

УДК 004.942

DOI: 10.22213/2410-9304-2024-2-69-79

Применение методов машинного обучения для прогнозирования в сельскохозяйственном секторе экономики

Р. В. Файзуллин, кандидат экономических наук, доцент, РТУ МИРЭА, Москва, Россия

О. А. Новикова, кандидат технических наук, РТУ МИРЭА, Москва, Россия

В статье исследуются вопросы применения методов машинного обучения для создания моделей для использования их применительно к сельскому хозяйству. Изучение вопросов адаптации передовых математических методов для сельскохозяйственного сектора экономики позволяет строить действенные модели управления показателями и прогнозированием их дальнейшего развития. Основная цель данной статьи заключается в анализе применения современных математических методов (линейная регрессия, градиентный бустинг, случайный лес, нейронные сети (Deep Learning), метод опорных векторов (SVR), машины опорных отсеков (One-Class SVM), K-ближайших соседей (KNN), алгоритм кластеризации K-средних (K-Means), метод главных компонент (PCA)) в сельскохозяйственном секторе с целью оптимизации производства, повышения эффективности агробизнеса и прогнозирования параметров. На основе анализа ключевых методов машинного обучения были получены выводы касательно их применимости в сельском хозяйстве. Отдельный акцент в исследовании сделан на разработке и улучшении модели прогнозирования урожайности в сельском хозяйстве с использованием методов машинного обучения. Для практической апробации выбрана базовая модель, основанная на Random Forest Regressor, позволяющая предсказывать урожайность, учитывая различные факторы, такие как погодные условия, уровень удобрений, тип почвы и другие значимые аспекты. В рамках исследования представлен алгоритм для проведения расчетов с использованием Python, позволяющий проводить обучение модели, визуализацию влияния параметров и статистический анализ. В процессе апробации была проведена оптимизация модели с использованием GridSearchCV, направленная на подбор оптимальных гиперпараметров. В рамках анализа выявлено, что возможно улучшить модель, однако с низкой статистической значимостью. Полученные результаты исследования показывают, что имеется возможность эффективно применять машинное обучение для повышения производительности прогнозирования ключевых параметров в сельском хозяйстве.

Ключевые слова: машинное обучение, умное сельское хозяйство, статистический анализ, искусственный интеллект, управление производством, эффективность производства.

Введение

Потребность в построении инновационной экономики обусловила ускоренное стимулирование производства нововведений и их внедрение в хозяйственную деятельность. Именно на основе постоянного увеличения роли науки и ее вклада в экономику происходит становление инновационной экономики [1]. Сельское хозяйство играет ключевую роль в экономическом развитии, обеспечивая продовольственную безопасность. В данном контексте актуализируется и потребность в инновационном обновлении методов управления агрокомплексами. Такая практика ставит вопрос о внедрении передовых разработок в практику хозяйствования агробизнеса [2].

Эффективное управление сельскохозяйственными ресурсами становится все более значимым аспектом, особенно в условиях изменяющихся климатических условий и роста спроса на продукты. Для этого предложено использовать экономико-математические ме-

тоды, направленные на оптимизацию производства, прогнозирование урожайности и повышение эффективности. К таким методам относятся и методы машинного обучения [3].

Непосредственно машинное обучение представляет собой набор методов, позволяющих компьютерам обучаться на основе данных и делать прогнозы или принимать решения без явного программирования. В различных сферах, таких как медицина, финансы и технологии, методы машинного обучения успешно применяются для анализа больших объемов информационных данных для принятия решений [4]. Например, в сельском хозяйстве методы машинного обучения могут быть использованы для оптимизации процессов посева, управления ресурсами и преодоления климатических вызовов [5].

Цель данного исследования – анализ применения методов машинного обучения в сельскохозяйственном секторе с целью оптимизации производства, повышения эффективности и

прогнозирования ключевых параметров. Для достижения поставленной цели были решены следующие задачи исследования: рассмотрено применение ключевых методов машинного обучения в сельском хозяйстве; проанализировано использование машинного обучения для прогнозирования; оценены прогностические возможности методов машинного обучения.

Теоретический анализ: потенциал применения машинного обучения в сельском хозяйстве

Функционирование всех отраслей хозяйства связано с интеллектуальными преобразованиями в глобальном макроэкономическом пространстве. Данные тенденции свойственны и агробизнесу, которому необходимо подстраиваться под новые реалии, обеспечивая эффективность сельскохозяйственной деятельности за счет использования нематериальных элементов и повышения инновационной активности [6].

Для этого следует разработать методические подходы к качественному повышению эффективности разных структурных элементов через выявление слабых мест. В сложившихся условиях недопущение снижения эффективности сельскохозяйственных производств обуславливает внедрение передовых математических методов в структуры управления [7; 8].

С позиции использования математического аппарата целесообразно определять факторы, препятствующие максимизации эффективности и производительности у предприятий сельскохозяйственного сектора в экономике. В случае использования эффективной системы мониторинга появляется возможность адаптировать разрозненные информационные показатели для составления необходимой базы данных.

На основе оценки эффективности использования ресурсов появляется возможность определять «точки роста», воздействие на которые окажет максимальный положительный эффект [9; 10].

На практике реализуются множественные экономические методы, ориентированные на анализ эффективности. Во многом именно методическое обоснование позволяет определять потенциальные направления для повышения эффективности управления на предприятиях сельского хозяйства. Нестабильные экономические условия ставят перед менеджментом

задачи по необходимости оптимизации производства, для которой можно использовать математические методы. Во многом достижение стратегических целей и задач обуславливает учет факторов, которые могут оказывать влияние на деятельность субъектов [11–13].

Для управления данными процессами рационализуется на основе применения современных компьютерных технологий в инновационной деятельности. Использование таких возможностей особенно актуализируется в связи с ростом вычислительных компьютерных мощностей, автоматизацией производства, усложнением моделируемых компьютерных систем, широкомасштабной разработкой систем искусственного интеллекта и появлением универсальной роботизированной техники [14; 15].

