

УДК 004.89  
Код ГРНТИ 28.23.25

doi 10.54708/19926502\_2024\_28410644

## Применение технологий машинного обучения для прогнозирования временных рядов с учетом влияния внешних факторов в агропромышленной сфере

М.Р. Ибатуллин, А.С. Ковтуненко\*

ФГБОУ ВО «Уфимский университет науки и технологий» (УУНиТ), г. Уфа, Россия

**Аннотация.** Работа посвящена разработке и применению методов прогнозирования временных рядов в АПК с учетом внешних факторов. Анализируются современные подходы, основанные на глубоком обучении (NeuralProphet) и градиентном бустинге (CatBoost), для повышения точности прогнозирования цен на сельскохозяйственную продукцию. Исследование включает в себя анализ влияния различных факторов, таких как стоимость кормов, правительственные регулирования, новости и сезонность, а также разработку стратегии кодирования категориальных признаков (One-Hot Encoding, Ordinal Encoding). Результаты демонстрируют существенное улучшение точности прогнозов по сравнению с традиционными методами, подчеркивая важность интеграции внешних данных для принятия эффективных управленческих решений в АПК.

**Ключевые слова:** Прогнозирование временных рядов, Агропромышленный комплекс, машинное обучение, NeuralProphet, CatBoost, внешние факторы, анализ новостей, кодирование признаков.

\*askovtunenکو@mail.ru

### Введение

Прогнозирование динамики рынков в агропромышленном комплексе (АПК) является важной задачей для принятия эффективных управленческих решений. Точные и своевременные прогнозы помогают производителям, переработчикам и поставщикам сельскохозяйственной продукции лучше планировать производство, закупки, логистику и ценообразование. Однако традиционные методы прогнозирования временных рядов, основанные исключительно на исторических данных, часто оказываются недостаточными для учета многочисленных внешних факторов, влияющих на динамику рынков АПК.

В последние годы в области прогнозирования временных рядов наблюдается значительный прогресс, связанный с учетом влияния внешних факторов. Эти подходы позволяют интегрировать в модели прогнозирования разнообразную информацию о событиях, явлениях и параметрах, которые оказывают влияние на исследуемые рыночные показатели. Такая интеграция внешних данных, включая новостные сообщения, экспертные оценки, статистические показатели и другие релевантные факторы, существенно повышает точность и обоснованность прогнозов.

Настоящее исследование посвящено разработке и применению передовых методов прогнозирования временных рядов с учетом анализа внешних факторов в сфере агропромышленного комплекса. Рассматриваются современные SOTA-подходы, демонстрирующие высокую эффективность в интеграции разнообразной информации для повышения точности прогнозирования динамики цен, спроса и предложения на ключевые виды сельскохозяйственной продукции.

### Современные подходы к прогнозированию временных рядов с использованием анализа внешних признаков

В последние годы в области прогнозирования временных рядов произошел настоящий прорыв благодаря использованию анализа внешних источников информации. Методы анализа

внешних источников информации повысили точность прогнозов по сравнению с традиционными подходами, которые опираются только на исторические данные.

Несколько подходов, которые показывают, как технологии машинного обучения позволяют учитывать некоторые внешние признаки для предсказания временного ряда:

1. Глубокие нейронные сети с анализом новостных текстов. В статье "Improving Stock Market Prediction with News Sentiment Analysis Using Attention-Based Recurrent Neural Networks" (опубликовано в журнале IEEE Access в 2021 году) [1], исследователи представили архитектуру, основанную на рекуррентных нейронных сетях с механизмом внимания. Эта модель анализирует тональность новостей для повышения точности прогнозирования стоимости акций на фондовом рынке.

2. Гибридные модели, сочетающие машинное обучение и анализ эмоциональной окраски. В работе "Hybrid Approach for Stock Market Prediction Using News Sentiment Analysis and Fuzzy Logic" (журнал IEEE Access, 2020) [2], авторы разработали гибридную модель, которая объединяет техники машинного обучения и нечёткую логику. Этот подход использует анализ настроений новостей для прогнозирования тенденций на фондовом рынке.

3. Многозадачное обучение для анализа настроений и прогнозирования цен на акции. В статье "Multi-Task Learning for Stock Market Prediction with Textual and Tabular Data" (журнал IEEE Transactions on Knowledge and Data Engineering, 2022) [3], представлена многозадачная модель глубокого обучения. Эта модель одновременно предсказывает цены акций и анализирует настроения, извлечённые из новостных текстов при помощи LSTM–Forest модели.

