

Торайғыров университетінің хабаршысы
ҒЫЛЫМИ ЖУРНАЛЫ

НАУЧНЫЙ ЖУРНАЛ
Вестник Торайғыров университета

Торайғыров университетінің ХАБАРШЫСЫ

Энергетикалық сериясы
1997 жылдан бастап шығады



ВЕСТНИК Торайғыров университета

Энергетическая серия
Издается с 1997 года

ISSN 2710-3420

№ 3 (2023)

ПАВЛОДАР

НАУЧНЫЙ ЖУРНАЛ
Вестник Торайгыров университета

Энергетическая серия
выходит 4 раза в год

СВИДЕТЕЛЬСТВО

о постановке на переучет периодического печатного издания,
информационного агентства и сетевого издания

№ 14310-Ж

выдано

Министерство информации и общественного развития
Республики Казахстан

Тематическая направленность

публикация материалов в области электроэнергетики,
электротехнологии, автоматизации, автоматизированных и
информационных систем, электромеханики и теплоэнергетики

Подписной индекс – 76136

<https://doi.org/10.48081/YBCY7199>

Бас редакторы – главный редактор

Кислов А. П.
к.т.н., доцент

Заместитель главного редактора

Талипов О. М., *доктор PhD*

Ответственный секретарь

Калтаев А.Г., *доктор PhD*

Редакция алқасы – Редакционная коллегия

Клецель М. Я.,	<i>д.т.н., профессор</i>
Новожилов А. Н.,	<i>д.т.н., профессор</i>
Никитин К. И.,	<i>д.т.н., профессор (Россия)</i>
Никифоров А. С.,	<i>д.т.н., профессор</i>
Новожилов Т. А.,	<i>д.т.н., профессор</i>
Алиферов А.И.,	<i>д.т.н., профессор (Россия)</i>
Кошеков К.Т.,	<i>д.т.н., профессор</i>
Приходько Е.В.,	<i>к.т.н., профессор</i>
Оспанова Н. Н.,	<i>к.п.н., доцент</i>
Нефтисов А. В.,	<i>доктор PhD</i>
Омарова А.Р.,	<i>технический редактор</i>

За достоверность материалов и рекламы ответственность несут авторы и рекламодатели

Редакция оставляет за собой право на отклонение материалов

При использовании материалов журнала ссылка на «Вестник Торайгыров университета» обязательна

© Торайгыров университет

<https://doi.org/10.48081/PXQO4695>

***А. С. Аканова¹, Н. Н. Оспанова², А. С. Казанбаева³,
Г. А. Анарбекова⁴, С. Е. Шарипова⁵**

^{1,4,5}Казахский агротехнический университет имени С. Сейфуллина,
Республика Казахстан, г. Астана,

²Торайғыров университет, Республика Казахстан, г. Павлодар

³Северо-Казахстанский университет имени М. Козыбаева,
Республика Казахстан, г. Петропавловск

e-mail: akerkegansaj@mail.ru

МЕТОДЫ ПРОГНОЗИРОВАНИЯ В МАШИННОМ ОБУЧЕНИИ: ОБЗОР И СРАВНЕНИЕ

В данной статье представлен обзор и сравнение различных методов прогнозирования в машинном обучении. Описаны методы линейной и логистической регрессии, деревьев решений, случайного леса, градиентного бустинга и метода опорных векторов. Также исследованы и проанализирована архитектура нейронных сетей, методы кластеризации, которые используются для выявления групп в данных на основе сходства. С целью выбора наилучшего метода для дальнейшего исследования, были рассмотрены преимущества и недостатки каждого метода прогнозирования. Проведен анализ статей опубликованных в журналах входящих в базу Scopus, чтобы получить обзор опубликованных методов прогнозирования урожайности. Это исследование показало, что в выборке из 101 статьи подходы, основанные на процессах и на основе данных, были представлены в равной степени. Из анализа литературы стало ясно, что основным препятствием для широкого внедрения комплексных подходов, основанных на данных, является отсутствие подходящих наборов данных, в то время как модели роста сельскохозяйственных культур сталкиваются с проблемой неспособности моделировать коэффициенты снижения, требуя часто многочисленные входные факторы. В целом, правильный выбор метода прогнозирования является ключевым фактором для достижения успешных результатов в машинном обучении.

Ключевые слова: методы прогнозирования, машинное обучение, нейронные сети, сельское хозяйство, урожайность.