При этом машинное обучение и искусственный интеллект предназначены не только для технических областей, с распространением устройств интернета вещей (IoT) и программного обеспечения для искусственного интеллекта данные технологии все чаще проникают в различные сферы, включая сельское хозяйство. Такие методы позволяют создавать точное сельское хозяйство (precision agriculture) [16].

В сельскохозяйственном производстве основной целью использования машинного обучения является оптимизация урожайности фруктов, овощей, а также производительности скота и т. д., где критически важно учитывать сложные факторы, влияющие на процессы и экономический результат.

Искусственный интеллект все чаще становится поддержкой сельскохозяйственных производителей в отслеживании изменений погодных условий, колебаний температуры и состава почвы. Восприятие его роли изменилось: от сокращения трудозатрат и увеличения производительности до решения более сложных задач и преодоления трудностей.

При этом повышение конкуренции приводит к потребности поиска новых решений. Так, приложения искусственного интеллекта в сельском хозяйстве предоставляют фермерам оперативные, масштабные и детальные данные об урожае [17]. В табл. 1 представлены примеры приложений искусственного интеллекта и машинного обучения в сельском хозяйстве.

Таблица 1. Приложения искусственного интеллекта и машинного обучения в сельском хозяйстве

Table 1. Applications of artificial intelligence and machine learning in agriculture

Область применения	Применение технологии	Примеры результатов
Определение качества почвы	Компьютерное зрение с использованием машинного обучения для оценки стадии роста растений и оценки почвы	Ранее трудоемкая оценка стадии роста растений, таких как пшеница, теперь может быть упрощена благодаря компьютерному зрению. Исследователи собирали изображения пшеницы на разных стадиях роста и при разных условиях освещения, что позволило создать модель искусственного интеллекта, превзошедшую человеческое наблюдение
Определение состояния посевов	Алгоритмы обучаются на изображениях, полученных с портативного микроскопа, для точной оценки качества земли, содержания песка и уровня органического вещества	Инновационные исследования показывают, что модель компьютерного зрения может с точностью оценивать характеристики почвы, не уступая точности лабораторным анализам
Прогнозирование болезней растений	Использование глубокого обучения и распознавания изображений для классификации и сегментации растений	Применение технологии распознавания изображений на основе глубокого обучения для подготовки нейронных сетей позволяет классифицировать и сегментировать изображения болезней растений
Роботы-тракторы	Автоматизированные тракторы с оборудованием и камерами, оснащенные искусственным интеллектом, для автономной обработки полей	Внедрение полностью автономных тракторов позволяет оптимизировать удобрения и уменьшить их потребление, повышая эффективность сельскохозяйственных работ
Роботы-беспилотники	Беспилотные летательные аппараты с искусственным интеллектом для распыления пестицидов и удобрений	Использование беспилотных опрыскивателей, оснащенных искусственным интеллектом, для распределения удобрений и пестицидов с высокой точностью. Распознавание целевых участков в реальном времени позволяет снизить риск загрязнения и повысить эффективность обработки полей
Управление сельскохозяйственной техникой	Системы автопилота на основе машинного обучения для оптимизации маршрутов и эффективного использования техники	Внедрение систем автопилота, обученных машинным обучением, для оптимального управления сельскохозяйственными машинами, уменьшая затраты топлива и повышая производительность
Оптимизация расходов	Алгоритмы машинного обучения для анализа данных о расходах и оптимизации бюджета в сельском хозяйстве.	Использование алгоритмов машинного обучения для автоматизированного анализа расходов на ресурсы, удобрения и топливо с целью оптимизации бюджета и повышения экономической эффективности
Прогнозирование цен на продукцию	Модели машинного обучения для прогнозирования цен на сельскохозяйственную продукцию с учетом внешних факторов	Разработка моделей машинного обучения, способных анализировать рыночные и климатические данные для прогнозирования цен на сельскохозяйственную продукцию

Перечисленные аспекты были собраны на основе анализа сети Интернет. Следует отметить, что для этих целей в сельском хозяйстве происходит внедрение корпоративного программного обеспечения, что предоставляет возможность визуально мониторить состояние посевов удаленно из разных мест. Системы предупреждений в реальном времени, созданные алгоритмами искусственного интеллекта, могут предупредить о возможных проблемах, способствуя более эффективному прогнозированию с оперативным контролем за результатами. Очевидно, что искусственный интеллект способствует решению множества проблем – от увеличения производительности до улучшения управления ресурсами и повышения устойчивости ведения сельского хозяйства.

Непосредственно алгоритмы машинного обучения обучаются анализу таких проблем, как выявление вредителей и болезней сельскохозяйственных культур путем анализа больших наборов

данных с изображениями здоровых и больных растений. Такое обучение позволяет алгоритмам распознавать закономерности и выявлять проблемы на ранних стадиях. Однако для эффективного машинного обучения необходим доступ к высококачественным данным, таким как исторические данные о погоде, состав почвы, данные об урожайности и другие аспекты. Собранные данные используются для обучения алгоритмов, что позволяет им делать точные прогнозы и рекомендации [18, 19]. Машинное обучение предсказывает влияние разных факторов на урожай, анализируя исторические данные и используя их для прогнозирования будущих погодных сценариев. Данные прогнозы помогают принимать обоснованные решения относительно стратегий управления агробизнесом. Можно сделать вывод, что машинное обучение применяется для точного сельскохозяйственного производства через анализ данных. Полученная в реальном времени инфор-

мация о качестве почвы, влажности, питательных веществ, инвестициях, рабочей силе и других значимых аспектах используется для оптимизации управления, повышая урожайность и экономическую эффективность.

Методы машинного обучения в прогнозировании сельского хозяйства

Машинное обучение может интегрироваться со многими цифровыми технологиями и обеспе-

чивать синергию разных методов для улучшения результатов сельского хозяйства.

На основе анализа источников из проведенного теоретического анализа были выделены ключевые методы машинного обучения, определен их тип и предоставлено краткое обоснование возможности их применения для прогнозирования в экономике, в частности в сельскохозяйственном секторе (табл. 2).