4. Трансферное обучение с использованием предварительно обученных языковых моделей для анализа настроений. В исследовании "Transfer Learning for Stock Market Prediction with Pretrained Language Models" (журнал IEEE Transactions on Neural Networks and Learning Systems, 2021) [4], авторы использовали предварительно обученные языковые модели для анализа настроений в новостях. Затем они применили трансферное обучение, чтобы повысить точность прогнозирования временных рядов на финансовых рынках. Этот подход показывает, как можно перенести знания LLM-модели для улучшения результатов в другой (финансовые прогнозы).

Суммируя представленные подходы использования технологий машинного обучения для прогнозирования временных рядов на финансовых рынках:

1. Глубокие нейронные сети с анализом новостных текстов:

**Преимущество:** анализ тональности новостей позволяет учитывать внешние факторы для улучшения точности прогнозирования акций.

**Недостаток:** не всегда возможно полностью учесть контекст и сложные взаимосвязи между новостными событиями и рыночными тенденциями.

2. Гибридные модели, сочетающие машинное обучение и анализ эмоциональной окраски:

**Преимущество:** использование нечёткой логики и анализа настроений новостей может улучшить прогнозирование рыночных тенденций.

**Недостаток:** не всегда эмоциональная окраска новостей точно отражает будущие изменения на финансовых рынках.

3. Многозадачное обучение для анализа настроений и прогнозирования цен на акции:

**Преимущество:** многозадачное обучение позволяет одновременно учитывать различные аспекты (цены акций, настроения) для прогнозирования.

**Недостаток:** может потребоваться большее количество данных для обучения эффективной многозадачной модели.

4. Трансферное обучение с использованием предварительно обученных языковых моделей для анализа настроений:

**Преимущество:** использование предварительно обученных моделей и трансферного обучения может повысить точность прогнозирования.

**Недостаток:** необходимость в доступе к качественным и релевантным данным для обучения модели.

Недостатком всех этих работ является отсутствие подобного опыта и исследований в сфере агропромышленного комплекса (АПК). В АПК такие модели и подходы могут столкнуться с уникальными проблемами, такими как влияние природных факторов на урожайность, а также специфика производственных процессов. Это означает, что применение и адаптация данных техник для сельского хозяйства может потребовать дополнительных исследований и настройки под конкретные потребности и особенности этой отрасли.

### Формализация проблемы

Математическая постановка задачи прогнозирования временных рядов с учетом анализа внешних факторов может быть представлена следующим образом.

Пусть дан временной ряд:

$$y_t, t = 1, 2, \dots, T, \quad (1)$$

где  $y_t$  – значение исследуемой величины в момент времени  $t$ .

Также пусть дана последовательность кодированных факторов

$$x_t, t = 1, 2, \dots, T, \quad (2)$$

где  $x_t$  – значение фактора, соответствующей моменту времени  $t$ .

Задача состоит в том, чтобы на основе имеющихся данных  $y_t$  и  $x_t$  построить модель, которая позволяет прогнозировать значение временного ряда  $y_{T+h}$  в момент времени  $T+h$ , где  $h > 0$  – горизонт прогнозирования.

Математически это можно записать в виде:

$$y_{T+h} = f(y_1, y_2, \dots, y_T, x_1, x_2, \dots, x_T, h), \quad (3)$$

где  $f$  – некоторая функция, которую необходимо определить.

Разные факторы в разной степени влияют на рыночные показатели, такие как цена, рост спроса или предложения. Почти никакие события не происходят сами по себе - им всегда предшествуют события-первопричины. Например, рост производства мяса птицы не происходит одновременно: сперва заключаются инвестиционные соглашения на создание новых цехов птицефабрик или новых площадок или реконструкции старых птицефабрик. Обычно эти события освещаются в региональных сми или в собственных информационных ресурсах компаний: корпоративный блог, сайт, страница компании в соцсетях.

Каждый тип событий в той или иной степени влияет на макроэкономические показатели рынка. Ежедневно, в России и в мире происходят сотни таких событий (Табл. 1).

**Таблица 1.** Факторы, влияющие на экономические показатели в АПК.

Фактор	Оказываемое влияние
Стоимость корма	Цена корма для животных напрямую влияет на стоимость производства мяса.
Затраты на труд, Транспортные расходы, Затраты на обработку и упаковку	Зарплата работников ферм и мясоперерабатывающих заводов также влияет на конечную стоимость мяса. Расстояние от фермы до места продажи и стоимость транспортировки могут существенно влиять на цену. Процесс обработки и упаковки мяса для розничной продажи также добавляет к стоимости.
Правительственные регулирования	Субсидии, налоги, импортные пошлины и другие регулирования могут снижать или повышать стоимость мяса.
Сезонность	В некоторых регионах цены на мясо могут колебаться в зависимости от сезона.

