

Планирование продаж и производства с применением технологий Big Data

Planning sales and production using Big Data technologies

Чугунов В.Р. (Chugunov V.R.)¹, Лычагин К.А. (Lychagin K.A.)²,
Агейкин М.А. (Ageykin M.A.)³, Жукова Л.В. (Zhukova L.V.)⁴

¹ЗАО «ЕС-лизинг» (JSC "EC-leasing"), руководитель отдела (head of department),
vchugunov@ec-leasing.ru

²ЗАО «ЕС-лизинг» (JSC "EC-leasing"), руководитель отдела (head of department),
klychagin@ec-leasing.ru

³ЗАО «ЕС-лизинг» (JSC "EC-leasing"), руководитель отдела (head of department),
mageykin@ec-leasing.ru

⁴ЗАО «ЕС-лизинг» (JSC "EC-leasing"), специалист (expert),
lvzhukova@mail.ru

Аннотация. При планировании производства и продаж произведенной продукции крайне необходимо найти баланс спроса и предложения и сформировать товарное предложение, дающее наилучшие финансовые результаты. В цепочке производитель-продавец важно понять, сколько и какой продукции необходимо произвести, чтобы не было недо/перепроизводства и забитых полок с товаром, не пользующимся спросом, а также недостатка ликвидного товара у продавца. Для решения этой задачи требуется собрать и обработать большой объем самых разнородных данных как из открытых источников сети интернет, так и от заказчика. В работе приведен опыт решения задач указанного класса с применением технологии Big Data.

Abstract. By planning production and marketing strategy it is necessary to find the balance between customer's preferences and manufacturability. It is very important to understand the manufacturer-vendor chain, i.e. to avoid over/underproduction, oversupply and shortage of goods. In order to solve this problem it is required to collect and process a great volume of different data both from open sources and from the customer. This article shows you the Big Data Technology solution experience.

Keywords: Ключевые слова: планирование продаж, открытие источники, машинное обучение, sales planning, open data, machine learning, big data.

В настоящий момент в области экономики и в области информационных технологий сложилась благоприятная ситуация, когда экономические потребности и технологические возможности позволяют решать такие задачи, которые на более ранних стадиях развития отраслей решались долго и требовали больших трудозатрат.

Например, маркетологи решают задачу обеспечения производства и реализации товаров, которые отвечают существующему и, что более важно, потенциальному спросу потребителей, что в свою очередь влияет на одну из ключевых задач производства: распланировать объем и сроки производства продукции с указанием конкретных количественных и качественных показателей по всей номенклатуре товаров.

Рынок потенциальных покупателей в розничной торговле, как и практически любой продукции крайне неоднороден. Существуют значительные различия в поведении покупателей, действующих на рынке розничной торговли, что обусловлено различием в экономическом положении потребителей, их предпочтений, покупательских привычках и традициях того места где потребители находятся. Поэтому немаловажно при решении задач увеличения продаж определить характеристики торговой привлекательности торговых точек по месту их положения используя открытые данные из интернет-среды для маркетинговых исследований.

Если смотреть на проблему увеличения продаж через оптимизацию ассортиментной матрицы, то здесь можно выделить анализ потребительских предпочтений как ключ к решению проблемы. Данной проблемой занимаются такие ученые как О.Н. Романенкова, Очковская М.С., Рыбалко М.А. Исследованиями методов сбора и использования маркетинговой информации занимается Б.Е. Токарев.

В процессе теоретического исследований данной проблемы было особое внимание уделено модели «Brand Key» (рисунок 1), которую использует компания Unilever (иногда ее называют модель Unilever Brand Key – UBK). В модели сначала анализируется конкурентное окружение и выделяется целевая аудитория, именно это и является одним из преимуществ модели: описание целевой аудитории и конкурентного окружения не вынесено за ее рамки, тем самым подчеркивается, что именно на основе этих данных создается бренд.

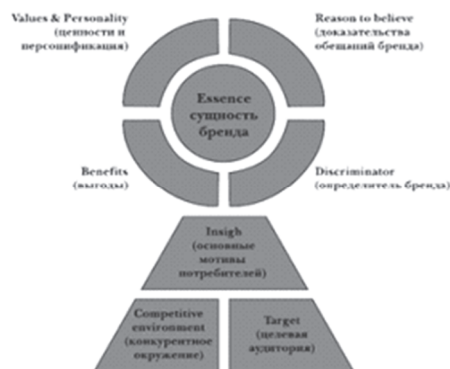


Рисунок 1.

Целью нашей работы является помощь высшему руководящему составу заказчика в принятии управленческих решений, направленных на увеличение объема продаж, путем оптимизация состава товаров индивидуальной ассортиментной матрицы (ИАМ) для каждой конкретной торговой точки (ТТ).

