Дневник практики

|  |  |
| --- | --- |
| Дата: | Выполненая работа: |
| 04.03.2016 - 10.03.2016 | Сбор материала и изучение алгоритмов прогнозирования |
| 10.03.2016 - 14.03.2016 | Сбор реальных данных по продажам за прошедшие годы для проведения экспериментов |
| 16.03.2016 - 18.03.2016 | Проведение экспериментов с алгоритмом который основан на базе логарифмического тренда |
| 20.03.2016 - 24.03.2016 | Проведение экспериментов с алгоритмом который основан на базе линейного тренда |
| 24.03.2016 - 28.03.2016 | Проведение экспериментов с алгоритмом который основан на базе степенного тренда |
| 01.04.2016 - 05.04.2016 | Проведение экспериментов с алгоритмом который основан на базе полиномиального тренда |
| 05.04.2016 - 15.04.2016 | Сравнении результатов проведенных экспериментов, и принятие решения какую математическую модель использовать |
| 15.04.2016 - 20.04.2016 | Разработка програмного обеспичения |

**СОДЕРЖАНИЕ**

**ВВЕДЕНИЕ**

Целю работы было изучить различные математические модели прогнозирования продаж. Который позже будет внедрен в экспертную систему ведения интернет торговли (ЭСИТ).

Были рассмотрены несколько вариантов алгоритмов, с которыми был проведен эксперемент на практике. Для того чтобы определиться с наиболее подходящей моделью сравнивались результаты эксперемента, а именно построенные прогнозы, метод давший наиболее точный прогноз, был принят как самый эффективный.

После этого начались работы по разработкке экспертной системы интернет торговли (ЭСИТ), и внедрение в эту систему выбраного алгоритма.

ЭСИТ предназначена для рассчета прогноза продаж путём использования уникального алгоритма, построенного в ходе разработки системы.

Исследуемые алгоритм прогноза опираются на статистику продаж по прошедшим периодам. Для работы со статистическими данными был разработан модуль управления базой знаний (МУБЗ). Помимо таких задач как внесения информации в базу данных и ее чтение, МУБЗ выполняет анализ актуальности данной информации. Информация является актуальной в течение определенного времени, обычно такое время составляет три прошедших периода. Понятие период для разной групы товаров может быть разным, поэтому длительность одного периода выставляется индивидуально пользователем в настройках.

ЭСИТ так же предоставляет инструментарий, который автоматизирует управление бизнес прецессами. Опираясь на результаты прогноза, система сама формирует список рекомендуемых товаров для закупки, то есть те товары, спрос на которые растет или, как минимум, остается стабильным, выводит списки пользователю на экран. При отсутствии такой автоматизации владельцу интернет-бизнеса придется самому выполнять этот рутинный анализ. Для интернет магазинов с широким ассортиментом товаров ручной анализ обходится довольно дорого как по времени, так и по финансам.

Экспертная система ведения интернет торговли - это не только запрограммированный алгоритм, а действительно целая система, состоящяя из нескольких частей, каждая из которых выполняет свою функцию.

На сервере находится основная часть разработанного програмного обеспечения, включая алгоритм прогнозирования, модуль управления базой знаний и непосредственно базу данных.   
 Источник данных - это место хранения информации о товаре, это могут быть данные как со склада компани, занимающейся оптовой торговлей, так и данные с собственного склада.   
 API предназначен для того, чтобы наладить взаимодействие с интернет магазинами, которые решили подключить разработанную систему. Именно посредством API клиенты имеют возможность внедрять готовые функции от ЭСИТ в собственные системы, например вывод списка товара для интернет магазина. Гибкость разработанного API позволяет внедрять функциональность не только в веб-версии своих проектов, а так же в мобильные или настольные приложения.  
 Клиентские системы представляют собой такие системы, в которые будет внедряться функциональность ЭСИТ. К ним относятся веб-версии интернет-магазинов, мобильные а также настольные приложения разработанные под различные операционные системы.

**1 АКТУАЛЬНОСТЬ**

На сегодняшниий день рынок электронной коммерции в России активно развивается. Согласно данным исследовательского агентства Data Insight, объем российского рынка электронной коммерции в 2011 году составил 310 млрд. рублей, в 2012 – 392 млрд. рублей (рост на 26%).

