ПРИМЕНЕНИЕ МЕТОДОВ МАШИННОГО ОБУЧЕНИЯ В ОБЛАСТИ ПРОГНОЗИРОВАНИЯ ОБЪЕМА ПРОДАЖ С УЧЕТОМ ДИНАМИЧЕСКИ ИЗМЕНЯЮЩИХСЯ ПРИЗНАКОВ

APPLICATION OF METHODS OF MACHINE TRAINING IN THE FIELD OF FORECASTING THE VOLUME OF SALES WITH ACCOUNT DYNAMICALLY CHANGING SIGNS

УДК 004.85

Валиахметова Юлия Ильясовна, кандидат технических наук, доцент кафедры «Вычислительной математики и кибернетики», Уфимский государственный авиационный технический университет**,** Россия, г. Уфа

Идрисова Эльмира Ильгизовна, студент, 2 курс, факультет «Информатики и робототехники», Уфимский государственный авиационный технический университет**,** Россия, г. Уфа

Valiahmetova YUliya Il'yasovna, e-mail: julikas@inbox.ru

Idrisova El'mira Il'gizovna, e-mail: eelemira@yandex.ru

Аннотация

Статья посвящена исследованию существующих математических решений и выбору наиболее подходящего метода для прогнозирования объема продаж с учетом признаков. Ниже будут представлены результаты проведенного анализа данных о продажах товаров компании «Walmart» за более чем пятилетний период и динамических признаков в виде цен на товары и календарных событий, результаты сравнительного анализа среди методов в области прогнозирования, который позволил выбрать метод и разработать алгоритмы для решения задачи прогнозирования с учётом признаков.

Annotation

The article is devoted to the study of existing mathematical solutions and the selection of the most suitable method for predicting sales volume taking into account signs. Below will be presented the results of the analysis of data on sales of Walmart goods for more than five years and dynamic signs in the form of prices for goods and calendar events, the results of a comparative analysis among methods in the field of forecasting, which made it possible to choose a method and develop algorithms for solving the problem forecasting taking into account signs

Ключевые слова: прогнозирование, предсказание, прогнозирование объема продаж с учетом динамически изменяющихся признаков, методы прогнозирования, обучение модели, обучающая выборка, валидационная выборка, анализ данных, визуализация данных, машинное обучение, кроссвалидация.

Keywords: forecasting, prediction, forecasting sales volume taking into account dynamically changing attributes, forecasting methods, model training, training set, validation sample, data analysis, data visualization, machine learning, cross-validation.

Актуальность

Построение качественного прогноза, которое учитывает не только данные о продажах, но и влияние признаков, таких как цены и различные события, влияет на планирование всего бизнеса: анализ продаж продукции и формирование оптимальных запасов, формирование бизнес-плана, планирование бюджета, оптимизацию и контроль расходов, распределение ресурсов компании для обеспечения ожидаемых продаж.

Практическая значимость заключается в том, что метод, предложенный в результате исследования, выявляет и учитывает закономерности между объемами продаж и признаками при построении прогноза. Повышается качество прогноза, которое положительно влияет на планирование бизнеса и принятие решений в неопределенных условиях.

Данные

В качестве реальных данных был взят датасет с международного конкурса по прогнозированию "M5 Forecasting – Accuracy" проходящий на платформе Kaggle [1], который представляет собой данные о продажах за 5,5 лет компании Walmart - американская компания, управляющая крупнейшей в мире сетью оптовой и розничной торговли.

Набор данных включает в себя продажи товаров с 2011 по 2016 год, которые классифицированы по 3 категориям (хобби, продукты питания и товары для дома) и 7 отделам. Товары продаются в десяти магазинах, расположенных в трех штатах США (Калифорния, Техас, Висконсин). Помимо данных временного ряда, он также включает в себя дополнительную информацию о ценах, датах, рекламных акциях и событиях (например, Суперкубок, День святого Валентина, Пасха и т.д.), которые влияют на продажи и используются для повышения точности прогнозирования.

Просмотр данных: структура таблиц и содержание

Набор данных Walmart состоит из 3 отдельных таблиц формата «csv».

Таблица продаж содержит данные о ежедневных продажах по каждому товару за весь период. Состоит из идентификаторов товара, отделов, категорий, магазинов и штатов. Каждый столбец имеет идентификатор в виде порядкового дня продажи, который хранит в себе количество проданных единиц соответствующего товара в день. Количество строк соответствует количеству товаров равное 30490, а количество столбцов советует количеству дней равное 1913. Структура таблицы продаж представлена на рисунке 1.

Таблица с ценами содержит информацию о средней цене проданных товаров в магазине за неделю. Состоит из идентификатора товара, магазина, которые связываю эти данные с другими таблицами. Количество строк соответствует количеству недель за весь период по каждому товару равное 684121. Структура таблицы отображена на рисунке 1.

