

ПОДГОТОВКА ДАННЫХ И ВЫБОР МЕТОДА ПРОГНОЗИРОВАНИЯ

А.О. Карташов, О.В. Оксюта, И.С. Кущева

ФГБОУ ВО «Воронежский государственный лесотехнический университет
имени Г.Ф. Морозова»

Аннотация. Использование методов прогнозирования позволяет компаниям и людям, отвечающим за выбор стратегии развития, приспосабливаться к изменениям до момента их наступления. Способность предвидеть перемены и предугадать направление развития позволяет принимать четкие и решительные действия в управлении предприятиями. Именно поэтому важность понимания того, какие методы прогнозирования существуют и как они работают, невозможно переоценить.

Ключевые слова: анализ данных, прогнозирование, данные, регрессионный анализ, Байесовские модели.

DATA PREPARATION AND SELECTION OF A FORECASTING METHOD

A.O. Kartashov, O.V. Oksyuta, I.S. Kushcheva

Voronezh State University of Forestry and Technologies named after G.F. Morozov

Abstract. The use of forecasting methods allows companies and individuals responsible for choosing development strategies to adapt to changes before they occur. The ability to anticipate changes and predict future directions enables businesses to take clear and decisive actions. Therefore, it is crucial to understand the different forecasting methods available and how they work.

Keywords: data analysis, forecasting, data, regression analysis, Bayesian models.

Введение

Анализ данных – это процесс, позволяющий взаимодействовать с массивом данных для нахождения необходимой информации. Получаемая информация может быть как простой, например, вывод максимального, минимального или среднего значения из набора данных. Так и сложной – создание математических формул для вычисления значений с несколькими независимыми переменными и формирование системы из них. Примером такой системы можно отнести расчет прибыли с учетом различных факторов. Такими факторами могут быть: учет сезонности, стоимости доставки, потерь от брака и так далее.

Следующим уровнем анализа является созданием моделей, которые способны прогнозировать данные на основе имеющейся информации. Такие модели являются незаменимыми в условиях постоянно меняющихся условий рынка, так как способны предсказать эти изменения до момента, когда они наступили.

Процесс подготовки данных для прогнозирования

Невозможно начинать создавать модель прогнозирования без предварительной обработки данных. Это связано с тем, что “новые данные”, которые мы получим в результате прогнозирования, являются результатом многоступенчатой обработки уже имеющейся у нас информации. Поэтому, если изначальный набор будет некорректным, то и на выходе мы получим неверные прогнозы, на которые нельзя будет полагаться в дальнейшем.

Чтобы избежать этого, необходимо ответственно подходить к сбору данных. Если сбор происходит самостоятельно, необходимо определить четкие критерии для отбора и внесения данных в итоговую выборку. Если же данные берутся из открытых источников, необходимо учитывать, что данные, хранящиеся там, могут быть некорректны или неадекватны, поэтому необходимо продумать методы и алгоритмы для дополнительной обработки.

Следующим шагом должна идти очистка данных, которая включает в себя: устранение пропусков (или замена их на условные значения), нахождение выбросов, аномалий и любых “сломанных” данных. В случае нахождения значений, которые в дальнейшем могут негативно повлиять на итоговый результат прогнозирования, необходимо предпринять меры для исправления изначальной выборки. Такими действиями могут быть:

1. Удаление неверных данных.

- Исправление ошибок.
- Игнорирование аномалий.

Языки программирования и СУБД, где хранятся данные, предоставляют большой арсенал инструментов, которые можно использовать для данных целей. Так функция `df.dropna` на языке Python позволяет очищать данные, удаляя пропуски, функция `DISTINCT` в SQL удаляет дубликаты, а `SELECT` позволяет просматривать все данные, чтобы оценить, насколько данные готовы для обработки.