Введение

Прогнозирование играет критическую роль во многих аспектах современной жизни, так как позволяет принимать информированные решения и строить стратегии развития на основе анализа данных и предсказания будущих тенденций. Машинное обучение, как один из основных инструментов для прогнозирования, предлагает ряд методов и алгоритмов для решения сложных задач в разных отраслях.

Машинное обучение играет важную роль в прогнозировании, так как предоставляет набор методов и алгоритмов, способных обрабатывать и анализировать большие объемы данных для выявления закономерностей и создания прогнозных моделей. В контексте прогнозирования машинное обучение помогает улучшить точность и скорость предсказаний, а также обеспечивает автоматизацию и оптимизацию процесса прогнозирования.

Методы машинного обучения набирают популярность в области прогнозирования. Однако понимание того, почему и как эти методы хорошо работают для прогнозирования, все еще находится на очень ранней стадии, отчасти из-за их сложности. Метод прогнозирования в машинном обучении на основе регрессии, которая предоставляет исследователям общий язык и абстракцию, помогает в их исследованиях. Чтобы продемонстрировать полезность фреймворка, [1] представлено, как его можно использовать для сопоставления и сравнения методов машинного обучения.

Деревья решений описывают тип метода машинного обучения, который широко используется в науках о Земле для автоматического извлечения шаблонов из сложных и [многомерных данных](#). Однако, как и любой метод, основанный на данных, применению препятствуют ограничения данных, такие как значительные погрешности, приводящие к потенциально физически нереалистичным результатам. [2] создали набор инструментов Python с открытым исходным кодом. Пользователи могут в интерактивном режиме создавать новые составные переменные, изменять переменную и пороговое значение для разделения, сокращения и группировки переменных на основе их физического значения.

Наличие загрязненных земель является неизбежным наследием промышленной деятельности, и управленческие решения, регулирующие рекультивацию этих земель, играют ключевую роль в минимизации экологических рисков и обеспечении безопасного и эффективного повторного использования земель. В этом контексте, прогнозировались оптимальные варианты реабилитации для будущих процессов принятия решений в области

устойчивого управления объектами, тем самым улучшив информационный обмен между заинтересованными сторонами. В исследовании [3] анализировалось 17 параметров чувствительности решений, и их влияние на модели управления загрязненными объектами определяется тремя алгоритмы дерева решений.

Машинное обучение на множествах с целью последовательного вывода является важной и повсеместной задачей. Приложения которой варьируются от языкового моделирования и мета-обучения до многоагентных стратегических игр и оптимизации электросетей. Сочетая элементы обучения представлению и структурированному прогнозированию, его две основные задачи включают в себя получение значимого представления множества, инвариантного к перестановкам, и последующее использование этого представления для вывода сложной целевой перестановки [4].

Методы машинного обучения применяются для объединения результатов в рамках федеративного патентного поиска. Несмотря на то, что было разработано несколько методов объединения результатов, ни один из них не был протестирован на патентных данных и не рассматривал несколько моделей машинного обучения. Таким образом, авторы экспериментируют с самыми современными методами используя патентные данные. Модель случайного леса обеспечивает наилучшие результаты по сравнению со всеми другими моделями и дает новое представление о проблеме объединения результатов. Доказано [5], что модели машинного обучения могут заменить другие стандартные методы и модели, которые использовались для объединения результатов в течение многих лет.

Методы ансамбля (алгоритмы обучения), которые создают набор классификаторов, а затем классифицируют новые точки данных путем взвешенного голосования по их прогнозам. Оригинальный метод ансамбля - байесовское усреднение, но более поздние алгоритмы включают кодирование выходных данных с исправлением ошибок. Рассмотрено [6], почему ансамбли часто могут работать лучше, чем другие методы прогнозирования.

Спектральные методы имеют фундаментальное значение в статистике и машинном обучении, поскольку они лежат в основе алгоритмов от классического анализа главных компонент до более современных подходов, использующих многообразную структуру. В большинстве случаев основная техническая проблема может быть сведена к вычислению низкорангового приближения к положительно определенному ядру. Однако для растущего числа приложений, имеющих дело с очень большими или многомерными наборами данных, оптимальное приближение, обеспечиваемое точным спектральным разложением, является слишком дорогостоящим, поскольку его сложность масштабируется как куб любого

количество обучающих примеров или их размерность. Представлен [7] 2 новых алгоритма для аппроксимации положительно-полуопределенных ядер вместе с оценками ошибок, которые улучшают результаты, приведенные в литературе.

