластеризация используются для идентификации групп датчиков с похожими условиями выращивания или для группировки датчиков на основе их временных характеристик влажности почвы, температуры и уровней освещенности. Это может помочь определить регионы с похожими условиями выращивания, что может быть полезно для принятия решений в сельском хозяйстве, управлении земельными ресурсами и других смежных областях.

4. *Визуализация данных.* Методы визуализации данных, такие как линейные диаграммы, гистограммы, тепловые и географические карты, можно использовать для визуального изучения и передачи закономерностей и тенденций в наборе данных обсерватории GROW. Визуализации могут обеспечить интуитивно понятное и информативное представление данных, упрощая выявление закономерностей, тенденций и выбросов. Визуализация также может помочь в передаче результатов заинтересованным сторонам и лицам, принимающим решения, и может способствовать принятию решений на основе данных в сельском хозяйстве и управлении земельными ресурсами.

Процесс интеллектуального анализа данных предполагает выполнение нескольких этапов, включая чтение данных, предобработку данных (удаление дубликатов, удаление или замена отсутствующих данных, свертывание, транспонирование и т.д.), построение модели, применение алгоритмов машинного обучения, проверку корректности и точности модели, а также применение полученной модели для решения конкретных задач в сельском хозяйстве.

Рынок интеллектуального анализа больших данных представлен широким выбором готовых инструментов. Однако готовые программные средства могут быть ограничены в функциональности и поддержке. Огромное количество библиотек и модулей для анализа данных, что позволяет гибко подходить к решению сложных задач, предоставляет Python. Этот язык обеспечивает большую свободу в выборе и комбинации различных инструментов для интеллектуального анализа данных, что способствует созданию настраиваемых решений, соответствующих требованиям проекта. В связи с этим Python был выбран для решения задачи кластеризации.

Для анализа климатических параметров почвы были использованы данные из двух файлов формата Excel: GrowTimeSeries и GrowLocations.

В ходе исследовательского анализа мы выяснили, в каких странах были установлены датчики. Для этого использовался сервис геокодирования Nominatim, поддерживаемый библиотеками Python. Результаты представлены в табл. 1.

Таблица 1. Количество датчиков по странам

Country	SensorCount
Croatia	290
Hungary	39
Ireland	30
Luxembourg	6
United Kingdom	12
Unknown	104

После импорта данных в датафрейм были проведены операции по удалению строк

дубликатов, всех строк, содержащих пропущенные значения или значения, не входящие в допустимый диапазон. Важной частью предобработки данных является обработка пропущенных значений (NaN). В данном случае NaN значения были заменены средними значениями показателей температуры и освещенности с датчиков, расположенных в радиусе 1 километра, а для показателей влажности был взят радиус 500 метров.

Предобработка данных предполагает применение группировки для разрозненных данных, в частности, показателей влажности, освещения и температуры с каждого датчика, и дальнейшего транспонирования в удобочитаемый формат (пример предобработанных данных представлен в табл. 2).

Таблица 2. Предобработанные данные

Serial	Latitude	Longitude	year_month	Air Temperature	Light	SoilMoisture
PI040290AA00000000	16.452	45.799	2019-07	28.352973	1.039683	1.592528
PI040290AA00000000	16.452	45.799	2019-08	27.564899	4.475712	21.689524
PI040290AA00000000	16.452	45.799	2019-09	19.078165	2.708485	11.389120
PI040297AA3I001108	19.813	48.004	2018-10	11.309444	9.646110	21.007143
PI040297AA3I001108	19.813	48.004	2018-11	17.158127	16.085093	20.663406
...

Визуализация данных является важным методом в процессе анализа данных, так как позволяет выявить различные тенденции, особенности или закономерности в данных. Приведенные на рис. 1 графики и диаграммы предоставляют информацию о динамике изменения влажности почвы и температуры воздуха.

