

Шабунин Глеб Вадимович

студент магистратуры

2 курс, направление «Информатика и вычислительная техника»

Национальный исследовательский технологический университет

«МИСиС»

Научный руководитель: Темкин Игорь Олегович

доктор технических наук, доцент

заведующий кафедрой «Автоматизированные системы управления»

Национальный исследовательский технологический университет

«МИСиС»

Россия, г. Москва

СЕГМЕНТАЦИЯ КЛИЕНТСКОЙ БАЗЫ И ПРОГНОЗИРОВАНИЕ ЕЕ ДИНАМИКИ НА ПРИМЕРЕ МЕТАЛЛУРГИЧЕСКОЙ КОМПАНИИ

***Аннотация:** В данной работе описана методика построения прогноза развития клиентской базы компании на основе сегментации клиентов с похожим поведением спроса. Для сегментации клиентов и построения прогноза динамики развития используются алгоритмы интеллектуального анализа данных, а также элементы предиктивной аналитики и машинного обучения. В качестве инструмента для реализации описанной в данной работе методики применяется высокоуровневый язык программирования Python.*

***Ключевые слова:** временной ряд, кластеризация, машинное обучение, прогнозирование, сегментация.*

***Annotation:** This paper describes a methodology for predicting the development of a company's customer base based on customer segmentation with similar demand behavior. For customer segmentation and forecasting the development dynamics, data mining algorithms are used, as well as elements of predictive analytics and machine*

learning. A high-level Python programming language is used as a tool for implementing the technique described in this paper.

Keywords: *time series, clustering, machine learning, forecasting, segmentation.*

Одним из важнейших аспектов управленческой деятельности компании является оптимальная работа с клиентской базой. Сегментация клиентов помогает компании оценить покупателя, выявить узкие места своих бизнес-процессов, определить неудовлетворенные потребности клиента и, как следствие, сократить расходы компании [1]. Мониторинг и прогнозирование состояния каждой из групп клиентов в совокупности со знанием о влиянии каждой группы на финансовые показатели представляет собой мощный инструмент для построения ценовой политики, разработки маркетинговых кампаний и выработке стратегий распространения товаров для каждого сегмента клиентов [2].

Актуальность данной работы обусловлена отсутствием комплексного решения задачи сегментации клиентской базы компании с последующим построением прогноза на будущие периоды для эффективного принятия управленческих решений.

В данной работе разрабатывается система сегментации и прогнозирования сегментов клиентской базы металлургической компании на основе транзакционных данных о продажах металлопроката. Данная система позволяет организовать взаимодействие с клиентом таким образом, чтобы минимизировать трудозатраты сотрудников на обработку заявок на розничную продажу, а также время на формирование бизнес-предложения.

Для разработки системы необходимо проанализировать существующие методы и подходы для управления клиентской базой компании, занимающейся розничными продажами металлопроката. Важно определить используемые модели для сегментации, прогнозирования, а также метрики для сравнения результатов работы моделей. Помимо этого, необходимо произвести анализ транзакционных данных о покупках клиентов, оценить важность финансовых

показателей, отраженных в транзакционных данных, для последующего использования в модели сегментации и прогнозирования.

С учетом результатов анализа предметной области и оценки существующих методов управления клиентской базой компании было принято решение о внедрении нового подхода, реализованного в системе сегментации и прогнозирования клиентской базы. В таком подходе применяется целый набор моделей сегментации и на основе сравнения метрик, определяющих качество сегментации, определяется подходящая модель. Опираясь на сформированные сегменты, производится прогнозирование размеров кластеров на будущие периоды. Такой подход позволяет менеджерам по продажам быстро определить стратегию работы с целыми группами клиентов, а также оценить текущее состояние перспективных групп [3].

Подход для прогнозирования кластеров клиентов подразумевает выполнение следующих шагов:

1. Первичная очистка анализируемых транзакционных данных о продажах металлопроката;
2. Определение агрегированных финансовых показателей;
3. Кластеризация клиентов по агрегированным показателям с использованием трех моделей кластеризации:
 - a. k-средних;
 - b. Иерархическая кластеризация методом «Уорда»;
 - c. Спектральная кластеризация
4. Оценка результатов моделей кластеризации при помощи индекса «Силуэта» и «Калинского-Харабаса» [4];
5. Выбор подходящей модели с наилучшими значениями индексов;
6. Определение размеров полученных кластеров за известные месяцы;
7. Прогнозирование размеров кластеров с помощью модели тройного экспоненциального сглаживания.

Транзакционные данные о продажах металлопроката содержат вещественные и категориальные признаки совершаемых клиентами сделок. К

вещественным признакам исходной выборки данных относятся: количество отгруженных тонн или оказанных услуг; выручка, полученная в результате сделки, дебиторская задолженность по сделке; маржинальная прибыль с выручки. К категориальным признакам относятся: уникальный идентификатор клиента компании; индустрия клиента; тип клиента (юридическое или физическое лицо); код продукта или услуги; дата совершения сделки; номер и позиция заказа. Вещественные признаки сделки были агрегированы в разрезе каждого анализируемого клиента. Помимо исходных признаков, было сформировано среднее количество заказов за анализируемый временной промежуток и среднее количество позиций в заказе клиента.

С учетом того, что дисперсия значений признаков слишком велика, были применены алгоритмы нормализации данных.