На рисунках 1 и 2 представлен набор данных в блокноте и csv формате соответственно. Данные в примерах отсортированы и отфильтрованы таким образом, чтобы дальнейшее взаимодействие с ними позволило создать информационную модель для прогнозирования.

Срок заказа	Дата	Год	Месяц	День недели	Категория	Товар	Цвет	Размер	Пол	Возрастная группа	Сегмент клиента	Тип ткани	Сложность пошива	Цена	Себестоимость	Количество	Процент брака					
Срок изготовления	Необходимость доставки	Способ доставки	Стоимость доставки	Стоимость заказа	Оценка пользователя	Источник заказа	Тип клиента	Способ оплаты	Регион	Затраты на рекламу	Сезон	Примечание	Температура	Тип ткани	Способ оплаты	Регион	Затраты на рекламу					
0RD-2025-001	02.01.2025	2025	1	Четверг	Пальто	Пальто зимнее с мехом	Черное	L	Женский	45+	Старшее поколение	Версть+мех	3	18000	12000	1	0.013	5	Да	Курьер СДЭК	500	9
0RD-2025-002	03.01.2025	2025	1	Пятница	Детская	Куртка зимняя	Синий	128	Детский	10-12 лет	Дети	Полиэстер+синтепон	2	4500	3000	2	0.006	3	Да	Почта России	400	9
0RD-2025-003	04.01.2025	2025	1	Суббота	Рубашка	Рубашка классическая	Белая	M	Мужской	25-34	Взрослые	Хлопок	2	3500	2300	3	0.004	2	Нет	Не требуется	0	9
0RD-2025-004	05.01.2025	2025	1	Воскресенье	Платье	Платье трикотажное	Черное	S	Женский	35-44	Взрослые	Версть	1	1800	1200	5	0.003	1	Да	Яндекс.Маркет	400	9
0RD-2025-005	06.01.2025	2025	1	Понедельник	Детская	Варежки	Красные	280	Детский	3-6 лет	Дети	Версть+флис	1	800	500	8	0.002	1	Да	Почта России	300	9
0RD-2025-006	07.01.2025	2025	1	Вторник	Пальто	Пальто с貉осом	Коричневое	XL	Мужской	45+	Старшее поколение	Версть	3	16500	11000	1	0.012	5	Да	Курьер СДЭК	700	9
0RD-2025-007	08.01.2025	2025	1	Среда	Платье	Платье трикотажное	Черное	M	Женский	25-34	Взрослые	Трикотаж	2	4200	2800	4	0.005	3	Нет	Не требуется	0	9
0RD-2025-008	09.01.2025	2025	1	Четверг	Брюки	Брюки классические	Черные	M	Мужской	35-44	Взрослые	Версть	2	4800	3200	3	0.004	2	Да	Почта России	300	9
0RD-2025-009	10.01.2025	2025	1	Пятница	Детская	Комбинезон зимний	Синий	116	Детский	7-9 лет	Дети	Полиэстер+синтепон	2	5200	3400	2	0.003	3	Да	Курьер СДЭК	700	9
0RD-2025-010	11.01.2025	2025	1	Суббота	Головные уборы	Шапка вязаная	Черная	M	Мужской	45+	Старшее поколение	Мех/шкура	2	3200	2100	4	0.006	2	Да	Яндекс.Маркет	400	9
0RD-2025-011	12.01.2025	2025	1	Воскресенье	Пальто	Пальто кашемировое	Бежевое	M	Женский	35-44	Взрослые	Кашемир	3	19000	12500	1	0.010	5	Да	Курьер СДЭК	500	9
0RD-2025-012	13.01.2025	2025	1	Понедельник	Детская	Свитер	Зеленый	122	Детский	7-9 лет	Дети	Версть	2	2800	1800	3	0.004	2	Да	Почта России	400	9
0RD-2025-013	14.01.2025	2025	1	Вторник	Рубашка	Рубашка классическая	Белая	L	Мужской	25-34	Взрослые	Фланель	1	3100	2000	5	0.003	1	Нет	Не требуется	0	9
0RD-2025-014	15.01.2025	2025	1	Среда	Аксесуары	Роскава	Черный	M	Унисекс	18-24	Молодые	Полиэстер	1	2500	1600	6	0.002	1	Да	Яндекс.Маркет	400	9
0RD-2025-015	16.01.2025	2025	1	Четверг	Платье	Платье с длинными рукавом	Синее	S	Женский	18-24	Молодые	Трикотаж	2	4500	3000	4	0.005	3	Да	Курьер СДЭК	500	9