Многим ведущим представителям розничной интернет-торговли удалось добиться увеличения продаж на 200-300%.  
 Важно отметить, что темпы роста данного сектора экономики значительно опережают темпы роста экономики страны: за 2012 год ВВП России вырос на 3,5%, а e-commerce – на 26%.  
 Согласно мнению экспертов компании   
J’son & Partners Consulting, в ближайшие годы ожидаются стабильные темпы роста на уровне 15-20% ежегодно, в то время как к 2020 году объем рынка электронной коммерции составит более 2180 млрд. руб. (рисунок 1).



**Рисунок 1 – Объем рынка электронной коммерции в России (млрд. руб.)**

Следует отметить тот факт, что сегмент электронной коммерции в экономике нашей страны стал заметен только за последние несколько лет. До 2011 года доля продаж посредством интернета составляла менее 1% от общего объема торговли России, в 2012 – около 2% (рисунок 2). Схожие показатели имели место в США и Великобритании в 2003 и 2005 годах, на сегодняшний день доля е-commerce в данных странах занимает более 10%, с чего мы можем сделать вывод о наличии определеннго потенциала для дальнейшего роста рынка электронной коммерции России.

Справедливо заметить, что на Российском рынке интернет-торговле присутствует дисбаланс в развитии (рисунок 3). Примерно 60% интернет-пректов ориентированы на большие город, такие как Москва или Санкт-Петербург , где проживают около 15% населения страны. Эксперты из Morgan Stanley прогнозируют что к 2019-2020 году объемы продаж через сеть интернет для вырастут на 30% для столичных городов , и примерно 55% для региональных городов.



Рисунок 3 – Прогноз распределения продаж через Интернет по регионам России

По результатам исследований за 2016 год, на рынке российской интернет коммерции преимущественно распростроняются цифровая продукция, доля которой 64% в рублях эта оценка состовляет 254,9 млрд. (рисунок 4). Большинство покупок в коммерческих онлайн сервисах это как правило бронь билетов, аудио, видео или любая другая цифровая продукция, бронь гостиниц или оплата путевок. Так же онлайн комерция часто используется для оплаты коммунальный услуг, мобильной связи или штрафов.



Рисунок 4 – Структура рынка электронной коммерции в России 2012 и 2017гг.

Что касается реально существующих товаров для которых необходима курьерская доставка их доля на рынке интернет торговли состовляет 36% от общей доли рынка, с особой интенсивностью растут продажи обуви, одежды, косметики а так же парфюмерии. Важно то что эта ниша активно развивается с того момента когда сформировался рынок, в конце 2011 - начале 2012 года общая сумма продаж состовляла 136,9 млрд. рублей, по результатам анализа за 2016 год, этот показатель вырос в уже двое. Эксперты прогнозируют что в ближайшие год-два доля продаж физических товаров закрепит свои позиции на 40 - 45% от общей суммы продаж на рынке интернет торговли. В Российской интернет коммерции присутствует ряд изьянов которые тормозят ее развитие. Одним из таких одним их таких изьянов является то что продвижением занимаются преимущественно, специалисты в области веб-разработки и ИТ, а не специалисты в области торговли и ведения бизнеса. Это приводит к большему количеству не жизнеспособных разработок, сложных в эксплуатации виртуальных магазинов. Из этого мы можем сделать вывод что на сегоднешний день все кто смог занять нишу российской электронной коммерции это мизерный процент от того количества предпренимателей которые пытались занять эту но их проекты оказались не жизни способными, из-за отсутствия опыта ведения интернет бизнеса. Согласно результатам опросса, многие придприниматели как начинающие так и опытные готовы вливать финансы на развитие собственного электронного предприятия. Однако в силу того что их опыт не позволяет сделать правильный выбор в сотовлении списков и растоновки им приоритетов для вывода на виртуальную ветрину, в каких обьемах выполнять закупки для владельцев собственных складов. А так же управлять другими процессами.

Проанализировав сложившуюся ситуацию, сделан вывод, учитывая количество только успешных владельцев электронного бизнеса это очень широая аудитория потенциальных потребителей разрабатываемого продукта.

