

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Тольяттинский государственный университет»

Институт математики, физики и информационных технологий

(наименование института полностью)

Кафедра «Прикладная математика и информатика»

(наименование кафедры)

02.03.03 Математическое обеспечение и администрирование
информационных систем

(код и наименование направления подготовки, специальности)

Технология программирование

(направленность (профиль)/специализация)

БАКАЛАВРСКАЯ РАБОТА

на тему Разработка модуля «Прогнозирование загрузки банкоматов»
комплексной системы автоматизации «СКАУТ – УКР/УИ»

Студент

И.Н. Мехмонов

(И.О. Фамилия)

_____ (личная подпись)

Руководитель

А.В. Очеповский

(И.О. Фамилия)

_____ (личная подпись)

Консультанты

Т.В. Маркелова

(И.О. Фамилия)

_____ (личная подпись)

Допустить к защите

Заведующий кафедрой к.т.н., доцент А.В. Очеповский

(ученая степень, звание, И.О. Фамилия)

_____ (личная подпись)

« _____ » _____ 20 _____ г.

АННОТАЦИЯ

Название бакалаврской работы – разработка модуля «Прогнозирование загрузки банкоматов» комплексной системы автоматизации «СКАУТ – УКР/УИ».

Объект работы: процесс хранения и манипулирования данными.

Предмет исследования: анализ эффективности банковской системы.

Цель работы: анализ автоматизированных банковских систем.

Для достижения необходимой цели нужно выполнить такие **задачи** как:

- анализ банковской системы;
- создание моделей Data Mining;
- разработка модуля, реализующий функции прогнозирования загрузки банкоматов.

В бакалаврской работе рассматривается процесс построения модуля прогнозирования для подсистемы планирования загрузки банкоматов, реализованный на базе Microsoft SQL Server 2012 и его компоненте Analysis Services с использованием алгоритмов интеллектуального анализа данных.

В работе освещены вопросы, связанные с организацией процесса прогнозирования загрузки банкоматов, способами хранения и аналитической обработки данных, рассмотрены технология интеллектуального анализа данных и сфера ее применения, алгоритмы интеллектуального анализа данных для решения задачи прогнозирования.

Бакалаврская работа содержит пояснительную записку объемом 49 страниц, включает в себя 15 рисунков, 2 таблиц и список использованной литературы, состоящий из 20 источника.

ABSTRACT

The title of the bachelor's work is "Development of the module" Forecasting the loading of ATMs "of the complex automation system" SCAUT-UKR / UI ".

Object of work: the process of storing and manipulating data.

Subject matter: banking system.

Objective: analysis of automated banking systems.

To achieve the necessary goal, you need to perform tasks such as:

- analysis of the banking system;
- creation of Data Mining models;
- analysis of execution of different algorithms for the created models

The title of the graduation work examines the process of building a forecasting module for the ATM load planning subsystem, implemented on the basis of Microsoft SQL Server 2012 and its Analysis Services component using data mining algorithms.

To date, ATM networks have become widespread. The number of users of plastic cards continues to grow, and, accordingly, the workload of the ATM network increases. Planning the process of collection allows you to solve these problems. To do this, it is necessary to monitor the cash flow in the ATM network and to predict the amount of cash to load ATMs.

The work deals with the organization of the process of forecasting the loading of ATMs, the methods of storage and analytical processing of data, the technology of data mining and its application, the algorithms of data mining for solving the prediction problem are considered.

The work contains an explanatory note on 49 pages, includes 15 figures, 2 tables and a list of used literature, consisting of 20 sources.

ОГЛАВЛЕНИЕ

ВВЕДЕНИЕ	6
ГЛАВА 1 АНАЛИЗ АВТОМАТИЗИРОВАННЫХ БАНКОВСКИХ СИСТЕМ..	
.....	8
1.1 Автоматизация деятельности коммерческого банка	8
1.2. Организация работы с банкоматами	9
1.3 Планирование загрузки банкоматов.....	10
1.3.1 Прогнозирование загрузки банкоматов.....	10
1.4 Способы хранения и аналитической обработки данных.....	13
1.4.1 Хранилища данных (Data Warehouse)	14
1.4.2 Оперативная аналитическая обработка данных (Online Analytical Processing (оперативная аналитическая обработка данных)).....	14
1.4.3 Интеллектуальный анализ данных (Data Mining)	16
1.5 Обзор существующих решений для управления загрузкой банкоматов	17
1.5.1 АТМ Cash Management Solution	17
1.5.2 СМ360: Оптимизация работы банкоматов	17
1.5.3 Управление оборотом наличных средств (Cash Management)	18
1.6 Постановка задачи для разработки модуля прогнозирования загрузки банкоматов.....	19
1.7 Вывод по главе 1.....	23
ГЛАВА 2 АНАЛИЗ ТРЕБОВАНИЙ И ВЫБОР АЛГОРИТМОВ ПРОГНОЗИРОВАНИЯ.....	24
2.1 Описание технологии инкассации банкоматов.....	24
2.2 Формирование требования банкомата.....	25
2.3 Технология аналитической обработки данных	26
2.4 Прогнозирование на основе алгоритма дерева принятия решений	28
2.5 Прогнозирование на основе алгоритма временных рядов.....	29
2.6 Вывод по главе 2.....	31
3.1 Выбор средств реализации Data Mining.....	33
3.2 Архитектура модуля прогнозирования менеджером.....	34

3.3 Проектирование Online Analytical Processing (оперативная аналитическая обработка данных) - куба.....	36
3.4 Построение модели прогнозирования на основе алгоритма временных рядов.....	39
3.5 Построение модели прогнозирования на основе алгоритма дерева принятия решений.....	42
3.6 Анализ результатов реализации модели - алгоритм дерева и алгоритм времени.....	44
3.7 Вывод по главе 3.....	46
ЗАКЛЮЧЕНИЕ	47
СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ.....	48

ВВЕДЕНИЕ

На сегодняшний день сети банкоматов получили широкое распространение. Число пользователей пластиковых карт продолжает расти, и, соответственно, увеличивается загруженность банкоматной сети. В связи с этим возникают следующие проблемы при организации работы сети банкоматов:

- увеличение объема информации, требующей обработки в процессинговых центрах;
- избыток отвлеченных денежных средств в банкоматах;
- внеплановые инкассации, которые могут возникнуть во время повышенного спроса на наличность или при единоразовом снятии клиентом со счета крупной суммы денег через банкомат;
- простои банкоматов, которые могут возникнуть как в загруженном состоянии, когда никто не пользуется данными банкоматами, так и при отсутствии в них денежных средств, например, в случае поломки.

Планирование процесса инкассации позволяет решить данные проблемы. Для этого необходимо отслеживать движение денежной наличности в банкоматной сети и прогнозировать суммы наличности для загрузки банкоматов.

