

Интеллектуальный анализ данных цен на смарт-часы

Гиренко Александр Евгеньевич

Приамурский государственный университет им.Шолом-Алейхема

Студент

Аннотация

В статье проводится интеллектуальный анализ данных для прогнозирования цен на смарт-часы на основе их технических характеристик. Сформирован и обработан набор данных из 500 устройств, включающий такие параметры, как тип экрана, тип часов, емкость аккумулятора, вес и класс пыли-влагозащиты. В практической части реализовано 25 экспериментов с использованием более чем 15 алгоритмов машинного обучения: линейной регрессии, гребневой регрессии, деревьев решений, случайного леса и градиентного бустинга. Проведен сравнительный анализ качества моделей, выполнена интерпретация значимости признаков и выявлены ключевые факторы ценообразования в сегменте смарт-часов.

Ключевые слова: Интеллектуальный анализ данных, машинное обучение, прогнозирование цен, смарт-часы, регрессионный анализ, XGBoost, важность признаков.

Intelligent analysis of smartwatch price data

Girenko Aleksandr Evgen'evich

Sholom-Aleichem Priamursky State University

Student

Abstract

The article provides an intelligent analysis of data on forecasting prices for smartwatches based on their technical characteristics. A dataset of 500 devices was compiled and processed, including parameters such as display type, device type, battery capacity, weight, and dust/water resistance class. In the practical part, 25 experiments were implemented using over 15 machine learning algorithms: linear regression, ridge regression, decision trees, random forest, and gradient boosting. A comparative analysis of model quality was conducted, feature importance was interpreted, and key pricing factors in the smartwatch segment were identified.

Keywords: Data mining, machine learning, price forecasting, smartwatches, regression analysis, XGBoost, feature importance.

1 Введение

1.1 Актуальность

Рынок смарт-часов является одним из наиболее динамично развивающихся секторов потребительской электроники. За последние пять лет

эти устройства превратились из нишевого аксессуара в массовый продукт, сочетающий в себе функции фитнес-трекера, мониторинга здоровья, коммуникационного хаба и элемента экосистемы «умного дома». Ежегодно производители выпускают сотни моделей, различающихся по дизайну, функционалу, техническим характеристикам и стоимости. В условиях высокой конкуренции и огромного ассортимента задача объективной оценки стоимости устройства на основе его параметров становится критически важной как для производителей и ритейлеров, так и для конечных потребителей. Интеллектуальный анализ данных (Data Mining) и алгоритмы машинного обучения позволяют выявить скрытые закономерности в ценообразовании и создать инструменты для точного прогнозирования рыночной стоимости.

1.2 Обзор исследований

В работе Ю.Н. Редченко и Т.В. Богачева проведен анализ и прогнозирование цен объектов недвижимости в регионах России с использованием методов интеллектуального анализа данных и библиотеки Scikit-learn [1]. Г.А. Амирханова, М.Е. Мансурова, Д.К. Кобегенова и А.М.Адилбекова исследуют применение методов интеллектуального анализа данных для оптимизации и визуализации расчета и интерпретации индекса потребительских цен, что позволяет повысить точность оценки инфляционных процессов [2]. Ю.В. Скутина рассматривает методы и инструменты для анализа цен на рынке недвижимости, систематизируя подходы к оценке стоимости объектов [3]. М.А. Михеенкова в своем исследовании описывает подход к интеллектуальному анализу данных в социальных науках, акцентируя внимание на методе JSM и формализованном качественном анализе [4]. С.В. Шайтура в своей работе представляет обзор методов и технологий интеллектуального анализа данных, рассматривая основные подходы к обработке данных [5]. Д.М. Назаров в монографии «Цифровые технологии и интеллектуальные модели» систематизирует современные подходы к применению интеллектуальных моделей в экономике [6]. Ю.Г.Чернышева, А.В. Хирная и А.А. Головки исследуют применение интеллектуального анализа данных (Data Mining) в бизнес-анализе, демонстрируя возможности использования методов обработки данных для решения прикладных задач в экономике [7].

1.3 Цель исследования

Целью исследования является проведение интеллектуального анализа данных (Data Mining) рынка смарт-часов для выявления ключевых технических факторов, влияющих на их стоимость, а также построение и сравнительный анализ прогностических моделей машинного обучения для высокоточного определения рыночной цены устройства на основе его характеристик.

2 Подготовка данных и разведочный анализ (EDA)

2.1 Общая характеристика и очистка набора данных

Для проведения исследования используется набор данных, содержащий информацию о 500 различных моделях смарт-часов. Датасет был собран автором путём парсинга Яндекс Маркета. В датасет включены следующие параметры: название модели, тип часов, цена в рублях, тип экрана, ёмкость аккумулятора, вес, класс пыле-влагозащиты и ссылка на страницу товара. (Рис.1).

✓ Файл загружен из: /content/sample_data/smart_watches_500.csv

Размер датасета: 500 строк, 9 столбцов

Названия столбцов:
['№', 'Название', 'Тип часов', 'Цена (₽)', 'Тип экрана', 'Ёмкость аккумулятора', 'Вес', 'Класс пыле-влагозащиты', 'Ссылка']

Проверка наличия URL товара:
✓ Столбец со ссылками найден: 'Ссылка'
Пример: <https://market.yandex.ru/card/umnyye-chasy-huawei-fit-se-chernyy/103825069752?do-waremd5=zhT5qWVojJe...>

Первые 5 строк:

№	Название	Тип часов
0	Умные часы HUAWEI FIT SE, (STA-B39), Сияющий ч...	Умные часы
1	Kingwear Мужские наручные смарт-часы черные с ...	Смарт часы
2	Смарт часы HUAWEI Band 11 Pro , Зеленый цвет, ...	Смарт часы
3	Умные смарт часы Watch X мужские /женские, для...	Умные смарт часы
4	Умные часы Smart Watch, фитнес-трекер, совмест...	Умные часы

	Цена (₽)	Тип экрана	Ёмкость аккумулятора	Вес	Класс пыле-влагозащиты
0	4013	amoled	144	21.0	WR50
1	4398	amoled	400	122.0	IP67
2	4467	amoled	300	18.0	5 ATM
3	5944	Super AMOLED	240	37.0	IP68
4	2277	LCD	420	115.0	IP67

	Ссылка
0	https://market.yandex.ru/card/umnyye-chasy-hua...
1	https://market.yandex.ru/card/kingwear-muzhski...
2	https://market.yandex.ru/card/fitnes-braslet-h...
3	https://market.yandex.ru/card/umnyye-smart-cha...
4	https://market.yandex.ru/card/umnyye-chasy-sma...

