

РАЗРАБОТКА МЕТОДОВ И АЛГОРИТМОВ РЕШЕНИЯ ЗАДАЧ УПРАВЛЕНИЯ В ОРГАНИЗАЦИОННЫХ СИСТЕМАХ ТОРГОВЫХ КОМПАНИЙ

DOI: 10.36724/2072-8735-2026-20-3-43-52

Леохин Юрий Львович,*Московский технический университет связи и информатики, Москва, Россия, y.l.leokhin@mtuci.ru***Дымкова Светлана Сергеевна,***Московский технический университет связи и информатики, Москва, Россия, s.s.dymkova@mtuci.ru***Фатхулин Тимур Джалилевич,***Московский технический университет связи и информатики, Москва, Россия, t.d.fatkhulin@mtuci.ru***Зозуля Ирина Сергеевна,***Московский технический университет связи и информатики, Москва, Россия***Manuscript received** 15 December 2025;**Accepted** 25 February 2026**Ключевые слова:** метод, алгоритм, задача управления, объемы продаж, организационная система торговых компаний, прогнозирование, показатели, эффективность

В статье показана разработка и экспериментальная проверка модифицированного метода на основе алгоритма XGBoost, применяемого для прогнозирования объемов продаж. Цель работы – повышение точности методов и алгоритмов прогнозирования объемов продаж, используемых в организационных системах торговых компаний. Актуальность работы обусловлена тем, что одной из основных проблем предметной области продаж является наличие шумов и аномалий в данных. При прогнозировании это приводит к снижению точности результатов и возникновению эффекта переобучения. Настройка существующих параметров в алгоритмах прогнозирования при управлении в организационных системах торговых компаний не всегда является решением данных проблем и приводит к необходимости дополнительных модификаций для уточнения прогнозов. В работе разработан метод модификации алгоритма градиентного усиления, использующий корректировки расчета градиента с помощью оценок аномальности образцов набора данных методом леса изоляции. Также использованы дополнительные гиперпараметры и метод остановки обучения при получении информации о начале переобучения. В ходе исследования были рассмотрены подходы к регуляризации и снижению влияния аномалий в данных на примере алгоритма XGBoost. Внесены дополнительные модификации в предлагаемый метод с применением кросс-валидации. Все модификации протестированы в разных комбинациях на разных наборах данных. Таким образом выявлена закономерность повышения точности при обучении на кросс-валидации с применением разработанного метода взвешивания аномалий. Практическая значимость работы заключается в разработке универсального подхода к модификации алгоритма XGBoost, который может быть применен для повышения точности прогнозирования продаж различных категорий товаров при управлении в организационных системах торговых компаний.

Информация об авторах:*Леохин Юрий Львович, Московский технический университет связи и информатики, профессор, д.т.н., orcid.org/0000-0003-3321-4497**Дымкова Светлана Сергеевна, Московский технический университет связи и информатики, к.т.н., orcid.org/0000-0003-1945-9850**Фатхулин Тимур Джалилевич, Московский технический университет связи и информатики, кафедра "Интеллектуальный анализ данных", доцент, к.т.н., orcid.org/0000-0003-0998-1055**Зозуля Ирина Сергеевна, Московский технический университет связи и информатики, кафедра "Математическая кибернетика и информационные технологии", магистрант, orcid.org/0009-0006-3445-8070***Для цитирования:***Леохин Ю.Л., Дымкова С.С., Фатхулин Т.Д., Зозуля И.С. Разработка методов и алгоритмов решения задач управления в организационных системах торговых компаний // Т-Comm: Телекоммуникации и транспорт. 2026. Том 20. №3. С. 43-52.***For citation:***Yu. L. Leokhin, S. S. Dymkova, T. D. Fatkhulin, I. S. Zozulya, "Development of Methods and Algorithms for Solving Management Problems in Organizational Systems of Trading Companies," T-Comm, 2026, vol. 20, no. 3, pp. 43-52. (in Russian)*

Введение

Среди ключевых задач управления в организационных системах торговых компаний выделяют своевременное прогнозирование объемов продаж товаров и услуг [5, 11]. В настоящее время задача управления объемами продаж различных товаров, в том числе телекоммуникационного или компьютерного оборудования, имеет первостепенное значение [3, 22-24]. Это связано как с резко возросшим спросом на такое оборудование из-за развития систем искусственного интеллекта, так и с санкционным давлением отдельных стран [7, 23-25]. В результате управление в организационных системах торговых компаний, которые подразумевают под собой формальную структуру компаний, происходящие в них процессы, реализуемые политики и процедуры, регулирующие деятельность торговых организаций, становится значительно затруднено. Оптимизация управления даст возможность повысить эффективность предприятий [26].

Существующие методы и алгоритмы прогнозирования различных показателей [1, 2, 14, 27] не всегда дают максимально эффективный результат [8, 14]. Это относится и к прогнозированию таких динамически изменяющихся величин, как пропускная способность сетей, построенных на определенном типе оборудования [6, 7, 15-21], так и экономических показателей торговых компаний, напрямую влияющих на результативность управления в этих организационных системах [19, 20]. В работах [8, 12-14] показано, что разработка модифицированного метода и алгоритма решения задач управления спросом в организационных системах телекоммуникационных компаний позволило значительно улучшить показатели таких систем.

Еще одним примером необходимости разработки новых методов и алгоритмов решения задач управления в разнообразных организационных системах является необходимость прогнозирования динамики изменения криптовалют [6]. В работе [7] показано, что разработка современных методов получения данных и прогнозирования валютных значений курсов криптовалют дает возможность эффективнее управлять организационными системами криптобирж на основе ретроспективной, текущей и экспертной информации.

Все это указывает на актуальность разработки методов и алгоритмов решения задач управления объемами продаж в организационных системах торговых компаний. Настоящее исследование строится на уже проведенном анализе существующих методов и алгоритмов прогнозирования объемов продаж, который позволил выявить наиболее перспективные из них для решения поставленной задачи [5, 11]. В данной работе будет предложен модифицированный вариант метода и алгоритма прогнозирования.

1 Постановка задачи

В современном обществе предоставляют много товаров и услуг. Для прогнозирования разных показателей требуется крайне много сил и затрат, как временных, так и человеческих ресурсов. На помощь приходят методы машинного обучения, которые систематизируют, находят закономерности и прогнозируют показатели, например, объемы продаж.

Среди множества алгоритмов машинного обучения алгоритм градиентного бустинга XGBoost (eXtreme Gradient

Boosting) зарекомендовал себя как один из наиболее эффективных инструментов для решения задач регрессии, в том числе и прогнозирования продаж [4, 5, 12, 13].

Однако, несмотря на свои преимущества, XGBoost обладает и определенными недостатками, главным из которых является склонность к переобучению, особенно при работе с определенными типами данных, имеющими высокий уровень шума, аномалий или недостаточный объем. Это ограничивает его прямое применение и создает предпосылки для дальнейшей модификации и адаптации алгоритма под специфику конкретных задач и наборов данных.