Таблица 2. Методы машинного обучения для прогнозирования

Table 2. Machine learning methods for forecasting

Метод	Тип	Описание	Применение для прогнозирования
Линейная регрессия	Обучение с учителем, регрессия	Метод, строящий линейную зависимость между входными признаками и выходной переменной	Помогает моделировать линейные зависимости между входными признаками и целевой переменной. Подходит для прогнозирования числовых значений
Градиентный бустинг	Обучение с учителем, ансамблевый метод, регрессия или классификация	Ансамблевый метод, комбинирующий прогнозы нескольких слабых моделей для улучшения общего результата. Градиентный бустинг обучает новые модели, фокусируясь на ошибках предыдущих	Эффективен для решения задач регрессии и классификации. Сочетает прогнозы нескольких слабых моделей, что улучшает общую точность прогнозирования. Помогает в выявлении сложных закономерностей в данных
Случайный лес	Обучение с учителем, ансамблевый метод, регрессия или классификация	Ансамблевый метод, строящий несколько решающих деревьев для улучшения качества предсказания. Каждое дерево голосует за результат, и затем выбирается наиболее часто встречающийся	Обеспечивает стабильные прогнозы, основанные на множестве решающих деревьев. Эффективен для работы с большими объемами данных и переменных
Нейронные сети (Deep Learning)	Обучение с учителем, глубокое обучение	Метод, моделирующий сложные нелинейные зависимости между входными и выходными данными с использованием искусственных нейронных сетей. Нейронные сети могут иметь несколько слоев и скрытых узлов	Мощный метод для моделирования сложных нелинейных зависимостей. Подходит для задач регрессии и классификации. Может автоматически извлекать признаки из данных. Используется в случаях, когда другие методы могут оказаться недостаточно гибкими
Метод опорных векторов (SVR)	Обучение с учителем, метод опорных векторов, регрессия	Метод, строящий гиперплоскость для максимального разделения целевой переменной. В случае регрессии, SVR ищет гиперплоскость, которая наилучшим образом подходит к точкам данных	Строит гиперплоскость для наилучшего приближения к точкам данных. Помогает в регрессии, особенно в случаях, когда данные нелинейны или имеют сложную структуру
Машины опорных отсеков (One-Class SVM)	Без учителя	Метод для обнаружения выбросов, строит границу, включающую большинство объектов	Применяются для обнаружения выбросов. Строят границу, включающую большинство объектов, что позволяет выявить аномалии
K-ближайших соседей (KNN)	Обучение с учителем или без учителя, классификация или регрессия	Метод, присваивающий объекту класс на основе большинства среди его K ближайших соседей	Простой метод, основанный на большинстве среди ближайших соседей. Подходит для задач классификации и регрессии. Может использоваться для прогнозирования на основе близости объектов в пространстве признаков
Алгоритм кластеризации K-средних (K-Means)	Без учителя	Разбивает данные на K кластеров, где каждый объект принадлежит к кластеру с ближайшим центром	Помогает выявить структуру данных путем разбиения на кластеры. Может использоваться для группировки схожих объектов, что полезно для прогнозирования на основе схожести
Метод главных компонент (PCA)	Без учителя	Метод снижения размерности данных путем проецирования их на главные компоненты, сохраняя при этом максимальную дисперсию	Помогает уменьшить размерность данных, удаляя избыточность. Применяется в качестве этапа предобработки перед использованием других методов
Рекуррентные нейронные сети (RNN)	Глубокое обучение	Нейронные сети, способные работать с последовательными данными, сохраняя информацию о предыдущих шагах	Способны работать с последовательными данными и сохранять информацию о предыдущих шагах. Подходят для задач, где важна последовательность данных, например в прогнозировании временных рядов

Для прогнозирования в сельском хозяйстве, где наблюдается наличие целого множества переменных и сложных зависимостей, можно обратиться к градиентному бустингу (например, XGBoost, LightGBM, CatBoost). Данный метод ансамблевого обучения эффективен для решения задач регрессии и классификации, обладает высокой точностью и хорошо работает с разнообразными данными. Градиентный бустинг обучает несколько слабых моделей, фокусируясь на ошибках предыдущих, что позволяет ему извлекать сложные закономерности из данных в сельском хозяйстве:

$$GBM(\theta) = \sum_{i=1}^N h_i(x; \theta), \quad (1)$$

где N – количество моделей в ансамбле; $h_i(x; \theta)$ – слабая модель, обученная на ошибках предыдущих моделей.

Далее предлагается составить алгоритм реализации данного метода машинного обучения на примере субъектов сельскохозяйственного сектора экономики с использованием программной среды Python. Расчеты могут быть реализованы в случае наличия базы данных по конкретным субъектам, что ставит первичную задачу по сбору данных. Непосредственно компьютерный анализ можно проводить в следующие этапы:

1. Установка библиотек Python. Допустимо для работы с данными и создания модели применять библиотеки: `pumpy` – для работы с массивами данных; `xgboost` – для регрессии и классификации. Для полного анализа устанавливаются и импортируются необходимые библиотеки, такие как `pumpy`, `xgboost`, `Random Forest Regressor`, `mean_squared_error`, `matplotlib` и `train_test_split`.

2. Сбор и подготовка данных. Сбор данных, моделирующих различные факторы, влияющие на урожайность, такие как погода, уровень удобрений и т. д. Объединение параметров в матрицу X и целевой переменной (урожайность) в вектор y . Разделение данных на обучающий и тестовый наборы для последующей оценки производительности модели. Допустимо для обучения и тестирования модели применять библиотеки: `pumpy` – для объединения данных; `train_test_split` – для разделения данных.

3. Обучение модели. Обучение модели `Random Forest Regressor` на обучающем наборе данных. Выбор `Random Forest` обусловлен тем, что ансамблевый метод способен комбинировать несколько деревьев решений для повышения точности прогнозов. Допустимо использовать `Random Forest Regressor` из библиотеки `sklearn`. Проводится обучение модели, способной комбинировать деревья решений.

Формула прогноза модели `Random Forest Regressor`:

$$\text{RandomForestRegressor}(X) = \frac{1}{N} \sum_{i=1}^N T_i(X), \quad (2)$$

где N – количество деревьев; T_i – дерево решений.