Типы данных:

№	int64
Название	object
Тип часов	object
Цена (₽)	int64
Тип экрана	object
Ёмкость аккумулятора	object
Вес	float64
Класс пыле-влагозащиты	object
Ссылка	object
dtype:	object

Рисунок 1 – Общая информация о датасете

Первым шагом выполнена очистка данных от неинформативных признаков. Столбцы, содержащие ссылки на веб-страницы и названия моделей, удалены, так как они не несут ценности для алгоритмов машинного обучения. Также проведен аудит пропущенных значений. По результатам аудита выявлены незначительные пропуски в трех столбцах (Рис. 2).

СТАТИСТИЧЕСКОЕ ОПИСАНИЕ ДАННЫХ			
	№	Цена (Р)	Вес
count	500.000000	500.000000	498.000000
mean	250.500000	5109.44600	58.756426
std	144.481833	4291.05617	32.720660
min	1.000000	1276.00000	16.500000
25%	125.750000	3346.00000	40.600000
50%	250.500000	3857.00000	48.000000
75%	375.250000	5225.50000	61.400000
max	500.000000	41768.00000	250.000000
АНАЛИЗ ПРОПУСКОВ			
	Пропуски		%
Емкость аккумулятора	2		0.4
Вес	2		0.4
Класс пыли-влагозащиты	8		1.6
ЗАПОЛНЕНИЕ ПРОПУСКОВ			
Емкость аккумулятора: заполнено модой (2 пропусков)			
Вес: заполнено медианой (2 пропусков)			
Класс пыли-влагозащиты: заполнено модой (8 пропусков)			

Рисунок 2 – Первичная обработка данных

Пропуски в категориальном признаке заполнены значением моды, в числовых — медианными значениями (Табл. 1).

Таблица 1 – Анализ пропусков и методы их заполнения

Признак	Тип данных	Пропущено значений	Доля, %	Метод заполнения
Емкость аккумулятора	Числовой	2	0.4%	Медиана
Вес	Числовой	2	0.4%	Медиана
Класс пыли-влагозащиты	Категориальный	8	1.6%	Мода

2.2 Изучение распределения целевой переменной

Для понимания природы ценообразования изучено распределение цен. С помощью гистограммы с наложенной кривой плотности (KDE) выявлено, что распределение имеет ярко выраженную правостороннюю асимметрию. Средняя цена (5 109.45 Р) значительно превышает медианную (3 857.00 Р), что указывает на наличие в выборке небольшого количества очень дорогих премиальных моделей при преобладании бюджетных решений (Рис. 3).