Продолжение табл. 1

Эпидемии болезней	Вспышки болезней, такие как птичий грипп или свиной грипп, могут сократить предложение и увеличить цены.
Экологические факторы	Погодные условия, такие как засухи или наводнения, могут влиять на производство кормов и доступность пастбищ, что в свою очередь влияет на стоимость мяса.
Мнение экспертов	Влияние на потребительское восприятие - эксперты часто формируют общественное мнение о качестве и безопасности продуктов питания. Их заявления могут повлиять на предпочтения потребителей и, как следствие, на спрос. Экспертные оценки в области зоонозов и эпидемий могут привести к изменениям в производственных практиках, что также влияет на стоимость мяса. Экспертные оценки текущих и будущих тенденций в сельском хозяйстве, экономике и торговле могут влиять на инвестиционные решения и производственные планы, что, в свою очередь, влияет на предложение мяса и его стоимость.
Новости об оснащенности производств, отчетность компаний	Положительные новости о модернизации производств могут увеличить инвестиции в мясную промышленность, что способствует улучшению технологий и снижению затрат на производство. Современное оборудование может повысить эффективность производства, что приводит к увеличению предложения и потенциальному снижению цен для потребителей. Улучшенное оборудование может также влиять на качество мяса, что может привести к изменению его стоимости на рынке. Новости о внедрении новых технологий, обеспечивающих безопасность продукции, могут повысить доверие потребителей и спрос на мясо. Современные технологии могут снижать вредное воздействие на окружающую среду, что также может повлиять на стоимость мяса через экологические налоги и субсидии.
Кредитная политика	Льготные кредиты и низкие процентные ставки могут позволить производителям мяса инвестировать в современное оборудование, улучшать производственные процессы и повышать эффективность, что может привести к снижению себестоимости и, как следствие, цен для потребителей. Улучшенные условия кредитования могут стимулировать производителей на расширение производства, что увеличивает предложение мяса на рынке и может привести к снижению цен. Кредитная политика также влияет на фермеров, которые могут получить доступ к кредитам для покупки кормов, скота и оборудования, что в долгосрочной перспективе может повлиять на стоимость мяса. С другой стороны, жесткая кредитная политика может уменьшить доступность кредитов, что может привести к сокращению производства и увеличению цен на мясо из-за снижения предложения.
Форсмажорные события	Нарушение сложившейся системы производства/реализации продукции

**Новости, как факторы влияющие на ключевые параметры производства  
и стоимость продукции в сфере АПК**

В таблице представлены факторы и пояснение к ним (Табл. 2).

**Таблица 2.** Входные данные для кодирования факторов.

Столбец	Пояснение
id Записи (Primary KEY)	id Записи в базе данных
Дата	Дата новости
Фактор	Выбирается один из исследуемых факторов: <ol style="list-style-type: none"> <li>1. Эпизоотика</li> <li>2. Комментарии экспертов</li> <li>3. Расширение/модернизация/логистика – обращать внимание на сроки реализации проекта – от этого зависит период влияния</li> <li>4. Статистические показатели</li> <li>5. Госрегулирование</li> <li>6. Ситуация на зерновом/кормовом рынках</li> <li>7. Кредитная политика</li> <li>8. Форсмажорные события (СВО, кризисы и т.д.)</li> <li>9. Отчетность предприятий</li> </ol>
Заголовок новости	
Ссылка на новость	
Влияние фактора	Оценивается сила влияния фактора по дискретной шкале от 0 до 3, где 0 – нет влияния 1 – слабое 2 – среднее 3 – сильное влияние
Продолжительность	Оценивается сила влияния фактора по дискретной шкале от 1 до 5, где 1 – неделя 2 – месяц 3 – 3 месяца 4 – 6 месяцев 5 – год
Регион события	
Регион влияния	Локальный уровень Республиканский Государственный СНГ
К чему приведет событие (снижение/повышение цены)	1 – если к повышению -1 – к понижению

Выходные данные представлены файлом в формате csv. Применяются следующие способы кодирования признаков:

1. Бинарное кодирование (One-Hot Encoding):

– Описание: Преобразует категориальные признаки в набор из бинарных переменных. Каждой категории соответствует свой бинарный признак, который равен 1, если объект принадлежит этой категории, и 0 в противном случае.

– Применение Подходит для большинства моделей, особенно для линейных моделей и дерева решений.

2. Порядковое кодирование (Ordinal Encoding):

– Описание: Присваивает каждой категории числовое значение в порядке их частоты или в произвольном порядке. Это подразумевает наличие порядка между категориями, что не всегда справедливо.