Таблица календарных событий содержит информацию о датах, событиях и периодах покупок с купонами SNAP. Также содержит

идентификатор недели, который связывает с таблицей цен и порядкового дня продажи, который связывает с таблицей продаж. Количество строк соответствует количеству дней за весь период обучения, а также предоставлена информация о событиях и акциях на следующие 2 месяца. Все события делятся на 4 основные категории: спортивные, культурные, национальные и религиозные. SNAP – расшифровывается как «Программа помощи в области дополнительного питания» и предоставляет льготы лицам и семьям с низким доходом, имеющим на это право, с помощью карты электронного перевода пособий. Структура таблицы представлена на рисунке 2.

	id	item_id	dept_id	cat_id	store_id	state_id	d_1	d_2	d_3	d_4	...	d_1904	d_1905	d_1906	d_1907
0	HOBBIES_1_001_CA_1_validation	HOBBIES_1_001	HOBBIES_1	HOBBIES	CA_1	CA	0	0	0	0	...	1	3	0	1
1	HOBBIES_1_002_CA_1_validation	HOBBIES_1_002	HOBBIES_1	HOBBIES	CA_1	CA	0	0	0	0	...	0	0	0	0
2	HOBBIES_1_003_CA_1_validation	HOBBIES_1_003	HOBBIES_1	HOBBIES	CA_1	CA	0	0	0	0	...	2	1	2	1
3	HOBBIES_1_004_CA_1_validation	HOBBIES_1_004	HOBBIES_1	HOBBIES	CA_1	CA	0	0	0	0	...	1	0	5	4
4	HOBBIES_1_005_CA_1_validation	HOBBIES_1_005	HOBBIES_1	HOBBIES	CA_1	CA	0	0	0	0	...	2	1	1	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...	...	...
30485	FOODS_3_823_WI_3_validation	FOODS_3_823	FOODS_3	FOODS	WI_3	WI	0	0	2	2	...	2	0	0	0
30486	FOODS_3_824_WI_3_validation	FOODS_3_824	FOODS_3	FOODS	WI_3	WI	0	0	0	0	...	0	0	0	0
30487	FOODS_3_825_WI_3_validation	FOODS_3_825	FOODS_3	FOODS	WI_3	WI	0	6	0	2	...	2	1	0	2
30488	FOODS_3_826_WI_3_validation	FOODS_3_826	FOODS_3	FOODS	WI_3	WI	0	0	0	0	...	0	0	1	0
30489	FOODS_3_827_WI_3_validation	FOODS_3_827	FOODS_3	FOODS	WI_3	WI	0	0	0	0	...	0	0	0	0

30490 rows × 1919 columns

Рисунок 1 – Структура таблицы продаж

	date	wm_yr_wk	weekday	wday	month	year	d	event_name_1	event_type_1	event_name_2	event_type_2	snap_CA	snap_TX
0	2011-01-29	11101	Saturday	1	1	2011	d_1	NaN	NaN	NaN	NaN	0	0
1	2011-01-30	11101	Sunday	2	1	2011	d_2	NaN	NaN	NaN	NaN	0	0
2	2011-01-31	11101	Monday	3	1	2011	d_3	NaN	NaN	NaN	NaN	0	0
3	2011-02-01	11101	Tuesday	4	2	2011	d_4	NaN	NaN	NaN	NaN	1	1
4	2011-02-02	11101	Wednesday	5	2	2011	d_5	NaN	NaN	NaN	NaN	1	0
...	...	...	...	...	...	...	...	...	...	...	...	...	...
1964	2016-06-15	11620	Wednesday	5	6	2016	d_1965	NaN	NaN	NaN	NaN	0	1
1965	2016-06-16	11620	Thursday	6	6	2016	d_1966	NaN	NaN	NaN	NaN	0	0
1966	2016-06-17	11620	Friday	7	6	2016	d_1967	NaN	NaN	NaN	NaN	0	0
1967	2016-06-18	11621	Saturday	1	6	2016	d_1968	NaN	NaN	NaN	NaN	0	0
1968	2016-06-19	11621	Sunday	2	6	2016	d_1969	NBAFinalsEnd	Sporting	Father's day	Cultural	0	0

1969 rows × 14 columns

	store_id	item_id	wm_yr_wk	sell_price
0	CA_1	HOBBIES_1_001	11325	9.58
1	CA_1	HOBBIES_1_001	11326	9.58
2	CA_1	HOBBIES_1_001	11327	8.26
3	CA_1	HOBBIES_1_001	11328	8.26
4	CA_1	HOBBIES_1_001	11329	8.26
...	...	...	...	...
6841116	WI_3	FOODS_3_827	11617	1.00
6841117	WI_3	FOODS_3_827	11618	1.00
6841118	WI_3	FOODS_3_827	11619	1.00
6841119	WI_3	FOODS_3_827	11620	1.00
6841120	WI_3	FOODS_3_827	11621	1.00

6841121 rows × 4 columns

Рисунок 2 – Структура таблицы с календарными событиями и ценами

Визуальный анализ данных

Анализ отдельных временных рядов показал, что большинство товаров имеют довольно низкую статистику ежедневных подсчетов и прогнозы будут иметь дело с довольно большим количеством нулевых значений. В товарах не прослеживается явная сезонность и периодичность,

отсутствуют тренды. На рисунке 3 представлены графики случайно выбранных товаров.

