

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное автономное образовательное учреждение
высшего образования
«Южно-Уральский государственный университет
(национальный исследовательский университет)»
Институт естественных и точных наук
Кафедра математического и компьютерного моделирования

РАБОТА ПРОВЕРЕНА

Рецензент, доцент кафедры УМФ,
к.ф.-м.н., доцент

_____/И.И. Иванов

« ____ » _____ 2019 г.

ДОПУСТИТЬ К ЗАЩИТЕ

Заведующий кафедрой,
д.ф.-м.н., доцент

_____/С.А. Загребина

« ____ » _____ 2019 г.

АНАЛИЗ И ПРОГНОЗИРОВАНИЕ ЦЕН АКЦИЙ SAMSUNG
ELECTRONICS CO.

ВЫПУСКНАЯ КВАЛИФИКАЦИОННАЯ РАБОТА ЮУрГУ –
02.03.01.2019.35.ПЗ ВКР

Руководитель работы, доцент
кафедры, к.ф.-м.н., доцент

_____/М.А Сагадеева

« ____ » _____ 2019 г.

Автор работы,
студент группы ЕТ-411

_____/А.С. Клепиков

« ____ » _____ 2019 г.

Нормоконтролер,
доцент кафедры, к.ф.-м.н.

_____/А.А. Акимова

« ____ » _____ 2019 г.

АННОТАЦИЯ

Клепиков А.С. Анализ и прогнозирование цен акций Samsung Electronics Co. – Челябинск: ЮУрГУ, ЕТ-411, 22 с., 4 ил., 0 табл., библиогр. список – 3 наим., 0 прил.

Аннотация.

ОГЛАВЛЕНИЕ

ВЕДЕНИЕ	5
1. ОБЩАЯ ИНФОРМАЦИЯ ОБ АКЦИЯХ, МЕТОДЫ РАЗРАБОТКИ ПРИЛОЖЕНИЙ, ОБРАБОТКА ИНФОРМАЦИИ	6
1.1. Информация об акциях	6
1.1.1. Категории акций	6
1.1.2. Стоимость акций	7
1.2. Общие сведения о биржах	8
1.3. Выбор средств разработки	9
1.3.1. Язык C#, платформа .Net Framework	9
1.3.2. Библиотеки	10
1.3.3. Система контроля версий Git	10
1.4. Выводы по разделу	11
2. РАЗРАБОТКА ПРИЛОЖЕНИЯ	12
2.1. Написание алгоритма работы программы	12
2.1.1. Проектирование внутренней структуры, применение паттернов разработки	13
2.1.2. Разработка аналитических модулей	13
2.2. Разработка пользовательского интерфейса	14
2.3. Выводы по разделу	14
3. АНАЛИЗ ДАННЫХ	15
3.1. Обзор методов прогнозирования данных	15
3.1.1. Простое скользящее среднее	15
3.1.2. Авторегрессионное скользящее среднее	15
3.2. Метод гусеницы	16
3.2.1. Развертка одномерного ряда в многомерный	17
3.2.2. Анализ главных компонент: сингулярное разложение выборочной ковариационной матрицы	17
3.2.3. Отбор главных компонент	18
3.2.4. Восстановление одномерного ряда	19

3.3. Реализация аналитических модулей	20
3.3.1. Блок-схема аналитических модулей	20
3.3.2. Пример работы программы	21
3.4. Анализ стоимостей акций Samsung Electronics Co. за 2018 год	21
3.5. Выводы по разделу	21
СПИСОК ЛИТЕРАТУРЫ	22

ВВЕДЕНИЕ

1.1. Информация об акциях

1.1.1. Категории акций

Различают обыкновенные и привилегированные акции:

1. Обыкновенные акции дают право на участие в управлении обществом (обычно, 1 акция соответствует одному голосу на собрании акционеров, за исключением проведения кумулятивного голосования) и участвуют в распределении прибыли акционерного общества. Источником выплаты дивидендов по обыкновенным акциям является чистая прибыль общества. Размер дивидендов определяется советом директоров предприятия и рекомендуется общему собранию акционеров, которое может только уменьшить размер дивидендов относительно рекомендованного советом директоров. Распределение дивидендов между владельцами обыкновенных акций осуществляется пропорционально вложенным средствам (в зависимости от количества купленных акций).
2. Привилегированные акции могут вносить ограничения на участие в управлении, а также могут давать дополнительные права в управлении (не обязательно), но по сравнению с обыкновенными акциями имеют ряд преимуществ: возможность получения гарантированного дохода, первоочередное выделение прибыли на выплату дивидендов, первоочередное погашение стоимости акции при ликвидации акционерного общества. Дивиденды часто фиксированы в виде определённой доли от бухгалтерской чистой прибыли или в абсолютном денежном выражении. Дивиденды по привилегированным акциям могут выплачиваться как из прибыли, так и из других источников — в соответствии с уставом общества. Привилегированные акции делятся на:
 - Привилегированные имеют ряд привилегий в обмен на право голоса. У их собственника определена величина дохода в момент выпуска и размещения ценных бумаг. Определён размер ликвидационной стоимости. Приоритет при начислении этих выплат

по отношению к обыкновенным.

- Кумулятивные (накапливающие). Привилегии — те же. Сохраняется и накапливается обязательство по выплате дивидендов. Фиксированный срок накопления дивидендов. При невыплате дивидендов обладатели этого вида акций получают право голоса на период до выплаты дивидендов.

Аналог привилегированных акций — акция учредительская (англ. *founders share*) — акция, распространяемая среди учредителей акционерных компаний и дающая им некоторые преимущественные права. Держатели таких акций могут:

- иметь дополнительное количество голосов на собрании акционеров;
- пользоваться первоочередным правом на получение акций в случае их последующих эмиссий;
- играть главную роль в решении всех вопросов, связанных с деятельностью акционерных компаний.

По именованным акциям данные об их владельцах регистрируются в реестре акционерного общества. В соответствии с законодательством физические и юридические лица могут быть владельцами именованных акций.

Акции на предъявителя допускают их свободную куплю-продажу на вторичном рынке без необходимости перерегистрации владельца. Российское законодательство допускало выпуск акций на предъявителя до 2002 года. С 2003 года акции могут выпускаться только в форме именной ценной бумаги.

1.1.2. Стоимость акций

Различают разные виды стоимости акций:

1. Номинальная стоимость акции — это то, что указано на её лицевой стороне (иногда её называют нарицательной стоимостью). Общая величина уставного капитала равна общей сумме номиналов всех выпущенных акций.

Номинальная стоимость всех обыкновенных акций общества должна быть одинаковой.

Номинальная стоимость не обязана отражать реальную ценность акций. Однако она часто используется для ряда операций (оценка пошлин, комиссий, тарифов), особенно на неразвитом, малоликвидном фондовом рынке. Цена акций при первичном размещении не должна быть ниже номинальной стоимости.

2. Эмиссионная стоимость акции — стоимость акций при их первичном размещении, по которой её приобретает первый держатель. Обычно эмиссионная цена акции превышает номинальную стоимость или равна ей. Превышение эмиссионной цены над номинальной стоимостью называется эмиссионной выручкой, или эмиссионным доходом.
3. Балансовая стоимость акций — частное от деления стоимости чистых активов компании (балансовой стоимости компании) на количество выпущенных акций, находящихся в обращении. Если рыночная цена ниже балансовой, то это является основой для будущего биржевого роста цены. Обычно балансовую стоимость определяют при аудиторских проверках.
4. Рыночная стоимость акции — это цена, по которой акция продаётся и покупается на вторичном рынке. Рыночная цена (котировка, курс) обычно формируется на торгах на фондовой бирже и отражает баланс спроса и предложения на данные акции. Для формирования рыночной цены важное значение имеет уровень ликвидности фондового рынка. Косвенно, рыночная стоимость акций отражает ликвидационную стоимость активов и пассивов компании.