В процессе анализа подходов к решению данной задачи было достигнуто понимание, что нам предстоит столкнуться с огромным объемом, порядка нескольких десятков терабайт, собираемой и обрабатываемой информации, как в структурированном, так и неструктурированном виде. Это связано с тем, что на вход системы планирования продаж-производства поступают данные о ТТ, которых порядка 1 миллиона. Данные, необходимые для достижения цели работы, поступают из информационных систем, уже находящихся в постоянной эксплуатации у заказчика, а также и из открытых источников сети интернет. Развитие информационных технологий в настоящий момент позволяет решать вышеописанные классические задачи, используя новые возможности, предоставляемые технологией Big Data.

Решение данной задачи можно разбить на следующие этапы:

- Этап 1 – Цель: формирование эталонной базы данных ТТ, которые уже находятся в БД заказчика, так и новых ТТ, найденных в процессе сбора информации из интернет.
- Этап 2 – Цель: анализ и подготовка списка признаков для объединение ТТ на основе схожих признаков.
- Этап 3 – Цель: создание профиля группы ТТ с определением значений показателей характерных для ТТ данной группы.
- Этап 4 – Цель: формирование состава показателей, характеризующих профиль продаж, потенциал ТТ и потребительские предпочтения из ассортимента товаров Заказчика.
- Этап 5 – Цель: создание индивидуальной ассортиментной матрицы для каждой ТТ.

На этапе 1 необходимо найти информацию о новых ТТ, в которые потенциально можно продавать продукцию производителя, собрать необходимую для оптимизации ИАМ информацию о ТТ из открытых источников сети интернет, объединить ее в единой БД, т.к. информация о ТТ поступает из различных информационных систем заказчика и открытых источников, очистить, нормализовать, обогатить и дедуплицировать, т.е. найти дублируемые записи. Сбор информации о ТТ из открытых источников сети интернет осуществляется при помощи специальных модулей, которые, в свою очередь, состоят из набора краулеров, написанных на языке python и библиотеки scrapy. Краулеры имитируют работу «любопытного пользователя», который открывает каждую ссылку на подключаемом сайте/странице соц. сети, вызывающую у него интерес. Краулеры обладают возможностью формирования очереди запросов для обслуживания нескольких параллельных потоков загрузки данных. Объединение информации о ТТ осуществляется средствами БД PostgreSQL после чего производится очистка адресов. При очистке адресов ТТ используются специальные правила, написанные на языке plpgsql в БД PostgreSQL. Каждое из правил направлено на удаление или замену тех или иных символов, букв, аббревиатур для придания

адресу каждой ТТ структурированного и понятного вида для последующего геокодинга. При геокодинге происходит обогащение адресов координатами широты, долготы, при распознавании – заполнение недостающих данных в адресе ТТ, например, город, область, страна для каждой торговой точки. Данная процедура подразумевает несколько циклов, включая обратную проверку валидности обогащенной информации. Далее обработка адресов заключается в поиске дублей (дедупликация) ТТ на основе правил, написанных на языке `python`. Объем данных из открытых источников на каждую ТТ составляет примерно 5 кбайт и как результат подготавливается эталонная база ТТ для дальнейшего расчета.

При реализации этапов 2 и 3 первым шагом является алгоритмизация, т.е. создание экономико-математической модели, и последующая группировка ТТ по стратам (представляет собой деление населенных пунктов, в которых расположены торговые точки, на группы (страты) по численности населения для последующей группировки ТТ) с формированием списков населенных пунктов для деления на страты с учетом городов-спутников. Главной целью данного шага является определить степень различия между стратами и состав признаков для каждой страты для оптимальной группировки ТТ по их местоположению. Алгоритмов деления ТТ на страты может быть несколько от простейшего алгоритма деления на 3 страты по численности населения городов и населенных пунктов (1 страта – более 1 млн. человек, 2 страта – более 250 тыс. человек и до 1 млн. человек, 3 страта – менее 250 тыс. человек) до деления на 7 страт согласно категоризации городов по статусу и значению: город федерального значения, города республиканского значения и т.д. При этом обоснованность разбиения численности населения городов и населенных пунктов проверялась экспериментально на основе статистического различия средних значений и медиан по 15 признакам по стратам между собой. Вторым шагом является проведение кластеризации ТТ. Кластерный анализ представляет собой класс методов, которые используются для классификации объектов или событий в достаточной степени однородные группы, которые и называются кластерами. Принципиальное значение при кластеризации является то, что объекты в кластерах обязаны быть похожими между собой, но при этом обязательно отличаться от объектов, находящихся в других кластерах. Существует два типа методов кластеризации: иерархические и неиерархические. Выбор метода кластеризации и выбор меры расстояния взаимосвязаны. В иерархической кластеризации важным критерием принятия решения о числе кластеров являются расстояния, при которых происходит объединение кластеров. Относительные размеры кластеров должны быть такими, чтобы имело смысл сохранить данный кластер, а не объединить его с другими. При этом расстояния рассчитываются на основе информации из открытых источников по 15 признакам, характеризующим каждую ТТ с использованием алгоритмов, написанных на языке `python` и библиотек `pandas`, `numpy`, `scikit-learn`.