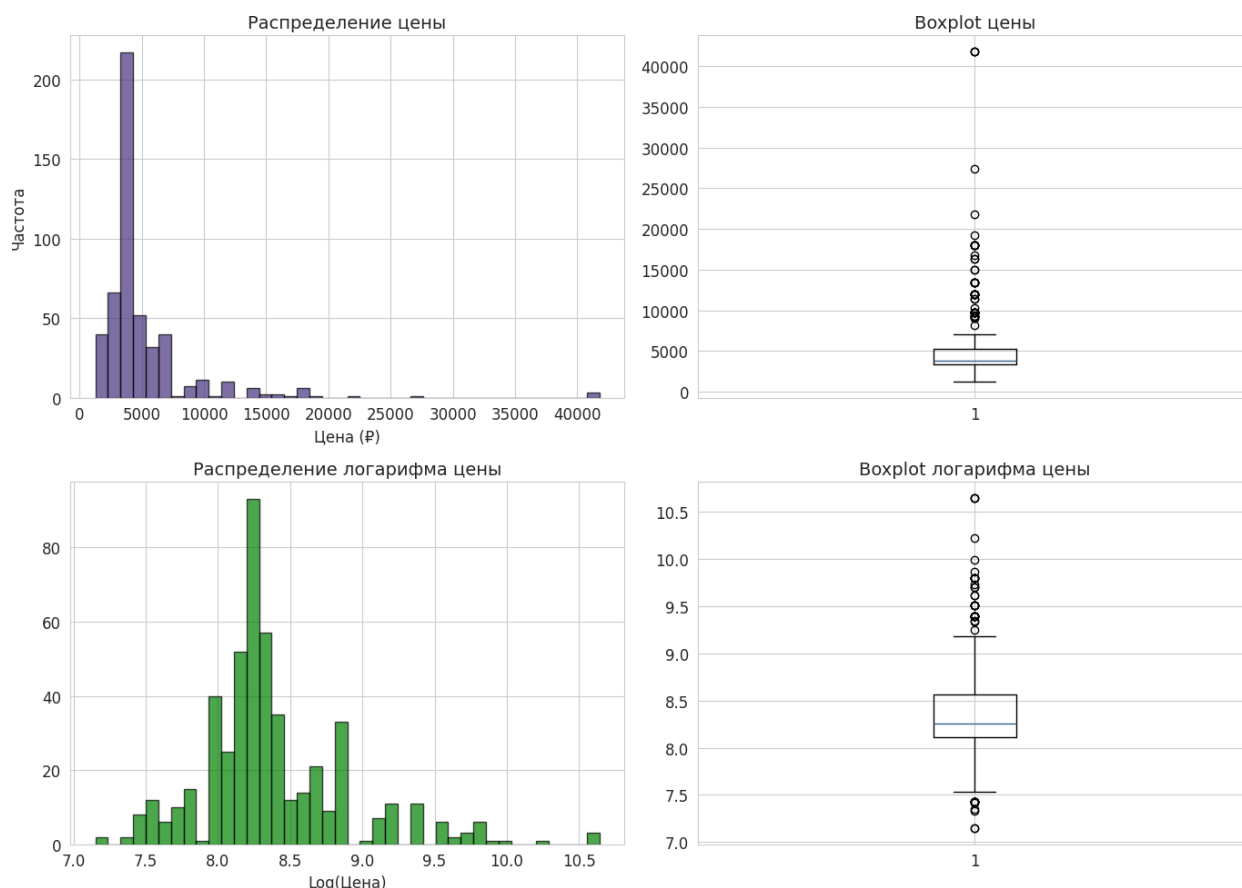


Рисунок 3 – График распределения цен смарт-часов

Сводная статистика по ключевым числовым признакам (табл. 2).

Таблица 2 – Статистическое описание числовых признаков датасета

Показатель	Цена, Р	Вес, г
Среднее	5 109.45	58.76
Стандартное отклонение	4 291.06	32.72
Минимум	1 276.00	16.50
25-й перцентиль	3 346.00	40.60
Медиана	3 857.00	48.00
75-й перцентиль	5 225.50	61.40
Максимум	41 768.00	250.00

2.3 Анализ и обработка выбросов

Важным этапом подготовки данных стал анализ аномалий (выбросов), которые могут негативно исказить результаты линейных моделей. Анализ показал наличие 53 записей за пределами установленных границ, что составило 10.60% от выборки. К ним относятся эксклюзивные модели премиум-сегмента, цена которых обусловлена не столько техническими характеристиками, сколько престижностью бренда. Для обеспечения стабильности моделей принято решение об обрезке выборки по 99-му перцентилю. Это позволило сконцентрировать обучение на массовом сегменте

рынка, повысив точность прогнозирования для абсолютного большинства устройств (Рис. 4, 5).

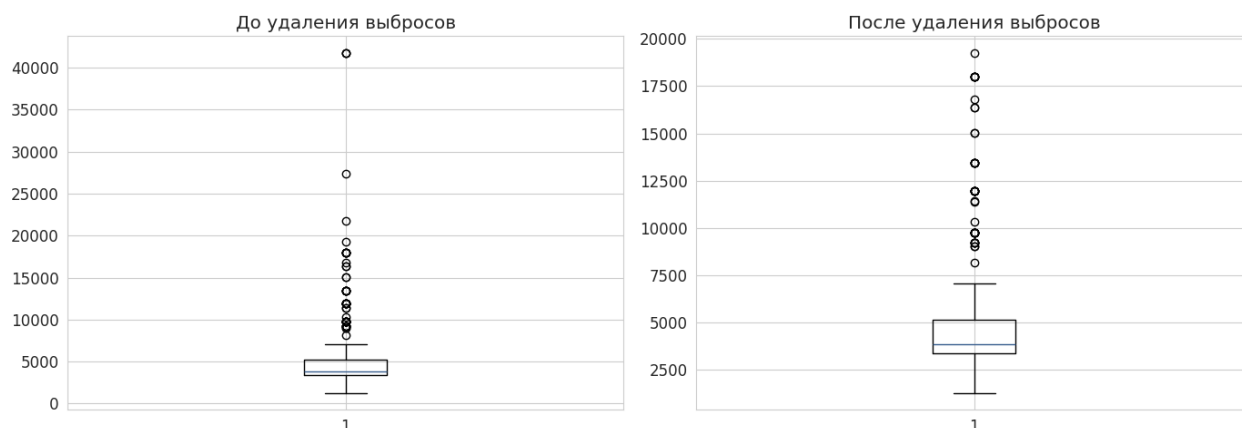


Рисунок 4 – Boxplot-диаграммы цены до и после удаления выбросов

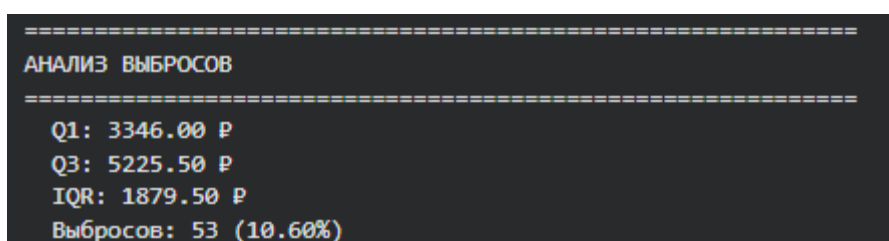


Рисунок 5 – Процентный анализ выбросов

2.4 Исследование парных зависимостей и корреляций

С помощью тепловой карты проанализирована корреляционная матрица всех числовых признаков. Коэффициент корреляции Пирсона между ценой и емкостью аккумулятора составил -0.08 , между ценой и весом — 0.10 . Анализ подтвердил отсутствие критической мультиколлинеарности между техническими параметрами, что делает набор данных пригодным для регрессионного анализа (Рис. 6).