1.2. Общие сведения о биржах

Фондовая биржа — это финансовый институт, обеспечивающий регулярное функционирование организованного рынка ценных бумаг.

В каждой стране мира с рыночной экономикой есть национальная фондовая биржа. Это организованный рынок для торговли ценными бумагами: акциями, облигациями и другими финансовыми инструментами.

Для допуска к торгам на бирже акции должны пройти процедуру листинга или быть допущены к торгам без прохождения процедуры листинга.

Участие акции в торгах позволяет эмитенту привлечь самый дешёвый и

самый долгосрочный капитал, повысить стоимость компании, снизить стоимость заимствований, поднять свой престиж, осуществлять дополнительную рекламу через биржевые каналы и успешно размещать последующие выпуски.

В данной работе будут рассматриваться биржи NASDAQ (сокр. от англ. National Association of Securities Dealers Automated Quotation, читается как «Насдак» — Автоматизированные котировки Национальной ассоциации дилеров по ценным бумагам) — американская биржа, специализирующаяся на акциях высокотехнологичных компаний (производство электроники, программного обеспечения и т. п.) и корейская биржа (англ. Korea Exchange, KRX) — крупнейшая в мире биржа по объему сделок с деривативами.

1.3. Выбор средств разработки

Для разработки аналитического приложения, способного самостоятельно собирать данные с интернет-ресурса в режиме реального времени нужна развитая платформа с большим набором инструментов. Для этих целей как раз подходит платформа Microsoft .Net Framework.

1.3.1. Язык C#, платформа .Net Framework

На сегодняшний день C# и платформа .Net Framework являются одними из наиболее популярных инструментов у разработчиков по всему миру. На этой платформе можно реализовать совершенно разные решения: от простейших консольных программ до сложнейших научных библиотек и микросервисов в веб-стеке.

Язык проектировался как полностью объектно-ориентированный, и его компания-разработчик Microsoft всегда выделяет это свойство языка в качестве его главных преимуществ. Ещё одним преимуществом является сравнительно низкий порог входа, обеспечивая вместе с тем большую потенциальную мощность разрабатываемых приложений.

Ещё одним преимуществом языка C# и платформы .Net Framework является развитое сообщество. Этот факт выводит возможности встроенного пакетного менеджера NuGet на новый уровень: любой желающий может

разработать своё собственное решение задачи и загрузить его на сервер Microsoft в открытый доступ.

1.3.2. Библиотеки

В разработанном приложении использованы следующие библиотеки:

1. HtmlAgilityPack. Данный пакет позволяет получать html-код веб-страницы и предлагает удобные методы для парсинга DOM-дерева с помощью XPath-строки;

