

Разработка алгоритма построения модели предсказания рекомендательной системы

Дерябина П.С.

Санкт-Петербургский государственный
электротехнический университет

Вендинговая машина

Вендинговая машина - автомат с небольшим набором продуктов примерно одного типа.

Особенности:

- нет информации о покупателях;
- покупка одного товара за один раз;
- специфический набор популярных продуктов для конкретной машины.



Цель

Разработать алгоритм построения модели рекомендательной системы, учитывающей особенности вендинговой торговли:

- нет информации о покупателях;
- покупка одного товара за один раз;
- специфический набор популярных продуктов для конкретной машины;
- небольшое число товаров примерно одного типа.

Задачи

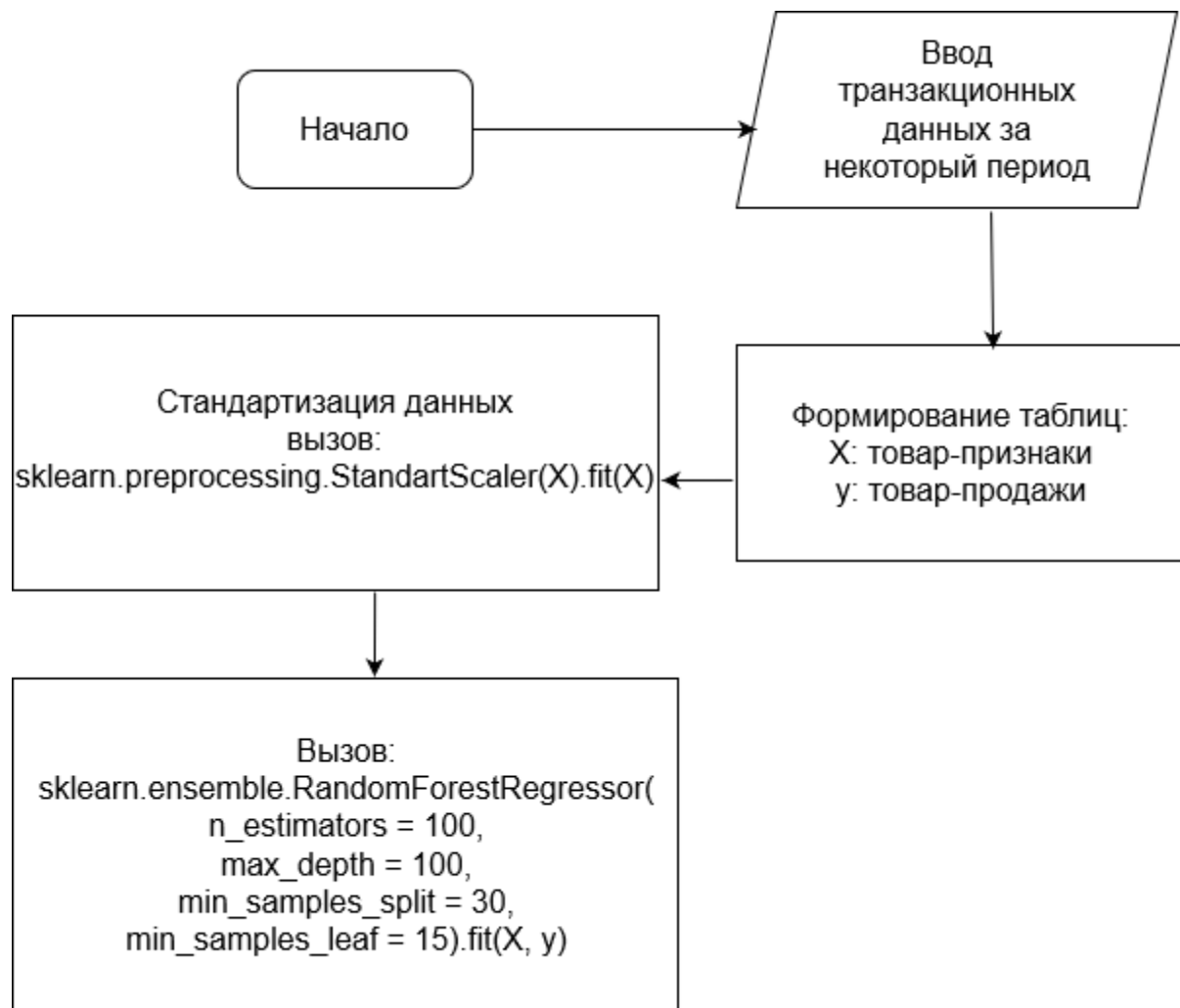
- Подбор методов регрессии для нахождения соответствия каждому товару его значения популярности (предположительное число продаж).
- Оценка метода на нескольких наборах разных данных с учетом особенностей вендинговой торговли.
- Формирование плана дальнейшего исследования.

Наборы данных

Были выбраны три набора данных со следующими характеристиками:

Набор данных	Число записей	Число признаков	Измерение продаж
Товары продуктового магазина	10000	3	Число проданных товаров
Товары супермаркета	15000	6	Число проданных товаров
Товары нижнего белья	90000	33	Рейтинг товара, присвоенный покупателем





Тестирование аналогов

	Товары продук- тового магазина	Товары супер- маркета	Товары нижнего белья
Линейная ре- грессия	0.22	0.32	0.12
Полиномиальная регрессия	0.32	0.34	0.19
Лассо-регрессия	0.21	0.21	0.07
Ридж-регрессия	0.19	0.25	0.11
Метод опорных векторов	0.56	0.59	0.12
Дерево решений	0.73	0.69	0.68
Нейронная сеть	0.55	0.32	0.13

Тестирование модели

Набор данных	Метрика <i>MSE</i>	Метрика <i>MAE</i>	Метрика <i>MAPE</i>	Метрика R^2
Товары продуктового магазина	0.04	0.03	0.006	0.81
Товары супермаркета	0.05	0.04	0.006	0.75
Товары нижнего белья	0.02	0.02	0.005	0.91

Дальнейшая работа

В дальнейшем предлагается рассмотреть следующие задачи

- исследования пространства параметров модели;
- исследование пространства признаков данных для поиска каких-либо зависимостей;
- поиск других подходящих данных для обучения с большим количеством признаков и записей.

Спасибо за внимание!