

Адаптивные подходы к оптимизации цены подшипниковой продукции

Выпускная работа

Динара Сабирова

29 ноября 2025 г.

Оглавление

- 1 Описание цели и задачи проекта
- 2 AS IS
- 3 TO BE
- 4 Методология работы
- 5 Результаты
- 6 Дальнейшее развитие проекта

Цель

Автоматизированное установление цен на подшипниковую продукцию.

Гипотеза

Внедрение модели машинного обучения в процесс ценообразования позволит повысить валовую прибыль по высоколиквидным позициям на 5 - 15% , за счет ускоренного процесса установки цены и ее дальнейшей оптимизации.

Задача

Разработка и внедрение модели для автоматизации процесса ценообразования на высоколиквидные товары в режиме реального времени.

Оглавление

- 1 Описание цели и задачи проекта
- 2 AS IS
- 3 TO BE
- 4 Методология работы
- 5 Результаты
- 6 Дальнейшее развитие проекта

Из чего формируется цена?



Формирование цены

Модель расчёта цены

Цена на подшипниковую продукцию формируется из перечня факторов влияющих на цену:

$$\text{Price} \sim (T; I; D; ER; C; M)$$

- **Transport** — расходы на доставку и логистику;
- **Insurance** — страхование поставки;
- **Duties** — таможенные пошлины;
- **Exchange Rates** — валютные колебания и риски;
- **Competitors** — цены конкурентов на аналогичную продукцию;
- **Markup** — фиксированная наценка компании.

Детали процесса

- Источники данных
 - Внутренние данные о продажах
 - Информация о конкурентах
 - Рыночные тренды и спрос на продукцию
- Методы анализа
 - Статистический анализ продаж
 - Сравнительный анализ с конкурентами
- Инструменты аналитики
 - Excel
- Процессы принятия решений
 - Ручное внесение изменений в цене (В среднем 3000 позиций, циклично обновляются в течении 15 дней, 150 позиций в день).

Участники процесса



Оглавление

- 1 Описание цели и задачи проекта
- 2 AS IS
- 3 TO BE
- 4 Методология работы
- 5 Результаты
- 6 Дальнейшее развитие проекта

Этапы

- Сбор данных
- Предварительная обработка данных
- Анализ данных и EDA (Exploratory Data Analysis)
- Разработка ML моделей установления цены
- Валидация и тестирование моделей
- Выбор финального пайплайна

Основные принципы процесса



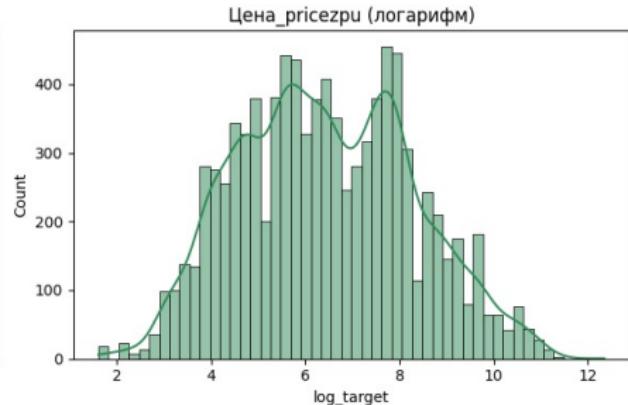
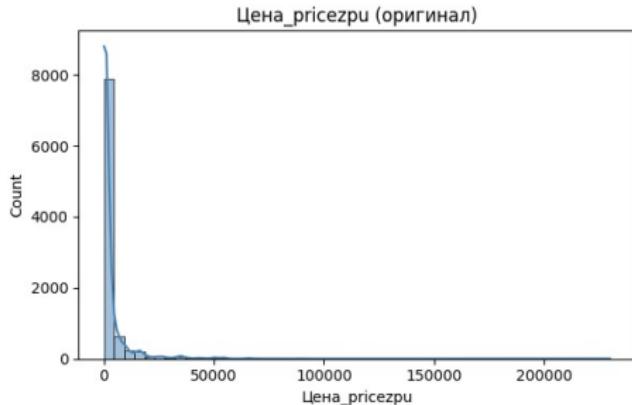
Оглавление