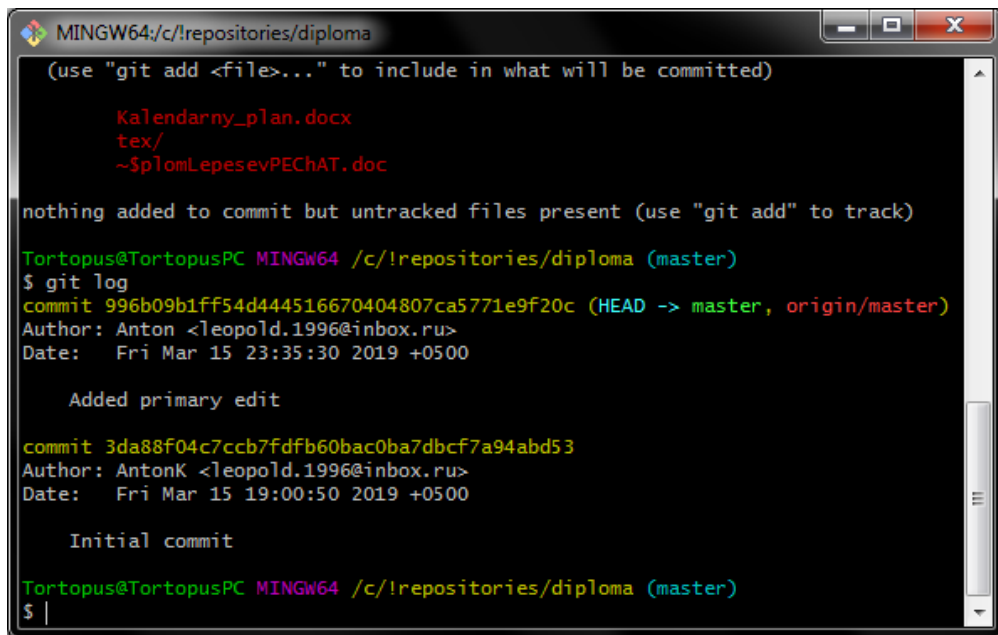
1.3.3. Система контроля версий Git

Система контроля версий — это система, записывающая изменения в файл или набор файлов в течение времени и позволяющая вернуться позже к определённой версии.

Ядро Git представляет собой набор команд для командной строки с параметром. Все настройки хранятся в текстовых файлах. Такая система позволяет сделать Git легко портируемым на любую платформу и легко интегрировать в любую систему.

Репозиторий Git представляет собой каталог на диске, содержащий файлы проекта и файлы конфигурации. Сам репозиторий загружается на удалённый сервер после вносимых изменений для возможности работы на разных устройствах.

Управление изменениями и репозиторием происходит посредством текстовых команд из командного окна: "git add "git commit "git push "git pull"и так далее. Пример представлен на 1.1.



```
MINGW64:/c:/repositories/diploma
(use "git add <file>..." to include in what will be committed)

Kalendarny_plan.docx
tex/
~$pomLepesevPEChAT.doc

nothing added to commit but untracked files present (use "git add" to track)

Tortopus@TortopusPC MINGW64 /c:/repositories/diploma (master)
$ git log
commit 996b09b1ff54d444516670404807ca5771e9f20c (HEAD -> master, origin/master)
Author: Anton <leopold.1996@inbox.ru>
Date:   Fri Mar 15 23:35:30 2019 +0500

    Added primary edit

commit 3da88f04c7ccb7Fdfb60bac0ba7dbcf7a94abd53
Author: AntonK <leopold.1996@inbox.ru>
Date:   Fri Mar 15 19:00:50 2019 +0500

    Initial commit

Tortopus@TortopusPC MINGW64 /c:/repositories/diploma (master)
$ |
```

Рисунок 1.1 – Командное окно Git Bash

1.4. Выводы по разделу

Из обзора видно, что необходима платформа, которая сможет совместить в себе возможность получения информации из интернет-ресурсов и мощные аналитические инструменты для работы с собранными данными.

В данном случае будет разработана программа, анализирующая временной ряд, образованный изменениями цен акций "Samsung Electronics Co.". Но разработав такую программу, её можно применить на любые аналогичные данные.

В качестве платформы для разработки выберем язык C# на платформе .Net Framework, так как он отвечает всем необходимым требованиям.

РАЗРАБОТКА ПРИЛОЖЕНИЯ

2.1. Написание алгоритма работы программы

Для разработки приложения, которое будет самостоятельно собирать данные с удалённых источников и в последствии их анализировать, необходимо реализовать следующие возможности:

- сбор данных непосредственно с удалённой веб-страницы;
- сериализация собранных данных;
- асинхронное чтение собранных ранее данных и запись их в память;
- прогноз будущих значений;
- отображение полученных данных на пользовательской форме.

Приложение разработано под ОС Windows. Для запуска необходимо выполнить исполняемый файл. Основной алгоритм представлен на 2.1.



Рисунок 2.1 – Основной алгоритм программы

2.1.1. Проектирование внутренней структуры, применение паттернов разработки

Основной рабочей структурой в программе являются аналитические модули. Они организованы в соответствии с паттерном проектирования "фабричный метод". На языке диаграмм UML можно представить структуру следующим образом (2.2).

