

# Использование методов машинного обучения для оценки прогнозирования продаж товара

Ю.А. Сердинская, email: serdinskaja2017@yandex.ru <sup>1</sup>

В.В. Мокшин, email: vladimir.mokshin@mail.ru <sup>1</sup>

<sup>1</sup> КАЗАНСКИЙ НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ  
ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ ИМ. А.Н. ТУПОЛЕВА – КАИ

***Аннотация.** В статье содержится описание этапов подготовки и анализа данных о продажах товаров продуктового магазина за один год. Практическая реализация произведена на языке программирования Python в среде разработки Jupyter Notebook.*

***Ключевые слова:** анализ данных, машинное обучение, визуализация данных, прогнозирование, Python..*

## Введение

Сегодня мировая экономика претерпевает колоссальные изменения. Возможность предсказать направления развития рынка и прогнозирования прибыли приобрело первостепенное значение. Для решения этих задач применяется машинное обучение (Machine Learning). Оно быстро развивается и вызывает сильный интерес у многих компаний. Но не всегда компаниям ясно, как можно применить эти технологии в повседневном бизнесе.

Анализ данных и построение корректного прогноза, учитывающего данные о продажах и влияние признаков, таких как цены и различные события, сильно влияют на планирование всего бизнеса. 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.

Также важен баланс между объемом поставок и продажами товаров, поэтому использование современных методов прогнозирования, основанных на технологиях машинного обучения, можно считать перспективным направлением.

## 1. Цель исследования

Цель данной научно-технической работы является исследование возможностей применения методов машинного обучения.

Для достижения цели необходимо выполнить ряд задач:

- выполнить предварительный анализ выборки (поиск пропусков, выбросов, аномалий и т.д.).
- выполнить временное прогнозирование.
- вычислить парные коэффициенты линейной корреляции.
- выполнить прогнозирование прибыли с помощью нейронной сети.
- оценить качество полученных результатов

## 2. Реализация

В качестве реальных данных был взят датасет с ежемесячным отчетом о продажах с указанием групп товаров, названий, количества, цен покупки, продажи и наценки, размещенный на платформе Kaggle [1], который представляет собой данные о продажах за 2018 год.

Выбранный набор данных представляет собой таблицу формата «csv», которая состоит из следующих столбцов: дата; идентификационный номер продукта; название группы продуктов; название продукта; количество товара; чистая цена единицы товара; общая чистая стоимость товаров; цена единицы товара с учетом прибыли; максимальная цена за единицу товара; общая максимальная стоимость товаров; валовая стоимость товара; валовая прибыль в процентах; процент маржи; прибыль за единицу товара; прибыль от продажи всех; доля в марже.

Количество строк соответствует количеству товаров и равно 23181. Структура таблицы продаж представлена на рис. 1.

Date	PKod	Pgroup	Pname	Pquantity	pce_zn	pwa_zn	pce_sn	pwa_sn	pce_sb	pwa_sb	puudzsb	pmarza	pmarzajedn	pkwmarza	
0	01.01.2018	5307	CHEMISTRY	PATYCZKI BELLA 200	1	1,77	1,77	2,57	2,57	2,78	2,78	0	31,13	0,8	0,8
1	01.01.2018	5986	CHEMISTRY	ACE 1 L	1	3,2	3,2	5,68	5,68	6,99	6,99	0,01	43,66	2,48	2,48
2	01.01.2018	399	DAIRY_CHESSSE	ALMETTE CZOSINEK NIEDZWIEDZI 150GR	4	2,68	10,72	3,8	15,2	3,99	15,96	0,02	29,47	1,12	4,48
3	01.01.2018	2246	DAIRY_CHESSSE	ALMETTE JOGURTOWY 150G HOCHLAND	4	2,68	10,72	3,8	15,2	3,99	15,96	0,02	29,47	1,12	4,48
4	01.01.2018	2247	DAIRY_CHESSSE	ALMETTE OGOREK/ZIOLA 150G	4	2,68	10,72	2,85	11,4	3,99	11,98	0,01	5,96	0,17	0,68

*Рис. 1.* Структура таблицы продаж

Для дальнейшей работы с данной таблицей были выполнены следующие преобразования: преобразованы столбец Date, заведен столбец Month, который содежит информацию о месяце продаж. Создан словарь групп товаров: каждой группе присвоено цифровое значение. Столбцы, имеющие тип Object, но имеющие числовое значение, которые преобразованы в тип данных float64.

Визуальный анализ временных рядов показал, что столбцы имеют хорошую наполненность значениями, а для пустых ячеек была проведена замена на среднее значение по столбцу [2].