0RD-2025-016	17.01.2025	2025	1	Пятница	Брюки	Брюки утяжеленные	Черные	M	Мужской	35-44	Взрослые	Версть+флис	2	4900	3200	3	0.004	2	Да	Почта России	400	9
0RD-2025-017	18.01.2025	2025	1	Суббота	Детская	Платье	Черное	L	Детский	3-6 лет	Дети	Версть	1	900	600	7	0.002	1	Нет	Не требуется	0	9
0RD-2025-018	19.01.2025	2025	1	Воскресенье	Головные уборы	Шапка фетровая	Черная	M	Женский	45+	Старшее поколение	Фетр	2	3800	2500	2	0.006	2	Да	Курьер СДЭК	700	9
0RD-2025-019	20.01.2025	2025	1	Понедельник	Пальто	Пальто с貉осом	Серое	XL	Мужской	45+	Старшее поколение	Версть+синтепон	3	17000	11000	1	0.011	5	Да	Почта России	400	9
0RD-2025-020	21.01.2025	2025	1	Вторник	Детская	Свитер	Зеленый	122	Детский	7-9 лет	Дети	Полиэстер	2	3900	2600	3	0.004	2	Да	Курьер СДЭК	700	9
0RD-2025-021	22.01.2025	2025	1	Среда	Рубашка	Рубашка классическая	Белая	L	Мужской	25-34	Взрослые	Хлопок	2	3400	2200	4	0.003	2	Нет	Не требуется	0	9
0RD-2025-022	23.01.2025	2025	1	Четверг	Платье	Платье с貉осом	Черное	M	Женский	35-44	Взрослые	Версть	1	1800	1200	5	0.003	1	Да	Яндекс.Маркет	400	9
0RD-2025-023	24.01.2025	2025	1	Пятница	Детская	Комбинезон зимний	Синий	116	Детский	7-9 лет	Дети	Полиэстер+синтепон	2	5200	3400	2	0.003	3	Да	Курьер СДЭК	700	9
0RD-2025-024	25.01.2025	2025	1	Суббота	Головные уборы	Шапка вязаная	Черная	M	Мужской	45+	Старшее поколение	Мех/шкура	2	3200	2100	4	0.006	2	Да	Яндекс.Маркет	400	9
0RD-2025-025	26.01.2025	2025	1	Воскресенье	Пальто	Пальто кашемировое	Бежевое	M	Женский	35-44	Взрослые	Кашемир	3	19000	12500	1	0.010	5	Да	Курьер СДЭК	500	9
0RD-2025-026	27.01.2025	2025	1	Понедельник	Детская	Свитер	Зеленый	122	Детский	7-9 лет	Дети	Версть	2	2800	1800	3	0.004	2	Да	Почта России	400	9
0RD-2025-027	28.01.2025	2025	1	Вторник	Рубашка	Рубашка классическая	Белая	L	Мужской	25-34	Взрослые	Фланель	1	3100	2000	5	0.003	1	Нет	Не требуется	0	9
0RD-2025-028	29.01.2025	2025	1	Среда	Аксесуары	Роскава	Черный	M	Унисекс	18-24	Молодые	Полиэстер	1	2500	1600	6	0.002	1	Да	Яндекс.Маркет	400	9
0RD-2025-029	30.01.2025	2025	1	Четверг	Платье	Платье с длинными рукавом	Синее	S	Женский	18-24	Молодые	Трикотаж	2	4500	3000	4	0.005	3	Да	Курьер СДЭК	500	9
0RD-2025-030	31.01.2025	2025	1	Пятница	Брюки	Брюки утяжеленные	Черные	M	Мужской	35-44	Взрослые	Версть+флис	2	4900	3200	3	0.004	2	Да	Почта России	400	9
0RD-2025-031	01.02.2025	2025	2	Суббота	Детская	Платье	Черное	L	Детский	3-6 лет	Дети	Версть	1	900	600	7	0.002	1	Нет	Не требуется	0	9
0RD-2025-032	02.02.2025	2025	2	Воскресенье	Головные уборы	Шапка фетровая	Черная	M	Женский	45+	Старшее поколение	Фетр	2	3800	2500	2	0.006	2	Да	Курьер СДЭК	700	9
0RD-2025-033	03.02.2025	2025	2	Понедельник	Пальто	Пальто с貉осом	Серое	XL	Мужской	45+	Старшее поколение	Версть+синтепон	3	17000	11000	1	0.011	5	Да	Почта России	400	9
0RD-2025-034	04.02.2025	2025	2	Вторник	Детская	Свитер	Зеленый	122	Детский	7-9 лет	Дети	Полиэстер	2	3900	2600	3	0.004	2	Да	Курьер СДЭК	700	9
0RD-2025-035	05.02.2025	2025	2	Среда	Рубашка	Рубашка классическая	Белая	L	Мужской	25-34	Взрослые	Хлопок	2	3400	2200	4	0.003	2	Нет	Не требуется	0	9