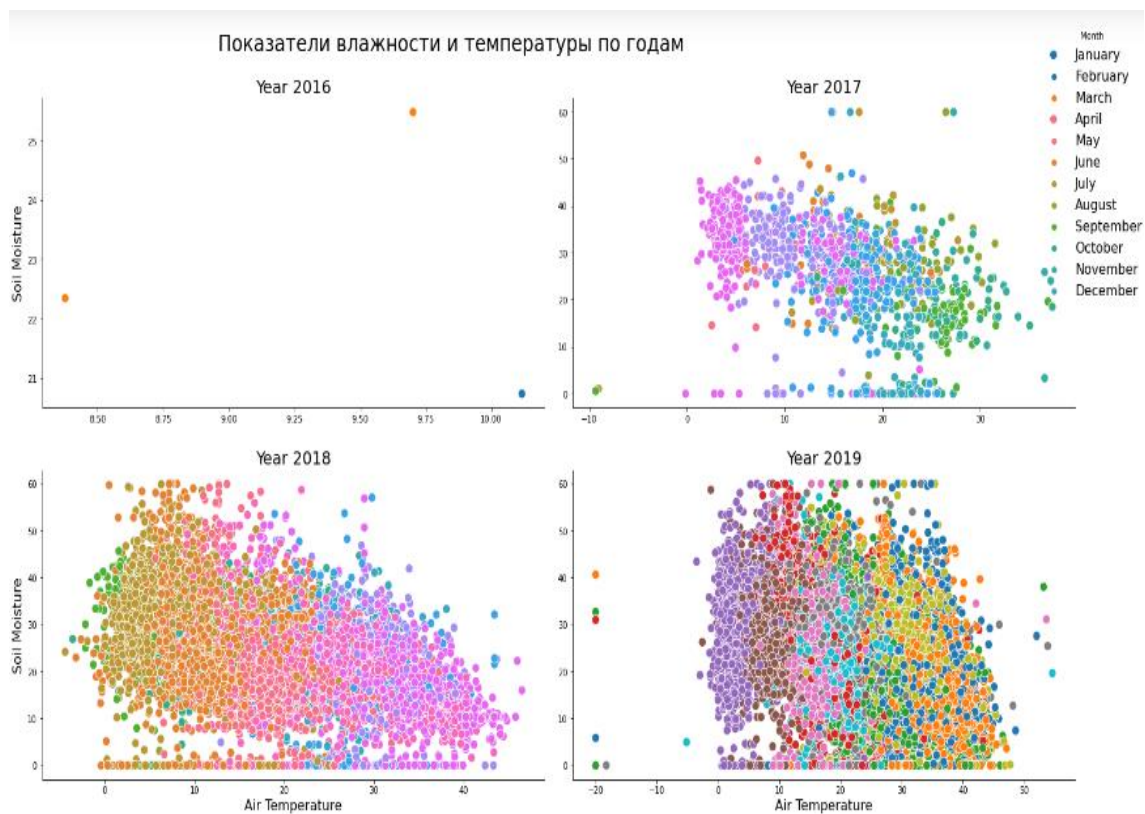


Рисунок 1. Динамика изменения влажности почвы и температуры воздуха по годам
Сезонные тенденции изменения показателей почвы визуализированы на рис. 2.