Представлен всесторонний обзор исследований зарубежных ученых, посвященных применению машинного обучения в системах сельскохозяйственного производства. Фильтрация и классификация представленных статей демонстрируют, как сельское хозяйство выиграет от технологий машинного обучения. Применяя машинное обучение к данным датчиков, системы управления фермами превращаются в программы с поддержкой искусственного интеллекта в реальном времени, которые предоставляют подробные рекомендации и аналитическую информацию для поддержки принятия решений и действий фермерами [8].

Возросшая доступность данных от регионального до полевого масштаба проложила путь для новых подходов, основанных на данных.

Целью этой статьи является сравнение методов прогнозирования урожайности в статьях опубликованных журналах входящих в базу Scopus.

Материалы и методы исследования

В исследовании были использованы следующие методы: анализ, научное обобщение сравнение. С помощью научного обобщения будут рассмотрены и обобщены результаты существующих опытов в области исследования урожайности.

Методы машинного обучения относятся к широкой категории, охватывающей подходы различной сложности. Несколько исследований были посвящены сравнению различных алгоритмов машинного обучения для прогнозирования урожайности. Например, в зарубежном исследовании [9] сравнили различные методы машинного обучения для прогнозирования урожайности орошаемых полей в орошаемом бассейне в Мексике. Другие ученые [10] протестировали несколько подходов к машинному обучению (искусственная нейронная сеть, регрессия опорных векторов, k-ближайшие соседи и случайный лес), используя региональные данные. Точно так же ученые из Кореи [11] сравнили различные методы (нейронные сети с различной архитектурой, машины опорных векторов, случайные леса) для прогнозирования урожайности кукурузы и сои на Среднем Западе США. В целом наблюдается, что подходы, характеризующиеся многими параметрами, требуют большого объема обучающих данных, например, во многих исследованиях нейронные сети использовались для прогнозирования урожайности с использованием большого набора данных [12], тогда как при наличии меньшего количества данных (и, как правило, с более высоким разрешением) более простые модели могут быть предпочтительнее.

Например, как [13] использовали деревья регрессии для прогнозирования урожайности на уровне полей, используя данные 48 полей сои в Китае и 48 полей пшеницы фермеров Бангладеш соответственно.

Машинное обучение является быстроразвивающейся областью, и методы прогнозирования не являются исключением. В Таблице 1 представлены методы прогнозирования, где слева преимущество метода, а справа соответственно приведены недостатки.

Таблица 1 – Методы прогнозирования (преимущества и недостатки)

Линейная и логистическая регрессия	
простая математическая формулировка и легко интерпретируема, что позволяет пользователям легко понимать влияние признаков на результат. обучение и предсказание требуют относительно мало вычислительных ресурсов, что делает их подходящими для работы с большими наборами данных легко реализовать и настроить с помощью существующих библиотек машинного обучения, таких как scikit-learn в Python	линейная зависимость между признаками и целевой переменной, что может быть неприемлемым для сложных, нелинейных данных чувствительны к выбросам и мультиколлинеарности (когда два или более признаков сильно коррелируют друг с другом) в данных, что может снижать качество модели логистическая регрессия ограничена бинарными классификационными задачами логистическая регрессия может столкнуться с проблемой локальных минимумов при оптимизации функции потерь
Деревья решений	
легко визуализировать и объяснить, что делает их полезными для обучения и презентаций могут работать с обоими типами данных без предварительной обработки не чувствительны к масштабу признаков. могут идентифицировать важные признаки во время обучения.	требует больше времени и ресурсов для обучения и предсказания, чем отдельные деревья решений. чувствительны к изменениям в данных, что может привести к значительным изменениям в структуре дерева. строятся жадным алгоритмом, который оптимизирует каждое разбиение локально, что может привести к подоптимальному глобальному решению.
Случайный лес	
ансамблевый подход случайного леса уменьшает риск переобучения. может достичь более высокой точности, чем отдельные деревья решений, благодаря усреднению предсказаний. может эффективно обрабатывать большие наборы данных и большое количество признаков.	требует больше времени и ресурсов для обучения и предсказания, чем отдельные деревья решений. хотя отдельные деревья решений легко интерпретировать, интерпретация ансамбля деревьев может быть сложнее.
Градиентный бустинг	

