

МЕТОДЫ МАШИННОГО ОБУЧЕНИЯ В РЕШЕНИИ ЗАДАЧИ ПРОГНОЗИРОВАНИЯ СПРОСА НА ОТДЕЛЬНЫЕ ВИДЫ ТОВАРОВ

Вишневыский Владимир Миронович,
Институт проблем управления им. В.А. Трапезникова РАН,
Москва, Россия, vishn@ipu.ru

Леохин Юрий Львович,
Московский технический университет связи и информатики,
Москва, Россия, y.l.leokhin@mtuci.ru

Фатхулин Тимур Джалилевич,
Московский технический университет связи и информатики,
Москва, Россия, t.d.fatkhulin@mtuci.ru

Занегин Алексей Вячеславович,
Московский технический университет связи и информатики,
Москва, Россия, vishn@ipu.ru

DOI: 10.36724/2072-8735-2024-18-10-34-43

Manuscript received 15 September 2024;
Accepted 14 October 2024

Ключевые слова: метод, алгоритм, машинное обучение, метрика, эффективность, спрос, показатель, товар

В работе рассматривается проблема повышения эффективности прогноза спроса на отдельные виды товаров с учетом информационных процессов в торговле и продажах. Цель работы – определение наиболее эффективного метода машинного обучения для прогнозирования спроса на отдельные виды товаров. Актуальность работы обусловлена тем, что традиционные статистические методы прогнозирования спроса не позволяют эффективно обрабатывать получаемые в настоящее время объемы информации о продажах и торговле, также не все методы машинного обучения для прогнозирования числовых величин подходят для рассматриваемой предметной области. Объектом исследования является возникающий спрос на отдельные виды товаров. Предметом исследования являются показатели эффективности (метрики) методов машинного обучения, используемых для прогнозирования спроса на отдельные виды товаров. В работе рассматриваются такие методы и алгоритмы, как линейная регрессия, случайный лес, различные виды градиентного усиления, а также рекуррентная нейронная сеть и нейронная сеть с долгосрочной-краткосрочной памятью. В качестве метрик, используемых для оценки эффективности того или иного метода, в ходе анализа предметной области выбраны среднеквадратичная ошибка, средняя абсолютная ошибка и коэффициент детерминации. В результате проведенного исследования были определены наиболее подходящие для рассматриваемой предметной области алгоритмы, которые позволяют эффективно прогнозировать спрос на различные виды товаров. Даны рекомендации по возможным настройкам параметров моделей, улучшающие показатели их работы. Сделаны выводы о применимости каждого из рассматриваемых алгоритмов и обозначены перспективы для дальнейших исследований. Методологическую основу работы составляют следующие методы: анализ, сопоставление, сравнение, эксперимент и обобщение.

Информация об авторах:

Вишневыский Владимир Миронович, Институт проблем управления им. В.А. Трапезникова РАН, Лаборатория 69, г.н.с., профессор, д.т.н., Москва, Россия. orcid.org/0000-0001-7373-4847

Леохин Юрий Львович, Московский технический университет связи и информатики, Ректорат, Проректор по научной работе, профессор, д.т.н., Москва, Россия. orcid.org/0000-0003-3321-4497

Фатхулин Тимур Джалилевич, Московский технический университет связи и информатики, кафедра "Математическая кибернетика и информационные технологии", доцент, к.т.н., Москва, Россия. orcid.org/0000-0003-0998-1055

Занегин Алексей Вячеславович, Московский технический университет связи и информатики, кафедра "Математическая кибернетика и информационные технологии", магистрант. orcid.org/0009-0005-5518-7882

Для цитирования:

Вишневыский 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. Том 18. №10. С. 34-43.

For citation:

Vishnevsky V.M., Leokhin Yu.L., Fatkhulin T.D., Zanegin A.V. 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)

Введение

Современное общество характеризуется значительным объемом продаж различных товаров и предоставляемых услуг [1, 2, 23-26]. Данные о продажах представляют собой непрерывную выгрузку в датасет (от англ. *dataset* – набор данных) [6, 7], где каждая группа «товар-магазин» представляет собой обособленный временной ряд, который практически невозможно по отдельности анализировать и представлять графически. Применение статистических методов является трудоемким и малоэффективным в данной задаче [27-31], особенно на больших наборах данных со сложными взаимосвязями и признаками.

Применение методов машинного обучения (МО) позволяет автоматизировать данный процесс. Тем не менее, их использование также не лишено недостатков, поскольку не всегда данное применение становится эффективным [21], а неоднозначность выбора конкретного метода связана с тем, что такой выбор зависит от множества факторов, таких как требования к скорости и точности вычислений, наличие вычислительных ресурсов и имеющихся данных [22-25].

Таким образом, для достижения указанной цели необходимо решить следующие задачи:

- проанализировать существующие методы и модели машинного обучения, используемые в рассматриваемой предметной области прогнозирования спроса;
- выделить и проанализировать основные этапы построения модели МО для прогноза спроса, а также определить используемые метрики оценки ее эффективности;
- проанализировать работу выбранных методов и моделей МО и сделать вывод об их эффективности.

Анализ методов машинного обучения, используемых для прогнозирования спроса

Проведем анализ методов машинного обучения, которые наиболее часто применяются в области прогнозирования спроса на различные виды товаров [4, 5, 8, 11, 13, 15-20, 22]. Далее будут приведены краткие характеристики рассматриваемых методов и их применимость для решения указанной задачи.

Linear Regression (линейная регрессия) – метод, с помощью которого находится отношение зависимой переменной от одной или нескольких независимых переменных (регрессоров) посредством линейной функции. Он позволяет прогнозировать новые значения при наличии линейной зависимости [15, 16]. Данный метод (алгоритм) является самым простым и одновременно самым быстрым. При наличии нелинейных связей он малоэффективен, однако в данной задаче применяется в качестве базового алгоритма для сравнения и оценки целесообразности применения более сложных моделей.

Алгоритм *Random Forest (случайный лес)* представляет собой ансамблевый метод машинного обучения, основу которого составляет более простой алгоритм дерева решений [15, 16]. За счет того, что происходит обучение множества деревьев параллельно и для оценки итогового результата используется их усреднение, алгоритм случайного леса имеет повышенную точность и устойчивость к выбросам и шуму в данных, что важно для решаемой задачи.

Алгоритмы градиентного усиления или градиентного «бустинга» (от англ. *boosting* – повышение, усиление) *XGBoost*, *CatBoost*, *LightGBM* основаны на последовательном обучении деревьев решений, где каждое последующее дерево минимизирует ошибку предыдущего. Среди этих алгоритмов есть неоднозначность выбора, поскольку каждый из них строит деревья по-разному, применяя различные техники оптимизации. Более того данные алгоритмы поддерживают использование графического процессора без каких-либо дополнительных библиотек [15, 16]. С учетом этого также необходимо рассмотреть его применимость и эффективность для решаемой задачи.

RNN (Recurrent neural network, рекуррентная нейронная сеть) и *LSTM (Long-Short-Term Memory, нейронная сеть с долгосрочной-краткосрочной памятью)* [3, 9, 12] – типы нейронных сетей, которые хорошо подходят для работы с последовательными данными, такими как временные ряды. RNN позволяет передавать информацию из предыдущих шагов времени в следующие, что помогает учесть зависимости и закономерности в данных. Однако, у обычных RNN возникает проблема затухания градиента, из-за которой они могут не учитывать выявленные зависимости в данных, полученные в прошлом. LSTM – тип RNN, который решает проблему затухания градиента. Он обладает специальной структурой, которая позволяет модели сохранять важную информацию на долгосрочной памяти и забывать ненужную. Это помогает улучшить способность модели к запоминанию долгосрочных зависимостей в данных [7, 14-16].