Объект работы: процесс хранения и манипулирования данными.

Предмет исследования: анализ эффективности банковской системы.

Цель работы: анализ автоматизированных банковских систем.

Для достижения необходимой цели нужно выполнить такие **задачи** как:

- анализ банковской системы;
- создание моделей Data Mining;
- разработка модуля, реализующий функции прогнозирования загрузки банкоматов.

В первой главе приведена организация работы с банкоматами и анализирована автоматизированная банковская система.

Во второй главе произведён выбор алгоритмов и анализ требований прогнозирования.

В третьей главе приведена реализация моделей data mining и анализ результатов реализации модели - алгоритм дерева и алгоритм времени.

ГЛАВА 1 АНАЛИЗ АВТОМАТИЗИРОВАННЫХ БАНКОВСКИХ СИСТЕМ

1.1 Автоматизация деятельности коммерческого банка

Эффективное взаимодействие между всеми звеньями инфраструктуры коммерческого банка достигается путем внедрения. Автоматизированные банковские системы - программно-технологический комплекс, охватывающий набор взаимосвязанных автоматизированных банковских операций и задач [1].

Структура Автоматизированные банковские системы и ее компоненты представлены на рис. 1.1:

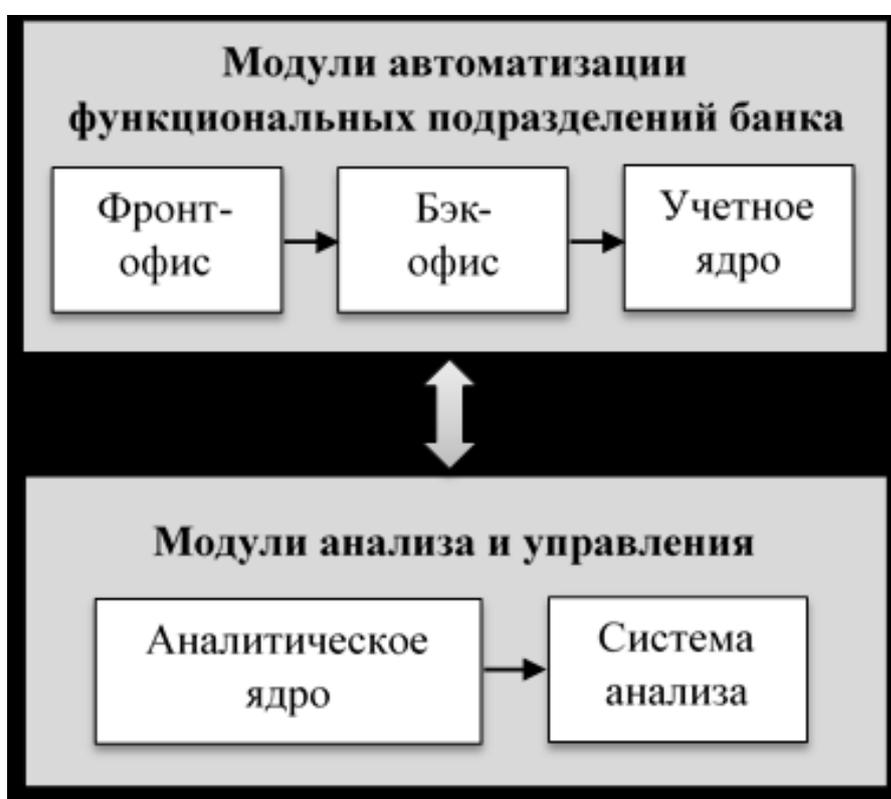


Рисунок 1.1- Структурная схема Автоматизированные банковские системы

Автоматизированные банковские системы строятся из совокупности функциональных модулей, которые соответствуют различным аспектам деятельности банка. Функциональные модули разделяются на три уровня: Фронт-офис, Бэк-офис и Учетное ядро системы [1].

Модули Фронт-офиса позволяют производить первичную обработку информации, которая регистрирует выполнение различных внутрибанковских

операций, а также обеспечивают внешнее взаимодействие банка с клиентами и партнерами. Бэк-офис – это совокупность автоматизированных рабочих мест сотрудников, обслуживающих совершенные в модулях Фронт-офиса банковские операции. В модулях учетного ядра организовано централизованное хранение общих для Бэк-офиса данных.

В отдельную подсистему выделяются аналитические модули, которые обеспечивают обработку и агрегирование информации, а также позволяют осуществить анализ текущего состояния банка, планирование и принятие решений. В данную подсистему может быть включен модуль прогнозирования загрузки банкоматов денежной наличностью.

1.2. Организация работы с банкоматами

Банкомат (АТМ - automated teller machine) - это многофункциональный программно-аппаратный комплекс, предназначенный для предоставления клиентам услуг по выпуску и инкассирования денежных средств во время транзакций с платежными карточками без присутствия персонала Банка [2].

Помимо выдачи наличности банкомат позволяет осуществлять внесение денежных средств на счет клиента, вывод состояния счета и перевод денежных средств на другой счет, формировать отчет о выполненных операциях за данный период, а также использовать различные информационно-справочные услуги.

Можно выделить следующие типы банкоматов:

1. Зарплатные. Основная масса денежных средств, которые клиенты снимают со счета с помощью данных банкоматов, приходится на дни получения зарплаты сотрудниками.

2. Уличные банкоматы. Банкоматы данного типа расположены в людных местах, например, в торговых центрах, метро, на улице.

3. Банкоматы, к которым ограничен доступ в зависимости от выходных и праздничных. Такие банкоматы установлены в офисах, торговых центрах, университетах и т.д.

По функциональным возможностям банкоматы можно разделить на однофункциональные, которые работают только на выдачу наличных по банковской карте, и многофункциональные, которые работают на выдачу и прием наличных, а также позволяют выполнять различные банковские операции. Кроме того, в отдельную категорию выделяются банкоматы с функцией замкнутого оборота наличности. С помощью таких банкоматов денежная наличность, внесённая одним клиентом, может быть получена другим клиентом.

На основе рассмотренной классификации можно производить наиболее точное прогнозирование загрузки банкоматов.

1.3 Планирование загрузки банкоматов

Общая схема планирования загрузки банкоматов состоит из следующих этапов:

1. Накопление данных по банкоматам (состояния банкоматов, данные о транзакциях).
2. Первичная обработка данных (обработка и агрегирование информации о состояниях банкоматов и транзакциях).
3. Прогнозирование спроса на денежную наличность в банкоматах с использованием методов аналитической обработки данных.
4. Планирование загрузки банкоматов на заданный период с учетом построенного прогноза.
5. Мониторинг текущего состояния банкоматов и оперативное оповещение о критическом состоянии.