<p>может достигать высокой точности прогноза за счет последовательного улучшения ошибки предыдущих моделей на каждом шаге.</p> <p>может обрабатывать категориальные и числовые признаки, а также текстовые данные, используя специальные методы кодирования.</p> <p>поддерживает регуляризацию, что помогает снизить переобучение и улучшить обобщающую способность модели.</p>	<p>требует больше времени и ресурсов для обучения и предсказания, чем отдельные деревья решений или случайный лес.</p> <p>может быть более подвержен переобучению, если гиперпараметры не настроены правильно.</p> <p>может быть чувствительным к выбросам в данных, что может привести к низкому качеству прогноза.</p>
Метод опорных векторов (Support Vector Machines, SVM)	
<p>Высокая точность и эффективность для решения задач классификации, регрессии и детектирования выбросов;</p> <p>Поддержка использования различных ядерных функций для обработки разных типов данных;</p> <p>Минимальная чувствительность к локальным экстремумам в функции ошибки.</p>	<p>Высокая сложность настройки гиперпараметров, таких как ширина ядра;</p> <p>Требуется масштабирование признаков для достижения оптимальной производительности;</p> <p>Неэффективность для работы с большими наборами данных, так как требует значительных вычислительных ресурсов.</p>
Нейронные сети	
<p>Высокая точность в решении сложных задач;</p> <p>Умение извлекать сложные закономерности из данных;</p> <p>Способность обрабатывать большие объемы данных;</p> <p>Адаптивность к новым данным и ситуациям.</p>	<p>Требовательность к вычислительным ресурсам и времени;</p> <p>Сложность настройки гиперпараметров и архитектуры нейронной сети;</p> <p>Часто требуется большой объем данных для эффективного обучения.</p>
Методы кластеризации (K-средних (K-means), Иерархическая кластеризация (Hierarchical clustering), DBSCAN)	
<p>Возможность выявлять скрытые структуры в данных;</p> <p>Не требуют заранее известного числа кластеров;</p> <p>Могут использоваться для обработки больших объемов данных.</p>	<p>Результаты могут зависеть от начальной инициализации и параметров алгоритма;</p> <p>Необходимость выбора подходящего метода и его настройки для конкретной задачи;</p> <p>Трудность интерпретации результатов.</p>

Чтобы получить обзор опубликованных методов прогнозирования урожайности, проведен обзор литературы, прочитали и извлекли информацию о методологии, использованной в 109 отобранных статьях. Цель систематического обзора состояла в том, чтобы понять масштаб и исходные данные, используемые для прогнозирования урожайности, основанных на данных и процессах. Систематический обзор был также направлен на выявление тематических исследований, в которых эти два подхода были объединены.

Проведен поиск в базе данных рефератов и цитирования Scopus, используя следующий запрос:

DOCTYPE (ar or re) AND TITLE-ABS-KEY (crop yield) AND TITLE-ABS-KEY(predict OR forecast* OR monitor* OR calculat*) AND TITLE-ABS-

KEY («decision support system» OR regional OR farm* OR tactical OR strategic OR scenario* OR policy* OR DSS OR (compar* AND method) OR review OR AI OR «machine learning» OR «deep learning» OR «random forest» OR «neural network» OR «support vector machines» OR «artificial intelligence» OR «big data» OR «statistic* model*») AND PUBYEAR > 2010.

Разбили запрос на следующие блоки, связанные оператором AND:

Опубликовано в период с 2009 по 2023 год (запрос был выполнен 12 мая 2023 года)

Включено «прогностическое» слово (прогнозировать ИЛИ прогнозировать* ИЛИ отслеживать* ИЛИ вычислять*)

Включено слово из списка слов цели (например, система поддержки принятия решений) или слово из списка слов метода (например, машинное обучение, модель роста урожая).

Из списка статей, отобранных по названию, исследованы 109 из них: восемь статей из 109 были обзорными, поэтому они были не включены в систематический обзор.

Для каждой из 101 включенной статьи зафиксировали следующую информацию:

1 Пространственный масштаб. Эта категория представляет пространственный масштаб данных. Информация устанавливается как двухуровневая переменная: масштаб поля или фермы или выше (обычно округ, провинция или штат).

2 Категория модели: эта информация представляла собой список методов, использованных в исследовании для прогнозирования урожайности, сократили неоднородность возможных методологий до двух категорий: управляемые данными (линейная и нелинейная регрессии) и основанные на процессах (модели роста урожая). В 12 случаях использовались оба подхода, все данные представлены на рисунке 1.