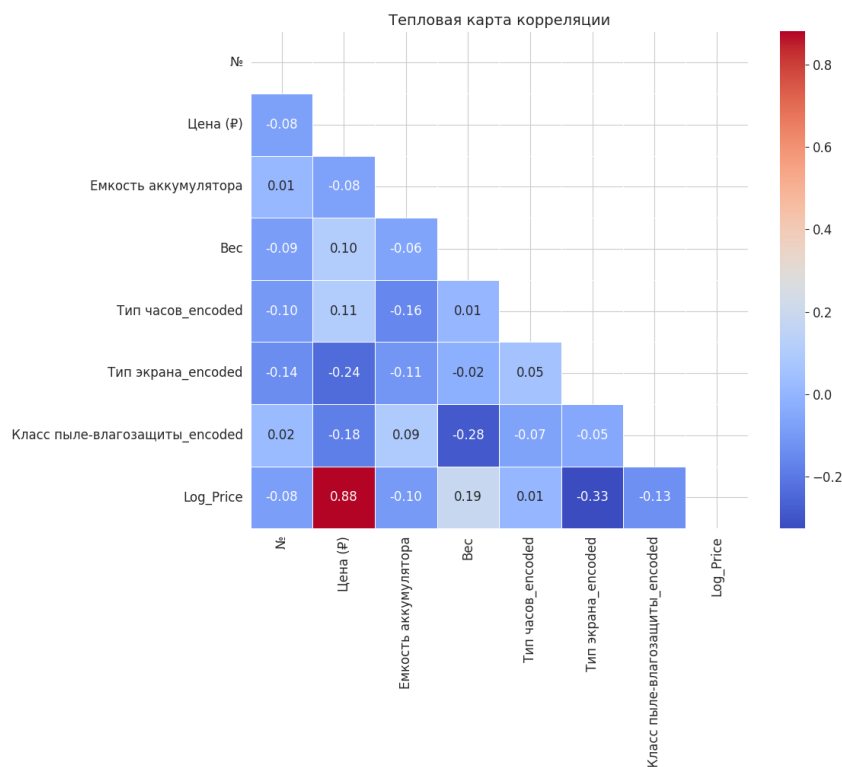


Рисунок 6 – Тепловая карта корреляции признаков

Для более детального изучения построены диаграммы рассеивания (scatter plots), визуализирующие зависимость цены от каждого числового признака. Точки на всех графиках образуют широкие облака, а не выстраиваются вдоль прямой линии, что подтверждает отсутствие линейных зависимостей (Рис. 7).

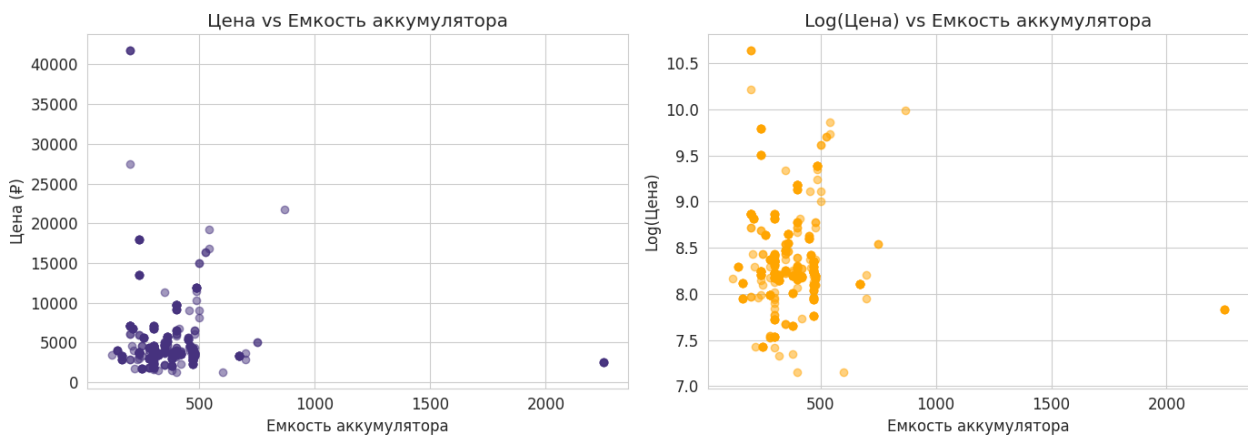


Рисунок 7.1– Диаграммы рассеивания цены и числовых признаков

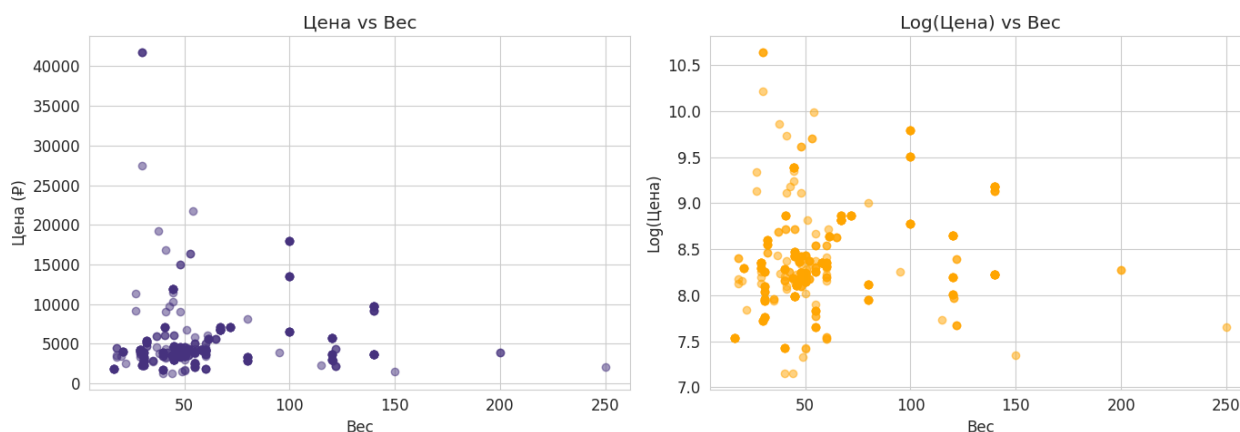


Рисунок 7.2 – Диаграммы рассеяния цены и числовых признаков

Для исследования влияния категориальных признаков на цену построены боксплоты, показывающие распределение цен внутри каждой категории. Анализ показал, что тип экрана демонстрирует четкую дифференциацию: модели с AMOLED-дисплеями имеют значительно более высокую медианную цену. Класс пыле-влагозащиты также выступает сильным ценовым индикатором. Тип часов показывает наименее выраженную дифференциацию (Рис. 8).

