



Банк России



Применение методов машинного обучения в задаче прогнозирования региональной инфляции на примере Ставропольского края

Серия докладов об экономических исследованиях

№ 157 / декабрь 2025

Н. Разин, Н. Крутов, А. Аразашвили, А. Нелюбина,
А. Копин, А. Смагин, Е. Харитонова

Николай Разин

Банк России, Департамент исследований и прогнозирования

E-mail: razinna@cbr.ru

Николай Крутов

Банк России, Департамент исследований и прогнозирования (в период работы в Банке России)

Антон Аразашвили

Банк России, Северо-Западное главное управление

E-mail: arazashviliat@cbr.ru

Алена Нелюбина

Банк России, Департамент исследований и прогнозирования

E-mail: NelyubinaAS1@cbr.ru

Артем Копин

Банк России, Департамент управления данными, Департамент данных, проектов и процессов

E-mail: kai6@cbr.ru

Александр Смагин

Банк России, Южное главное управление¹, Департамент данных, проектов и процессов

E-mail: SmaginAA01@cbr.ru

Екатерина Харитонова

Банк России, Департамент данных, проектов и процессов (в период работы в Банке России)

¹ Отделение по Ставропольскому краю.

Доклады Банка России проходят процедуру анонимного рецензирования со стороны членов Консультативного исследовательского совета Банка России и внешних рецензентов.

Авторы выражают признательность Сергею Селезневу и Андрею Синякову (ДИП БР) за неоценимый вклад в формирование методологического подхода этой работы, а также за консультации и предложения по улучшению работы. Мы также благодарим коллег за обсуждение результатов на внутренних семинарах, проведенных в Банке России в течение 2023 и 2024 годов.

Содержание настоящего доклада по экономическим исследованиям отражает личную позицию авторов. Результаты исследования являются предварительными и публикуются с целью стимулировать обсуждение и получить комментарии для возможной дальнейшей доработки материала. Содержание и результаты исследования не следует рассматривать, в том числе цитировать в каких-либо изданиях, как официальную позицию Банка России или указание на официальную политику или решения регулятора. Любые ошибки в данном материале являются исключительно авторскими.

Все права защищены. Воспроизведение представленных материалов допускается только с разрешения авторов.

Фото на обложке: Shutterstock/FOTODOM

Адрес: 107016, Москва, ул. Неглинная, 12, к. В

Тел.: +7 499 300-30-00, +7 495 621-64-65 (факс)

Официальный сайт Банка России: www.cbr.ru

Оглавление

Аннотация	5
1. Введение	6
2. Обзор литературы	8
3. Описание использованных данных	11
4. Описание методологии исследования	13
4.1. Описание методологии, использованной при псевдореальном прогнозировании	13
4.1.1. Подготовка данных для DMA и других байесовских моделей	14
4.1.2. DMA и DMS	15
4.1.3. Другие байесовские модели	16
4.1.4. Модель с ненаблюдаемыми компонентами и случайной волатильностью	17
4.1.5. ARIMA	17
4.1.6. Методы машинного обучения, примененные в работе	17
4.2. Методология интерпретируемости построенных моделей	19
4.2.1. Описание подхода через встроенный в XGBoost-метод оценки важности признаков	19
4.2.2. Описание SHAP-метода	19
5. Результаты экспериментов	20
5.1. Результаты сравнения метрик качества прогнозов моделей	20
5.2. Результаты экспериментов с встроенным в XGBoost-методом оценки важности признаков	23
5.3. Результаты экспериментов с модулями значений SHAP в глобальном контексте	25
5.4. Результаты экспериментов с модулями значений SHAP в локальном контексте	25
6. Заключение	29
Список литературы	30
Приложение	33
Приложение А. Результаты теста Диболда – Мариано	33
Приложение Б. Визуализация важности признаков в соответствии с встроенным в XGBoost-методом	34
Приложение В. Визуализация важности признаков в соответствии с методом SHAP	36
Приложение Г. Описание признаков, использованных при обучении моделей	38

Аннотация

В работе оценивается качество прогнозирования инфляции с применением методов машинного обучения на примере данных Ставропольского края. Исследование содержит сравнительный анализ псевдовневыборочного (pseudo-out-of-sample) прогноза часто используемых эконометрических моделей и ряда популярных методов машинного обучения, таких как бустинг над решающими деревьями, случайный лес и другие. Показано, что наиболее точным среди рассматриваемых в работе методов машинного обучения является бустинг над решающими деревьями (XGBoost). Также показано, что отличие в точности XGBoost по сравнению с эконометрическими моделями является статистически незначимым с точки зрения теста Диболда – Мариано. Кроме того, при помощи встроенных в XGBoost декомпозиций и декомпозиций Шепли был проведен анализ значимости предикторов с целью исследования возможностей интерпретируемости результатов прогноза алгоритмами машинного обучения. Проиллюстрирована возможность оценки вклада некоторых переменных в динамику прогноза инфляции в периоды значимых для российской экономики шоков. Сравнивая с классическими эконометрическими моделями точность прогнозов и логичная (но менее прозрачная) интерпретируемость алгоритмов машинного обучения на текущем этапе развития позволяют рассматривать их как полезный дополнительный инструмент анализа и интерпретации данных, но не в качестве способа значительного улучшения качества прогноза региональной инфляции.

Ключевые слова: региональная инфляция, машинное обучение, эконометрика, прогнозирование временных рядов, интерпретируемость моделей.

JEL-коды: E31, E37, E52, E58.

1. Введение

Построение точных прогнозов инфляции – это важная задача для всех участников финансового рынка. Целью прогнозирования в центральном банке является поддержка принятия решений по денежно-кредитной политике (ДКП), а также закоривание инфляционных ожиданий населения. Точные прогнозы инфляции позволяют регуляторам принимать решения, способствующие поддержанию стабильности цен и финансовых рынков, в особенности в условиях проведения политики таргетирования инфляции. Дополнительной сложностью в принятии подобных решений является то, что денежно-кредитная политика действует с лагами ввиду инерционности трансмиссионных механизмов, уровень инфляции реагирует на ужесточение или смягчение ДКП не мгновенно, а с некоторой задержкой (Smets, Wouters, 2003; Gertler, Karadi, 2015).

Во-вторых, коммуникационная политика в последнее время играет все большую роль в повышении эффективности монетарной политики. Она способствует лучшему пониманию экономическими агентами действий центрального банка, что помогает повысить уровень доверия и закорить инфляционные ожидания. Если центральный банк публикует свои прогнозы и экономические агенты им доверяют, они включают эти прогнозы в свои ожидания и принимают решения о будущем потреблении, сбережениях и инвестициях исходя из этой информации (Bernanke, 2024).

Построение точных прогнозов инфляции – сложная задача (Lenza et al., 2023; Ozgur et al., 2021, Medeiros et al., 2016; Байбуза, 2018), для решения которой в настоящее время применяется широкий спектр алгоритмов (Stock, Watson, 1999; Iversen et al., 2016). На сегодня уже существует обширный эконометрический и иной инструментарий, доступный центральным банкам и позволяющий получать качественные прогнозы. Однако в последнее время подобные задачи все чаще решают с помощью методов машинного обучения (Chakraborty, Joseph, 2017; Байбуза, 2018; Семитуркин, Шевелев, 2022; Lenza et al., 2023; Латышов и др., 2024). Одним из основных преимуществ этих методов является возможность улавливания нелинейных взаимосвязей между переменными, однако, как отмечается в ряде работ (Lenza et al., 2023), влияние этой нелинейности хоть и значимо, но невелико и качество прогнозов нелинейных моделей машинного обучения в макроэкономических задачах соотносится с линейными моделями state-of-the-art. Более того, макроэкономисты в развивающихся странах обычно имеют в распоряжении не так много данных во временном разрезе, чтобы нелинейности могли бы быть хорошо идентифицированы. Другое преимущество, которое также присуще моделям машинного обучения, – возможность легкого с технической точки зрения и без значительного увеличения времени расчетов добавления в модель большого количества потенциальных предикторов. Структурные (DSGE – ДСМОП, динамическая стохастическая модель общего равновесия) и полуструктурные (QPM – КПМ, Квартальная прогнозная модель) модели позволяют включать при необходимости десятки переменных. Однако с расширением их количества растет сложность и время подготовки прогноза. В отличие от классических моделей, алгоритмы машинного обучения позволяют использовать широкий спектр экономических предикторов (Monteforte, Moretti, 2010; Cicceri et al., 2020) и обучать модели, обладающие хорошей обобщающей способностью (Chakraborty, Joseph, 2017; Medeiros et al., 2019).

Вместе с тем у моделей искусственного интеллекта есть и свои слабые стороны. Например, в отличие от результатов применения традиционного эконометрического подхода, результаты алгоритмов машинного обучения сложнее интерпретировать, что создает дополнительные трудности для практического использования таких прогнозов в ДКП (Грищенко, Крылов, 2024). Именно поэтому повышение объяснимости и интерпретируемости моделей машинного обучения является важной областью исследований, и для этого разработано множество различных методов (Molnar, 2025; Renard et al., 2019). Среди этих методов можно отметить локальные интерпретируемые объяснения, не зависящие от модели (LIME) и аддитивные объяснения Шепли (SHAP), описанные в работе (Lundberg, Lee, 2017). Такие инструменты позволяют выделить из множества признаков самые важные, то есть те, которые обладают наибольшей объясняющей силой. Так, у экспертов появляется возможность сверить, насколько эти основные переменные, вносящие вклад в прогноз, соответствуют экономической логике.

Настоящая работа посвящена исследованию применения методов машинного обучения в задаче прогнозирования региональной инфляции. Авторы на примере данных Ставропольского края исследуют 2 прикладных вопроса:

- Каковы отличия в точности прогнозирования инфляции моделей машинного обучения и классических эконометрических моделей?
- Что лучшая из обученных ML-моделей² может «рассказать» про данные, на которых она обучалась?

Ответ на первый вопрос дается через исследование точности выбранных в этой работе ML-алгоритмов³ и их сравнение с часто применяемыми эконометрическими моделями⁴. Оценка точности работы отобранных алгоритмов в настоящем исследовании выполняется на горизонтах прогнозирования $h = 1, 3, 6$ и 12 месяцев с применением классических метрик (RMSE⁵, MAE⁶). Полученные данные в целом говорят о том, что результаты прогнозирования инфляции с помощью ML-алгоритмов статистически незначимо отличаются от результатов широко используемых моделей эконометрики на данных с точки зрения метрик, использованных в нашем материале.

Ответ на второй вопрос дается при помощи анализа интерпретируемости лучшей ML-модели (забегая вперед, отметим, что это XGBoost⁷), обученной для предсказания региональной инфляции. Информация о взаимосвязях (корреляциях) между переменными набора данных выражается в виде количественной оценки величины вклада конкретного признака в прогноз целевой переменной, то есть прогноз региональной инфляции. Множество значимых факторов в соответствии с экспериментами включает как макропеременные (например, процентные ставки – RUONIA, MIACR, обменный курс и так далее), так и микрокомпоненты инфляции, динамика которых наибольшим образом характеризовала изменения прогноза совокупного показателя региональной инфляции.

Также на примере ряда экономических шоков, значимых для российской экономики, рассмотрены наиболее важные факторы, влияющие на региональную инфляцию. Показано соответствие результатов анализа вкладов в прогноз региональной инфляции предикторов, выявленных ML-моделью, общей логике данных событий. Например, в период COVID-19 некоторые компоненты, почти не воздействовавшие на инфляцию ранее (в частности, медицинские товары), начали вносить существенный вклад в прогнозируемую переменную. Используемый инструментарий анализа важности признаков не является полноценной альтернативой декомпозициям, которые аналитики строят на основе результатов эконометрических или структурных теоретических моделей. Однако он может быть некоторым шагом и в сторону повышения уровня интерпретируемости моделей машинного обучения.

Работа имеет следующую структуру. В разделе 2 приведен обзор литературы. Раздел 3 содержит описание данных и аспектов их подготовки в рамках настоящего исследования. В разделе 4 описана методология. Результаты экспериментов, в том числе оценка и сравнение точности всех участвовавших в сравнении алгоритмов, а также исследование подходов к построению наглядной

² ML-модель – machine learning модель, модель машинного обучения.

³ Использовались следующие машинного обучения: бустинг над решающими деревьями (XGBoost), эластичная сеть (Elastic Net), случайный лес (Random Forest), модели линейной регрессии и другие.

⁴ Использовались классические эконометрические алгоритмы прогнозирования: авторегрессия второго порядка AR(2), авторегрессия на основании скользящего среднего (ARIMA), алгоритм динамического усреднения моделей (Dynamic Model Averaging, Raftery et al., 2010), модель с ненаблюдаемыми компонентами (UCSV, Stock and Watson, 2007) и другие.

⁵ RMSE (Root mean square error) – рассчитывается как корень средней квадратичной ошибки, метрика, применяемая для оценки качества моделей восстановления числовых зависимостей.

⁶ MAE (Mean absolute error) – рассчитывается как среднее модулей отклонений прогнозов модели в задаче восстановления числовых зависимостей.

⁷ XGBoost – конкретная реализация бустинга над решающими деревьями, алгоритма, относящегося к ансамблевым методам.

интерпретации XGBoost – лучшего ML-алгоритма среди всех, рассмотренных авторами, отражены в разделе 5. Раздел 6 содержит заключение работы и основные выводы.

2. Обзор литературы

Методы машинного обучения применяются для решения различных, характерных для центральных банков задач (Chakraborty, Joseph, 2017), в том числе для прогнозирования макроэкономических переменных (ВВП, инфляция, обменный курс и другие). При этом большое внимание уделяется прогнозированию инфляции (Garcia et al., 2017; Ozgur et al., 2021).

Необходимо, чтобы прогнозы, которые будут использоваться для формирования макроэкономической политики, обладали двумя свойствами. Во-первых, это точность, то есть ошибка (отклонение модельной траектории от фактических данных) должна быть низкой. Во-вторых, это интерпретируемость. То есть модели могут помочь аналитикам понять, почему прогноз выглядит именно так и за счет каких факторов.

Существует ряд работ, в которых исследуются различные аспекты применения методов машинного обучения в задаче прогнозирования инфляции – в частности, возможность учета нелинейных взаимосвязей между целевой переменной и предикторами, возможность использования большого количества данных для формирования более точных прогнозов. Например, в (Lenza et al., 2023) методы машинного обучения применяются для выделения нелинейных взаимосвязей между инфляцией и предикторами. В ряде работ применение методов машинного обучения улучшает результаты прогнозирования (Theoharidis et al., 2023; Шибанов, 2024), в то время как существуют работы, показывающие, что применение методов машинного обучения не приносит значимой прибавки к точности (Lenza et al., 2023). В работе (Байбуза, 2018) проводится сравнительная оценка точности методов машинного обучения и традиционных эконометрических методов. Показано, что некоторые модели искусственного интеллекта дают прогнозы лучше, чем базовые эконометрические модели. Некоторые методы⁸ машинного обучения позволяют без дополнительных усилий со стороны исследователей существенно увеличить количество предикторов для прогнозирования инфляции в сравнении с традиционными эконометрическими методами, что также является предметом исследований в части повышения точности прогнозирования (Medeiros et al., 2016; Medeiros et al., 2019).