Так же обязательно нужно принять во внимание приведенный выше факт о том что много людей готовых вливать финансы в интернет комерцию но опасаются это делать в силу отсутствия навыков ведения интернет бизнесса. Это так же очень широкая аудитория потенциальных потребителей. Процент сбыта среди такого рода аудитории гараздо выше, так как они получают экспертную систему которая выполняет все статистические анализы расчитывает прогноз, сортирует списки товаров которые стоит закупать в будующем а какие нет. Так же обладет другими функциями с которыми можно ознакомиться в документации.

Разрабатываемая система практически полностью автоматизирует управление бизнес процессом, что гораздо снизит порог входимости новых предпринимателей на рынок интернет комерции.

Разработка подобной экспертной системы действительно актуальна, более того экспертная система с такой функциональностью востребованна на рынке, по той причине что эта экспертная система позволяет делать расчеты на основе которых повышается точность принятых решений.

**2 РАЗРАБОТКА АЛГОРИТМА РАСЧЕТА ПРОГНОЗА ПРОДАЖ (РПП)**

Сегодня уже достаточно много разработано различных технологий в области прогнозирования.

Активно применяются алгоритмы нейросетевого прогнозирования. Так же довольно часто применяются методы нечёткой логики. Большинство таких задач решается, по средствам применения методов исследования операций, к ним относятся теория игр, имитационное моделирование, трендовый а так же регрессионный анализ. При разрабоке системы РПП был выбран один из возможных алгоритмов строящих прогноз объёма продаж для товароы с сезонным характером.

Важно понимать, такой величиной как “сезон” в области прогнозировании могут быть любые временные колебания - сезонные вариации, так например, в случае когда проводится исследование товарооборота в течение одной недели, то в этом случае в качестве “сезона” выступает один рабочий день. К тому же, длительность цикла колебаний не обязательно один год, эта величина может варьироваться. После того как выявлена величина цикла колебаний, следующим строится прогноз используя для этого временной ряд в котором была выявлена величина цикла, с применением мультипликативных а так же аддитивных моделей.

Аддитивная модель имеет вид:

(#)

где:

*F* – прогнозируемое значение;

*Т* – тренд;

*S* – сезонная компонента;

*Е* – ошибка прогноза.

В формуле все значения указанны в денежных еденицах.

Мультипликативные модели применяются, когда во временных рядах наблюдается значение сезонной компоненты которое являются частью трендового значения. Такие модели имеют вид:

(#)

Аддитивную модель отличить от мультипликативной возможно ориентируясь на величину сезонной вариации. Для Аддитивной модели характерна почти постоянная сезонная вариация, в то время как для мультипликативной модели сезонная вариация возрастает или убывает, на графиках это выражается в изменении амплитуды колебания сезонного фактора, пример изображен нарисунке 1.



Рисунок 1 − Аддитивная и мультипликативные модели прогнозирования

**3 ОПРЕДИЛЕНИЕ ТРЕНДА**

Существует множество различных взаимозаменяемых методов определения тренда. Наиболее часто встречаимые и рекомендуемые в различных литературных источниках это логарифмический, линейный, степенной, экспонинцеальный и полиномиальный методы. Задача - выбрать метод который дает наиболее точную оценку. Выбор метода в итоге, по результатам расчета, даст наиболее точный прогноз. Прежде чем определиться с выбором метода, был изучен и опробован каждый из выше перечисленных.  
 **3.1 Логарифмический тренд**

Логарифмического тренд имеет следующие свойства:

• когда b>0, уровни тренда растут с замедлением

• когда b<0, уровни тренда уменьшаются с замедлением.

• абсолютные изменения уровней по модулю всегда уменьшаются со временем.

• ускорения абсолютных изменений обозначаются знаком противоположности самим абсолютным изменениям, в то время как по модулю постепенно уменьшаются.

• цепные темпы изменения при t>? плавно приближаются к 100 % .

Напрашивается вывод, о том что логарифмический тренд, как и описанный ниже гиперболический тренд, отражает постепенно затухающий процесс изменений. Различие заключается в том, что затухание по гиперболе происходит довольно быстро при приближении к конечному пределу, в случае с логарифмическом трендом затухающий процесс продолжается без ограничения значительно медленнее.

Система уравнений для определения коэффициентов уравнения регрессии имеет вид:

(#)

для функции вида:

(#)

(#)

При выполнении анализа рядов динамики большую роль играет выявление сезонных колебаний. Таким колебаниям свойственны практически постоянные изменения уровней ряда по внутригодовым периодам: месяцам, кварталам. Для того что бы выявить сезонные колебания используют специальные показатели – индексы сезонности (Is).