1.3.1 Прогнозирование загрузки банкоматов

Прогнозирование является важным этапом процесса планирования загрузки банкоматов. Прогнозирование - это метод, который использует как прошлый опыт, так и текущие предположения о будущем для его определения.

Процесс прогнозирования можно представить в виде последовательности этапов (Рисунок 1.2):

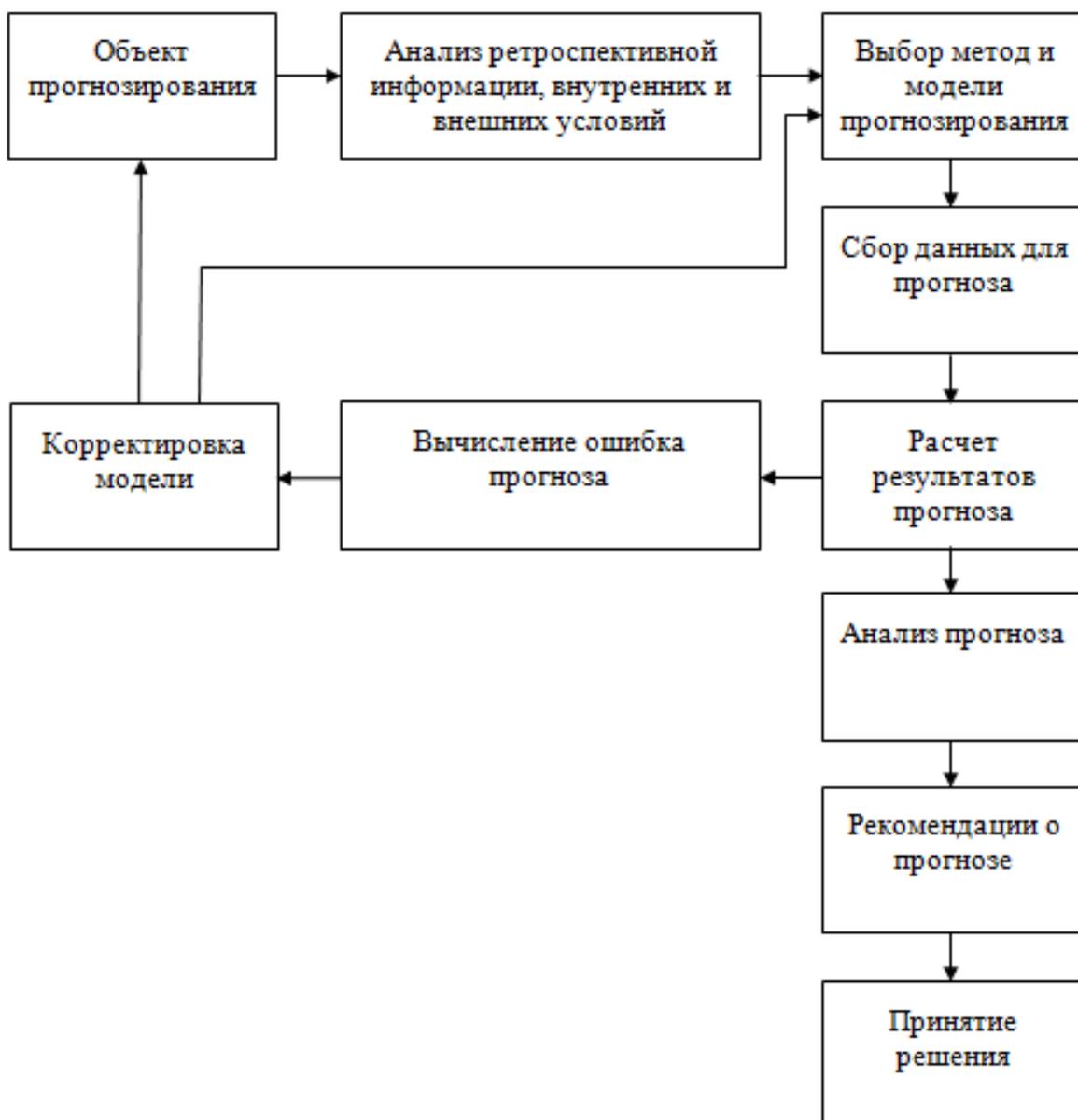


Рисунок 1.2 - Общая схема процесса прогнозирования

Для прогнозирования требуется использование определенных методов, которые представляют собой набор приёмов, позволяющих сделать заключения относительно будущего объекта на основе анализа исторических данных, а также с учётом внешних и внутренних условий.

Методы прогнозирования можно разбить на две группы [3]:

1. Формализованные методы, которые основаны на фактической информации о состоянии объекта прогнозирования (в прошлом и настоящем) и используют математический аппарат.

2. Интуитивные методы, которые используются тогда, когда невозможно использование математического аппарата ввиду сложности объекта прогнозирования. В данном случае применяются обработанные суждения экспертов.

Для построения математические модели прогнозирования, используются свойственные конкретной предметной области [3].

Классификация методов и моделей прогнозирования представлена на рисунке 1.3.