В случае использования большого количества предикторов с точки зрения интерпретируемости модели в литературе по машинному обучению пока уделяется меньше внимания. В то же время прозрачное понимание логики модели (ее интерпретируемость) может помочь понять, как устроены сами данные (в случае большого количества переменных это особенно актуально и нетривиально) и как – сама модель, насколько ее поведение предсказуемо в зависимости от того, что у нее на входе. Необходимо отметить работы по прогнозированию инфляции, в которых приводится оценка значимости используемых предикторов в целях интерпретируемости полученных результатов (Chakraborty, Joseph, 2017; Семитуркин, Шевелев, 2023). В них предложен инструментарий, который дает возможность извлекать факторы, обладающие наибольшей объясняющей силой. Рассмотрим детальнее некоторые из перечисленных выше работ.

В работе (Chakraborty, Joseph, 2017) приведены примеры решения задач центральных банков с помощью методов машинного обучения. Авторы предлагают читателю обзор базовых понятий методов машинного обучения, подробно останавливаются на некоторых конкретных моделях (наивный байесовский классификатор, метод k ближайших соседей, решающие деревья и случайный лес, нейронные сети и другие), описывают принципы валидации моделей и подбора их гиперпараметров. Авторы приводят в качестве примера применения этих алгоритмов решение 3 задач: выявление компаний с нестандартными характеристиками отчетности, прогнозирование инфляции на 2-летнем горизонте и выделение компаний с высоким потенциалом влияния на финансовые технологии. Задача по прогнозированию инфляции Великобритании является задачей восстановления числовой зависимости (регрессии) и решается на квартальных данных, имеющихся в открытом доступе. Набор данных включает сведения в период с 1988 по 2015 г. для 12 показателей, включая прогнозируемый – значение темпа роста потребительских цен. Учитывая, что период

⁸ Например, такие как случайный лес, машина опорных векторов, бустинг над решающими деревьями и другие.

используемых данных включает кризис 2008 г., авторы разделяют набор данных на 3 интервала: I квартал 1990 – IV квартал 2004 г. как первоначальный период для обучения модели и определения гиперпараметров, I квартал 2005 – III квартал 2010 г. как период прогнозирования «до кризиса» и IV квартал 2010 – IV квартал 2015 г. как период прогнозирования «после кризиса». В результате сравнительного анализа точности использованных моделей машинного обучения (нейронные сети, метод опорных векторов, случайный лес и другие) и традиционно используемых для этих целей методов (AR(1), Ridge-регрессия и другие) авторы приходят к выводу, что методы машинного обучения позволяют добиться меньшего значения ошибки на каждом из интервалов «до кризиса», «после кризиса», а также на объединенном интервале I квартал 2005 – IV квартал 2015 года.

В работе (Medeiros et al., 2016) рассматривается применение методов машинного обучения к прогнозированию инфляции Бразилии. При этом применяются модели, способные работать с наборами данных высокой размерности и включающие отбор признаков, такие как LASSO⁹ (Tibshirani et al., 1995) и adaLASSO¹⁰ (Zou, 2006). Полученные результаты прогнозирования сравниваются с авторегрессионной моделью, а также иными эконометрическими моделями, предикторы которых отобраны на основании метода главных компонент. В качестве целевых переменных в настоящей работе используются следующие показатели инфляции Бразилии – индексы IPCA¹¹ и IGP-M¹². Используемый набор данных включает 102 предиктора, данные по которым представлены с шагом 1 месяц с января 2000 по декабрь 2013 года. Обе целевые переменные прогнозируются на глубину 1, 3, 6 и 12 месяцев. В результате экспериментов авторы делают вывод, что в задаче прогнозирования IPCA лучшими моделями являются LASSO и adaLASSO на горизонтах до 3 месяцев. С точки зрения теста Giacomini – White (Giacomini, White, 2006) результаты всех использованных моделей статистически не отличаются друг от друга. При этом отобранные признаки относились к группе характеристик государственного долга и инфляции. При прогнозировании индекса IGP-M статистически лучшей оказалась модель adaLASSO на горизонте прогнозирования 1 месяц, авторегрессионная модель – на горизонтах 3, 6 и 12 месяцев. В множество отобранных в этом случае переменных вошли те же самые, что и в случае IPCA, но с чуть меньшими весами, а также переменные, характеризующие объем наличных денег и государственных облигаций. Авторы на основании результатов применения методов машинного обучения к прогнозу инфляции делают следующие выводы. Во-первых, экономические параметры не обладают хорошей предсказательной силой при прогнозировании инфляции на длительные горизонты. Во-вторых, на коротком интервале (до 3 месяцев) параметры государственного долга являются хорошими предикторами. И, в-третьих, рассмотренные инфляционные индексы Бразилии – сложные функции, зависящие от множества факторов. Они не очень хорошо описываются стандартной кривой Филлипса.

Работа (Medeiros et al., 2019) посвящена прогнозированию инфляции в США с применением методов машинного обучения. В работе исследуется сразу несколько методов машинного обучения, но особое внимание уделяется модели случайного леса. Для формирования обучающей совокупности используются помесечные значения за период с января 1960 по декабрь 2015 г. 122 переменных, доступных в базе данных FRED-MD. Итоговый набор данных включает дополнительные временные ряды, рассчитанные на основании 122 переменных, лаг целевой переменной как авторегрессионный предиктор и 4 лаговых значения для каждой из получившихся переменных. Таким образом, итоговый набор данных представляет собой матрицу 508 x 672 значений предикторов. Глубина прогнозирования инфляции – от 1 до 12 месяцев. В дополнение прогнозируется еще 3 величины – совокупная инфляция за последующие 3, 6 и 12 месяцев. Точность моделей сравнивается для 2 периодов, характеризующихся разным значением волатильности – с января 1990 по декабрь 2000 г. и с января 2001 по декабрь 2015 года. Авторы приходят к выводу, что методы машинного обучения, такие как случайный лес, позволяют строить более точные прогнозы по сравнению со стандартными бенчмарками за счет использования большого количества предикторов. Авторы обращают внимание, что отобранные методом RF предикторы стабильны на

⁹ LASSO – least absolute shrinkage and selection operator.

¹⁰ adaLASSO – adaptive LASSO.

¹¹ IPCA – The Extended National Consumer Price Index – основной показатель инфляции Бразилии.

¹² IGP-M – General Index of Market-Prices – показатель инфляции, характеризующий ряд выделенных секторов экономики, рассчитываемый и публикуемый ежемесячно за период с 21 числа текущего месяца по 20 число следующего месяца.

всех горизонтах прогнозирования и преимущественно выбраны из категорий: цены, процентные ставки, обменный курс, рынки жилья и труда. Случайный лес также показывает лучшую точность на всех периодах тестирования, в том числе характеризующихся повышенной неопределенностью и периодом, следующим сразу после Великой рецессии.

Прогнозирование макропоказателей России, в том числе инфляции, имеет свою специфику. С одной стороны, как отмечено в (Шибанов, 2024), центральные банки крупных экономик, таких как ЕЦБ и ФРС США, в последние десятилетия (2010–2019 гг.) имели дело с низкой инфляцией и обучали свои модели на таких данных. В результате они не смогли предсказать масштаб и инерционность инфляции в ситуации шоков 2020–2023 годов. Для России такая проблема может стоять менее остро, поскольку наша инфляция постоянно испытывает на себе внешние шоки, связанные с санкциями, перестройкой экономики, изменением цен на экспортные товары и так далее. С другой стороны, успех моделей машинного обучения во многом связан с большими данными. При этом макроэкономисты такими массивами не обладают (Грищенко, Крылов, 2024). Особенно это касается России, в которой ряды макроэкономических показателей короткие относительно других стран.

Исследование (Байбуза, 2018) посвящено применению методов машинного обучения (LASSO, Ridge, Elastic Net, случайный лес и бустинг) в задаче прогнозирования инфляции России, а также включает сравнительный анализ их точности с классическими эконометрическими методами, такими как случайное блуждание (Random Walk) и авторегрессионная модель первого порядка. В качестве прогнозируемой переменной используется индекс потребительских цен на глубину 1, 2, 3, 6, 12, 18 и 24 месяца. Набор данных включает 93 временных ряда показателей (включая саму целевую переменную), значения которых представлены с шагом 1 месяц в период с февраля 2002 по июнь 2016 года. Согласно результатам исследования методы случайного леса и бустинга над решающими деревьями показывают результаты существенно выше в сравнении с AR (1) на глубине прогнозирования от 2 месяцев и при этом сравнимую точность с AR (1) на глубине прогнозирования 1 месяц. Авторы также показывают в работе, что модели случайного леса и бустинга над решающими деревьями отличаются большей точностью на необработанных данных, нежели на преобразованных.

В исследовании (Павлов, 2020) для прогнозирования инфляции России также применяются методы машинного обучения (ridge regression, neural networks, support vector machine) и сравниваются их ошибки с ошибкой простой эконометрической модели (AR). Используются ежемесячные данные с 2002 по 2018 г. 10 макроэкономических рядов: инфляция, ВВП, производительность труда, денежная масса, объем выданных кредитов, норма безработицы, экспорт, цена нефти, реальный располагаемый доход, процентная ставка. Автор показал, что нейронная сеть и метод опорных векторов справились с прогнозами не хуже традиционных подходов. Дополнительно применялась декомпозиция Шепли для интерпретации результатов.

В работе (Семитуркин, Шевелев, 2022) исследуется применение методов машинного обучения для прогнозирования региональной инфляции макрорегиона Сибирь, а также выполняется сравнение со стандартными подходами – в частности, моделями ARIMA и ARIMAX. В основе набора данных лежат 2 ряда за период с февраля 2002 по декабрь 2020 г. по макрорегиону Сибирь: инфляции и курса национальной валюты. Итоговый набор данных формируется из лаговых переменных от 1 до 12 месяцев, также средних значений и стандартных отклонений за предыдущие 3, 6, 9 и 12 месяцев. Прогнозируемая переменная – величина инфляции с глубиной 3, 6, 9, 12 и 24 месяца. В целом авторы делают вывод, что качество методов машинного обучения при прогнозировании инфляции для макрорегиона Сибирь сравнимо с традиционными эконометрическими методами, такими как ARIMA и ARIMAX. Однако на горизонте прогнозирования 2 года точность методов машинного обучения для большинства регионов и для макрорегиона Сибирь выше стандартных методов, в отличие от прогнозов на срок до года. При этом авторы приводят величины значимости признаков, отобранных при обучении градиентного бустинга, и отмечают, что большее влияние на целевую переменную оказывают среднее значение инфляции за последние 12 месяцев, предыдущее значение инфляции (месяц назад), порядковый номер месяца, выступающий моделью для учета сезонности, а также среднее значение курса рубля за последние 6 месяцев.

3. Описание использованных данных

Выбор переменных осуществлялся исходя из общей экономической логики и на основе изучения существующей литературы по прогнозированию инфляции. В работе использован стандартный набор макроэкономических переменных, см., например, (Байбуза, 2018; Стырин, 2019; Medeiros et al., 2016; Medeiros et al., 2019; Wang et al., 2012; Kim, Swanson, 2018; An et al., 2019; Lenza et al., 2024). Временные ряды охватывают данные с января 2002 по июнь 2022 года.

Поскольку мы разрабатываем алгоритмы для прогнозирования инфляции Ставропольского края, в нашем случае прирост месячного индекса потребительских цен является целевой переменной. Это означает, что при прогнозировании целевой переменной на глубины $h = 1, 3, 6$ и 12 месяцев в нашей работе строится прогноз прироста месячного индекса потребительских цен через $1, 3, 6$ и 12 месяцев соответственно от момента прогнозирования.

В качестве признаков мы используем сотни переменных, но для упрощения изложения опишем их по блокам. Детальный перечень всех 732 использованных показателей приведен в Приложении Г.

Блок данных по инфляции включает 453 микрокомпонента инфляции продовольственных, непродовольственных товаров и услуг Ставропольского края в терминах месяц к предыдущему месяцу. Предполагается, что поведение отдельных компонентов может предсказать динамику совокупного ряда ИПЦ. Например, в ситуации неопределенности потребители могут проявлять ажиотажный спрос на некоторые товары первой необходимости. В таком случае изменение темпов роста их цен будет при прочих равных характеризовать динамику общего ряда. Кроме того, включение микрокомпонентов позволяет расширить набор данных. Для моделей машинного обучения это важно, поскольку «успех многих современных алгоритмов ML объясняется огромным размером обучающей выборки» (Грищенко, Крылов, 2024) и чем больше признаков они используют, тем выше вероятность того, что точность работы модели будет выше. Дополнительно для обозначения периодов высокой инфляции была рассчитана бинарная переменная, которая принимает значение 1, если показатель инфляции в текущем месяце превышает пороговое значение, и 0 – иначе. Пороговое значение тут – это 80 перцентиль временного ряда.

Одной из важных макропеременных, необходимых для объяснения динамики инфляции, является обменный курс. В нашей работе использованы различные вариации показателя, отражающего номинальную цену доллара США в российских рублях. Изменение курса валют влияет на темпы прироста цен по 2 каналам. Во-первых, через цены импортируемых товаров и услуг, доля которых в последнее время снижается, но остается высокой. «Статистические данные о доле импорта в ИПЦ отсутствуют, но представление о ней в части товаров может дать удельный вес импорта в структуре товарных ресурсов розничной торговли (в последние годы – около 44%)»¹³. Во-вторых, валютный курс воздействует на инфляцию через совокупный спрос. Так, ослабление национальной валюты влечет за собой удорожание стоимости импорта. Местные товары становятся для потребителей более привлекательными, что стимулирует импортозамещение. Кроме того, повышается конкурентоспособность российской продукции за рубежом. Это приводит к росту чистого экспорта и совокупного спроса, что в свою очередь создает повышательное давление на темпы роста цен. Также в модели включена группа данных, отражающих цены и объемы торговли нефти. Эти переменные аппроксимируют динамику нефтегазовых экспортных доходов на ежемесячной основе. Рост цен и/или объемов продажи при прочих равных увеличивает количество валюты, поступающей в страну, укрепляя курс¹⁴. Данные торгов нефти и доллара преобразованы от

¹³ Банк России. Что такое инфляция:

https://cbr.ru/dkp/about_inflation/#highlight=%D0%B8%D0%BC%D0%BF%D0%BE%D1%80%D1%82%D0%B0

¹⁴ В условиях существования бюджетного правила связь нефть-курс уже не такая явная. Однако она по-прежнему существует, поскольку экспортеры продают валюту не только для уплаты налогов.

ежедневных к месячным. При этом при агрегации показателей произведен расчет средних, максимальных и минимальных значений и суммы за каждый месяц.