Для ряда внутригодовой динамики, в которой основная тенденция роста

незначительна (или она не наблюдается совсем), изучение сезонности базируется на методе постоянной средней: являющейся средней из всех рассматриваемых уровней. Один из способов получить индекс сезонности состоит в следующем: для каждого года отдельно рассчитывается средний уровень, после чего выполняется сопоставление(в процентах) уровня каждого месяца. Получаемое процентное отношение является индексом сезонности:

(#)

В большинстве случаев при выполнении моделирования рядов динамики с использованием полиномов или экспонециальной функции не получают необходимых результатов, это происходит по той пречине что эти ряды динамики содержат в себе периодические колебания вокруг общей тенденции.

В таких случаях как правило приходится использовать гармонический анализ.

Цель данного анализа заключается в нахождении и измерении периодических колебаний врядах динамики. Функцию, которая была заданна в каждой точке рассматриваемого интервала времени, представляют в виде бесконечного ряда синусоидальных и косинусоидальных функций.

Гармонический анализ - это операция выражения заданной периодической функции в виде ряда Фурье по гармоникам разных порядков.

Каждый член ряда - это слагаемое постоянной величины с функциями синусов и косинусов некоторого периода.

Аппроксимация динамики финансовых явлений по средствам ряда Фурье заключается в отборе таких гармонических колебаний, которые при наложении друг на друга отражают периодические колебания динамического ряда. Используя ряды Фурье представляют динамику явлений в виде некоторой функции во времени, в которой слагаемые расставлены по убыванию периодов:

Методом наименьших квадратов рассчитывают параметры уравнений:

На графиках изображены варианты зависимостей факторного признакаY от факторного Х, где Х – временной фактор.

По данным таблицы 1, был построен график который изображен на рисунке 1.

  
Рисунок 1 - Логарифмический тренд

Выполнив расчет коэффициента детерминации был получен результат.

**3.2 Линейный тренд**

Линейный тренд представляется как линейная зависимость анализируемой величины вида

где

y – исследуемая переменная (например, производительность);

x – значение, указывающее позицию (порядковый номер) года в периоде прогнозирования.

Когда выполняется линейная аппроксимация связи двух параметров, для того что бы найти эмпирические коэффициенты линейной функции, чаще всего используют метод наименьших квадратов. Суть этого метода заключается в том, что линейная функция «наивысшего соответствия» проходит через точки графика, которые соответствуют минимуму суммы квадратов отклонений измеряемого параметра. Имеет вид:

Уравнения расчитывающее параметры линейной парной регресии имеют следующий вид:

где n – объем совокупности (исследуемой).



Рисунок X − Построение тренда методом наименьших квадратов

Константы a и b или значение коэффициента при переменной Х а также свободного члена уравнения рассчитывается по следующей формуле:

а также:

В таблице # приведен пример вычисления линейного тренда по данным.

Таблица #. Вычисление линейного тренда

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| Исходные данные | | XY |  |  |
| Период времени X | Обьем продажи Y |
| 1 | 10 | 10 | 1 | 14,68 |
| 2 | 12 | 24 | 4 | 15,42 |
| 3 | 13 | 39 | 9 | 16,16 |
| 4 | 16 | 64 | 16 | 16,9 |
| 5 | 19 | 95 | 25 | 17,64 |
| 6 | 23 | 138 | 36 | 18,36 |
| 7 | 26 | 182 | 49 | 19,12 |
| 8 | 30 | 240 | 64 | 19,86 |
| 9 | 28 | 252 | 81 | 20,6 |
| 10 | 18 | 180 | 100 | 21,34 |
| 11 | 16 | 176 | 121 | 22,08 |
| 12 | 14 | 168 | 144 | 22,82 |
|  |  |  |  |  |
| Среднее 6,5 | 18,75 |  |  |  |

, ; , ; ;

= =;

= 18,75 − 0,74 \* 6,5 = 18,25 − 4,81 = 13,94

уравнение тренда:

= 13,94 + 0,74

Методы сглаживания колебаний. В случае когда наблюдаются большие расхождения между соседними значениями, тренд расчитанный методом регресии, сложно анализировать. В случае когда при построении прогноза ряд состоит из данных с высокой амплитудой колебаний между ближайшими значениями, выполняется сглаживание используя специальные методы. Существует довольно много таких методов самые известные среди них: метод экспоненциального сглаживания иметод скользящих средних.