Разработка экономико-математического обоснования деления на страты, то есть из всех предложенных и проанализированных алгоритмов выбран наиболее оптимальный.

По результатам анализа кластеризации определены наиболее чувствительные признаки кластеризации и которые впоследствии были использованы в наборе характеристик ТТ, по которым впоследствии определены потребительские предпочтения в кластере и далее сформирована ассортиментная матрица, то есть новое – это такая комбинация дала точные ответы во сколько раз можно увеличить продажи достичь потенциала при текущей ассортиментной матрице и также новое – это при какой именно комбинации товарных позиций можно расширить границы потенциальных продаж. Новшество такого подхода заключается в установлении взаимосвязи значений наиболее чувствительных признаков кластеризации (характеристик ТТ) – потребительских предпочтений – возможностей увеличения продаж. Набор характеристик ТТ, как перечень показателей торговой привлекательности по человеку-поток, согласно значению признаков кластеризации, свидетельствует об:

- экономической активности (количество и расстояние до ближайших мест притяжения населения);
- торговой активности (количество и расстояние до ближайших ТТ ka-сетей и не ka-сетей);
- наличия транспортной инфраструктуры (количество и расстояние до ближайших объектов типа остановок, вокзалов, метро).

На этапе 4 определяются показатели, характеризующие профиль продаж, потенциал ТТ, потребительские предпочтения и алгоритмы их расчета. Показатели получают по результатам установления взаимосвязи между особенностями ТТ и результативными показателями, характеризующими продажи, выявляются закономерности в динамике продаж, с использованием различных экономических и статистических методов анализа.

На этапе 5 формируются основные элементы индивидуальной ассортиментной матрицы, состоящей из матрицы потребительских предпочтений, и матрица потенциала ТТ. Матрица потребительских предпочтений состоит из: предпочитаемых наименований категорий товаров в группе продаж, предпочитаемого объема продаж, предпочитаемого диапазона объема продаж, предпочитаемого количества наименований товаров, предпочитаемого количества SKU, предпочитаемого диапазона количества SKU. Матрица потенциала ТТ, состоит из: относительного и абсолютного потенциального размера выручки от реализации, потенциального объема продаж, потенциального количество наименований товаров.

Полученная в итоге индивидуальная ассортиментная матрица представляет собой список из рекомендованных значений, ориентируясь на которые, можно значительно повысить качество принимаемых решений. При этом в автоматизации этих процессов принятия управленческих решений учтены как индивидуальные, так и групповые особенности ТТ.

Литература

1. Мандель И.Д. Кластерный анализ, 1988 год
2. Н.Паклин. «Кластеризация данных: масштабируемый алгоритм CLOPE».
3. Дэниал Фасуло «Анализ последних работ по алгоритмам кластеризации».
4. Н. Паклин «Алгоритмы кластеризации на службе Data Mining».
5. Вятчин Д. А. Нечёткие методы автоматической классификации. — Минск: Техно-принт, 2004. — 320 с.
6. И. А. Чубукова Data Mining. Учебное пособие. – М.: Интернет-Университет Информационных технологий;
7. Ян Янсон «Моделирование».
8. И. А. Чубукова Data Mining. Учебное пособие, 2006.
9. «Доступная аналитика данных», Anil Maheshwari
10. Кеннет Кекьер «Большие данные: революция, которая изменит нашу жизнь, работу и мысли»
11. Кэти О'нейл и Рейчел Шутт “Наука о данных”
12. Под общей редакцией О. Н. Романенковой “Маркетинговые исследования. Теория и практика” Издательство Юрайт, 2014. – 315 с. – Серия: Бакалавр. Базовый курс.
13. Токарев Б.Е. Методы сбора и использования маркетинговой информации: Учебно-практич. пособие. – М.: Экономистъ, 2004. – 256 с.
14. Очковская М.С., Рыбалко М.А. Маркетинг: новые тенденции и перспективы: Учебное пособие. – М.: МАКС Пресс, 2012. – 196 с.