

## **Технологии Big Data в торговле товарами через виртуальные доски объявлений**

**Молоствов Н.И.**, магистрант, ФГБОУ ВО «Московский политехнический университет», Москва, Россия

**Суворов С.В.**, профессор, кандидат экономических наук, ФГБОУ ВО «Московский политехнический университет», Москва, Россия

**Евтихов В.Г.**, доцент, кандидат технических наук, ФГБОУ ВО «Московский политехнический университет», Москва, Россия

**Аннотация.** В данной статье рассматривается проблематика взаимосвязи спроса и полного описания онлайн-объявлений о товарах и услугах, размещенных на виртуальных досках объявлений. Данные были предоставлены интернет-сервисом для размещения объявлений. С помощью технологии Data Mining был произведен первичный анализа больших данных и их визуализация.

**Ключевые слова:** большие данные, интеллектуальный анализ больших данных, доска объявлений, онлайн-объявление, Python 3, Deductor

## **Big data analysis in merchandise trading through classified advertisements website**

**Molostvov N.I.**, undergraduate, Moscow Polytechnic University, Moscow, Russia

**Suvorov S.V.**, professor, candidate of Economic Sciences, Moscow Polytechnic University, Moscow, Russia

**Evtihov V.G.**, associate professor, candidate of Technical Sciences, Moscow Polytechnic University, Moscow, Russia

**Annotation.** This article discusses the relationship between demand and a full description of online ads for goods and services posted on classified advertisements

website. Data was provided by an online service for posting ads. Using Data Mining technology, an initial analysis of big data and its visualization was carried out.

**Keywords:** big data, big data mining, classified advertisements website, online ad, Python 3, Deductor

## **Введение**

Рынок товаров и услуг неизменно связан с нашей жизнью [1]. Он возник еще в глубокой древности. На протяжении всей истории рынка люди давали различные объявления на досках объявлений.

Доска объявлений – место, на котором размещаются объявления, при этом не взималась какая-либо плата, но срок был ограничен. Исторически она представляла собой обозначенную поверхность, куда прикрепляли физические носители (таблички, листки бумаги и т. п.) с нужной информацией. До недавнего времени очень популярны были объявления в газетах, где под доску объявлений был выделен отдельный раздел. В наше время, когда компьютеры и интернет появились практически в каждом доме, появилась возможность создания виртуальной доски объявлений в виде веб-сайта [2]. Первая такая доска появилась более 40 лет назад, в 1978 году [3]. Сейчас по всему миру их счет пошел на тысячи.

Самым многочисленным типом объявлений на современных сайтах – досках объявлений является продажа личных вещей от частных лиц. Оно и понятно, у частного лица, в отличие от компании, нет ресурсов, на которых можно было бы выложить рекламу своего товара. Первыми на ум приходят как раз доски объявлений. Но для успешной продажи мало просто подать объявление.

При продаже товаров в интернете, особенно подержанных, большое значение могут иметь детали описания товара. Описание может быть неинформативным: содержать мало информации или, наоборот, много лишнего. Даже грамматические ошибки способны отпугнуть потенциального покупателя.

Немаловажное влияние на успешную продажу оказывает правильная фотография товара. Фотографии плохого качества или неправильные ракурсы товара уменьшают его привлекательность. И даже если покупателю все понравилось, то его может смутить цена, которую продавец поставил без учета актуальных рыночных цен. Также на спрос могут и другие параметры, не относящиеся напрямую к товару. Например, исторический спрос в определенной географической области или наличие похожих объявлений.

### **Основная часть**

Виртуальные доски объявлений в настоящее время очень популярны. Ежедневно по всему миру публикуются миллионы объявлений. На одной площадке счет может идти на сотни тысяч. Поэтому собрать данные возможно без особых временных затрат. Но современный рынок может быть довольно волатильным [4], поэтому важно использовать не слишком устаревшие данные.