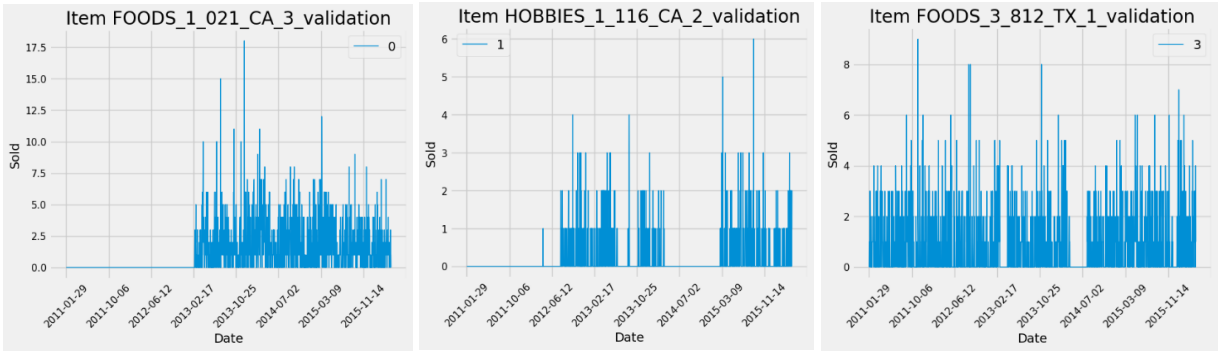

Рисунок 3 – Графики случайно выбранных товаров

При анализе агрегированных временных рядов по штатам (см. рисунок 4) и категориям (см. рисунок 5) была отчетливо выявлена годовая сезонность и ежемесячная периодичность. Были замечены падения на каждое Рождество, которое является единственным днем в году, когда магазины закрыты и провалы в 2014 году у всех магазинов, которые, случаются не всегда, но они в первую очередь отражают годовую сезонность, которая уже была замечена. В целом продажи растут, что является хорошим показателем для Walmart. Больше всего товаров продает штат Калифорния (CA) а самой распространенной категория является «еда», за которой следует категория «товары для дома» и затем категория «хобби».