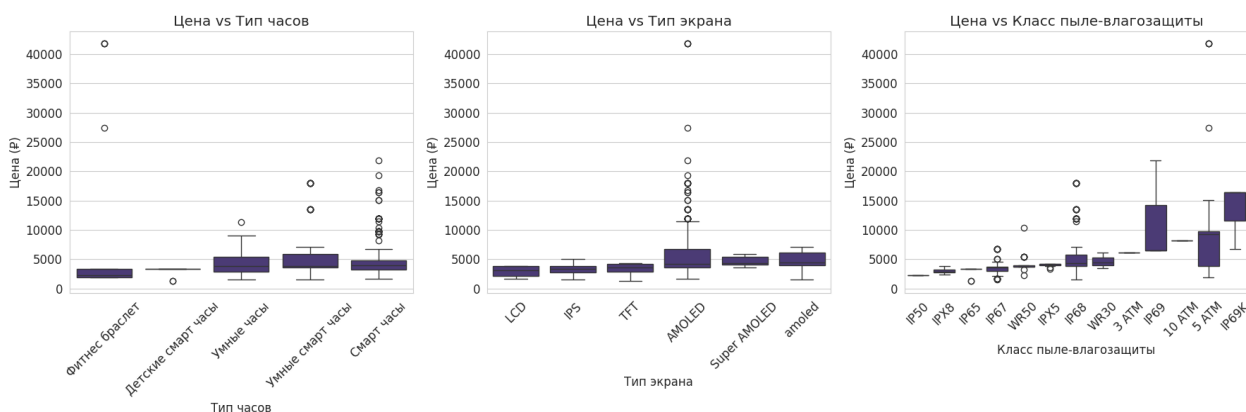


Рисунок 8 – Боксплоты для категориальных признаков

3 Экспериментальное исследование моделей

3.1 Подготовка признаков и трансформация данных

Для корректной работы алгоритмов машинного обучения качественные признаки преобразованы в числовой формат. С помощью метода Label Encoding трем категориальным столбцам присвоены уникальные целочисленные метки. Данные разделены на обучающую и тестовую выборки в соотношении 80/20 (Рис. 9).

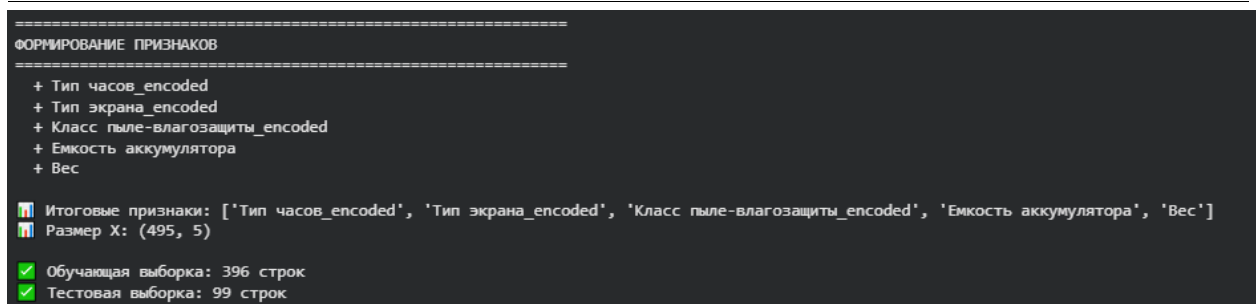


Рисунок 9 – Разделение данных на обучающую и тестовую выборки

Для устранения различий в масштабах числовых признаков (емкость аккумулятора измеряется в мАч, а вес — в граммах) проведена стандартизация данных. Эта процедура особенно критична для линейных моделей и метода К-ближайших соседей, поскольку данные алгоритмы чувствительны к масштабу входных переменных (Рис. 10).

```

if scale:
    scaler = StandardScaler()
    X_train = scaler.fit_transform(X_train)
    X_test = scaler.transform(X_test)

model.fit(X_train, y_train)
y_pred = model.predict(X_test)

return {
    'Model': model_name,
    'R2': r2_score(y_test, y_pred),
    'MSE': mean_squared_error(y_test, y_pred),
    'RMSE': np.sqrt(mean_squared_error(y_test, y_pred)),
    'MAE': mean_absolute_error(y_test, y_pred)
}

```

Рисунок 10 – Масштабирование числовых признаков

3.2 Реализация алгоритмов машинного обучения

На этапе практической реализации проведено 25 серий экспериментов, охватывающих более 15 различных архитектур регрессионных моделей. Основная задача заключалась не только в поиске алгоритма с максимальной предсказательной способностью, но и в исследовании влияния трансформации целевой переменной на итоговую точность. Для этого обучение проводилось в двух параллельных режимах: прогнозирование исходной цены и прогнозирование логарифмированной цены. Использование натурального логарифма позволило сгладить значительный разброс цен между бюджетными и флагманскими устройствами, выявленный на этапе разведочного анализа.