Этапы построения моделей машинного обучения

При построении моделей машинного обучения с применением выбранных ранее методов (алгоритмов) необходимо определить этапы, на которых можно повлиять на эффективность и точность модели в целом. Для этого рассмотрим основные этапы построения модели МО.

Предобработка данных является базовым этапом, от качества выполнения которого зависит эффективность всех последующих действий. Применительно к данному этапу самым важным является работа с пропусками и отсутствующими значениями в данных. Данные во временных рядах должны быть представлены через равные промежутки времени. Так, при отсутствии продаж за этот день должны использоваться нулевые значения.

Проектирование производных признаков позволяет модели лучше адаптироваться к данным о продажах и воспринимать их временную природу. Среди используемых производных признаков в ходе проведенного анализа предметной области и информационных процессов в торговле и продажах выделены следующие:

- *признаки времени* – номер дня, недели, месяца, года, которые идентифицируют временную характеристику значений продаж для модели машинного обучения, поскольку алгоритмы не воспринимают временные зависимости автоматически;
- *«лаговые» переменные* (от англ. *lag* – отставание) – отображают зависимости текущих продаж от прошлых;
- *скользящие статистики* – помогают сгладить выбросы и шумы в данных, а также выявить тенденции;

- *разности «лагов»* – выявляют динамику изменения данных, таких как скорость увеличения или падения продаж;
- *доли продаж от общего объема продаж* – позволяют выявить взаимодействие и структуру между отдельными временными рядами.

При проектировании производных признаков со сдвигом во времени, необходимо руководствоваться принципом избегания «подглядывания в будущее». С этой целью производные признаки генерируются на горизонт, равный горизонту прогнозирования, иначе в тестовых данных появляются подсказки для прогноза, а модель становится переоцененной.

Подготовка данных являясь техническим этапом, также влияет на поведение модели. Кодирование категориальных признаков необходимо для корректного восприятия данных. Для отдельных алгоритмов, таких как LSTM, необходимо масштабирование переменных.

Немаловажно также отметить проблему разбиения данных. Обучение и прогноз только на тестовой выборке приведет к искажению представления о возможностях алгоритма. Обучение с настройкой гиперпараметров, т.е. глобальных параметров, устанавливаемых вручную для управления процессом обучения, обычно проводится на проверочной или так называемой «валидационной» выборке (от англ. *validation* – проверка), а прогнозы осуществляются на тестовой выборке для более объективной оценки эффективности их работы.

Этап настройки гиперпараметров помогает алгоритмам адаптироваться к данным, оптимизировать скорость и точность работы. Для рассматриваемых алгоритмов в ходе анализа литературы (информация из открытых ресурсов <https://xgboost.readthedocs.io/en/stable/>, <https://catboost.ai/en/docs/references/training-parameters/>, <https://lightgbm.readthedocs.io/en/latest/pythonapi/lightgbm.LGBMModel.html> и <https://keras.io/api/>) были выбраны основные гиперпараметры и диапазоны для настройки, которые будут представлены далее.

Выбор большого количества гиперпараметров не обязательно приведет к более точной настройке. Для поиска оптимальных значений даже с помощью автоматических библиотек требуется большое количество итераций и вычислительных мощностей, в противном случае поиск становится бессмысленным. Более того, настроенные гиперпараметры приводят к ограничению работы алгоритма, а неправильный выбор диапазона поиска может привести к отрицательным результатам.

Стоит отметить, что алгоритмы, в основе которых лежат деревья решений, имеют схожие настройки для сокращения переобучения, уменьшения влияния шумов и выбросов в данных, несмотря на разные подходы к формированию деревьев. Настройки программного обеспечения (ПО), реализующего конкретный алгоритм МО, в таком случае связаны с количеством деревьев в ансамбле, их глубиной и ограничениями по разделению [9, 13, 14].

Этап оценки эффективности проводится с помощью стандартных *метрик* *RMSE* (Root Mean Squared Error – среднеквадратичная ошибка), *MAE* (Mean Absolute Error – средняя абсолютная ошибка) и *R2* (коэффициент детерминации) [10, 15, 16, 18]. Несмотря на то, что есть и более продвинутые метрики, данные метрики являются универсальными, а их использование вместе позволяет более полно оценить поведение моделей, что было определено в ходе анализа рассматриваемой предметной области.

Также стоит учесть фактор времени работы метода (алгоритма) МО как при обучении, так и при прогнозе на тестовой выборке, так как даже самый точный алгоритм не может считаться эффективным, если требуется большое количество вычислительных ресурсов, а результаты получаются с опозданием, не в режиме реального времени. Таким образом, проанализированы основные этапы построения модели МО, а также определено, какие метрики позволят оценить эффективность рассматриваемых моделей МО.

Экспериментальные исследования

В соответствии с представленными выше этапами построения модели прогнозирования с применением методов машинного обучения, рассматриваемые алгоритмы были обучены на данных, взятых из датасетов открытой платформы «Kaggle».

В таблице 1 приведен список выбранных гиперпараметров для настройки рассматриваемых методов (алгоритмов) МО, а также примененный диапазон для поиска наилучших значений с помощью библиотеки «Optuna».

Таблица 1

Гиперпараметры для настройки рассматриваемых методов (алгоритмов) машинного обучения

| Алгоритм | Наименование гиперпараметра | Тип значения | Диапазон поиска гиперпараметра |
|--------------------------|-----------------------------|--------------|--------------------------------|
| <i>Linear Regression</i> | - | - | - |
| <i>Random Forest</i> | n_estimators | int | 100 – 500, шаг 100 |
| | max_depth | int | 1 – 20 |
| | max_features | float | 0,05 – 1 |
| | min_samples_split | int | 2 – 20 |
| <i>XGBoost</i> | 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 |
| <i>LightGBM</i> | min_child_weight | int | 1 – 20 |
| | n_estimators | int | 200 – 1000, шаг 200 |
| | num_leaves | int | 2 – 2**10 |
| | learning_rate | float | 0,001 – 1 |
| | subsample | float | 0,05 – 1 |
| | colsample_bytree | float | 0,05 – 1 |
| <i>CatBoost</i> | min_child_samples | int | 1 – 20 |
| | iterations | int | 200 – 1000, шаг 200 |
| | depth | int | 1 – 10 |
| | learning_rate | float | 0,001 – 1 |
| | subsample | float | 0,05 – 1 |
| <i>LSTM</i> | colsample_bylevel | float | 0,05 – 1 |
| | min_data_in_leaf | int | 1 – 20 |
| | n_units | int | 10 – 200, шаг 100 |
| | dropout_rate | float | 0,2 – 0,5, шаг 0,1 |
| | learning_rate | float | 0,001 – 0,1 |
| | batch_size | int | 30, 60, 90, 120, 150 |
| | epochs | int | 10 – 50, шаг 10 |

Ниже приведено описание реальных практических задач на нескольких наборах данных, проведена оценка эффективности работы алгоритмов и сделаны выводы.