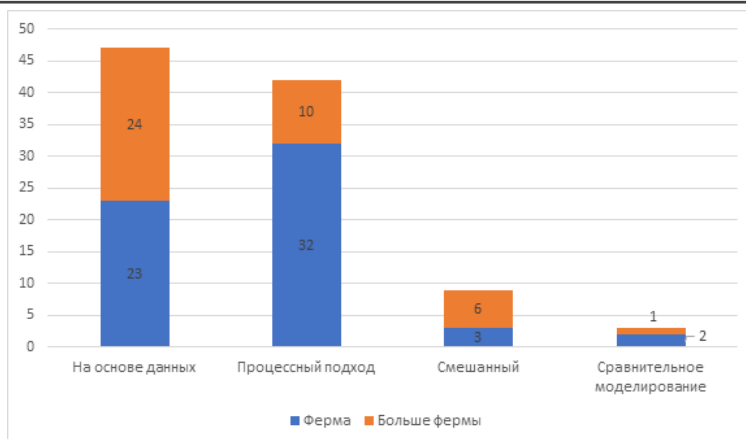


Рисунок 1 – Таблица непредвиденных обстоятельств статей по масштабу и подходу.

3 Входные данные: погода, почва, управление (включая орошение, удобрение, дату посадки), сорт и отражательная способность растительного покрова, биологические переменные культуры (например, дата [цветения](#) или содержание азота в листьях) и другие (для входных данных, которые использовались менее чем в двух исследованиях). Входные данные считались самыми необработанными, а дальнейшие переменные, полученные из исходных входных данных, не учитывались.

Результаты и обсуждение

Анализ пространственного масштаба прогнозов показал, что модели, основанные на процессах, чаще использовались в масштабе поля или ниже (32 против 10 случаев, рисунок 1). Напротив, модели, управляемые данными, использовались одинаково в разных пространственных масштабах (23 против 24). Преобладающее использование моделей, основанных на процессах, в масштабе поля или ниже, вероятно, отражает потребность в управленческой информации (например, о внесении удобрений, дате посева) для их запуска. Очевидно, что модели, основанные на процессах, можно запускать в региональном масштабе с учетом только погодных условий, однако в этих случаях пользователю необходимо делать предположения об [управлении культурами](#).

В целом обзор литературы выявил, по крайней мере, три основных направления развития исследований моделей роста сельскохозяйственных культур: ассимиляция данных в моделях роста сельскохозяйственных культур для улучшения внутрисезонного прогнозирования, интеграция моделей роста

сельскохозяйственных культур и генетики, включение факторов снижения урожайности.

Правильный выбор метода прогнозирования является ключевым фактором для достижения успешных результатов в машинном обучении. Это помогает обеспечить высокую точность и эффективность прогнозирования, а также улучшить принятие решений и повысить эффективность бизнес-процессов.

Информация о финансировании (при наличии)

Данное исследование финансируется Комитетом науки Министерства науки и высшего образования Республики Казахстан (грант №. AP19675312).

Выводы

Это исследование показало, что в выборке из 101 статьи подходы, основанные на процессах и на основе данных, были представлены в равной степени. Из анализа литературы стало ясно, что основным препятствием для широкого внедрения комплексных подходов, основанных на данных, является отсутствие подходящих наборов данных, в то время как модели роста сельскохозяйственных культур сталкиваются с проблемой неспособности моделировать коэффициенты снижения, требуя часто многочисленные входные факторы. Мы нашли частичное подтверждение того, что в нескольких исследованиях использовалось сочетание этих двух подходов для прогнозирования урожайности, и мы определили три основных способа смешивания двух подходов: использование моделей роста урожая для разработки характеристик или индексов, использование моделей, управляемых данными, для оценки недостающего урожая, ввод моделей роста, создание метамоделей для снижения вычислительной нагрузки.

Исходя из анализа и сравнения получаем, что исследование с применением нейросети остается открытым вопросом и требует дополнительного тщательного исследования.

Наконец, важно помнить, что использование методов прогнозирования в машинном обучении является всего лишь инструментом для решения задач, и правильный выбор метода и правильная обработка данных не гарантируют успеха. Важно учитывать контекст задачи и принимать во внимание все необходимые факторы при принятии решений.