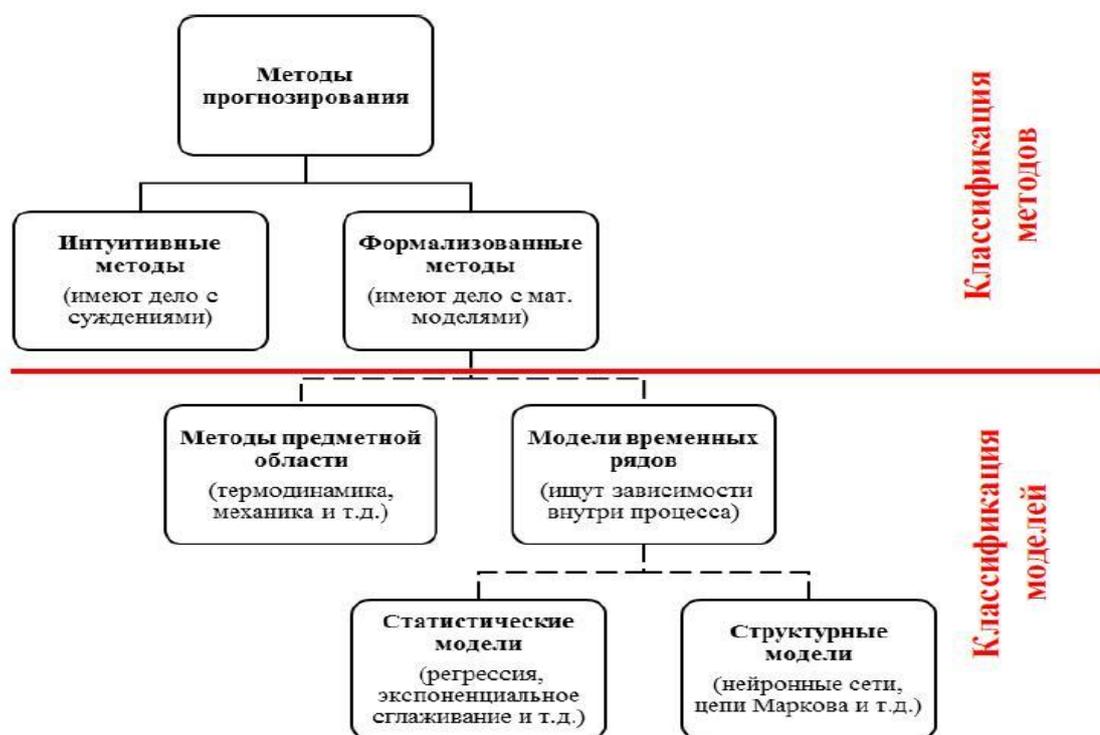


Рисунок 1.3 - Классификация методов и моделей прогнозирования

Ежедневный расход наличности в банкоматах можно рассматривать как временной ряд, поэтому для прогнозирования загрузки необходимо использовать модели временных рядов. Модели временных рядов — математические модели прогнозирования, которые используются для нахождения зависимости будущего значения от исторической информации о процессе и по этой зависимости вычислить прогноз. Эти модели являются универсальными для применения в различных областях. Они в свою очередь делятся на статистические и структурные.

В статистических моделях зависимость будущей ценности от прошлого дается в виде некоторого уравнения. В структурных моделях зависимость будущей ценности от прошлого дается в виде некоторой структуры и правил перехода.

При прогнозировании сумм загрузки банкоматов необходимо учитывать следующие факторы:

- неравномерность снятия наличных через банкомат в зависимости от дня недели;
- неравномерность снятия наличных через банкомат в зависимости от времени суток;
- местоположение банкомата;
- различные купюрные наборы в кассетах и режим работы сотрудников кассового центра;
- технические неисправности банкоматов;
- сезонность, выходные и праздничные дни.

На указанные выше факторы влияет непредсказуемость поведения клиентов.

1.4 Способы хранения и аналитической обработки данных

В процессе сбора информации по банкоматам накапливаются большие объемы информации. На основе исходных данных можно обнаружить скрытые закономерности, необходимые для прогнозирования сумм загрузки. Таким образом, при прогнозировании возникает проблема первичной подготовки данных.

Рассмотрим концепции хранения и анализа данных, получившие на сегодняшний день наибольшее распространение.

Чтобы существующие хранилища данных способствовали принятию управленческих решений, информация должна быть представлена аналитику в правильной форме, то есть он должен иметь разработанные инструменты для доступа и обработки хранилища данных.

1.4.1 Хранилища данных (Data Warehouse)

Хранилище данных - это комплексный, интегрированный по времени и неизменный набор данных, предназначенный для поддержки принятия решений [4].

Целью использования хранилище данных является анализ накопленных данных, т.е. структура данных должна быть организована так, чтобы они могли быть использованы для статистического анализа и формирования аналитических отчетов.

Выделяются следующие свойства хранилище данных [5]:

- предметная ориентация (организация данных в соответствии с назначенным видом деятельности организации);
 - интегрированность данных (агрегация данных из разных источников);
 - стабильность, инвариантность во времени (записи в хранилище данных являются отпечатками данных в определенное время);
 - минимальная избыточность информации (перед загрузкой в хранилище данных данные фильтруются и хранятся в определенной последовательности)
- хранилище данных и реляционная база данных имеют отличия.

Реляционные база данных используются пользователями для выполнения повседневной работы, в то время как хранилище данных предназначены для принятия решений. В реляционные база данных вносятся частые изменения, а данные в хранилище обычно обновляются по расписанию. Кроме того, реляционные база данных чаще всего являются источником данных для формирования хранилища.

1.4.2 Оперативная аналитическая обработка данных (Online Analytical Processing (оперативная аналитическая обработка данных))

Online Analytical Processing (оперативная аналитическая обработка данных) – технология комплексного многомерного анализа данных. Многомерное концептуальное представление представляет собой множественную перспективу, состоящую из нескольких независимых

измерений, вдоль которых могут быть проанализированы данные [6]. Многомерный анализ – это анализ данных одновременно по нескольким измерениям. Целью применения Online Analytical Processing (оперативная аналитическая обработка данных) является анализ данных и отображение результатов этого анализа в удобном для принятия решений виде.

Преимущества использования Online Analytical Processing (оперативная аналитическая обработка данных):

- наглядное представление анализируемых данных как функций от большого числа их характеризующих параметров;
- создание аналитических отчетов без участия программистов;
- быстрое построение аналитических отчетов благодаря предварительной агрегации данных;
- мощный аналитический и вычислительный аппарат: статистические функции, функции над множествами, сложные функции агрегирования.

Исходные и агрегированные данные могут храниться в реляционных и многомерных структурах. По этой причине многомерное представление данных бывает трех видов [7]:

1. Monline Analytical (Multidimensional Online Analytical Processing) - исходные и агрегированные данные хранятся в многомерной баз данных, что позволяет работать с данными как с многомерным массивом.

2. Ronline Analytical (Relational Online Analytical Processing) - исходные и агрегированные данные хранятся в реляционной база данных, при этом агрегированные данные помещаются в служебные таблицы.

3. Nonline Analytical Processing (оперативная аналитическая обработка данных) (Hybrid Online Analytical Processing (оперативная аналитическая обработка данных)) - исходные данные хранятся в исходной реляционной база данных, а агрегированные данные помещаются в многомерную базу данных.

Выбор многомерного представления данных зависит от объема и структуры исходных данных, требований к скорости выполнения запросов и

частоты обновления Online Analytical Processing (оперативная аналитическая обработка данных) - кубов.

1.4.3 Интеллектуальный анализ данных (Data Mining)

Data Mining представляет собой технологию анализа данных, цель которой – нахождение в данных скрытых шаблонов, неизвестных ранее знаний и их взаимосвязей, необходимых для принятия решений в сложных системах [7]. Отличие Data Mining от Online Analytical Processing (оперативная аналитическая обработка данных) заключается в том, что задача анализа и поиска знаний в данных переложена с человека на компьютер.

Data Mining позволяет решать следующие задачи:

1. Классификация - поиск признаков, которые отличают классы объектов и являются основной для выявления принадлежности объекта к определённому классу.

2. Регрессия - поиск зависимости определённого параметра (числового значения) по известным его характеристикам.

3. Прогнозирование - определение будущих значений (состояний) объекта на основе ретроспективных данных.

4. Кластеризация – разделение множества объектов на группы схожих по параметрам. Это более сложная задача, чем классификация, так как классы объектов изначально не известны.