Следующий блок данных содержит различные ставки межбанковского кредитования группы MIACR и RUONIA, которые могут варьироваться в зависимости от срочности, валюты кредитования или кредитного рейтинга банка-заемщика. Передача импульса от ставки денежно-кредитной политики к ставкам денежного рынка является первым звеном процентного канала трансмиссионного механизма. Далее происходит трансляция от межбанковских ставок в ставки по более длинным операциям для реального сектора: доходности облигаций, ставки по банковским кредитам и депозитам. В результате изменяется спрос на кредиты, инвестиции, сбережения, что отражается в совокупном спросе и, соответственно, инфляции (Могилат, 2017). В качестве характеристики спроса в работе используются данные по обороту розничной торговли в терминах месяц к месяцу.

Также база включает данные о дефиците/профиците ликвидности и связанных с ними показателях: объемы кредитов Банка России кредитным организациям, требования по операциям РЕПО, средствам кредитных организаций на счетах в Банке России. Наличие, например, структурного дефицита ликвидности будет приводить к тому, что ставки межбанковского рынка окажутся выше ключевой ставки. Может возникнуть ситуация, при которой центральный банк проводит смягчающую политику, рыночные ставки также снижаются, но не на такую же величину, так как присутствует дефицит ликвидности. Такие ситуации могут исказить действие трансмиссионного механизма.

Для данных, выгруженных с ресурсов Банка России и характеризующих денежный рынок, ликвидность банковского сектора и денежно-кредитную политику регулятора, при преобразовании из ежедневного разреза в ежемесячный используется следующий подход: в качестве значения каждой переменной за конкретный месяц берется последнее заполненное дневное значение этой переменной в этом месяце. Если все дневные значения за некоторый месяц в отдельной переменной пропущены, ее месячное значение также остается пропущенным в преобразованном наборе данных.

Дополнительно в качестве признаков для выбранных переменных рассчитаны:

- лагированные¹⁵ значения;
- скользящие средние за разные периоды, продемонстрировавшие наилучшие результаты;
- аналоги производных¹⁶, характеризующие импульс движения временного ряда.

Для показателей, представляющих аналоги первой и второй производной, посчитаны лаги и скользящие средние, как для целевых переменных.

Ни одна из переменных не очищалась от сезонности. При этом были введены идентификаторы порядкового номера каждого месяца (от 1 до 12) и квартала (от 1 до 4), чтобы уловить сезонность. Кроме того, был добавлен такой признак, как порядковый номер календарного года (диапазон значений: от 2002 до 2022). Он также важен, поскольку в разные годы экономика проходила разные фазы экономического цикла (подъемы и кризисы), а также подвергалась нескольким структурным сдвигам (2014, 2020, 2022 гг.). Все использованные в работе данные являются открытыми. У некоторых показателей дата начала ряда может отличаться от января 2002 г., что связано с различиями во времени начала публикации или трудностями при выгрузке этих данных.

¹⁵ Представляют собой значения оригинального ряда, но смещенные на определенный интервал по времени, называемый лагом. Все лагированные переменные приведены в Приложении Г, названия переменных содержат префикс lag_.

¹⁶ Рассчитываются при помощи метода `numpy.gradient`, <https://numpy.org/doc/stable/reference/generated/numpy.gradient.html>.

Важно отметить, что при построении прогноза может возникнуть проблема «неровных краев» (Mamedli, Shibitov, 2021). Она характерна для временных рядов и объясняется неравномерной публикацией информации статистическими органами. В текущей работе этой проблемы нет, так как мы реализовывали прогнозы в псевдореальном времени. В будущем при использовании моделей, продемонстрированных в исследовании, необходимо будет заполнять пропуски в данных. Для этого существуют различные подходы, однако они не являются предметом данного исследования.

4. Описание методологии исследования

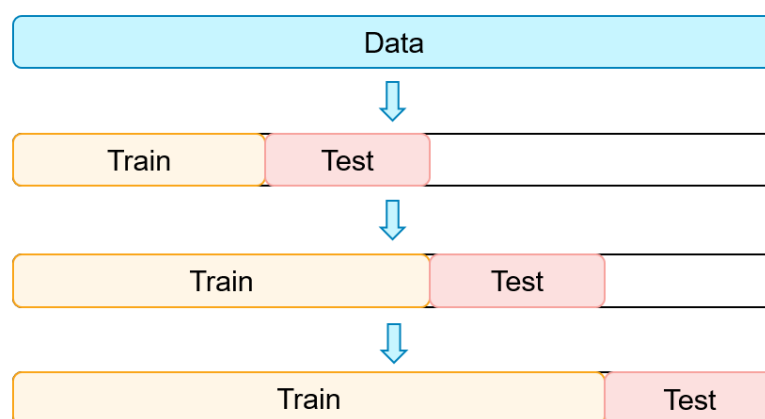
4.1. Описание методологии, использованной при псевдореальном прогнозировании

Для сравнения качества прогнозирования региональной инфляции с использованием моделей, примененных в рамках настоящей работы, строятся вневыборочные прогнозы в псевдореальном времени (pseudo-out-of-sample).

Традиционные методы кросс-валидации, такие как K-Fold, не подходят для временных рядов, поскольку они предполагают, что точки данных независимы и одинаково распределены. Однако в этой задаче порядок точек имеет значение и между ними существует временная зависимость. В подобных задачах для кросс-валидации используется специальный метод разбиения `TimeSeriesSplit`¹⁷, согласно которому мы имеем n отрезков временного ряда, где каждый последующий обучающий набор расширяется и состоит только из наблюдений, не включающих следующий тестовый набор. Таким образом, никакие будущие наблюдения не могут быть использованы при построении прогноза. Соответственно, в каждой итерации обучающая выборка используется для подбора гиперпараметров модели, а на тестовой мы получаем результат.

Следующая диаграмма (рис. 1) иллюстрирует серию обучающих и тестовых наборов, где желтые наблюдения формируют обучающие наборы, а красные – тестовые наборы.

Рис. 1. Диаграмма обучения и тестирования алгоритма с указанием принципа разбиения выборки на обучающую и тестовую



Прогнозы строятся отдельно для горизонтов прогнозирования $h = 1, 3, 6$ и 12 месяцев. Данные с 2002 и до конца 2014 г. используются исключительно для обучения моделей, а первый прогноз для горизонта $h = 1$ месяц строится на январь 2015 г., для горизонта $h = 3$ месяца – на март 2015 г., для $h = 6$ месяцев – на июнь, а для $h = 12$ месяцев – на декабрь 2015 года. Последний прогноз для всех рассматриваемых горизонтов производится на июнь 2022 года. На каждом шаге делается прогноз из момента времени t на h месяцев вперед, то есть прогноз величины на месяц $t + h$. Затем

¹⁷ https://scikit-learn.org/1.6/modules/generated/sklearn.model_selection.TimeSeriesSplit.html

обучающая выборка расширяется: к ней добавляются значения предикторов и прогнозируемой переменной за месяц $t + 1$. После чего модель обучается заново на расширенной тренировочной выборке и строит прогноз на месяц $t + h + 1$. Эти действия последовательно повторяются с первого месяца прогноза для соответствующего горизонта до последнего месяца – июня 2022 года.

В результате получаются последовательные значения прогнозов модели для каждого из месяцев прогнозного периода. Затем вычисляются абсолютные и квадратичные ошибки прогнозов модели и усредняются по всем месяцам, на которые осуществлялся прогноз. В итоге для каждой модели подсчитываются значения среднеквадратичной ошибки (RMSE) и средней абсолютной ошибки (MAE) для каждого горизонта прогнозирования.

4.1.1. Подготовка данных для DMA и других байесовских моделей

В качестве одного из методов для прогнозирования региональной инфляции было использовано динамическое усреднение моделей (DMA – Dynamic Model Averaging). Оно было разработано и впервые описано в (Raftery et al., 2010). Применение этого метода, а также сравнение его с несколькими бенчмарками, являющимися упрощениями DMA, основано на методологии, описанной в (Стырин, 2019). Качество моделей оценивалось с помощью пакета eDMA для языка программирования R (Catania, Nonejad, 2018).

Частью использования метода DMA и оценки основанных на нем бенчмарков является дополнительная предобработка данных. Ее необходимость объясняется прежде всего неспособностью применяемых алгоритмов автоматически качественно обрабатывать категориальные переменные и пропущенные значения в данных.

Для заполнения пропущенных значений в признаках используемого набора данных применялся следующий подход: если в ряде значений предиктора до данного пропущенного значения встречались заполненные значения, то текущее пропущенное заполнялось средним арифметическим всех предыдущих заполненных значений. Если до данного пропущенного значения были также лишь пропущенные значения во всем временном ряду рассматриваемого признака, то такое пропущенное значение заполнялось нулем. Значения временных признаков, таких как номер месяца, номер квартала, порядковый номер недели в году и так далее (кроме номера календарного года), заменялись на средние значения прогнозируемой переменной, отвечающие соответствующему значению заполняемой временной переменной до замены. Причем усреднение для такого заполнения производилось по значениям прогнозируемой переменной, доступным на дату, для которой осуществлялось заполнение временного признака. То есть, например, значение «1» для переменной с номером месяца всюду заменено на среднее значение январской инфляции в период с начала 2002 г. до момента, на который производится заполнение включительно.

Все признаки, кроме лагов инфляции и ее скользящих средних, приведены к нулевому среднему и единичной дисперсии. Для обучения моделей использовано фиксированное для каждого из горизонтов прогнозирования свое подмножество предикторов. Количество переменных в этих подмножествах для каждого из горизонтов равняется 20 из соображений оптимизации времени, требующегося для обучения и оценки модели DMA. Из них 2 предиктора – инфляция на последний месяц, с которого осуществляется прогноз, и на предшествующий ему месяц – присутствуют в каждом из подмножеств, отвечающих конкретному горизонту прогнозирования. Для выполнения отбора остальных 18 признаков были построены регрессионные зависимости с использованием каждой переменной из набора данных до 2015 г., где в качестве зависимой переменной была прогнозируемая региональная инфляция, а в качестве прогнозирующих переменных использованы свободный член и рассматриваемый признак из набора данных. Затем все признаки отранжированы по убыванию модуля значения t -статистики. Порядок переменных, полученных после ранжирования, отличается для разных горизонтов. Затем производился отбор 18 признаков, имеющих высший ранг, но при этом таким образом, чтобы попарные коэффициенты корреляции Пирсона для этого подмножества предикторов не превышали значения 0,9. Использование при

построении регрессий только данных до 2015 г. (с 2002 по 2014 г. включительно) обеспечивает отсутствие преимущества ретроспективного анализа, то есть использования при отборе данных, заведомо недоступных на время, с которого осуществляется прогноз. Также для каждой из моделей DMA и основанных на ней бенчмарков в качестве дополнительного предиктора используется свободный член, помимо 20 основных предсказывающих переменных. Описанная стратегия отбора применяется для всех оцениваемых байесовских моделей, которые используют в качестве предикторов не только лаги прогнозируемой переменной, но и другие признаки.

4.1.2. DMA и DMS

Прогноз DMA для каждого месяца – усредненное с некоторыми весами значение прогнозов отдельных моделей, использующих разные подмножества предикторов из 21 заранее отобранного (включая два лага инфляции и свободный член). Каждая индивидуальная модель использует свое фиксированное подмножество признаков для прогнозирования целевой переменной. При этом в исследовании каждое из таких подмножеств содержит первый лаг инфляции и свободный член. Усреднение прогнозов таких отдельных моделей производится с весами, равными апостериорным вероятностям того, что эта индивидуальная модель является процессом генерации данных для прогнозного месяца. Таким образом, количество отдельных линейных моделей равняется 2^{19} , то есть количеству подмножеств элементов, выбираемых из 19 элементов множества. Количество признаков в множестве, из которого формируются всевозможные подмножества, равняется 19, так как было фиксировано обязательное использование свободного члена и первого лага инфляции, а всего в предварительно отобранных множествах содержится по 21 предиктору.

Если некоторая индивидуальная модель использует подмножество предикторов $x_t^{(k)} \subseteq x_t, k = 1, \dots, K$, то она записывается в виде:

$$y_t = x_t^{(k)\top} \theta_t^{(k)} + \epsilon_t^{(k)},$$

где:

ошибка нормально распределена: $\epsilon_t^{(k)} \sim \text{i. i. d. } \mathcal{N}(0, V_t^{(k)})$;

$\theta_t^{(k)}$ – вектор следующих процессу случайного блуждания параметров.

Оценивание значения дисперсии ошибки $\hat{V}_t^{(k)}$ осуществляется в соответствии с (Коор, Korobilis, 2012) с постоянным коэффициентом сглаживания μ . Значение его выбрано равным 0,96 по образцу (Стырин, 2019). Значение вероятности того, что на момент t процессом генерации данных будет модель k при условии известных значений прогнозируемой переменной y в моменты времени вплоть до s обозначим за $w_{t|s,k}$. Тогда уравнение обновления этих значений, согласно (Raftery et al., 2010), имеет вид:

$$w_{t|t,k} = \frac{w_{t|t-1,k} p_k(y_t | y^{t-1})}{\sum_{j=1}^K w_{t|t-1,j} p_j(y_t | y^{t-1})},$$

где:

$p_k(y_t | y^{t-1})$ – плотность вероятности прогноза индивидуальной модели k . Значения для следующего шага вычисляются путем аппроксимации с использованием константы α , называемой фактором забвения:

$$w_{t|t-1,k} = \frac{w_{t-1|t-1,k}^\alpha}{\sum_{j=1}^K w_{t-1|t-1,j}^\alpha}.$$

Значение этой константы предполагается меньшим или равным 1. Для DMA и DMS оно выбрано равным 0,99 по аналогии с (Стырин, 2019). Прогноз DMA получается усреднением прогнозов индивидуальных моделей с весами $w_{t|t-1,k}$.

Для метода динамического отбора модели (DMS – Dynamic Model Selection), также применяемого в настоящем исследовании, в качестве прогноза берется не усредненное значение прогнозов индивидуальных линейных моделей, а просто прогноз одной отдельной модели, апостериорная вероятность которой наибольшая.

Для DMS и DMA по аналогии с (Dangl and Halling, 2012) производится также усреднение еще одного уровня. Оно осуществляется по различным значениям коэффициента сглаживания λ , который используется при аппроксимации ковариационной матрицы вектора возмущений из уравнения состояния для $\theta_t^{(k)}$. Усредняются прогнозы, отвечающие значениям λ из множества $\{0,90;0,91;0,92;\dots;1,00\}$, с весами, равными апостериорным вероятностям.

4.1.3. Другие байесовские модели

Байесовское усреднение моделей (BMA – Bayesian Model Averaging) – метод, также предполагающий усреднение значений индивидуальных моделей, однако, в отличие от DMA, параметры отдельных моделей, а также веса их усреднения считаются неизменными во времени (нединамическими). Так как BMA можно рассматривать как частный случай DMA, оценка этого метода также производилась с помощью библиотеки eDMA, с использованием параметров $\alpha = 1$, $\mu = 1$ и $\lambda = 1$.