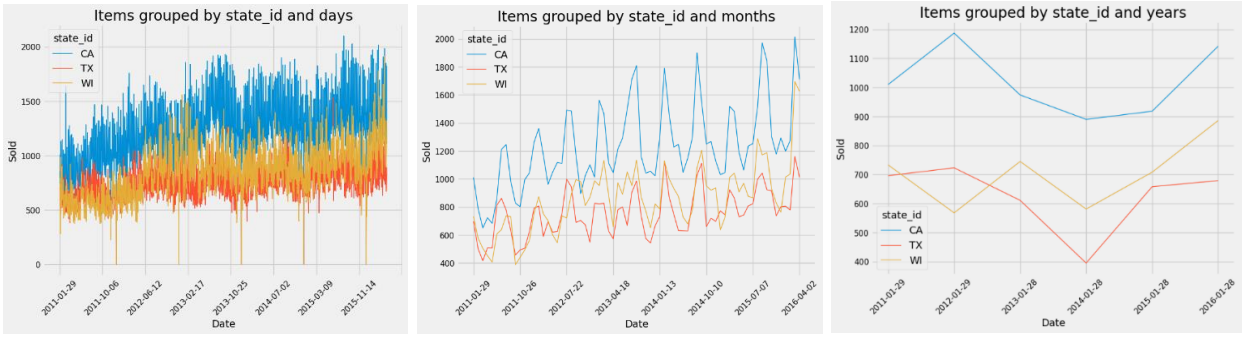

Рисунок 4 – Графики ежедневных, ежемесячных и ежегодных продаж по штатам

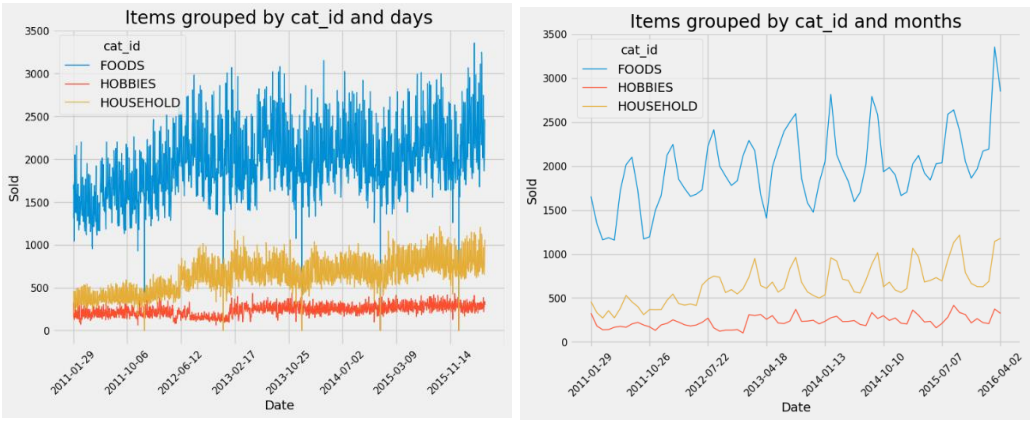

Рисунок 5 – Графики ежедневных, ежемесячных продаж по категориям

Анализ признаков: календарные события и цены

Анализ календарных событий (см. рисунок 6) показал, что дни с событиями занимают около 8% всех дней за год, большинство событий относится к религиозным мероприятиям, далее идут национальные праздники, затем культурные и спортивные. Количество дней, когда разрешены покупки с талонами на питание SNAP, одинаковы для всех штатов, на что указывает одинаковый процент одинаковый для всех 3 штатов 33%. Данный вывод подкрепляется построением графика в виде календаря (см. рисунок 6), на котором выделены оранжевым цветом дни покупок с талонами SNAP за первое полугодие, который показывает, что дни SNAP всегда являются одинаковыми и проходят в первой половине месяца не позднее 15 числа. Их всегда 10, и конкретные дни отличаются от штата к штату. Это, безусловно, помогает измерить их влияние и сделать прогнозы более надежными.

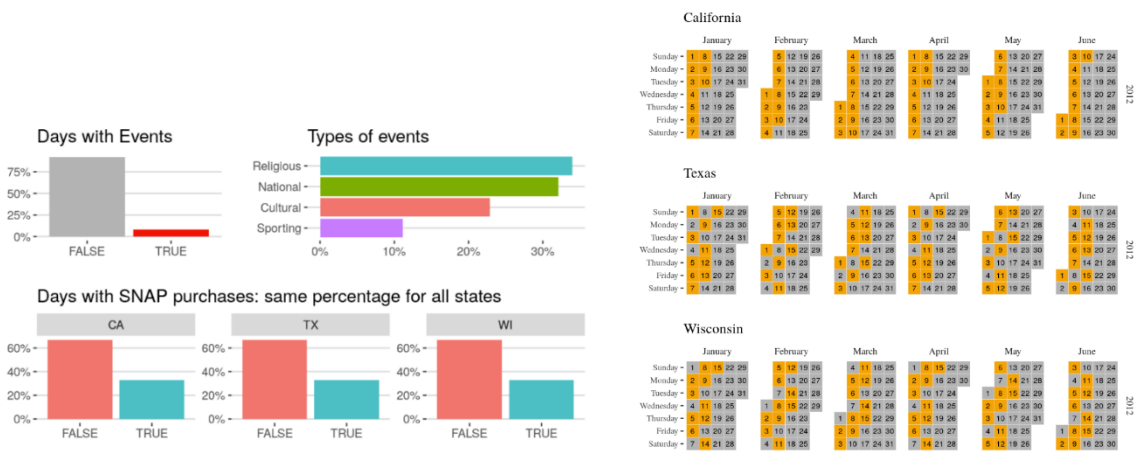

Рисунок 6 – Слева диаграммы: дни с событиями; типы событий; дни с покупками SNAP, справа: календарь покупок с талонами SNAP

Влияние признаков на продажи товаров

Теперь, когда есть представление о динамических признаках, посмотрим, как они влияют на данные временных рядов. Соединим продажи товаров с календарными событиями и покупками с купонами SNAP. Продажи будут выведены в виде линейных графиков, серым фоном будут отображены периоды использования SNAP, индикаторы событий будут отображены в виде черных точек результат построения графиков представлен слева на рисунке 7. Для соединения продажи товаров с изменением цен выведем изменение цен за весь период в виде тепловой карты и наложим на них график продаж для тех же 3х товаров разной категории. Более светлые цвета означают более низкие цены и наоборот.