Рисунок 1 – Данные о заказе в блокноте

ID	Заказ №	Дата	Год	Месяц	День недели	Категория	Товар	Цвет	Размер	Пол	Возрастная группа	Тип ткани	Сложность	Цена	Себестоимость	Количество	Процент брака	Срок и вид	Метод доставки	Способ доставки	Стоимость	Оценка	На источник	Тип клиента	Ссылка	Рейтинг	Запр.		
1	0RD-2025	02.01.2025	2025	1	Четверг	Пальто	Пальто зимнее с мехом	Черное	L	Женский	45+	Старшее поколение	Версть+мех	3	18000	12000	1	0.013	5	Да	Курьер СДЭК	500	9	Telegram	Оскара	СБС	СБС-руф Москва		
2	0RD-2025	03.01.2025	2025	1	Пятница	Детская	Куртка зимняя	Синяя	128	Детский	10-12 лет	Дети	Полиэстер+синтепон	2	4500	3000	2	0.006	3	Да	Почта России	400	9	Telegram	Оскара	СБС	СБС-руф Москва		
3	0RD-2025	04.01.2025	2025	1	Суббота	Рубашка	Рубашка классическая	Белая	M	Мужской	25-34	Взрослые	Хлопок	2	3500	2300	3	0.004	2	Нет	Не требуется	0	9	Офлайн	Офлайн	Наличие	Самант-Пет		
4	0RD-2025	05.01.2025	2025	1	Воскресенье	Платье	Платье трикотажное	Черное	S	Женский	35-44	Взрослые	Версть	1	1800	1200	5	0.003	1	Да	Яндекс.Маркет	400	9	Telegram	Оскара	СБС	СБС-руф Казань		
5	0RD-2025	06.01.2025	2025	1	Понедельник	Детская	Варежки	Черные	L	Детский	3-6 лет	Дети	Версть+флис	1	800	500	8	0.002	1	Да	Почта России	300	9	ВКонтакте	Оскара	СБС	Наличие	Самант-Пет	
6	0RD-2025	07.01.2025	2025	1	Вторник	Пальто	Пальто с貉осом	Коричневое	XL	Мужской	45+	Старшее поколение	Версть	3	16500	11000	1	0.012	5	Да	Курьер СДЭК	700	9	Telegram	Оскара	СБС	СБС-руф Москва		
7	0RD-2025	08.01.2025	2025	1	Среда	Платье	Платье трикотажное	Черное	M	Женский	25-34	Взрослые	Трикотаж	2	4200	2800	4	0.005	3	Нет	Не требуется	0	9	Офлайн	Офлайн	Наличие	Красода		
8	0RD-2025	09.01.2025	2025	1	Четверг	Брюки	Брюки классические	Черные	M	Мужской	35-44	Взрослые	Версть	2	4800	3200	3	0.004	2	Да	Почта России	300	9	ВКонтакте	Оскара	СБС	Наличие	Москва	
9	0RD-2025	10.01.2025	2025	1	Пятница	Детская	Комбинезон зимний	Синий	116	Детский	7-9 лет	Дети	Полиэстер+синтепон	2	5200	3400	2	0.003	3	Да	Курьер СДЭК	700	9	Telegram	Оскара	СБС	СБС-руф Новосибирск		
10	0RD-2025	11.01.2025	2025	1	Суббота	Головные уборы	Шапка вязаная	Черная	M	Мужской	45+	Старшее поколение	Мех/шкура	2	3200	2100	4	0.006	2	Да	Яндекс.Маркет	400	9	Telegram	Оскара	СБС	СБС-руф Москва		
