

УДК 330.43

<http://doi.org/10.35854/1998-1627-2026-4-506-514>

Предсказательное моделирование социально-экономических процессов на основе интеллектуального анализа данных

Сергей Александрович Игнатьев^{1✉}, Кирилл Александрович Буслов²,
Александр Александрович Ткаченко³

^{1, 2, 3} Институт проблем передачи информации имени А. А. Харкевича Российской академии наук, Москва, Россия

¹ ignatev-sa@iitp.ru[✉], <https://orcid.org/0009-0002-0450-1913>

² buslov-ka@iitp.ru, <https://orcid.org/0009-0008-6175-3774>

³ tkachenko-aa@iitp.ru, <https://orcid.org/0009-0003-4751-7568>

Аннотация

Цель. Сформировать воспроизводимую методологию предсказательного моделирования социально-экономических процессов на основе интеллектуального анализа данных, пригодную для академических исследований и практического внедрения в контуры управления.

Задачи. Обосновать переход от «узких» эконометрических спецификаций к гибридным схемам, сочетающим эконометрику, регуляризацию и методы машинного обучения / глубокого обучения; охарактеризовать протокол построения прогностического конвейера для временных рядов с учетом ограничений реальных данных; задать правила корректного сравнения моделей на временной оси и предотвращения утечек данных; раскрыть роль интерпретируемости и мониторинга устойчивости для социально-экономических приложений; показать методологию на репликациях прогнозирования инфляции.

Методология. Исследование построено как воспроизводимый modeling pipeline: согласование частот и календаря, предобработка и инженерия признаков (лаги, разности, агрегаты), обучение и настройка моделей (классические временные ряды, регуляризованные регрессии, деревья и ансамбли, нейросети), time-aware валидация (pseudo-out-of-sample, rolling/expanding windows), оценка качества (MAE, RMSE), интерпретация результатов (SHAP/LIME), а также процедуры мониторинга деградации и дрейфа.

Результаты. Показано, что качество прогнозов определено не только выбором алгоритма, но и информационным пространством факторов и корректностью протокола оценки. В постановках data-rich современные ML/DL-модели способны давать существенное снижение ошибок на коротких горизонтах. Однако эффект неоднороден по горизонтам и конфигурациям, а отдельные ансамблевые модели могут не превосходить простые бенчмарки без строгой настройки и проверки. Методы динамического усреднения/выбора моделей демонстрируют преимущества точно и не гарантируют систематического доминирования.

Выводы. Предсказательное моделирование социально-экономических процессов требует строгого воспроизводимого протокола: временно-корректной валидации, контроля утечек, сопоставимых метрик и обязательной интерпретируемости. Выигрыш современных методов является режимно-зависимым и должен подтверждаться псевдовыборочным тестированием и стресс-проверками; внедрение моделей должно сопровождаться мониторингом устойчивости и управлением рисками применения.

Ключевые слова: предсказательное моделирование, социально-экономические процессы, интеллектуальный анализ данных, машинное обучение, временные ряды, интерпретируемость

© Игнатьев С. А., Буслов К. А., Ткаченко А. А., 2026

Данная статья является статьей открытого доступа, распространяемой в соответствии с условиями лицензии Creative Commons Attribution 4.0 License (CC BY 4.0) (<https://creativecommons.org/licenses/by/4.0/>).

Predictive modeling of socioeconomic processes based on data mining

Sergei A. Ignatev^{1✉}, Kirill A. Buslov², Alexander A. Tkachenko³

^{1, 2, 3} Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute), Moscow, Russia

¹ ignatev-sa@iitp.ru✉, <https://orcid.org/0009-0002-0450-1913>

² buslov-ka@iitp.ru, <https://orcid.org/0009-0008-6175-3774>

³ tkachenko-aa@iitp.ru, <https://orcid.org/0009-0003-4751-7568>

Abstract

Aim. This study aimed to develop a reproducible methodology for predictive modeling of socioeconomic processes based on data mining, which is suitable for academic research and practical implementation within management frameworks.

Objectives. To justify the transition from narrow econometric specifications to hybrid schemes combining econometrics, regularization, and machine learning/deep learning methods; to characterize a protocol for constructing a predictive pipeline for time series given real-world data constraints; to establish rules for the correct comparison of models along the temporal axis and for preventing data leakage; and to elucidate the role of interpretability and stability monitoring for socioeconomic applications; to demonstrate the methodology using replicable inflation forecasting examples.