Эксперимент №1. За основу был взят датасет фирмы «1С» о продажах сети магазинов в различных регионах для отдельных товаров (информация из открытых ресурсов платформы «Kaggle»: <https://www.kaggle.com/competitions/competitive-data-science-predict-future-sales>). Данные были агрегированы ежемесячно до товарных групп в одном регионе. В итоге получилось 16 249 записей из 67 товарных групп в 13 магазинах за 34 месяца. Проверочная и тестовая выборка составляют один месяц каждая. Особую сложность для алгоритмов машинного обучения в данном эксперименте представляет укрупнение данных ежемесячно, с учетом этого, при обучении используются короткие временные ряды. В таблице 2 выделены наиболее значимые показатели отдельных алгоритмов для рассматриваемого случая. В таблицах 2-4 под названиями столбцов «train» и «test» понимается результат работы алгоритма на обучающей и тестовой выборке соответственно. Лучшие показатели выделены жирным шрифтом во всех таблицах.

Таблица 2

Результаты работы алгоритмов на полных данных для Эксперимента №1

| Алгоритм | Метрика | С настройками | | Без настроек | | Уменьшение ошибок/Прирост точности | |
|-------------------------|---------|---------------|--------------|--------------|--------------|------------------------------------|------|
| | | train | test | train | test | train | test |
| <i>CatBoost</i> | MAE | 12,49 | 37,48 | 14,41 | 34,17 | -13% | 10% |
| LightGBM | MAE | 3,29 | 24,17 | 14,07 | 25,33 | -77% | -5% |
| <i>LinearRegression</i> | MAE | 30,37 | 35,05 | 30,37 | 35,05 | 0% | 0% |
| <i>LSTM</i> | | 54,88 | 55,44 | 61,60 | 54,19 | -11% | 2% |
| RandomForest | MAE | 16,55 | 23,44 | 7,62 | 25,19 | 117% | -7% |
| XGBoost | MAE | 4,60 | 24,66 | 6,44 | 26,39 | -29% | -7% |
| <i>CatBoost</i> | R2 | 0,97 | 0,67 | 0,97 | 0,70 | 0% | -4% |
| LightGBM | R2 | 0,99 | 0,78 | 0,97 | 0,74 | 3% | 5% |
| <i>LinearRegression</i> | R2 | 0,81 | 0,67 | 0,81 | 0,67 | 0% | 0% |
| <i>LSTM</i> | | 0,41 | 0,49 | 0,37 | 0,56 | 10% | -12% |
| RandomForest | R2 | 0,94 | 0,83 | 0,99 | 0,81 | -5% | 2% |
| XGBoost | R2 | 1,00 | 0,82 | 1,00 | 0,80 | 0% | 2% |
| <i>CatBoost</i> | RMSE | 33,56 | 81,07 | 32,98 | 77,93 | 2% | 4% |
| LightGBM | RMSE | 17,50 | 66,62 | 34,91 | 72,01 | -50% | -7% |
| <i>LinearRegression</i> | RMSE | 84,44 | 81,69 | 84,44 | 81,69 | 0% | 0% |
| <i>LSTM</i> | | 148,02 | 101,19 | 152,75 | 94,04 | -3% | 8% |
| RandomForest | RMSE | 47,88 | 58,79 | 21,79 | 61,62 | 120% | -5% |
| XGBoost | RMSE | 8,27 | 59,74 | 10,48 | 62,96 | -21% | -5% |

На рисунке 1 визуализированы данные, представленные в таблице 2. По результатам анализа данных, представленных в таблице 2 и визуализированных на рисунке 1, можно сделать следующие выводы. Наиболее точными алгоритмами для решаемой задачи прогнозирования спроса на отдельные виды

товаров являются Random Forest, XGBoost и LightGBM. Дополнительная настройка гиперпараметров помогла еще больше повысить эффективность прогнозов. Наибольшей точностью, в таком случае, отличается RandomForest. Алгоритм CatBoost выдал средние показатели, сравнимые с базовым алгоритмом линейной регрессии. Стоит также отметить, что настройка гиперпараметров только ухудшила его показатели, что свидетельствует о необходимости более тонкой настройки алгоритма по сравнению с остальными алгоритмами градиентного усиления, которые при настройке только самых основных гиперпараметров показали лучшие результаты. Наихудшие результаты показала нейронная сеть LSTM, не превзойдя по точности даже базовый алгоритм.

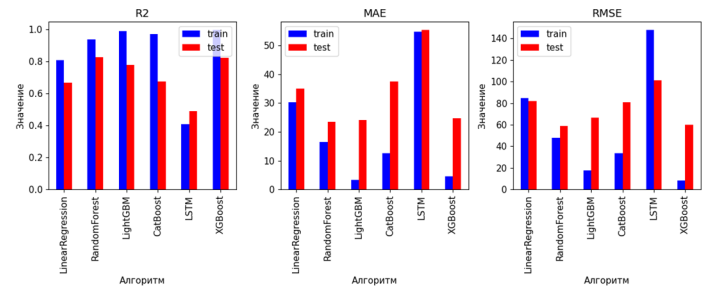


Рис. 1. Графики метрик для Эксперимента №1

При более детальном анализе результатов эксперимента можно отметить следующее. На рисунке 1 представлено сравнение метрик моделей с настроенными гиперпараметрами на обучающей и тестовой выборках. На основании анализа роста ошибок на тестовом наборе данных можно судить о присутствии переобучения моделей, что является неизбежным при применении алгоритмов машинного обучения, особенно при наличии шумов и выбросов в датасете. Имеется необходимость введения дополнительных объясняющих признаков в датасет с одной стороны и необходимость более тонкой настройки алгоритмов с другой.

Модель LSTM при худших результатах в целом, наоборот показала улучшение показателей на тестовых данных, что может свидетельствовать о недообучении. Можно сделать вывод о необходимости конструирования более сложных слоев под конкретные данные и их настройке. Также для модели LSTM может не хватать объема данных при большом количестве признаков в датасете. Однако увеличение объема выборки и усложнение самой модели приведет к значительному росту потребления ресурсов и времени обучения, в то время как другие алгоритмы работают лучше уже при имеющихся данных.

В целом у алгоритмов, показавших лучшую точность, минимальный разброс между значениями метрик тренировочной и тестовой выборки имеет Random Forest, в отличие от XGBoost и LightGBM, которые несмотря на низкие показатели ошибок показали большее переобучение. Однако, алгоритм Random Forest имеет гораздо меньший набор настроек по сравнению с градиентным усилением и в данном случае алгоритм показал порог своей эффективности.

Алгоритмы градиентного усиления имеют множество дополнительных параметров для более тонкой подгонки под данные в части дополнительной регуляризации для устранения влияния выбросов. Их введение, настройка и тестирование гораздо менее ресурсоемки по сравнению с конструированием нейронной сети LSTM.

Среди представленных на рисунке 1 метрик моделей градиентного усиления модель XGBoost, имеет лучшие показатели и больший потенциал к увеличению точности и обобщающей способности на тестовой выборке.

На графике ошибок моделей (рис. 2) видна наибольшая схожесть моделей с лучшими показателями: XGBoost и Random Forest. В целом заметны аномальные выбросы у всех моделей с разной степенью разброса. Модель LSTM стабильно занижает прогнозы относительно фактических значений, что подтверждает необходимость ее доработки, о чем было указано выше.