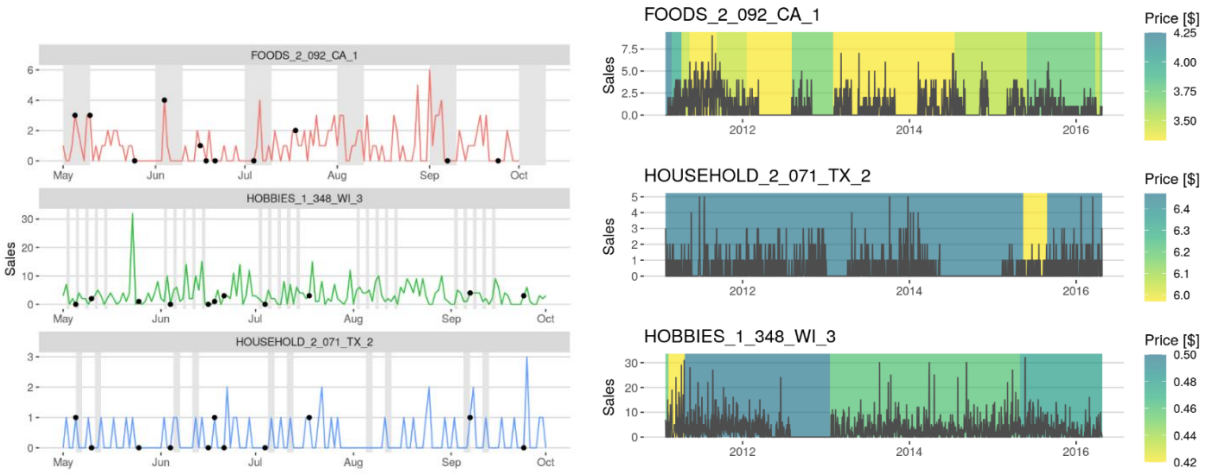

Рисунок 7 – Слева: график наложения продаж и календарных событий, справа: график наложения продаж и изменения цен

Итоги анализа данных

Наибольшее количество продаж происходит в течение периодов SNAP (особенно для категории FOODS). Для предметов HOBBIES и HOUSEHOLD нет никаких прямых признаков того, что дни SNAP обеспечивают определенный рост продаж. События не сильно влияют на рост продаж, но иногда в день события наблюдаются всплески продаж (например, FOOOS в

мае и начале июня).

У некоторых товаров чаще меняется цена, чем у других. Продукт FOODS, взятый случайным образом меняет цену примерно 10 раз, тогда как товар HOUSEHOLD имеет только один короткий период высокой цены, а затем возвращается к своей первоначальной цене. В некоторых случаях изменения цен коррелируют с ростом спроса после заметного разрыва нулевых продаж. Показательный пример: продажи продуктов FOODS в 2013 году и продажи продуктов HOBBIES в то же время. В обоих случаях изменение цены во время разрыва совпадает с увеличением продаж, которое длится в течение определенного периода времени, не только 1 или 2 дня.

Обзор методов и моделей прогнозирования

Для построения прогноза были рассмотрены следующие модели:

- Модели временных рядов ARIMA, SARIMAX;
- Аддитивная регрессионная модель;
- Нейронная сеть LSTM;
- Градиентный бустинг, использующий деревья решений (далее алгоритм Catboost);

Для оценки моделей были взяты следующие метрики:

- коэффициент детерминации R^2;
- средняя абсолютная ошибка прогнозирования MAE;
- корень квадратный из средней квадратичной ошибки прогнозирования RMSE.

На этапе выбора модели прогнозирования следует учитывать, что чем проще представленная модель прогнозирования, тем легче она будет для понимания ее результатов, однако простота модели не всегда может гарантировать высокое качество прогноза.

Анализ данных показал, что отдельные временные ряды для каждого товара не содержат явной сезонности, а тренд отсутствует, также в них имеется большое количество нулевых значений, однако, прогнозирование

объема продаж для конкретного товара актуально и является основным заданием конкурса.

Напротив, агрегирование товаров по штатам, категориям и отделам позволяет получить заполненные временные ряды с годовыми, месячными и недельными сезонностями. В добавок сгруппированные временные ряды отображают общую картину по продажам, что очень важно для планирования, анализа и принятия решений.

Учтивая вышеописанное для определения подходящей модели прогнозирования было решено проводить вычислительный эксперимент, в несколько этапов:

1) Прогнозирование и оценка моделей временных рядов:
- ARIMA,
- SARIMAX,
- Аддитивная регрессионная модель;

2) Прогнозирование и оценка методов машинного обучения:
- Нейронная сеть LSTM;
- Catboost - алгоритм, использующий деревья решений;

При этом каждый этап прогнозирования и оценки модели был рассмотрен на отдельных и агрегированных временных рядах.

Подготовка моделей временных рядов

Проведем анализ с помощью аддитивной регрессионной модели, используя разложение на компоненты временного ряда «Сумма всех продаж в день» с периодом 365 и 7 (см. рисунок 8). В таблице 1 представлены оценки аддитивной регрессионной модели.