Таким образом, возникает проблема, заключающаяся в необходимости повышения устойчивости и обобщающей способности метода XGBoost для задач прогнозирования продаж за счет разработки и применения комплекса модификаций, направленных на борьбу с переобучением и адаптацией к различным особенностям исходных данных.

2 Определение вариантов модификаций для повышения точности метода XGBoost

По результатам экспериментов [5] оптимальным методом для прогнозирования объемов продаж товаров признан алгоритм градиентного бустинга XGBoost. Этот алгоритм характеризуется высокой точностью, быстротой вычислений и возможностью тонкой настройки параметров [1-2, 13-16], хотя и обладает недостатками такими, как склонность к переобучению и снижению точности на определенных типах данных, что открывает перспективы дальнейшего совершенствования.

Основным критерием при проведении модификации метода градиентного усиления является критерий точности. При этом стоит учитывать факт того, что модификация метода и алгоритма предполагает не только увеличение точности, но и возможное увеличение скорости и сложности построенной модели. Поэтому при выборе модификации стоит минимизировать указанные риски.

При проведении экспериментов и анализе полученных данных были выявлены проблемы переобучения и снижения точности, связанные с недостаточностью обрабатываемых данных, наличием шумов и аномалий. Всё это приводит к искажению обучения [11]. Под шумами понимают случайные колебания продаж, которые не несут полезной информации для обучающего алгоритма.

Модифицировать алгоритм можно несколькими способами. Рассмотрим некоторые из них.

1. Оптимизация гиперпараметров

Гиперпараметры определяют поведение модели. Грамотная настройка может заметно улучшить качество и производительность модели [9, 10]. Рассмотрим ключевые параметры:

- `max_depth`: ограничивает глубину деревьев, предотвращая переусложнение модели. Чем меньше глубина, тем ниже риск переобучения. Обычно оптимальное значение находится в диапазоне от 3 до 10.
- `min_child_weight`: определяет минимальный суммарный вес (вес сумм всех объектов в узле) для формирования узла дерева. Увеличение `min_child_weight` препятствует созданию небольших узлов, снижая переобучение [17].

- `subsample`: регулирует долю объектов, выбираемых случайно для каждого отдельного дерева. Установив `subsample` менее 1, можно уменьшить эффект переобучения и повысить устойчивость модели.

- `colsample_bytree`: контролирует долю признаков, используемых при построении каждого дерева. Установка `colsample_bytree` менее 1 уменьшает размерность пространства признаков, ускоряя обучение и уменьшая риск переобучения.

- `learning_rate`: шаг обучения влияет на скорость и точность модели. Маленькие `learning_rate` замедляют обучение, но снижают вероятность переобучения. Большие шаги увеличивают скорость, но повышают риск скачков в процессе обучения.

2. Регуляризация

Регуляризация помогает бороться с переобучением, добавляя штраф в случае сложной модели. В XGBoost поддерживаются два основных вида регуляризации [8]:

- `alpha` (L1 regularization): L1-регуляризация поощряет редкие признаки, приводя некоторые веса к нулю, тем самым устраняя лишние признаки.

- `lambda` (L2 regularization): L2-регуляризация равномерно распределяет штрафы, уменьшая влияние экстремально высоких весов.

3. Ранняя остановка обучения (Early Stopping)

Если качество модели перестаёт улучшаться, то ранняя остановка останавливает обучение модели. Данный подход позволяет эффективно предотвращать переобучение и экономить время. Для включения ранней остановки необходимо задать аргумент `early_stopping_rounds`.

4. Параллельное обучение

XGBoost изначально поддерживает параллельную обработку, что позволяет существенно ускорить выполнение алгоритма. Параметр `n_jobs` управляет количеством потоков.

Также можно использовать графические ускорители (GPU), установив соответствующий флажок:

```
rgb_model = xgb.XGBRegressor(tree_method='gpu_hist').
```

Этот режим полезен для больших наборов данных, так как позволяет существенно ускорить процесс обучения.

5. Выбор правильного формата данных

Чтобы ускорить работу XGBoost, важно правильно хранить данные. Лучший вариант – конвертировать их во внутренний формат библиотеки, `DMatrix`, который заточен под высокую скорость обработки. Также стоит сохранять исходные данные в эффективных форматах вроде `Parquet` или `ORC` – это уменьшит их объем и ускорит чтение с диска, что положительно скажется на общей производительности.

6. Кросс-валидация

Кросс-валидация – это важный метод оценки качества модели, помогающий определить оптимальные параметры и выявить потенциальные проблемы переобучения. Выполнять кросс-валидацию стоит на небольшом количестве данных [3-9].

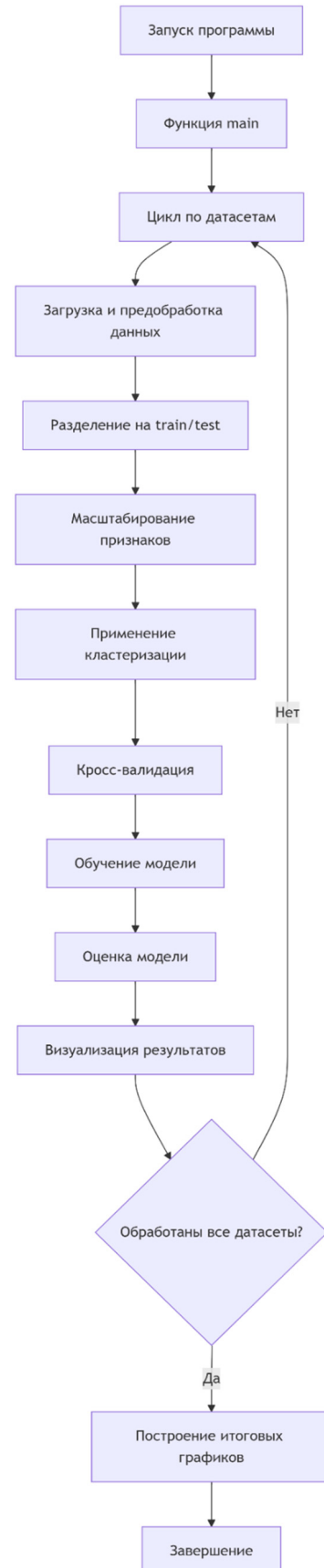


Рис. 1. Блок-схема алгоритма модифицированного метода XGBoost

7. Балансировка классов

Несбалансированные классы приводят к смещению модели в сторону преобладающего класса. Используются следующие техники для борьбы с дисбалансом:

- **Class Weights:** установка весов для классов позволяет сбалансировать важность редких классов.
- **Oversampling:** искусственное увеличение числа представителей редкого класса.
- **Undersampling:** сокращение числа представителей преобладающего класса.

Балансировка классов поможет избежать смещения модели и обеспечить лучшее качество прогнозов.

8. Выявление и устранение лишней корреляции

Коррелированные признаки создают излишнюю избыточность и ухудшают интерпретируемость модели. Для диагностики корреляции строится матрица корреляций, и исключаются признаки с высоким уровнем взаимозависимости.