5. Поиск ассоциативных правил - определение закономерности между взаимосвязанными событиями, происходящими одновременно.

6. Анализ последовательностей используется для обнаружения закономерностей в последовательностях событий.

7. Анализ отклонений используется для обнаружение данных, которые сильно отличаются от общего набора данных, и их анализа.

В интеллектуальном анализе данных используется математический анализ для определения моделей и тенденций в данных.

1.5 Обзор существующих решений для управления загрузкой банкоматов

1.5.1 ATM Cash Management Solution

Решение ATM Cash Management Solution от «СмартТехнологии» предназначено для прогнозирования загрузки банкоматов.

Возможности и особенности данного решения [8]:

- построение прогнозов загрузки банкоматов с высокой точностью;
- накопление и анализ исторических данных по каждому банкомату при прогнозировании;
- выявление дней пиковых нагрузок на банкоматы;
- учет стоимости инкассации и отвлеченных денежных средств при планировании загрузки;
- оптимизация размещения банкоматов;
- управление инкассацией и контроль инкассаторской службы;
- составление маршрутов для службы инкассации;
- принятия решения о загрузке банкоматов.

Данная система поддерживает любой тип банкоматов и позволяет работать с сетью банкоматов неограниченного размера. Кроме того, существует возможность прогнозирования сразу по нескольким валютам.

1.5.2 CM360: Оптимизация работы банкоматов

Данное решение от Industrial Data Systems предназначено для минимизации затрат на загрузку банкоматной сети. Система позволяет осуществлять загрузку банкоматов с учетом их местоположения, стоимости обслуживания, расписания работы инкассаторской службы, а также прогнозировать выход из строя значимых узлов сети банкоматов [9].

Решение основано на прогнозировании сумм загрузки банкоматов. Используются специальные математические алгоритмы, которые учитывают историю работы банкоматов и данные о прошлых загрузках, вычисляются оптимальные суммы для загрузки в будущем. Алгоритмы адаптируются и

обучаются в режиме реального времени. Применение алгоритмических композиций, методов отбора значимых факторов и методов вычленения значимой информации из накопленных исторических данных позволяет эффективно строить прогнозирующие модели высокого качества без участия эксперта.

Функциональные возможности:

- мониторинг необходимости в загрузке и среднесрочное планирование инкассаций;
- формирование рекомендаций по суммам загрузки;
- оформление заявок инкассаторской службе с учетом всех ограничений;
- оценка влияния внешних и внутренних факторов с помощью инструментария сценарного анализа;
- генерация настраиваемых автоматических сигналов для ситуаций, требующих реагирования;
- формирование отчетов по эффективности функционирования банкоматной сети.

1.5.3 Управление оборотом наличных средств (Cash Management)

«Энвижн Груп» предлагает систему, обеспечивающую современный подход к планированию денежных средств в банкоматной сети [10].

Данная система разработана на базе продуктов Oracle и состоит из несколько модулей, отвечающих за прогнозирование загрузки, планирования запасов, оптимизацию логистики и построение аналитических отчетов.

Функциональные возможности решения:

1. Прогнозирование спроса на денежную наличность на основе исторических данных и внешних факторов.
2. Определение плана инкассации по срокам и оптимальным суммам загрузки.
3. Расчет купюрной разбивки суммы подкрепления банкоматов на основании исторических данных о спросе и влияния внешних факторов.

4. Определение стоимости проведения инкассации.

Система разделена на три модуля и интерфейс пользователей:

- Oracle Demantra Management – модуль прогнозирования загрузки;
- Oracle Strategic Network Optimization – модуль планирования и оптимизация логистики;
- Oracle Business Intelligence – аналитический модуль;
- GUI – интерфейс пользователя.

1.6 Постановка задачи для разработки модуля прогнозирования загрузки банкоматов

Рассмотренные решения имеют закрытые алгоритмы, что не позволяет изучить методы, которые используются в них для анализа данных и прогнозирования. Это обуславливает сложность их внедрения в уже существующую систему в качестве отдельных аналитических модулей для управления загрузкой банкоматов.

В связи с этим необходимо разработать модуль для прогнозирования загрузки банкоматов.

Для решения поставленной задачи необходимо:

1. Провести рассмотрение процесса загрузки банкоматов и сформировать требования к модулю.
2. Осуществить выбор модели прогнозирования загрузки банкоматов.
3. Осуществить выбор показателей качества оценки результатов прогнозирования.
4. Осуществить выбор среды для реализации модуля.
5. Реализовать базовые алгоритмы модуля.

"СМ360: Оптимизация работы банкоматов" разработанный со всеми деталями бизнес-процессов управления денежными потоками и позволяющий решать широкий круг задач:

- платежные терминалы и сбор банкоматов с учетом их местонахождения (включая группы), стоимости обслуживания, возможностей сбора, наличия устройств и других условий контракта с услугой сбора;

- инкассация доп. офисов, обменных пунктов, отделений банка, клиентов – юридических лиц, депозитных сейфов;

- выхода прогнозирование из строя значимых узлов устройств самообслуживания.

"СМ360: Оптимизация работы банкоматов" реализован на платформе X360, позволяет оптимизировать потоки наличных денежных средств между денежными хранилищами, кассовыми узлами, депозитными сейфами, устройства самообслуживания позволяют сократить объем средств, хранящихся в устройствах, офисах и филиалах самообслуживания, в кассе, в результате, снизить стоимость страхования наличных средств и перевозимых сумм.

Учитываются возможности:

- хранение средств, планируемое прибытие наличных, стоимость обслуживания, время работы; поддержание минимально необходимых резервов наличности в кассовом хранилище с выбором оптимального источника пополнения (наличными); планирование в области управления потоками наличных средств с помощью сценарного управленческого анализа;

- согласование баланса денежных потоков и автоматическое обнаружение нарушений, требующих внимания;

- мониторинг потоков денежных средств в настоящем и мониторинг процессов управления наличностью с помощью аналитических инструментов.

"СМ360: Оптимизация работы банкоматов" основан на прогнозировании движения денежных средств. Алгоритмы адаптированы и обучены в современном режиме. Использование алгоритмических композиций, методов выбора значимых факторов и методов выделения содержательной информации из накопленных исторических данных позволяет эффективно строить прогностические модели высокого качества без участия эксперта.

Система в режиме реального времени создает оптимальные с точки зрения рекомендаций по минимизации затрат для реализации коллекций с учетом требований уровня обслуживания и всех бизнес-ограничений. Рекомендации включают в себя: оптимальную дату и время, суммы и комплекты сбора денежных средств. Наряду с этим система решает многие связанные задачи и предоставляет богатый инструмент для управления, мониторинга и анализа денежных потоков, планирования ремонтных работ устройств самообслуживания. Система управляется с помощью интуитивно понятного веб-интерфейса.