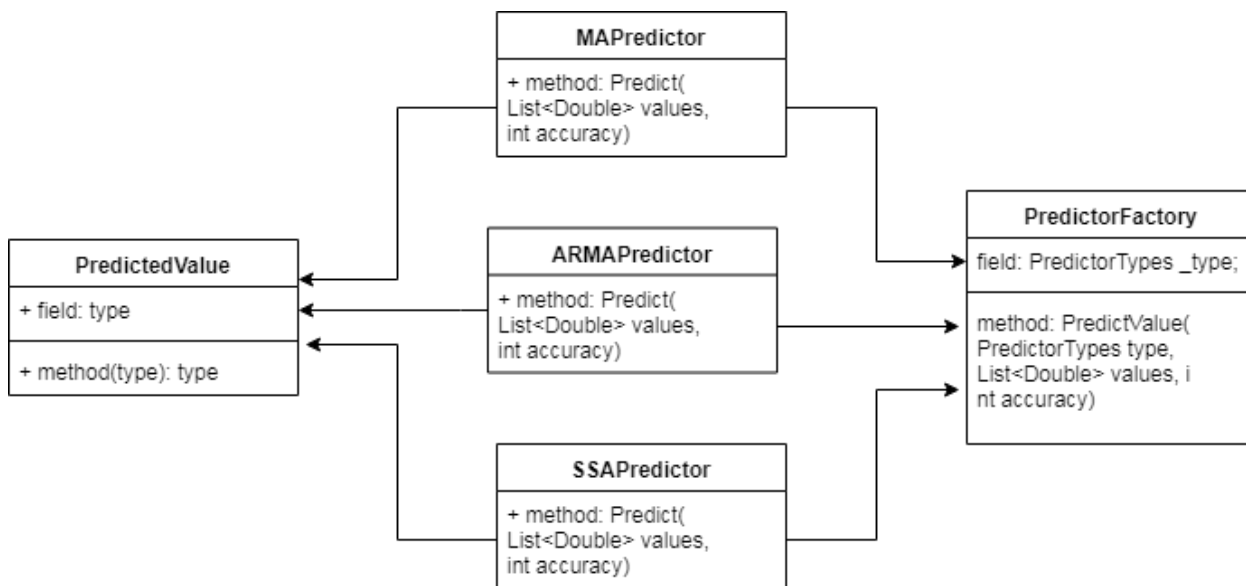


Рисунок 2.2 – Описание паттерна проектирования "фабричный метод"

2.1.2. Разработка аналитических модулей

В программе имеется три аналитических модуля, реализующие различные прогнозирующие методы:

- метод простого скользящего среднего;
- метод авторегрессионного скользящего среднего;
- метод гусеницы.

Доступ к модулям осуществляется из класса-сборщика PredictorFactory. На вход помещается временной ряд, требуемая точность прогнозирования и тип прогнозирующего метода.

Далее, в зависимости от типа, выбирается соответствующий метод, и ему в обработку передаётся временной ряд и заданная точность.

2.2. Разработка пользовательского интерфейса

Для базового управления процессом и вывода результатов пользователю используется технология WinForms.

2.3. Выводы по разделу

В этом разделе было разработано приложение, следящее за изменением цен на акции "Samsung Electronics Co. и анализирующее собранные результаты тремя разным методами прогнозирования.

АНАЛИЗ ДАННЫХ

3.1. Обзор методов прогнозирования данных

Технический анализ – наиболее распространенный вид анализа валютного рынка. Он дарит ощущение полного контроля в прогнозах торгующего человека. Суть его заключается в анализе прошлого, так трейдеры анализируют прошедшие котировки и уровни цены, и на основе этого делают выводы о будущем.

Методология состоит из анализа ценовых фигур, которые состоят из прошедших котировок валютной пары, анализа линий сопротивления/поддержки (линии, которые цене пройти наиболее тяжело), а также использования специальных математических индикаторов, которые сигнализируют о состоянии на валютном рынке.

