

УДК 330.4, 519.8, 004.94

1.4. Проектирование информационно-аналитической системы прогнозирования индекса потребительских цен в Российской Федерации: архитектура, методы и инструменты

Тодуа А.Н., Магистрант, РЭУ им. Г.В. Плеханова
Китов В.А., доцент кафедры информатики, к.т.н., РЭУ им. Г.В. Плеханова

Статья посвящена разработке концепции информационно-аналитической системы (ИАС) для прогнозирования индекса потребительских цен (ИПЦ) в Российской Федерации. В рамках исследования предложена архитектура ИАС, определены функциональные и технические требования, а также выбран стек технологий для её реализации. Проведён сравнительный анализ методов прогнозирования временных рядов — от классических статистических до современных методов машинного обучения и нейросетевых моделей. Оценены возможности различных аналитических платформ, таких как Microsoft Excel, Яндекс DataLens, Foresight Analytics Platform и OptiMA, в контексте решения задач прогнозирования ИПЦ. Сформулированы рекомендации по выбору модели и программной реализации системы в зависимости от характеристик данных и доступных ресурсов.

Введение

Стабильность цен является ключевым фактором макроэкономического регулирования. Индекс потребительских цен (ИПЦ) служит важнейшим индикатором уровня инфляции, влияя на принятие решений в государственном и корпоративном управлении. В условиях цифровой трансформации экономики актуальной становится задача построения информационно-аналитических систем (ИАС), способных не только собирать и обрабатывать данные о ценах, но и осуществлять их прогнозирование с использованием современных математических моделей и программных решений. Существующие подходы к прогнозированию ИПЦ варьируются от простых статистических моделей (например, ARIMA) до сложных алгоритмов машинного и глубокого обучения (LSTM, XGBoost и др.). Вместе с тем существует потребность в комплексной системе, объединяющей обработку данных, прогнозную аналитику и визуализацию, ориентированной на прикладные задачи в органах статистики, исследовательских центрах и экономических департаментах. Цель статьи — разработать архитектуру ИАС для прогнозирования ИПЦ в РФ, определить стек технологий, а также провести сравнительный анализ применимых методов и инструментов с позиции их эффективности, точности и реализуемости в рамках выбранной архитектуры.

Задачи исследования

1. Проанализировать особенности динамики и источников данных по ИПЦ в РФ.
2. Классифицировать методы прогнозирования временных рядов, применимые к задаче анализа ИПЦ.
3. Провести обзор современных инструментов визуализации и обработки данных, применимых в ИАС.
4. Разработать архитектуру ИАС для решения задачи прогнозирования ИПЦ.
5. Определить функциональные и нефункциональные требования к системе.
6. Провести сравнительный анализ применимых моделей прогнозирования (ARIMA, XGBoost, LSTM и др.)
7. Сформулировать рекомендации по выбору методов и технологического стека для решения задач прогнозирования ИПЦ.

Методика исследования

1. Анализ предметной области и постановка задачи

Индекс потребительских цен (ИПЦ) представляет собой ключевой показатель инфляции и один из важнейших макроэкономических индикаторов, используемых при анализе потребительского сектора экономики. Он отражает изменение совокупного уровня цен на товары и услуги, входящие в потребительскую корзину, и используется для оценки покупательной способности национальной валюты, мониторинга ценовой стабильности и выработки социальной политики. В методологическом плане ИПЦ представляет собой агрегированный индекс, рассчитываемый на основе фиксированных весов и цен более чем 500 позиций, собираемых в территориально распределённой выборке. В России расчет ИПЦ осуществляется Федеральной службой государственной статистики по единой методике, соответствующей международным стандартам [1].

Методология расчёта базируется на определении отношения стоимости фиксированного набора товаров и услуг в текущих ценах к его стоимости в базисном периоде.

Базовая формула расчёта ИПЦ [3]:

$$\text{ИПЦ} = \frac{\sum p_t q_0}{\sum p_0 q_0} \cdot 100\%,$$

где p_t – цены текущего периода;

p_0 – цены базисного периода;

q_0 – количество товаров и услуг в базисном периоде.

Практическое значение ИПЦ заключается в его широкой применимости в сфере государственного и корпоративного регулирования. В частности, Банк России ориентируется на ИПЦ при реализации политики инфляционного таргетирования и установлении ключевой процентной ставки, оказывающей влияние на стоимость кредитования и динамику сбережений населения [2]. Кроме того, на основе ИПЦ осуществляется индексация социальных выплат, тарифов и налоговых ставок. В условиях высокой чувствительности индекса к сезонным, внешнеэкономическим и геополитическим колебаниям возникает необходимость в его регулярном прогнозировании. Это особенно важно в периоды экономической нестабильности, когда устойчивость потребительских цен служит критерием эффективности государственной макроэкономической политики.