При проектировании "СМ360: Оптимизация работы банкоматов" учитывался богатый опыт ведущих банков.

Стратегия управления коллекциями осуществляется путем выбора наилучшего соотношения ключевых показателей, таких как уровень обслуживания и стоимость средств, а также поддержание собранных объектов. Система позволяет установить баланс ключевых индикаторов для всей сети, а также для отдельных групп собранных объектов. Например, вы можете установить повышенный уровень обслуживания для VIP-устройств. Для пользователей системы, которые должны выполнять сбор, анализируется аналитическая информация, которая позволяет сделать процесс формирования рекомендаций максимально доступным. Аналитик находит возможность не только проверить правильность предлагаемой рекомендации, но и получить подробную информацию о причинах ее формирования.

"СМ360: Оптимизация банкоматов" дает возможность централизованно и эффективно управлять графиком инкассации. Когда он используется, затраты на обслуживание банкоматов и нехватку средств значительно сокращаются.

С помощью "СМ360: Оптимизация работы банкоматов" Вы имеете возможность выбрать наилучшую стратегию управления инкассациями: оптимальное соотношение эффект обслуживания клиентов и затрат на содержание сети банкоматов.

Повседневные решения о том, какие устройства, какие суммы наличными, принимаются на централизованной основе, при условии согласованной политики управления устройствами (ограничения на загрузку). Это приводит к тому, что количество сотрудников уменьшается, которые вовлечены в процессы мониторинга и сбора баланса, что, в свою очередь, повышает операционные риски и затраты на рабочую силу.

Процесс сбора оказывается более очевидным, при помощи Системы вы можете вычислять текущие затраты на обслуживание и обслуживание устройств и уровень обслуживания всей сети, также отдельных групп устройств. СМ360 даёт вам возможность понять, каковы наименьшие затраты на обслуживание банкоматов на различных уровнях обслуживания.

Основные задачи, выполняемые с помощью Системы:

- поддерживать принятие решений, выбирая стратегию управления инкассацией;
- оперативный мониторинг необходимости сбора наличных средств и планирования среднесрочного сбора наличных средств;
- составление рекомендаций по сумме и разбивке банкнот собранных средств и обработке заявок на услуги по сбору платежей с учетом всех ограничений для бизнеса;
- управление денежными средствами в денежной и кассовой единице;
- оценка влияния внешних и внутренних факторов посредством инструментария сценарного анализа;
- генерация настраиваемых автоматических сигналов для ситуаций, вызывающие реагирования;
- приготовление отчетов по динамике степени качества управления инкассациями и качества работы сети устройств самообслуживания.

Рекомендации, выполняемые Системой, могут быть дополнительно поправлены вашими специалистами на основании их личного опыта и внешних факторов, не проверяемые Системой.

Прогнозы, добавленные пользователем, содержатся вместе с системными прогнозами, что даёт возможность сравнивать их продуктивность друг с другом, а также устанавливать значение влияния корректировок, которые пользователи сделали в запросах на сбор, рекомендованных система.

Все результаты системы хранятся в хранилище данных и открываются для дальнейшего анализа.

1.7 Вывод по главе 1

В данной главе анализирована автоматизированная банковская система и приведена организация работы с банкоматами разных типов (зарплатные, уличные банкоматы и банкоматы, к которым ограничен доступ в зависимости от выходных и праздничных). После планирование загрузки банкоматов и их прогнозирование с помощью определённых методов приведены способы хранения и аналитической обработки данных.

Произведена постановка задачи для разработки модуля прогнозирования загрузки банкоматов.

ГЛАВА 2 АНАЛИЗ ТРЕБОВАНИЙ И ВЫБОР АЛГОРИТМОВ ПРОГНОЗИРОВАНИЯ

2.1 Описание технологии инкассации банкоматов

Для формирования требований к модулю нужно анализировать процесс при подкреплении банкоматной сети денежной наличностью, который включает в себя прогнозирование загрузки банкоматов.

Прогноз потребность на денежную наличность можно создать, используя исторические данные по состояниям банкоматов и транзакциях. На основе полученного прогноза образовывается по срокам инкассации и суммам денежной наличности для загрузки.

Процесс подготовки и проведения инкассации банкоматов изображён на рисунок 2.1.



Рисунок 2.1 - Схема проведения инкассации банкоматов

1. В процессинговом центре выполняется сбережения данных по банкоматам (состояния банкоматов, данные о транзакциях).

2. Данные из процессингового центра поступают на первичную обработку (агрегирование информации о состояниях банкоматов и транзакциях).

3. Формируется прогноз спроса на денежные средства в банкоматах с использованием методов аналитической обработки данных.

4. На основе полученного прогноза планируется загрузка банкоматов на заданный период с учетом:

- стоимости обслуживания банкоматов;
- режима работы денежного хранилища банка и инкассаторских служб;
- состояния банкоматов и их конфигурации;
- местоположения банкоматов.

5. Формируется заявка службе инкассации на обслуживание банкоматов.

6. Служба инкассации получает чек о состоянии счетчиков согласно данным процессингового центра.

7. Формируется заявка кассовому центру на подготовку кассет с денежной наличностью.

8. Кассовый центр формирует кассеты с денежной наличностью, одновременно с этим служба инкассации формирует группу инкассаторов на определённый маршрут.

9. В ходе проведения инкассации банкомата изымаются кассеты с остатками наличности и загружаются новые кассеты, вводятся данные о количестве и номинале купюр в новых кассетах, осуществляется печать чеков выгрузки и загрузки кассет банкомата.

2.2 Формирование требования банкомата

На основе рассмотренного процесса инкассации были сформулированы следующие требования к модулю прогнозирования загрузки банкоматов:

1. Учёт состояния банкомата:

- выбор банкоматов, требующих загрузки.

2. Используемые данные:

- дата;
- месяц;

- часть месяца;
- день месяца;
- день недели;
- исторические данные по суммам загрузки.

3. Прогнозируемое значение:

- общая сумма денежной наличности для загрузки банкомата.

4. Представление информации по банкоматам:

- ежедневное отражение данных;
- общая сумма прогноза денежной наличности;

5. Функционирование в рамках подсистемы планирования инкассации банкоматов.

2.3 Технология аналитической обработки данных

Средством представления и многомерного анализа данных является Online Analytical Processing (оперативная аналитическая обработка данных) - куб.

Проектирование Online Analytical Processing - куба проводится на основе атрибутов таблиц хранилища данных в следующей последовательности [5]:

1. Определение и согласование измерений. Измерения являются критериями для анализа и определяют многомерное пространство Online Analytical Processing -куба. За счет фиксирования значений одного или нескольких измерений строятся срезы. Каждый срез представляет собой запрос к данным, включающий агрегации.