Methods. The study is structured as a reproducible modeling pipeline: frequency and calendar alignment, preprocessing and feature engineering (lags, differences, aggregates), model training and tuning (classical time series models, regularized regressions, trees and ensembles, neural networks), time-aware validation (pseudo-out-of-sample, rolling/expanding windows), quality assessment (MAE, RMSE), result interpretation (SHAP/LIME), as well as procedures for monitoring degradation and drift.

Results. Forecast quality is determined not only by the choice of algorithm but also by the informational space of predictors and the correctness of the evaluation protocol. In data-rich settings, modern ML/DL models can achieve substantial error reduction over short horizons. However, the effect is heterogeneous across horizons and configurations, and some ensemble models may not outperform simple benchmarks without rigorous tuning and validation. Dynamic model averaging/model selection methods demonstrate advantages in certain cases but do not guarantee systematic dominance.

Conclusion. Predictive modeling of socioeconomic processes requires a strict reproducible protocol: time-correct validation, leakage control, comparable metrics, and mandatory interpretability. The gains from modern methods are regime-dependent and must be confirmed through pseudo-out-of-sample testing and stress testing; model implementation should be accompanied by stability monitoring and risk management of application.

Keywords: predictive modeling, socioeconomic processes, data mining, machine learning, time series, interpretability

For citation: Ignatev S.A., Buslov K.A., Tkachenko A.A. Predictive modeling of socioeconomic processes based on data mining. *Ekonomika i upravlenie* = *Economics and Management*. 2026;32(4):506-514. (In Russ.). <http://doi.org/10.35854/1998-1627-2026-4-506-514>

Введение

Социально-экономические процессы отличаются сложной структурой динамики, многофакторностью и неоднородностью источников информации. Их развитие определяется одновременно краткосрочными рыночными колебаниями, институциональными решениями, внешними шоками и долгосрочны-

ми структурными трендами. Доступность данных ограничена временными лагами публикации, пересмотрами статистики и различиями в методологии расчета показателей. В подобных условиях прогнозирование не может быть рассмотрено как формальная процедура подбора уравнения; оно представляет собой сложную инженерно-аналитическую систему, в которой важны

не только модельные предпосылки, но и архитектура обработки данных, стратегия оценки качества и механизмы интерпретации результатов.

Интеллектуальный анализ данных в социально-экономическом контексте предполагает системное объединение процедур сбора, очистки, трансформации и моделирования информации. В отличие от традиционного эконометрического подхода, сосредоточенного преимущественно на спецификации модели, парадигма Data Mining акцентирует внимание на процессе извлечения закономерностей из больших и разнородных массивов данных, что особенно актуально для современной цифровой экономики [1]. В течение последних лет происходит постепенный переход от «узких» эконометрических моделей к гибридным системам, сочетающим классические методы временных рядов с алгоритмами машинного обучения и ансамблевыми подходами.

Однако эмпирические исследования показывают, что устойчивое превосходство сложных моделей над простыми бенчмарками достигается далеко не всегда. В области макроэкономического прогнозирования, в частности при анализе инфляции, авторегрессионные модели нередко демонстрируют сопоставимое или даже лучшее качество на отдельных горизонтах. Это обстоятельство указывает на необходимость строгого методологического сравнения и осторожности в интерпретации преимуществ современных алгоритмов.

Одновременно исследования в среде data-rich подтверждают, что при расширении пространства факторов и корректной регуляризации модели машинного обучения способны выявлять нелинейности и сложные взаимодействия, что приводит к приросту прогностической точности. Машинное обучение рассматривают как инструмент предсказания, а не как альтернативу экономической теории или причинному анализу.

Цель настоящей статьи заключается в формировании методологически целостной схемы предсказательного моделирования социально-экономических процессов, объединяющей эконометрические методы, алгоритмы машинного обучения, инструменты интерпретируемости и процедуры управления рисками применения моделей.

Методология

Методология исследования построена как воспроизводимый аналитический конвейер (pipeline), направленный на прогнозирование показателя y_{t+h} , то есть значения экономического индикатора в будущем. В данном случае t — текущий момент времени, h — горизонт прогноза (например, 1 или 12 месяцев вперед). Иными словами, задача состоит в том, чтобы по данным, доступным сегодня, оценить, каким будет показатель через определенный период.

Принципиально важно соблюдать временную структуру данных: при построении прогноза использовать следует только такую информацию, которая доступна на момент времени t . Это исключает «утечку данных» и обеспечивает корректность оценки качества модели [2; 3]. Методология апробируется на задаче прогнозирования инфляции в России с использованием открытых данных и опубликованных исследований [4; 5; 6; 7], а также с анализом методов динамического усреднения моделей (DMA/DMS) [8].