В первую группу исследуемых методов вошли классические линейные регрессоры, такие как стандартная линейная регрессия, Ridge, Lasso и ElasticNet. Эти алгоритмы показали низкий, но положительный коэффициент детерминации ($R^2 \approx 0.13$), что свидетельствует о том, что простые линейные зависимости способны уловить лишь малую часть вариации цен (Рис. 11).

```
print("\n 📊 Линейные модели:")

# Linear Regression
results.append(evaluate_model(LinearRegression(), X_train, X_test, y_train, y_test, 'Linear Regression'))
results.append(evaluate_model(LinearRegression(), X_train, X_test, y_train, y_test, 'Linear Regression (scaled)', scale=True))

# Ridge (ребенная)
results.append(evaluate_model(Ridge(alpha=1.0, random_state=42), X_train, X_test, y_train, y_test, 'Ridge', scale=True))

# Lasso (L1-регуляризация)
results.append(evaluate_model(Lasso(alpha=0.001, random_state=42), X_train, X_test, y_train, y_test, 'Lasso', scale=True))

# ElasticNet
results.append(evaluate_model(ElasticNet(alpha=0.001, l1_ratio=0.5, random_state=42), X_train, X_test, y_train, y_test, 'ElasticNet', scale=True))
print(" ✓ 5 линейных моделей")
```

Рисунок 11 – Определение и обучение линейных регрессионных моделей

Параллельно тестировались деревья решений с глубиной 5, 10 и без ограничений, а также три конфигурации метода К-ближайших соседей. Их результаты (R^2 от 0.70 до 0.85) показали, что они способны улавливать нелинейности, но уступают ансамблевым подходам (Рис. 12).

```
# ===== 2. ДЕРЕВЬЯ РЕШЕНИЙ =====
print("\n 📊 Деревья решений:")

results.append(evaluate_model(DecisionTreeRegressor(max_depth=5, random_state=42), X_train, X_test, y_train, y_test, 'Decision Tree (depth=5)'))
results.append(evaluate_model(DecisionTreeRegressor(max_depth=10, random_state=42), X_train, X_test, y_train, y_test, 'Decision Tree (depth=10)'))
results.append(evaluate_model(DecisionTreeRegressor(max_depth=None, random_state=42), X_train, X_test, y_train, y_test, 'Decision Tree (full)'))
print(" ✓ 3 модели дерева решений")

# ===== 3. МЕТРИЧЕСКИЕ МЕТОДЫ =====
print("\n 📊 KNN:")

results.append(evaluate_model(KNeighborsRegressor(n_neighbors=3), X_train, X_test, y_train, y_test, 'KNN (k=3)', scale=True))
results.append(evaluate_model(KNeighborsRegressor(n_neighbors=5), X_train, X_test, y_train, y_test, 'KNN (k=5)', scale=True))
results.append(evaluate_model(KNeighborsRegressor(n_neighbors=10), X_train, X_test, y_train, y_test, 'KNN (k=10)', scale=True))
print(" ✓ 3 модели KNN")
```

Рисунок 12 – Обучение деревьев решений и KNN-моделей

Наилучшие показатели точности во всем исследовании продемонстрировали ансамблевые методы, основанные на деревьях решений. Группа алгоритмов, включающая Random Forest, Bagging, Extra Trees, AdaBoost и XGBoost, стабильно показывала точность выше 0.80. Абсолютными лидерами стали модели XGBoost, достигшие значений коэффициента детерминации 0.91. Такая высокая точность обусловлена способностью данных моделей улавливать сложные нелинейные зависимости и взаимодействия между признаками, что характерно для рынка потребительской электроники (Рис. 13).

```

# ===== 4. АНСАМБЛЕВЫЕ МЕТОДЫ =====
print("\n📊 Ансамблевые методы (Random Forest):")

results.append(evaluate_model(RandomForestRegressor(n_estimators=50, random_state=42, n_jobs=-1), X_train, X_test, y_train, y_test, 'Random Forest (50)'))
results.append(evaluate_model(RandomForestRegressor(n_estimators=100, random_state=42, n_jobs=-1), X_train, X_test, y_train, y_test, 'Random Forest (100)'))
results.append(evaluate_model(RandomForestRegressor(n_estimators=200, random_state=42, n_jobs=-1), X_train, X_test, y_train, y_test, 'Random Forest (200)'))
print("\n✓ 3 модели Random Forest")

print("\n📊 Бустинг и бэггинг:")

# XGBoost
results.append(evaluate_model(XGBRegressor(n_estimators=50, random_state=42, verbosity=0), X_train, X_test, y_train, y_test, 'XGBoost (50)'))
results.append(evaluate_model(XGBRegressor(n_estimators=100, random_state=42, verbosity=0), X_train, X_test, y_train, y_test, 'XGBoost (100)'))
results.append(evaluate_model(XGBRegressor(n_estimators=200, random_state=42, verbosity=0), X_train, X_test, y_train, y_test, 'XGBoost (200)'))

# Extra Trees
results.append(evaluate_model(ExtraTreesRegressor(n_estimators=100, random_state=42, n_jobs=-1), X_train, X_test, y_train, y_test, 'Extra Trees'))

# AdaBoost
results.append(evaluate_model(AdaBoostRegressor(n_estimators=100, random_state=42), X_train, X_test, y_train, y_test, 'AdaBoost'))

# Bagging
results.append(evaluate_model(BaggingRegressor(n_estimators=100, random_state=42, n_jobs=-1), X_train, X_test, y_train, y_test, 'Bagging'))
print("\n✓ 7 моделей на основе бустинга/бэггинга")

# Создаем DataFrame
results_df = pd.DataFrame(results)

print(f"\n📊 Всего обучено моделей: {len(results_df)}")

```

Рисунок 13 – Обучение ансамблевых моделей

3.3 Оценка качества и анализ прогнозных способностей моделей

Для объективной проверки точности обученных алгоритмов использовалась отложенная тестовая выборка, которая не участвовала в процессе обучения (Табл. 3). Основным критерием качества выступал коэффициент детерминации (R^2), однако для понимания практической применимости моделей также анализировались показатели средней абсолютной ошибки (MAE) и среднеквадратичной ошибки (RMSE).