11	0RD-2025	12.01.2025	2025	1	Воскресенье	Пальто	Пальто кашемировое	Бежевое	M	Женский	35-44	Взрослые	Кашемир	3	19000	12500	1	0.010	5	Да	Курьер СДЭК	500	9	Telegram	Оскара	СБС	СБС-руф Санкт-Петербург		
12	0RD-2025	13.01.2025	2025	1	Понедельник	Детская	Свитер	Зеленый	122	Детский	7-9 лет	Дети	Версть	2	2800	1800	3	0.004	2	Да	Почта России	400	9	ВКонтакте	Оскара	СБС	Наличие	Самант-Пет	
13	0RD-2025	14.01.2025	2025	1	Вторник	Рубашка	Рубашка классическая	Белая	L	Мужской	25-34	Взрослые	Фланель	1	3100	2000	5	0.003	1	Нет	Не требуется	0	9	Офлайн	Офлайн	Наличие	Самант-Пет		
14	0RD-2025	15.01.2025	2025	1	Среда	Аксесуары	Роскава	Черный	M	Унисекс	18-24	Молодые	Полиэстер	1	2500	1600	6	0.002	1	Да	Яндекс.Маркет	400	9	Telegram	Оскара	СБС	СБС-руф Ростов-на-Дону		
15	0RD-2025	16.01.2025	2025	1	Четверг	Платье	Платье с貉осом	Синее	S	Женский	18-24	Молодые	Трикотаж	2	4500	3000	4	0.005	3	Да	Курьер СДЭК	500	9	Telegram	Оскара	СБС	СБС-руф Москва		
16	0RD-2025	17.01.2025	2025	1	Пятница	Брюки	Брюки утяжеленные	Черные	M	Мужской	35-44	Взрослые	Версть+флис	2	4900	3200	3	0.004	2	Да	Почта России	400	9	ВКонтакте	Оскара	СБС	Наличие	Самант-Пет	
17	0RD-2025	18.01.2025	2025	1	Суббота	Детская	Платье	Черное	L	Детский	3-6 лет	Дети	Версть	1	900	600	7	0.002	1	Нет	Не требуется	0	9	Офлайн	Офлайн	Наличие	Самант-Пет		
18	0RD-2025	19.01.2025	2025	1	Воскресенье	Головные уборы	Шапка фетровая	Черная	M	Женский	45+	Старшее поколение	Фетр	2	3800	2500	2	0.006	2	Да	Курьер СДЭК	700	9	Telegram	Оскара	СБС	СБС-руф Казань		
19	0RD-2025	20.01.2025	2025	1	Понедельник	Пальто	Пальто с貉осом	Серое	XL	Мужской	45+	Старшее поколение	Версть+синтепон	3	17000	11000	1	0.011	5	Да	Почта России	400	9	ВКонтакте	Оскара	СБС	Наличие	Самант-Пет	
20	0RD-2025	21.01.2025	2025	1	Вторник	Детская	Свитер	Зеленый	122	Детский	7-9 лет	Дети	Полиэстер	2	3900	2600	3	0.004	2	Да	Курьер СДЭК	700	9	Telegram	Оскара	СБС	СБС-руф Новосибирск		
21	0RD-2025	22.01.2025	2025	1	Среда	Рубашка	Рубашка классическая	Белая	L	Мужской	25-34	Взрослые	Хлопок	2	3400	2200	4	0.003	2	Нет	Не требуется	0	9	Офлайн	Офлайн	Наличие	Самант-Пет		
22	0RD-2025	23.01.2025	2025	1	Четверг	Платье	Платье с貉осом	Черное	M	Женский	35-44	Взрослые	Версть	1	1800	1200	5	0.003	1	Да	Яндекс.Маркет	400	9	Telegram	Оскара	СБС	СБС-руф Краснодар		