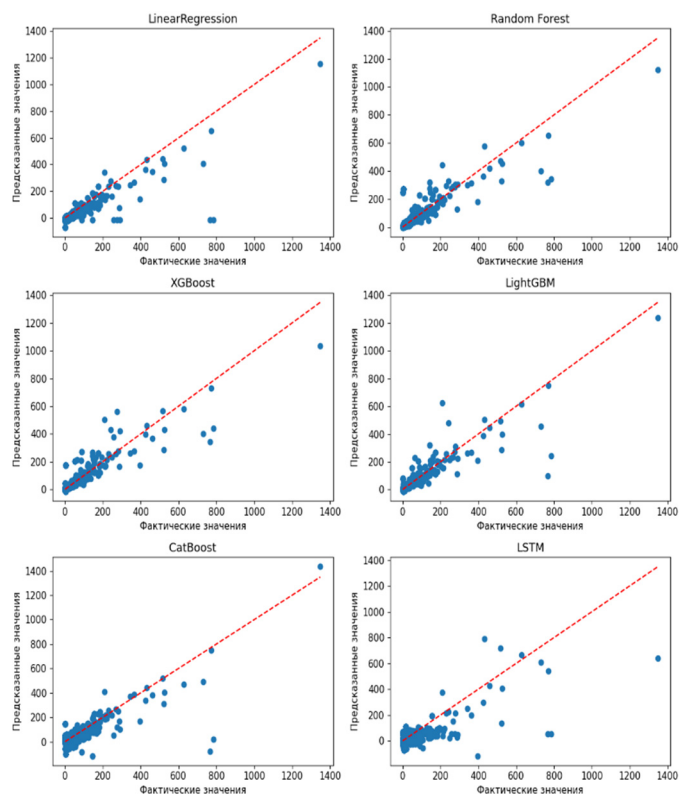


Рис. 2. Графики ошибок предсказаний моделей в Эксперименте №1

Эксперимент №2. В данном эксперименте был использован датасет крупной розничной сети Эквадора (информация из открытых ресурсов платформы «Kaggle»: <https://www.kaggle.com/competitions/store-sales-time-series-forecasting>).

Помимо данных о продажах сети магазинов в датасете предоставлены внешние признаки: цены на нефть и данные о праздниках. Была сделана укрупненная до вида товара выборка за 2 года. Она составила 24 090 записей, включающих 33 товарных групп за 730 дней. Прогноз осуществляется на 30 дней по каждому дню. Тестовая и проверочная выборки содержат по 30 дней. Результаты работы алгоритмов на полных данных для Эксперимента №2 представлены в таблице 3.

На рисунке 3 визуализированы данные, представленные в таблице 3. Графики ошибок предсказаний моделей в Эксперименте №2 приведены на рисунке 4.

Таблица 3

Результаты работы алгоритмов на полных данных для Эксперимента №2

| Алгоритм | Метрика | С настройками | | Без настроек | | Уменьшение ошибки/Прирост точности | |
|------------------|---------|---------------|----------------|--------------|----------------|------------------------------------|------|
| | | train | test | train | test | train | test |
| LinearRegression | MAE | 6293,56 | 4720,75 | 6293,56 | 4720,75 | 0% | 0% |
| RandomForest | MAE | 1067,36 | 2117,09 | 771,51 | 2136,86 | 38% | -1% |
| LightGBM | MAE | 47,59 | 2136,77 | 1092,12 | 2346,57 | -96% | -9% |
| CatBoost | MAE | 1043,07 | 2697,66 | 1474,00 | 2260,66 | -29% | 19% |
| LSTM | | 12204,46 | 7898,41 | 12190,19 | 7761,87 | 0% | 2% |
| XGBoost | MAE | 971,88 | 1846,36 | 156,82 | 3334,45 | 520% | -45% |
| LinearRegression | R2 | 0,90 | 0,98 | 0,90 | 0,98 | 0% | 0% |
| RandomForest | R2 | 0,99 | 0,99 | 1,00 | 0,99 | 0% | 0% |
| LightGBM | R2 | 1,00 | 0,99 | 1,00 | 0,99 | 0% | 0% |
| CatBoost | R2 | 1,00 | 0,99 | 1,00 | 0,99 | 0% | 0% |
| LSTM | | 0,71 | 0,89 | 0,73 | 0,94 | -3% | -5% |
| XGBoost | R2 | 1,00 | 0,99 | 1,00 | 0,96 | 0% | 3% |
| LinearRegression | RMSE | 16994,01 | 7746,31 | 16994,01 | 7746,31 | 0% | 0% |
| RandomForest | MSE | 4222,55 | 6703,20 | 3217,47 | 6840,75 | 31% | -2% |
| LightGBM | MSE | 381,10 | 6522,57 | 2641,55 | 6734,77 | -86% | -3% |
| CatBoost | RMSE | 2365,29 | 6266,11 | 3415,55 | 5968,56 | -31% | 5% |
| LSTM | MSE | 29684,64 | 18365,90 | 28581,47 | 14229,86 | 4% | 29% |
| XGBoost | RMSE | 3320,69 | 5609,92 | 245,25 | 10800,73 | 1254% | -48% |

По результатам анализа данных, представленных в таблице 3 и визуализированных на рисунке 3, можно сделать следующие выводы. На данном датасете в целом все модели показали хороший результат по коэффициенту детерминации R2, однако, наилучшим алгоритмом с настроенными гиперпараметрами оказался XGBoost, при этом настройка гиперпараметров внесла значительный вклад в точность по сравнению со стандартными настройками. CatBoost в данном случае также не лишился проблемы из предыдущего эксперимента относительно ухудшения точности после настройки.

Согласно графику метрик, представленному на рисунке 3, модель XGBoost имеет наилучшие показатели, превзойдя Random Forest. Модель LightGBM оказалась сильно подвержена переобучению, не ухудшив при этом метрику на тестовом наборе данных относительно показателей остальных алгоритмов. Модель LSTM оказалась хуже базового алгоритма Linear Regression.

График ошибок предсказаний (рисунок 4) наглядно демонстрирует указанные показатели, где у модели XGBoost виден наименьший разброс и смещение значений предсказаний среди остальных моделей на основе деревьев решений, а модель LSTM показала аналогичные Эксперименту №1 проблемы, занижая предсказанные значения.

Однако это становится гораздо менее выраженным, поскольку на данном датасете объясняющих данных на каждый вид товара приходится больше.