Формула обучения модели градиентного бустинга:

$$\theta^* = \arg \min_{\theta} \sum_{i=1}^n L(y_i, GBM(\theta; x_i)), \quad (3)$$

где L – функция потерь; (x_i, y_i) – обучающая выборка.

4. Визуализация влияния параметров. Создание графиков, демонстрирующих влияние каждого параметра (погода, уровень удобрений и т. д.) на урожайность. Такой шаг позволяет визуализировать, как изменения в параметрах влияют на результаты модели. Допустимо использовать `matplotlib` для построения графиков. Проводится визуализация влияния параметров на урожайность для понимания влияния факторов.

5. Визуализация важности признаков. Создание графика, отображающего важность каждого признака в модели. Важность признаков определяется влиянием на результаты модели. Такой шаг позволяет визуализировать, какие факторы влияют сильнее. Допустимо использовать `feature_importances_` из модели `Random Forest Regressor`.

6. Оценка модели. Вычисление среднеквадратичной ошибки (MSE) модели на тестовом наборе данных. Кросс-валидация используется для проверки устойчивости модели на разных наборах данных. Допустимо использовать `mean_squared_error` из библиотеки `sklearn` и проводить кросс-валидацию с использованием `cross_val_score`.

7. Оптимизация модели. Оптимизация гиперпараметров модели с использованием `GridSearchCV` позволяет подобрать такие параметры, которые максимально улучшат производительность модели. Допустимо использовать `GridSearchCV` из библиотеки `sklearn`.

8. Прогноз с оптимизированной моделью. Создание и обучение модели с использованием найденных оптимальных гиперпараметров. Например, прогнозирование урожайности с использованием оптимизированной модели. Следует использовать оптимальные гиперпараметры для создания новой модели.

9. Статистический анализ. Сравнение среднеквадратичных ошибок между исходной и оптимизированной моделями с использованием критерия Уилкоксона. Требуется использовать критерий Уилкоксона для сравнения ошибок, что позволяет определить, является ли улучшение производительности статистически значимым. Проводится

оценка статистической значимости улучшения производительности модели.

В представленном алгоритме реализации градиентного бустинга в сельском хозяйстве через Python были определены этапы, начиная от установки необходимых библиотек и сбора данных до оценки производительности модели. Акцент сделан на использовании градиентного бустинга, ансамблевого метода, способного извлекать сложные закономерности из разнообразных данных в сельском хозяйстве. Данные шаги, включая обучение модели, визуализацию влияния параметров и важности признаков, а также статистический анализ, дают представление о процессе применения метода машинного обучения для оптимизации модели и прогнозирования урожайности.

Результаты исследования

Разработанная модель урожайности в сельском хозяйстве, основанная на методах машинного

обучения, была протестирована на субъектах сельскохозяйственного сектора экономики.

В процессе сбора данных были проведены расчеты базовой модели с использованием Random Forest Regressor, что позволило предсказать урожайность, учитывая основные факторы. Модель была обучена на обучающем наборе данных и готова к использованию для прогнозирования урожайности на основе входных параметров.

Данный алгоритм представляет собой цикл разработки модели с использованием данных и визуализации результатов. Графики на рис. 1–9 демонстрируют влияние различных параметров (погода, уровень удобрений, тип почвы и др.) на урожайность субъектов сельскохозяйственного сектора экономики. Например, можно наблюдать, как изменения в уровне удобрений или типе почвы влияют на прогнозируемую урожайность.

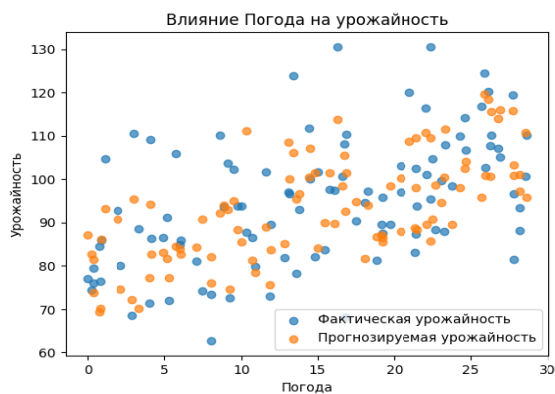


Рис. 1. Влияние фактора «Погода» на урожайность

Fig. 1. The influence of the "Weather" factor on productivity

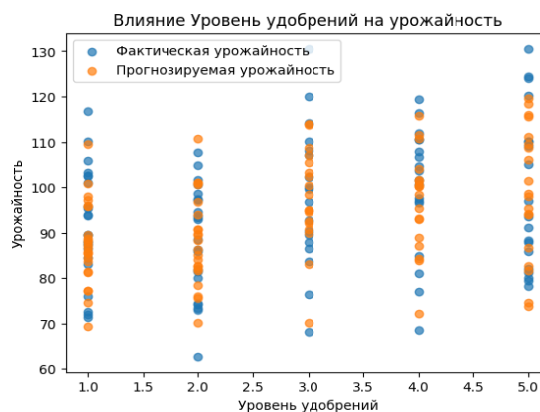


Рис. 2. Влияние фактора «Уровень удобрений» (выделено 5 базовых уровней) на урожайность

Fig. 2. The influence of the factor "Fertilizer level" (5 basic levels) on productivity

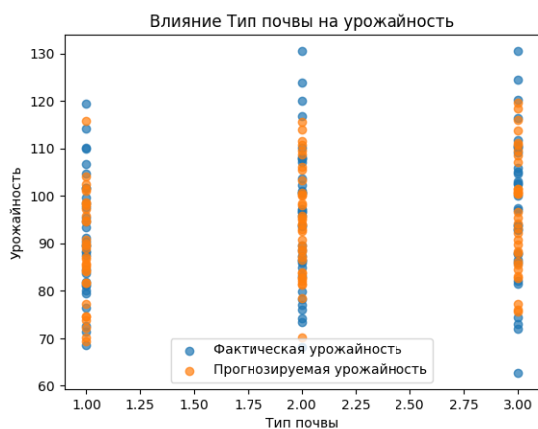


Рис. 3. Влияние фактора «Тип почвы» (3 базовых уровня) на урожайность

Fig. 3. The influence of the "Soil type" factor (3 basic levels) on productivity

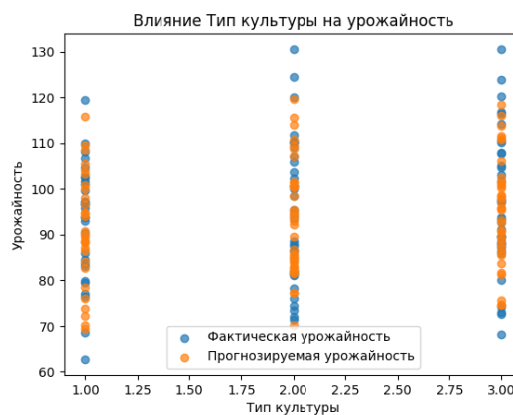


Рис. 4. Влияние фактора «Тип культуры» (выделено 3 базовых уровня) на урожайность