23	0RD-2025	24.01.2025	2025	1	Пятница	Детская	Комбинезон зимний	Синий	116	Детский	7-9 лет	Дети	Полиэстер+синтепон	2	4500	3000	4	0.005	3	Да	Курьер СДЭК	500	9	Telegram	Оскара	СБС	СБС-руф Москва		
24	0RD-2025	25.01.2025	2025	1	Суббота	Детская	Комбинезон зимний	Синий	116	Детский	7-9 лет	Дети	Полиэстер+синтепон	2	5200	3400	2	0.003	3	Да	Курьер СДЭК	700	9	Telegram	Оскара	СБС	СБС-руф Москва		
25	0RD-2025	26.01.2025	2025	1	Воскресенье	Головные уборы	Шапка меховая	Черная	M	Мужской	45+	Старшее поколение	Мех/шкура	2	3200	2100	4	0.006	2	Да	Яндекс.Маркет	400	9	Telegram	Оскара	СБС	СБС-руф Екатеринбург		
26	0RD-2025	27.01.2025	2025	1	Понедельник	Пальто	Пальто кашемировое	Бежевое	M	Женский	35-44	Взрослые	Кашемир	3	19000	12500	1	0.010	5	Да	Почта России	400	9	Telegram	Оскара	СБС	СБС-руф Санкт-Петербург		
27	0RD-2025	28.01.2025	2025	1	Вторник	Детская	Свитер	Зеленый	122	Детский	7-9 лет	Дети	Полиэстер	2	3900	2600	3	0.004	2	Да	Почта России	400	9	Telegram	Оскара	СБС	СБС-руф Новосибирск		
28	0RD-2025	29.01.2025	2025	1	Среда	Рубашка	Рубашка классическая	Белая	L	Мужской	25-34	Взрослые	Фланель	1	3100	2000	5	0.003	1	Нет	Не требуется	0	9	Офлайн	Офлайн	Наличие	Самант-Пет		
29	0RD-2025	30.01.2025	2025	1	Четверг	Аксесуары	Роскава	Черный	M	Унисекс	18-24	Молодые	Полиэстер	1	2500	1600	6	0.002	1	Да	Яндекс.Маркет	400	9	Telegram	Оскара	СБС	СБС-руф Ростов-на-Дону		
30	0RD-2025	31.01.2025	2025	1	Пятница	Платье	Платье с貉осом	Синее	S	Женский	18-24	Молодые	Трикотаж	2	4500	3000	4	0.005	3	Да	Курьер СДЭК	500	9	Telegram	Оскара	СБС	СБС-руф Москва		
31	0RD-2025	01.02.2025	2025	2	Суббота	Детская	Комбинезон зимний	Синий	116	Детский	7-9 лет	Дети	Полиэстер+синтепон	2	5200	3400	2	0.003	3	Да	Курьер СДЭК	700	9	Telegram	Оскара	СБС	СБС-руф Краснодар		
32	0RD-2025	02.02.2025	2025	2	Воскресенье	Головные уборы	Шапка вязаная	Черная	M	Мужской	45+	Старшее поколение	Мех/шкура	2	3200	2100	4	0.006	2	Нет	Не требуется	0	9	Офлайн	Офлайн	Наличие	Самант-Пет		
33	0RD-2025	03.02.2025	2025	2	Понедельник	Пальто	Пальто кашемировое	Бежевое	M	Женский	35-44	Взрослые	Кашемир	3	19000	12500	1	0.010	5	Да	Почта России	400	9	Telegram	Оскара	СБС	СБС-руф Санкт-Петербург		
34	0RD-2025	04.02.2025	2025	2	Вторник	Детская	Свитер	Зеленый	122	Детский	7-9 лет	Дети	Полиэстер	2	3900	2600	3	0.004	2	Да	Почта России	400	9	ВКонтакте	Оскара	СБС	Наличие	Самант-Пет	