Байесовский выбор модели (BMS – Bayesian Model Averaging) – модель, являющаяся упрощением BMA в таком же смысле, в каком DMS является упрощением для DMA. Вместо усреднения прогнозов моделей, производимого в BMA, метод BMS просто выбирает модель с самой высокой апостериорной вероятностью и использует ее прогноз в качестве финального. Оценка с помощью eDMA произведена с использованием тех же параметров, что и для BMA: библиотека позволяет оценить сразу обе эти модели.

Байесовская авторегрессия второго порядка (AR (2) – Autoregression) – модель без динамических параметров, в которой используются в качестве предикторов 2 лага инфляции и свободный член.

Авторегрессия второго порядка с меняющимися во времени параметрами (TVP-AR(2) – Time-Varying Parameters – Autoregression) – модель, которая использует тот же набор предикторов из двух лагов инфляции и свободного члена, но коэффициенты при предсказывающих переменных могут меняться со временем.

Динамическое усреднение моделей с использованием в качестве множества признаков только двух лагов инфляции и свободного члена (DMA-AR(2) – модель DMA, которая использует указанное ограниченное множество предикторов и тем самым получает на вход столько же информации, сколько и обычная авторегрессия.

Регрессия с динамическими параметрами по всем 21 из заранее отобранного множества предикторов, включая свободный член и 2 лага инфляции (TVP-KS – Time-Varying Parameter – Kitchen Sink) – модель, которая всегда содержит фиксированное множество всех предикторов, то есть усреднения прогнозов моделей не происходит, но коэффициенты регрессии меняются со временем.

4.1.4. Модель с ненаблюдаемыми компонентами и случайной волатильностью

Прогнозирование с помощью UCSV (Stock, Watson, 2007) осуществлено с опорой на пример применения этой модели из (Kroese, Chan, 2014). Модель опирается на следующие предположения для процесса порождения данных прогнозируемой переменной y :

$$\begin{aligned} y_t &= \tau_t + e^{h_t/2} \epsilon_t, \\ \tau_t &= \tau_{t-1} + \xi_t, \\ h_t &= h_{t-1} + \omega_t, \end{aligned}$$

где:

$$\epsilon_t \sim \text{i.i.d. } \mathcal{N}(0, 1);$$

$$\xi_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_\tau^2);$$

$\omega_t \sim \text{i.i.d. } \mathcal{N}(0, \sigma_h^2)$, а начальные распределения для процессов из уравнений задаются таким образом: $\tau_1 \sim \mathcal{N}(0, 9)$, $h_1 \sim \mathcal{N}(0, 9)$.

Для априорных распределений дисперсий предполагалось, что они имеют обратное гамма-распределение: $\sigma_\tau^2 \sim \text{Inv} - \text{Gamma}(\alpha_\tau, \beta_\tau)$ и $\sigma_h^2 \sim \text{Inv} - \text{Gamma}(\alpha_h, \beta_h)$ со значениями: $\alpha_\tau = \alpha_h = 10$, $\beta_\tau = 0,25^2(\alpha_\tau - 1)$, $\beta_h = 0,2^2(\alpha_h - 1)$. Прогноз же на горизонт h производится таким образом: $\mathbb{E}_t y_{t+h} = \mathbb{E}_t \tau_t$.

4.1.5. ARIMA

Модель ARIMA (Autoregressive Integrated Moving Average) – один из самых популярных методов анализа и прогнозирования временных рядов. ARIMA использует три основных параметра (p , d , q), которые выражаются как целые числа и параметризуют глубину авторегрессионной зависимости, порядок разностей целевой переменной и авторегрессионную компоненту остатков. Мы использовали библиотеку `autoarima` (Python) для автоматического подбора параметров модели с нашими данными.

4.1.6. Методы машинного обучения, примененные в работе

XGBoost (Extreme Gradient Boosting) – это алгоритм машинного обучения, основанный на деревьях решений и использующий градиентный бустинг как метод оптимизации целевого критерия. Он отлично подходит для решения задач машинного обучения, в которых используются структурированные или табличные данные разного размера, позволяет решать задачи прогнозирования, ранжирования, классификации и регрессии (Chen, Guestrin, 2016).

Реализация XGBoost на Python (библиотека XGBoost Python Package) дает доступ к большому количеству внутренних гиперпараметров алгоритма, которые являются предметом определения в каждой конкретной задаче. В нашей работе это делается с помощью метода `GridSearch`, который, по своей сути, обучает несколько моделей со всеми возможными комбинациями указанных гиперпараметров и в итоге выбирает лучший вариант.

Переборная сетка значений гиперпараметров:

- `base_score`: [0,35; 0,4; 0,45]
- `max_depth`: [2, 3]
- `early_stopping_rounds`: [30]
- `learning_rate`: [0,017; 0,014; 0,012; 0,01; 0,008; 0,006; 0,003]

- `n_estimators: [800, 600]`

Для определения лучшей комбинации гиперпараметров применяется процедура кросс-валидации, устроенная следующим образом: каждый блок данных в кросс-валидации состоит из одного года (это делается из-за годовой сезонности исследуемых данных). Таким образом, в рамках процедуры кросс-валидации для каждого тренировочного набора данных формируется n кросс-валидационных наборов данных (фолдов), где n равно количеству полных лет в тренировочном наборе данных. Модель обучается последовательно на данных периода с января 2002 по январь 2014 г. по следующему принципу. Первый валидационный набор данных формируется из первого фолда (январь 2002 – декабрь 2002 г.) как обучающего, и следующего за ним фолда (январь 2003 – декабрь 2003 г.) как тестового. Второй валидационный набор данных формируется уже из первых двух фолдов (январь 2002 – декабрь 2002 г. и январь 2003 – декабрь 2003 г.), и следующего за ним фолда (январь 2004 – декабрь 2004 г.). Процедура повторяется до конца тренировочной выборки (до фолда январь 2014 – декабрь 2014 г.), рассчитывается средняя ошибка предсказания по всем валидационным наборам данных. Таким образом, расчет средней ошибки выполняется для каждой комбинации гиперпараметров, а далее выбирается такая из них, которая соответствует наименьшему значению ошибки (в этой работе применяется RMSE). Таким образом, с помощью этого метода у нас есть возможность найти на отложенных выборках кросс-валидации наилучшие параметры и в дальнейшем именно их использовать при обучении и использовании модели.

В настоящем исследовании выполняется сравнение модели XGBoost с другими моделями, также широко используемыми в литературе при решении задачи прогнозирования инфляции (полный список моделей приводится в табл. 1, среди них – Random Forest, Elastic Net, GradientBoostingRegressor). Результаты моделей не только сравниваются с точки зрения величин метрик точности (RMSE, MAE), но и определяется статистическая значимость преимущества алгоритма XGBoost в сравнении со всеми остальными моделями.

Random Forest – популярный и эффективный алгоритм машинного обучения, основанный на композиции множества решающих деревьев. Сам прогноз получается в результате агрегирования ответов множества деревьев. Метод широко используется для задач классификации и регрессионного прогнозного моделирования со структурированными (табличными) наборами данных, в том числе в задачах прогнозирования временных рядов.

Elastic Net (Zou, Hastie, 2005) представляет собой комбинацию двух наиболее популярных регуляризованных вариантов линейной регрессии: Ridge и Lasso. Ridge использует регуляризацию L2, а Lasso использует регуляризацию L1. Elastic Net использует особенности обоих методов, извлекая уроки из их недостатков.

GradientBoostingRegressor – решение градиентного бустинга от библиотеки sklearn и по своей сути аналогичен решению от библиотеки XGBoost. Однако в практическом смысле есть очень существенные различия. XGBoost включает дополнительную регуляризацию, что сильно снижает его уязвимость к переобучению, а также эффективно справляется с недостающими значениями, что не предусмотрено в аналогичном методе библиотеки sklearn. Кроме этого, необходимо отметить, что GradientBoostingRegressor работает медленнее, чем XGBoost, а также менее эффективен с точки зрения использования памяти и может быть не распараллелен по умолчанию из библиотеки.

4.2. Методология интерпретируемости построенных моделей

4.2.1. Описание подхода через встроенный в XGBoost-метод оценки важности признаков

При подсчете значений значимости признаков (`feature_importances`)¹⁸ в качестве значения параметра `importance_type` использовано `total_gain`, что дает для каждой из предсказывающих переменных в качестве оценки ее важности сумму изменения функции потерь при каждой вершине по всем деревьям. Полученные таким образом значения оценки важности признаков представляют собой функцию от обучающей выборки и отражают глобальную значимость конкретных признаков для предсказаний модели. Этот метод оценки важности, однако не позволяет локально понять, какой вклад вносит отдельный признак в прогноз на некоторый конкретный месяц из числа тех, на которые осуществлялся прогноз. То есть используется исключительно в глобальном контексте.

4.2.2. Описание SHAP-метода

SHAP (SHapley Additive exPlanations, метод Шепли) – метод, позволяющий интерпретировать результаты работы моделей путем оценки величины вклада каждого признака в итоговый прогноз. Одним из основных преимуществ этого метода является то, что он может быть применен к любой модели (*model agnostic*), строящей взаимосвязь между признаками и целевой переменной, и не зависит от конкретного семейства алгоритмов, которому принадлежит модель.

В нашей работе значения SHAP были рассчитаны для всех обученных моделей семейства XGBoost для всех глубин предсказания, то есть для $h = 1, 3, 6$ и 12 месяцев. При этом значения SHAP анализировались в двух аспектах – в глобальном и локальном контексте.

Глобальный контекст – это усредненные значения модулей SHAP по всей использованной для обучения выборке, отражающие средний вклад каждого признака, применяемого для прогнозирования, в прогнозируемую переменную по всей обучающей совокупности. Вектор значений SHAP рассчитывается в этом случае как средняя величина модулей значений Шепли, вычисленных для каждого объекта обучающей совокупности.

Локальный контекст – это значения SHAP для последней точки временного ряда, по которой предсказывается инфляция на следующем шаге, и в этом случае играет значение уже величина влияния конкретного признака на предсказываемое значение инфляции, что позволяет локально оценить, какие признаки, с точки зрения модели, сильнее других влияют на целевую переменную.

Значения SHAP позволяют интерпретировать результаты работы моделей машинного обучения благодаря следующим свойствам:

- Значения SHAP являются аддитивными, что означает, что вклад каждой функции в окончательный прогноз можно вычислить независимо, а затем суммировать. Это свойство позволяет эффективно вычислять значения SHAP даже для многомерных наборов данных.
- Значения SHAP равны нулю для отсутствующих или нерелевантных для прогноза признаков. Это делает значения SHAP устойчивыми к отсутствующим данным и гарантирует, что нерелевантные признаки не исказят интерпретацию.
- Значения SHAP не изменяются если вклад признака не меняется. Это означает, что значения SHAP устойчивы к изменению архитектуры модели либо ее гиперпараметрам при условии, что вклад признаков остается тем же самым.

¹⁸ См. Python API Reference, https://xgboost.readthedocs.io/en/latest/python/python_api.html

В целом значения SHAP обеспечивают последовательный и формальный способ получить представление о том, как модель машинного обучения делает прогнозы и какие признаки оказывают наибольшее влияние.

При визуализации значимостей признаков они объединены в группы для простоты представления (см. перечень групп в Приложении Г). При этом значимость, рассчитанная для каждой группы признаков, представляет собой усредненное значение значимостей всех признаков, входящих в эту группу. Таким образом, группы удобно сравнивать между собой с точки зрения среднего вклада каждого признака по отдельности. Для того чтобы увидеть общий вклад группы, достаточно взять средний вклад от каждого признака и умножить на количество признаков в группе.

5. Результаты экспериментов

Эксперименты, проведенные авторами, можно разделить на 2 группы.

Первая группа экспериментов посвящена исследованию точности выбранных в настоящей работе ML-алгоритмов и их сравнению с различными эконометрическими моделями. Полученные результаты в целом говорят о том, что точность прогнозирования инфляции с помощью ML-алгоритмов статистически незначимо отличается от точности моделей, часто используемых в эконометрике.

Вторая группа экспериментов посвящена исследованию интерпретируемости лучшей с точки зрения метрик ML-модели, обученной для предсказания региональной инфляции. В этой группе экспериментов ML-модель рассматривается как объект, который может что-то «рассказать» про данные, на которых он обучался. Информация о взаимосвязях (корреляциях) между переменными набора данных выражается в виде количественной оценки величины вклада группы признаков в прогноз целевой переменной, то есть прогноз региональной инфляции. Важно отметить, что большой вклад в прогноз переменной конкретным признаком не стоит интерпретировать как причинно-следственную связь между этим признаком и целевой переменной. Тем не менее найденные моделью взаимосвязи между целевой переменной и предикторами могут говорить о косвенной роли данного предиктора, передающего истинное влияние некоторого фактора на инфляцию. Информация о найденных ML-моделью взаимосвязях между признаками и целевой переменной в дальнейшем может быть использована при анализе и моделировании причинно-следственных связей.

Предложенный инструментальный анализ важности признаков на текущем этапе не является полноценной альтернативой декомпозициям, которые аналитики строят на основе результатов эконометрических или структурных теоретических моделей. Однако он может рассматриваться как хороший первый шаг для понимания результатов и повышения уровня интерпретируемости моделей машинного обучения, используемых в макроэкономическом прогнозировании.

В разделе 5.1 приведены детали экспериментов по анализу точности ML-моделей, а также их сравнению со стандартными эконометрическими моделями. В разделах 5.2 и 5.3 приводятся результаты исследования интерпретируемости одной из ML-моделей с помощью подходов feature importances (XGBoost) и SHAP в глобальном смысле. В разделе 5.4 приводится разбор значимостей SHAP в локальном смысле в том числе на примере поведения этого показателя в период конкретных экономических событий.

5.1. Результаты сравнения метрик качества прогнозов моделей

В табл. 1 представлены сравнительные результаты 14 алгоритмов прогнозирования. По горизонтали приводятся метрики (RMSE и MAE) и горизонты прогнозирования, по вертикали – примененный метод прогнозирования на данных, используемых в рамках данного исследования. При этом для методов с использованием эконометрических моделей DMA, DMS, BMA, BMS и TVP-

KS приводятся результаты как для экспериментов со всеми 734 предикторами, так и для применения методов к набору предсказывающих переменных, из которого заранее удалены компоненты инфляции. На пересечении строки (конкретная модель) и столбца (конкретная глубина прогнозирования и метрика) приводится значение этой метрики для соответствующего алгоритма, прогнозирующего на соответствующую глубину. Цветом обозначены лучшие 5 моделей для каждой пары глубина-метрика, оранжевый – лучшая модель, желтый – следующие топ-4 после лучшей.