Метод скользящих средних или МСС. МСС сглаживает ряд значений что позволяет выявить тренд. Для проведения расчетов данным методом используется среднее арифметическое значение определенного количества значений. Например трехточечное скользящее среднее, для это берутся первые три значения из списка например данные за январь февраль и март и считается среднее арифметическое значение которое ставится в центре диапазона то есть в феврале. Следующим выполняется сдвиг на один месяц и берется три следующих значения, то есть февраль, март, апрель, и считается среднее арифметическое значение уже между этими тремя значениями. По такому принципу обрабатывается весь ряд. Новые значения которые были получены в результате полной обработки ряда составляют новый ряд. Чем больше ряд содержит значений тем сильнее выполняется сглаживание. Пример построения тренда предоставлен в таблице # а так же изображен на рисунке #

Таблица 5.2 Расчет тренда методом трехточечного скользящего среднего

|  |  |  |
| --- | --- | --- |
| Месяц | Объем продаж, тыс. ед. | Трехточечное скользящее среднее объема продаж, тыс. ед. |
| Январь | 10 | (10 + 12 + 13) / 3 = 11,67 |
| Февраль | 12 | (12 + 13 + 16) / 3 = 13,67 |
| Март | 13 | (13 + 16 + 19) / 3 = 16 |
| Апрель | 16 | (16 + 19 + 23) / 3 = 19,33 |
| Май | 19 | (19 + 23 + 26) / 3 = 22,67 |
| Июнь | 23 | (23 + 26 + 30) / 3 = 26,33 |
| Июль | 26 | (23 + 26 + 30) / 3 = 26,33 |
| Август | 30 | (26 + 30 + 28) / 3 = 28 |
| Сентябрь | 28 | (30 + 28 + 18) / 3 = 25,33 |
| Октябрь | 18 | (28 + 18 + 16) / 3 = 20,67 |
| Ноябрь | 16 | (18 + 16 + 14) / 3 = 16 |
| Декабрь | 14 |  |

Динамика колебаний исходных значений, которые были полученны методом скользящего среднего изображен на рисунке #. Результаты сравнения графиков рядов исходных значений и трехточечных скользящих средних показывают что удается сгладить колебания. Чем больше точек содержится в диапозоне вычисления тем четче будт вырисовываться тренд. Стоит отметить что операция укрупнения диапозона ведет к уменьшению количества конечных значений что приводит к менее точному прогнозу.



Рис. 5.4.

На рисунке(#) Избражена динамика изменения обьема продаж: ряд №3 - исходные данные; ряд №4 - скользящее среднее; ряд №2 - экспоненциальное среднее; ряд №1 - тренд, построенный методом регрессии.

Метод экспоненциального сглаживания. Еще один подход к уменьшению разброса значений ряда. Такое значение ему было присвоено из-за того что каждое значение периодов, уходящих в прошлое, уменьшается на множитель то есть 1 – α. Для расчета каждого значения используется формуа:

St =aYt +(1−α)St−1,

где St – текущее сглаженное значение;

Yt – текущее значение временного ряда; St – 1 – предыдущее сглаженное значение; α – сглаживающая константа, 0 ≤ α ≤ 1.

Чем меньше значение константы α , тем менее оно чувствительно к изменениям тренда в данном временном ряду.

Посчитав коэффициент детерминации был получен результат R(квадрат) = 0,0166;

**3.4 Выбор наиболее точного расчета тренда**

По резултатам исследования и выполненых эксперементов был выбран полиномиальный тренд. Результаты показали что данный тренд в сравнении с таким трендами как степенной а так же логарифмический или экспоненциальный оказался наиболее эффективным. Так как он наиболее удовлетворительно аппроксимирует фактические значения, и имеет самы высокий коэффицент детерминации:

· логарифмический R2 = 0,0166;

· степенной R2 =0,0197;

· экспоненциальный R2 =8Е-05.

Полниномиальный же R2 = 0,7435, что позволяет в разы уменьшить вероятность ошибки при прогнозировании.