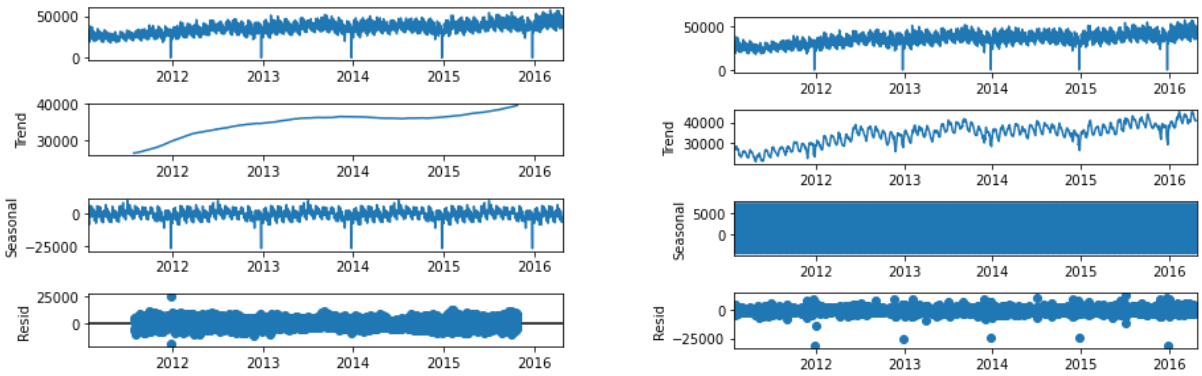

Рисунок 8 – Сумма всех продаж в день с периодом 365 и 7

Таблица 1 – Оценки для временного ряда «Сумма всех продаж в день с периодом 365 и 7»

Временной ряд	**Оценки**			
	R2	**RMSE**	**MSE**	**MAE**
Сумма всех продаж в день с периодом 365	0.532	4567.452	20861619.551	3741.786
Сумма всех продаж в день с периодом 7	0.868	2662.987	7091503.013	1678.756

Исследование на корреляцию

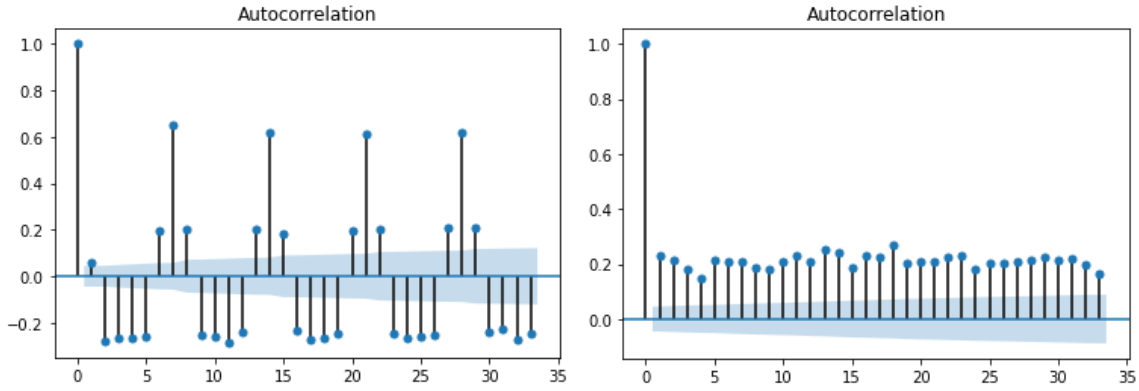

Рисунок 9 – Автокорреляционная функция для суммы продаж и для отдельного товара

Для временного ряда «Сумма всех продаж в день» была построена АКФ (см. рисунок 9), которая показала, что продажи за текущий день очень хорошо коррелированы с продажами 7 дней назад - недельный цикл налицо. Аналогичным образом была построена АКФ для товара (см. рисунок 9 справа)

и оказалось, что сезонной корреляции в таких рядах не обнаружено. Таким образом, можно сделать вывод о том, что недельная периодичность выражена намного ярче годовой. Будем в дальнейшем строить модель SARIMAX с недельной периодичностью.

Подготовка методов машинного обучения

В ходе изучения было выявлено, что дополнительные признаки влияют на продажи, поэтому было решено создать новую таблицу, которая бы объединяла все данные. К полученной таблице со всеми данными были добавлены дополнительные категориальные признаки (см. таблицу 2).

Таблица 2 - Дополнительные признаки

Признак	**Описание**
iteam_sold_avg	Среднее кол-во продаж каждого товара за все время
state_sold_avg	Среднее кол-во продаж каждого штата за все время
store_sold_avg	Среднее кол-во продаж каждого магазина за все время
cat_sold_avg	Среднее кол-во продаж каждой категории за все время
dept_sold_avg	Среднее кол-во продаж каждого отдела за все время
cat_dept_sold_avg	Среднее кол-во продаж по категории и отделу
store_item_sold_avg	Среднее кол-во продаж по магазину и товару
cat_item_sold_avg	Среднее кол-во продаж по категории и товару
dept_item_sold_avg	Среднее кол-во продаж по отделу и отделу
state_store_sold_avg	Среднее кол-во продаж по штату и магазину
state_store_cat_sold_avg	Среднее кол-во продаж по штату, магазину и категории
store_cat_dept_sold_avg	Среднее кол-во продаж по штату, категории и отделу

Таким образом была создана таблица, содержащая все исходные данные и дополнительные признаки (см. рисунок 10).