Табл. 1. Сравнение точности методов прогнозирования инфляции

Модель	h = 1		h = 3		h = 6		h = 12	
	RMSE	MAE	RMSE	MAE	RMSE	MAE	RMSE	MAE
XGBoost	0,85	0,44	0,72	0,42	0,71	0,40	0,75	0,41
Random Forest	0,83	0,47	0,74	0,46	0,77	0,50	0,78	0,46
Elastic Net	1,06	0,60	0,85	0,56	0,97	0,61	0,91	0,55
GradientBoostRegressor	0,82	0,48	0,78	0,47	0,81	0,48	0,89	0,53
Arima	0,92	0,55	0,93	0,59	0,96	0,66	1,01	0,66
DMA	1,07	0,54	0,94	0,57	0,79	0,47	0,77	0,44
DMS	1,11	0,57	1,05	0,62	0,85	0,52	0,83	0,49
BMA	0,87	0,48	0,85	0,56	0,82	0,52	0,78	0,48
BMS	0,87	0,49	0,89	0,58	0,83	0,54	0,80	0,48
DMA – без компонентов	1,02	0,57	0,89	0,51	0,85	0,51	0,78	0,45
DMS – без компонентов	1,18	0,65	0,96	0,52	1,03	0,60	0,80	0,46
BMA – без компонентов	0,88	0,49	0,78	0,47	0,81	0,48	0,78	0,46
BMS – без компонентов	0,90	0,50	0,80	0,49	0,81	0,48	0,81	0,47
AR(2)	0,85	0,49	0,81	0,52	1,00	0,75	0,79	0,47
TVP-AR(2)	0,90	0,49	0,89	0,52	1,01	0,53	0,78	0,44
DMA-AR(2)	0,90	0,49	0,80	0,47	0,90	0,47	0,78	0,43
UCSV	0,86	0,48	0,81	0,52	0,72	0,41	0,80	0,46
TVP-KS	0,90	0,53	0,89	0,58	0,84	0,53	1,01	0,74
TVP-KS – без компонентов	0,99	0,58	0,86	0,54	0,91	0,57	0,85	0,57

Главным наблюдением по итогам сравнения метрик качества прогнозов моделей является высокое качество модели XGBoost при прогнозировании региональной инфляции. По метрике RMSE эта модель оказывается лучшей для горизонтов прогнозирования h = 3, 6, 12 месяцев, а с точки зрения метрики MAE модель XGBoost становится самой точной для всех рассмотренных горизонтов. Оценка статистической значимости, однако показывает разную степень надежности

полученных результатов сравнения качества моделей. Так, с уровнем значимости 0,1 качество прогноза XGBoost с точки зрения теста Диболда – Мариано значительно отличается от качества прогноза модели Random Forest¹⁹, также основанной на решающих деревьях, лишь для глубины прогноза $h = 6$ месяцев. Полный перечень величин р-значений теста Диболда – Мариано, проведенного для сравнения прогнозов алгоритма XGBoost и каждого из приведенных в табл. 1 алгоритмов, доступен в Приложении А. Табл. 2 включает величины р-значений для теста Диболда – Мариано для сравнения прогнозов алгоритмов XGBoost и Random Forest.

Табл. 2. Результаты теста Диболда – Мариано для сравнения прогнозов XGBoost и Random Forest

Горизонт прогнозирования	$h = 1$	$h = 3$	$h = 6$	$h = 12$
р-значения	0,7236	0,8047	0,0806	0,3195

Для горизонта прогнозирования $h = 1$ месяц наибольшее качество по метрике RMSE – 0,82 – показывает модель GradientBoostRegressor, являющаяся альтернативным по отношению к XGBoost методом градиентного бустинга, рассмотренным в настоящем исследовании. В топ-5 по метрике RMSE также попадает еще одна ML-модель – Random Forest со значением метрики RMSE равным 0,83. XGBoost также входит в пятерку лучших по RMSE-моделей, однако значение метрики для него, равное 0,85, совпадает со значением для модели-бенчмарка – авторегрессии второго порядка AR(2). Это вполне соотносится с общей закономерностью, наблюдаемой в различных исследованиях (Семитуркин, Шевелев, 2023), при сравнении моделей машинного обучения с эконометрическими моделями для прогнозирования инфляции: на коротких горизонтах добиться значительного превосходства ML-моделей, как правило, не удастся. Действительно, отличие качества, оцениваемого по метрике RMSE, между самой лучшей для этого горизонта моделью GradientBoostRegressor и простым линейным бенчмарком AR (2) очень мало. Замыкает пятерку лучших по RMSE моделей для $h = 1$ месяц модель UCSV с качеством 0,86. По метрике MAE, как было подчеркнуто ранее, лучшей моделью оказывается XGBoost с качеством 0,44. В топ-5 по MAE не входит AR (2), зато попадает другая линейная эконометрическая модель – ВМА (метод байесовского усреднения моделей), запущенная на наборе данных, включающем компоненты инфляции в качестве предикторов.

На горизонте прогнозирования $h = 3$ месяца XGBoost показывает качество 0,72 и 0,42 по RMSE и MAE соответственно. В топ-5 моделей по RMSE в этот раз, помимо XGBoost, Random Forest, GradientBoostRegressor, входят ВМА и ВМС, прогнозирующие инфляцию без использования лагов ее компонентов. Улучшение качества прогнозов некоторых моделей по сравнению с горизонтом, равным 1, наблюдается из-за различия в выборках, для которых строится прогноз для разных h . Так, прогноз на 1 месяц вперед строится начиная с января 2015 г., а прогноз на 3 месяца вперед – с марта. При этом в самом начале 2015 г. инфляция остается достаточно высокой, а затем снижается. Учет в метриках для моделей, прогнозирующих на 1 месяц вперед, этих периодов высоких значений инфляции приводит к худшему наблюдаемому качеству в сравнении с моделями, прогнозирующими на 3 и более месяцев вперед.

Значения метрик RMSE и MAE модели XGBoost для прогнозирования на полгода вперед равны 0,71 и 0,40. Можно заметить, что эти показатели несильно изменились по сравнению с предыдущим горизонтом прогнозирования. В пятерку лучших по RMSE-моделей в этот раз возвращается UCSV, отсутствовавшая в ней для $h = 3$. Интересно заметить, что топ-5 по этой метрике впервые попадает DMA (динамическое усреднение моделей), прогнозирующая с использованием компонентов, со значением качества, равным 0,79. Более того, только на таком

¹⁹ Модель выбрана для сравнения, поскольку тоже принадлежит классу ансамблевых методов.

горизонте прогнозирования этому сложному эконометрическому методу удастся обогнать по обеим оцениваемым метрикам бенчмарк AR(2).

Качество прогноза самой точной модели XGBoost на горизонте $h = 12$ месяцев составляет 0,75 по RMSE и 0,41 по MAE. На втором месте по метрике RMSE располагается DMA с использованием компонентов со значением этой метрики, равным 0,77, а третье место со значением этой метрики 0,78 делят еще 2 модификации DMA – DMA-AR(2) и DMA без компонентов, а также ансамблевый метод Random Forest. Небольшая величина превосходства XGBoost над линейной моделью DMA и ее модификациями – в частности, DMA-AR(2), для годового горизонта прогнозирования может рассматриваться как косвенное свидетельство того, что зависимости между предикторами и прогнозируемой переменной по большей части носят линейный характер. Это соответствует и основным выводам, сделанным в работе (Lenza et al., 2023).

В то же время в топ-5 по RMSE попадает с качеством 0,78 метод DMA-AR(2) – линейная модель, использующая в качестве предикторов только 2 подряд идущих лага прогнозируемой переменной. Для горизонта $h = 12$ месяцев это будут значения региональной инфляции за год до прогнозируемого и за 13 месяцев (за год и 1 месяц) до прогнозируемого. Прогноз, опирающийся только на 2 значения инфляции год назад, почти не уступает по качеству прогнозам, учитывающим и другие предикторы.

5.2. Результаты экспериментов с встроенным в XGBoost-методом оценки важности признаков

В разделе ниже рассматривается, какие факторы вносили наибольший вклад в прогнозную динамику инфляции Ставропольского края в соответствии с встроенным в XGBoost подходом – feature importances. Для каждого из горизонтов прогнозирования приводятся признаки с наибольшей значимостью для последней («наиболее поздней», с наибольшей обучающей выборкой) из обученных моделей XGBoost.

Отметим, что в каждой подгруппе оказались значимы как макроэкономические факторы, например, процентные ставки, объем торгов нефтью, обменный курс и другие, так и микрокомпоненты инфляции, динамика которых наибольшим образом характеризовала изменения совокупного показателя региональной инфляции. Если в список наиболее важных признаков попал какой-то компонент инфляции, то это означает, что его значение в текущий момент влияет на прогноз общей инфляции. В некотором смысле это можно трактовать как инерционность.

Перечислим сначала макрофакторы, которые оказались важной частью прогноза инфляции Ставропольского края на каждом прогнозном промежутке. В топ-20²⁰ важных признаков для каждого горизонта прогнозирования попали группы признаков RUONIA и MIACR. В краткосрочной перспективе, когда речь идет об одном или нескольких месяцах, действия центрального банка транслируются в первую очередь через канал ожиданий, канал обменного курса и процентный канал. На более длинном горизонте (как правило, от 6 месяцев) изменения ставок начинают в полной мере отражаться в инфляции²¹.

Стоит обратить внимание на то, что для каждого горизонта прогнозирования в топ-20 входят лаги инфляции и аналогов ее производных (приращения в единицу времени). Их наличие можно интерпретировать как вклад в прогноз инерционных и нелинейных инфляционных ожиданий.

²⁰ В некоторых графиках число признаков может быть меньше из-за постфильтрации по пороговому значению, используемому для отсеивания незначимых признаков.

²¹ Сроки полного проявления эффектов от решений Банка России на экономику и инфляцию растянуты во времени. Диапазон реализации составляет 3–6 кварталов. При этом такой интервал не означает, что в первые 2 квартала ничего не происходит. Он означает, что основная масса эффектов реализуется в этом интервале. Но при более сильных изменениях ставки и в первые два квартала эффекты будут более заметны. Из интервью А.Б. Заботкина <https://www.cbr.ru/press/event/?id=17127>

В прогнозе на 6 и 12 месяцев высокой значимостью (входит в топ-3 признаков) обладает признак, характеризующий оборот розничной торговли, который аппроксимирует спрос в данном случае (признак «Оборот торговли РФ»). Поскольку спрос влияет на инфляцию с некоторыми лагами, этот фактор проявляется на горизонте от 2 кварталов включительно. Практически для всех периодов оказалась важной характеристика порядкового номера месяца, что означает, что модель улавливает сезонность.

Примеры визуального представления значимостей признаков для глубин прогнозирования $h = 3, 6$ и 12 месяцев приведены в Приложении Б. В этом разделе мы ограничимся приведением примера визуализации значимости признаков для горизонта прогнозирования $h = 1$ (рис. 2 и 3).

Рис. 2 и 3 дают представление о том, как распределяется значимость топ-10 предикторов в динамике псевдореального эксперимента (рис. 3) и значимость топ-20 предикторов в последней точке псевдореального эксперимента (рис. 2). Динамика компонентов отображает поведение значимости во времени для фиксированного множества из топ-10 по значимости предикторов модели, обученной на всем наборе данных вплоть до марта 2022 года. Значимость для топ-20 предикторов приведена в виде столбчатой диаграммы для простоты и наглядности представления. Например, видно, что среди наиболее важных признаков с точки зрения встроенного в XGBoost-метода оценивания значимости оказываются «Все товары и услуги (без овощей)», «Базовый ИПЦ», группы продовольственных товаров, сезонность, лаги инфляции и ее изменения, «RUONIA», «MIACR» и другие (рис. 2). Отметим, что таким образом определенная важность не говорит напрямую о причинно-следственных связях между инфляцией и вышеперечисленными предикторами, а говорит о том, что эти предикторы могут передавать истинное влияние иных факторов, определение которых является отдельной исследовательской задачей.

Рис. 2. 20 наиболее важных признаков для модели, обученной на данных до мая 2022 г. включительно, с горизонтом прогнозирования $h = 1$ месяц

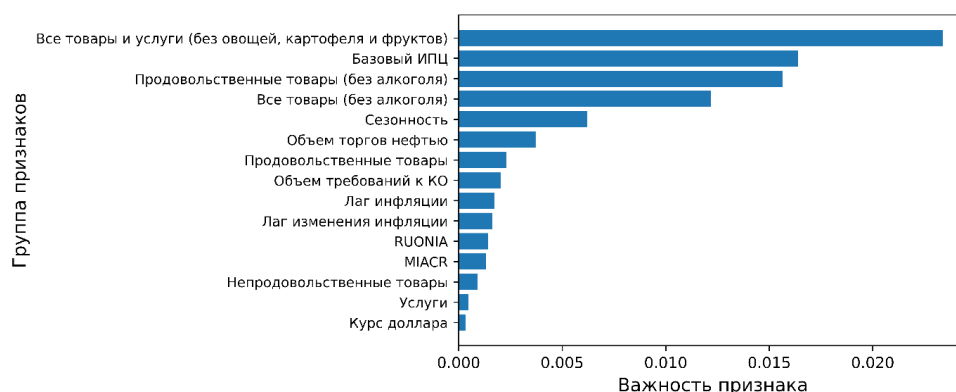
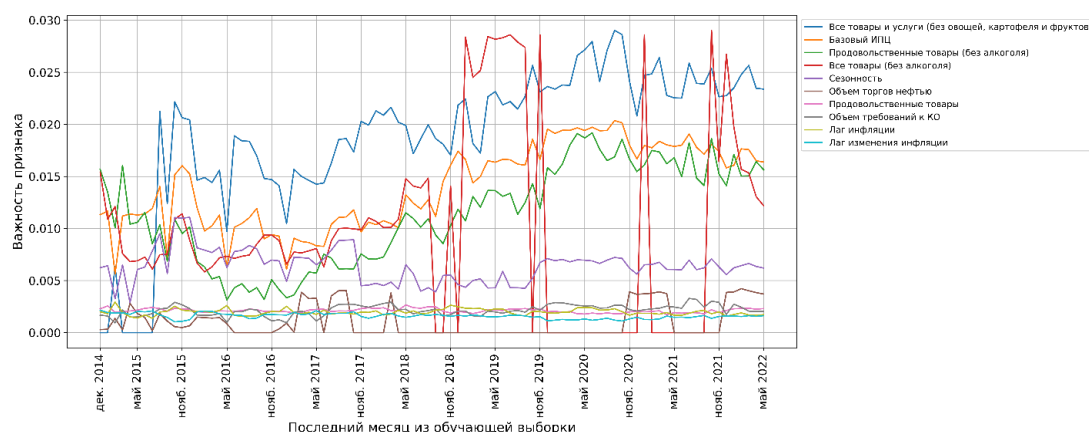


Рис. 3. Динамика важностей признаков (по рейтингу на последний месяц) при расширении обучающей выборки для горизонта прогнозирования $h = 1$ месяц



5.3. Результаты экспериментов с модулями значений SHAP в глобальном контексте

В предыдущем разделе описано, как выявлялись наиболее важные признаки при помощи встроенного в XGBoost подхода – feature importances. Это подход для оценки значимости признаков в глобальном смысле, то есть значимость здесь следует интерпретировать как среднюю величину на совокупном наборе данных, использованном для обучения. SHAP-значения предполагают возможность расчета как в глобальном, так и в локальном смысле. Эксперименты по расчету значений SHAP в глобальном смысле на использованных в работе данных показывают схожую динамику, что и в рассчитанных в том же глобальном смысле показатели в разделе 5.2. В связи с этим результаты упомянутых экспериментов в нашей работе не приводятся.

