



Возможности применения интеллектуальных платформ анализа больших данных на предприятии¹

Демичев В.В., к.э.н, доцент, доцент кафедры статистики и кибернетики,
ФГБОУ ВО Российский государственный аграрный университет – МСХА
им. К.А. Тимирязева, Москва, Россия

Невзоров А.С., ассистент кафедры статистики и кибернетики,
ФГБОУ ВО Российский государственный аграрный университет – МСХА
им. К.А. Тимирязева, Москва, Россия

Федоров Д.А., к.с.-х.н., преподаватель кафедры овощеводства
ФГБОУ ВО Российский государственный аграрный университет – МСХА
им. К.А. Тимирязева, Москва, Россия

Аннотация. Одной из ключевых задач управления на любом предприятии является обеспечение эффективного производства на основе оптимизации затрат. Целью настоящего исследования является разработка и демонстрация возможностей интеллектуальных платформ анализа больших данных по прогнозированию валового сбора томата и поиска резерва снижения затрат на производство.

В качестве объекта исследования выступает тепличный комплекс, осуществляющий выращивание томата в промышленных масштабах. Основная идея исследования заключается в том, что тщательный учет, обработка и анализ данных результатов деятельности и условий выращивания, в данном случае томата, могут существенно увеличить точность прогнозируемых значений и повысить эффективность производства.

¹ Данное исследование проведено при поддержке Фонда содействия инновациям (ФСИ) в рамках конкурса «Студенческий стартап».

Ключевые слова: большие данные, интеллектуальный анализ данных, платформа, веб-приложение, прогнозирование, оптимизация, овощеводство защищенного грунта, машинное обучение.

Possibility of using intelligent big data analysis platforms in the enterprise

Demichev V.V., candidate of Economics, Associate Professor of the Department of Statistics and Cybernetics,

Russian State Agrarian University – MTAA, Moscow, Russia

Nevzorov A.S., assistant of the Department of Statistics and Cybernetics, Russian State Agrarian University – MTAA, Moscow, Russia

Fedorov D.A., candidate of agricultural sciences, teacher of the Department of Vegetable Growing, Russian State Agrarian University – MTAA, Moscow, Russia

Annotation. One of the key management tasks at any enterprise is to ensure efficient production based on cost optimization. The purpose of this study is to develop and demonstrate the capabilities of intelligent big data analysis platforms for forecasting the gross tomato harvest and finding a reserve for reducing production costs.

The object of the study is a greenhouse complex that grows tomatoes on an industrial scale. The main idea of the study is that careful accounting, processing and analysis of data on the results of activities and growing conditions, in this case tomatoes, can significantly increase the accuracy of predicted values and improve production efficiency.

Key words: big data, data mining, platform, web application, forecasting, optimization, protected ground vegetable growing, machine learning.

Введение. Инфляционное давление постоянно вынуждает производителей сельскохозяйственной продукции искать резервы снижения производственной себестоимости, что позволяет увеличивать или сохранять доходность предприятий, а также сдерживать рост цен для потребителей. Оптимальное использование ресурсов позволяет не только снизить затраты, но и добиться

повышения урожайности и максимального потенциала растений. Установленные датчики в теплицах (предусмотрены производителями теплиц) позволяют собирать обширные данные, анализировать их и разрабатывать рекомендации в процессе выращивания растений для повышения их качественных характеристик и урожайности². Именно использование методов искусственного интеллекта и нейросетевого анализа, как в разработанной платформе, в противовес традиционной интуитивной оценке состояния роста растений, так называемой «green finger» оценки, позволят максимально проявить потенциал растений в виде роста их урожайности и повышения качества.

Цель исследования. Реализовать прототип интеллектуальной платформы анализа больших данных и прогнозирования на предприятии, осуществляющем деятельность в овощеводстве защищенного грунта с применением методов машинного обучения.

Задачи исследования:

- 1) Осуществить сбор и подготовку данных для тренировки и тестирования моделей прогнозирования;
- 2) Обучить модели для решения задач прогнозирования по временным рядам, регрессии, кластеризации и факторного анализа;
- 3) Разработать пользовательский интерфейс интеллектуальной платформы;
- 4) Предложить форму сбора данных на предприятии в формате веб-страницы.