Затем был проведен анализ продаж по группам товаров за год, при этом оценивался ассортимент продаваемых товаров и прибыли, получаемой от их продажи, результаты представлены на рис. 2.

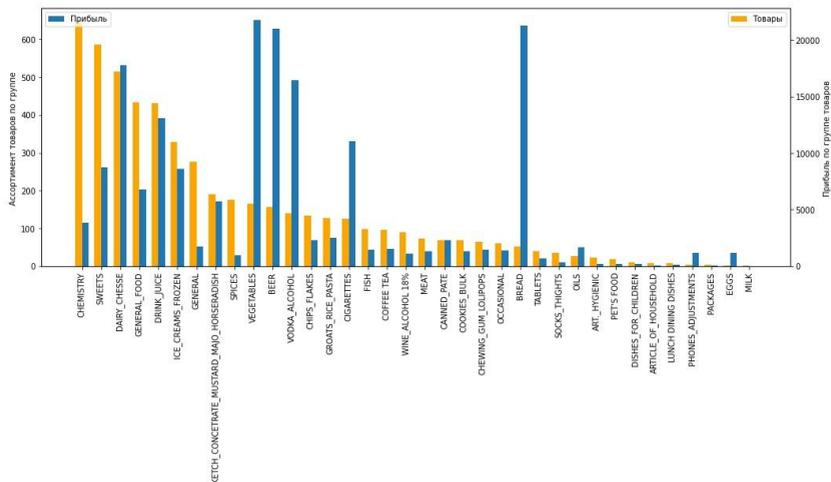


Рис. 2. Анализ продаж по группам товаров за год.

Полученные результаты позволяют сделать вывод, что размер прибыли не зависит от разнообразия ассортимента по группе.

В разрезе по месяцам за год была проанализирована суммированная прибыль от продажи товаров по месяцам (рис. 3).

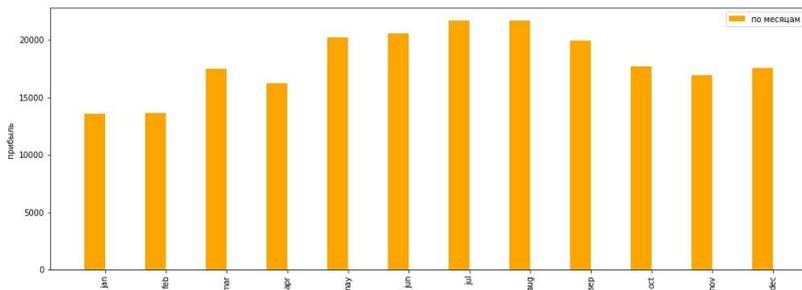


Рис. 3. Суммированная прибыль от продажи товаров по месяцам

Анализ показал, что наиболее высокий уровень прибыли приходился на летний период года.

Для выявления взаимосвязи факторов был проведен корреляционный анализ данных [3] (рис.4) .



Рис. 4. Корреляционный анализ данных

Отдельно рассмотрена линейная корреляция между прибылью и количеством проданных товаров ( $Pquantity$   $pudzmarza = 0.766460$ ). График представлен на рис. 5.

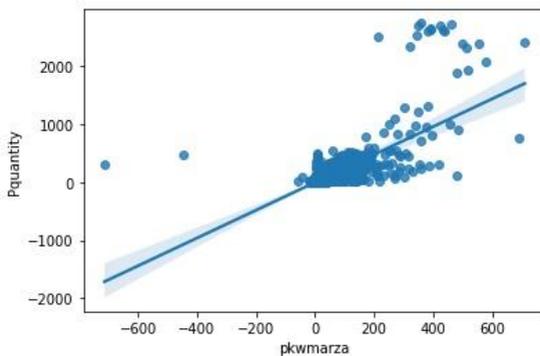


Рис. 5. Корреляция между прибылью и количеством проданных товаров

На рис. 6 представлена линейная корреляция между прибылью и общей суммой продаж ( $\text{pwa\_sn}$   $\text{pkwmarza} = 0.774149$ ).

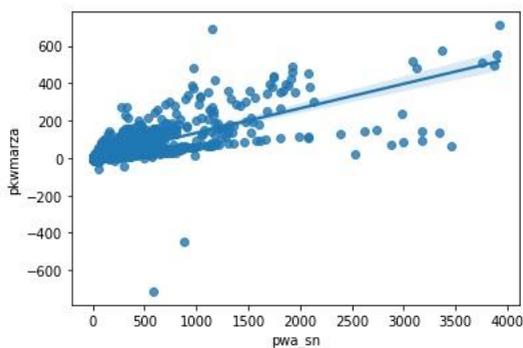


Рис. 6. Корреляция между прибылью и общей суммой продаж

Итоги анализа данных: общая прибыль продаж линейно зависит от количества проданного товара и его стоимости.