id	item_id	dept_id	cat_id	store_id	state_id	d	sold	wm_yr_wk	weekday	...	cat_dept_sold_avg	store_item_sol
14370	1437	3	1	0	0	36	0	11106	2	...	0.703125	0.313721
14380	1438	3	1	0	0	36	0	11106	2	...	0.703125	0.257812
14390	1439	3	1	0	0	36	0	11106	2	...	0.703125	0.150513
14400	1440	3	1	0	0	36	0	11106	2	...	0.703125	1.718750
14410	1441	3	1	0	0	36	0	11106	2	...	0.703125	0.966309
14420	1442	3	1	0	0	36	0	11106	2	...	0.703125	0.858398
14430	1443	3	1	0	0	36	0	11106	2	...	0.703125	0.221680
14440	1444	3	1	0	0	36	0	11106	2	...	0.703125	7.230469
14450	1445	3	1	0	0	36	7	11106	2	...	0.703125	1.186523
14460	1446	3	1	0	0	36	0	11106	2	...	0.703125	0.719238

Рисунок 10 – Структура итоговой таблицы для вычислительного эксперимента

После проведения дополнительного анализа и подготовки данных были выполнены расчеты, представленные в общей таблице 3.

Таблица 3 - Итоговые результаты исследования

Методы	Вычислительный эксперимент	R2	RMSE	MAE
Модели временных рядов: ARIMA, SARIMAX; Аддитивная регресионная модель.	на агрегированных временных рядах SARIMAX с недельной периодичностью	0.865	2605.7811	1929.533
	на отдельных временных рядах SARIMAX с недельной периодичностью	0.007	1.113	0.860
Методы машинного обучения: Нейронная сеть LSTM; Catboost - алгоритм, использующий деревья решений;	на агрегированных временных рядах LSTM	0.599	3.039	1.492
	на агрегированных временных рядах CATBOOST	0.851	4.835	4.185
	на отдельных временных рядах LSTM	-0.379	1.402	0.733
	на отдельных временных рядах CATBOOST	0.976	0.664	0.220

Вычислительный эксперимент

Алгоритм CatBoost для решения сложных задач по прогнозированию объема продаж с учетом динамически изменяющихся признаков был реализован на языке программирования Python и протестирован на платформе Kaggle, которая предоставляет участникам удаленную среду для тестирования с параметрами:16Гб GPU – видеокарта, 3Гб оперативная память, 5Гб на диске.

Для вычислительного эксперимента были использованы реальные данные о продажах компании Walmart за 5,5 лет с 2011 по 2016 год, которые предоставили организаторы конкурса для проведения исследований. Так же были предоставлены валидационные данные о продажах на следующие 28 дней, для которых строился прогноз по исходным данным.

Для предсказания к исходным данным были добавлены дополнительные категориальные признаки. Алгоритм обучался при следующих параметрах: глубина дерева 8 и для обучения были взяты данные за 1 прошедший год. В таблице 4 представлены метрики для рассчитанного прогноза.

Таблица 4 – Результаты численного эксперимента

R2	MAE	MSE	RMSE
0.723	1.044	4.429	2.104

R2 показывает не плохой результат, равный 0,723 что говорит об адекватности модели. Так же в задании конкурса предусмотрена загрузка решения в виде прогноза, по которой определяется score-метрика (WRMSSE – взвешенная среднеквадратичная ошибка), это значит, что в оценочных данных, с которыми будет сравниваться загруженный прогноз разным товарам советуют разные веса, которые определены компанией и не известны участникам конкурса. В результате загрузки решения рассчитанная метрика WRMSSE равна 0,63, что показывает неплохой результат на взвешенных данных. На рисунке 11 представлены графики с результатами получившего прогноза для общих продаж, на которых видно, что предсказанные значения

почти совпадают с реальными. Обратите внимание, что предсказанное значение на 7 день (см. рисунок 11 слева) сильно отклонилось от реального, однако на продолжение этого графика справа, можно увидеть, как модель улавливает падение, которые следуют за резкими скачками. На рисунке 12 предсказанные значения не так сильно совпадают с реальными, как на графике рисунка 11, однако модель улавливает и сигнализирует о том, в какие дни ожидается повышение продаж.