Уменьшение размерности пространства признаков упрощает модель и ускорит её обучение.

При корректном использовании нескольких методов, возможно улучшить как точность прогнозирования, так и не увеличить скорость обработки информации.

3 Алгоритм модифицированного метода XGBoost

В результате выбора оптимального сочетания методов улучшения алгоритма был получен модифицированный алгоритм, который будет протестирован с применением датасетов, использованных в экспериментах 1-3 [5] (рис. 1). С целью оценки качества прогнозов необходимо вычислить следующие метрики: RMSE, MAE и R^2 [5].

4 Экспериментальная апробация разработанных методов и алгоритмов

При обучении рассматриваемых в исследовании моделей были взяты данные из датасетов открытой платформы «Kaggle» (открытый ресурс: <https://www.kaggle.com/datasets>) [5]. Данные выбирались релевантно рассматриваемой предметной области и согласно решаемой практической задаче. В таблице 1 приведены варианты гиперпараметров, используемых при настройке исследуемых методов (алгоритмов) машинного обучения.

В дальнейшем будут показаны расчеты значения метрик в каждом эксперименте для всех моделей.

Эксперимент №1. Рассмотрим датасет продаж (информация из открытых ресурсов платформы «Kaggle»: <https://www.kaggle.com/datasets/ihelon/coffee-sales>).

Данные представляют собой продажи кофейных напитков (капучино, горячий шоколад, какао, латте, американо и др.) в вендинговых аппаратах. Данные были предобработаны и приведены к единому числовому формату.

Таблица 1

Значения гиперпараметров для применяемых алгоритмов МО

Алгоритм	Наименование гиперпараметра	Тип значения	Диапазон поиска гиперпараметра
UMAP	n_neighbors	int	2 - 10
	n_components	int	2 - 10
	min_dist	float	0,0 - 0,5
KMeans	n_clusters	int	2 - 11
XGBoost	n_estimators	int	200 - 1000, шаг 200
	max_depth	int	1 - 10
	learning_rate	float	0,001 - 1
	subsample	float	0,05 - 1
	colsample_bytree	float	0,05 - 1
XGBoost с регуляризацией L1, L2, gamma	min_child_weight	int	1 - 20
	reg_alpha	float	0 - 10
	reg_lambda	float	0 - 10
	gamma	float	0 - 10
Isolation Forest	n_estimators	int	200 - 1000, шаг 200
	max_samples	float	0,05 - 1
	max_features	float	0,05 - 1
	bootstrap	bool	True, False

Всего в датасете указаны 262 факта продажи в разные временные интервалы. Задача спрогнозировать объёмы продаж в ближайшем будущем.

Таблица 2 демонстрирует работу алгоритмов по различным метрикам. Результаты приведены как на тестовой, так и на обучающей выборках (рис. 2). Ранее при рассмотрении алгоритмов машинного обучения на этом датасете оптимальные результаты показал алгоритм XGBoost.

Далее в таблицах 2-4 рассмотрены оценки изменения точности и показатели результатов на метриках.

Как видно из таблицы 2, модели с CV демонстрируют более сбалансированные результаты между тренировочной и тестовой выборками. Также необходимо отметить позитивное влияние кросс-валидации, которая приводит к лучшим результатам метрик модификаций. При этом традиционный вариант с регуляризацией L1, L2, gamma увеличивает ошибки модели на тестовых данных. Таким образом, этот подход является неэффективным.

В моделях с UMAP по отдельности отмечается катастрофическое ухудшение метрик (R^2 снижается до 0.47), однако в комбинации с Isolation Forest и кросс-валидацией наблюдается некоторое улучшение показателей. Модели с Isolation Forest демонстрируют стабильно хорошие результаты, что свидетельствует о наличии выбросов в данных и эффективности их фильтрации.

Когда в комбинированные подходы добавляем кросс-валидацию, наблюдается повышение показателей точности в сравнении с базовыми модификациями. Следовательно, правильная комбинация предлагаемых модификаций приводит к тому, что они дополняют друг друга.

Таблица 2

Полученные значения работы алгоритмов в моделях в Эксперименте №1

Модель	Метрика	Значения на выборке	
		train	test
<i>XGBoost</i>	<i>MAE</i>	0,0080	0,1375
<i>XGBoost + L1, L2, gamma</i>		0,2120	0,3162
<i>XGBoost+IsolationForest</i>		0,0010	0,1599
<i>XGBoost+UMAP</i>		0,0008	1,5044
<i>XGBoost+IsolationForest+UMAP</i>		0,0006	1,9260
<i>XGBoost+CV</i>		0,0204	0,1335
<i>XGBoost+IsolationForest+CV</i>		0,1390	0,2692
<i>XGBoost+UMAP+CV</i>		0,0349	1,5203
<i>XGBoost+IsolationForest+UMAP+CV</i>		0,0234	2,3413
<i>XGBoost</i>	<i>R²</i>	1,0000	0,9905
<i>XGBoost + L1, L2, gamma</i>		0,9923	0,9852
<i>XGBoost+IsolationForest</i>		1,0000	0,9899
<i>XGBoost+UMAP</i>		1,0000	0,4718
<i>XGBoost+IsolationForest+UMAP</i>		1,0000	0,3121
<i>XGBoost+CV</i>		0,9999	0,9898
<i>XGBoost+IsolationForest+CV</i>		0,9955	0,9861
<i>XGBoost+UMAP+CV</i>		0,9999	0,5498
<i>XGBoost+IsolationForest+UMAP+CV</i>		0,9999	-0,0157
<i>XGBoost</i>	<i>RMSE</i>	0,0126	0,4152
<i>XGBoost + L1, L2, gamma</i>		0,3705	0,5176
<i>XGBoost+IsolationForest</i>		0,0016	0,4282
<i>XGBoost+UMAP</i>		0,0011	3,0915
<i>XGBoost+IsolationForest+UMAP</i>		0,0010	3,5281
<i>XGBoost+CV</i>		0,0384	0,4303
<i>XGBoost+IsolationForest+CV</i>		0,2845	0,5203
<i>XGBoost+UMAP+CV</i>		0,0502	2,8540
<i>XGBoost+IsolationForest+UMAP+CV</i>		0,0320	4,2869

Проанализируем графики метрик для Эксперимента №1. Самый лучший результат получен при использовании XGBoost. Таким образом, эта модель подходит для поставленной задачи. Подходы с UMAP продемонстрировали очень низкие показатели ошибок на тренировочных выборках. Можно сделать вывод о явной тенденции к переобучению, так как на тестовой выборке показатели ошибок экстремально высоки. При этом более сбалансированные результаты выдают модели с кросс-валидацией, которая уменьшает переобучение и эффективна при применении сложных модификаций.