5.4. Результаты экспериментов с модулями значений SHAP в локальном контексте

В разделах 5.2 и 5.3 описано, каким образом наиболее важные признаки выявлялись с точки зрения оценки значимости в глобальном смысле. В этом разделе описаны результаты альтернативного инструмента определения наиболее значимых факторов – локального метода SHAP. Этот подход локальный в том смысле, что отражает логику модели локально, в окрестности текущей точки, из которой строится прогноз. Так, у нас появляется возможность сравнить 2 подхода к оценке значимостей признаков.

В целом можно отметить, что в наиболее значимые признаки снова попали и макроэкономические факторы, и микрокомпоненты региональной инфляции. При этом на горизонте 1 и 3 месяца на прогноз в основном влияет инерция, то есть лаги и микрокомпоненты инфляции.

На горизонте 6 месяцев (рис. 18 в Приложении В) и 12 месяцев (рис. 8) в список важных показателей включает больше макрофакторов. Среди них процентные ставки и иные переменные денежного рынка, отражающие изменение денежно-кредитных условий; обменный курс и объем торгов нефтью; а также оборот розничной торговли, характеризующий спрос (признак «Оборот торговли РФ»). Эти признаки проявляются все больше с увеличением горизонта прогнозирования, так как они находят свое отражение в инфляции только с некоторыми лагами. Также значимой группой признаков для всех периодов оказалась «Сезонность» (включает порядковый номер месяца, недели и иные характеристики временного периода), что говорит о том, что модель уловила сезонность.

Методика визуального представления полностью аналогична описанной в предыдущем разделе. В этом разделе приведены примеры для горизонта прогнозирования $h = 1$ и 12 месяцев, а также приведен углубленный анализ значимостей признаков в отдельные периоды времени. Визуализация для остальных горизонтов прогнозирования (3 и 6 месяцев) приведены в Приложении В.

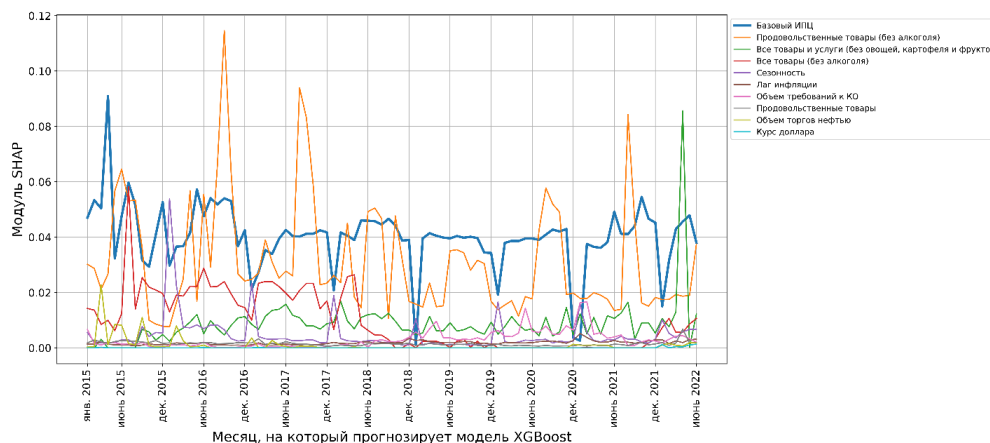
В июне 2022 г. (последней точке нашей выборки) самыми большими модулями значений SHAP для модели с горизонтом прогнозирования $h = 1$ месяц (рис. 4) обладает группа «Базовый ИПЦ», а также группы, характеризующие продовольственные товары, лаги инфляции и сезонность. Кроме того, для месячного горизонта прогнозирования в топ-10 признаков по модулю значения SHAP попадает предиктор, характеризующий изменения курса доллара США к рублю и объем торгов нефтью.

Рис. 4. Значения 20 наибольших модулей SHAP для прогноза на июнь 2022 г. с горизонтом прогнозирования $h = 1$ месяц



Для глубины прогнозирования $h = 1$ месяц характерно резкое изменение векторов модулей SHAP в январях (рис. 5). В течение года набор товаров, которые входят в потребительскую корзину, отслеживаемую Росстатом, неизменен. Фиксированными остаются также веса этих товаров и услуг в ИПЦ (вес рассчитывается как доля расходов на определенный товар в расходах потребителей). Однако из-за того, что потребление тех или иных товаров и услуг изменяется в домохозяйствах, их веса в расходах также меняются. Некоторые товары вообще уходят из потребления (например, Росстат уже много лет не наблюдает цены на DVD-диски). В связи с этим раз в год статистический орган корректирует состав индекса²². Это, наряду с сезонностью, может быть одним из факторов, объясняющих подобный эффект.

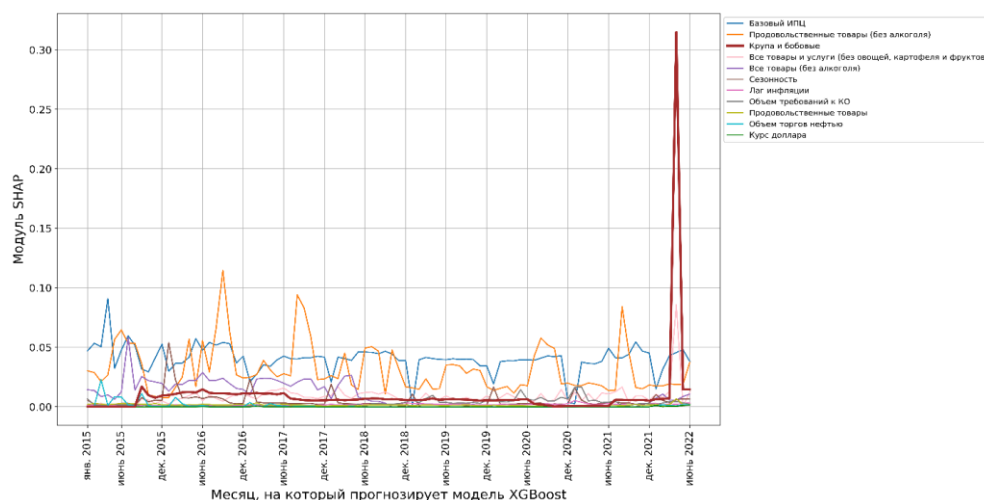
Рис. 5. Динамика модулей SHAP (по рейтингу на последний месяц) для горизонта прогнозирования $h = 1$ месяц



Другим интересным наблюдением для прогноза на месяц вперед является высокое влияние на прогноз компоненты инфляции «крупы и бобовые». Ощущение неопределенности подталкивало домохозяйства в апреле 2022 г. к формированию запасов продуктов длительного хранения. Резкий рост модуля SHAP для этого компонента (рис. 6) демонстрирует то, как модель может неявно учитывать потребительское поведение при прогнозировании инфляции. Действительно, можно считать, что желание потребителей создать запасы продуктов на фоне неопределенности в экономике являлось хорошим предиктором инфляции в апреле 2022 года.

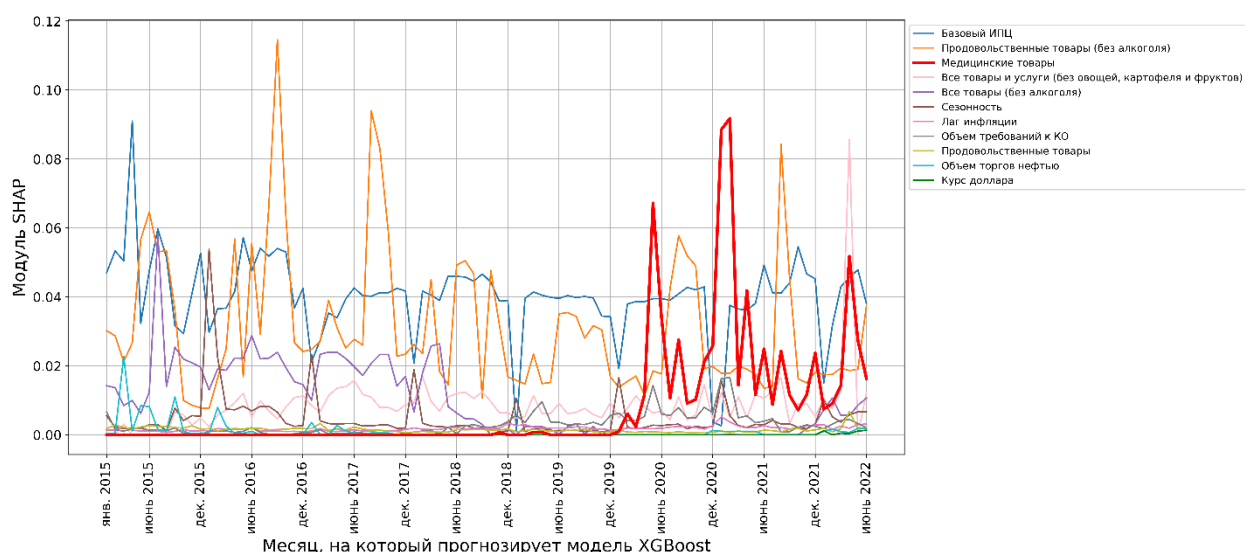
²² [Как Росстат считает инфляцию?](#)

Рис. 6. Всплеск влияния компонента «Крупа и бобовые» в апреле 2022 г. на итоговый прогноз инфляции, глубина прогнозирования $h = 1$ месяц



Для горизонта прогнозирования $h = 1$ месяц также можно заметить резкое увеличение модулей SHAP для компонента «медицинские товары» примерно с начала 2020 г. (рис. 7), что совпадает со временем начала пандемии COVID-19.

Рис. 7. Зависимость степени влияния компонента «Медицинские товары» от времени (глубина прогнозирования $h = 1$ месяц) на итоговую инфляцию



Пандемия привела к росту спроса на медицинское оборудование и препараты, а также вызвала проблемы в снабжении и логистике, что повлияло на уровень цен, способствовало увеличению инфляции в этом потребительском сегменте и стало хорошим предиктором для будущих значений инфляции, как прокси к периоду пандемии.

Для горизонта прогнозирования, равного 1 году ($h = 12$ месяцев), в число 20 предикторов с самыми большими модулями SHAP в июне 2022 г. попадают сразу несколько признаков, характеризующих состояние денежного рынка: признаки групп RUONIA и MIACR, а также группы «Объем требований к КО». Как было отмечено выше, на данном периоде накопленный эффект от изменения основного инструмента денежно-кредитной политики, с учетом всех лагов, начинает в полной мере транслироваться в совокупный спрос и, соответственно, инфляцию.

Рис. 8. Значения 20 наибольших модулей SHAP для прогноза на июнь 2022 г. с горизонтом прогнозирования $h = 12$ месяцев

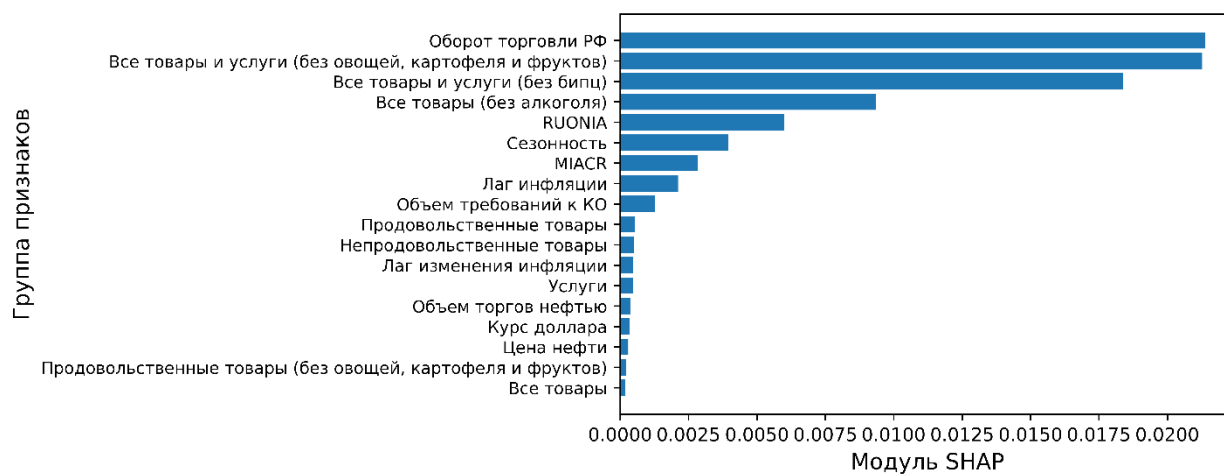
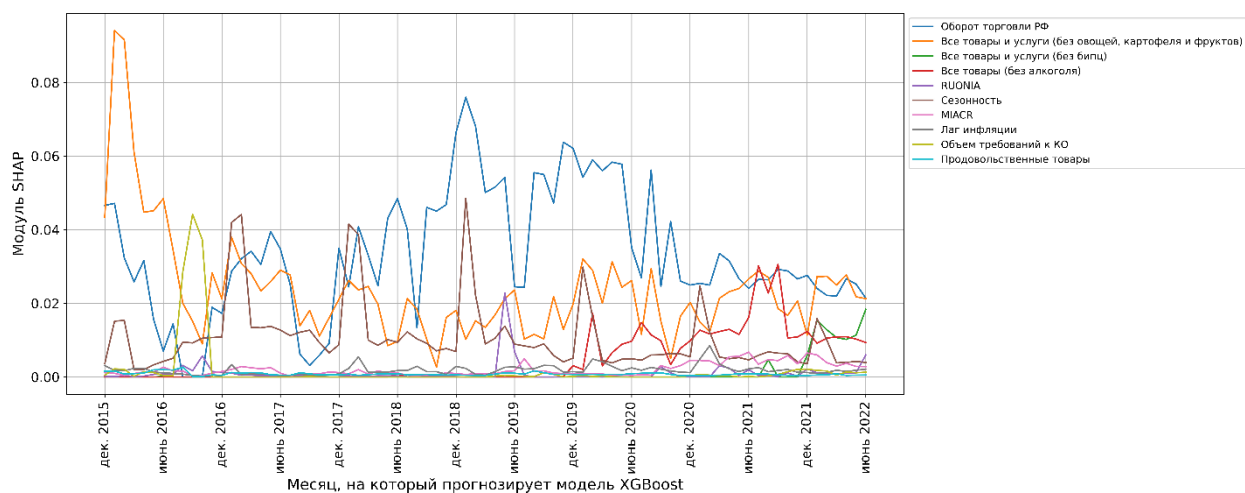


Рис. 9. Динамика модулей SHAP (по рейтингу на последний месяц) для горизонта прогнозирования $h = 12$ месяцев



6. Заключение

Прогнозирование инфляции – важная задача для центральных банков. Понимание будущих траекторий темпов прироста цен необходимо для принятия решений по денежно-кредитной политике в условиях существования лагов ДКП и инерционности макропоказателей, а также для коммуникации этих решений в публичное пространство. В последнее время все более активно обсуждается возможность применения методов машинного обучения для построения прогнозов. При этом модели, которые используются центральными банками должны, во-первых, давать точные прогнозы, а во-вторых – возможность их интерпретировать. На текущем этапе развития методы машинного обучения с первой задачей справляются лучше, чем со второй.