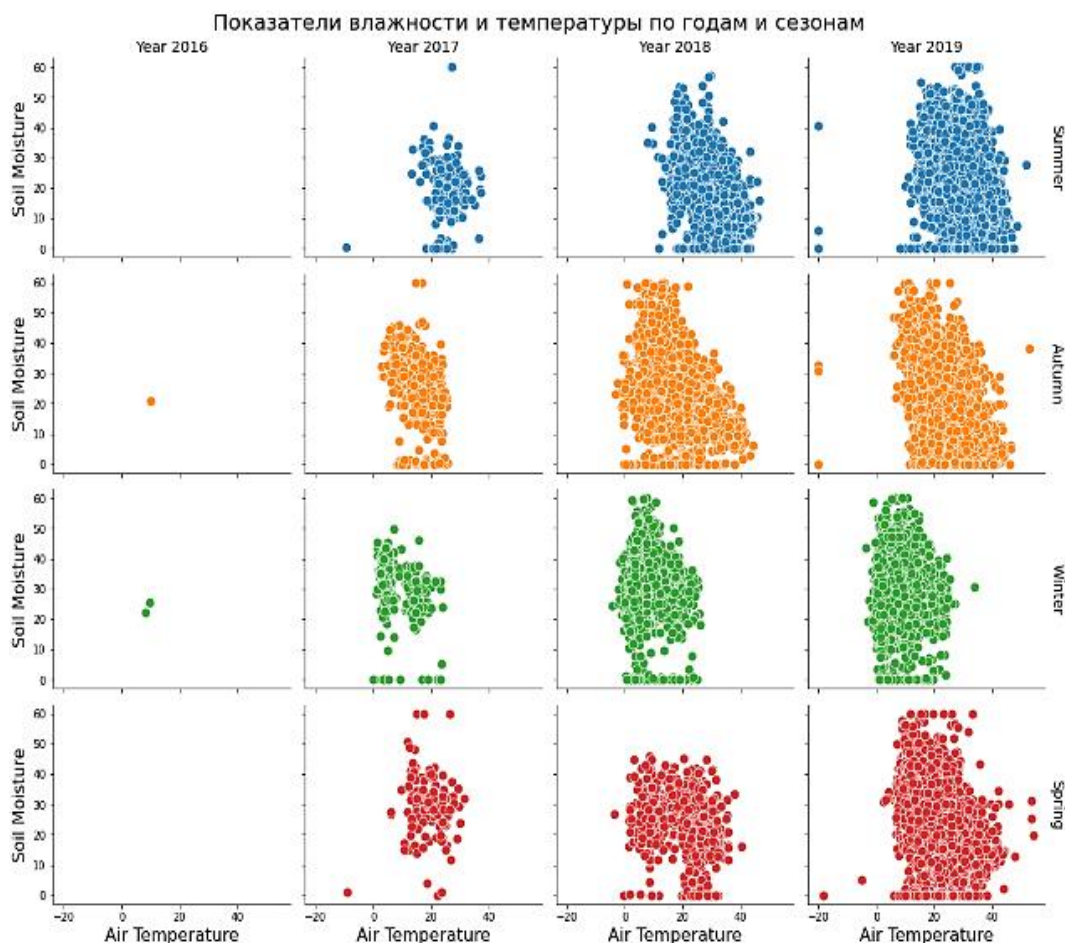


Рисунок 2. Сезонные тенденции изменения показателей почвы

Для кластеризации по схожим условиям влажности и температуры воздуха использовался метод K-Means (рис. 3).

В данной работе были использованы методы прогнозирования временных рядов по параметрам влажности почвы, освещенности и температуры воздуха:

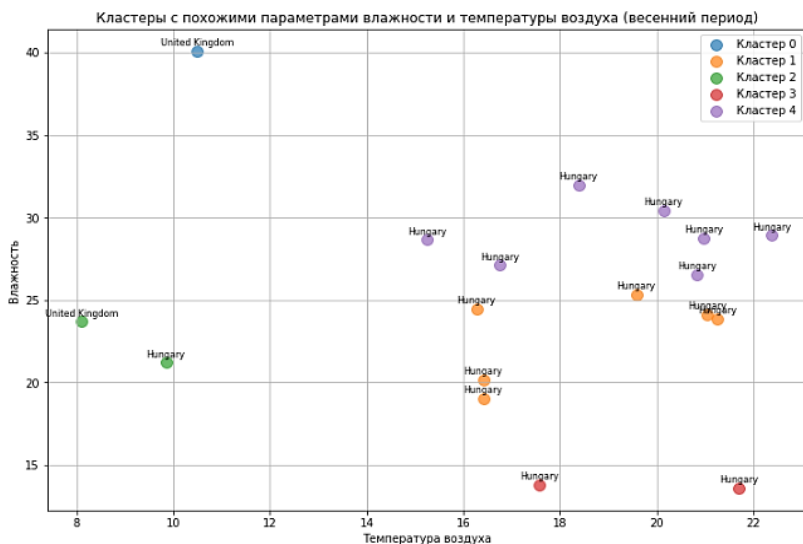
- ARIMA (авторегрессионная интегрированная скользящая средняя). Этот метод хорошо подходит для прогнозирования временных рядов, если есть некоторая стационарность и сезонность в данных;

- LSTM (долгая краткосрочная память). Это рекуррентные нейронные сети, которые могут хорошо работать с последовательными данными, такими как временные ряды.