Fig. 4. The influence of the "Type of crop" factor (3 basic levels) on productivity

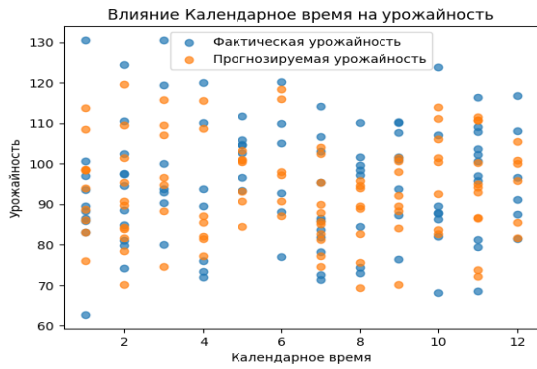


Рис. 5. Влияние фактора «Календарное время» (разбито на 12 месяцев) на урожайность

Fig. 5. The influence of the "Calendar time" factor (divided into 12 months) on productivity

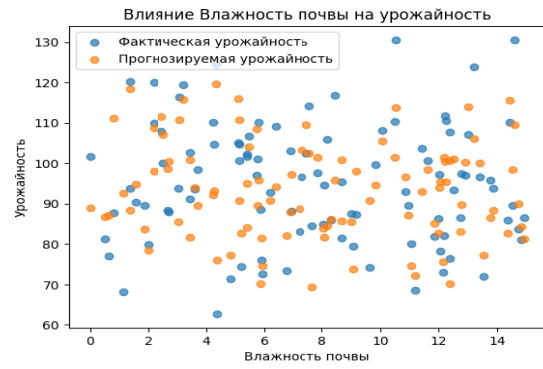


Рис. 6. Влияние фактора «Влажность» на урожайность

Fig. 6. The influence of the "Humidity" factor on productivity

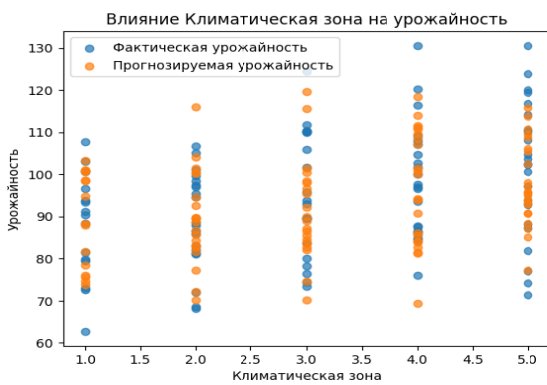


Рис. 7. Влияние фактора «Климатическая зона» (выделено 5 базовых уровней) на урожайность

Fig. 7. The influence of the "Climate zone" factor (5 basic levels) on productivity

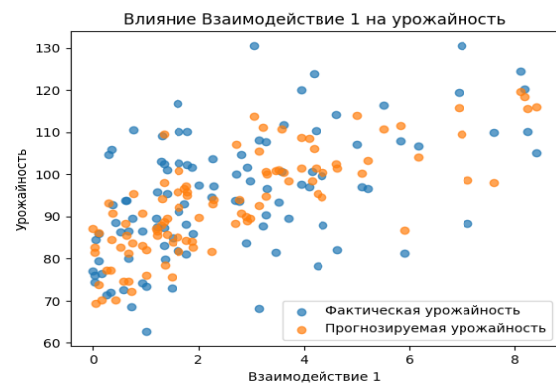


Рис. 8. Влияние фактора взаимодействия 1 (погодные условия и уровень удобрений; разбито на 10 базовых блоков) на урожайность

Fig. 8. The effect of Interaction factor 1 (weather conditions and fertilizer level; divided into 10 basic blocks) on productivity

Также предлагается оценить влияние взаимодействия между различными факторами, которое также оказывает влияние, что подчеркивает необходимость учета взаимодействий для повышения качества прогнозов. Были определены взаимодействия 1 и 2, объединяющие различный комплекс условий хозяйствования.

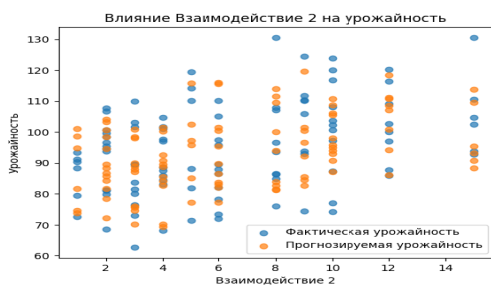


Рис. 9. Влияние фактора «Взаимодействие 2» (влажность почвы и календарное время; разбито на 20 базовых блоков) на урожайность

Fig. 9. The effect of Interaction factor 2 (soil moisture and calendar time; divided into 20 basic blocks) on productivity

Так, для каждого параметра был создан график, отображающий влияние каждого параметра на урожайность. На рис. 10 представлен анализ важности признаков в модели.