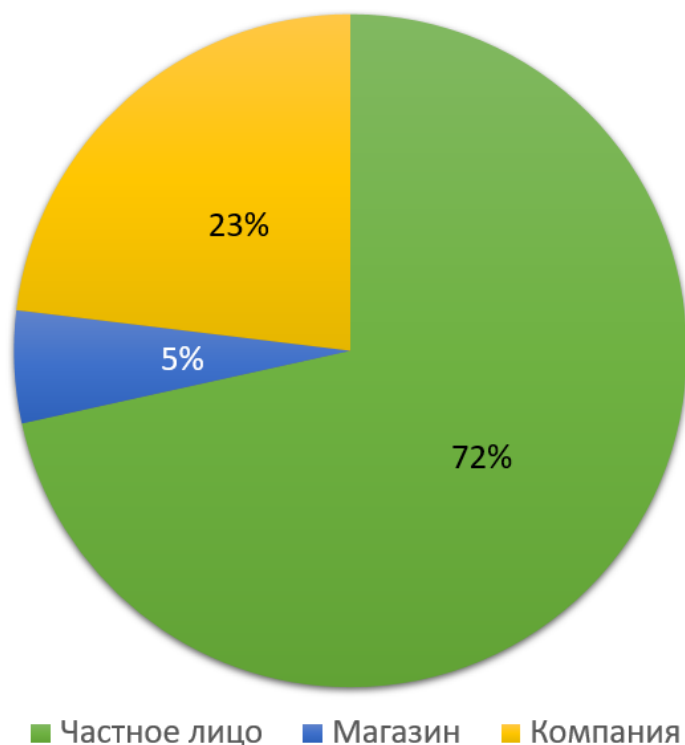
Анализируемые здесь данные представляют собой датасет размером более 1,5 миллиона объявлений, поданных в одном временном промежутке в течение двух недель, и включает в себя следующие характеристики:

- Регион
- Город
- Категория товара
- Подкатегории товара, включающие в себя до 4 уровней вложенности

в зависимости от категории

- Заголовок объявления
- Описание
- Цена
- Фотография
- Дата активации объявления
- Тип продавца

Почти три четверти объявлений публикуются частными лицами, чуть меньше четверти опубликованы компаниями, и небольшой процент принадлежит магазинам (рис. 1).



**Рис. 1 – Распределение объявлений по типу продавцов**

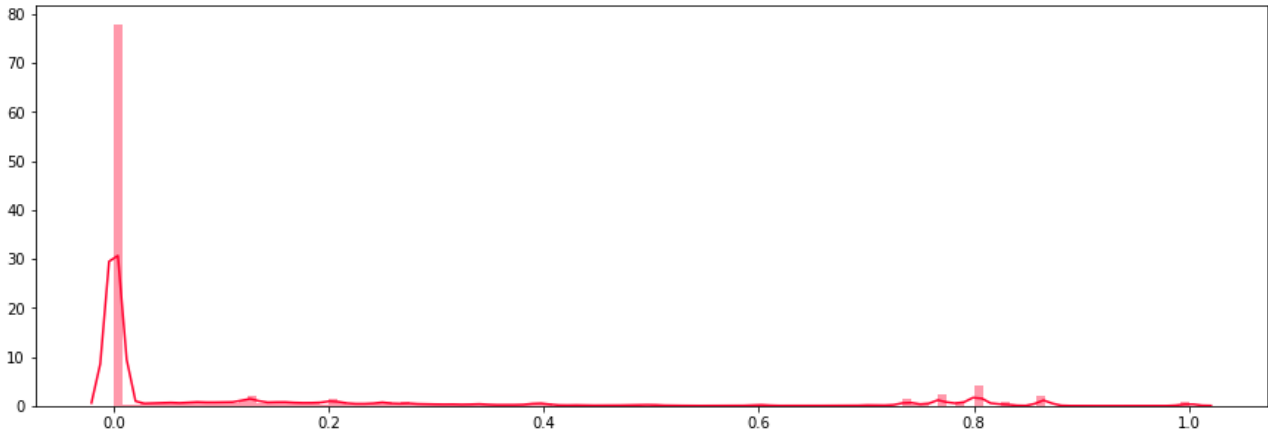
Самой популярной категорией продаваемых товаров, почти половина (46%), являются личные вещи. Эта категория включает в себя одежду и обувь, товары для детей и игрушки, различные украшения, косметику и прочие, в основном небольшие вещи. Так же эту категорию продавцы обычно указывают, когда не смогли определить позицию в представленном на сайте списке. Следом идут товары для дома (12%), включающие в себя бытовую технику, мебель, посуду и стройматериалы; электроника (12%) и недвижимость (10%). Меньший объем числа объявлений занимают хобби (6%), транспорт (5%), услуги (4%), животные (4%) и товары для бизнеса (1%) (рис. 2). Для тематических досок объявлений возможно преобладание товаров по представленной в ней тематике.