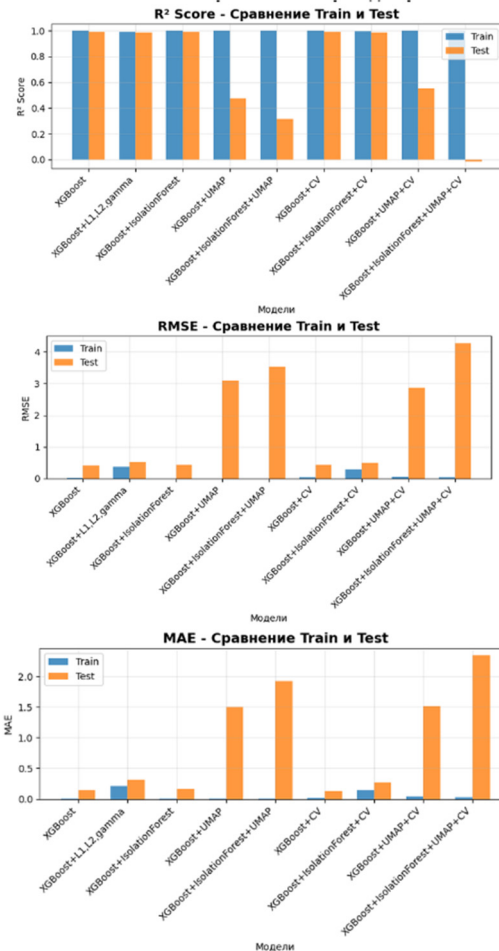


Рис. 2. Значения метрик для Эксперимента №1

Эксперимент №2. Эксперимент проводится на следующем датасете продаж (информация из открытых ресурсов платформы «Kaggle»: <https://www.kaggle.com/datasets/xontoloyo/data-penjualan-zara>).

Данные представляют собой продажи одежды, аксессуаров и иных вещей из магазина Zara. Данные были преобразованы и приведены к единому числовому формату. Всего в датасете указаны 252 факта продажи в разные временные рамки. Задача спрогнозировать объёмы продаж в ближайшем будущем.

В таблице 3 показаны увеличение и уменьшение точности дальнейших модификаций. Оценка производится в сравнении с показателями предыдущего эксперимента, где на этом же датасете оптимальные результаты продемонстрировала модель с алгоритмом XGBoost (рис. 3).

Аналогично предыдущему эксперименту, регуляризация с помощью штрафования L1 и L2 ухудшила показатели точности, однако в данном случае проявился иной эффект - метрики на тренировочных данных значительно ухудшились (R^2 снизился с 0.9997 до 0.9932), что свидетельствует о недостаточной гибкости модели при сильной регуляризации. Следовательно, настраивать параметры регуляризации надо с применением более тонкого подхода.

Таблица 3

Полученные значения работы алгоритмов в моделях Эксперимента №2

Модель	Метрика	Значения на выборке	
		train	test
<i>XGBoost</i>	<i>MAE</i>	9,4271	674,4052
<i>XGBoost + L1, L2, gamma</i>		42,4052	667,6066
<i>XGBoost+IsolationForest</i>		0,2391	771,7242
<i>XGBoost+UMAP</i>		1,2992	728,8567
<i>XGBoost+IsolationForest+UMAP</i>		1,7943	635,1774
<i>XGBoost+CV</i>		535,9774	568,9708
<i>XGBoost+IsolationForest+CV</i>		518,4127	575,4599
<i>XGBoost+UMAP+CV</i>		554,0367	583,7050
<i>XGBoost+IsolationForest+UMAP+CV</i>		530,1759	573,8359
<i>XGBoost</i>		<i>R²</i>	0,9997
<i>XGBoost + L1, L2, gamma</i>	0,9932		-0,5012
<i>XGBoost+IsolationForest</i>	1,00		-0,8270
<i>XGBoost+UMAP</i>	1,00		-0,7925
<i>XGBoost+IsolationForest+UMAP</i>	1,00		-0,4902
<i>XGBoost+CV</i>	0,1763		-0,0399
<i>XGBoost+IsolationForest+CV</i>	0,1903		-0,0508
<i>XGBoost+UMAP+CV</i>	0,1433		-0,0185
<i>XGBoost+IsolationForest+UMAP+CV</i>	0,1919		-0,0552
<i>XGBoost</i>	<i>RMSE</i>		12,7994
<i>XGBoost + L1, L2, gamma</i>		57,6514	834,4577
<i>XGBoost+IsolationForest</i>		0,3392	920,5702
<i>XGBoost+UMAP</i>		1,7389	911,8384
<i>XGBoost+IsolationForest+UMAP</i>		2,3661	831,3951
<i>XGBoost+CV</i>		634,9137	694,5244
<i>XGBoost+IsolationForest+CV</i>		621,2661	698,1325
<i>XGBoost+UMAP+CV</i>		647,5251	687,3219
<i>XGBoost+IsolationForest+UMAP+CV</i>		620,6437	699,6118

Каждая по отдельности модели с Isolation Forest и UMAP показали идеальные результаты на тренировочных данных ($R^2 = 1.0000$), однако на тестовой выборке продемонстрировали катастрофическое ухудшение метрик (R^2 от -0.79 до -0.83), что указывает на сильное переобучение. Их совместное применение несколько улучшило ситуацию ($R^2 = -0.49$), но не позволило достичь положительных значений, что может свидетельствовать о неадекватной обработке аномалий при текущем разделении данных.

Применение кросс-валидации с XGBoost существенно повысило точность прогнозов на тестовой выборке – модель XGBoost+CV показала $R^2 = -0.0399$ против $R^2 = -0.5753$ у базового XGBoost, что подтверждает эффективность кросс-валидации для снижения переобучения.

Однако добавление Isolation Forest к кросс-валидации незначительно ухудшило результат ($R^2 = -0.0508$), что может

проявляться в необходимости более тонкой настройки параметров обнаружения аномалий. Дальнейшее добавление UMAP к кросс-валидации улучшило результат (XGBoost+UMAP+CV с $R^2 = -0.0185$), а модель XGBoost+IsolationForest+UMAP+CV показала $R^2 = -0.0552$.

Наилучший результат среди всех моделей продемонстрировала XGBoost+UMAP+CV, что свидетельствует о положительном влиянии снижения размерности в комбинации с кросс-валидацией. Одновременное снижение ошибок на тренировочной и тестовой выборках у моделей с кросс-валидацией по сравнению с базовыми подходами подтверждает эффективность методов борьбы с переобучением, хотя ни одной модели не удалось достичь положительного значения R^2 на тестовой выборке, что указывает на необходимость дополнительного анализа данных и feature engineering (рис. 3).