Таблица 3 – Сводная таблица результатов моделей

№	Модель	R^2	RMSE	MAE
1	XGBoost (50)	0.9058	1 092.10	509.17
2	XGBoost (100)	0.9049	1 097.01	504.82
3	XGBoost (200)	0.9048	1 097.76	505.06
4	Random Forest (100)	0.8862	1 200.25	570.33
5	Bagging	0.8861	1 200.70	570.96
6	Random Forest (200)	0.8848	1 207.93	571.02
7	Random Forest (50)	0.8830	1 217.24	578.47
8	Decision Tree (depth=10)	0.8507	1 374.73	600.60
9	Decision Tree (full)	0.8357	1 442.35	629.61
10	KNN (k=3)	0.8116	1 544.34	716.23
11	Extra Trees	0.8112	1 546.22	670.43
12	AdaBoost	0.8016	1 584.90	1 213.97
13	KNN (k=5)	0.7953	1 610.03	756.30
14	Decision Tree (depth=5)	0.7842	1 653.08	910.50
15	KNN (k=10)	0.6971	1 958.28	1 154.30
16	Linear Regression	0.1289	3 321.02	2 006.30
17	Linear Regression (scaled)	0.1289	3 321.02	2 006.30
18	Lasso	0.1289	3 321.02	2 006.30
19	ElasticNet	0.1289	3 321.03	2 006.28

20	Ridge	0.1288	3 321.08	2 006.22
----	-------	--------	----------	----------

Анализ показал, что при переходе от простейших линейных методов к градиентному бустингу величина средней ошибки сократилась более чем в 6 раз, что подтверждает высокую эффективность ансамблевых подходов при работе со сложными структурами данных. Семь лучших моделей — это ансамблевые методы. Лидером является XGBoost с 50 оценщиками, ошибающийся в среднем менее чем на 510 рублей.

Отдельный интерес представляют результаты пяти моделей, обученных на логарифмированной целевой переменной. Наилучшая модель для логарифмированной цены (Random Forest, $R^2 = 0.8079$) уступает лучшей модели для абсолютной цены (XGBoost, $R^2 = 0.9058$), что говорит о предпочтительности прогнозирования в исходном масштабе цен для данной задачи.

Таблица 4 – Результаты моделей для логарифмированной цены

№	Модель	R^2	MAE	RMSE
1	Random Forest (log)	0.8079	0.1122	0.2363
2	XGBoost (log)	0.8025	0.1128	0.2396
3	KNN (log)	0.7658	0.1380	0.2610
4	Decision Tree (log)	0.6663	0.1432	0.3115
5	Linear Regression (log)	0.1911	0.3582	0.4850

3.4 Интерпретация важности признаков в формировании цены

Заключительным этапом исследования стал анализ значимости переменных (Feature Importance), проведенный на основе весов лучшей обученной модели. Данный анализ позволяет понять, какие именно технические спецификации вносят наибольший вклад в рыночную стоимость смарт-часов. Процедура оценки важности показала, что структура ценообразования на рынке носимой электроники носит иерархический характер (Табл. 5).

Таблица 5 – Анализ важности признаков для модели XGBoost

Ранг	Признак	Важность, %	Интерпретация влияния
1	Емкость аккумулятора	35.36	Определяющий фактор. Время работы — ключевая ценность
2	Тип экрана	25.83	Технология дисплея как маркер премиальности
3	Класс пыле-влагозащиты	18.53	Индикатор защищенности и позиционирования
4	Вес	18.46	Косвенный признак материалов корпуса

5	Тип часов	1.84	Наименее значимый, дублируемый другими
---	-----------	------	--

Наглядная визуализация сравнительной значимости признаков(Рис.14).