В этой связи особую значимость приобретает задача построения краткосрочного прогноза индекса потребительских цен, особенно на месячном горизонте. Решение данной задачи необходимо для обеспечения своевременной адаптации налогово-бюджетной, денежно-кредитной и социальной политики, а также для целей макроэкономического и бюджетного планирования. Согласно методологическим положениям официальной статистики, утверждённым приказом Росстата № 915 от 15 декабря 2021 года, ИПЦ представляет собой агрегированный показатель, рассчитываемый на основе репрезентативной выборки потребительских товаров и услуг, и отражает изменение совокупного уровня цен во времени по отношению к базовому периоду [3]. Учитывая высокую чувствительность ИПЦ к нерегулярным и внешним шокам, а также его выраженную сезонность и краткосрочные флуктуации, построение прогноза требует применения адаптивных моделей, способных учитывать скрытые тренды, устойчивых к шумам и подходящих для автоматизированного внедрения в информационно-аналитические системы (ИАС). В связи с этим в рамках настоящего исследования ставится задача разработки методики краткосрочного прогнозирования ИПЦ на основе интерпретируемой и ресурсно-эффективной модели, обеспечивающей высокую точность в условиях ограниченных вычислительных ресурсов и нестабильности наблюдаемых данных.

2. Источники и структура данных

В качестве основного источника статистических данных для анализа использовались официальные ежемесячные и еженедельные данные Федеральной службы государственной статистики (Росстат) по индексу потребительских цен (ИПЦ) на товары и услуги. Согласно приказу Росстата № 727 от 18.12.2025 в 2026 году утверждено 558 товаров и услуг для ежемесячного наблюдения за ценами и тарифами, а для еженедельного отчета установлено 110 товаров и услуг для расчета индекса потребительских цен [4]. Источниками официальных статистических данных Росстата являются: Единая межведомственная информационно-статистическая система (ЕМИСС) [5], витрина данных, где публикуются массивы в формате CSV и JSON [6], а также BI-инструменты Росстата, позволяющие строить графики, диаграммы и интерактивные панели по различным индикаторам, включая ИПЦ. В качестве дополнительных регрессоров используются макроэкономические индикаторы: курсы валют (прежде всего USD и EUR), ключевая ставка Центрального банка РФ, мировые цены на нефть (например, Brent), а также международные индексы цен, такие как CPI США или глобальные продовольственные индексы FAO. Все данные хранятся в формате временных рядов в СУБД PostgreSQL, с возможностью альтернативного представления в виде JSON-структур для API-интеграции, а также экспортируются в CSV-формате для последующей загрузки в BI-системы.

3. Обоснование выбора методов для прогнозирования

Современные подходы к прогнозированию индекса потребительских цен (ИПЦ) опираются как на традиционные эконометрические методы, так и на алгоритмы машинного обучения. Классические модели, включая авторегрессию (AR), ARIMA и SARIMA, продолжают широко использоваться в аналитике макроэкономических показателей, благодаря их формальной интерпретируемости и устоявшейся методологии [7]. Однако их применимость ограничивается линейными взаимосвязями и требует строгих предпосылок о стационарности временного ряда.

В то же время методы машинного обучения, такие как случайный лес, градиентный бустинг (XGBoost, LightGBM) и CatBoost, демонстрируют высокую точность на табличных и высокоизмеримых данных. Согласно исследованию Банка России, XGBoost позволил достичь более высокой точности прогнозирования регионального ИПЦ по сравнению с ARIMA-моделями, особенно при наличии сложных взаимосвязей между переменными [8]. Байбуза также показывает, что использование Ridge, LASSO, а также ансамблевых методов в задачах инфляционного прогнозирования обеспечивает устойчивые результаты при высокой доле регрессоров.

В ряде российских публикаций обоснована применимость и CatBoost — особенно для задач, где признаки представлены в категориальной форме (например, товарные группы, регионы, типы услуг). Отмечается потенциал использования нейросетевых моделей (в частности, LSTM) в случае наличия больших объёмов исторических данных и необходимости учёта сезонных эффектов [9].