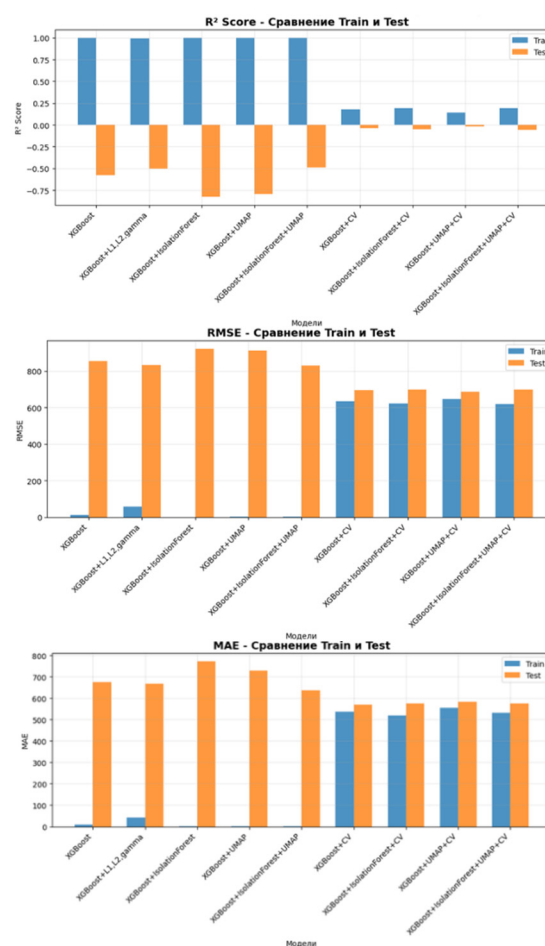


Рис. 3. Значения метрик для Эксперимента №2

Эксперимент №3. Эксперимент проводится на следующем датасете (информация из открытых ресурсов платформы «Kaggle»: <https://www.kaggle.com/datasets/atharvasoundankar/chocolate-sales>). Данные представляют собой продажи шоколадных изделий (белого, молочного и темного шоколада). Данные были предобработаны и приведены к единому числовому формату. Всего в датасете указаны 1094 факта продажи в разные временные рамки. Задача спрогнозировать объемы продаж в ближайшем будущем. Полученные значения по результатам работы алгоритмов как на тестовых, так и на обучающих выборках отражены в таблице 4.

Таблица 4

Полученные значения работы алгоритмов в моделях для Эксперимента №3

Модель	Метрика	Значения на выборке	
		train	test
<i>XGBoost</i>	<i>MAE</i>	681,002	4003,608
<i>XGBoost + L1, L2, gamma</i>		1132,416	3819,624
<i>XGBoost+IsolationForest</i>		181,073	4052,540
<i>XGBoost+UMAP</i>		60,437	3629,583
<i>XGBoost+IsolationForest+UMAP</i>		49,958	3832,292
<i>XGBoost+CV</i>		3223,656	3221,571
<i>XGBoost+IsolationForest+CV</i>		3194,351	3222,103
<i>XGBoost+UMAP+CV</i>		3217,339	3253,444
<i>XGBoost+IsolationForest+UMAP+CV</i>		3180,561	3227,209
<i>XGBoost</i>	<i>R²</i>	0,952	-0,594
<i>XGBoost + L1, L2, gamma</i>		0,872	-0,488
<i>XGBoost+IsolationForest</i>		0,996	-0,680
<i>XGBoost+UMAP</i>		0,999	-0,316
<i>XGBoost+IsolationForest+UMAP</i>		0,999	-0,446
<i>XGBoost+CV</i>		0,041	-0,005
<i>XGBoost+IsolationForest+CV</i>		0,047	-0,002
<i>XGBoost+UMAP+CV</i>		0,054	-0,018
<i>XGBoost+IsolationForest+UMAP+CV</i>		0,070	0,002
<i>XGBoost</i>	<i>RMSE</i>	899,253	5086,646
<i>XGBoost + L1, L2, gamma</i>		1471,762	4914,831
<i>XGBoost+IsolationForest</i>		237,797	5223,173
<i>XGBoost+UMAP</i>		81,637	4622,154
<i>XGBoost+IsolationForest+UMAP</i>		69,812	4845,172
<i>XGBoost+CV</i>		4031,295	4039,579
<i>XGBoost+IsolationForest+CV</i>		3987,638	4032,989
<i>XGBoost+UMAP+CV</i>		4003,164	4066,136
<i>XGBoost+IsolationForest+UMAP+CV</i>		3941,093	4024,302

Оптимальным для построения точной модели по результатам эксперимента оказался алгоритм XGBoost (рис. 4).

Таким образом, все основные моменты Эксперимента №2 остались неизменными. Как и в предыдущий раз, отсутствует положительный эффект на обучение при регуляризации посредством L1, L2 и gamma. Модель XGBoost+L1,L2,gamma показывает $R^2 = -0.4883$ на тестовой выборке против $R^2 = -0.5942$ у базового XGBoost, что свидетельствует об ухудшении обобщающей способности. Если отдельно дополнить модель XGBoost алгоритмом Isolation Forest, то $R^2 = -0.6809$. Дополнение модели XGBoost алгоритмом

UMAP выдает $R^2 = -0.3163$. Таким образом, метрики качества ухудшаются.

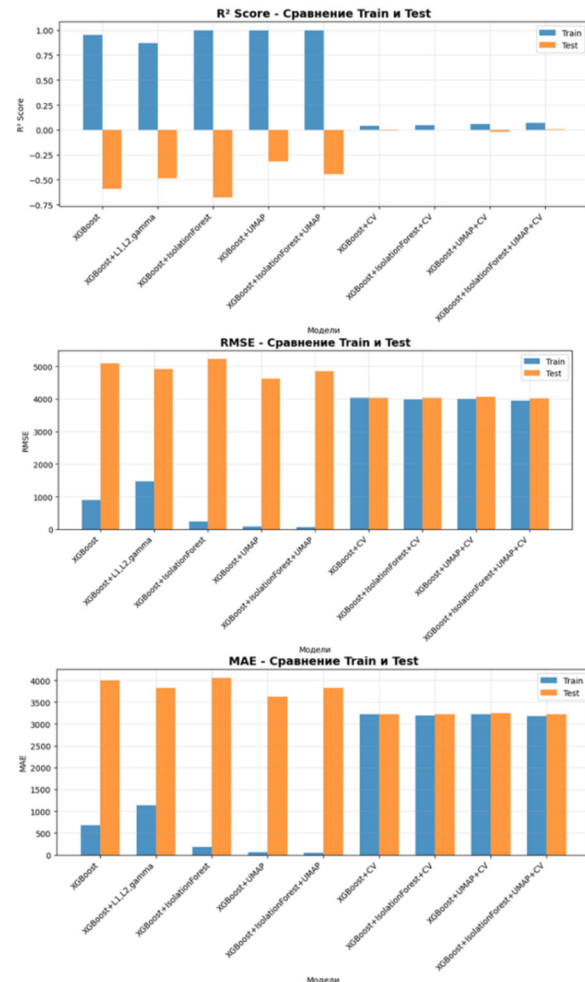


Рис. 4. Значения метрик для Эксперимента №3

Точность итоговой модели значительно улучшается при кросс-валидации: все модели с ней показывают существенное улучшение тестовых метрик.