2. Выбор набора мер. Меры — это атрибуты измерения Online Analytical Processing - куба, сопоставляемые с атрибутами таблиц хранилища данных.

3. Сохранение результатов предварительных расчетов в таблице фактов;

В кубе находятся отношения обобщения и специализации (roll-up/drill-down) по иерархиям измерений. Ячейка высшей степени иерархии может «спускаться» (drill-down) к ячейке нижней степени и в обратном порядке, «подняться» (roll-up).

Одним из основных этапов построения Online Analytical Processing – куба является задание функций агрегирования. Так как цель Online Analytical Processing – реализация многомерной структуры данных, данные на уровнях, отличных от фактического, должны быть заранее агрегированы. Для каждого измерения есть вероятность задание собственной функции агрегации. Поэтому, в случае куба с n измерениями функция агрегации:

$$f(x) = (\{f_{1,1}, \dots, f_{1,k_1}\}, \dots, \{f_{n,1}, \dots, f_{n,k_n}\}) \quad (2.1)$$

где x – точка куба, а $f_{i,j}$ – j -ая функция агрегирования по i -ому измерению.

Для визуализации данных, хранящихся в Online Analytical Processing – кубе, применяются табличные представления, которые имеют сложные иерархические заголовки строк и столбцов.

Общую схему, отражающую принцип многомерного анализа данных, можно представить в следующем виде (рисунок 2.3):

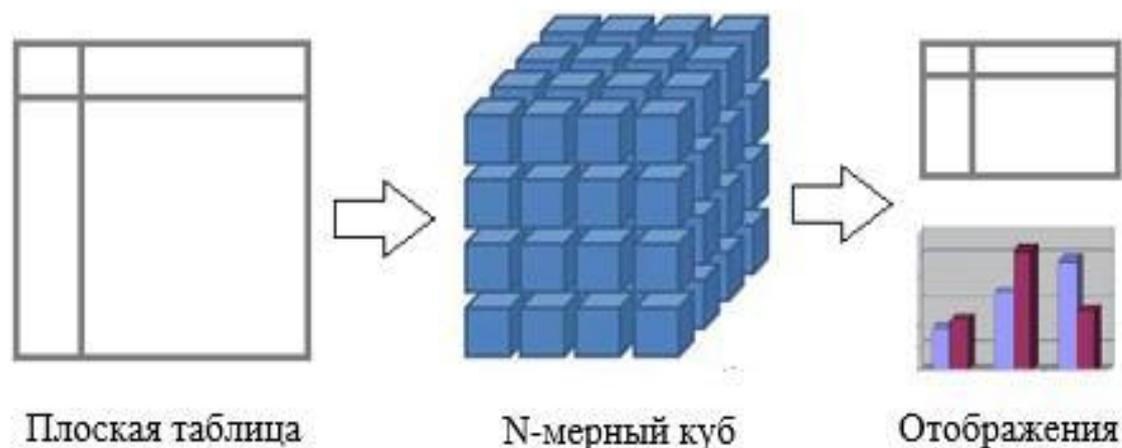


Рисунок 2.3 - Схема многомерного анализа данных

Она реализует следующий алгоритм:

1. Построение хранилища данных путем представления исходных данных в виде плоской таблицы или результата именованного запроса.
2. Извлечение и преобразование информации из хранилища данных в куб.
3. Отображение построенного куба при помощи кросс-таблиц, диаграмм, графиков.

В данной работе используется способ обработки данных, основанный на совместном использовании Online Analytical Processing и технологии Data Mining.

Одними из направлений Data Mining являются поиск в большом наборе данных скрытых тенденций и решение задач прогнозирования. Процесс Data Mining состоит из следующих этапов [5]:

1. Создание структуры и модели с помощью определенного алгоритма.
2. Обучение модели с использованием тренировочных данных (в тренировочных данных указываются исходные атрибуты и атрибуты, которые будут предсказываться).
3. Подача на вход модели data mining неизвестных предсказываемых атрибутов с последующим определением их значений.

2.4 Прогнозирование на основе алгоритма дерева принятия решений

Алгоритм дерева принятия решений используется для прогнозирования атрибутов путем установления зависимости прогнозируемого атрибута от независимых входных переменных.

Принцип построения дерева следующий. Дерево строится сверху вниз от корня. В первую очередь необходимо выявить атрибут для проверки в корне дерева. Для этого необходимо определить, насколько хорошо каждый атрибут разделяет данные на группы в зависимости от прогнозируемого атрибута. Далее для каждого значения выбранного атрибута создается ветвь дерева, и набор данных разделяется дальше в соответствии со значением. Данный процесс повторяется рекурсивно для каждой ветви. Также осуществляется проверка критерия остановки. Когда этот рекурсивный процесс заканчивается, формируется дерево решений.

Исходный набор данных имеет следующий вид:

$$(x, Y) = (x_1, x_2, \dots, x_k, Y), \quad (2.2)$$

Зависимая переменная Y является целевой непрерывной переменной, значения которой необходимо прогнозировать. Вектор x состоит из входных

переменных x_1, x_2, x_3 и т. д., которые используются для выполнения этой задачи.

Непрерывные переменные могут быть спрогнозированы при помощи формул линейной регрессии, основанных на указанных регрессорах. Регрессор - это непрерывный входной атрибут, который используется для моделирования непрерывного прогнозируемого атрибута линейным образом.

Классическая формула линейной регрессии:

$$Y = a + b * X + e \quad (2.3)$$

где Y – непрерывный прогнозируемый атрибут;

X – регрессор;

a, b – коэффициенты регрессии;

e – случайная ошибка.

Рассмотренный алгоритм в каждом узле дерева содержит формулу линейной регрессии [13].

2.5 Прогнозирование на основе алгоритма временных рядов

Для прогнозирования загрузки банкоматов можно использовать исторические данные, а на их основе оценивать будущий спрос на денежную наличность в банкомате для выбора оптимальной суммы загрузки. Рассмотренному подходу совпадают алгоритм Data Mining временных рядов (Time-Series Data Mining).

Алгоритм временных рядов предназначен для прогнозирования непрерывных значений. В отличие от алгоритма дерева принятия решений, модель временных рядов не требует дополнительных сведений для прогнозирования. Построение прогноза осуществляется с помощью закономерностей, извлекаемых алгоритмом из исходных данных.

Временной ряд – это ряд последовательных значений, характеризующих изменения каких-либо параметров исследуемого процесса во времени. Временные ряды обладают некоторыми важными характеристиками [12]:

- основной уровень, определяющее среднее значение временного ряда;
- тренд, который отображает, как временные ряды изменяются от одного периода к другому;
- сезонные колебания. Некоторые значения имеют тенденцию повышения или уменьшения в зависимости от разных периодов времени, например, от дня месяца или года.