Как можно заметить, в исходных данных нет пропусков. Сами данные отфильтрованы по различным категориям: ID, категория, пол, размер, цена, стоимость, оценка, региона и так далее.

Создание оптимальных форматов для массива данных, позволяющих повысить эффективность и сократить избыточностью, называется структурированием и агрегацией. Необходимо учитывать, какой метод прогнозирования будет применяться в той или иной модели, чтобы привести значения к единой структуре. Например, для модели прогнозирования ARIMA данные должны быть в формате временных рядов, а регрессия должна представлять собой таблицу с независимыми и зависимыми переменными.

Агрегация, в свою очередь, позволяет уменьшать объемы данных и выделять основные параметры. Этого можно достичь за счет сбора данных в более широкие группы (отчеты о рейтингах и затратах за длинный временной промежуток вместо короткого). К методам агрегации относится:

1. Нахождение максимального значения
2. Нахождение минимального значения
3. Нахождение среднего значения
4. Подсчет количества данных по определенным критериям
5. Вычисление медианного значения
6. Суммирование по категориям
7. Группировка данных
8. Сортировка данных

Однако необходимо учитывать, что потеря детализации или ошибки группировки также может создавать проблемы в процессе работы с данными. Поэтому необходимо правильно использовать библиотеки, запросы и функции, которые позволяют взаимодействовать с функциями агрегации в языках программирования или СУБД.

Методы прогнозирования

Выбор правильного метода прогнозирования позволяет создать четкую модель, способную работать с ней на постоянной основе Основными методами являются:

1. Временные ряды.
2. Регрессионный анализ.
3. Байесовские методы.

Временные ряды могут использоваться для прогноза продаж, обращаясь к данным за определенный промежуток времени (месяц, полгода, года и так далее). Однако необходимо учитывать, что если набор данных содержит множество аномалий или зависимостей, то итоговый результат может быть неверным.

Регрессионный анализ формирует зависимость между независимыми и зависимыми переменными. Одним из примеров использования данного анализа является прогноз клиентского поведения на основе данных о рекламе и продажах.

Байесовские модели позволяют постоянно обновлять прогнозы с учетом обновления, замены и удаления данных. Но использование данного метода сопряжено с рисками вычислительной сложности, а также необходимостью настройки распределений.

Выбор между методами машинного обучения (ML) и гибридными методами, включающими в себя традиционные статистические подходы с применением ML, может быть крайне затруднителен. В таблице 1 представлено сравнение по различным критериям.