Материалы, методы и условия проведения исследования. Объектом исследования выступил один из тепличных комплексов Ставропольского края. Дополнительные сведения о предприятии не могут быть сообщены по причине наличия NDA у авторов исследования. Базой исследования являлись большие данные теплицы по сбору томатов, данные факторов производства томатов и ряд других показателей³. Исследования проводились с помощью методов статистики

² Бартлетт, Р. Машинное обучение: практические методы анализа данных. М.: Диалектика, 2019

³ Джеффри, У. Большие данные: принципы и практики построения масштабируемых систем обработки данных. М.: Вильямс, 2016

и машинного обучения: авторегрессия, кластерный анализ, регрессионный анализ, факторный анализ. В качестве инструмента реализации платформы, ее элементов и моделей машинного обучения выбран язык программирования Python.

Результаты исследования и их обсуждение.

Разработка клиентской части интеллектуальной платформы. В результате проведенной работы создан пользовательский интерфейс (Рис. 1) на основе библиотеки Python, позволяющий осуществить выбор метода анализа с последующей загрузкой данных.

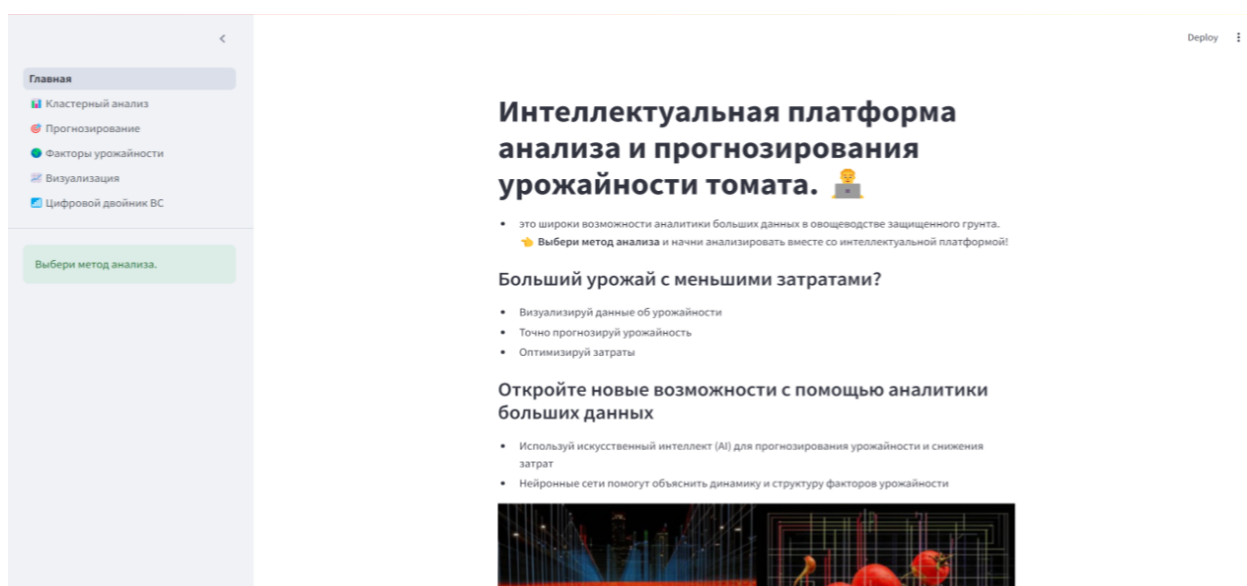


Рис. 1 – Главная страница приложения

На главной странице предоставляется возможность выбора метода анализа («Кластерный анализ», «Прогнозирование», «Факторы урожайности», «Визуализация», «Цифровой двойник ВС»).

Ввиду необходимости сбора дополнительных данных о растениях, в частности данных по фенологии авторами исследования разработана веб-форма (Рис. 2) для сбора параметров сбора, условий содержания и роста томата, поскольку эти данные по факту находятся в разнородных источниках или вовсе не собираются.

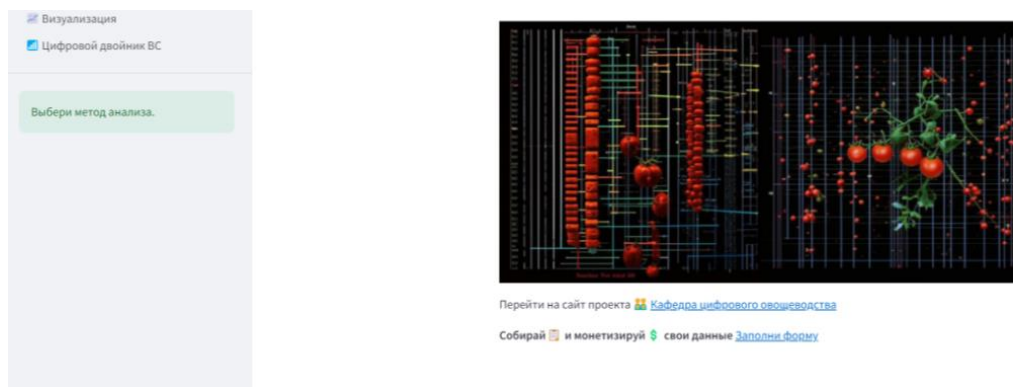


Рис. 2 – Ссылка на форму для сбора данных на главной страниц

При использовании веб-формы (Рис. 3), агрономы-овощеводы могут собирать данные из разнородных источников, что существенно повысит скорость, эффективность и качество разрабатываемых моделей машинного обучения, снизит затраты труда на скорость и обработку данных.