Работа с социально-экономическими данными начинается с согласования частот наблюдений и синхронизации календарей публикации показателей. Игнорирование этих аспектов может приводить к скрытым утечкам информации и завышению оценок качества. Особое внимание уделено учету винтажей данных и возможных пересмотров статистики, поскольку использование «последней версии» данных [9] в историческом тестировании способно искусственно улучшить результаты¹.

Предобработка включает в себя обработку пропусков и выбросов. К тому же следует избегать так называемого заглядывания в будущее. Все операции импутации и стандартизации должны опираться исключительно на информацию, доступную в соответствующий период. Валидность кросс-валидации для временных рядов обсуждается в специализированной литературе, при этом речь идет о необходимости использования псевдовыборочных схем оценки [3].

Инженерия признаков строится вокруг лаговых зависимостей, разностей и агрегированных индикаторов, отражающих экономическую инерцию и влияние внешних факторов. В постановке data-rich в модель

¹ Tabassi E. Artificial Intelligence Risk Management Framework (AI RMF 1.0) // National Institute of Standards and Technology (NIST). January 26. 2023. URL: <https://www.nist.gov/publications/artificial-intelligence-risk-management-framework-ai-rmf-10> (дата обращения: 20.12.2025).

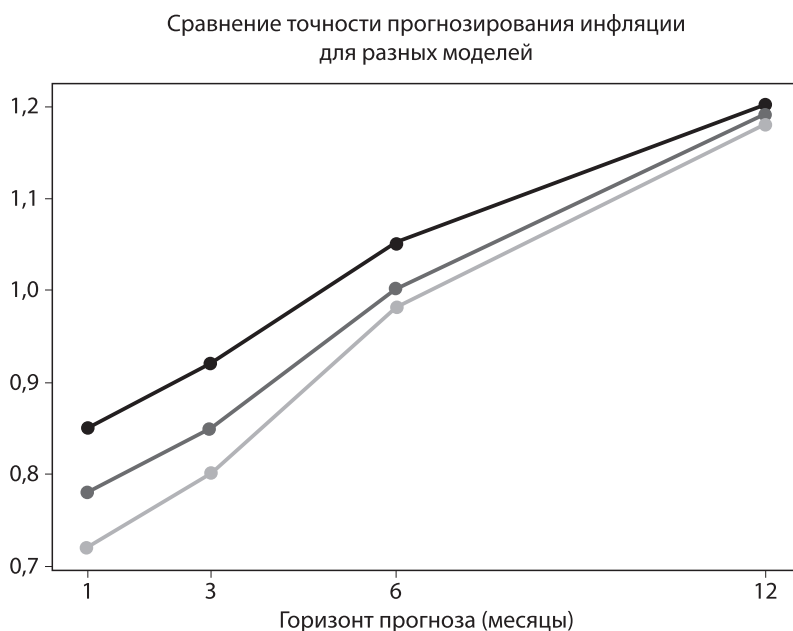


Рис. 1. Сравнение среднеквадратичной ошибки прогноза (RMSE) инфляции для различных моделей (AR, ML, DMA/DMS) на горизонтах 1–12 месяцев
Comparison of root mean square error (RMSE) of inflation forecasts for various models (AR, ML, DMA/DMS) over horizons of 1–12 months

Источник: составлено авторами.

включаются дополнительные макроэкономические и финансовые переменные, что позволяет учитывать более сложную структуру взаимосвязей [4; 5; 10].

В исследовании сопоставлены различные классы моделей. Классические временные ряды и state-space подходы обеспечивают прозрачную интерпретацию динамики и статистическую строгость [8; 11]. Регуляризованные регрессии позволяют работать с высокомерными пространствами признаков и снижать риск переобучения [9; 12]. Нелинейные ансамбли, такие как случайный лес и градиентный бустинг, отражают высокую эффективность при работе с табличными данными [13; 14], а современные реализации (XGBoost, LightGBM, CatBoost) обеспечивают масштабируемость и гибкость настройки [15; 16; 17]. Нейросетевые архитектуры, в том числе LSTM и GRU, предназначены для моделирования сложных временных зависимостей [4; 9].