- 1 Описание цели и задачи проекта
- 2 AS IS
- 3 TO BE
- 4 Методология работы
- 5 Результаты
- 6 Дальнейшее развитие проекта

Описание данных

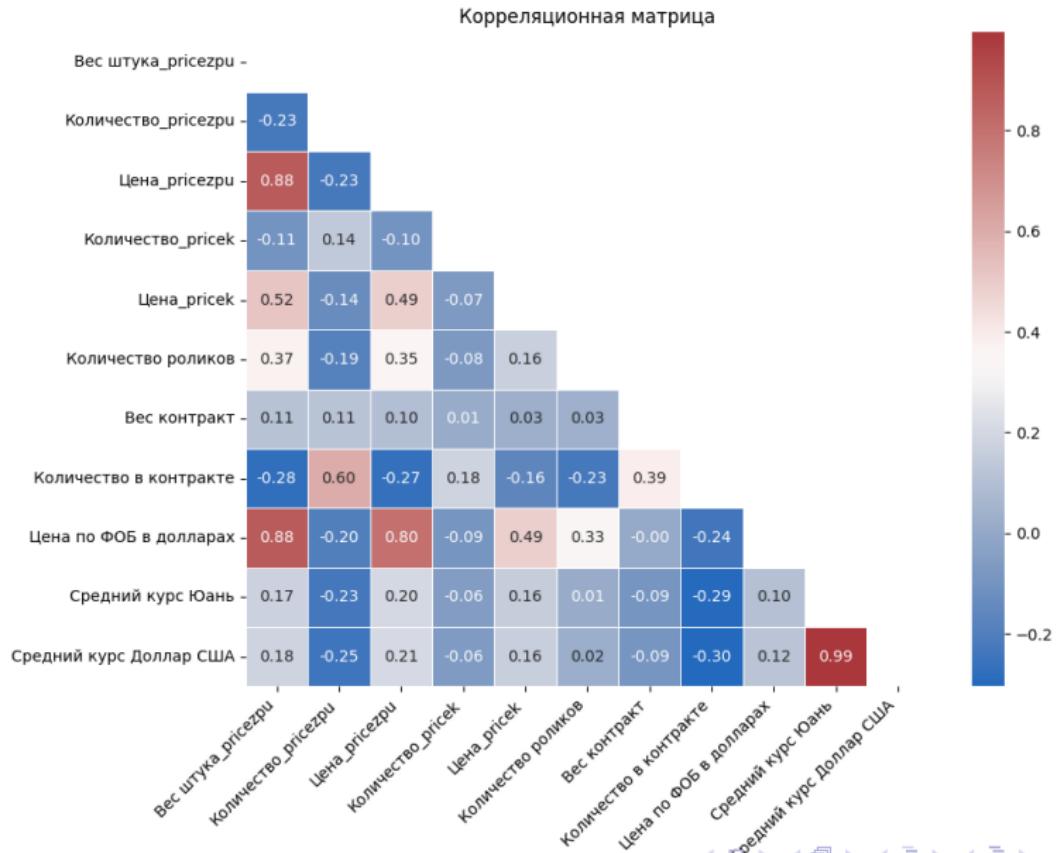
- Датасет:
 - 9 346 строк, 40+ признаков. Источник — исторические продажи подшипниковой продукции и внешние прайс-листы конкурентов за период 03.06.2024 - 30.05.2025.
- Основной таргет:
 - Price — фактическая цена реализации за единицу продукции.
- Типы признаков:
 - непрерывные: вес, количество в контракте, себестоимость, курс валюты, наценка;
 - категориальные: тип подшипника, производитель, страна, клиент.

EDA: Распределение таргета



- Вывод: Большинство цен сконцентрированы в диапазоне до 30 000 ₽, распределение имеет длинный правый хвост, что говорит о наличии выбросов в виде более дорогостоящих позиций. Для наглядности выведен график с логарифмированным таргетом.