Так как наше приложение свободно манипулирует большим количеством текущих и прошлых данных, этот вид анализа наиболее подходящий.

3.1.1. Простое скользящее среднее

Метод простого скользящего среднего (МА) основан на выборе некоторого числа последних элементов и подсчёте их среднего арифметического. Такое число считается следующим порядковым числом, идущим за выборкой. Модель скользящего среднего порядка q обозначается $MA(q)$ и записывается следующим образом:

$$X_t = \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \varepsilon_t,$$

где $\theta_1, \dots, \theta_q$ – параметры модели, а $\varepsilon_t, \dots, \varepsilon_{t-q}$ – ошибки.

3.1.2. Авторегрессионное скользящее среднее

Авторегрессионное скользящее среднее (ARMA), также называемая иногда моделью Бокса-Дженкинса, является объединением двух моделей исследования временных рядов: авторегрессионной (AR) и модели простого

скользящего среднего (МА). Под обозначением ARMA(p,q) понимается модель, содержащая p авторегрессионных составляющих и q скользящих средних. Точнее модель ARMA(p,q) включает в себя модели AR(p) и MA(q):

$$X_t = c + \sum_{i=1}^q \theta_i \varepsilon_{t-i} + \sum_{i=1}^p \phi_i X_{t-i},$$

где ϕ_1, \dots, ϕ_p — параметры авторегрессионной модели, а c — константа. Для простоты константу, как правило, опускают.

Обычно значения ошибки ε_t полагают независимыми одинаково распределёнными случайными величинами, взятыми из нормального распределения с нулевым средним: $\varepsilon_t \sim N(0, \sigma^2)$, где σ^2 — дисперсия.

3.2. Метод гусеницы

Метод гусеницы (SSA) в базовом варианте состоит в:

1. развертка одномерного ряда в многомерный;
2. анализ главных компонент: сингулярное разложение выборочной ковариационной матрицы;
3. отбор главных компонент;
4. восстановление одномерного ряда.

Таким образом, результатом применения метода является разложение временного ряда на простые компоненты. Полученное разложение может служить основой прогнозирования как самого ряда, так и его отдельных составляющих. "Гусеница" допускает естественное обобщение на многомерные временные ряды, а также на случай анализа изображений.

Рассмотрим временной ряд $x_{i=1}^N$, образованный последовательностью N равноотстоящих значений некоторой (возможно, случайной) функции $f(t)$:

$$x_i = f((i-1)\delta t),$$

где $i = 1, 2, \dots, N$.

Отметим, что существует несколько вариантов базового алгоритма, здесь мы опишем лишь минимальный.

3.2.1. Развертка одномерного ряда в многомерный

Выберем некоторое число $M < N$, называемое длиной гусеницы, и представим первые M значений последовательности f в качестве первой строки матрицы X . В качестве второй строки матрицы берем значения последовательности с x_2 по x_{M+1} . Последней строкой с номером $k = N - M + 1$ будут последние M элементов последовательности. Эту матрицу, элементы которой равны $x_{ij} = x_{i+j-1}$, можно рассматривать как M -мерную выборку объема k или M -мерный временной ряд, которому соответствует M -мерная траектория (ломаная в M -мерном пространстве из $k - 1$ звена). Отметим, что матрица X (будем называть ее матрицей ряда) представлена в традиционном для прикладной статистики виде "строка - индивид, столбец - признак". При изложении теоретических аспектов естественным является сопоставление индивиду столбца.

Далее по обычной схеме (за исключением стандартизации признаков) проводится анализ главных компонент (АГК).

3.2.2. Анализ главных компонент: сингулярное разложение выборочной ковариационной матрицы

Сначала вычисляется матрица $V = (1/k)X^T X$. Несмотря на то, что ее элементы не центрированы, мы будем называть ее ковариационной матрицей, иногда добавляя слово "нецентральная".