Наилучший результат демонстрирует модель XGBoost+IsolationForest+UMAP+CV с $R^2 = 0.0022$, что является единственным положительным значением среди всех подходов. Также заметное улучшение наблюдается в модели XGBoost+IsolationForest+CV с $R^2 = -0.0021$, что подтверждает эффективность комбинации методов борьбы с переобучением.

Особенности датасета могут оказать значительное влияние на результаты эксперимента. Так, неэффективность методов снижения размерности может стать результатом того, что в датасете есть определенное количество схожих по характеристикам товарных групп. В этом случае кластеризация с применением алгоритма UMAP практически не повысила точность при использовании без CV. Это отдельный случай определенного датасета. В целом же применение кластеризации весьма эффективно для обеспечения универсальности модели на разнообразных данных.

Следует отметить, что у модифицированных подходов с кросс-валидацией наблюдается уменьшение переобучения по сравнению с другими моделями. Это выражается в выравнивании ошибок как на тестовых, так и на тренировочных данных, что показано на рисунке 4.

Модели без CV демонстрируют экстремальное переобучение с $R^2 \approx 1.0$ на тренировочных данных и отрицательными значениями на тестовых, в то время как подходы с CV показывают более сбалансированные результаты, что подтверждает их эффективность для улучшения обобщающей способности моделей.

Эксперименты 1-3 позволяют определить основные тенденции и закономерности, на основе чего можно сделать ряд выводов.

Эффект переобучения возможен при обучении моделей и их модификации на недостаточном количестве данных. Эксперимент №2 ярко продемонстрировал данный эффект переобучения, когда на тестовой выборке все значения R^2 оказались отрицательными у всех моделей.

Эксперимент №3 доказывает, что комбинация различных алгоритмов в моделях приводит к улучшению метрик, при последовательном добавлении методов растет точность.

Таким образом, наиболее эффективной стратегией оказалось комбинирование методов обработки аномалий, снижения размерности и кросс-валидации, что позволило достичь сбалансированных результатов и улучшить обобщающую способность моделей.

5 Выводы

Таким образом, в настоящей работе осуществлена разработка методов и алгоритмов решения задач прогнозирования объемов продаж в организационных системах торговых компаний. В исследовании изучена проблема прогнозирования объемов продаж различных (в том числе телекоммуникационных) товаров с применением метода градиентного усиления.

Был проведен ряд модификаций этого алгоритма для решения задач управления в организационных системах торговых компаний: взвешивание образцов выборки с применением IsolationForest, кластеризация данных с использованием UMAP, кросс-валидация. Комплекс перечисленных модификаций позволил существенно улучшить точность прогнозирования. Перечисленные модификации, использующиеся в ходе разработки методов и алгоритмов решения задач управления в организационных системах торговых компаний, позволили улучшить обучение алгоритма XGBoost за счет фильтрации выбросов и снижения их негативного влияния на процесс обучения, учесть специфику данных, улучшить обработку разнонаправленных закономерностей в распределениях шумов, обеспечить стабильность результатов обучения, повысить обобщающую способность моделей.

В результате проведения настоящего исследования можно сделать выводы о том, что разработанные методы и алгоритмы могут быть применены для повышения эффективности решения задач прогнозирования продаж, что используется при управлении в организационных системах торговых компаний.

Литература

1. John D. Kelleher, Deep Learning. The MIT Press Essential Knowledge series, MIT Press, 2019.
2. Simon J.D. Prince, Understanding Deep Learning. MIT Press, 2023.

3. Леохин Ю.Л., Фатхулин Т.Д., Маслов К.В. Разработка методов системного анализа бизнес-процессов в банковской сфере для принятия решений о кредитовании различных организаций // Научные технологии в космических исследованиях Земли. 2025. Т. 17. № 5. С. 59-71. DOI: 10.36724/2409-5419-2025-17-5-59-71.

4. Леохин Ю.Л., Дымкова С.С., Фатхулин Т.Д., Мяличева А.А. Методы прогнозирования дефектов в программных продуктах на основе ретроспективной и текущей информации // Т-Comm: Телекоммуникации и транспорт. 2026. Том . № 1. С. XX-УУ.

5. Леохин Ю.Л., Дымкова С.С., Фатхулин Т.Д., Зозуля И.С. Методы и алгоритмы интеллектуальной поддержки принятия управленческих решений в организационных системах торговых компаний // Т-Comm: Телекоммуникации и транспорт. 2025. Том 19. №12. С. 44-50.

6. Leokhin Y., Fatkhulin T., Dymkova S. Research of Methods for Forecasting Dynamically Changing Data in Solving Applied Problems // 2025 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED), Moscow, Russian Federation, 2025, pp. 1-12.

7. Леохин Ю.Л., Дымкова С.С., Фатхулин Т.Д. Методы машинного обучения в прикладных задачах прогнозирования динамично изменяющихся данных // Т-Comm: Телекоммуникации и транспорт. 2025. Т. 19. №8. С. 49-63.

8. Леохин Ю.Л., Фатхулин Т.Д., Занегин А.В. Модификация метода градиентного усиления для прогнозирования спроса на отдельные виды товаров // Научные технологии в космических исследованиях Земли. 2025. Т. 17. № 2. С. 32-41. DOI: 10.36724/2409-5419-2025-17-2-32-41.

9. Leokhin Y., Fatkhulin T., Boitsov K. Computer Vision Methods in Applied Problems of Classifying Objects in Images // 2025 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF), St. Petersburg, Russian Federation, 2025, pp. 1-10, DOI: 10.1109/WECONF65186.2025.11017109.

10. Леохин Ю.Л., Дымкова С.С., Фатхулин Т.Д. Исследование и разработка инструментальных средств повышения качества изображений // Т-Comm: Телекоммуникации и транспорт. 2025. Т. 19. №4. С. 45-56. (in English).

11. Фатхулин Т.Д., Зозуля И.С. Исследование методов прогнозирования показателей бизнес-процессов // REDS: Телекоммуникационные устройства и системы. 2025. Т. 15, № 2. С. 12-17. EDN ZXTFFD.

12. Fatkhulin T., Leokhin Y., Zanegin A., Rakhmatova A. Development and Research of a Modified Gradient Boosting Method Effectiveness to Solve Applied Problems of Time-Series Forecasting // 2025 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, Russian Federation, 2025, pp. 1-10, DOI: 10.1109/IEEECONF64229.2025.10948023.

13. Leokhin Y., Fatkhulin T., Zanegin A., Rakhmatova A. Researching the Efficiency of Machine Learning Methods Used in Forecasting Demand for Certain Types of Goods // 2025 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, Russian Federation, 2025, pp. 1-8, DOI: 10.1109/IEEECONF64229.2025.10948113.

14. Вишневецкий В.М., Леохин Ю.Л., Фатхулин Т.Д., Занегин А.В. Методы машинного обучения в решении задачи прогнозирования спроса на отдельные виды товаров // Т-Comm: Телекоммуникации и транспорт. 2024. Т. 18. №10. С. 34-43.

15. Leokhin Y.L., Fatkhulin T.D. Evaluation of Service Availability in Software-Defined Optical Network // 2021 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, Russia, 2021, pp. 1-6, DOI: 10.1109/IEEECONF51389.2021.9416122.