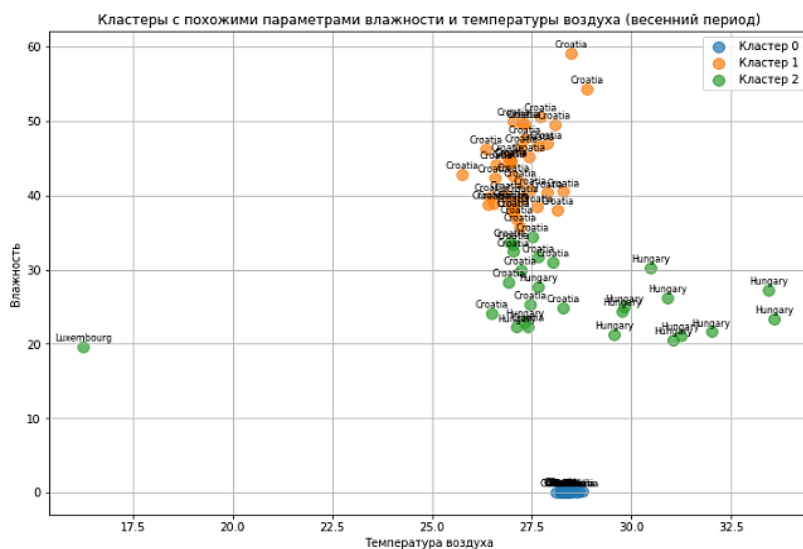
Для определения точности модели использовались методы:

- Mean Squared Error (MSE) – среднеквадратичная ошибка, которая измеряет среднее квадратичное отклонение прогнозных значений от фактических (чем меньше значение MSE, тем лучше точность модели);

- Mean Absolute Error (MAE) – средняя абсолютная ошибка, которая измеряет среднее абсолютное отклонение прогнозных значений от фактических. Эта метрика также представляет собой меру точности прогнозов.



а



б

Рисунок 3. Кластеризация по параметрам влажности и температуры воздуха в весенние (а) и летние (б) месяцы

Результаты и их обсуждения. В рамках исследования было рассмотрено применение интеллектуального анализа данных для оптимального управления земельными ресурсами.

Визуализация предоставила интуитивно понятное и информативное представление данных, упрощая выявление закономерностей и тенденций. В ходе анализа диаграмм и графиков данных было выявлено, что изменение влажности почвы в течение года имеет ярко выраженные сезонные колебания. Влажность почвы увеличивается во время осенних и весенних периодов, когда количество осадков обычно выше, а затем постепенно снижается в течение летнего периода. Эти сезонные колебания влажности почвы оказывают существенное влияние и на рост и развитие растений, и на земледелие в целом.

Результаты точности моделей ARIMA и LSTM представлены в табл. 3.

Таблица 3. Показатели меры точности моделей прогнозирования влажности и температуры

Методы прогнозирования	ARIMA	LSTM
Методы оценки точности модели		
MSE	70,09	65,23
MAE	66, 65	73, 7

Таким образом, модель прогнозирования LSTM дает более точные результаты.

Из рис. 2 видно, что количество устанавливаемых датчиков увеличивалось с каждым годом. Поэтому для кластерного анализа были взяты данные за последние 2 года. При кластеризации был применен метод K-Means, который является одним из самых популярных методов кластеризации. На рис. 3 отчетливо видны районы Хорватии и Венгрии, которые имеют схожие параметры влажности и температуры воздуха.

Заключение. В данной статье было проведено исследование данных о климатических параметрах почвы с использованием методов интеллектуального анализа данных. Сельское хозяйство – одна из областей, активно применяющих технологии интернета вещей (IoT) и технологии больших данных. Применение методов интеллектуального анализа данных позволяет принимать информированные решения, оптимизировать производственные процессы и минимизировать потери в аграрном секторе.