Для прогнозирования валовой прибыли в данной работе использована библиотека глубокого обучения Keras, которая обеспечивает взаимодействие с искусственными нейронными сетями [4]. Для анализа данных была выбрана последовательная модель Sequential, которая представляет собой линейную совокупность слоев [5].

Были выбраны две модели с различным количеством выходных узлов в слое и с разным количеством слоев. Исходные данные были

разделены на две части, 80% данные для обучения нейронных сетей и 20% для тестирования и получения прогноза по получению валовой прибыли [6].

Прогнозы, полученные этими моделями, представлены на рис.7 и рис.8.

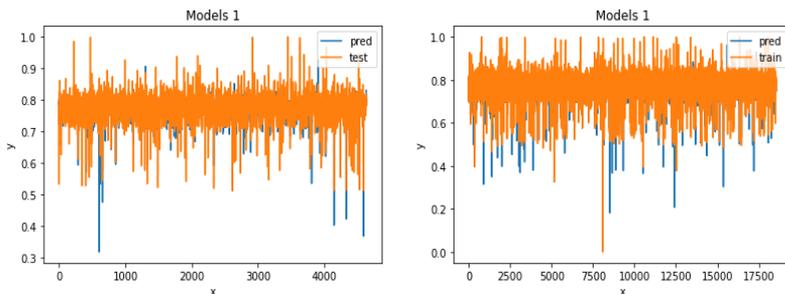


Рис. 7. Сравнение реальных и предсказанных данных моделью 1

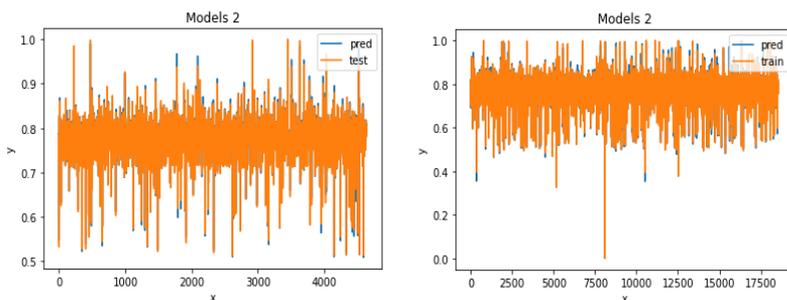


Рис. 8. Сравнение реальных и предсказанных данных моделью 2

Вторая модель, содержащая большее число узлов в слоях и большее количество слоев, показала более высокую точность прогнозирования.

### Заключение

Проведенная работа еще раз подтвердила высокий потенциал использования методов машинного обучения при решении задач прогнозирования. Рассмотренный метод анализа данных применим к любому набору данных. На базе полученных результатов можно провести построение более высокоуровневых моделей для применения в исследовательских целях, а так же для создания систем анализа и прогнозирования в сфере продаж.

### Список литературы

1. Система организации конкурсов по исследованию данных: сайт платформы Kaggle. [Электронный ресурс]. URL: <https://www.kaggle.com/> (дата обращения: 15.11.2021)
2. Параллельный генетический алгоритм отбора значимых факторов, влияющих на эволюцию сложной системы / Мокшин В.В. Вестник Казанского государственного технического университета им. А.Н. Туполева. – 2009. – №3. – С.89-93.
3. Мокшин В.В., Якимов И.М. Метод формирования модели анализа сложной системы / В.В. Мокшин, И.М. Якимов // Информационные технологии. – 2011. – №5. – С. 46–51.
4. Применение математических моделей и алгоритмов при планировании и прогнозировании потребления водных ресурсов / Мокшин В.В., Спиридонова А.В., Спиридонов Г.В. Вестник Воронежского государственного технического университета. – 2021. – №4. – Т.17. – С.57-64.
5. Рекурсивно-регрессионная самоорганизация моделей анализа и контроля сложных систем / Мокшин В.В., Якимов И.М., Юльметьев Р.М., Мокшин А.В. Нелинейный мир. – 2009. – №1. – Т.7. – С.66-76.
6. Рекурсивный алгоритм построения регрессионных моделей сложных вероятностных объектов / Мокшин В.В., Сайфудинов И.Р., Кирпичников А.П. Вестник Технологического университета. –2017. – №9. – Т.20. – С.112-116.