Наша работа вносит вклад в литературу, посвященную точности и интерпретируемости прогнозов, полученных при помощи методов машинного обучения. В исследовании приведены результаты работы таких алгоритмов машинного обучения, как бустинг над решающими деревьями (XGBoost, GradientBoostRegressor), Random Forest, Elastic Net на примере решения задачи по прогнозированию инфляции Ставропольского края. Используются ежемесячные данные с 2002 по 2022 г., которые охватывают множество различных шоков. Эти методы были сопоставлены с рядом эконометрических моделей. Показано, что XGBoost дает наименьшие ошибки прогноза среди примененных методов машинного обучения, статистически незначимо превосходя по этому показателю классические эконометрические методы.

Для того чтобы понять, за счет каких факторов формируется динамика прогноза, мы использовали встроенные в XGBoost декомпозиции и декомпозиции Шепли с представлением анализа значимости предикторов. Показано, что для разных горизонтов значение имели как макроэкономические факторы, например, процентные ставки, показатели денежного рынка, спрос, обменный курс и так далее, так и микрокомпоненты инфляции, динамика которых наибольшим образом характеризовала изменения совокупного показателя региональной инфляции. Этот подход к интерпретируемости может быть полезен как дополнительный инструмент для исследователя в выявлении возможных неочевидных взаимосвязей между инфляцией и применяемыми предикторами. Дальнейшие исследования могут быть посвящены как улучшению качества прогнозирования, так и проработке новых подходов, позволяющих оценивать, что именно модель может рассказать о данных, на которых обучалась, помимо точности прогноза. Также перспективным видится обогащение данных более широким спектром макро- и микропоказателей и изучение степени влияния категорий переменных, размеченных экспертом, на итоговый прогноз, по аналогии с методами, изложенными в (Renard et al., 2019).

Список литературы

1. Байбуза И. (2018). Прогнозирование инфляции с помощью методов машинного обучения. Деньги и кредит, 4, 42–59.
2. Грищенко В., Крылов И. (2024). Новые подходы к измерению, изучению и прогнозированию цен: обзор совместного семинара Банка России, РЭШ и НИУ «Высшая школа экономики». Деньги и кредит, 83(2), 92–111.
3. Латыпов Р., Ахмедова Е., Постолиит Е., Микитчук М. (2024). Прогнозирование компонент инфляции методами машинного обучения. Деньги и кредит, 83(3), 23–44.
4. Могилат А. (2017). Обзор основных каналов трансмиссионного механизма денежно-кредитной политики и инструментов их анализа в Банке России. Деньги и кредит, 9, 3–9.
5. Павлов Е. (2020). Прогнозирование инфляции в России с помощью нейронных сетей. Деньги и кредит, 79(1), 57–73.
6. Семитуркин О., Шевелев А. (2022). Прогнозирование региональной инфляции с помощью методов машинного обучения на примере макрорегиона Сибирь. Банк России.
7. Семитуркин О., Шевелев А. (2023). Корректное сравнение предиктивных свойств моделей машинного обучения на примере прогнозирования инфляции в Сибири. Деньги и кредит, 1, 87–103.
8. Стырин К. (2019). Прогнозирование инфляции в России методов динамического усреднения моделей. Деньги и кредит, 78(1), 3–18.
9. Шибанов О. К. (2024). Предсказания инфляции при помощи ML: достижения последних лет. Банк России.
10. Шибанов О. К. (2024). Уроки для центральных банков: инфляция 2021–2023 гг. Журнал Новой экономической ассоциации, 1(62), 240–245.
11. An J., Mikhaylov A., Moiseev N. (2019). Oil price predictors: machine learning approach. International Journal of Energy Economics and Policy, 9(5), 1–6.
12. Bernanke B. S. (2024). Forecasting for monetary policy making and communication at the Bank of England: a review. Bank of England.
13. Catania L., Nonejad N. (2018). Dynamic model averaging for practitioners in economics and finance: the eDMA package. Journal of Statistical Software, 84(11), 1–39.
14. Chakraborty C., Joseph A. (2017). Machine learning at central banks. Bank of England Working Paper, 674.
15. Chen T., Guestrin C. (2016). Xgboost: a scalable tree boosting system. KDD'16: Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, 785–794.
16. Cicceri G., Inserra G., Limosani M. (2020). A machine learning approach to forecast economic recessions - an Italian case study. Mathematics, 8(2), 241.
17. Diebold Francis X., Mariano Roberto S. (1995). Comparing predictive accuracy. Journal of Business & Economic Statistics, 13(3), 253–263.
18. Garcia M., Medeiros M., Vasconcelos G. (2017). Real-time inflation forecasting with high-dimensional models: the case of Brazil. International Journal of Forecasting, 33, 679–693.

19. Gertler M., Karadi P. (2015). Monetary policy surprises, credit costs, and economic activity. *American Economic Journal: Macroeconomics*, 7(1), 44–76.
20. Giacomini R., White H. (2006). Tests of Conditional Predictive Ability. *Econometrica*, 74(6), 1545–1578.
21. Iversen J., Laseen S., Lundvall H., Soderstrom U. (2016). Real-time forecasting for monetary policy analysis: the case of Sveriges Riksbank. *Sveriges Riksbank Research Paper Series*, 142.
22. Kim H., Swanson N. (2018). Mining big data using parsimonious factor, machine learning variable selection and shrinkage methods. *International Journal of Forecasting*, 34(2), 339–354.
23. Lenza M., Moutachaker I., Paredes J. (2023). Density forecasts of inflation: a quantile regression forest approach. *European Central Bank Working Paper Series*, 2830.
24. Lundberg, S., Lee S. (2017). A unified approach to interpreting model predictions, 31st Conference on Neural Information Processing Systems (NIPS 2017), Long Beach, CA, USA, 4768–4777.
25. Mamedli M., Shibitov D (2021). Forecasting Russian CPI with data vintages and machine learning techniques. *The Bank of Russia Working Paper Series*, 37.
26. Medeiros M., Vasconcelos G., De Freitas E.H. (2016). Forecasting Brazilian inflation with high dimensional models. *Brazilian Review of Econometrics*, 36(2), 223–254.
27. Medeiros M., Vasconcelos G., Veiga A., Zilberman E. (2019). Forecasting inflation in a data-rich environment: the benefits of machine learning methods. Available at SSRN 3155480.
28. Molna, C. (2025). *Interpretable Machine Learning: A Guide for Making Black Box Models Explainable* (3rd ed.).
29. Monteforti L., Moretti G. (2010). Real time forecasts of inflation: the role of financial variables. *Banca D'Italia Temi di discussione* (working papers), 767.
30. Ozgur O., Akkoc U. (2021). Inflation forecasting in an emerging economy: selecting variables with machine learning algorithms. *International Journal of Emerging Markets*, 17(8), 1889–1908.
31. Raftery, A., Kárný M., Ettler P. (2010). Online prediction under model uncertainty via dynamic model averaging: application to a cold rolling mill. *Technometrics*, 52(1), 52–66.
32. Renard, X., Woloszko N., Aigrain J., Detynieck M. (2019). Concept tree: high-level representation of variables for more interpretable surrogate decision trees, available at: <http://arxiv.org/abs/1906.01297> (Актуальна на 03.03.2025).
33. Smets, F., Wouters R. (2003). An estimated dynamic stochastic general equilibrium model of the euro area. *Journal of the European Economic Association*, 1(5), 1123–1175.
34. Stock J.H., Watson M.W. (1999). Forecasting inflation. *Journal of Monetary Economics*, 44(2), 293–335.
35. Stock J.H., Watson MW. (2007). Why has US inflation become harder to forecast? *Journal of Money, Credit and Banking*, 39(s1), 3–33.
36. TechCrunch (2015), CrunchBase Database, available at: <https://www.crunchbase.com> (Актуальна на 03.03.2025).
37. Theoharidis A., Guillen D.A., Lopes, H. (2023). Deep learning models for inflation forecasting. *Applied Stochastic Models in Business and Industry*, 39(3), 447–470.

38. Tibshirani R. (1995). Regression shrinkage and selection via the Lasso. *Journal of the Royal Statistical Society: Series B (Methodological)*, 58(1), 267–288.
39. Wang Y., Wang B., Zhang X. (2012). A new application of the support vector regression on the construction of financial conditions index to CPI prediction. *Procedia Computer Science*, 9, 1263–1272.
40. Zou H. (2006). The adaptive lasso and its oracle properties. *Journal of the American Statistical Association, Theory and Methods*, 101(476), 1418–1429.
41. Zou H., Hastie T. (2005). Regularization and variable selection via the elastic net. *Journal of the Royal Statistical Society: Series B (Methodology)*, 67(2), 301–320.

Приложение

Приложение А. Результаты теста Диболда – Мариано

Результаты теста Диболда – Мариано – р-значения для сравнения прогнозов XGBoost с прогнозами других моделей (RMSE).

Метод	h = 1	h = 3	h = 6	h = 12
Random Forest	0,724	0,805	0,081	0,320
Elastic Net	0,205	0,009	0,002	0,106
GradientBoostRegressor	0,557	0,196	0,381	0,090
Arima	0,376	0,065	0,053	0,003
DMA	0,354	0,115	0,047	0,526
DMS	0,343	0,096	0,029	0,111
BMA	0,704	0,056	0,004	8,54E-19
BMS	0,636	0,093	0,002	8,24E-05
DMA – без компонентов	0,272	0,107	0,135	0,495
DMS – без компонентов	0,051	0,057	0,057	0,064
BMA – без компонентов	0,634	0,046	0,017	0,123
BMS – без компонентов	0,566	0,011	0,012	0,049
AR(2)	0,969	0,056	0,001	0,188
TVP-AR(2)	0,494	0,133	0,225	0,497
DMA-AR(2)	0,441	0,172	0,237	0,597
UCSV	0,887	0,223	0,850	0,491
TVP-KS	0,672	0,055	6,69E-05	0,108
TVP-KS – без компонентов	0,179	0,073	0,015	0,092

Приложение Б. Визуализация важности признаков в соответствии с встроенным в XGBoost-методом

Рис. 10. 20 наиболее важных признаков для модели, обученной на данных до марта 2022 г. включительно, с горизонтом прогнозирования $h = 3$ месяца

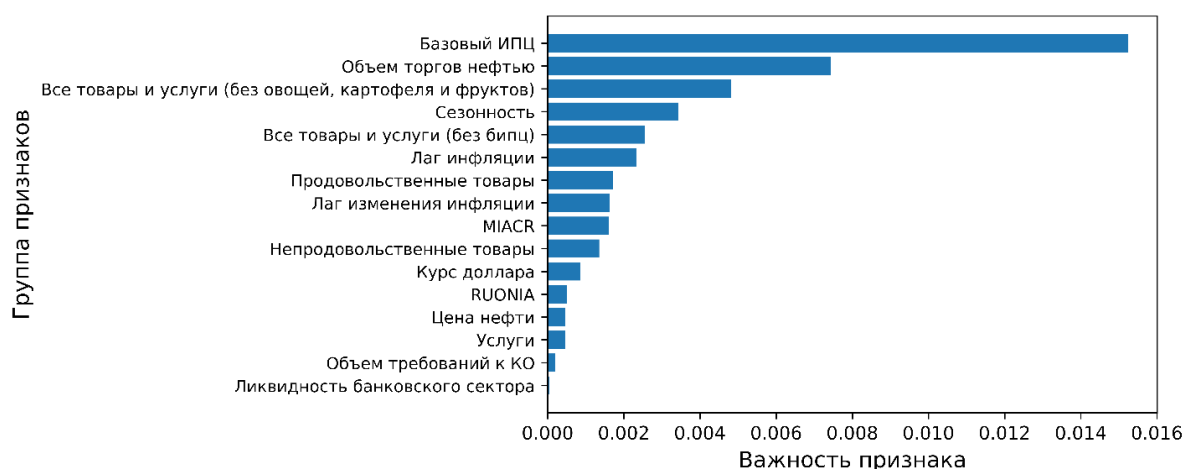


Рис. 11. Динамика важностей признаков (по рейтингу на последний месяц) при расширении обучающей выборки для горизонта $h = 3$ месяца

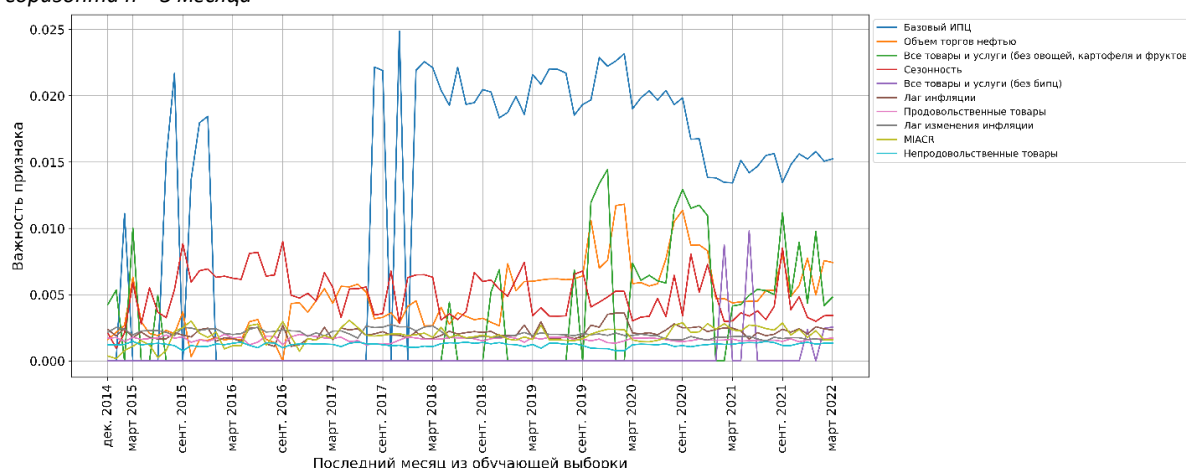


Рис. 12. 20 наиболее важных признаков для модели, обученной на данных до декабря 2021 г. включительно, с горизонтом прогнозирования $h = 6$ месяцев