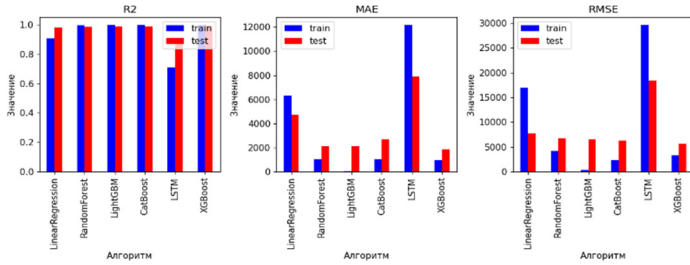


Рис. 3. Графики метрик для Эксперимента №2

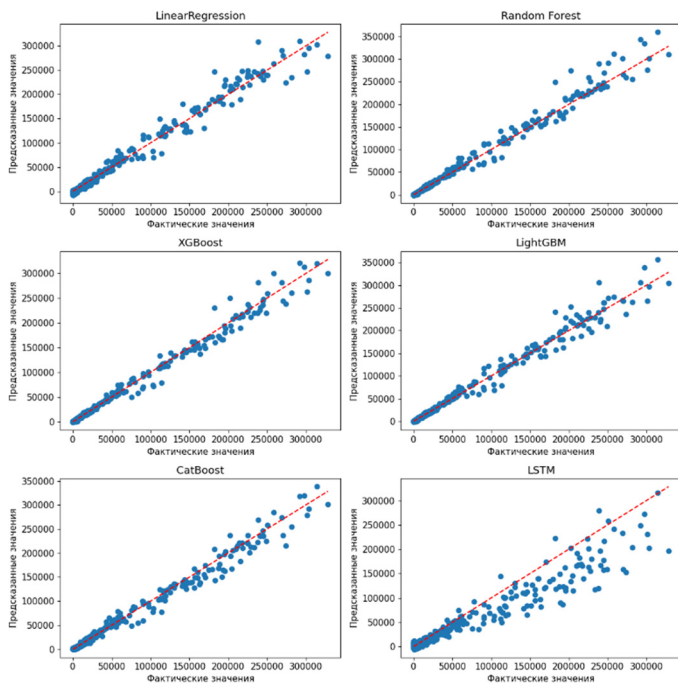


Рис. 4. Графики ошибок предсказаний моделей в Эксперименте №2

Эксперимент №3. На основе представленного датасета (информация из открытых ресурсов платформы «Kaggle»: <https://www.kaggle.com/datasets/mragpavank/predicting-the-sales-of-products-of-a-retail-chain>) о продажах сети розничных магазинов в различных регионах была сформирована выборка по 1 товарному отделу в 1 регионе. Выборка из 15 800 записей содержит информацию о 4 видах товара в 5 магазинах за 790 дней. Прогноз осуществляется на 30 дней с шагом в 1 день. Размер тестовой и проверочной выборок равен 30 дням. Результаты работы алгоритмов на полных данных для Эксперимента №3 представлены в таблице 4. На рисунке 5 визуализированы данные, представленные в таблице 4. Графики ошибок предсказаний моделей в Эксперименте №3 приведены на рисунке 6.

Согласно результатам работы алгоритмов (табл. 4) и графику метрик (рис. 5) можно сделать выводы о сохранении общих закономерностей поведения алгоритмов из предыдущих экспериментов: о худших показателях модели LSTM в целом и небольшом ухудшении качества работы модели CatBoost после настройки гиперпараметров. Однако в данном случае даже отмеченные в таблице показатели на тестовых данных

нельзя назвать лучшими, поскольку все они находятся на одном уровне по эффективности с базовым алгоритмом Linear Regression, который имеет низкие показатели точности.

Таблица 4

Результаты работы алгоритмов на полных данных для Эксперимента №3

| Алгоритм | Метрика | С настройками | | Без настроек | | Уменьшение ошибки/Прирост точности | |
|------------------|---------|---------------|-------------|--------------|-------------|------------------------------------|-------|
| | | train | test | train | test | train | test |
| LinearRegression | MAE | 2,13 | 1,87 | 2,13 | 1,87 | 0% | 0% |
| RandomForest | MAE | 2,08 | 1,97 | 0,74 | 1,97 | 182% | 0% |
| LightGBM | MAE | 0,98 | 1,94 | 1,44 | 2,06 | -32% | -6% |
| CatBoost | MAE | 1,90 | 2,06 | 1,71 | 1,97 | 11% | 5% |
| LSTM | | 2,49 | 2,61 | 2,38 | 2,63 | 5% | -1% |
| XGBoost | MAE | 1,00 | 2,03 | 0,51 | 2,41 | 96% | -16% |
| LinearRegression | R2 | 0,50 | 0,58 | 0,50 | 0,58 | 0% | 0% |
| RandomForest | R2 | 0,53 | 0,59 | 0,93 | 0,59 | -43% | 0% |
| LightGBM | R2 | 0,91 | 0,58 | 0,81 | 0,53 | 12% | 10% |
| CatBoost | R2 | 0,62 | 0,56 | 0,70 | 0,58 | -12% | -4% |
| LSTM | | 0,11 | -0,01 | 0,18 | 0,02 | -40% | -132% |
| XGBoost | R | 0,90 | 0,53 | 0,97 | 0,35 | -7% | 53% |
| LinearRegression | RMSE | 3,45 | 2,94 | 3,45 | 2,94 | 0% | 0% |
| RandomForest | RMSE | 3,32 | 2,90 | 1,25 | 2,91 | 167% | 0% |
| LightGBM | RMSE | 1,49 | 2,95 | 2,14 | 3,14 | -30% | -6% |
| CatBoost | RMSE | 3,01 | 3,03 | 2,65 | 2,95 | 14% | 3% |
| LSTM | | 4,59 | 4,57 | 4,40 | 4,52 | 4% | 1% |
| XGBoost | RMSE | 1,51 | 3,13 | 0,81 | 3,69 | 86% | -15% |

Алгоритмы LightGBM и XGBoost в моделях имеют высокие показатели точности на тренировочных данных, что свидетельствует о наличии у данных алгоритмов большей перспективы для устранения переобучения и повышения точности на тестовой выборке.

Графики ошибок моделей на тестовых данных, представленные на рисунке 6, подтверждают выводы по метрикам, описанные выше. Модели имеют схожие значительные разбросы и смещение, а у модели LSTM присутствует стабильное занижение прогноза.

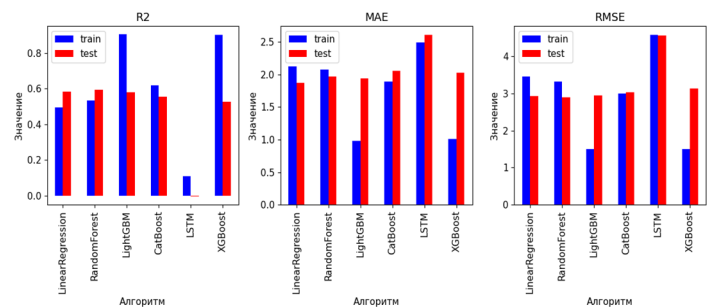


Рис. 5. Графики метрик для Эксперимента №3

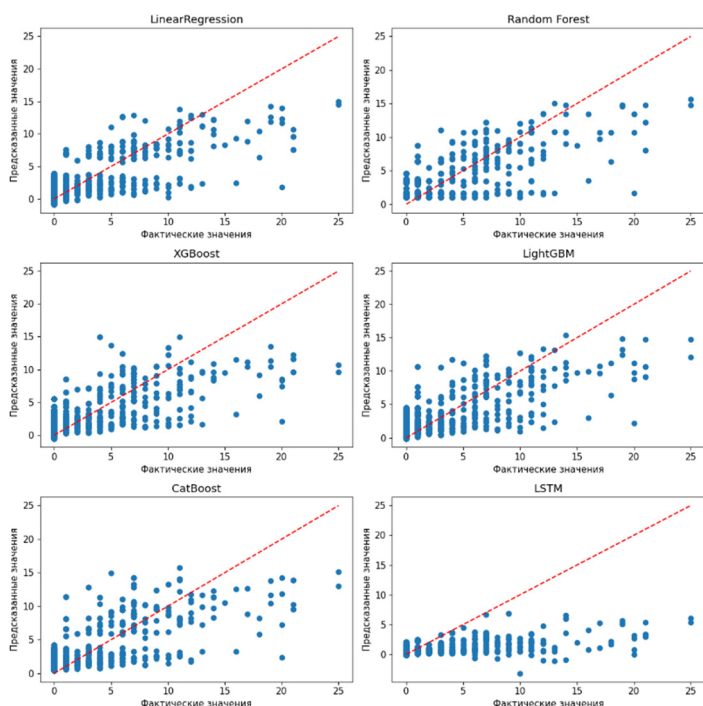


Рис. 6. Графики ошибок предсказаний моделей в Эксперименте №3

По результатам анализа данных из таблицы 4, а также рисунков 5 и 6, можно сделать следующие выводы. В Эксперименте №3 ни одна модель не показала высокую точность. Она для всех моделей оказалась близка по точности к базовому алгоритму Linear Regression. Для корректного прогноза необходимо представить больше переменных, объясняющих шум в данных, поскольку, несмотря на настройку и оптимизацию гиперпараметров, систематические скачки приводят к неправильной работе алгоритмов, которые не могут их интерпретировать ввиду отсутствия соответствующих признаков.