Следующий шаг, как обычно в АГК, состоит в вычислении собственных чисел и собственных векторов матрицы V , т.е. разложение ее $V = P\Lambda P^T$, где Λ - диагональная матрица, на диагонали которой стоят упорядоченные по убыванию собственные числа, а P - ортогональная матрица собственных векторов матрицы V .

Матрицы L и P совместно имеют множество интерпретаций, основанных на АГК. В частности, матрицу P можно рассматривать как матрицу перехода к главным компонентам $XP = Y = (y_1, y_2, \dots, y_M)$.

Если изучается выборка из случайной совокупности, то собственные числа матрицы V являются выборочными дисперсиями соответствующих главных компонент, а квадратные корни из них - выборочными стандарта-

ми. Графическое представление собственных чисел и некоторых функций от них в АГК традиционно используется для выявления структуры исследуемой совокупности и отбора и интерпретации главных компонент.

Заметим также, что при выборе длины гусеницы, равной $N - M + 1$, собственные вектора и главные компоненты (с точностью до нормировки) просто меняются местами.

3.2.3. *Отбор главных компонент*

В силу свойств матрицы P мы можем представить матрицу ряда X как $X = YP^T$. Таким образом, мы получаем разложение матрицы ряда по ортогональным составляющим (главным компонентам).

В то же время преобразование $y_j = Xp_j$ является линейным преобразованием исходного процесса с помощью дискретного оператора свертки, т.е.

$$y_j[l] = \sum_{q=1}^M x_{lq} p_{jq} = \sum_{q=1}^M x_{l+q-1} p_{jq}.$$

Таким образом, процедура "Гусеница" порождает набор линейных фильтров, настроенных на составляющие исходного процесса. При этом собственные векторы матрицы V выступают в роли переходных функций соответствующих фильтров.

Визуальное и аналитическое изучение как собственных векторов, так и главных компонент, полученных в результате линейной фильтрации, может дать много интересной информации о структуре изучаемого процесса и свойствах составляющих его слагаемых.

В частности, среди главных компонент можно выделить

1. относящиеся к тренду (медленно меняющиеся);
2. периодические;
3. шумовые.

Для нахождения периодических составляющих чрезвычайно большую визуальную информацию дает изучение двумерных графиков, аналогичных фигурам Лиссажу, когда по осям x и y откладываются различные пары собственных векторов или главных компонент. Известно, что, если по осям

откладывать значения синусоиды одной и той же частоты, но с разными фазами, то на плоскости получается эллипс. Из ортогональности собственных векторов и главных компонент следует, что сдвиг фаз между такими парами обязательно будет равен $\pm\pi/2$ и эллипс переходит в окружность.

3.2.4. Восстановление одномерного ряда

Следующим ключевым элементом метода "Гусеница" является процедура восстановления. Эта процедура основана на разложении $X = YP^T$. Будем говорить, что восстановление проводится по данному набору главным компонентам, если при применении формулы восстановления $X = YP^T$ матрица Y получена из матрицы Y обнулением всех не входящих в набор главных компонент. Таким образом, мы можем получить интересующее нас приближение матрицы ряда или интерпретируемую часть этой матрицы.

Переход к исходному ряду формально может быть осуществлен усреднением матрицы ряда по побочным диагоналям и может привести к некоторому искажению полученной структуры.

Отметим сразу одну очень важную особенность описанного метода - его интерактивность, то есть использование диалога исследователя и ЭВМ в процессе применения метода. Причем эта интерактивность не должна рассматриваться как следствие общей тенденции развития современного программного обеспечения персональных ЭВМ. Скорее, наоборот, эффективная реализация алгоритма стала возможной только благодаря возможностям современных ПК. Его интерактивность связана с типично статистическим свойством алгоритма - необходимостью интерпретации промежуточных результатов и управлением работой алгоритма в процессе многоэтапной процедуры обработки. Опыт многолетнего использования различных реализаций алгоритма на ЭВМ разных поколений показал, что чем больше промежуточных результатов удастся "увидеть тем более полным оказывается решение поставленной задачи.