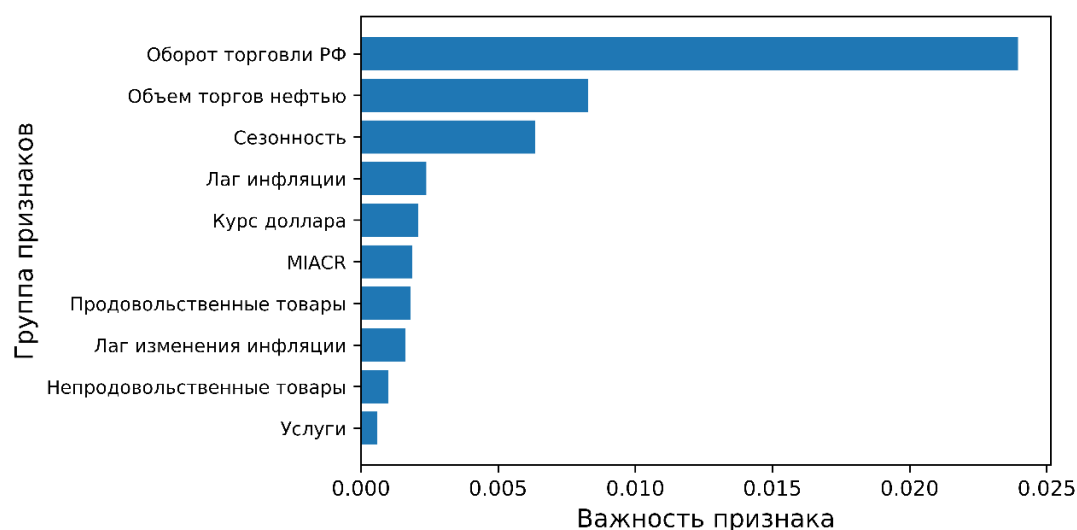


Рис. 13. Динамика важностей признаков (по рейтингу на последний месяц) при расширении обучающей выборки для горизонта прогнозирования $h = 6$ месяцев

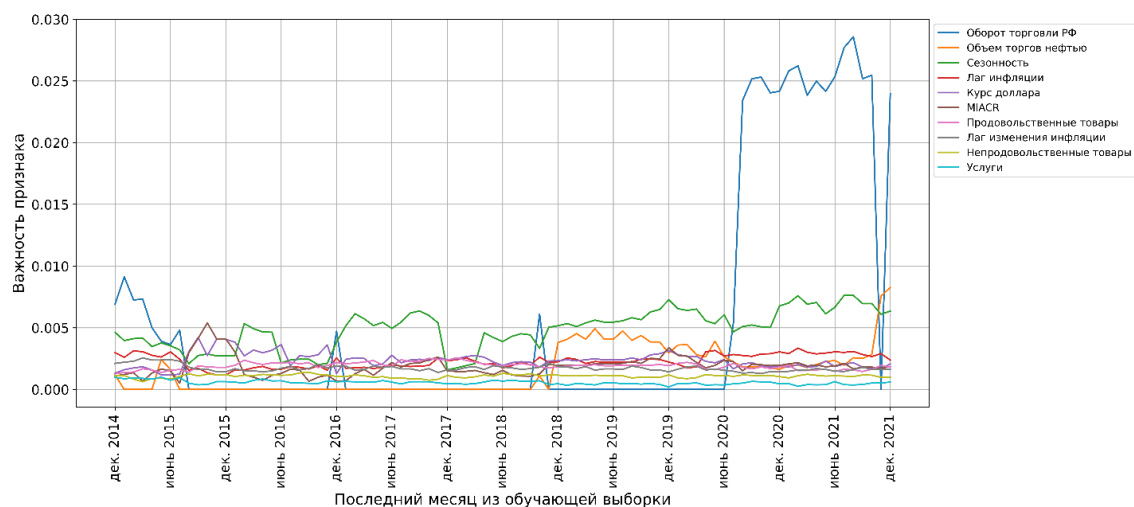


Рис. 14. 20 наиболее важных признаков для модели, обученной на данных до июня 2021 г. включительно, с горизонтом прогнозирования $h = 12$ месяцев

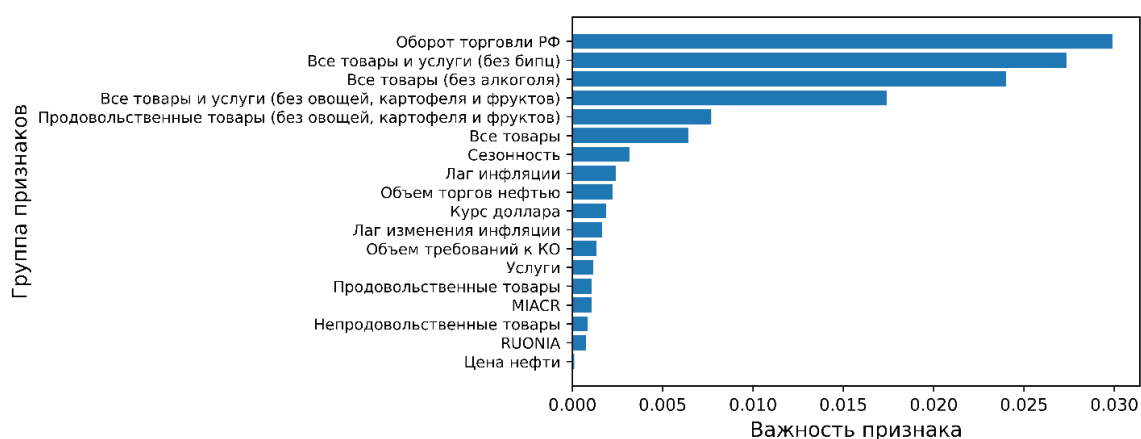
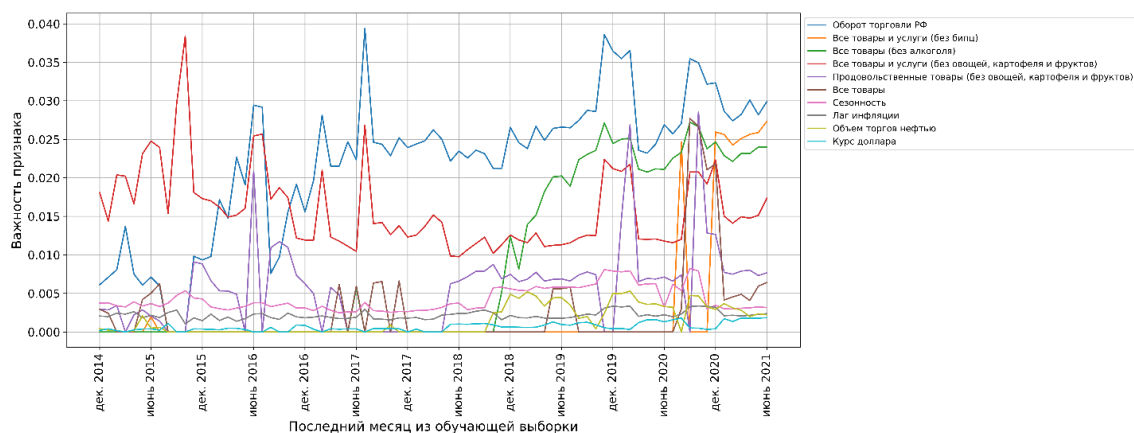


Рис. 15. Динамика важностей признаков (по рейтингу на последний месяц) при расширении обучающей выборки для горизонта прогнозирования $h = 12$ месяцев



Приложение В. Визуализация важности признаков в соответствии с методом SHAP

Рис. 16. Значения 20 наибольших модулей SHAP для прогноза на июнь 2022 г. с горизонтом прогнозирования $h = 3$

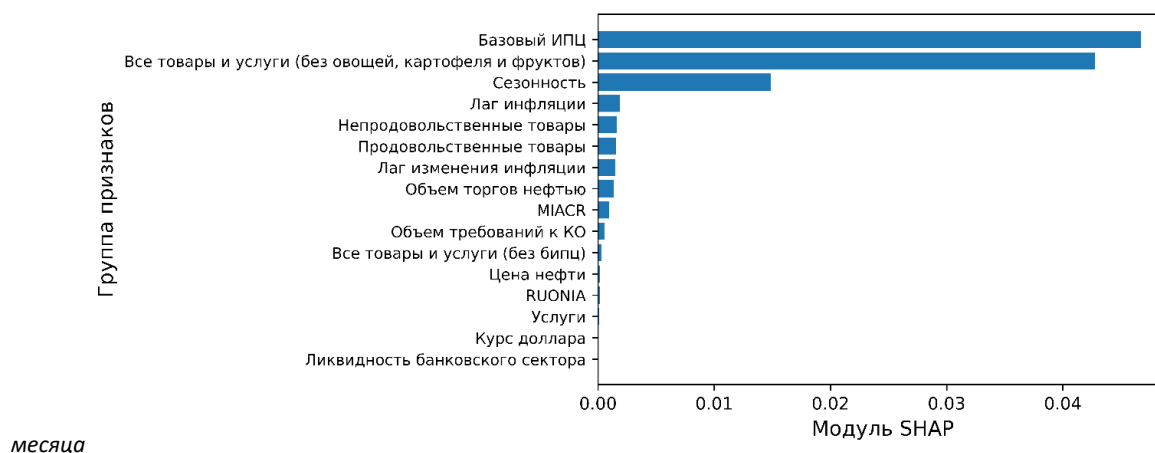


Рис. 17. Динамика модулей SHAP (по рейтингу на последний месяц) для горизонта прогнозирования $h = 3$ месяца

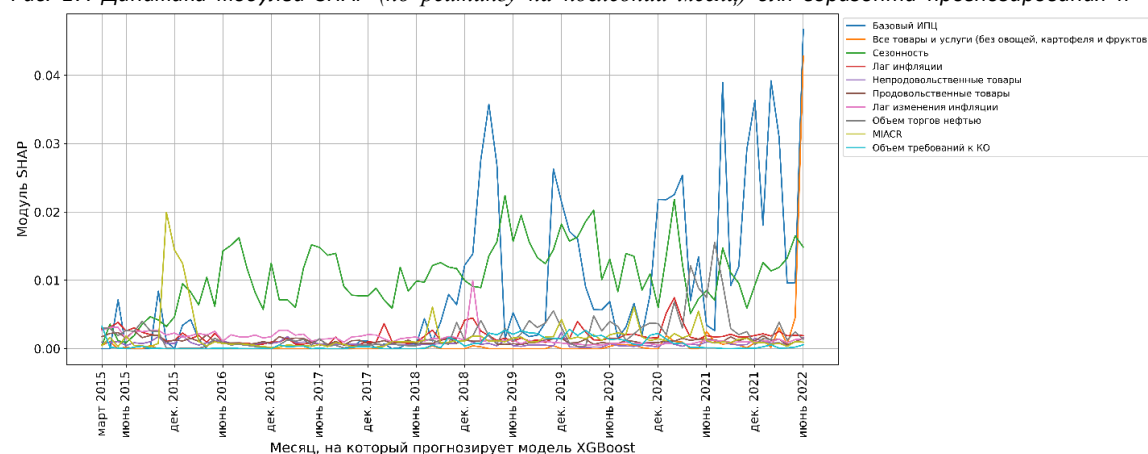


Рис. 18. Значения 20 наибольших модулей SHAP для прогноза на июнь 2022 г. с горизонтом прогнозирования $h = 6$ месяцев

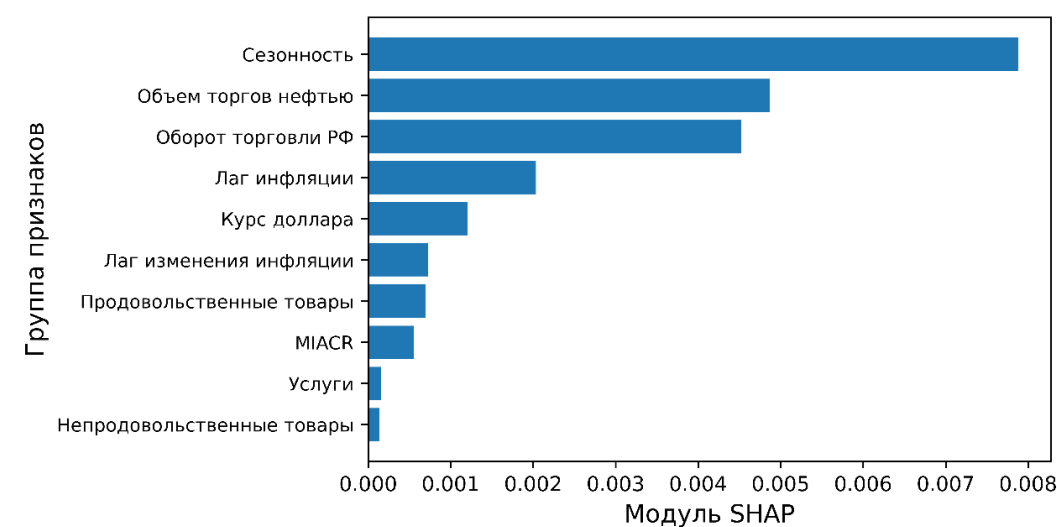
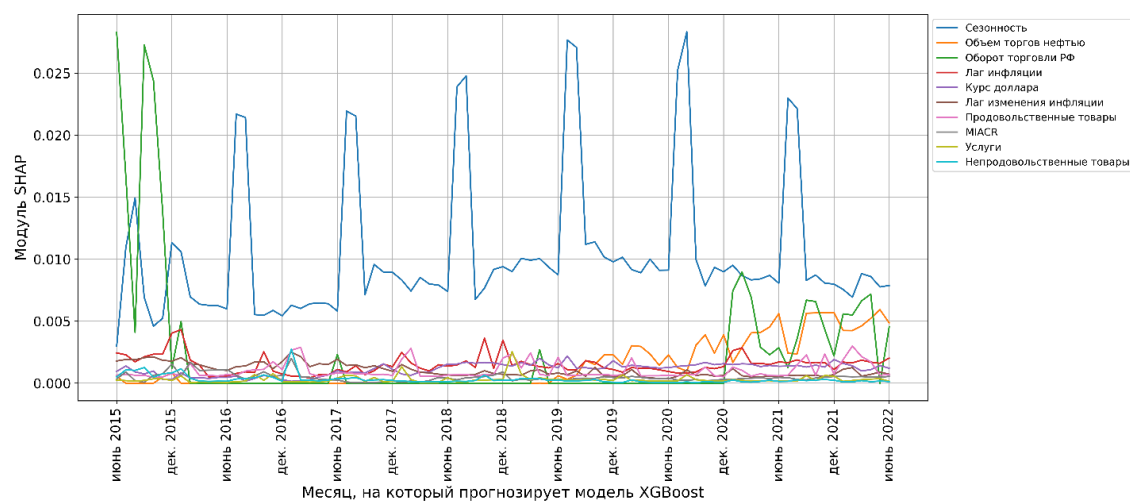


Рис. 19. Динамика модулей SHAP (по рейтингу на последний месяц) для горизонта прогнозирования $h = 6$ месяцев



Приложение Г. Описание признаков, использованных при обучении моделей

[Ссылка](#) на файл с описанием признаков.