В целом на основании проведенных экспериментов можно сделать выводы о применимости рассматриваемых алгоритмов машинного обучения в задаче прогнозирования спроса на товары, однако важно уделять внимание всем этапам. К примеру, для данных с систематическим шумом, как в Эксперименте №3, необходимо ввести объясняющую переменную для корректных прогнозов.

Важно отметить, что алгоритм RandomForest доказал свою эффективность при относительно небольшом количестве настроек. Параллельное построение деревьев и усреднение результатов позволяют значительно снизить влияние выбросов. Однако, одновременное построение деревьев вызывает проблемы со скоростью его работы. Алгоритмы градиентного усиления XGBoost, CatBoost, LightGBM показали наибольшую гибкость. Относительная сложность и необходимость настройки параметров компенсируется скоростью и точностью результатов. Тем не менее, при настройке гиперпараметров следует проверять эффективность работы на стандартных значениях, во избежание падения эффективности как в случае с CatBoost.

Применение нейронной сети LSTM оказалось неоднозначной. Можно сделать вывод о том, что в общем случае ее применение к данной задаче будет затруднительно. Помимо того, что необходимо большое количество данных и длительное время на обучение, также необходима детальная настройка и

проектирование слоев, подходящих под конкретные данные. Проектирование под каждую конкретную задачу снижает универсальность применения и возможность быстрой адаптации к изменениям.

Делая выбор относительно наиболее универсального алгоритма для рассматриваемой задачи, восприимчивого к доработкам скорости работы, стоит обратить внимание на алгоритмы градиентного усиления, в частности, на XGBoost. Более детальная настройка параметров, а также правильный подбор признаков в данных позволит увеличить эффективность и точность построенных им прогнозов, что будет сделано в дальнейших исследованиях.

Заключение

В настоящей работе была исследована эффективность работы наиболее популярных алгоритмов машинного обучения, применяемых в решении задачи прогнозирования спроса на товары с использованием различных датасетов. Определены основные гиперпараметры и диапазоны значений для настройки алгоритмов и оценен эффект от их настройки. Выявлены систематические недостатки работы отдельных моделей на различных датасетах и определен наилучший алгоритм, который может быть положен в основу дальнейших работ.

Несмотря на то, что некоторые алгоритмы могут быть адаптированы на работу для прогнозирования спроса на отдельные виды товаров без дополнительных настроек, эффективность их работы даже с минимальными настройками повышается, поскольку каждый датасет имеет уникальные характеристики и особенности.

В дальнейшем планируется модифицировать алгоритм градиентного усиления XGBoost для улучшения эффективности его работы при решении задачи прогнозирования спроса на товары с использованием различных датасетов. Планируется провести исследования эффективности работы модифицированной версии алгоритма градиентного усиления XGBoost по сравнению с его базовой версией и рассмотренными в настоящей работе другими алгоритмами и методами МО.

Литература

1. Азаров В.Н., Аниськина Н.Н., Леохин Ю.Л. Методы и этапы проектирования цифрового предприятия // Качество. Инновации. Образование. 2022. № 6(182). С. 3-22. DOI 10.31145/1999-513x-2022-6-03-22. EDN FTHRFF.
2. Азаров В.Н., Бойцов Б.В., Леохин Ю.Л., Чекартев А.В. Цифровая трансформация в период пандемии и постпандемии // Качество и жизнь. 2021. № 3(31). С. 34-52. DOI 10.34214/2312-5209-2021-31-3-34-52. EDN MKVVAS.
3. Вострикова П.В., Рыбка С.О., Рыжкова У.С., Фатхулин Т.Д. Анализ нейросетевых технологий, используемых для улучшения качества изображений // REDS: Телекоммуникационные устройства и системы. 2024. Т. 14, № 1. С. 57-65. EDN WVBDNR.
4. Леохин Ю.Л., Сыроежко А.А. Метод анализа влияния изменений в типах данных на функции разрабатываемого программного обеспечения // ИТ-Стандарт. 2024. № 1(38). С. 13-17. EDN DEFOFB.
5. Леохин Ю.Л., Фатхулин Т.Д. Разработка методов и алгоритма формализации текстового запроса к онлайн-сервисам, генерирующим изображения посредством нейросетевых технологий // Вестник Рязанского государственного радиотехнического университета. 2023. № 85. С. 82-95. DOI 10.21667/1995-4565-2023-85-82-95. EDN PZWYZV.