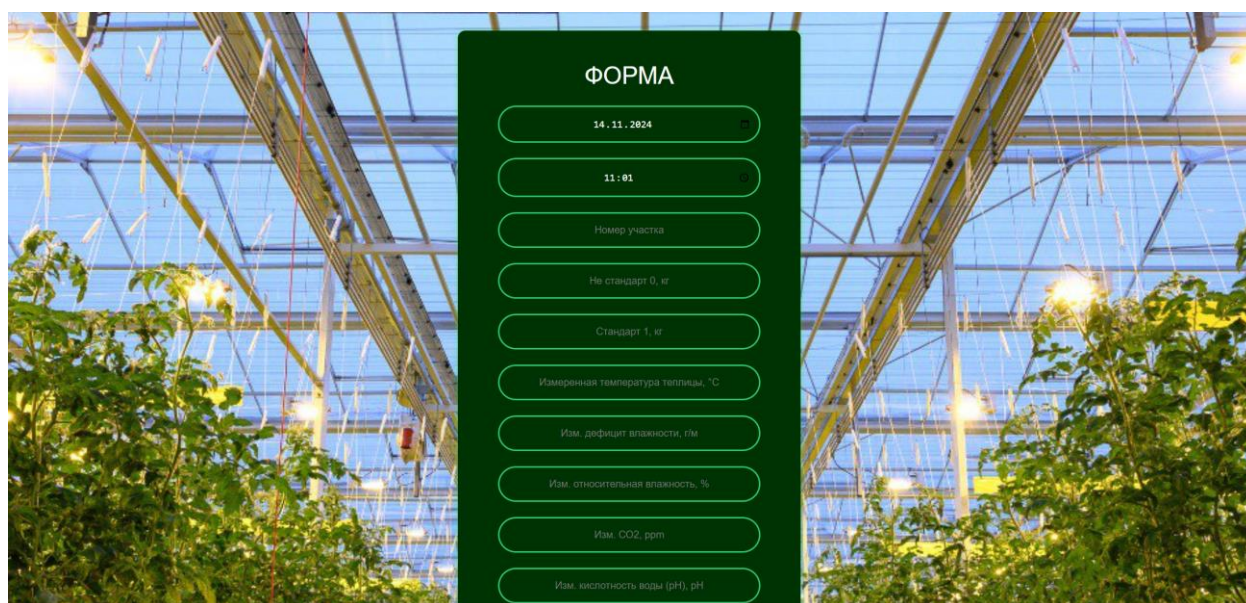


Рис. 3 – Главная страница формы для сбора данных

Если количество показателей, которое заявлено в форме, недостаточно, то предусмотрена возможность добавления показателя с указанием его наименования и единицы измерения⁴. Детальный учет, проведение измерений и дальнейший анализ данных позволят осуществлять лучшее оперативное, тактическое и даже стратегическое управление производственными процессами

⁴ Зегжда, Д. П., Семенов, Ю. А. Информационная безопасность и защита данных. М.: Горячая линия-Телеком, 2017

на предприятии за счет более «прозрачного» понимания причинно-следственных взаимоотношений между результатом, условиями и факторами производства.

Название показателя

Показатель

Название показателя

Показатель

Название показателя

Показатель

Название показателя

Показатель

Название показателя

Показатель

Рис. 4 – Дополнительная форма для сбора показателей

После заполнения показателей данные сохраняются в csv-формате и могут быть использованы для дальнейшего анализа.

На основе имеющихся или собранных данных может быть осуществлен прогноз на основе авторегрессионной модели машинного обучения⁵.

Прогнозирование

Выбрать файл



Drag and drop file here
Limit 200MB per file

Browse files

Рис. 5 – Страница «Прогнозирование»

⁵ Иванов, А. В. «Применение методов анализа больших данных в управлении производством». Вестник Московского университета. Серия 12: Экономика, № 11, 2019, стр. 56–70

В каждой вкладке предусмотрена загрузка файлов с данными и в таком же виде представлены и остальные страницы: «Кластерный анализ», «Факторы урожайности», «Визуализация», «Цифровой двойник ВС».