4. Определение метрик качества

С целью верификации применимости различных подходов к прогнозированию индекса потребительских цен (ИПЦ) была проведена эмпирическая оценка точности четырёх моделей: линейной регрессии, случайного леса (Random Forest), метода опорных векторов (SVR) и ARIMA. Анализ выполнен на основе авторского датасета, включающего значения ИПЦ в ежемесячной разбивке, где каждый период представляет собой относительное изменение показателя по сравнению с предыдущим месяцем. Для упрощения формализации временная переменная была закодирована численно в виде последовательности от 1 до 20.

В ходе расчётов выборка была разделена на обучающую (70%) и тестовую (30%) подвыборки. Качество моделей оценивалось по трём метрикам: MAE (средняя абсолютная ошибка), MAPE (средняя абсолютная процентная ошибка), и R^2 (коэффициент детерминации). Результаты представлены в сравнительной таблице 2

Таблица 2. Сравнительная оценка моделей

Model	MAE	MAPE	R^2
Linear Regression	0.2042	0.0020	-0.1487
Random Forest	0.1928	0.0019	-0.0661
SVR	0.2891	0.0029	-1.6038
ARIMA	0.2997	0.0030	0.2044

Источник: составлено автором

Наиболее низкие значения MAE и MAPE продемонстрировала модель Random Forest, что указывает на её преимущество в контексте малых абсолютных и относительных отклонений от фактических значений. В то же время, наилучшее значение коэффициента детерминации R^2 показала модель ARIMA, единственная продемонстрировавшая положительное значение этой метрики, что свидетельствует о более высокой доле объяснённой вариации зависимой переменной. Негативные значения R^2 у остальных моделей указывают на слабую способность данных моделей воспроизводить динамику ИПЦ на отложенной выборке. Особенно низкие показатели по всем метрикам продемонстрировал SVR, что делает его наименее эффективным в текущем контексте.

Исходя из данного анализа было определено, что с точки зрения абсолютной точности прогнозов оптимальным выбором является метод Random Forest, в то время как ARIMA может рассматриваться как подход, обеспечивающий лучшее соответствие общей трендовой структуре временного ряда.

5. Анализ инструментов визуализации

С целью выбора подходящей платформы для визуального представления результатов прогнозирования индекса потребительских цен был проведён сравнительный анализ четырёх BI-инструментов, наиболее широко применяемых в практике прикладной аналитики. Основное внимание было уделено следующим критериям: уровень поддержки формата CSV, наличие инструментов интерактивной визуализации, возможности отображения прогнозных данных и условия локализации с точки зрения стоимости владения. Результаты анализа представлены в 1.

Таблица 1 - Сравнительная оценка BI-платформ

Критерий	Microsoft Excel	Power BI	Foresight Analytics Platform	Яндекс DataLens
Интеграция с CSV	Поддерживается напрямую, ограничена по объёму и стабильности	Осуществляется через Power Query, требует предварительной настройки	Импорт возможен, предпочтение отдаётся корпоративным хранилищам	Полная поддержка, реализована через мастер загрузки
Интерактивные дашборды	Ограничены стандартными таблицами и графиками	Поддерживаются, включают фильтры, связи и визуальные компоненты	Поддерживаются, имеются встроенные панели мониторинга	Поддерживаются, возможна гибкая настройка визуальных элементов
Визуализация прогнозов	Линейная аппроксимация и скользящее среднее через встроенные функции	Требует подключения внешних моделей или использования DAX-выражений	Встроенные прогнозные модули, доступны в базовой конфигурации	Реализуется средствами SQL и фильтрации временных рядов
Локализация и стоимость	Полная локализация, входит в состав офисного пакета	Частичная локализация, базовый функционал доступен бесплатно	Полная локализация, лицензируемое решение	Полная локализация, бесплатный доступ к основному функционалу

Результаты сравнения показывают, что Microsoft Excel, несмотря на распространённость, демонстрирует ограниченную гибкость при работе с прогнозными данными и дашбордами. Power BI предлагает расширенный функционал, однако требует дополнительной настройки для интеграции CSV и частично ограничен в локализации. Foresight Analytics Platform ориентирован на корпоративные внедрения, обладает встроенными средствами прогнозирования, но предполагает лицензирование. Оптимальное соотношение простоты интеграции, доступности, локализации и функциональности продемонстрировал Яндекс DataLens, что позволяет рассматривать его в качестве предпочтительного инструмента для визуального слоя создаваемой информационно-аналитической системы.