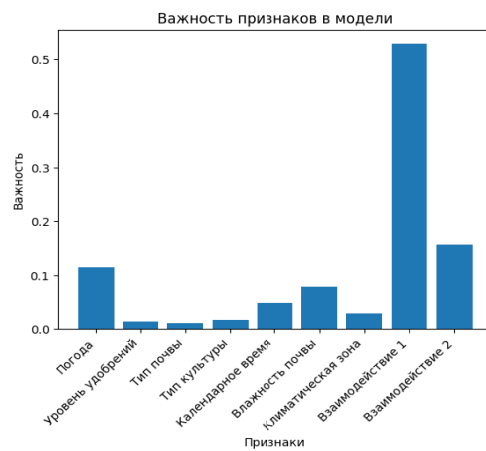


Рис. 10. Полученные результаты анализа важности факторов в модели

Fig. 10. The obtained results of the analysis of the importance of factors in the model

Анализ важности признаков

Выявлено, что «взаимодействия 1 и 2» между различными факторами оказывают существенное влияние на урожайность. В первую очередь комбинация факторов «погодные условия» и «уровень удобрений» оказали наибольшее влияние на урожайность, а следовательно, на максимизацию экономического эффекта сельского хозяйственного производства. Также высокое влияние наблюдается от комбинаторного эффекта факторов «влажность почвы» и «календарное время». Следовательно, в случае ориентации на комбинаторные параметры появляется возможность проанализировать способы реализации управленческого воздействия.

Погода, календарное время, влажность почвы также выступают значимыми факторами с позиции влияния на урожайность. В то же время без анализа комбинаторных взаимодействий их влияние остается недостаточным. При этом вызывает интерес обстоятельство, что уровень удобрений без сочетания с погодными условиями имеет низкую значимость для модели.

На рис. 11 представлен график разброса фактических и прогнозируемых значений. График разброса показывает, что прогнозы хорошо согласуются с фактическими значениями, что подтверждает адекватность модели.

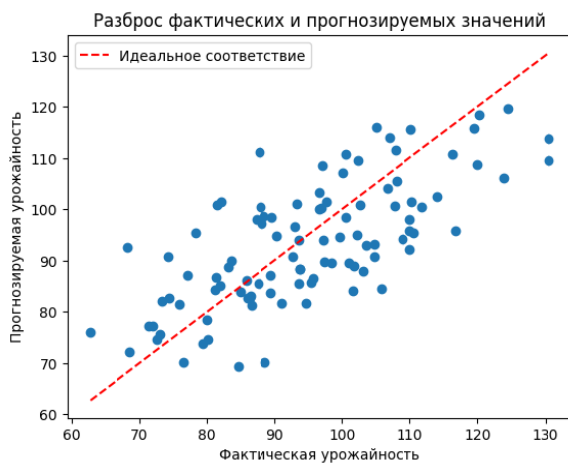


Рис. 11. Анализ разброса фактических и прогнозируемых значений

Fig. 11. Analysis of the spread of actual and predicted values

Можно сделать вывод, что модель успешно оценивает влияние различных факторов на урожайность. Важные факторы включают погодные условия, уровень удобрений, тип почвы, тип культуры, влажность почвы, климатическую зону и взаимодействие между ними. Прогнозы модели согласуются с фактическими значениями,

что поддерживает ее надежность и применимость для прогнозирования урожайности.

Результаты апробации модели:

1. Среднеквадратичная ошибка (MSE):

– Исходная модель MSE на тестовом наборе данных: 113.22.

– Средняя MSE при кросс-валидации (оценка производительности на различных наборах данных): 116.65.

Данные значения указывают на относительно хорошую точность модели.

2. Лучшие гиперпараметры: {'max_depth': 20, 'n_estimators': 150}. Полученные значения параметров указывают на то, какие значения гиперпараметров (глубина деревьев и количество деревьев) были определены как оптимальные в результате поиска по сетке (GridSearchCV). Они позволяют достичь лучшей производительности модели.

3. Среднеквадратичная ошибка с оптимизированной моделью: 111.73. Полученное значение представляет собой среднеквадратичную ошибку после оптимизации модели с использованием найденных лучших гиперпараметров.

4. Статистика и p -значение критерия Уилкоксона: (0.0, 1.0). Критерий Уилкоксона используется для сравнения среднеквадратичных ошибок между исходной и оптимизированной моделями. Значение 0.0 свидетельствует о том, что между двумя моделями есть статистически значимые различия, p -значение 1.0 не позволяет отвергнуть нулевую гипотезу, что может быть связано с ограниченным объемом данных.

Полученные результаты свидетельствуют о том, что модель Random Forest Regressor демонстрирует неплохую производительность, особенно после оптимизации. Полученные оценки валидации, такие как среднеквадратичная ошибка (MSE), указывают на относительно хорошую точность модели.

Оптимизация гиперпараметров позволила улучшить производительность. Несмотря на наличие различий между исходной и оптимизированной моделями, ограниченный объем данных позволяет успешно моделировать влияние различных факторов на урожайность. При этом оптимизированная модель позволяет добиться лучшей точности прогнозов, что подтверждается результатами апробации.

Заключение

В контексте развивающейся цифровой трансформации сельского хозяйства, использование алгоритмов машинного обучения представляет собой значимый шаг в направлении повышения эффективности сельскохозяйствен-

ного производства. В настоящей статье представлен анализ применения методов машинного обучения в прогнозировании в сельском хозяйстве, в частности прогнозирование урожайности.

Разработанная модель, основанная на алгоритмах машинного обучения, успешно оценивает влияние различных факторов, таких как погодные условия, уровень удобрений, тип почвы и другие. Прогнозы модели согласуются с фактическими значениями, что подчеркивает ее значимость для оптимизации сельскохозяйственного производства.

В процессе апробации модели выявлены обнадеживающие результаты. MSE на тестовом наборе данных и при кросс-валидации свидетельствует о высокой точности модели. Дальнейшая оптимизация гиперпараметров позволила улучшить производительность. Таким образом, модель, основанная на Random Forest Regressor, успешно моделирует сложные взаимосвязи между различными факторами.

В заключение можно отметить, что применение методов машинного обучения в анализе сельскохозяйственных данных улучшает прогнозирование урожайности и предоставляет возможность для более эффективного управления ресурсами и повышения эффективности сельскохозяйственного производства.