6. *Леохин Ю.Л., Фатхулин Т.Д., Кожанов М.С.* Анализ и исследование применения нейросетевых технологий для генерации программного кода // Вестник Рязанского государственного радиотехнического университета. 2024. № 87. С. 41-53. DOI 10.21667/1995-4565-2024-87-41-53. EDN НКЕОФХ.
7. *Леохин Ю.Л., Фатхулин Т.Д., Ментус М.В.* Разработка и применение методов распознавания зашумленных аудиофайлов посредством нейросетевых технологий // Вестник Рязанского государственного радиотехнического университета. 2024. № 88. С. 65-73. DOI 10.21667/1995-4565-2024-88-65-73. EDN НМХАСИ.
8. *Митрофанов А.О., Степанов М.Н., Фатхулин Т.Д.* Анализ нейросетевых методов генерации изображения по текстовому запросу // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2022. № 1. С. 19-23. EDN CWRLQA.
9. *Фатхулин Т.Д., Климов Н.Ю., Гежин С.А.* Анализ нейросетевых технологий, позволяющих генерировать текст // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2023. № 1. С. 123-127. EDN МХСТВО.
10. *Фатхулин Т.Д., Кожанов М.С.* Проблемы хранения и управления результатами статического анализа в жизненном цикле разработки безопасного ПО и пути их решения // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2023. № 1. С. 116-122. EDN KFDCNG.
11. *Фатхулин Т.Д., Леонова В.О., Трemasова Л.А.* Анализ нейросетевых технологий, применяемых для web-разработки // REDS: Телекоммуникационные устройства и системы. 2024. Т. 14. № 2. С. 35-41. EDN SDCNKM.
12. *Фатхулин Т.Д., Лушин Е.А.* Анализ развития автоматической генерации кода для web-сервисов // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2023. № 1. С. 128-132. EDN JUEGXP.
13. *Фатхулин Т.Д., Смирнов Д.А., Разумов И.В.* и др. Анализ влияния составляемых текстовых запросов (промптов) на качество изображений, генерируемых нейросетевыми технологиями // Системы синхронизации, формирования и обработки сигналов. 2024. Т. 15. № 2. С. 52-57. EDN TSVMSK.
14. *Фатхулин Т.Д., Фатхулина Г.Г., Ментус М.В.* Разработка методики формирования запроса к нейросети с целью генерации изображений с учетом рекомендаций компьютерной лингвистики // Труды Северо-Кавказского филиала Московского технического университета связи и информатики. 2023. № 1. С. 133-139. EDN PPRTOM.
15. *Polyantseva K., Gorodnichev M., Mitrokhin Y., Moseva M.* On the Problem of Developing a Realistic Road Infrastructure Simulator for Reinforced Learning // 2023 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF), St. Petersburg, Russian Federation, 2023, pp. 1-8, DOI: 10.1109/WECONF57201.2023.10147983.
16. *Moseva M.S., Kharrasov K.R., Gorodnichev M.G.* Neural Network Method for Removing the Effect of Atmospheric Noise in an Image // 2024 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, Russian Federation, 2024, pp. 1-4, DOI: 10.1109/IEEECONF60226.2024.10496730.
17. *Fatkhulin T., Leokhin Y., Mentus M., Kulikova A., Alshawi R.* Analysis of the Basic Image Generation Methods by Neural Networks // 2023 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED), Moscow, Russian Federation, 2023, pp. 1-7, DOI: 10.1109/TIRVED58506.2023.10332668.
18. *Fatkhulin T., Leokhin Y., Mentus M., Salova A., Tremasova L.* Development and Research of Neural Network Methods for Recognizing Noisy Speech Audio Files // 2024 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, Russian Federation, 2024, pp. 1-5, DOI: 10.1109/IEEECONF60226.2024.10496716.
19. *Leokhin Y., Fatkhulin T., Kulikova A.* Research of the Influence of the Semantics and Syntax of a Text Request on the Image Quality Generated by Neural Network Techniques // 2023 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED), Moscow, Russian Federation, 2023, pp. 1-7, DOI: 10.1109/TIRVED58506.2023.10332723.
20. *Leokhin Y., Fatkhulin T., Kozhanov M.* Research of Static Application Security Testing Technique Problems and Methods for Solving Them // 2024 Systems of Signals Generating and Processing in the Field of on Board Communications, Moscow, Russian Federation, 2024, pp. 1-7, DOI: 10.1109/IEEECONF60226.2024.10496755.
21. *Saksonov E.A., Leokhin Y.L., Azarov V.N.* Organization of Information Security in Industrial Internet of Things Systems // 2019 International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS), Sochi, Russia, 2019, pp. 3-7, DOI: 10.1109/ITQMIS.2019.8928442.
22. *Портнов Э.Л., Фатхулин Т.Д.* Технологии достижения высоких скоростей передачи в современных когерентных DWDM-системах связи // T-Comm: Телекоммуникации и транспорт. 2015. Т. 9. № 8. С. 34-37. EDN ULFNRR.
23. *Деарт В.Ю., Фатхулин Т.Д.* Анализ современного состояния транспортных сетей с целью внедрения технологии программно-конфигурируемых сетей (SDN) // T-Comm: Телекоммуникации и транспорт. 2017. Т. 11, № 6. С. 4-9. EDN ZAOKIH.
24. *Деарт В.Ю., Фатхулин Т.Д.* Анализ транспортных программно-конфигурируемых сетей (T-SDN) с управляемым оптическим уровнем с целью получения модели, позволяющей оценить возможность предоставления сервиса Bandwidth on Demand // T-Comm: Телекоммуникации и транспорт. 2018. Т. 12, № 4. С. 35-42. DOI 10.24411/2072-8735-2018-10063. EDN UQHXY.
25. *Деарт В.Ю., Фатхулин Т.Д.* Анализ процесса создания суперканала с необходимой пропускной способностью в сети, построенной по технологии транспортных программно-конфигурируемых сетей (T-SDN) // T-Comm: Телекоммуникации и транспорт. 2018. Т. 12, № 10. С. 23-30. DOI 10.24411/2072-8735-2018-10152. EDN YMWVON.
26. *Богачков И.В.* Обнаружение натяжённых участков в оптических волокнах на основе метода бриллюэновской рефлектометрии // T-Comm: Телекоммуникации и транспорт. 2016. Т. 10. № 12. С. 85-91.
27. *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.
28. *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.
29. *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.
30. *Бугаев А.С., Таташев А.Г., Яшина М.В., Лаэров О.С., Носов Е.А.* Восстановление динамики транспортного потока на основе детерминированно-стохастической модели и данных с интеллектуально транспортными системами // T-Comm: Телекоммуникации и транспорт. 2019. Т. 13. № 10. С. 35-44.
31. *Буслаев А.П., Кучелев Д.А., Яшина М.В.* Динамические системы и математические модели трафика информации // T-Comm: Телекоммуникации и транспорт. 2018. Т. 12. № 3. С. 22-38.

MACHINE LEARNING METHODS IN SOLVING THE PROBLEM OF FORECASTING DEMAND FOR SPECIFIC TYPES OF GOODS

Vladimir M. Vishnevsky, V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Moscow, Russia, vishn@ipu.ru

Yuri L. Leokhin, Moscow Technical University of Communications and Informatics, Moscow, Russia, y.l.leokhin@mtuci.ru

Timur D. Fatkhulin, Moscow Technical University of Communications and Informatics, Moscow, Russia, t.d.fatkhulin@mtuci.ru

Vladimir M. Vishnevsky, Moscow Technical University of Communications and Informatics, Moscow, Russia

Abstract

The paper examines the problem of increasing the efficiency of forecasting demand for certain types of goods, taking into account information processes in trade and sales. The aim of the work is to determine the most effective machine learning method for forecasting demand for certain types of goods. The relevance of the work is due to the fact that traditional statistical methods for forecasting demand do not allow for efficient processing of the currently received volumes of information on sales and trade; also, not all machine learning methods for predicting numerical values are suitable for the subject area under consideration. The object of the study is the emerging demand for certain types of goods. The subject of the study is the effectiveness indicators (metrics) of machine learning methods used to forecast demand for certain types of goods. The work discusses methods and algorithms such as linear regression, random forest, various types of gradient boosting, as well as a recurrent neural network and a neural network with long-short-term memory. The root mean square error, mean absolute error and coefficient of determination were selected as metrics used to evaluate the effectiveness of a particular method during the analysis of the subject area. As a result of the study, the most suitable algorithms for the subject area under consideration were identified, which make it possible to effectively forecast demand for various types of goods. Recommendations are given on possible settings of model parameters that improve their performance. Finally, conclusions are drawn about the applicability of each of the algorithms under consideration and prospects for further research are outlined. The methodological basis of the work consists of the following methods: analysis, matching, comparison, experiment and generalization.

Keywords: method, algorithm, machine learning, metric, efficiency, demand, indicator, good.