6. Архитектура ИАС и требования

В рамках исследования была разработана модульная архитектура информационно-аналитической системы (ИАС), предназначенной для прогнозирования индекса потребительских цен (ИПЦ) и представлена на Рисунке 1.

Архитектура ИАС для прогнозирования ИПЦ

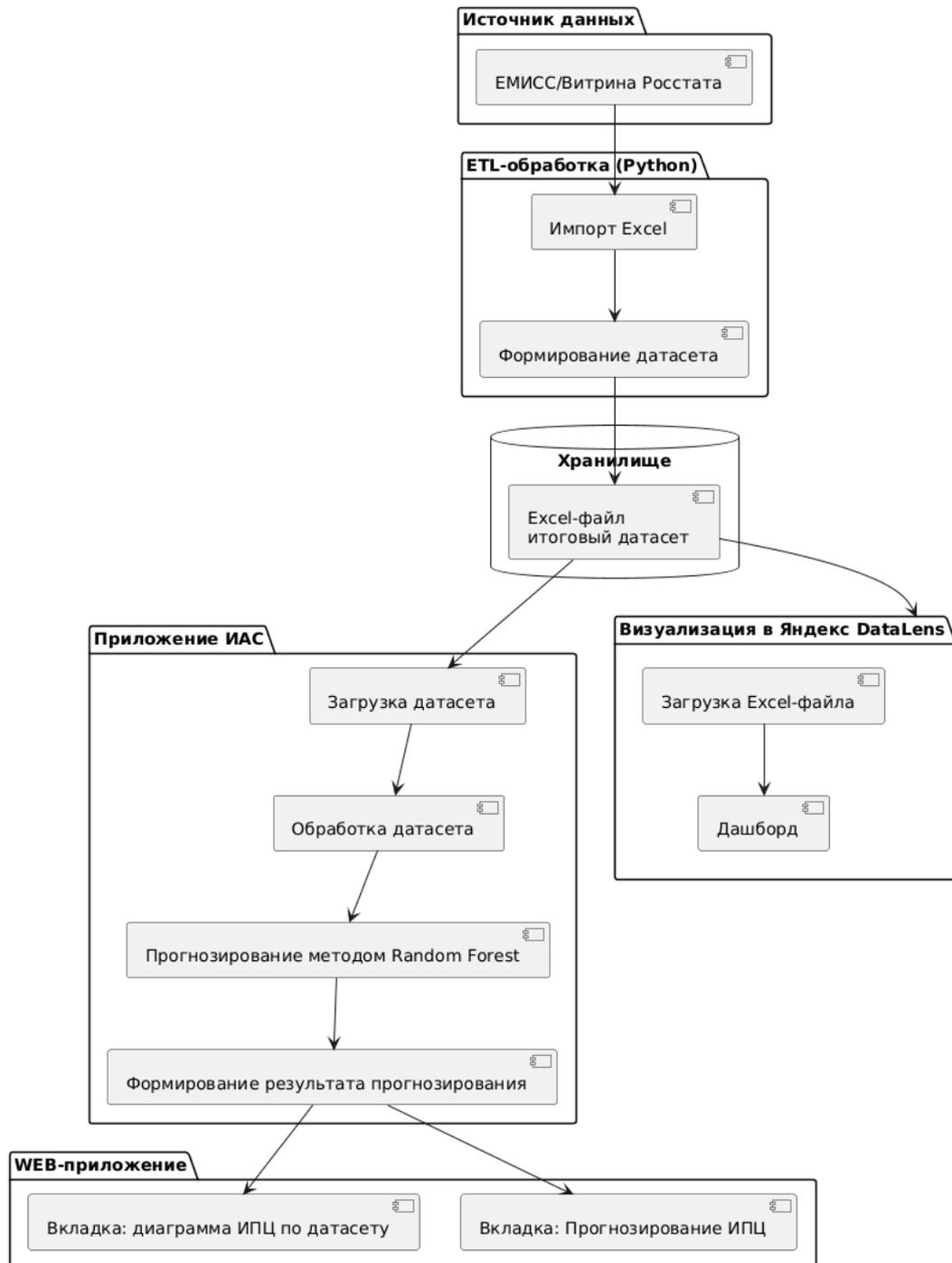


Рисунок 1 - Архитектура информационно-аналитической системы

В качестве основного источника данных используется витрина Росстата (ЕМИСС), откуда ежемесячные значения ИПЦ выгружаются в формате Excel. Далее реализуется этап ETL-обработки с использованием Python-скриптов, в ходе которого производится очистка, нормализация и формирование итогового датасета. Готовый файл служит входным для аналитического модуля и системы визуализации.