Оценка качества осуществляется с использованием временно-корректной стратегии pseudo-out-of-sample, предполагающей обучение модели на прошлом и тестирование на будущем [4; 5; 8]. Основными метриками выступают MAE и RMSE, а для инфляцион-

ных прогнозов применяют их модификации в терминах forecast error [8]. Интерпретируемость результатов обеспечивается использованием методов SHAP и LIME, позволяющих оценивать вклад отдельных факторов в прогноз [2; 18]. Современные исследования в области объяснимого искусственного интеллекта (далее — ИИ) [19] свидетельствуют о значимости вопроса прозрачности моделей в управленческих контекстах¹.

Результаты

Анализ репликаций прогнозирования инфляции показывает, что преимущество моделей машинного обучения перед классическими авторегрессионными подходами зависит от широты информационного пространства. В «узкой» постановке, основанной преимущественно на лаговых зависимостях, улучшение качества носит умеренный характер и может варьироваться по горизонтам [5]. Сравнительный анализ среднеквадратичной ошибки прогноза (RMSE) на различных горизонтах свидетельствует о том, что преимущество ML-моделей носит в основном краткосрочный характер и сокращается по мере увеличения горизонта

¹ Tabassi E. Op. cit.

прогнозирования, как видно на рисунке 1.

На рисунке 1 синяя линия соответствует авторегрессионной модели (AR), оранжевая — модели машинного обучения (ML), зеленая — методам динамического усреднения/выбора моделей (DMA/DMS). Сопоставление динамики этих линий по оси горизонта прогнозирования помогает визуально оценить изменение точности различных подходов во времени.

Вместе с тем при расширении набора факторов (среда data-rich) отдельные нейросетевые и ансамблевые модели демонстрируют заметное снижение ошибок на краткосрочных горизонтах [4; 5]. Это свидетельствует о том, что нелинейные алгоритмы способны эффективнее использовать дополнительную информацию, особенно в условиях повышенной неопределенности. Но методы динамического усреднения моделей (DMA/DMS) не обеспечивают систематического доминирования над более простыми бенчмарками [8]. Их преимущества проявляются точно, зависят от горизонта прогноза и структуры данных, что подтверждает необходимость строгого сравнительного анализа.

Интерпретация через SHAP показывает, что значимыми предикторами инфляции выступают лаговые значения показателя, глобальные ценовые индикаторы и внутренние макроэкономические факторы [5; 18]. Эти результаты согласуются с выводами о роли нелинейных взаимодействий [9] в макроэкономическом прогнозировании¹.

Обсуждение

С учетом полученных результатов становится очевидным тот факт, что эффективность применения методов машинного обучения в макроэкономическом прогнозировании определена не столько формальной «мощностью» алгоритма, сколько комплексом методологических условий, в которых его применяют. Ключевая роль отведена сочетанию корректной регуляризации, позволяющей контролировать переобучение в высокомерных пространствах признаков, учета нелинейных зависимостей и взаимодействий факторов, а также строгой процедуры временно-корректной валидации [9]. Именно регуляризация и дисциплинированная настройка гиперпараметров обеспечи-

вают баланс между гибкостью модели и ее устойчивостью к шуму, а учет нелинейностей помогает выявлять эффекты, не фиксируемые в рамках линейных спецификаций. Без адекватной псевдовыборочной проверки (pseudo-out-of-sample) и сопоставимого сравнения с бенчмарками даже продвинутые ML-модели могут демонстрировать завышенные показатели качества, обусловленные спецификой обучающей выборки, а не реальной способностью к обобщению [9].

Дополнительную сложность в макроэкономическом контексте создают структурные сдвиги, изменения режима политики, внешние шоки и пересмотры статистики, которые приводят к феномену concept drift, то есть изменению связи между предикторами и целевой переменной во времени. В таких условиях модель, показавшая высокое качество на историческом интервале, может быстро утратить прогностическую силу при изменении экономической среды. Поэтому необходимы процедуры постоянного мониторинга качества, диагностики деградации и адаптации параметров или архитектуры модели [20]. Это предполагает регулярную переоценку на основе новых данных, использование адаптивных схем обучения и проведение стресс-тестов на различных исторических режимах. Игнорирование динамической природы экономической среды может привести к институционально значимым ошибкам прогноза, особенно в задачах денежно-кредитной политики и бюджетного планирования.