**Рис. 2 – Распределение товаров по категориям**

Каждое объявление имеет целевую переменную, рассчитанную после факта совершения сделки. Это вероятность того, что товар действительно был продан. Невозможно с уверенностью проверить совершенную сделку, поэтому значение этой переменной лежит в диапазоне от нуля до единицы, где единица – совершенная сделка, а ноль – отсутствие сделки за представленные в датасете две недели.

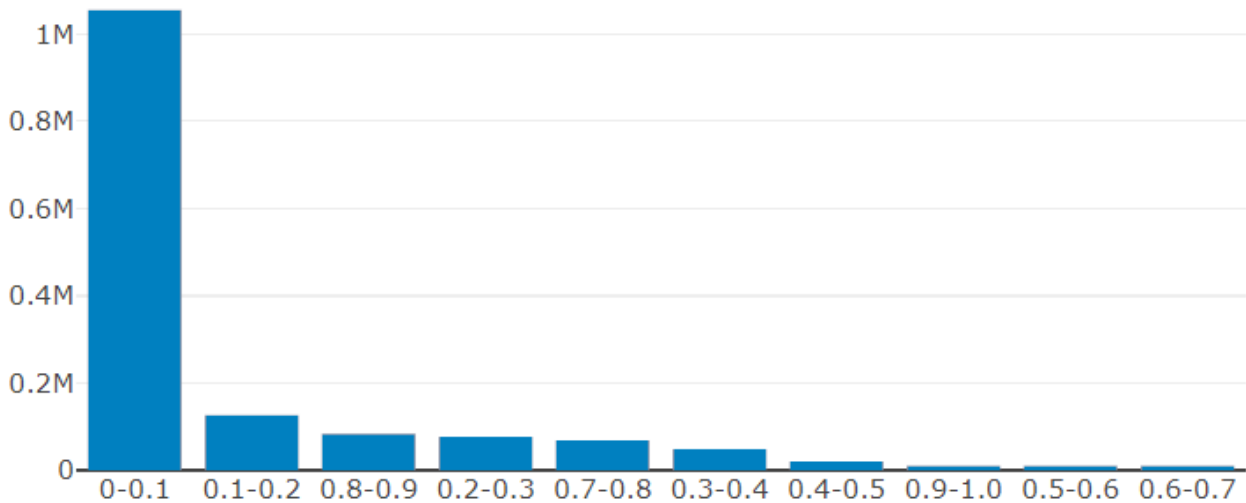
Из графика распределения вероятностей сделки (рис. 3) видно, что большинство предметов, почти 80%, имеют крайне низкую вероятность сделки, очень немногие значения имеют вероятность сделки 0,5 или более.



**Рис. 3 – График распределения вероятностей сделки**

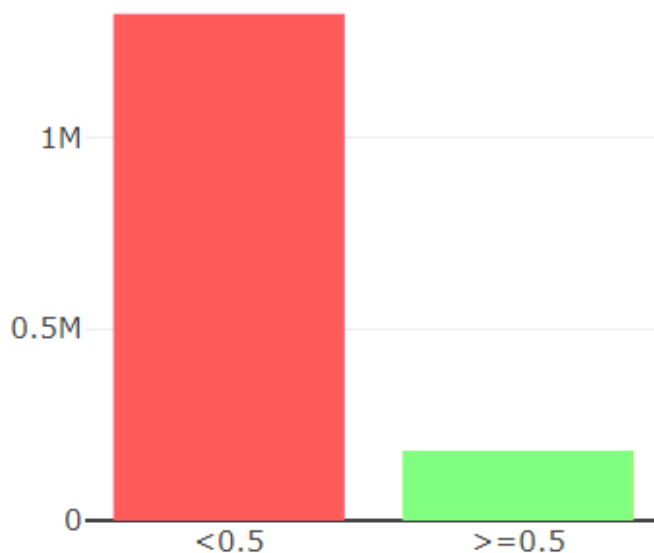
Малое количество объявлений наблюдается вблизи вероятности сделки 1,0, что указывает на то, что в наборе данных есть некоторые элементы, имеющие очень высокое значение вероятности сделки.

Наиболее доминирующими категориями являются предметы, имеющие вероятность сделки: 0 - 0,2; 0,8 - 0,9 (рис. 4).