Аналитическое ядро системы построено на модели Random Forest, показавшей наилучшие результаты по метрикам MAE и MAPE. Прогнозирование осуществляется с горизонтом в 1–3 месяца, после чего данные визуализируются в веб-интерфейсе на двух вкладках – «Динамика ИПЦ» и «Прогнозирование». Дополнительно обеспечена интеграция с Яндекс DataLens, который подключается к итоговому Excel-файлу и предоставляет дашборд для мониторинга ключевых макроэкономических показателей.

К функциональным требованиям системы относятся автоматическое обновление данных, построение краткосрочного прогноза, визуализация фактических и прогнозных значений, а также формирование отчётов. Нефункциональные требования включают точность прогноза на уровне MAPE < 10%, производительность (время генерации прогноза не более 3 секунд), поддержку интеграции через Excel и API, а также разграничение прав доступа

Результаты исследования

В рамках данного исследования разработана модульная архитектура ИАС, сочетающая гибкость Python-аналитики с визуализацией через BI-инструменты. Система реализует процесс ETL-обработки, прогнозирование методом Random Forest и визуализацию через Яндекс DataLens, обеспечивая автоматизацию сбора, обработки и анализа данных ИПЦ.

Сравнительный анализ методов прогнозирования был проведён на основе трёх метрик: MAE (средняя абсолютная ошибка), MAPE (средняя абсолютная процентная ошибка) и R^2 (коэффициент детерминации). Наилучшие результаты продемонстрировала модель Random Forest с MAE = 0.1928, MAPE = 0.0019 и $R^2 = -0.0661$. Модель Linear Regression показала MAE = 0.2042, MAPE = 0.0020 и $R^2 = -0.1487$, а SVR: MAE = 0.2891, MAPE = 0.0029 и $R^2 = -1.6038$. Модель ARIMA продемонстрировала худшие MAE (0.2997) и MAPE (0.0030), несмотря на положительное значение $R^2 = 0.2044$. По совокупности показателей наиболее точной является модель Random Forest.

На основании сравнительного анализа BI-платформ было выбрано решение Яндекс DataLens, поскольку оно сочетает удобную работу с Excel-источниками, интуитивный интерфейс и возможность встраивания дашбордов в веб-приложения. В отличие от конкурентов, DataLens обеспечивает быструю визуализацию без установки дополнительного ПО, а также гибкость фильтрации и интерактивности, это критически важно для построения масштабируемой ИАС с ограниченным бюджетом.

Критерии выбора методов и инструментов включали в себя следующие основные признаки и возможности: наличие готовых реализаций в Python, возможность оценки точности прогноза с помощью метрик, устойчивость к сезонности, а также интеграцию с архитектурой ИАС. Выбор Random Forest объясняется его устойчивостью к шуму, высокой точностью и эффективностью при работе с ограниченным числом признаков.

Обсуждение и заключение

Значимость полученных результатов заключается в систематизации и классификации методов прогнозирования временных рядов, релевантных для анализа динамики ИПЦ. В работе проведён сравнительный анализ BI-инструментов с акцентом на удобство визуализации, интеграцию с Excel и гибкость настройки, что позволило обоснованно выбрать Яндекс DataLens. Разработана архитектура информационно-аналитической системы, ориентированной на автоматизацию процесса прогнозирования ИПЦ, включая обоснованный выбор технологий, моделей машинного обучения и среды визуализации. Итогом исследования стали практические рекомендации по подбору инструментов и методов в зависимости от характеристик данных и аналитических задач.

В отличие от узкоспециализированных исследований, направленных исключительно на точность прогнозной модели, данная работа предлагает комплексный подход: от выбора источников данных и построения архитектуры ИАС до сравнения аналитических методов и визуальных платформ. Статья фокусируется на прикладной задаче в государственном управлении и демонстрирует интеграцию моделей в полноценную систему поддержки решений.

Исследование опирается на открытые исторические данные без учёта региональной детализации и влияния нестандартизированных макрофакторов, что ограничивает точность прогнозов. Все тестирования проводились в локальной среде, без масштабируемых вычислений. Модели не учитывали внешние шоки, такие как санкции или изменения в государственной экономической политике, что может снизить их применимость в условиях нестабильности.