EDA: Корреляционная матрица



Вывод

- Цена по FOB в долларах ↔ Цена_pricek = 0.49.
Отражает рыночную концентрацию на прибыльных сегментах.
- Количество_pricezpri ↔ Количество_pricek = 0.14.
Это отражает разную стратегию управления ассортиментом — ЗПУ делает ставку на стабильные модели, а рынок более динамичен.
- Цена_pricezpri ↔ Количество_pricek = -0.10. Это отражает конкурентное давление в массовом сегменте — ЗПУ снижает цену, чтобы удерживать позиции в категориях с активной конкуренцией.

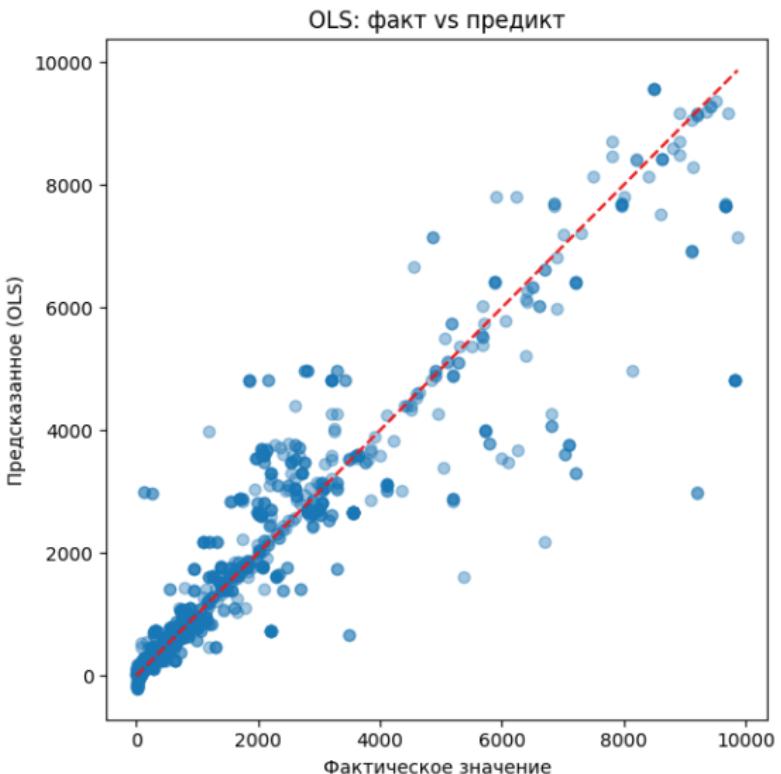
Бейзлайн (OLS): результаты

Уравнение линейной регрессии

$$\hat{Y} = \beta_0 + \beta_1 \cdot \text{В} + \beta_2 \cdot \text{С} + \beta_3 \cdot \text{Кол} + \beta_4 \cdot \text{Курс} + \varepsilon$$

- R^2_{train} : 0.84; R^2_{test} : 0.87
- **Durbin-Wu-Hausman**: $p < 0.05 \rightarrow$ не принимаем H_0
→ переменные эндогенны → нужна IV-регрессия
- **Breusch-Pagan**: $p < 0.05 \rightarrow$ не принимаем $H_0 \rightarrow$ присутствует гетероскедастичность → нужны
робастные ошибки (HC3)
- **Основные значимые факторы**: Цена по FOB в
долларах: 0.64, Цена_pricek: 0.23, Вес
штука_pricezpu: 0.18, Количество_pricezpu: 0.05

OLS: факт vs предикт



Модели

Описание

Для сравнения с бейзлайном были опробованы следующие модели:

- **Ансамбли:** Bagging, Gradient Boosting;
- **Нейросеть (MLP);**
- Цель — улучшить точность прогнозирования относительно OLS.