Принципиально важным считаем разграничение предсказательной и причинной интерпретации результатов. Методы объяснимого ИИ дают возможность анализировать вклад ряда признаков в формирование прогноза и тем самым повышают прозрачность алгоритма, но они отражают лишь структуру зависимостей, «выученных» моделью на исторических данных [12]. Такая интерпретация не эквивалентна установлению причинно-следственных связей и не заменяет экономическую теорию или экспериментальный дизайн. Предсказательная значимость фактора может быть следствием корреляционной структуры данных, а не его структурной роли в экономическом механизме. Поэтому использование ML-моделей в аналитике требует осознанного разграни-

¹ United States Macroeconomic data for 1959Q1 — 2009Q3 (macrodata dataset documentation) // Statsmodels. URL: <https://www.statsmodels.org/stable/datasets/generated/macrodata.html> (дата обращения: 19.02.2026).

чения между задачами повышения точности прогноза и задачей выявления причинных эффектов, особенно при разработке управленческих решений [12].

В условиях практического применения прогнозные системы должны соответствовать принципам ответственного ИИ, предполагающим прозрачность, воспроизводимость, документирование моделей и управление рисками их использования¹. Это включает в себя внедрение механизмов аудита, отслеживание смещений и нестабильности, анализ чувствительности результатов к изменению входных данных и гиперпараметров, а также формирование процедур реагирования на ухудшение качества. В социально-экономической сфере, в которой прогнозы влияют на распределение ресурсов и стратегические решения, такие требования приобретают не только техническое, но и институциональное значение. Следовательно, эффективность машинного обучения в макроэкономике определена не только точностью, но и степенью встроенности моделей [9] в систему методологического контроля, интерпретации и управления рисками².

Выводы

Предсказательное моделирование социально-экономических процессов на основе интеллектуального анализа данных требует не просто выбора «наиболее мощного» алгоритма, а построения целостной методологической архитектуры исследования. Такая архитектура должна предусматривать корректную временно-ориентированную (time-aware) валидацию, строгий контроль утечек информации и обязательную интерпретируемость результатов как часть научного и прикладного протокола. В отличие от стандартных задач машинного обучения с независимыми наблюдениями, социально-экономические ряды обладают внутренней зависимостью, нестационарностью и структурными сдвигами, что делает применение случайной кросс-валидации методологически некорректным. Псевдовыборочная стратегия (pseudo-out-of-sample), скользящие и расширяющиеся окна обучения, а также четкая фиксация момента доступности информации становятся необходимыми

условиями достоверной оценки обобщающей способности модели. Без соблюдения этих принципов даже высокоточные алгоритмы могут отражать иллюзорное преимущество, обусловленное скрытыми утечками данных или ретроспективной информацией, недоступной в реальном времени.

Современные модели машинного и глубокого обучения (ML/DL) действительно способны обеспечивать прирост точности прогнозов, особенно в расширенных (data-rich) постановках, при которых учитывают широкий спектр макроэкономических, финансовых и внешних факторов. Их преимущество во многом связано со способностью выявлять нелинейные зависимости, учитывать сложные взаимодействия признаков и адаптироваться к высокой размерности предикторного пространства. Однако эмпирические исследования говорят о том, что этот выигрыш не является универсальным: он проявляется преимущественно в определенных режимах экономики (например, в условиях повышенной волатильности или структурных сдвигов) и существенно зависит от корректности процедуры настройки и оценки. В «узких» постановках, при которых динамика показателя определена главным образом лаговой инерцией, традиционные авторегрессионные модели могут показывать сопоставимое качество, а иногда и превосходство над более сложными архитектурами. Таким образом, преимущество ML/DL-моделей является режимно-зависимым и методологически чувствительным: оно требует строгой дисциплины валидации, отбора признаков и сопоставимого сравнения с базовыми бенчмарками.

Усложнение модели само по себе не служит гарантией улучшения качества прогноза без учета специфики данных, горизонта прогнозирования и характера структурных зависимостей. Методы динамического усреднения и выбора моделей (например, DMA/DMS) теоретически ориентированы на адаптацию к изменяющейся структуре экономики. Однако практические результаты демонстрируют, что их преимущества носят точечный характер и не обеспечивают систематического доминирования над более простыми альтернативами. Это подтверждает более общий вывод: сложность модели должна быть функционально оправдана

¹ Tabassi E. Op. cit.

² Ibid.

информационной насыщенностью данных и реальной необходимостью учета нелинейностей. В противном случае растет риск переобучения, нестабильности параметров и ухудшения обобщающей способности на новых данных.