Дальнейшее развитие ИАС предполагает тестирование современных нейросетевых моделей и AutoML-платформ, на расширенных и более разнообразных наборах данных. Перспективным направлением является добавление новых функций: сценарного анализа, регионального прогнозирования и детализации по товарным группам. Важной задачей остаётся количественная оценка экономического эффекта от внедрения ИАС в практику принятия решений.

Литература

1. Как Росстат считает инфляцию? // Официальный сайт Росстата URL: <https://ps.rosstat.gov.ru/inflation> (дата обращения: 17.01.2026).
2. Что такое инфляция // Банк России URL: https://www.cbr.ru/dkp/about_inflation (дата обращения: 17.01.2026).

3. Приказ «Об утверждении Официальной статистической методологии наблюдения за потребительскими ценами на товары и услуги и расчета индексов потребительских цен» от 15.12.2021 № 915 // Официальный сайт Росстата. – 2021
4. Приказ «Об утверждении наборов потребительских товаров и услуг и перечня базовых городов Российской Федерации для наблюдения за ценами и тарифами» от 18.12.2025 № 727 // Официальный сайт Росстата. – 2025
5. Официальные статистические показатели // ЕМИСС URL: <https://www.fedstat.ru/> (дата обращения: 17.01.2026).
6. Витрина статистических данных URL: <https://showdata.rosstat.gov.ru/finder/> (дата обращения: 17.01.2026).
7. Букина Т.В., Кашин Д.В. Прогнозирование региональной инфляции: эконометрические модели или методы машинного обучения? // Экономический журнал Высшей школы экономики. – 2024. – №1 – Том 28. – С. 81-107.
8. Применение методов машинного обучения в задаче прогнозирования региональной инфляции на примере Ставропольского края // Банк России URL: https://www.cbr.ru/ec_research/ser/wp_15_/ (дата обращения: 17.01.2026).
9. Альчаков В.В., Крамарь В.А. ОЦЕНКА МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ ДЛЯ ПРОГНОЗИРОВАНИЯ СЕЗОННЫХ ВРЕМЕННЫХ РЯДОВ // ИЗВЕСТИЯ ЮФУ. ТЕХНИЧЕСКИЕ НАУКИ. – 2023. – №2 (232). – С. 250-263.

References in Cyrillics

1. Bukina T.V., Kashin D.V. Prognozirovanie regional'noj inflyacii: e'konometricheskie modeli ili metody mashinnogo obucheniya? // E'konomicheskij zhurnal Vy'sshej shkoly e'konomiki. – 2024. – №1 – Том 28. – С. 81-107.
2. Al'chakov V.V., Kramar' V.A. OCENKA METODOV MASHINNOGO OBUChENIYa DLYa PROGNOZIROVANIYa SEZONNY'X VREMENNY'X RYaDOV // IZVESTIYa YuFU. TEXNICHESKIE NAUKI. – 2023. – №2 (232). – С. 250-263..

Ключевые слова:

индекс потребительских цен, прогнозирование, информационно-аналитическая система, временные ряды, ARIMA, LSTM, BI-инструменты, архитектура ИАС.

Тодуа Алекс Николаевич

магистрант, РЭУ им. Г.В. Плеханова (115054, Россия, Москва, Стремянный переулок 36),
ORCID: 0009-0003-7650-9663, todua.alex@gmail.com

Китов Владимир Анатольевич

кандидат технических наук, доцент кафедры информатики РЭУ им. Г.В. Плеханова (115054,
Россия, Москва, Стремянный переулок 36),
ORCID: 0000-0002-4821-779X, Kitov.VA@rea.ru

Aleks N. Todua, Vladimir A. Kitov. Design of an information and analytical system for forecasting the consumer price index in the Russian Federation: architecture, methods and tools.

Keywords

consumer price index, forecasting, information and analytical system, time series, ARIMA, LSTM, BI tools, IAS architecture

DOI: 10.34706/DE-2026-01-04

JEL classification M15 Деловое администрирование: управление информационными технологиями

Abstract

This article explores the development of a concept for an information and analytical system (IAS) for forecasting the consumer price index (CPI) in the Russian Federation. The study proposes IAS architecture, defines its functional and technical requirements, and selects a technology stack for its implementation. A comparative analysis of time series forecasting methods, from classical statistical approaches to modern machine learning and neural network models, is conducted. The capabilities of various analytical platforms, such as Microsoft Excel, Yandex DataLens, Foresight Analytics Platform, and OptiMA, are assessed for solving CPI forecasting problems. Recommendations are provided for selecting a model and software implementation based on data characteristics and available resources.