После загрузки данных (в виде временного ряда) может быть получен прогноз урожайности на основе модели авторегрессии (Рис. 6).

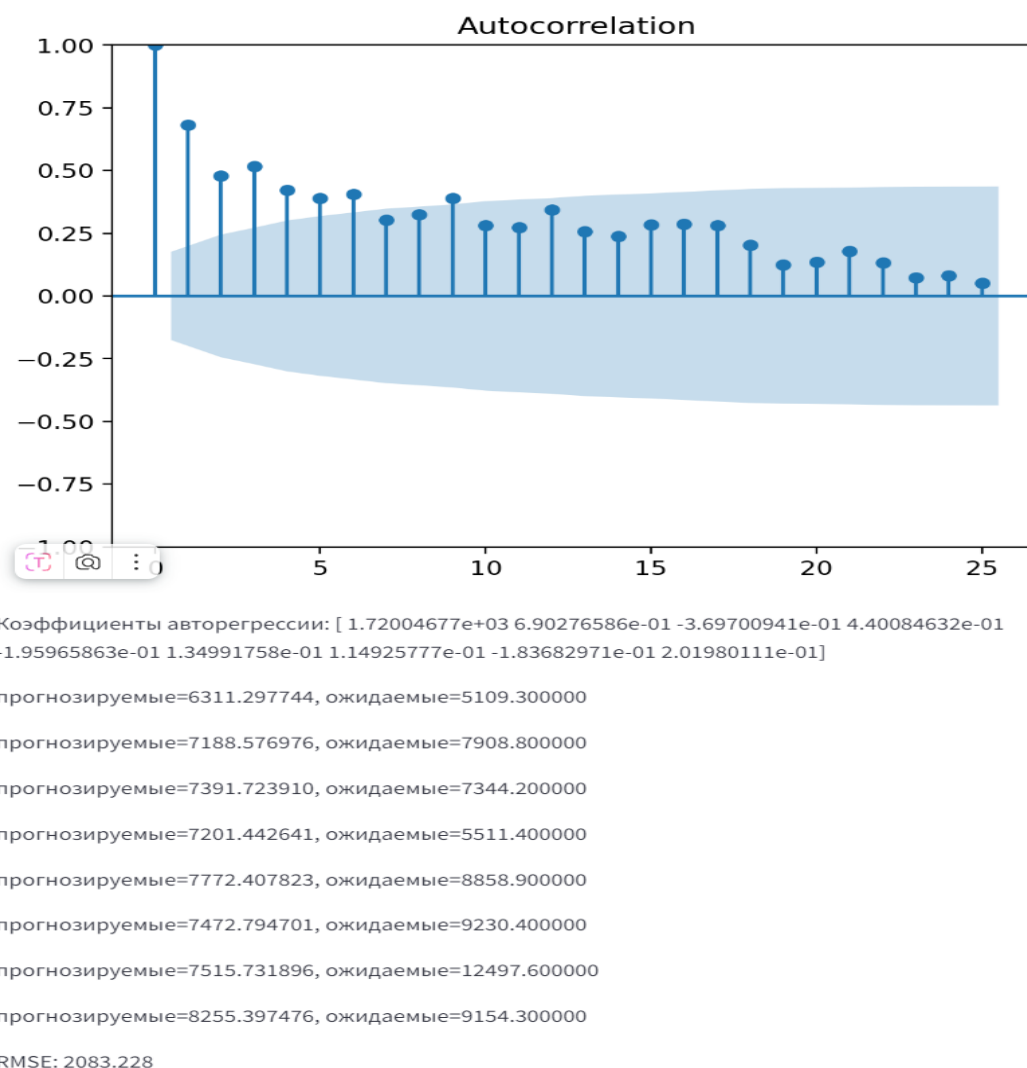


Рис.6 – Автокорреляционная функция и прогноз валового сбора томата на 8 дней вперед