References

- [1] V.N. Azarov, N.N. Anis'kina, Yu.L. Leokhin, "Methods and Stages of Designing a Digital Enterprise," *Kachestvo. Innovacii. Obrazovanie*. 2022. No. 6. pp. 3-22, DOI 10.31145/1999-513x-2022-6-03-22. (in Russian)
- [2] V.N. Azarov, B.V. Bojcov, Yu.L. Leokhin, A.V. Chekmarev, "Digital Transformation in the Period of Pandemic and Post-Pandemic," *Kachestvo i zhizn'*. 2021. No. 3, pp. 34-52, DOI 10.34214/2312-5209-2021-31-3-34-52. (in Russian)
- [3] P.V. Vostrikova, S.O. Rybka, U.S. Ryzhkova, T.D. Fatkhulin, "Analysis of Neural Network Technologies Used to Improve Image Quality," *REDS: Telekommunikacionnye ustrojstva i sistemy*. 2024. Vol. 14. No. 1, pp. 57-65. (in Russian)
- [4] Yu.L. Leokhin, A.A. Syroezhko, "Method for Analyzing the Impact of Changes in Data Types on the Functions of the Software Being Developed," *IT-Standart*. 2024. No. 1, pp. 13-17. (in Russian)
- [5] Yu.L. Leokhin, T.D. Fatkhulin, "Development of Methods and Algorithm for Formalizing a Text Request to Online Services that Generate Images Using Neural Network Technologies," *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*. 2023. No. 85, pp. 82-95, DOI 10.21667/1995-4565-2023-85-82-95. (in Russian)
- [6] Yu.L. Leokhin, T.D. Fatkhulin, M.S. Kozhanov, "Analysis and Research of the Use of Neural Network Technologies for Generating Program Code," *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*. 2024. No. 87, pp. 41-53, DOI 10.21667/1995-4565-2024-87-41-53. (in Russian)
- [7] Yu.L. Leokhin, T.D. Fatkhulin, M.V. Mentus, "Development and Application of Methods for Recognizing Noisy Audio Files Through Neural Network Technologies," *Vestnik Rjazanskogo gosudarstvennogo radiotekhnicheskogo universiteta*. 2024. No. 88, pp. 65-73, DOI 10.21667/1995-4565-2024-88-65-73. (in Russian)
- [8] A.O. Mitrofanov, M.N. Stepanov, T.D. Fatkhulin, "Analysis of Neural Network Methods of Image Generation by Text Requests," *Trudy Severo-Kavkazskogo filiala Moskovskogo tehničeskogo universiteta svjazi i informatiki*. 2022. No. 1, pp. 19-23. (in Russian)
- [9] T.D. Fatkhulin, N.Yu. Klimov, S.A. Gezhin, "Analysis of Neural Network Technologies that Allow Generating Text," *Trudy Severo-Kavkazskogo filiala Moskovskogo tehničeskogo universiteta svjazi i informatiki*. 2023. No. 1, pp. 123-127. (in Russian)
- [10] T.D. Fatkhulin, M.S. Kozhanov, "Problems of Storage and Management of Static Analysis Results in the Life Cycle of Secure Software Development and Ways to Solve Them," *Trudy Severo-Kavkazskogo filiala Moskovskogo tehničeskogo universiteta svjazi i informatiki*. 2023. No. 1, pp. 116-122. (in Russian)
- [11] T.D. Fatkhulin, V.O. Leonova, L.A. Tremasova, "Analysis of Neural Network Technologies used for Web Development," *REDS: Telekommunikacionnye ustrojstva i sistemy*. 2024. Vol. 14. No. 2, pp. 35-41. (in Russian)
- [12] T.D. Fatkhulin, E.A. Lushin, "Analysis of the Development of Automatic Code Generation for Web Services," *Trudy Severo-Kavkazskogo filiala Moskovskogo tehničeskogo universiteta svjazi i informatiki*. 2023. No. 1, pp. 128-132. (in Russian)
- [13] T.D. Fatkhulin et al., "Analysis of the Influence of Compiled Text Requests (Prompt) on the Quality of Images Generated by Neural Network Technologies," *Sistemy sinhronizacii, formirovaniya i obrabotki signalov*. 2024. Vol. 15. No. 2, pp. 52-57. (in Russian)

- [14] T.D. Fatkhulin, G.G. Fatkhulina, M.V. Mentus, "Development of a Methodology for Forming a Request to a Neural Network in Order to Generate Images Taking Into Account the Recommendations of Computational Linguistics," *Trudy Severo-Kavkazskogo filiala Moskovskogo tehnikeskogo universiteta svyazi i informatiki*. 2023. No. 1, pp. 133-139. (in Russian)
- [15] K. Polyantseva, M. Gorodnichev, Y. Mitrokhin and M. Moseva, "On the Problem of Developing a Realistic Road Infrastructure Simulator for Reinforced Learning," *2023 Wave Electronics and its Application in Information and Telecommunication Systems (WECONF)*, St. Petersburg, Russian Federation, 2023, pp. 1-8, DOI: 10.1109/WECONF57201.2023.10147983.
- [16] M.S. Moseva, K.R. Kharrasov and M.G. Gorodnichev, "Neural Network Method for Removing the Effect of Atmospheric Noise in an Image," *2024 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, Russian Federation, 2024, pp. 1-4, DOI: 10.1109/IEEECONF60226.2024.10496730.
- [17] T. Fatkhulin, Y. Leokhin, M. Mentus, A. Kulikova and R. Alshawi, "Analysis of the Basic Image Generation Methods by Neural Networks," *2023 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED)*, Moscow, Russian Federation, 2023, pp. 1-7, DOI: 10.1109/TIRVED58506.2023.10332668.
- [18] T. Fatkhulin, Y. Leokhin, M. Mentus, A. Salova and L. Tremasova, "Development and Research of Neural Network Methods for Recognizing Noisy Speech Audio Files," *2024 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, Russian Federation, 2024, pp. 1-5, DOI: 10.1109/IEEECONF60226.2024.10496716.
- [19] Y. Leokhin, T. Fatkhulin and A. Kulikova, "Research of the Influence of the Semantics and Syntax of a Text Request on the Image Quality Generated by Neural Network Techniques," *2023 Intelligent Technologies and Electronic Devices in Vehicle and Road Transport Complex (TIRVED)*, Moscow, Russian Federation, 2023, pp. 1-7, DOI: 10.1109/TIRVED58506.2023.10332723.
- [20] Y. Leokhin, T. Fatkhulin and M. Kozhanov, "Research of Static Application Security Testing Technique Problems and Methods for Solving Them," *2024 Systems of Signals Generating and Processing in the Field of on Board Communications*, Moscow, Russian Federation, 2024, pp. 1-7, DOI: 10.1109/IEEECONF60226.2024.10496755.
- [21] E.A. Saksonov, Y.L. Leokhin and V.N. Azarov, "Organization of Information Security in Industrial Internet of Things Systems," *2019 International Conference "Quality Management, Transport and Information Security, Information Technologies" (IT&QM&IS)*, Sochi, Russia, 2019, pp. 3-7, DOI: 10.1109/ITQMIS.2019.8928442.
- [22] 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)
- [23] V.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)
- [24] V.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)
- [25] 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)
- [26] I.V. Bogachkov, "Detection of strained areas in optical fibers based on the Brillouin reflectometry method," *T-Comm*. 2016. Vol. 10. No. 12, pp. 85-91.
- [27] 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.
- [28] 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.
- [29] 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.
- [30] A.S. Bugaev, A.G. Tatashev, M.V. Yashin , O.S. Lavrov, E.A. Nosov, "Reconstruction of the dynamics of traffic flow based on a deterministic-stochastic model and data from intelligent transport systems," *T-Comm*. 2019. Vol. 13. No. 10, pp. 35-44.
- [31] A.P. Buslaev, D.A. Kuchelev, M.V. Yashina, "Dynamic systems and mathematical models of information traffic," *T-Comm*. 2018. Vol. 12. No. 3, pp. 22-38.

Information about authors:

Vladimir M. Vishnevsky, V. A. Trapeznikov Institute of Control Sciences of Russian Academy of Sciences, Laboratory 69, Principal Investigator, full professor, Dr. Sc. (Tech.), Moscow, Russia. orcid.org/0000-0001-7373-4847

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

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

Alexey V. Zanegin, Moscow Technical University of Communications and Informatics, Dpt. of MC and IT, master's student, Moscow, Russia. orcid.org/0009-0005-5518-7882