СПИСОК ИСПОЛЬЗОВАННЫХ ИСТОЧНИКОВ

1 **Casper, S. B.** Understanding machine learning-based forecasting methods: A decomposition framework and research opportunities [Text] / S.B. Casper // International Journal of Forecasting. – Volume 38. – Issue 4. – 2022, – P. 1555–1561.

2 **Georgios, S.** Integrating scientific knowledge into machine learning using interactive decision trees [Text] / S. Georgios, W. Thorsten, P. Francesca // Computers & Geosciences. – Volume 170. – 2023. – P. 105248.

3 **Xiaonuo, L.** Sustainable decision-making for contaminated site risk management: A decision tree model using machine learning algorithms [Text] / L. Xiaonuo, Y. Shiyi, B.C. Andrew, Ch. Weiping // Journal of Cleaner Production. – Volume 371, 2022. – P. 133612.

4 **Jurewicz, M.** Set-to-sequence methods in machine learning: A review [Text] / M. Jurewicz, L. Derezynski // Journal of Artificial Intelligence Research, 2021. – Volume 71. – P. 885–924.

5 **Stamatis, V.** Machine learning methods for results merging in patent retrieval [Text] / V. Stamatis, M. Salampassis, K. Diamantaras // Data Technologies and Applications. – 2023. – P. 2514.

6 **Dietterich, T.** PAC Guarantees and Effective Algorithms for Detecting Novel Categories [Text] / T.G [Dietterich](#), S. [Liu](#) // [Journal of Machine Learning Research](#). – 2022. – P. 23.

7 **Belabbas, M.** Spectral methods in machine learning and new strategies for very large datasets [Text] / M. Belabbas, P.J. Wolfe // Proceedings of the National Academy of Sciences of the United States of America. – 106(2). – 2009. – P. 369–374.

8 **Liakos, K.** Machine learning in agriculture: A review [Text] / K.G. Liakos, P. Busato, D. Moshou, S. Pearson, D. Bochtis // Sensors (Switzerland). – 18(8). – 2018. – P. 2674.

9 **Gonzalez-Sanchez, A.** Attribute Selection Impact on Linear and Nonlinear Regression Models for Crop Yield Prediction [Text] / A. Gonzalez-Sanchez, J. Frausto-Solis, W. Ojeda-Bustamante // The Scientific World Journal. – Vol. 2014. – 2014. – P. 10.

10 **Maya Gopal P.** Performance Evaluation of Best Feature Subsets for Crop Yield Prediction Using Machine Learning Algorithms [Text] / P.S. Maya Gopal, R. Bhargavi // Applied Artificial Intelligence. – 33:7. – 2019. – P. 621–642.

11 **Kim N. A** comparison between major artificial intelligence models for crop yield prediction: Case study of the midwestern United States, 2006–2015 [Text] / N. Kim, N.W. Park // ISPRS International Journal of Geo-Information 8.5. – 2019. – P. 240.

12 **Crane-Droesch, A.** Machine learning methods for crop yield prediction and climate change impact assessment in agriculture [Text] / A. Crane-Droesch // Environmental Research Letters. – 2018. – Volume 13. – №. 11. – P. 114003.

13 **Haifeng, Zh.** Classification and regression tree (CART) for analysis of soybean yield variability among fields in Northeast China: The importance of phosphorus application rates under drought conditions [Text] / Zh. Haifeng,

14 Ch. Liding, H. Xiaozeng, Zh. Xinfeng, M. Yan // Agriculture, Ecosystems & Environment. – Volume 132. – Issues 1–2, 2009. – P. 98-105.

Принято к изданию 18.09.23.

*А. С. Аканова¹, Н. Н. Оспанова², А. С. Казанбаева³,
Г. А. Анарбекова⁴, С. Е. Шарипова⁵

^{1,4,5}С. Сейфуллин атындағы Қазақ агротехникалық университеті,
Қазақстан Республикасы, Астана қ;

²Торайғыров университеті, Қазақстан Республикасы, Павлодар қ;

³М. Қозыбаев атындағы Солтүстік Қазақстан университеті,
Қазақстан Республикасы, Петропавл қ.

e-mail: akerkegansaj@mail.ru

Басып шығаруға 18.09.23 қабылданды.