Разработка искусственного интеллекта и нейронной сети для решения задач прогнозирования на основе регрессии. При реализации моделей необходимо составить список показателей, на основе которых и будет строиться дальнейший анализ и прогнозирование с применением методов регрессии.

```

25 # Может быть использовано, если файл с данными доступен:
26 dataframe = pd.read_csv(uploaded_file, delimiter=',', index_col=0, encoding='windows-1251')
27 dataframe.rename(columns={'Участок': 'Участки',
28                           'Нестандарт 0': 'Нестандарт 0, кг',
29                           'Стандарт 1': 'Стандарт 1, кг',
30                           'изм. ОВ -- Greenhouse:Среднее по Отд 51,52,53,54 -- %': 'Изм. относительная влажность, %',
31                           'изм. темп.теплицы -- Greenhouse:Среднее по Отд 51,52,53,54 -- °C': 'Измеренная температура теплицы, °C',
32                           'измер CO2 -- Greenhouse:Среднее по Отд 51,52,53,54 -- ppm': 'Изм. CO2, ppm',
33                           'изм. pH -- Greenhouse:ВОД.СИС 50 -- pH': 'Изм. кислотность воды (pH), pH',
34                           'изм. ЕС -- Greenhouse:ВОД.СИС 50 -- мС/см': 'Изм. концентрация питательного раствора, мС/см',
35                           'наружная темп. -- Greenhouse:METEО -- °C': 'Наружная температура, °C',
36                           'скорость ветра -- Greenhouse:METEО -- м/с': 'Скорость ветра, м/с',
37                           'наружный ДВ -- Greenhouse:METEО -- г/м³': 'Наружный ДВ, г/м³',
38                           'освещенность -- Greenhouse:METEО ДАТЧ.ОСВЕЩ -- Вт/м²': 'Освещенность, Вт/м²',
39                           'сумма освещенности -- Greenhouse:METEО ДАТЧ.ОСВЕЩ -- Дж/см²': 'Сумма освещенности, Дж/см²',
40                           'наружная ОВ -- Greenhouse:METEО -- %': 'Наружная ОВ, %',
41                           'изм. ДВ -- Greenhouse:Среднее по Отд 51,52,53,54 -- г/м³': 'Изм. дефицит влажности, г/м³'}, inplace=True)
42
43 st.write(dataframe['Стандарт 1, кг'])

```

Рис. 7 – Показатели валового сбора и условий выращивания томата

В списке присутствуют показатели: влажность, температура, CO₂, наружная температура и др.

На базе текущих показателей и консультации с экспертами-агрономами построена модель регрессии, которая строит прогноз урожайности⁶. На рисунке 8 представлен вариант реализации одной из моделей множественной регрессии.

```

# Построение модели регрессии
model = sm.OLS(endog=np.asarray(Y), exog=np.asarray(X_1)) # Модель регрессии
model_results = model.fit() # Результаты регрессии
print('\nSummary:\n', model_results.summary()) # Отображение основных результатов
print('\nParameters:\n', model_results.params) # Отображение параметров уравнения

# Прогноз
X_values = [] # Предполагаемые значения факторов
for i in range(len(x_names_list)):
    x = float(st.number_input(label=f'Введите значение x{i + 1}', value=0, key=i))
    X_values.append(x)
st.write('X_values', X_values)

# Таблица для получения прогноза
X_for_predict = pd.DataFrame({'const' : [1]})
for i in range(len(x_names_list)):
    X_for_predict[x_names_list[i]] = [X_values[i]]
st.write('X_for_predict', X_for_predict)

Y_predict = round(model_results.predict(X_for_predict)[0]) # Прогноз по модели
# y = a + b1 * x1 + b2 * x2 ...
n_tonn = round(Y_predict/1000)
st.write(f'При факторах X = {X_values} **валовой сбор** составит: {Y_predict}', 'кг')
image1 = Image.open('E:\ПРИЛОЖЕНИЕ\pages\Коробка_томат.jpg')
st.image([image1]*n_tonn, width=150)

```

Рис. 8 – Реализация множественной регрессии

Регрессионная зависимость предполагает, что каждому значению переменной x могут соответствовать различные значения y ⁷. В нашем случае мы

⁶ Майерс, Дж. Д. Анализ больших данных: принципы и практика обработки больших объемов информации. М.: Вильямс, 2018

⁷ Петров, О. Н. «Использование интеллектуальных платформ для оптимизации бизнес-процессов». Экономический журнал, № 18, 2018, стр. 34–45

ищем зависимости в большом количестве показателей, которые влияют на урожайность томата.

Дальнейшее развитие платформы требует тренировки и внедрения нейронных сетей, позволяющих определять нелинейные зависимости и тем самым более точно осуществлять прогнозирование.

Возможности платформы по визуализации собираемых данных.

Интеллектуальная платформа также позволяет визуализировать собираемые данные, в частности по результатам производства (сбор стандартного и нестандартного томата в данном исследовании) и условий выращивания растений (Рис. 9).