**Рис. 4 – Категории вероятностей сделки**

Успешные и неуспешные продажи можно условно разделить медианным значением целевой переменной, которое будет равно 0,5. Из 1,5 миллионов объявлений около 1,3 млн. имеют вероятность сделки менее 0,5, в это время только около 200 тысяч объявлений с вероятностью сделки более 0,5 (рис. 5).



*Рис. 5 – Неуспешные и успешные продажи*

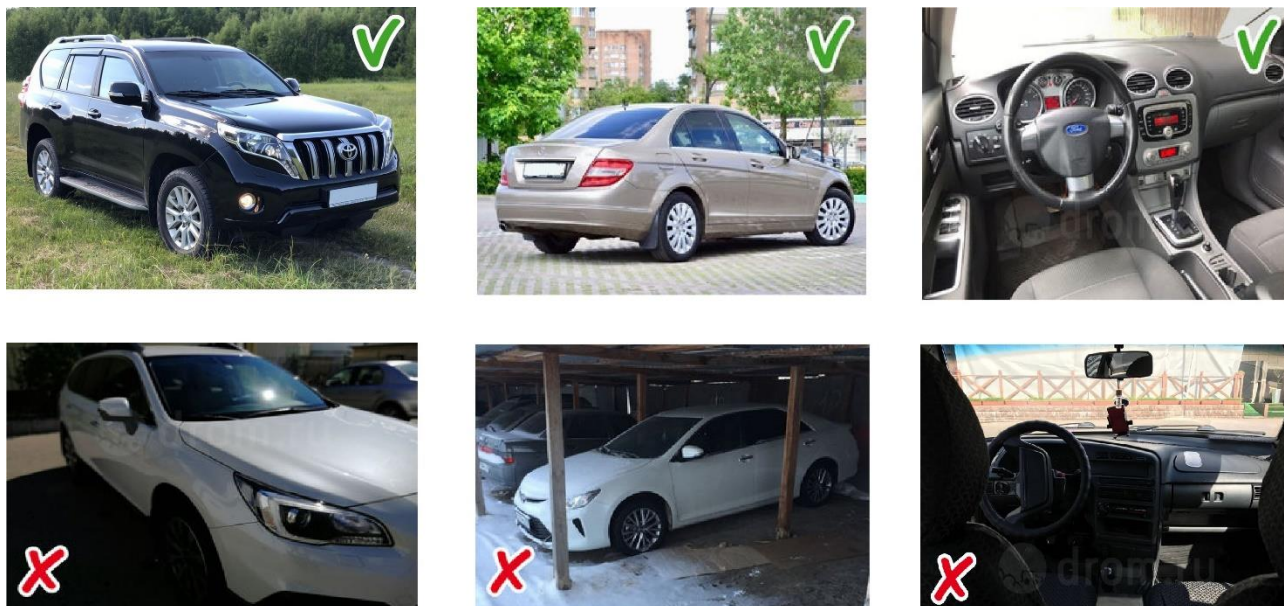
Как видно большинство продавцов терпят неудачу в продаже своих товаров. Им придется сбавить цену или ждать покупателя более длительный срок. Почему же так происходит?

Конечно, это может быть из-за отсутствия спроса на данный вид товара, обусловленное сезонностью, географическими особенностями или моральным устареванием. Но если таких условий нет, то дело скорей всего в самом объявлении.

Еще на этапе прокрутки списка объявлений покупатель смотрит в первую очередь на цену. Большая вероятность, что он даже не откроет объявление, если его не устроит стоимость. Если цена будет слишком большая, то покупатель предпочтет более дешевые варианты. А если слишком низкая, то закрадутся подозрения в качестве и состоянии товара.

Но как обычному человеку, не владеющему информацией о текущем рынке товаров, которые он хочет продать, узнать оптимальную цену? В этом поможет технология интеллектуального анализа данных (Data Mining) [5]. Проанализировав последние проданные аналогичные товары за определенный срок, можно понять допустимый диапазон цен, который не оттолкнет покупателя.

Аналогичной задачей для Data Mining является анализ описания объявления. Если сезон и географическое положение продавцу изменить не под силу, то подправить описание или сделать другую фотографию вполне возможно. Продавец может быть предупрежден о низкой информативности или плохом восприятии описания объявления и о плохой или неподходящей фотографии (рис. 6) еще на этапе подачи объявления.