16. Leokhin Y.L., Fatkhulin T.D. Approach to Estimating the Probability of Providing "Cloud" Services in the SDN // 2020 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, Russia, 2020, pp. 1-9, DOI: 10.1109/IEEECONF48371.2020.9078593.

17. *Деарт В.Ю., Фатхулин Т.Д.* Анализ процесса создания суперканала с необходимой пропускной способностью в сети, построенной по технологии транспортных программно-конфигурируемых сетей (Т-SDN) // Т-Comm: телекоммуникации и транспорт. 2018. Т. 12. №10. С. 23-30.
18. *Деарт В.Ю., Фатхулин Т.Д.* Анализ транспортных программно-конфигурируемых сетей (Т-SDN) с управляемым оптическим уровнем с целью получения модели, позволяющей оценить возможность предоставления сервиса Bandwidth on Demand // Т-Comm: Телекоммуникации и транспорт. 2018. Т.12. №4. С.35-42.
19. *Deart V., Fatkhulin T.* Analysis of the functioning of a multi-domain transport software-defined network with controlled optical layer // 2017 21st Conference of Open Innovations Association (FRUCT), Helsinki, Finland, 2017, pp. 79-87, DOI: 10.23919/FRUCT.2017.8250168.
20. *Деарт В.Ю., Фатхулин Т.Д.* Анализ современного состояния транспортных сетей с целью внедрения технологии программно-конфигурируемых сетей (SDN) // Т-Comm: Телекоммуникации и транспорт. 2017. Т.11. №6. С. 4-9.
21. *Портнов Э.Л., Фатхулин Т.Д.* Технологии достижения высоких скоростей передачи в современных когерентных DWDM-системах связи // Т-Comm: Телекоммуникации и транспорт. 2015. Т.9. №8. С. 34-37.
22. *Ринас Н.А., Золкин А.Л., Каберова А.Р., Скибин Ю.В.* Влияние автоматизации и искусственного интеллекта на социальное неравенство // Экономика и управление: проблемы, решения. 2025. Т. 7, № 1(154). С. 116-125. DOI 10.36871/ek.up.p.r.2025.01.07.015. EDN BGBKZZ.
23. *Беспалова В.В., Каберова А.Р., Белинская Д.Б.* и др. Методический подход к управлению устойчивостью развития региона // Экономика и управление: проблемы, решения. 2024. Т. 11, № 11(152). С. 88-93. DOI 10.36871/ek.up.p.r.2024.11.11.011. EDN GDEJYC.
24. *Драгуленко В.В., Золкин А.Л., Есина О.И., Каберова А.Р.* Влияние численности населения на экономический рост и развитие стран // Экономика и управление: проблемы, решения. 2024. Т. 11, № 9(150). С. 67-75. DOI 10.36871/ek.up.p.r.2024.09.11.009. EDN ANEUBG.
25. *Pavlov S.V., Dokuchaev V.A., Maklachkova V.V., Mytenkov S.S.* Features of supporting decision making in modern enterprise infocommunication systems // Т-Comm: Телекоммуникации и транспорт. 2019. Т. 13. № 3. С. 71-74.
26. *Бурлов В.Г., Грачев М.И.* Аналитическо-динамическая модель управленческого решения в социально-экономических системах на примере руководителя учебного заведения высшего образования // Т-Comm: Телекоммуникации и транспорт. 2019. Т. 13. № 10. С. 27-34.
27. *Шелухин О.И., Раковский Д.И.* Бинарная классификация многоатрибутных размеченных аномальных событий компьютерных систем с помощью алгоритма SVDD // Научные исследования в космических исследованиях Земли. 2021. Т. 13. № 2. С. 74-84.

DEVELOPMENT OF METHODS AND ALGORITHMS FOR SOLVING MANAGEMENT PROBLEMS IN ORGANIZATIONAL SYSTEMS OF TRADING COMPANIES

Yuri L. Leokhin, Moscow Technical University of Communications and Informatics, Moscow, Russia, y.l.leokhin@mtuci.ru
Svetlana S. Dymkova, Moscow Technical University of Communications and Informatics, Moscow, Russia, s.s.dymkova@mtuci.ru
Timur D. Fatkhulin, Moscow Technical University of Communications and Informatics, Moscow, Russia, t.d.fatkhulin@mtuci.ru
Irina S. Zozulya, Moscow Technical University of Communications and Informatics, Moscow, Russia

Abstract

This paper presents the development and experimental validation of a modified method based on the XGBoost algorithm for sales forecasting. The objective of the study is to improve the accuracy of sales forecasting methods and algorithms employed in organizational systems of trading companies. The relevance of this study lies in the fact that one of the key challenges in the sales domain is the presence of noise and anomalies in data. This leads to a decrease in the accuracy of forecasting results and the occurrence of overfitting. Tuning existing parameters in forecasting algorithms for management in organizational systems of trading companies does not always solve these problems and leads to the need for additional modifications to refine forecasts. This paper develops a method for modifying the gradient boosting algorithm using adjustments to the gradient calculation using anomaly estimates for dataset samples using the isolation forest method. Additional hyperparameters and a method for stopping training upon receiving information about the onset of overfitting are also used. The study examined approaches to regularization and mitigating the impact of data anomalies using the XGBoost algorithm as an example. Additional modifications to the proposed method are made using cross-validation. All modifications were tested in various combinations on different datasets. Thus, a pattern of increased accuracy was identified when training on cross-validation using the developed anomaly weighting method. The practical significance of this work lies in the development of a universal approach to modifying the XGBoost algorithm, which can be applied to improve the accuracy of sales forecasting for various product categories in organizational management systems of retail companies.

Keywords: method, algorithm, management problem, sales volumes, organizational system of trading companies, forecasting, indicators, efficiency