Визуализация

Выбрать файл

Drag and drop file here
Limit 200MB per file

Browse files

5_20-21_harv_factors_days.csv 39.7KB

Данные

	Дата	Участки	Нестандарт 0, кг
43	2021-05-24	01_5 02_5 03_5 04_5 05_5 06_5 07_5 08_5 09_5 10_5 11_5 12_5 13_5 14_5	1,004.7
44	2021-05-26	01_5 02_5 03_5 04_5 05_5 06_5 07_5 08_5 09_5 10_5 11_5 12_5 13_5 14_5	1,019.5
45	2021-05-28	01_5 02_5 03_5 04_5 05_5 06_5 07_5 08_5 09_5 10_5 11_5 12_5 13_5 14_5	849.8
46	2021-05-31	01_5 02_5 03_5 04_5 05_5 06_5 07_5 08_5 09_5 10_5 11_5 12_5 13_5 14_5	1,010.6
47	2021-06-02	01_5 02_5 03_5 04_5 05_5 06_5 07_5 08_5 09_5 10_5 11_5 12_5 13_5 14_5	751.2
48	2021-06-04	01_5 02_5 03_5 04_5 05_5 06_5 07_5 08_5 09_5 10_5 11_5 12_5 13_5 14_5	865.5
49	2021-06-07	01_5 02_5 03_5 04_5 05_5 06_5 07_5 08_5 09_5 10_5 11_5 12_5 13_5 14_5	837.7
50	2021-06-09	01_5 02_5 03_5 04_5 05_5 06_5 07_5 08_5 09_5 10_5 11_5 12_5 13_5 14_5	644.7
51	2021-06-11	01_5 02_5 03_5 04_5 05_5 06_5 07_5 08_5 09_5 10_5 11_5 12_5 13_5 14_5	668.3
52	2021-06-15	01_5 02_5 03_5 04_5 05_5 06_5 07_5 08_5 09_5 10_5 11_5 12_5 13_5 14_5	884.6

Рис. 9 – Исходные данные для визуализации

Данные предварительно были обработаны: убраны пустые значения, проведена нормализация, убраны резко выделяющиеся значения, также построен макет для вывода таблиц. На рис. 9 продемонстрирована возможность просмотра данных после их загрузки и перед возможностью использования для визуализации и построения моделей машинного обучения⁸.

⁸ Сидоров, Е. Г. «Роль больших данных в стратегическом управлении предприятием». Вопросы экономики, № 23, 2017, стр. 78–89

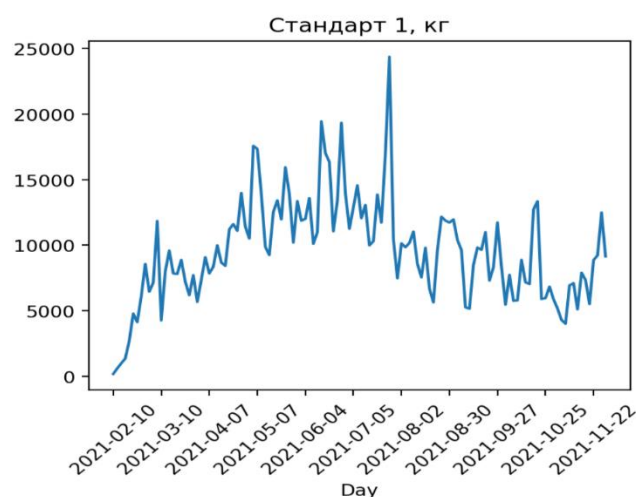


Рис. 10 – Динамика сборов томата «Стандарт 1»

Построенный временной ряд демонстрирует сильную вариативность валового сбора томата в течении всего периода сбора культуры в теплице.

Помимо таблиц и графиков, результатом работы платформы является статистические показатели, такие как корреляция, коэффициент детерминации и др.