В социально-экономическом контексте особое значение приобретает интерпретируемость моделей и мониторинг их устойчивости во времени. Прогнозы используют в управленческих решениях, бюджетном планировании, денежно-кредитной политике и распределении ресурсов, поэтому они должны быть прозрачными и поддаваться аудиту. Методы объяснимого ИИ, такие как SHAP и LIME, позволяют анализировать вклад отдельных факторов в формирование прогноза и выявлять изменения в структуре значимости признаков. Но интерпретируемость не сводится к визуализации важностей; она служит элементом инсти-

туциональной ответственности и управления рисками применения алгоритмов. В условиях concept drift и структурных разрывов мониторинг деградации качества и регулярная переоценка модели становятся необходимыми элементами аналитической инфраструктуры.

Таким образом, практическая применимость предсказательных систем в социально-экономической сфере обусловлена их точностью и совокупностью методологических требований: корректной временной валидацией, контролем утечек, сопоставимостью бенчмарков, прозрачностью интерпретации и механизмами адаптации к изменяющейся среде. Современные ML/DL-подходы обладают значительным потенциалом, но их эффективность раскрывается лишь в рамках строгой воспроизводимой процедуры и при осознании ограничений, связанных с режимной природой экономических процессов.

Список источников

1. Барсегян А. А., Куприянов М. С., Степаненко В. В., Холод И. И. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP: учеб. пособие. 2-е изд., перераб. и доп. СПб.: БХВ-Петербург, 2007. 384 с.
2. Ribeiro M. T., Singh S., Guestrin C. "Why should I trust you?": Explaining the predictions of any classifier // Proc. of the 22nd ACM SIGKDD Int. conf. on knowledge discovery and data mining (San Francisco, 13–17 August, 2016). New York, NY: Association for Computing Machinery (ACM), 2016. P. 1135–1144. <https://doi.org/10.1145/2939672.2939778>
3. Bergmeir C., Hyndman R. J., Koo B. A note on the validity of cross-validation for evaluating autoregressive time series prediction // Computational Statistics and Data Analysis. 2018. Vol. 120. P. 70–83. <https://doi.org/10.1016/j.csda.2017.11.003>
4. Джункеев У. Прогнозирование инфляции в России на основе градиентного бустинга и нейронных сетей // Деньги и кредит. 2024. Т. 83. № 1. С. 53–76.
5. Астраханцева И. А., Герасимов А. С. Прогнозирование региональной инфляции на основе гибридной модели машинного обучения: градиентный бустинг и случайный лес // Научные труды Вольного экономического общества России. 2023. Т. 243. № 5. С. 200–226. <https://doi.org/10.38197/2072-2060-2023-243-5-200-226>
6. Павлов Е. Прогнозирование инфляции в России с помощью нейронных сетей // Деньги и кредит. 2020. Т. 79. № 1. С. 57–73. <https://doi.org/10.31477/rjmf.202001.57>
7. Уланова Д. В., Шиккульский М. И. Интеллектуальный анализ и прогнозирование социально-экономического развития субъекта Российской Федерации // Вестник Астраханского государственного технического университета. 2017. № 1. С. 62–68. <https://doi.org/10.24143/1812-9498-2017-1-62-68>
8. Стырин К. Прогнозирование инфляции в России методом динамического усреднения моделей // Деньги и кредит. 2019. Т. 78. № 1. С. 3–18. <https://doi.org/10.31477/rjmf.201901.03>
9. Coulombe P. G., Leroux M., Stevanovic D., Surprenant S. How is machine learning useful for macroeconomic forecasting? // Journal of Applied Econometrics. 2022. Vol. 37. No. 5. P. 920–964. <https://doi.org/10.1002/jae.2910>
10. Medeiros M. C., Vasconcelos G. F. R., Veiga A., Zilberman E. Forecasting inflation in a data-rich environment: The benefits of machine learning methods // Journal of Business and Economic Statistics. 2021. Vol. 39. No. 1. P. 98–119. <https://doi.org/10.1080/07350015.2019.1637745>
11. Айвазян С. А. Методы эконометрики: учебник. М.: Магистр: Инфра-М, 2010. 506 с.
12. Athey S., Imbens G. W. Machine learning methods that economists should know about // Annual Review of Economics. 2019. Vol. 11. P. 685–725. <https://doi.org/10.1146/annurev-economics-080217-053433>
13. Breiman L. Random forests // Machine Learning. 2001. Vol. 45. No. 1. P. 5–32. <https://doi.org/10.1023/A:1010933404324>