Заметим, что при программной реализации этапы приведенного выше алгоритма не воспроизводятся непосредственно. Например, этап построения матрицы X отсутствует, а формулы преобразованы к виду удобному для проведения ускоренных вычислений.

3.3. Реализация аналитических модулей

Для реализации описанных выше методов был создан ряд классов, отражающих сущности операций и математических объектов.

3.3.1. Блок-схема аналитических модулей

На блок-схеме (3.1) можно увидеть процесс поступления на обработку данных, процесса выбора нужного метода прогнозирования и возврата прогнозируемого значения.

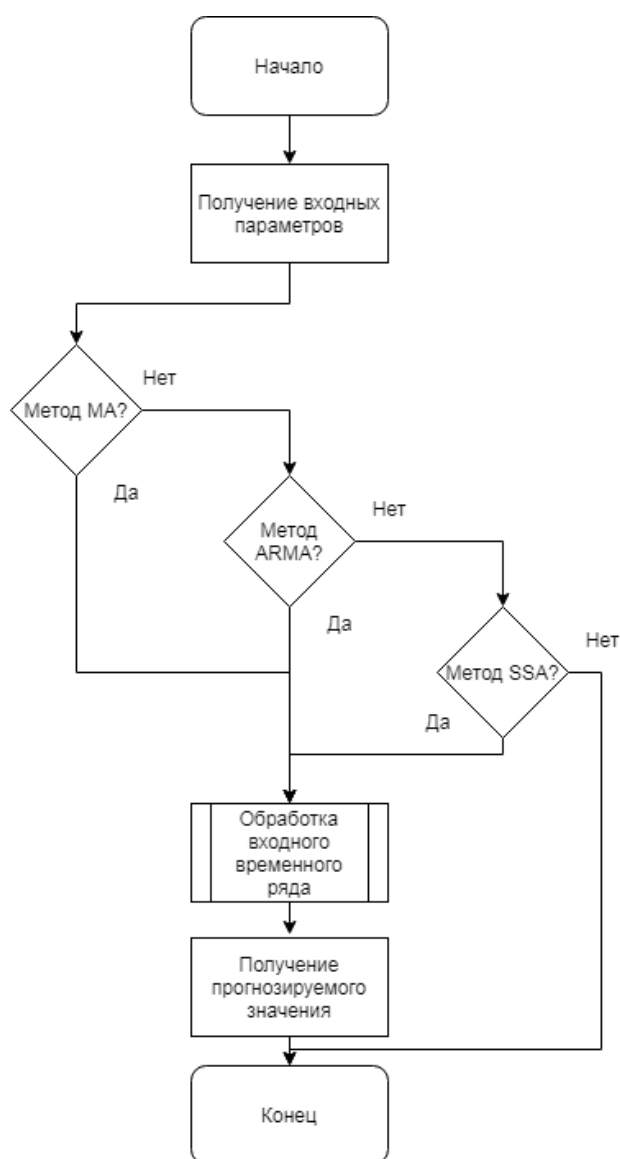


Рисунок 3.1 – "Блок-схема работы аналитических модулей"

3.3.2. Пример работы программы

todo

3.4. Анализ стоимостей акций Samsung Electronics Co. за 2018 год

todo

3.5. Выводы по разделу

СПИСОК ЛИТЕРАТУРЫ

1. GitHub.com, доступ: 2. <https://github.com/zzzprojects/html-agility-pack/tree/master/docs2/pages/documentations> (дата обращения – 7 июня 2019).
2. А. А. Жиглявский, В. Н. Солнцев: "Главные компоненты временных рядов: метод "Гусеница".
3. Тихомиров, Н. П. Эконометрика учеб. для вузов по специальности "Мат. методы в экономике" Н. П. Тихомиров, Е. Ю. Дорохина ; Рос. экон. акад. им. Г. В. Плеханова. - 2-е изд., стер. - М.: Экзамен, 2007. - 510 с. граф.