МАШИНАЛЫҚ ОҚЫТУДАҒЫ БОЛЖАУ ӘДІСТЕРІ: ШОЛУ ЖӘНЕ САЛЫСТЫРУ

Бұл мақалада машиналық оқытудағы әртүрлі болжау әдістеріне шолу және салыстыру берілген. Сызықтық және логистикалық регрессия әдістері, шешім ағаштары, кездейсоқ орман, градиентті күшейту және тірек векторлық әдіс сипатталған. Сондай-ақ, нейрондық желілердің архитектурасы, ұқсастыққа негізделген деректердегі топтарды анықтау үшін қолданылатын кластерлеу әдістері зерттеліп, талданды. Әрі қарай зерттеу үшін ең жақсы әдісті таңдау мақсатында әр болжау әдісінің артықшылықтары мен кемшіліктері қарастырылды. Жарияланған өнімділікті болжау әдістеріне шолу жасау үшін Scopus базасына кіретін журналдарда жарияланған мақалаларға талдау жасалды. Бұл зерттеу 101 мақаланың үлгісінде процестерге негізделген және деректерге негізделген тәсілдер бірдей ұсынылғанын көрсетті. Әдебиеттерді талдаудан деректерге негізделген кешенді тәсілдерді кеңінен енгізудің негізгі кедергісі қолайлы деректер жиынтығының болмауы екені белгілі болды, ал дақылдардың осу үлгілері төмендеу коэффициенттерін модельдей алмау мәселесіне тап болып, жиі талап етеді. көптеген кіріс факторлары. Тұтастай алғанда, болжау әдісін дұрыс таңдау машиналық оқытуда сәтті нәтижелерге қол жеткізудің негізгі факторы болып табылады.

Кілтті сөздер: болжау әдістері, машиналық оқыту, нейрондық желілер, ауыл шаруашылығы, өнімділік.

*A. S. Akanova¹, N. N. Ospanova², A. S. Kazanbayeva³,
G. A. Anarbekova⁴, S. E. Sharipova⁵

^{1,4,5}S. Seifullin Kazakh Agrotechnical University,
Republic of Kazakhstan, Astana

²Toraighyrov University, Republic of Kazakhstan, Pavlodar

³North Kazakhstan University named after M. Kozybayev,
Republic of Kazakhstan, Petropavlovsk

e-mail: akerkegansaj@mail.ru

Accepted for publication on 18.09.23

FORECASTING METHODS IN MACHINE LEARNING: OVERVIEW AND COMPARISON

This article provides an overview and comparison of various forecasting methods in machine learning. The methods of linear and logistic regression, decision trees, random forest, gradient boosting and the support vector method are described. The architecture of neural networks, clustering methods that are used to identify groups in data based on similarity are also investigated and analyzed. In order to choose the best method for further research, the advantages and disadvantages of each forecasting method were considered. The analysis of articles published in journals included in the Scopus database was carried out in order to get an overview of published methods for forecasting yields. This study showed that in a sample of 101 articles, process-based and data-based approaches were equally represented. From the analysis of the literature, it became clear that the main obstacle to the widespread implementation of integrated data-based approaches is the lack of suitable data sets, while crop growth models face the problem of inability to model reduction coefficients, often requiring multiple input factors. In general, the correct choice of the forecasting method is a key factor for achieving successful results in machine learning.

Keywords: forecasting methods, machine learning, neural networks, agriculture, productivity.

Теруге 18.09.2023 ж. жіберілді. Басуға 29.09.2023 ж. қол қойылды.

Электронды баспа

29.9 Мб RAM

Шартты баспа табағы 22,2. Таралымы 300 дана. Бағасы келісім бойынша.

Компьютерде беттеген: А. К. Мыржикова

Корректор: А. Р. Омарова, Д. А. Кожас

Тапсырыс №4140

Сдано в набор 18.09.2023 г. Подписано в печать 29.09.2023 г.

Электронное издание

29.9 Мб RAM

Усл. печ. л. 22,2. Тираж 300 экз. Цена договорная.

Компьютерная верстка: А. К. Мыржикова

Корректор: А. Р. Омарова, Д. А. Кожас

Заказ № 4140

«Toraighyrov University» баспасынан басылып шығарылған

Торайғыров университеті

140008, Павлодар қ., Ломов к., 64, 137 каб.

«Toraighyrov University» баспасы

Торайғыров университеті

140008, Павлодар қ., Ломов к., 64, 137 каб.

67-36-69

E-mail: kereku@tou.edu.kz

www.vestnik-energy.tou.edu.kz