Интеграция машинного обучения в агропромышленный сектор также открывает перспективы для современных подходов к управлению ресурсами. Рациональное распределение средств на основе прогнозов способствует эффективному использованию финансовых ресурсов, повышая устойчивость сельскохозяйственных предприятий в условиях переменчивости внешних условий. Полученные результаты открывают перспективы для дальнейших исследований и внедрения инновационных решений в сельское хозяйство.

Библиографические ссылки

1. Тюленева А. В., Джой Е. С. Инновационная экономика и ее развитие в России // Экономика сегодня: современное состояние и перспективы развития (Вектор-2021) : сборник материалов конференции. 2021. С. 287–291.

2. Талерчик С. М., Зайцев А. А., Шаванов М. В. Обеспечение экономической безопасности в контексте устойчивого инновационного развития агропромышленного комплекса в регионах России // Фундаментальные исследования. 2021. № 2. С. 57–65.

3. Cammarano D., van Evert F.K., Kempenaar C. Precision Agriculture: Modelling. Springer Cham, 2023. 301 p.

4. Кетова К. В., Касаткина Е. В., Вавилова Д. Д. Кластеризация регионов Российской Федерации по уровню социально-экономического развития с использованием методов машинного обучения // Экономические и социальные перемены: факты, тенденции, прогноз. 2021. Т. 14. № 6. С. 70–85. DOI: 10.15838/esc.2021.6.78.4.

5. Krishnan S., Anand A.J., Prasanth N., Goundar S., Ananth C. Predictive Analytics in Smart Agriculture (1st ed.). CRC Press, 2023. 312 p.

6. Ильченко С. В., Дубаневич Л. Э., Кубарский А. В. Перспективы использования интеллектуального капитала в отечественном агробизнесе // Modern Economy Success. 2020. № 6. С. 237–243.

7. Кубарский А. В., Трофимова Н. Н., Чиченков И. И. Перспективы использования КРІ в сельском хозяйстве // Эпомен. 2021. № 52. С. 54–60.

8. Трофимова Н. Н., Чиченков И. И., Домарацкая Е. А. Развитие сельского хозяйства в условиях экономической нестабильности // Modern Economy Success. 2020. № 6. С. 260–266.

9. Кирица А. А. Техническая оснащенность и оценка уровня доходности сельскохозяйственных организаций Московской области // Агроинженерия. 2020. № 5. С. 43–48.

10. Файзуллин Р. В., Симченко О. Л., Чиченков И. И. Методика оценки эффективности использования ресурсов предприятиями агрохолдинга // Наука и инновации XXI века: сборник статей конференции. 2020. С. 306–310.

11. Ходыревская В. Н., Припадчева И. В., Заикин Е. Н. Повышение эффективности управления сбытом готовой продукции на сельскохозяйственном предприятии // Вестник Курской государственной сельскохозяйственной академии. 2014. № 6. С. 11–13.

12. Зайцев А. А., Дмитриев Н. Д. Использование стресс-тестирования для повышения эффективности финансового планирования на предприятии // Цифровая экономика и Индустрия 4.0: новые вызовы: сборник статей конференции. 2019. С. 362–368.

13. Дмитриев Н. Д., Зайцев А. А. Математические методы в управлении промышленным производством // Фундаментальные и прикладные исследования в области управления, экономики и торговли : сборник трудов конференции. 2019. С. 208–212.

14. Алферьев Д. А., Родионов Д. Г. Место современных информационных компьютерных технологий в управлении инновационной деятельностью промышленных предприятий // Вестник Алтайской академии экономики и права. 2020. № 9-2. С. 199–203.

15. Vavilova D. D. Development of an effective adaptive forecasting system based on the combination of neural network and genetic algorithm // Journal of Physics: Conference Series. II International Scientific Conference on Metrological Support of Innovative Technologies (ICMSIT II-2021). Krasnoyarsk, 2021. 32029. DOI: 10.1088/1742-6596/1889/3/032029.

16. Kwaghtyo D. K., Eke C. I. Smart farming prediction models for precision agriculture: a comprehensive survey // *Artificial Intelligence Review*. 2023. Vol. 56. pp. 5729-5772.

17. Ding Y., Wang L., Li Y., Li D. Model predictive control and its application in agriculture: A review // *Computers and Electronics in Agriculture*. 2018. Vol. 151. Pp. 104-117.

18. Mourtzinis S., Esker P.D., Specht J.E. Advancing agricultural research using machine learning algorithms // *Scientific Reports*. 2021. Vol. 11. P. 17879.

19. Zaman A., Maitra S. Crop modeling: a tool for agricultural research // *MOJ Food Process Technol*. 2018. № 6(4). pp. 350-353.

References

1. Tyuleneva A.V., Joy E.S. *Innovatsionnaya ekonomika i ee razvitie v Rossii* [Innovative economy and its development in Russia]. *Ekonomika segodnya: sovremennoe sostoyanie i perspektivy razvitiya (Vektor-2021) : sbornik materialov konferentsii* [Proc. Economics today: current state and development prospects (Vector-2021): collection of conference materials]. 2021. Pp. 287-291 (in Russ.).

2. Talerchik S.M., Zaitsev A.A., Shavanov M.V. [Ensuring economic security in the context of sustainable innovative development of the agro-industrial complex in the regions of Russia]. *Fundamental'nye issledovaniya*. 2021. No. 2. Pp. 57-65 (in Russ.).

3. Cammarano D., van Evert F.K., Kempenaar C. *Precision Agriculture: Modeling*. Springer Cham, 2023. 301 p.

4. Ketova K.V., Kasatkina E.V., Vavilova D.D. [Clustering Russian Federation Regions According to the Level of Socio-Economic Development with the Use of Machine Learning Methods]. *Ekonomicheskie i sotsial'nye peremeny: fakty, tendentsii, prognoz*. 2021. Vol. 14, no. 6. Pp. 70-85 (in Russ.). DOI: 10.15838/esc.2021.6.78.4.

5. Krishnan S., Anand A.J., Prasanth N., Goundar S., Ananth C. *Predictive Analytics in Smart Agriculture* (1st ed.). CRC Press, 2023. 312 p.

6. Ilchenko S.V., Dubanevich L.E., Kubarsky A.V. [Prospects for the use of intellectual capital in domestic agribusiness]. *Modern Economy Success*. 2020. No. 6. Pp. 237-243 (in Russ.).