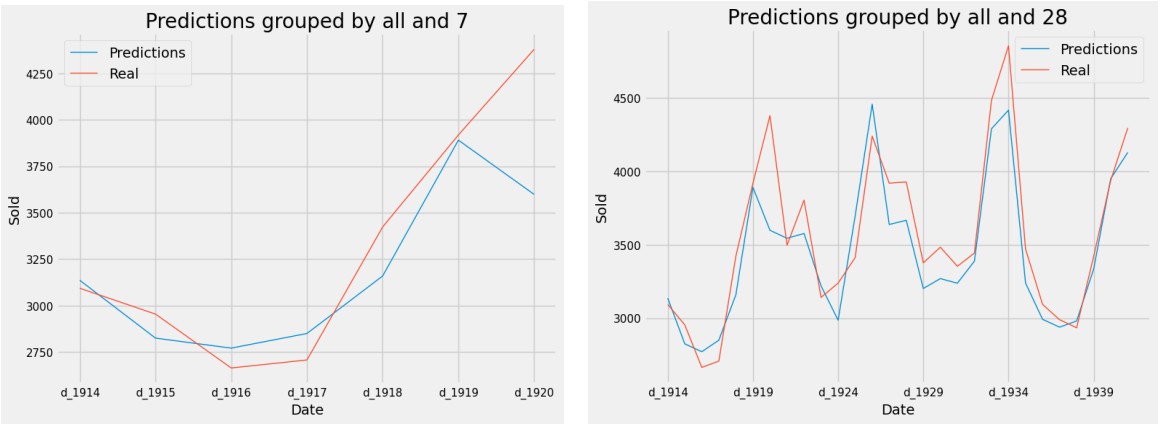

Рисунок 11 – Прогноз для общих продаж на следующие 7 и 28 дней

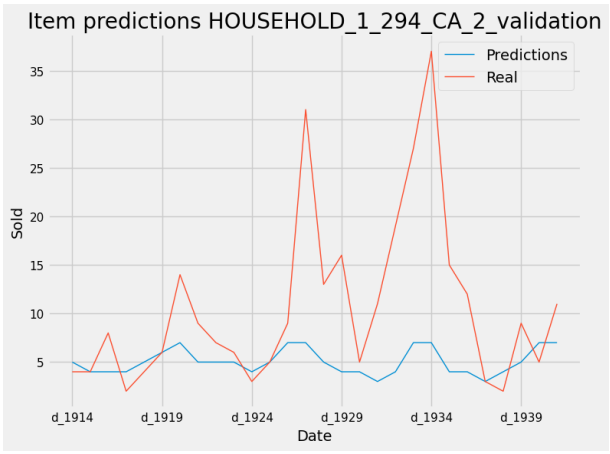

Рисунок 12 – Прогноз для товара «HOUSEHOLD_1_294_CA_2_validation» на следующие 28 дней

Итоги исследования

В результате сравнительного анализа моделей и методов было выявлено что регрессионные модели показывают хорошие результаты на сгруппированных временных рядах. Хорошие оценки показала модель SARIMAX которая в отличии от ARIMA учитывает периодичность в 365 и 7 дней, но не подошла для прогнозирования временных рядов по отдельному

товара. Еще одним недостатком регрессионных моделей является то, что они на вход принимают только одномерные временные ряды. Такой подход использования модели создает проблему для подсчитывания 30 тысяч временных рядов для каждого товара и хранения 30 тысяч обученных моделей.

В отличии от регрессионных моделей методы машинного обучения: нейронная сеть LSTM принимает на вход последовательность векторов-признаков предыдущих дней, а алгоритм CatBoos принимает на вход вектор категориальных признаков. В ходе испытаний этих методов на сгруппированных и отдельных временных рядах алгоритм CatBoost показал наилучшие результаты.

Выбор алгоритма подкреплен вычислительным экспериментом, в ходе которого был построен прогноз на следующие 28 дней, который показал приемлемые результаты.

Таким образом, наилучшим методом для прогнозирования объема продаж с учетом динамических изменяющихся признаков в виде цен на товары и календарных событий оказался метод машинного обучения - градиентный бустинг, использующий деревья решений или алгоритм CatBoost.

Литература

1. Система организации конкурсов по исследованию данных: сайт платформы Kaggle. [Электронный ресурс]. URL: https://www.kaggle.com/ (дата обращения: 10.04.2020).
2. Э.Е. Тихонов. Методы прогнозирования в условиях рынка: учебное пособие. - Невинномысск, 2006. - 221 с.
3. Tiflin C. LSTM Recurrent Neural Networks for Signature Verification. Saarbruecken, LAP Lambert Academic Publishing, 2012, 104 p – ISBN 9783846589946.
4. Алгоритм CatBoost: сайт Википедии. [Электронный ресурс] URL: https://neerc.ifmo.ru/wiki/index.php?title=CatBoost (дата обращения: 03.05.2020).

5. Granger C. W. J. Forecasting in business and economics. – Academic Press, 2014.
6. Васильев А. А. МЕТОДЫ И МОДЕЛИ ПРОГНОЗИРОВАНИЯ ОБЪЕМА ПРОДАЖ В МАРКЕТИНГЕ. – 2013.

Literature

1. The system of organizing competitions for data research: the site of the Kaggle platform. [Electronic resource]. URL: https://www.kaggle.com/ (accessed date: 04/10/2020).
2. E.E. Tikhonov. Forecasting methods in market conditions: a training manual. - Nevinnomyssk, 2006 .-- 221 p.
3. Tiflin C. LSTM Recurrent Neural Networks for Signature Verification. Saarbruecken, LAP Lambert Academic Publishing, 2012, 104 p - ISBN 9783846589946.
4. CatBoost Algorithm: Wikipedia site. [Electronic resource] URL: https://neerc.ifmo.ru/wiki/index.php?title=CatBoost (accessed 03.05.2020).
5. Granger C. W. J. Forecasting in business and economics. - Academic Press, 2014.
6. Vasiliev A. A. METHODS AND MODELS OF FORECASTING THE VOLUME OF SALES IN MARKETING. - 2013.