Для анализа данных использовался язык программирования Python, который обладает обширным набором библиотек и модулей для анализа данных. Python позволяет гибко подходить к решению сложных задач и создавать настраиваемые решения, наилучшим образом соответствующие требованиям проекта.

В ходе исследования были проведены пространственный и сезонный анализы данных. Пространственный анализ показал, что большая часть датчиков установлена в европейских странах. Надо отметить, что инструменты анализа геоданных в языке Python имеют низкую производительность, что вызвало трудности при геокодировании данных широты и долготы.

Применение методов визуализации позволило выявить сезонные тенденции изменения показателей почвы, что является важным аспектом для принятия решений в сельском хозяйстве.

Исследование данных позволило получить ценные результаты, которые могут быть использованы в дальнейшем в Казахстане. С этой целью необходимо развивать центры по сбору данных, подобные GrowPlace Observatory, и в нашей стране.

Благодарности. Выражаем благодарность профессорам университета Данди, в частности профессору Владимиру Яничу, декану Школы естественных наук и инженерии Хосе Фиадейро, директору по исследованиям последипломного образования Думитру Труку, оказавшим большое содействие в проведении исследований.

References

1. Islam S. IOT-Smart Agriculture: Comparative Study on farming applications and disease prediction of Apple crop using machine learning // Iraqi Journal of Science. – 2022. – V. 63. – № 12. – P. 5520-5533. – <https://doi.org/10.24996/ijcs.2022.63.12.37>.
2. Corbari C., Paciolla N., Ben Charfi I., Skokovic D., Sobrino J. A., Woods M. Citizen science supporting agricultural monitoring with hundreds of low-cost sensors in comparison to remote sensing data // European Journal of Remote Sensing – 2022. – V. 55. – № 1. – P. 388-408. – <https://doi.org/10.1080/22797254.2022.2084643>.
3. Priyanka P. Chandak, A. J. Agrawal. Smart Farming System Using Data Mining // International Journal of Applied Engineering Research - 2017. – V. 12. – № 11. – P. 2788-2791.

