

Вахрушева Инна Алексеевна

доцент

Кукарцева Светлана Владиславовна

студентка

Щелинцева Светлана Вячеславовна

доцент

ФГБОУ ВО «Российский государственный
аграрный университет – МСХА им. К.А. Тимирязева»
г. Москва

ПРИМЕНЕНИЕ МЕТОДОВ BIG DATA ANALYTICS И МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ УРОЖАЙНОСТИ В СЕЛЬСКОМ ХОЗЯЙСТВЕ

***Аннотация:** в статье рассматривается проблема применения технологий Big Data для оптимизации сельхозпроизводства в условиях климатической неустойчивости. Проведена обработка данных (устранение пропусков, one-hot encoding, нормализация), разработано интерактивное приложение на Streamlit для анализа распределения культур, погодных условий и агротехнических факторов. Расчёт экономической эффективности показал наибольшую рентабельность риса (25,44%) и хлопка (22,14%), убыточность ячменя и сои. Оптимизация структуры посевных площадей на основе Big Data позволяет повысить рентабельность на 15–25%.*

***Ключевые слова:** большие данные, предиктивная аналитика, сельское хозяйство, агропромышленный комплекс, машинное обучение, обработка данных, Streamlit, визуализация данных, оптимизация посевных площадей, региональный анализ, климатические факторы, агротехнические мероприятия.*

Машинное обучение представляет собой область искусственного интеллекта, которая позволяет компьютерам обучаться на основе данных и делать предсказания или принимать решения без явного программирования. В последние годы применение машинного обучения в технических и исследовательских

областях стало активно развиваться, особенно в сфере контроля целостности и обнаружения отклонений в работе сложных систем. Это связано с тем, что современные научные исследования генерируют огромные объемы данных, включая результаты тестирований, показания сенсоров и структурную информацию об объектах. Использование алгоритмов машинного обучения позволяет извлекать значимую информацию из этих данных и повышать точность выявления нештатных ситуаций, что в свою очередь может привести к более эффективным методам обслуживания и предотвращения сбоев.

Машинное обучение находит широкое применение в различных аспектах мониторинга состояния объектов. Например, алгоритмы машинного обучения могут анализировать визуальные данные, такие как снимки поверхностей, топографические сканы и ультразвуковые проверки структуры, для выявления дефектов, которые могут указывать на нарушения целостности, такие как коррозия или усталость материала. Одним из наиболее известных примеров является использование глубоких нейронных сетей для анализа изображений, где модели обучаются на больших наборах данных и демонстрируют высокую точность в идентификации критических элементов. Кроме того, машинное обучение также используется для анализа цифровых журналов событий (лог-файлов) и архивов состояний, позволяя инженерам выявлять закономерности и предсказывать потенциальные сбои в работе оборудования на основе истории его эксплуатации и сопутствующих параметров.

Несмотря на значительные достижения в применении машинного обучения для диагностики состояния систем, существует множество вызовов, которые необходимо учитывать. Во-первых, качество и объем данных, используемых для обучения моделей, могут значительно варьироваться, что влияет на точность предсказаний. Необходимость обеспечения конфиденциальности и безопасности критически важных данных также создает дополнительные сложности. Во-вторых, технические специалисты должны понимать, как интерпретировать результаты, полученные с помощью методов машинного обучения, чтобы быть уверенными в правильности оценки ситуации. Наконец, необходимо учитывать

этические аспекты применения технологий, включая вопросы ответственности при ошибочных заключениях и необходимость соблюдения стандартов профессиональной практики. В будущем, с улучшением алгоритмов и увеличением объема разнообразных данных, использование машинного обучения в анализе состояния систем будет продолжать развиваться и изменять подходы к обеспечению надежности.

Современное сельское хозяйство сталкивается с вызовами, связанными с изменением климатических условий, истощением почв и необходимостью повышения эффективности производства. Ежегодно агропромышленный комплекс генерирует огромное количество информации от спутниковых снимков до датчиков на технике.

Анализ текущего состояния АПК в России показывает высокую степень региональной дифференциации. Лидерами по валовому сбору остаются Центрально-Черноземный регион и Южный федеральный округ, где черноземы и благоприятный климат позволяют получать стабильно высокие урожаи. В то же время регионы Поволжья и Сибири сталкиваются с рисками засухи и коротким вегетационным периодом. Есть два ключевых фактора, влияющих на урожайность, природные – это количество осадков, температура, содержание гумуса в почве и агротехнические – это использование удобрений, полив, качество семян. Исследования подтверждают, что даже при высоком плодородии почв (гумус повышает урожайность на 10–15%), без учета прочих переменных невозможно достичь максимальной эффективности.

В ходе работы был проведен анализ датасета, содержащего данные о регионах, типах почв, культурах, осадках, температуре и агротехнических мероприятиях. Первым этапом стала предобработка данных: проверка на пропуски и дубликаты, преобразование категориальных переменных в числовой формат (one-hot encoding) и нормализация числовых признаков с помощью StandardScaler.

Для визуализации результатов была выбрана библиотека Streamlit, которая позволяет быстро создавать интерактивные веб-приложения на Python. Разрабо-

танное приложение дает возможность: анализировать распределение культур по частоте встречаемости в выборке, сравнивать среднюю температуру по различным регионам, оценивать количество осадков в зависимости от типа почв, изучать влияние удобрений на сроки созревания культур с помощью box-plot диаграмм.

Заключительным этапом стал расчет экономической эффективности на основе полученных данных. Используя фиксированные рыночные цены и рассчитанную себестоимость, были определены показатели рентабельности для различных культур.

Результаты анализа показали, что наибольшую рентабельность демонстрируют рис (25.44%) и хлопок (22.14%), в то время как ячмень и соя при заданных условиях оказались убыточными культурами. Данный анализ подтверждает, что оптимизация структуры посевных площадей на основе данных Big Data может повысить рентабельность производства на 15–25%.

Список литературы

1. Информационные системы и технологии в АПК: учебник / А.В. Бабкина, И.Е. Быстренина, М.И. Горбачев [и др.]. – М.: Российский государственный аграрный университет, 2025. – 615 с. EDN ZJCSXJU

2. Высоцкий Р.Н. Разработка программного обеспечения для автоматического контроля качества фруктов и овощей на базе малых сельхозпредприятий / Р.Н. Высоцкий, А.С. Прудкий, С.В. Кукарцева // Стратегии устойчивого развития: социальные, экономические и юридические аспекты: материалы VII Всерос. науч.-практ. конф. с междунар. участием. – Чебоксары: Среда, 2025. – С. 133–137. EDN PHOWEI

3. Прогнозирование тепловых процессов в подземных горных сооружениях с использованием рекуррентных нейронных сетей / М.В. Журавлев, О.А. Антамошкин, М.Н. Степанцевич, Е.Ф. Малыха // Известия Тульского государственного университета. Науки о Земле. – 2025. – №3. – С. 205–211. EDN LZFRCG

4. Содержание структурных углеводов и лигнина в многолетних злаковых кормовых травах в зависимости от фазы роста и цифровые инновации в анализе состава корма / Х.К. Худякова, Е.В. Худякова, М.Н. Степанцевич [и др.] // Тимирязевский биологический журнал. – 2023. – №4. – С. 107–115. DOI 10.26897/2949-4710-2023-4-107-115. EDN PHTЕAA