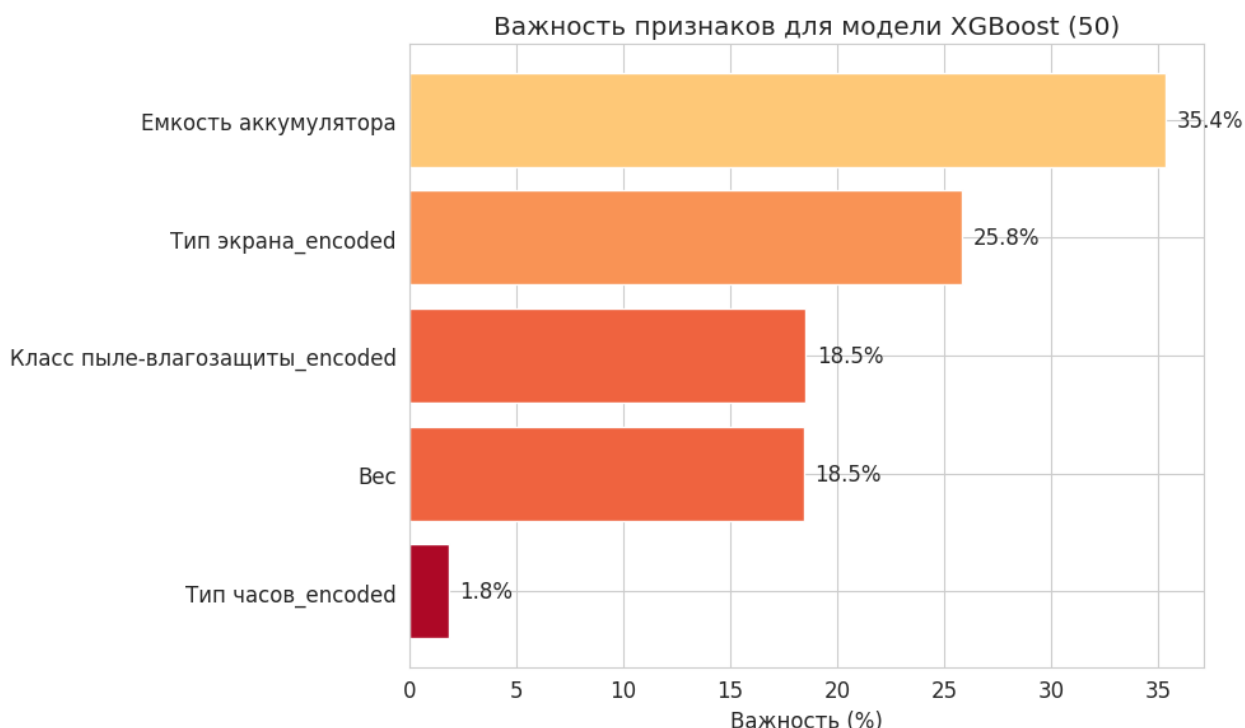


Рисунок 14 – График важности признаков для модели XGBoost

Первостепенное значение имеет емкость аккумулятора (35.36%), что логично коррелирует с разделением устройств на бюджетный, средний и премиальный классы. Тип экрана (25.83%) и класс пыли-влагозащиты (18.53%) суммарно формируют более 44% важности, выступая ключевыми дифференциаторами ценовых сегментов. Примечательно, что такой параметр, как тип часов, оказался наименее значимым предиктором (1.84%), что подтверждает гипотезу о вторичности этого фактора при ценообразовании. Полученные результаты интерпретации не только подтверждают адекватность работы построенных моделей, но и могут быть использованы ритейлерами для оптимизации ценовой политики на основе наиболее востребованных характеристик оборудования.

4 Выводы

В ходе проведенного исследования реализован полный цикл интеллектуального анализа данных, направленный на решение задачи прогнозирования рыночной стоимости смарт-часов на основе их технических характеристик. На этапе подготовки данных применение метода межквартильного размаха и обрезки по 99-му перцентилю позволило эффективно очистить выборку от аномалий и ошибок парсинга, обеспечив стабильность обучения для массового сегмента устройств. Проведенный

разведочный анализ подтвердил наличие сложных нелинейных зависимостей между аппаратным обеспечением и ценой, что предопределило выбор в пользу современных ансамблевых методов машинного обучения.

Сравнительный анализ 25 различных моделей показал значительное превосходство композиционных алгоритмов над классическими линейными подходами. Согласно полученным метрикам, наилучшую точность продемонстрировал алгоритм XGBoost, достигнув коэффициента детерминации $R^2 = 0.91$ и минимальной средней абсолютной ошибки MAE = 509.17 Р. Данные результаты свидетельствуют о том, что построенная модель способна объяснять более 90% дисперсии цен, что является высоким показателем для динамичного рынка электроники.

Анализ важности признаков позволил установить, что ключевыми факторами ценообразования являются емкость аккумулятора (35.36%), тип экрана (25.83%) и класс пыле-влагозащиты (18.53%). Математически подтверждено, что именно эти параметры выступают основными дифференциаторами ценовых сегментов. Разработанный инструментарий может быть практически применен для автоматизации оценки стоимости устройств, оптимизации закупочных цен в ритейле или создания рекомендательных систем для покупателей. Перспективой дальнейшего исследования является интеграция в модель данных о престижности бренда и анализ динамики цен во времени.

Библиографический список

1. Редченко Ю.Н., Богачев Т.В. Анализ и прогнозирование цен объектов недвижимости в регионах России методами интеллектуального анализа данных // Технологическое предпринимательство: проблемы проектирования, применения и безопасности информационных систем в условиях цифровой экономики и реализации инновационного и технологического предпринимательства: Материалы международной студенческой научно-практической конференции. Ростов-на-Дону: Ростовский государственный экономический университет (РИНХ), 2023. С. 166–171. URL: <https://elibrary.ru/item.asp?id=68607778>
2. Амирханова Г.А., Мансурова М.Е., Кобегенова Д.К., Адильбекова А.М. Применение методов интеллектуального анализа данных для оптимизации и визуализации расчета и интерпретации индекса потребительских цен // Статистическое образование в России: интеллектуальный анализ данных: материалы международной конференции (конгресса). Оренбург: Оренбургский государственный университет, 2023. С. 349–359. URL: <https://elibrary.ru/item.asp?id=64569737>
3. Скутина Ю.В. Методы и инструменты для анализа цен на рынке недвижимости // Молодежь и XXI век - 2016: Материалы VI Международной молодежной научной конференции: в 4-х томах / Ответственный редактор Горохов А.А. Т. 3. Курск: Закрытое акционерное общество «Университетская книга», 2016. С. 110–113. URL:

- <https://elibrary.ru/item.asp?id=25605199>
4. Mikheyenkova M.A. On the approach to intelligent data analysis in the social sciences // Открытые семантические технологии проектирования интеллектуальных систем. 2020. № 4. С. 129–132. URL: <https://elibrary.ru/item.asp?id=43856159>
 5. Шайтура С.В. Интеллектуальный анализ данных // Славянский форум. 2015. № 2 (8). С. 341–350. URL: <https://elibrary.ru/item.asp?id=23417284>
 6. Назаров Д.М. Цифровые технологии и интеллектуальные модели: монография. Казань: Бук, 2022. 192 с. URL: <https://elibrary.ru/item.asp?id=49948365>
 7. Чернышева Ю.Г., Хирная А.В., Головкин А.А. Интеллектуальный анализ данных (Data Mining) и его применение в бизнес-анализе // Бухгалтерский учет, анализ, аудит и статистика: информационные инструменты достижения целей устойчивого развития экономики: Материалы Международной научно-практической конференции, посвященной 70-летию Учетно-экономического факультета. Ростов-на-Дону: АзовПринт, 2023. С. 212–217. URL: <https://elibrary.ru/item.asp?id=59990888>
 8. Google Colab Notebook (рабочая среда исследования). URL: https://colab.research.google.com/drive/1gtz4pdbRN0_FJyuz0XofgfgA_AKd6IDA?usp=sharing (дата обращения 22.04.2026).