Чем ближе значение к 1 или -1, тем выше связь между показателями

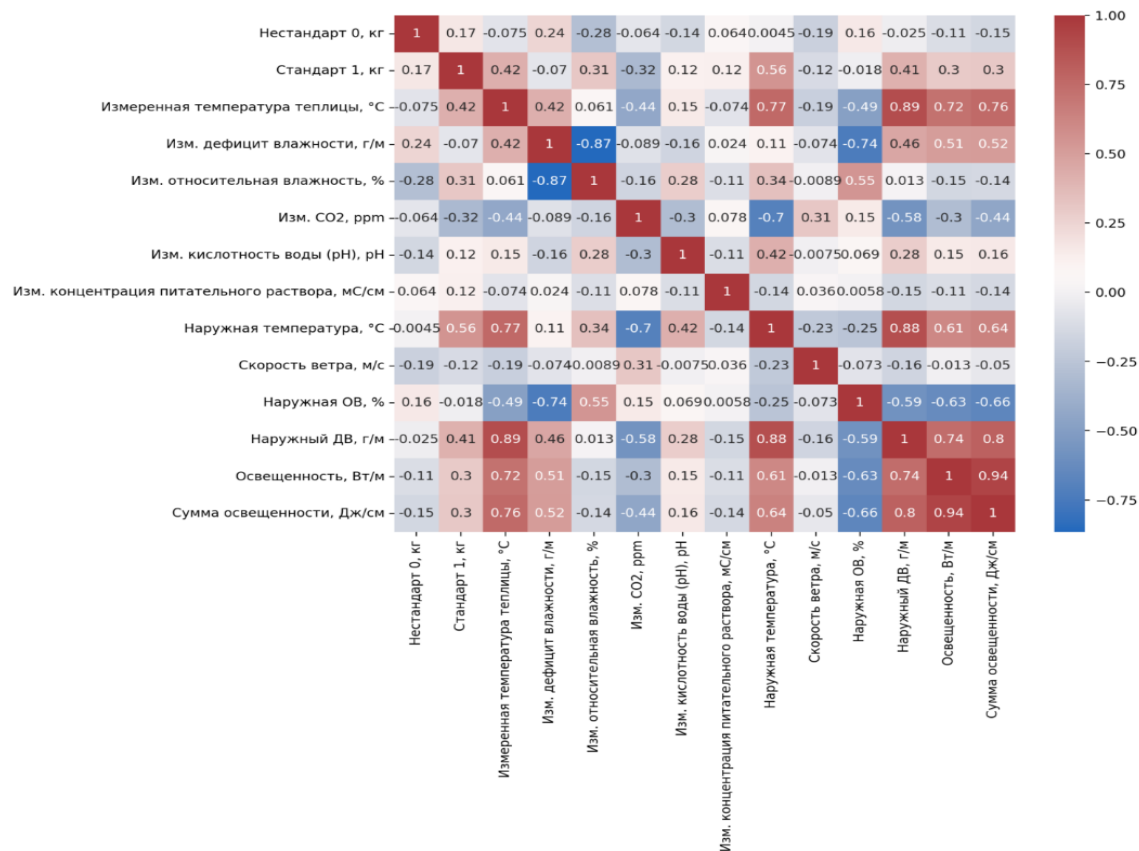


Рис. 11 – Матрица корреляций

Степень взаимосвязи показателей между собой показывает нам наиболее значимые для построения моделей факторы. Корреляция находится в диапазоне от -1 до 1, и чем ближе к -1 или 1, тем связь показателей выше.

Тестирование и отладка интеллектуальной платформы. Тестирование является неотъемлемой частью разработки любого программного продукта. По сути, вся разработка происходит параллельно с тестированием, ведь сам разработчик регулярно проверяет корректность функционирования написанного кода⁹. Тем не менее, отследить все возникающие проблемы и ошибки самостоятельно невозможно.

В задачу тестирования входило:

Функциональное тестирование: проверены все основные функции платформы, включая обработку данных, оптимизацию факторов, визуализацию и безопасность данных. Проведены тесты на предмет совместимости с различными браузерами и устройствами.

Тестирование на производительность: проведены нагрузочные тесты, чтобы оценить способность платформы работать при высокой нагрузке, например, когда множество агрономов одновременно отправляют обращения или обрабатывают данные¹⁰.

По итогам всех этапов тестирования были выявлены области, требующие доработки. На их основе код был отредактирован, исправлены найденные «баги» в функционале и недостатки в интерфейсе приложения, также была улучшена общая производительность системы¹¹.

Важно отметить, что процесс тестирования, доработки и изменения кода является неотъемлемой частью жизненного цикла у всех современных онлайн-платформ. Именно постоянная поддержка и улучшение позволяет продукту оставаться востребованным и отвечать меняющимся требованиям времени, продолжая тем самым предоставлять качественный пользовательский опыт.

⁹ Смирнов, Л.А. «Интеллектуальные системы поддержки принятия решений на основе анализа больших данных». Труды Института проблем управления РАН, № 16, 2016, стр. 123–135

¹⁰ Тан, П.-Н., Стейнбах, М., Карпентер, В., Штайнбах, М. Введение в анализ данных. СПб.: Питер, 2020

¹¹ Фролова, Н. Б. «Современные технологии анализа данных в сельском хозяйстве». Агроинформатика, № 14, 2018, стр. 67–79

Заключение. Деятельность товаропроизводителя направлена прежде всего на производство продукции, но при этом важно стремиться к повышению эффективности и снижению затрат. В этой связи использование методов анализа данных открывает значительные перспективы для достижения этих целей. На примере сельского хозяйства, в частности производства томатов в защищенном грунте, продемонстрировано, что интеллектуальные платформы могут эффективно применяться для оптимизации производственных процессов и повышения их результативности. Таким образом, анализ больших данных становится важным инструментом для улучшения производительности предприятий независимо от сферы их деятельности.