References

- [1] John D. Kelleher, *Deep Learning*. The MIT Press Essential Knowledge series, MIT Press, 2019.
- [2] Simon J.D. Prince, *Understanding Deep Learning*. MIT Press, 2023.
- [3] Yu.L. Leokhin, T.D. Fatkhulin, K.V. Maslov, "Development of business processes system analysis methods in the banking sphere for making decisions on lending various organizations," *H&ES Reserch*. 2025. Vol. 17. No. 5, pp. 59-71 DOI: 10.36724/2409-5419-2025-17-5-59-71. (In Rus)
- [4] Yu. L. Leokhin, S. S. Dymkova, T. D. Fatkhulin, A. A. Myalicheva, "Methods of predicting defects in software products based on retrospective and current information," *T-Comm*, 2026, vol. 20, no.1. pp. 52-59. (in Russian)
- [5] Yu.L. Leokhin, S.S. Dymkova, T.D. Fatkhulin, I.S. Zozulya, "Methods and algorithms of intellectual support for making management decisions in organizational systems of trading companies," *T-Comm*, 2025, vol. 19, no.12, pp. 44-50. (in Russian)
- [6] Y. Leokhin, T. Fatkhulin and S. Dymkova, "Research of Methods for Forecasting Dynamically Changing Data in Solving Applied Problems," *2025 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED)*, Moscow, Russian Federation, 2025, pp. 1-12, in press.
- [7] Yu.L. Leokhin, S.S. Dymkova, T.D. Fatkhulin, "Machine learning methods in applied problems of forecasting dynamically changing data," *T-Comm*, 2025, vol. 19, no.8, pp. 49-63. (in Russian)
- [8] Yu.L. Leokhin, T.D. Fatkhulin, A.V. Zanegin, "The gradient boosting method modification to forecast demand for individual types of goods," *H&ES Reserch*. 2025. Vol. 17. No. 2, pp. 32-41. DOI: 10.36724/2409-5419-2025-17-2-32-41. (In Rus)
- [9] Y. Leokhin, T. Fatkhulin and K. Boitsov, "Computer Vision Methods in Applied Problems of Classifying Objects in Images," *2025 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, St. Petersburg, Russian Federation, 2025, pp. 1-10, DOI: 10.1109/WECONF65186.2025.11017109.
- [10] Yu.L. Leokhin, S.S. Dymkova, T.D. Fatkhulin, "Research and development of image improvement tools," *T-Comm*, 2025, vol. 19, no. 4, pp. 45-56.
- [11] Fatkhulin T.D., Zozulya I.S., "Research of methods for forecasting business process indicators," *REDS: Telecommunication devices and systems*. - 2025. Vol. 15, No. 2, pp. 12-17. (in Russian)
- [12] T. Fatkhulin, Y. Leokhin, A. Zanegin and A. Rakhmatova, "Development and Research of a Modified Gradient Boosting Method Effectiveness to Solve Applied Problems of Time-Series Forecasting," *2025 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, Russian Federation, 2025, pp. 1-10, DOI: 10.1109/IEEECONF64229.2025.10948023.
- [13] Y. Leokhin, T. Fatkhulin, A. Zanegin and A. Rakhmatova, "Researching the Efficiency of Machine Learning Methods Used in Forecasting Demand for Certain Types of Goods," *2025 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, Russian Federation, 2025, pp. 1-8, DOI: 10.1109/IEEECONF64229.2025.10948113.
- [14] V.M. Vishnevsky, Yu.L. Leokhin, T.D. Fatkhulin, A.V. Zanegin, "Machine learning methods in solving the problem of forecasting demand for specific types of goods," *T-Comm*, vol. 18, no. 10, pp. 34-43. (in Russian)
- [15] Y. L. Leokhin and T. D. Fatkhulin, "Evaluation of Service Availability in Software-Defined Optical Network," *2021 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, Russia, 2021, pp. 1-6, DOI: 10.1109/IEEECONF51389.2021.9416122.
- [16] Y. L. Leokhin and T. D. Fatkhulin, "Approach to Estimating the Probability of Providing "Cloud" Services in the SDN," *2020 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, Russia, 2020, pp. 1-9, DOI: 10.1109/IEEECONF48371.2020.9078593.
- [17] V.Yu. Deart, T.D. Fatkhulin, "Analysis of the process of creating a superchannel with the necessary capacity in the network organized according to transport software-defined networks (T-SDN) technology," *T-Comm*, 2018, vol. 12, no.10, pp. 23-30. (in Russian).
- [18] Yu. Deart, T.D. Fatkhulin, "Analysis of transport software-defined networks (T-SDN) with controlled optical layer to obtain a model providing assesment of the possibility of bandwidth on demand service," *T-Comm*, 2018, vol. 12, no.4, pp. 35-42. (in Russian).
- [19] V. Deart and T. Fatkhulin, "Analysis of the functioning of a multi-domain transport software-defined network with controlled optical layer," *2017 21st Conference of Open Innovations Association (FRUCT)*, Helsinki, Finland, 2017, pp. 79-87, DOI: 10.23919/FRUCT.2017.8250168.
- [20] Yu. Deart, T.D. Fatkhulin, "Analysis of current state of transport networks with the purpose of introducing software defined networks (SDN) technology," *T-Comm*, 2017, vol. 11, no.6, pp. 4-9. (in Russian)
- [21] E.L. Portnov, T.D. Fatkhulin, "Technologies aimed at achieving high speed transmission in modern coherent DWDM communication systems," *T-Comm*. 2015. Vol 9. No.8, pp. 34-37. (in Russian)
- [22] N. A. Rinas, A. L. Zolkin, A. R. Kaberova and Yu. V. Skibin, "The Impact of Automation and Artificial Intelligence on Social Inequality," *Economy and Management: Problems, Solutions*, 2025, Vol. 7, No. 1(154), pp. 116-125. DOI 10.36871/ek.up.p.r.2025.01.07.015. (in Russian)
- [23] V. V. Bepalova, A. R. Kaberova, D. B. Belinskaya [et al.], "Methodological approach to managing the sustainability of regional development," *Economy and Management: Problems, Solutions*, 2024, Vol. 11, No. 11(152), pp. 88-93. DOI 10.36871/ek.up.p.r.2024.11.11.011. (in Russian)
- [24] V. V. Dragulenko, A. L. Zolkin, O. I. Esina and A. R. Kaberova, "The Impact of Population on Economic Growth and Development of Countries," *Economy and Management: Problems, Solutions*, 2024, Vol. 11, No. 9(150), pp. 67-75. DOI 10.36871/ek.up.p.r.2024.09.11.009. (in Russian)
- [25] S.V. Pavlov, V.A. Dokuchaev, V.V. Maklachkova, S.S. Mytenkov, "Features of supporting decision making in modern enterprise infocommunication systems," *T-Comm*. 2019, vol. 13, no. 3, pp. 71-74.
- [26] V.G. Burlov, M.I. Grachev, "Analytical-dynamic model of management decision in socio-economic systems on the example of the head of a educational institution of higher education," *T-Comm*, 2019, vol. 13, no.10, pp. 27-34.
- [27] O.I. Sheluhin, D.I. Rakovskiy, "Binary classification of multi-attribute tagged data about anomalous events in computer systems using the SVDD algorithm," *H&ES Research*. 2021, vol. 13, no. 2, pp. 74-84. Doi: 10.36724/2409-5419-2021-13-2-74-84.

Information about authors:

Yuri L. Leokhin, Moscow Technical University of Communications and Informatics, Rectorate, Vice-Rector for Scientific Work, full professor, Dr. Sc. (Tech.), orcid.org/0000-0003-3321-4497

Svetlana S. Dymkova, Moscow Technical University of Communications and Informatics, Candidate Sc. (Tech.), orcid.org/0000-0003-1945-9850

Timur D. Fatkhulin, Moscow Technical University of Communications and Informatics, Dpt. of IDA, Docent, Candidate Sc. (Tech.), orcid.org/0000-0003-0998-1055

Irina S. Zozulya, Moscow Technical University of Communications and Informatics, Dpt. of MC and IT, master's student, orcid.org/0009-0006-3445-8070