– Применение Подходит для моделей, где важно сохранить порядок категорий, например, для регрессионных моделей, если категории действительно имеют естественный порядок.

### Прогнозирование с учетом праздников

В качестве примера попробуем описанные подходы применения машинного обучения и кодирования факторов были закодированы календарные праздники РФ как фактор и спрогнозирована цена на говяжью голяшку в Москве.

В качестве моделей выбраны:

1. Prophet – предназначенная для прогнозирования временных рядов на основе аддитивной модели, которая учитывает тренды, сезонность и праздники.

2. NeuralProphet – это модернизированная версия Prophet, основанная на глубоком обучении, которая сочетает классические подходы к временным рядам с мощностью рекуррентных нейронных сетей и механизмов внимания.

3. CatBoost – это эффективная реализация градиентного бустинга над решающими деревьями от Яндекса, оптимизированная для работы с категориальными и числовыми признаками и предназначенная для различных задач машинного обучения, включая прогнозирование временных рядов.

Они показали следующие результаты (Табл. 3, Рис. 1):

**Таблица 3.** Оценка моделей.

Модель	Средняя абсолютная ошибка, руб	Средняя относительная ошибка, %
Prophet	88.91	20.33
NeuralProphet	17.54	4.34
CatBoost	35.54	8.20

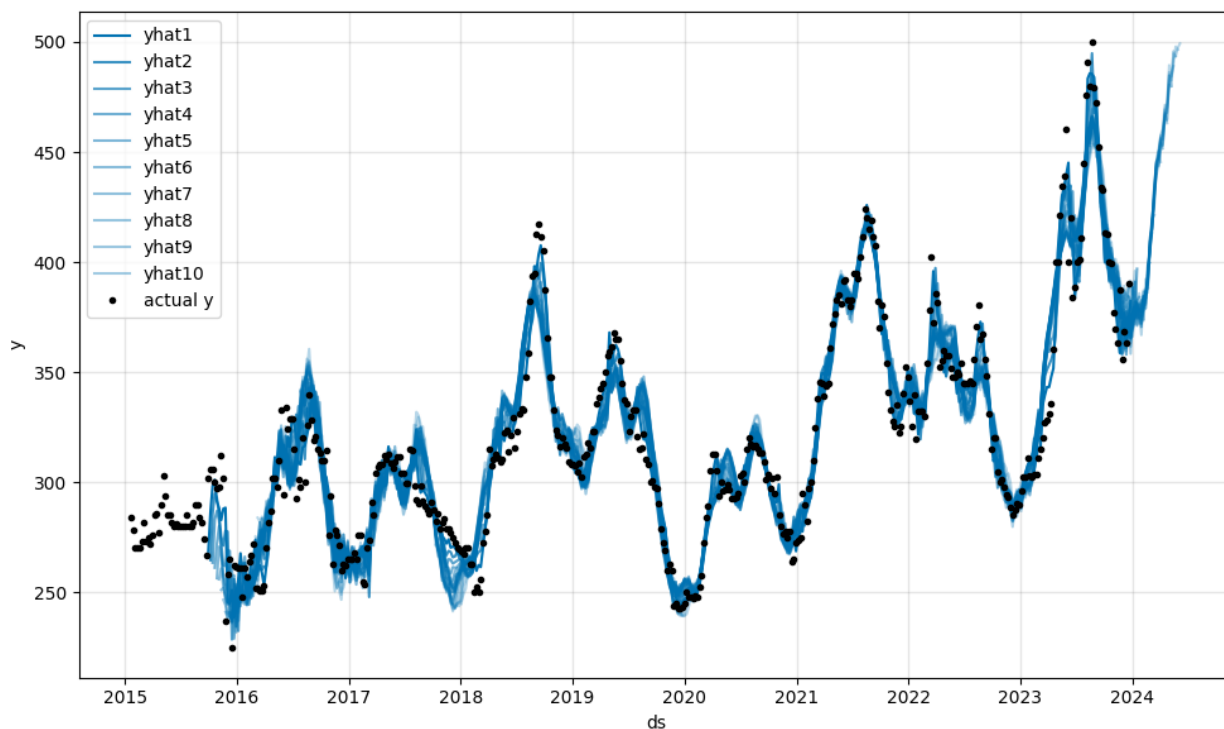


Рисунок 1. Результат прогноза Neural Prophet.

### Аналитическое заключение

Сравнительный анализ трех моделей показал, что NeuralProphet является наиболее эффективным инструментом для прогнозирования цен на говяжью голяшку с учетом праздников, обеспечивая наименьшую среднюю абсолютную и относительную ошибку. Это подчеркивает важность использования методов глубокого обучения в задачах, связанных с временными рядами, особенно когда данные имеют сложные паттерны и требуют учета множества факторов.