4. R. Manavalan. Towards an intelligent approaches for cotton diseases detection: A review // *Computers and Electronics in Agriculture*. – 2022. – V. 200. – P. 330-346. – <https://doi.org/10.1016/j.compag.2022.107255>.
 5. Rahul Sharma. Challenges of Data mining techniques into Agriculture // *International Journal of Advances in Engineering and Management (IJAEM)*. – 2022. – V. 4. – № 4. – P. 413-421. – DOI: 10.35629/5252-0404413421.
 6. Kingsley Ukhurebor, Charles Adetunji, Olugbemi Olaniyan, Wilson Nwankwo. Precision agriculture: Weather forecasting for future farming. In *AI, Edge and IoT-Based Smart Agriculture*; Academic Press: Cambridge, MA, USA, 2022; Chapter 6. – Pp. 101–121.
 7. K Aishwarya, MA Jabbar. Data Mining Analysis for Precision Agriculture: A Comprehensive Survey // *The Electrochemical Society. ECS Transactions*. – 2022. – V. 107. – № 1. – DOI 10.1149/10701.17769ecst.
 8. James Todd, Edwis Dufrene, Herman Waguespack, Collins Kimbeng, Michael Pontif, Debbie Boykin. Data mining sugarcane breeding yield data for ratoon yield prediction // *Euphytica* – 2021. – V. 217. – № 54. – P. 47-54. - <https://link.springer.com/article/10.1007/s10681-021-02786-z>.
 9. <https://discovery.dundee.ac.uk/en/datasets/grow-soil-moisture-data>
 10. Anthony Suarez, Barjinder Singh, Firas Almkhtar, Rajnish Kler, Sonali Vyas, Karthikeyan Kaliyaperumal. Identifying Smart Strategies for Effective Agriculture Solution Using Data Mining Techniques. *Artificial Intelligence in Food Quality Improvement 2021*. – 2022. – <https://doi.org/10.1155/2022/6600049>.
 11. Akshatha, Shailesh Shetty S, Anet P James, Athira M Saseendran, Chaitra M Poojary. Crop Analysis and Profit Prediction using Data Mining Techniques (Id:39) // *International Journal of Engineering Research & Technology (IJERT)*. – 2019. – V. 7. – DOI:10.17577/IJERTCONV7IS08024
 12. Ravesa Akhter, Shabir Ahmad Sofi. Precision agriculture using IoT data analytics and machine learning. // *Journal of King Saud University - Computer and Information Sciences*. – 2022. – V. 34, Issue 8. – P. 5602-5618. – <https://doi.org/10.1016/j.jksuci.2021.05.013>.
 13. Blekanov, A. Molin, D. Zhang, E. Mitrofanov, O. Mitrofanova, Yin Li. Monitoring of grain crops nitrogen status from uav multispectral images coupled with deep learning approaches // *Computers and Electronics in Agriculture*. – 2023, № 212. – <https://doi.org/10.1016/j.compag.2023.108047>.
 14. Ivana V., Dorijan R., Mladen J., Antonela M., Kulundžić, Manda A. Prediction of sugar beet yield and quality parameters with varying nitrogen fertilization using ensemble decision trees and artificial neural networks panel // *Computers and Electronics in Agriculture*. – 2023. – № 212. – <https://doi.org/10.1016/j.compag.2023.108076>.
 15. Mohammad Taghi Sattari, Anca M., Avram, Halit Apaydin, Oliviu Matei. Soil Temperature Estimation with Meteorological Parameters by Using Tree-Based Hybrid Data Mining Models // *In Mathematics* – 2020. - <https://doi.org/10.3390/math8091407>.
 16. Orazbayev B.B., Ospanov Ye.A., Orazbayeva K.N., Makhatova V.E., Urazgaliyeva M.K. Shagayeva A.B. (2019) Development of mathematical models of R-1 reactor hydrotreatment unit using available information of various types. *Journal of Physics: Conference Series*. 1399(45) 156053. <https://doi.org/10.1088/1742-6596/1399/4/044024>.
 17. Zhiwei Rao, Jie Yuan. Data mining and statistics issues of precision and intelligent agriculture based on big data analysis. // *Acta Agriculturae Scandinavica, Section B – Soil & Plant Science*. - 2021. – V. 71, Issue 9. - <https://doi.org/10.1080/09064710.2021.1954684>
 18. G. Buvaanyaa, Dr. S. Radhimeenakshi. The impact of data mining in machine learning algorithm for crop recommendation to yield – A review. // *Journal of Emerging Technologies and Innovative Research*. – 2022. – V. 9, Issue 7. – P. 578-587.
 19. Cai Y. Integrating satellite and climate data to predict wheat yield in Australia using machine learning approaches // *Agricultural and Forest Meteorology*. – 2019. – <https://doi.org/10.1016/j.agrformet.2019.03.010>.
 20. Chlingaryan A. Machine learning approaches for crop yield prediction and nitrogen status estimation in precision agriculture: a review // *Comput. Electron. Agric.* – 2019. – <https://doi.org/10.1016/j.compag.2018.05.012>.
 21. Aasim M. Machine learning (ML) algorithms and artificial neural network for optimizing in vitro germination and growth indices of industrial hemp (*Cannabis sativa* L.) // *Ind. Crop. Prod.* – 2022. <http://dx.doi.org/10.1016/j.indcrop.2022.114801>.
 22. Brédy J. Water table depth forecasting in cranberry fields using two decision-tree-modeling approaches // *Agric. Water Manag.* – 2020. <https://doi.org/10.1016/j.agwat.2020.106090>.
 23. Khalifani S. Prediction of sunflower grain yield under normal and salinity stress by RBF-K, MLP and CNN models // *Ind. Crops Prod.* – 2022. <http://dx.doi.org/10.1016/j.indcrop.2022.115762>.
 24. Gómez D., Salvador P., Sanz J., Casanova J. Potato Yield Prediction Using Machine Learning Techniques and Sentinel 2 Data. *Remote Sens.* 2019;11:1745. doi: 10.3390/rs11151745.
-
-