14. Friedman J. H. Greedy function approximation: A gradient boosting machine // *The Annals of Statistics*. 2001. Vol. 29. No. 5. P. 1189–1232. <https://doi.org/10.1214/aos/1013203451>
15. Chen T., Guestrin C. XGBoost: A scalable tree boosting system // *Proc. of the 22nd ACM SIGKDD Int. conf. on knowledge discovery and data mining* (San Francisco, 13–17 August, 2016). New York, NY: Association for Computing Machinery (ACM), 2016. P. 785–794. <https://doi.org/10.1145/2939672.2939785>
16. Ke G., Meng Q., Finley T. et al. LightGBM: A highly efficient gradient boosting decision tree // *Proc. of the 31st Int. conf. on neural information processing systems (NIPS 2017)*. Long Beach, CA: Curran Associates Inc., 2017. P. 3149–3157.
17. Prokhorenkova L., Gusev G., Vorobev A., Dorogush A. V., Gulin A. CatBoost: Unbiased boosting with categorical features // *Proc. of the 32nd Int. conf. on neural information processing systems (NeurIPS 2018)*. Montréal: Neural Information Processing Systems Foundation, 2018. P. 6638–6648.
18. Lundberg S. M., Lee S.-L. A unified approach to interpreting model predictions (SHAP) // *31st Conf. on neural information processing systems (NIPS 2017)*. Long Beach, CA: Curran Associates Inc., 2017. P. 1–10. <https://doi.org/10.48550/arXiv.1705.07874>
19. Arrieta A. B., Díaz-Rodríguez N., Del Ser J. et al. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI // *Information Fusion*. 2020. Vol. 58. P. 82–115. <https://doi.org/10.1016/j.inffus.2019.12.012>
20. Gama J., Žliobaitė I., Bifet A., Pechenizkiy M., Bouchachia A. A survey on concept drift adaptation // *ACM Computing Surveys*. 2014. Vol. 46. No. 4. P. 1–37. <https://doi.org/10.1145/2523813>

References

1. Barsegyan A.A., Kupriyanov M.S., Stepanenko V.V., Kholod I.I. Data analysis technologies: Data mining, visual mining, text mining, OLAP. St. Petersburg: BHV-Peterburg; 2007. 384 p. (In Russ.).
2. Ribeiro M.T., Singh S., Guestrin C. “Why should I trust you?”: Explaining the predictions of any classifier. In: *Proc. 22nd ACM SIGKDD Int. conf. on knowledge discovery and data mining* (San Francisco, August 13-17, 2016). New York, NY: Association for Computing Machinery (ACM); 2016:1135-1144. <https://doi.org/10.1145/2939672.2939778>
3. Bergmeir C., Hyndman R.J., Koo B. A note on the validity of cross-validation for evaluating autoregressive time series prediction. *Computational Statistics and Data Analysis*. 2018;120:70-83. <https://doi.org/10.1016/j.csda.2017.11.003>
4. Dzhunkeev U. Forecasting inflation in Russia using gradient boosting and neural networks. *Den'gi i kredit = Russian Journal of Money and Finance*. 2024;83(1):53-76. (In Russ.).
5. Astrakhantseva I.A., Gerasimov A.S. Regional inflation prediction based on hybrid machine learning model: Gradient boosting and random forest. *Nauchnye trudy Vol'nogo ekonomicheskogo obshchestva Rossii = Scientific Works of the Free Economic Society of Russia*. 2023;243(5):200-226. (In Russ.). <https://doi.org/10.38197/2072-2060-2023-243-5-200-226>
6. Pavlov E. Forecasting inflation in Russia using neural networks. *Den'gi i kredit = Russian Journal of Money and Finance*. 2000;79(1):57-73. (In Russ.). <https://doi.org/10.31477/rjmf.202001.57>
7. Ulanova D.V., Shikulskiy M.I. Predictive analysis and forecasting socio-economic development of the subject of the Russian Federation. *Vestnik Astrakhanskogo gosudarstvennogo tekhnicheskogo universiteta = Vestnik of Astrakhan State Technical University*. 2017;(1): 62-68. (In Russ.). <https://doi.org/10.24143/1812-9498-2017-1-62-68>
8. Styryn K. Forecasting inflation in Russia using dynamic model averaging. *Den'gi i kredit = Russian Journal of Money and Finance*. 2019;78(1):3-18. (In Russ.). 18. <https://doi.org/10.31477/rjmf.201901.03>
9. Coulombe P.G., Leroux M., Stevanovic D., Surprenant S. How is machine learning useful for macroeconomic forecasting? *Journal of Applied Econometrics*. 2022;37(5):920-964. <https://doi.org/10.1002/jae.2910>
10. Medeiros M.C., Vasconcelos G.F.R., Veiga Á., Zilberman E. Forecasting inflation in a data-rich environment: The benefits of machine learning. *Journal of Business and Economic Statistics*. 2021;39(1):98-119. <https://doi.org/10.1080/07350015.2019.1637745>
11. Aivazyan S.A. Econometric methods. Moscow: Magistr: Infra-M; 2010. 506 p. (In Russ.).
12. Athey S., Imbens G.W. Machine learning methods that economists should know about. *Annual Review of Economics*. 2019;11:685-725. <https://doi.org/10.1146/annurev-economics-080217-053433>
13. Breiman L. Random forests. *Machine Learning*. 2001;45(1):5-32. <https://doi.org/10.1023/A:1010933404324>
14. Friedman J.H. Greedy function approximation: A gradient boosting machine. *The Annals of Statistics*. 2001;29(5):1189-1232. <https://doi.org/10.1214/aos/1013203451>