К сформированному набору признаков были применены модели k-средних, иерархической кластеризации методом «Уорда» и спектральной кластеризации с использованием элементов машинного обучения [5]. Реализация моделей была произведена с использованием высокоуровневого языка программирования Python и библиотеки для интеллектуального анализа данных scikit-learn. В результате построения моделей были получены значения индексов для выбора подходящей модели, представленные в таблице 1.

Таблица 1 – Значения индексов результатов кластеризации

№	Модель	Индекс «Силуэта»	Индекс «Калинского-Харабаса»
1	k-средних	0,27	1019,39
2	Иерархическая кластеризация методом «Уорда»	0,38	1829,20
3	Спектральная кластеризация	0,29	1534,22

Выбрав модель с наилучшими значениями индексов и ее результаты кластеризации, была произведена кластеризация клиентов в ежемесячной разбивке. Динамика изменений размеров кластеров может быть представлена в виде набора временных рядов. Набор данных был разбит на обучающую и

тестовую выборку, после чего была использована модель тройного экспоненциального сглаживания (модель Хольта-Уинтерса) для прогнозирования размеров кластеров на будущие временные периоды. Модель Хольта-Уинтерса (тройного экспоненциального сглаживания), используемая для прогнозирования процессов, обладающих трендом и сезонностью, является одной из популярных моделей для долгосрочного прогнозирования.

В результате применения модели прогнозирования была получена динамика изменения кластеров клиентов, которая представлена на рисунке 1.

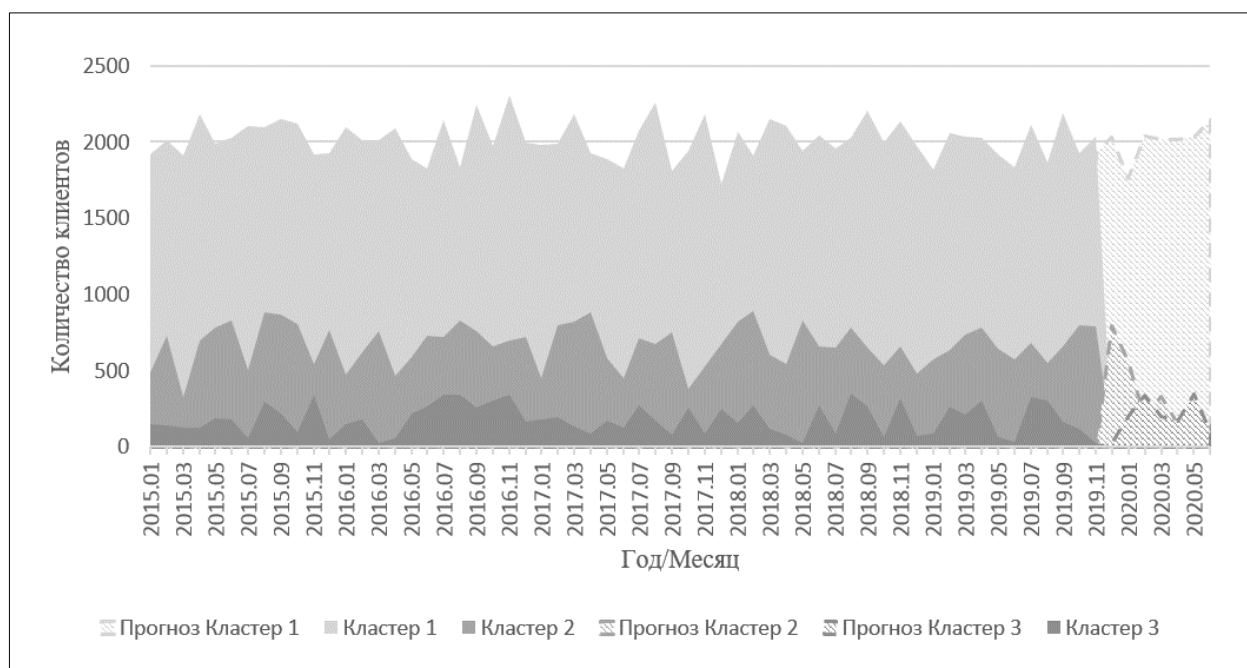


Рисунок 1. Динамика изменения кластеров

Преимуществом предлагаемого подхода и созданной системы кластеризации и прогнозирования клиентской базы компании является упрощение и ускорение процесса формирования бизнес-предложения на розничную продажу металлопрокатных изделий.

Снижение временных затрат сотрудников, построение стратегий продаж в отношении групп клиентов, а также возможность оценки будущего состояния клиентской базы компании позволяют оптимизировать процесс

продажи товара, увеличить выручку компании, сократить затраты на сотрудников компании.

Список использованных источников:

1. Kotler P., Armstrong G. Principles of Marketing, 18th Edition. – Prentice Hall International, 2020. – P. 201-229.

2. Ойнер О.К. Управление результативностью маркетинга: учебник и практикум для бакалавриата и магистратуры, 2-е издание. – М.: Издательство Юрайт, 2019. – 122 с.

3. Kumar V., Werner J. Reinartz. Customer Relationship Management: Concept, Strategy, and Tools. – Springer Texts in Business and Economics, 2012. – P. 102-105.

4. Desgraupes B. Clustering Indices. – University Paris Ouest Lab Modal'X, 2017. – P. 9.

5. Bruce P., Bruce A. Practical Statistics for Data Scientists: 50 Essential Concepts. – O'Reilly Media, 2017 – P. 98.