Таблица 1 – Сравнительная таблица

	Гибридные методы	Машинное обучение
Точность	Около 75-85%	Около 90% в сложных системах
Объемы данных	Можно использовать в диапазоне от нескольких сотен до нескольких тысяч записей	Нельзя использовать, если количество записей меньше нескольких тысяч
Устойчивость к шуму	Средняя	Высокая
Возможности для применения	Планирование на основе исходных данных, спрос с учетом сезонности	Прогнозирование рейтингов, продаж, спроса

Практическая значимость исследования заключается в выявлении процессов, которые могут помочь в подготовке данных для дальнейшего взаимодействия с ними информационной системе. В ходе исследования сделаны выводы о том, какие методы прогнозирования могут применяться к набору данных с учетом точности, объема данных, устойчивости к шуму, а также итоговому назначению.

Список литературы

1. Савочкина, Д. С. Исследование технологий визуализации информации и данных и разработка прототипа модуля визуализации многомерных данных в системах аналитики / Д. С. Савочкина // Международная научно-практическая конференция по компьютерной и информационной безопасности (INFSEC 2023): сборник статей, Екатеринбург, 30 июня 2023 года / ООО «Институт цифровой экономики и права». – Екатеринбург: Общество с ограниченной ответственностью "Институт Цифровой Экономики и Права", 2023. – С. 108-109.
2. Дудорова, А. В. Применение Bootstrap метода в анализе поведенческих данных и метод визуализации данных выбора из трех объектов / А. В. Дудорова, Д. А. Подгрудков, Е. П. Крученкова // Зоологический журнал. – 2023. – Т. 102, № 3. – С. 349-353. – DOI 10.31857/S0044513423020046.
3. Аксенов, М. С. Краткие сведения о power Bi: загрузка данных, визуализация данных / М. С. Аксенов, Р. Е. Солдатов // Педагогическое образование на Алтае. – 2017. – № 2. – С. 7-13.
4. Гамаюнова, О. А. Анализ данных и BI-системы в маркетинге: визуализация и интерпретация данных для принятия решений / О. А. Гамаюнова // Достижения науки и технологий-ДНиТ-11-2023 : Сборник научных статей по материалам II Всероссийской научной конференции, Красноярск, 27–28 февраля 2023 года. Том Выпуск 7. – Красноярск: Общественное учреждение "Красноярский краевой Дом науки и техники Российского союза научных и инженерных общественных объединений", 2023. – С. 364-368.

References

1. Savochkina, D. S. Research of information and data visualization technologies and development of a prototype module for visualization of multidimensional data in analytics systems / D. S. Savochkina // International Scientific and Practical Conference on Computer and Information Security (INFSEC 2023) : collection of articles, Yekaterinburg, June 30, 2023 / Institute of Digital Economics and Law, LLC. Yekaterinburg: Limited Liability Company "Institute of Digital Economics and Law", 2023, pp. 108-109.
2. Dudorova, A.V. Application of the Bootstrap method in the analysis of behavioral data and a method for visualizing data from a choice of three objects / A.V. Dudorova, D. A. Podgrudkov, E. P. Kruchenkova // Zoological Journal. – 2023. – Vol. 102, No. 3. – pp. 349-353. – DOI 10.31857/S0044513423020046.
3. Aksenov, M. S. Brief information about power Bi: data loading, data visualization / M. S. Aksenov, R. E. Soldatov // Pedagogical education in Altai. - 2017. – No. 2. – pp. 7-13.
4. Gamayunova O. A. Data analysis and BI systems in marketing: visualization and interpretation of data for decision-making / O. A. Gamayunova // Achievements of Science and technology-DNiT-11-2023 : Collection of scientific articles based on the materials of the II All-Russian Scientific Conference, Krasnoyarsk, February 27-28, 2023. Volume 7. Krasnoyarsk: Krasnoyarsk Regional House of Science and Technology of the Russian Union of Scientific and Engineering Public Associations, Krasnoyarsk, 2023, pp. 364-368.