15. Chen T., Guestrin C. XGBoost: A scalable tree boosting system. In: Proc. 22nd ACM SIGKDD Int. conf. on knowledge discovery and data mining (San Francisco, CA, August 13-17, 2016). New York, NY: Association for Computing Machinery; 2016:785-794. <https://doi.org/10.1145/2939672.2939785>
16. Ke G., Meng Q., Finley T., et al. LightGBM: A highly efficient gradient boosting decision tree. In: Proc. 31st Int. conf. on neural information processing systems (NIPS 2017). Long Beach, CA: Curran Associates Inc.; 2017:3149-3157.
17. Prokhorenkova L., Gusev G., Vorobev A., Dorogush A.V., Gulin A. CatBoost: Unbiased boosting with categorical features. In: Proc. 32nd Int. conf. on neural information processing systems (NeurIPS 2018). Montréal: Neural Information Processing Systems Foundation; 2018:6638-6648.
18. Lundberg S.M., Lee S.-L. A unified approach to interpreting model predictions (SHAP). In: 31st Conf. on neural information processing systems (NIPS 2017). Long Beach, CA: Curran Associates Inc.; 2017:1-10. <https://doi.org/10.48550/arXiv.1705.07874>
19. Arrieta A.B., Díaz-Rodríguez N., Del Ser J., et al. Explainable artificial intelligence (XAI): Concepts, taxonomies, opportunities and challenges toward responsible AI. *Information Fusion*. 2020;58:82-115. <https://doi.org/10.1016/j.inffus.2019.12.012>
20. Gama J., Žliobaitė I., Bifet A., Pechenizkiy M., Bouchachia A. A survey on concept drift adaptation. *ACM Computing Surveys*. 2014;46(4):1-37. <https://doi.org/10.1145/2523813>

Информация об авторах

Сергей Александрович Игнатьев

заведующий лабораторией информационных процессов в сложных социальных системах

Институт проблем передачи информации имени А. А. Харкевича Российской академии наук

127051, Москва, Большой Каретный пер., д. 19, стр. 1

Кирилл Александрович Буслов

научный сотрудник лаборатории информационных процессов в сложных социальных системах

Институт проблем передачи информации имени А. А. Харкевича Российской академии наук

127051, Москва, Большой Каретный пер., д. 19, стр. 1

Александр Александрович Ткаченко

младший научный сотрудник лаборатории обработки и передачи информации в когнитивных системах

Институт проблем передачи информации имени А. А. Харкевича Российской академии наук

127051, Москва, Большой Каретный пер., д. 19, стр. 1

Поступила в редакцию 02.03.2026
Прошла рецензирование 23.03.2026
Подписана в печать 27.04.2026

Information about the authors

Sergei A. Ignatov

head of the Laboratory of Information Processes in Complex Social Systems

Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute)

19 Bolshoy Karetnyy ln., bldg. 1, Moscow 127051, Russia

Kirill A. Buslov

researcher of the Laboratory of Information Processes in Complex Social Systems

Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute)

19 Bolshoy Karetnyy ln., bldg. 1, Moscow 127051, Russia

Alexander A. Tkachenko

junior researcher of the Laboratory for Information Processing and Transmission in Cognitive Systems

Institute for Information Transmission Problems of the Russian Academy of Sciences (Kharkevich Institute)

19 Bolshoy Karetnyy ln., bldg. 1, Moscow 127051, Russia

Received 02.03.2026
Revised 23.03.2026
Accepted 27.04.2026

Конфликт интересов: авторы декларируют отсутствие конфликта интересов, связанных с публикацией данной статьи.

Conflict of interest: the authors declare no conflict of interest related to the publication of this article.