Оглавление

- 1 Описание цели и задачи проекта
- 2 AS IS
- 3 TO BE
- 4 Методология работы
- 5 Результаты
- 6 Дальнейшее развитие проекта

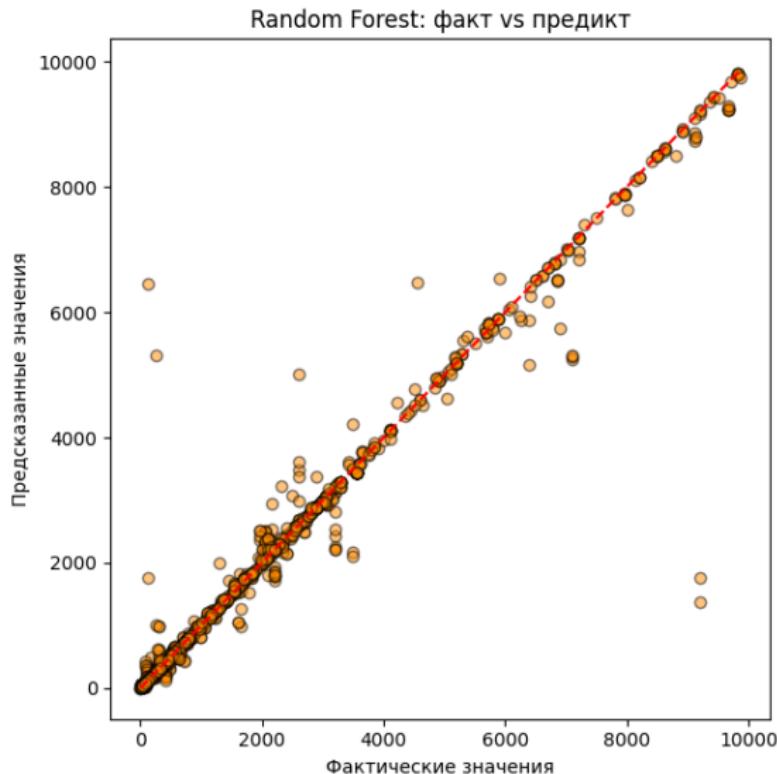
Расширенные модели: OLS, Bagging, Boosting, NN

Сравнение

Model	R ² (test)	RMSE (₽)	MAPE (%)
OLS	0.87	725.27	65.18
Random Forest	0.97	369.27	13.71
Gradient Boosting	0.94	496.68	37.81
Neural Network (MLP)	0.87	718.51	39.35

- **Характеристики целевой переменной:** mean = 1384.73, 25-й перцентиль= 128.00, 50-й перцентиль = 470.00, 75-й перцентиль = 1900.00

Random Forest: факт vs предикт



Вывод

- На формирование цены влияют Цена по FOB в долларах, Цена_pricek, Вес штука_pricezpu, Количество_pricezpu
- Для проведения пилота по набору метрик лучший результат показала модель Random Forest.

Оглавление

- 1 Описание цели и задачи проекта
- 2 AS IS
- 3 TO BE
- 4 Методология работы
- 5 Результаты
- 6 Дальнейшее развитие проекта

Планы

ИНТЕГРАЦИЯ И АВТОМАТИЗАЦИЯ

ETL-конвейер для автоматического обновления данных из ERP/1С, прайсов конкурентов и валютных API.
Ежедневный пересчёт цен и выгрузка в CRM / систему КП.
Дашборд мониторинга: прибыль, скорость обновления, аномалии.



РАСШИРЕНИЕ ДАННЫХ И ФАКТОРОВ

Подключение новых источников (индексы, тарифы, сезонность).
Сегментация моделей по категориям товаров.
Применение временных рядов (LSTM) для прогноза цен на 1–3 месяца.



ПИЛОТ

Тестирование модели на ограниченной группе товаров.
Сравнение прогнозных и фактических цен, оценка экономического эффекта.
Корректировка параметров модели и бизнес-логики на основе обратной связи.



ПРОМ

Внедрение системы в промышленную эксплуатацию.
Интеграция с 1С, CRM и BI-панелью в реальном времени.
Автоматическая оптимизация цен с учётом валют, спроса и конкурентов.



Спасибо за внимание!