В данном исследовании наглядно продемонстрированы функции разработанного прототипа интеллектуальной платформы, которые в полной мере позволяют обеспечить производителя возможностью прогнозировать результативный показатель (в данном случае валовой сбор томатов); повысить эффективность факторов производства и снизить затраты; повысить качество производимых товаров. Данная платформа может быть адаптирована для организаций, осуществляющих свою деятельность в сфере защищенного грунта, а также других сельскохозяйственных и несельскохозяйственных предприятий, получающих, при условии внедрения платформы, конкурентные преимущества на рынке.

Библиографический список:

- 1) Бартлетт, Р. Машинное обучение: практические методы анализа данных. М.: Диалектика, – 2019.
- 2) Джеффри, У. Большие данные: принципы и практики построения масштабируемых систем обработки данных. М.: Вильямс, – 2016.
- 3) Зегжда, Д.П., Семенов, Ю.А. Информационная безопасность и защита данных. М.: Горячая линия-Телеком, – 2017.
- 4) Иванов, А.В. «Применение методов анализа больших данных в управлении производством». Вестник Московского университета. Серия 12: Экономика, № 11, – 2019, – стр. 56–70.

5) Майерс, Дж.Д. Анализ больших данных: принципы и практика обработки больших объемов информации. М.: Вильямс, – 2018.

6) Петров, О.Н. «Использование интеллектуальных платформ для оптимизации бизнес-процессов». Экономический журнал, – № 18, – 2018, – стр. 34–45.

7) Сидоров, Е.Г. «Роль больших данных в стратегическом управлении предприятием». Вопросы экономики, – № 23, – 2017, – стр. 78–89.

8) Смирнов, Л.А. «Интеллектуальные системы поддержки принятия решений на основе анализа больших данных». Труды Института проблем управления РАН, – № 16, – 2016, – стр. 123–135.

9) Тан, П.-Н., Стейнбах, М., Карпентер, В., Штайнбах, М. Введение в анализ данных. СПб.: Питер, 2020.

10) Фролова, Н.Б. «Современные технологии анализа данных в сельском хозяйстве». Агроинформатика, – № 14, – 2018, – стр. 67–79.

References:

1) Bartlett, R. Machine Learning: Practical Methods for Data Analysis. Moscow: Dialectics, – 2019.

2) Jeffrey, W. Big Data: Principles and Practices for Building Scalable Data Processing Systems. Moscow: Williams, – 2016.

3) Zegzhda, D.P., Semenov, Yu.A. Information Security and Data Protection. Moscow: Goryachaya Liniya-Telecom, – 2017.

4) Ivanov, A.V. «Application of Big Data Analysis Methods in Production Management». Bulletin of Moscow University. Series 12: Economics, – № 11, – 2019, pp. 56–70.

5) Myers, J. D. Big Data Analysis: Principles and Practice of Processing Large Volumes of Information. M.: Williams, – 2018.

6) Petrov, O.N. «Using Intelligent Platforms to Optimize Business Processes». Economic Journal, – № 18, – 2018, – pp. 34–45.

- 7) Sidorov, E.G. «The Role of Big Data in Strategic Enterprise Management». Voprosy Ekonomiki, – № 23, – 2017, – pp. 78–89.
- 8) Smirnov, L. A. «Intelligent Decision Support Systems Based on Big Data Analysis». Proceedings of the Institute of Control Sciences of the Russian Academy of Sciences, – № 16, – 2016, – pp. 123–135.
- 9) Tan, P.-N., Steinbach, M., Carpenter, W., Steinbach, M. Introduction to Data Analysis. SPb.: Piter, 2020.
- 10) Frolova, N.B. «Modern technologies of data analysis in agriculture». Aggroinformatics, – № 14, – 2018, – pp. 67–79.

Для цитирования: Демичев В.В., Возможности применения интеллектуальных платформ анализа больших данных на предприятии / Демичев В.В., Невзоров А.С., Федоров Д.А. // Российский экономический интернет-журнал. – 2024. – № 4. URL:

© Демичев В.В., Невзоров А.С., Федоров Д.А., Российский экономический интернет-журнал 2024, № 4