7. Kubarsky A.V., Trofimova N. N., Chichenkov I. I. [Prospects of using KPI in agriculture]. *Epomen*. 2021. No. 52. Pp. 54-60 (in Russ.).

8. Trofimova N.N., Chichenkov I.I., Domaratskaya E.A. [Development of agriculture in conditions of economic instability]. *Modern Economy Success*. 2020. No. 6. Pp. 260-266 (in Russ.).

9. Kiritsa A. A. [Technical equipment and assessment of the level of profitability of agricultural organizations in the Moscow region]. *Agroinzheneriya*. 2020. No. 5. Pp. 43-48 (in Russ.).

10. Fayzullin R.V., Simchenko O.L., Chichenkov I.I. *Metodika otsenki effektivnosti ispol'zovaniya resursov predpriyatiyami agrokholdinga* [Methodology for evaluating the efficiency of resource use by agricultural holding enterprises]. *Nauka i innovatsii XXI veka: sbornik statei konferentsii* [Proc. Science and innovations of the XXI century: collection of conference articles]. 2020. Pp. 306-310 (in Russ.).

11. Khodyrevskaya V.N., Pripadcheva I.V., Zaikin E.N. [Improving the efficiency of sales management of finished products at an agricultural enterprise]. *Vestnik Kurskoi gosudarstvennoi sel'skokhozyaistvennoi akademii*. 2014. No. 6. pp. 11-13 (in Russ.).

12. Zaitsev A.A., Dmitriev N.D. *Ispol'zovanie stress-testirovaniya dlya povysheniya effektivnosti finansovogo planirovaniya na predpriyatii* [The use of stress testing to improve the effectiveness of financial planning in an enterprise]. *Tsifrovaya ekonomika i Industriya 4.0: novye vyzovy: sbornik statei konferentsii* [Proc. Digital economy and Industry 4.0: new challenges: a collection of conference articles]. 2019. Pp. 362-368 (in Russ.).

13. Dmitriev N.D., Zaitsev A.A. *Matematicheskie metody v upravlenii promyshlennym proizvodstvom* [Mathematical methods in industrial production management]. *Fundamental'nye i prikladnye issledovaniya v oblasti upravleniya, ekonomiki i trgovli : sbornik trudov konferentsii* [Fundamental and applied research in the field of management, economics and trade: proceedings of the conference]. 2019. Pp. 208-212 (in Russ.).

14. Alferyev D.A., Rodionov D.G. [The place of modern information computer technologies in the management of innovative activities of industrial enterprises]. *Vestnik Altaiskoi akademii ekonomiki i prava*. 2020. No. 9-2. Pp. 199-203 (in Russ.).

15. Vavilova D.D. Development of an effective adaptive forecasting system based on the combination of neural network and genetic algorithm. In *Journal of Physics: Conference Series. II International Scientific Conference on Metrological Support of Innovative Technologies (ICMSIT II-2021)*. Krasnoyarsk, 2021. 32029. DOI: 10.1088/1742-6596/1889/3/032029.

16. Kwaghtyo D.K., Eke C.I. Smart farming prediction models for precision agriculture: a comprehensive survey. In *Artificial Intelligence Review*. 2023. Vol. 56. Pp. 5729-5772.

17. Ding Y., Wang L., Li Y., Li D. Model predictive control and its application in agriculture: A review. In *Computers and Electronics in Agriculture*. 2018. Vol. 151. Pp. 104-117.

18. Mourtzinis S., Esker P.D., Specht J.E. Advancing agricultural research using machine learning algorithms. In *Scientific Reports*. 2021. Vol. 11. P. 17879.

19. Zaman A., Maitra S. Crop modeling: a tool for agricultural research. In *MOJ Food Process Technol*. 2018. No. 6. Pp. 350-353.

Application of Machine Learning Methods for Forecasting in the Agricultural Sector of the Economy

R. V. Fayzullin, PhD in Economics, Associate Professor, MIREA – Russian Technological University, Moscow, Russia

O. A. Novikova, PhD in Engineering, MIREA – Russian Technological University, Moscow, Russia

This article is aimed at researching machine learning methods in the context of agriculture. Studying the issues of advanced mathematical method adaptation for the agricultural sector of the economy allows us to build effective management models and predict further development. The main purpose of this article is to analyze the application of advanced machine learning methods (linear regression, gradient boosting, random forest, neural networks (Deep Learning), Support Vector Machines (SVR), One-Class SVM, KNN, K-Means, principal component analysis (PCA)) in the agricultural sector in order to optimize production, increase the efficiency of agribusiness and predict parameters. Based on the analysis of key machine learning methods, conclusions regarding their applicability in agriculture were drawn. A special emphasis in the study is placed on the development and improvement of a model for predicting agricultural yields using machine learning methods. For practical testing, a basic model based on RandomForestRegressor was chosen, which allows predicting yield, taking into account various factors such as weather conditions, fertilizer level, soil type and other significant aspects. As part of the study, an algorithm for performing calculations using Python is presented, which allows for model training, visualization of the influence of parameters and statistical analysis. During the testing process, the model was optimized using GridSearchCV, aimed at selecting optimal hyperparameters. The analysis revealed the possibility of model improvement, yet, at low statistical significance. The obtained research results show that it is possible to apply machine learning to improve the productivity of forecasting key parameters in agriculture effectively.

Keywords: machine learning, smart agriculture, statistical analysis, artificial intelligence, production management, production efficiency.

Получено: 08.02.24

Образец цитирования

Файзуллин Р. В., Новикова О. А. Применение методов машинного обучения для прогнозирования в сельскохозяйственном секторе экономики // Интеллектуальные системы в производстве. 2024. Т. 22, № 2. С. 69–79. DOI: 10.22213/2410-9304-2024-2-69-79.

For Citation

Fayzullin R.V., Novikova O.A. [Application of Machine Learning Methods for Forecasting in the Agricultural Sector of the Economy]. *Intellektual'nye sistemy v proizvodstve*. 2024, vol. 22, no. 2, pp. 69-79. DOI: 10.22213/2410-9304-2024-2-69-79.