*Рис. 6 – Хорошие и плохие фотографии при продаже автомобиля*

Для анализа потребуется подготовить имеющиеся данные. Это очень важная часть аналитического процесса и может занимать большую его часть. Необходимо найти неподходящие данные с «мусором» или с пропущенными значениями и убрать их из датасета.

Для первой задачи, определения оптимального диапазона цен товара, потребуется убрать из датасета записи с экстремальными ценами. У около 6% объявлений цена составляет менее 10 рублей или вовсе отсутствует. Вероятно, такие объявления не отображают реальную цену. Еще около 6% объявлений имеют цену более миллиона рублей. В основном это недвижимость и автомобили с высокой ценовой дисперсией. Возможно, стоит игнорировать и



такие объявления для упрощения и точности анализа. В целом отбрасывается чуть более 12 % данных.

Высокая ценовая дисперсия для определенной категории товара может указывать на существование неявных подкатегорий. Трудно получить точную оценку цены для предмета, если мы знаем только широкую категорию, к которой он относится. Такое имеет место быть, когда при подаче объявления предлагается недостаточная классификация. Определение подкатегории товара, если она не была задана явно, является еще одной отдельной задачей для интеллектуального анализа данных.

Следует помнить, что рынок не статический, и происходит постоянное изменение цен. Для выдачи актуальных данных необходимо периодически обновлять данные. Сложно назвать оптимальный период, так как разные категории товаров имеют разную волатильность. Можно взять усредненный показатель и обновлять данные раз в год. Но для повышения актуальности предпочтительней рассчитать период для каждой категории отдельно.

## **Заключение**

Некорректные и неинформативные объявления имеют маленький спрос, даже если товар востребован. Это может негативно сказываться на развитии сектора рынка, представленного досками объявлений, и, в частности, на финансовом положении продавца. Проведенный анализ поможет продавцам быстрее достичь результата, уменьшив срок продажи.

Определение корректности объявления может быть представлено в виде сервиса, интегрированного в виртуальную доску объявлений [6]. Алгоритм вынесет свой вердикт до момента публикации объявления и у продавца появится возможность исправить недочеты.

## **Библиографический список**

1. Sullivan Arthur, Steven M. Sheffrin. Economics: Principles in action. – Prentice Hall, 2009.
2. Электронная доска объявлений [Электронный ресурс]. – URL: [https://ru.wikipedia.org/wiki/Электронная\\_доска\\_объявлений](https://ru.wikipedia.org/wiki/Электронная_доска_объявлений) (дата обращения: 25.02.2020).
3. Feb. 16, 1978: Bulletin Board Goes Electronic [Электронный ресурс]. – URL: <https://www.wired.com/2010/02/0216cbbs-first-bbs-bulletin-board/> (дата обращения: 25.02.2020).
4. С. Вайн. Опционы: Полный курс для профессионалов. – М.: Альпина Паблишер, 2016. – 436 с.
5. А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. Технологии анализа данных: Data Mining, Text Mining, Visual Mining, OLAP. 2 изд. – С-П.: БХВ-Петербург, 2008 – 384 с.
6. М.Р. Когаловский. Методы интеграции данных в информационных системах. – М.: Институт проблем рынка РАН, 2010.

## **References**

1. Sullivan Arthur, Steven M. Sheffrin. Economics: Principles in action. – Prentice Hall, 2009.
2. Electronic bulletin board [Electronic resource]. – URL: [https://ru.wikipedia.org/wiki/Электронная\\_доска\\_объявлений](https://ru.wikipedia.org/wiki/Электронная_доска_объявлений) (accessed date: 25/02/2020).
3. Feb. 16, 1978: Bulletin Board Goes Electronic [Electronic resource]. – URL: <https://www.wired.com/2010/02/0216cbbs-first-bbs-bulletin-board/> (accessed date: 25/02/2020).
4. S. Vine. Options: Complete Course for Professionals. – М.: Alpina Publisher, 2016. – P. 436.

5. A.A. Barseghyan, M.S. Kupriyanov, V.V. Stepanenko, I.I. Cold. Data Analysis Technologies: Data Mining, Text Mining, Visual Mining, OLAP. 2nd ed. – S-P.: BHV-Petersburg, 2008 – P. 384.

6. M.R. Kogalovsky. Data Integration Methods in Information Systems. – M.: Institute for Market Problems RAS, 2010.