Prophet, хотя и является хорошим инструментом для базового прогнозирования, может быть недостаточно эффективен для более сложных задач, где необходимо учитывать взаимодействие различных факторов. CatBoost, в свою очередь, также показал достойные результаты, однако его эффективность может быть улучшена при более глубоком анализе признаков и параметров модели.

### Заключение

Можно сформулировать основные выводы.

1. Использование внешних факторов, таких как стоимость кормов, затраты на труд, правительственные регулирования, и другие, позволяет значительно улучшить точность прогнозов. Эти факторы предоставляют дополнительную информацию, которая помогает моделям более точно отражать реальные экономические процессы и реагировать на изменения в окружающей среде и рыночной ситуации.

2. Модели машинного обучения, такие как NeuralProphet и CatBoost, демонстрируют высокую эффективность в задачах прогнозирования временных рядов благодаря своей способности обрабатывать большие объемы данных и интегрировать различные типы данных. NeuralProphet особенно эффективен благодаря использованию рекуррентных нейронных сетей и механизмов внимания, которые позволяют учитывать долгосрочные зависимости и изменения в данных.

3. Одним из ключевых аспектов успешного прогнозирования является правильная обработка и трансформация входных данных. Методы, такие как One-Hot Encoding и Ordinal Encoding, позволяют преобразовать категориальные и текстовые данные в формат, пригодный

для обработки моделями машинного обучения. Это обеспечивает возможность использовать полный спектр доступной информации для улучшения качества прогнозов.

4. Включение информации о праздниках и сезонных колебаниях позволяет дополнительно уточнить модели, делая их более чувствительными к периодическим изменениям в данных. Это особенно важно в сфере АПК, где многие процессы имеют выраженную сезонную природу.

В целом прогнозирование временных рядов с учетом внешних факторов является важным и актуальным направлением исследований в области анализа данных и предиктивной аналитики.

#### **Литература:**

1. Zhang H., Li Y., Wang X. Improving Stock Market Prediction with News Sentiment Analysis Using Attention-Based Recurrent Neural Networks // *IEEE Access*. 2021. Vol. 9. P. 47701–47712.
2. Kumar R., Singh P., Gupta S. Hybrid Approach for Stock Market Prediction Using News Sentiment Analysis and Fuzzy Logic // *IEEE Access*. 2020. Vol. 8. P. 189024–189035.
3. Chen L., Zhao X., Liu Y. Multi-Task Learning for Stock Market Prediction with Textual and Tabular Data // *IEEE Transactions on Knowledge and Data Engineering*. 2022. Vol. 34. № 7. P. 3421–3432.
4. Wang T., Zhou K., Hu J. Transfer Learning for Stock Market Prediction with Pretrained Language Models // *IEEE Transactions on Neural Networks and Learning Systems*. 2021. Vol. 32. № 10. P. 4705–4716.

#### **Об авторах:**

**ИБАТУЛЛИН Марат Русланович**, аспирант кафедры информатики ФГБОУ ВО «УУНиТ»; ibatullin.m@gmail.com

**КОВТУНЕНКО Алексей Сергеевич**, к.т.н., доцент, доцент кафедры информатики ФГБОУ ВО «УУНиТ»; askovtunenکو@mail.ru

#### **Metadata:**

**Title:** Machine learning in Time Series Forecasting Taking into Account the Influence of External Factors in the Agro-Industrial Area

**Author 1:** Ibatullin Marat Ruslanovich, postgraduate student, Department of Computer Science, UUST.

**Author 2:** Kovtunenکو Alexey Sergeevich, Cand. Sci., Associate Professor, Department of Computer Science, UUST, <https://orcid.org/0000-0002-3814-7310>, ААН-5198-2019, 57204176565.

**Abstract:** The work is devoted to the development and application of time series forecasting methods in agriculture, taking into account external factors. Modern approaches based on deep learning (Neural Prophet) and gradient boosting (Cat Boost) are analyzed to improve the accuracy of forecasting prices for agricultural products. The study includes an analysis of the impact of various factors such as feed costs, government regulations, news and seasonality, as well as the development of a strategy for encoding categorical features (One-Hot Encoding, Ordinal Encoding). The results demonstrate a significant improvement in the accuracy of forecasts compared to traditional methods, emphasizing the importance of integrating external data for making effective management decisions in the agro-industrial complex.

**Keywords:** Time series forecasting, agro-industrial complex, machine learning, NeuralProphet, Cat-Boost, external factors, news analysis, feature coding