Таким образом, назначая тренд, накладывая линию тренда на базовый уровень и выявляя сезонную составляющую, можно получить модель для составления прогноза значений: Значение прогноза + Уровень базовый + Трендовый показатель + Составляющая сезона.

Для анализа временных рядов можно использовать совокупность двух алгоритмов:

1. «Авторегрессионное дерево с перекрестным прогнозированием» (ARTXP), который предназначен для использования в краткосрочном прогнозировании и обеспечивает наилучшие значения прогноза ближайшего временного отрезка за последним фактическим.

2. «Интегрированные скользящие средние авторегрессии» (ARIMA), который предназначен для использования в краткосрочном прогнозировании.

Алгоритм ARTXP существенно менее точен в долгосрочных прогнозах по сравнению с ARIMA.

ARIMA (или, Модель авторегрессии и интегрированного скользящего среднего) модель также называется моделью Бокса-Дженкинса и расшифровывается, как модель авторегрессии и интегрированного скользящего среднего.

В обозначениях Бокса и Дженкинса модель записывается как ARIMA(p, d, q):

$$\Delta^d Y_t = c + \sum_{i=1}^p \alpha_i \Delta^d Y_{t-1} + \sum_{i=1}^q b_i \varepsilon_{t-1} + \varepsilon_t, \quad (2.4)$$

ARIMA (p,d,q) описывает три основных процесса динамического ряда:

1. AR (параметр p) — авторегрессия.

Модель авторегрессии – это модель временных рядов, в которой значения временного ряда в данный момент линейно зависят от предыдущих значений этого же ряда. Авторегрессионный процесс порядка p определяется следующим образом:

$$Y_t = \sum_{i=1}^p a_i Y_{t-i} + \varepsilon_t, \quad (2.5)$$

a_1, a_2, \dots, a_p – коэффициенты авторегрессии; ε_t – случайная ошибка.

2. Параметр d — число разностей, необходимых для приведения процесса к стационарности. Модель ARIMA требует, чтобы серия была стационарной (средняя постоянная и дисперсия выборки и автокорреляция со временем не изменяются). Поэтому обычно необходимо принять разность рядов до тех пор, пока он не станет стационарным (часто для стабилизации дисперсии используется также логарифмическое преобразование).

3. MA (параметр q) — скользящее среднее. В отличие от процесса авторегрессии, в процессе скользящего среднего каждый элемент ряда подвержен суммарному воздействию предыдущих ошибок. Модель скользящего среднего q -го порядка — модель временного ряда следующего вида:

$$\sum_{i=1}^q b_i \varepsilon_{t-i} \quad (2.6)$$

где b_1, b_2, \dots, b_q – параметры модели; ε_t – случайная ошибка.

ARIMA модели охватывают широкий спектр временных рядов, и при небольшой модификации этих моделей позволяют достаточно точно описывать временные ряды с сезонностью.

2.6 Вывод по главе 2

В этой главе произведён анализ требований и выбор алгоритмов прогнозирования, где приведена процесс подготовки и проведения инкассации

банкоматов, также сформулированы требования к модулю прогнозирования их загрузки. Для прогнозирования атрибутов путем установления зависимости прогнозируемого атрибута от независимых входных переменных используется алгоритм дерева принятия решений, а также прогнозирование производится на основе алгоритма временных рядов. Основой современного подхода к построению аналитических систем является идея интегрированного хранилища данных, обеспечивающая единый логический вид и доступ к информации, разбросанной по различным операционным системам организации и поступающей из внешних источников. В то же время важно, чтобы данные в хранилище имели исторический характер, то есть включали не только гетерогенные источники, но и архивные данные, возникающие в ходе работы конкретной операционной системы.

ГЛАВА 3 РЕАЛИЗАЦИЯ МОДЕЛЕЙ DATA MINING

3.1 Выбор средств реализации Data Mining

Необходимо выбрать средство реализации, которое предоставляет следующие возможности:

1. Управление источниками внешних данных и их представлениями;
2. Предварительное агрегирование данных;
3. Проектирование многомерных Online Analytical Processing - кубов для проведения анализа;
4. Построение модели Data Mining для прогнозирования временных рядов;
5. Наличие библиотек для доступа к Online Analytical Processing - кубам и моделям Data Mining из приложений.

Заданным требованиям отвечает MS SQL Server, который предлагает службы Analysis Services для построения многомерных структур данных. Службы Analysis Services позволяют строить модели на основе Online Analytical Processing - куба или исходной таблицы, развертывать модели в виде базы данных на экземпляре службы Analysis Services, обрабатывать базы данных для загрузки в них данных. Доступ к полученной модели данных может быть предоставлен любым клиентским приложениям, которые поддерживают Analysis Services в качестве источника данных.

MS SQL Server Analysis Services включает среду разработки Business Intelligence Development Studio, которая позволяет создавать различные решения с использованием многомерных Online Analytical Processing - кубов и моделей Data Mining. Она позволяет настраивать источники данных и представления источников данных, создавать и обрабатывать структуры и модели Data Mining, а также строить прогнозирующие запросы при помощи языка Data Mining Extensions (DMX), просматривать и сохранять результаты.

Data Mining Extensions возможности интеллектуального анализа данных (DMX) - это язык, который можно использовать для создания и работы с

моделями интеллектуального анализа данных в Microsoft SQL Server Analysis Services.

На следующем этапе разработки необходимо осуществить проектирование Online Analytical Processing - кубов и тестирование моделей Data Mining на реальных данных. Так как к началу этапа разработки отсутствовали реальные данные по банкоматам, было принято решение о работе с реальными данными задачи из сферы ЖКХ.

3.2 Архитектура модуля прогнозирования менеджером

Одноуровневая архитектура. Все сетевые узлы связаны друг с другом одним Менеджером.

Протокол, взаимодействующий с интегрирующим компонентом на лексическом. Данная архитектура реализует основополагающие функции управления, которые являются общими для разных периферийных устройств.

Это архитектура неэффективна для интегрированной корпоративной сети, из-за большого количества специализированного сетевого оборудования и разнотипных протоколов связи.

Иерархическая архитектура. Эта архитектура основана на иерархии диспетчера. Каждый сегмент корпоративной сети имеет контрольную станцию для мониторинга состояния сетевых устройств в сегменте.

С помощью элементарного сетевого менеджера, установленного администратором на административной станции, администратор управляет сетевыми устройствами. Эти станции управления взаимодействуют с корпоративными станциями управления, которые, в свою очередь, связаны с интеграцией станций управления. Поэтому управляющий директор иерархии станций нижнего уровня может быть как агентом, так и менеджером. Для разработки модуля прогнозирования был использован комплексный подход по выполнению технологий Online Analytical Processing (оперативная аналитическая обработка данных) и Data Mining. Архитектура модуля прогнозирования изображена на рисунок 3.1.

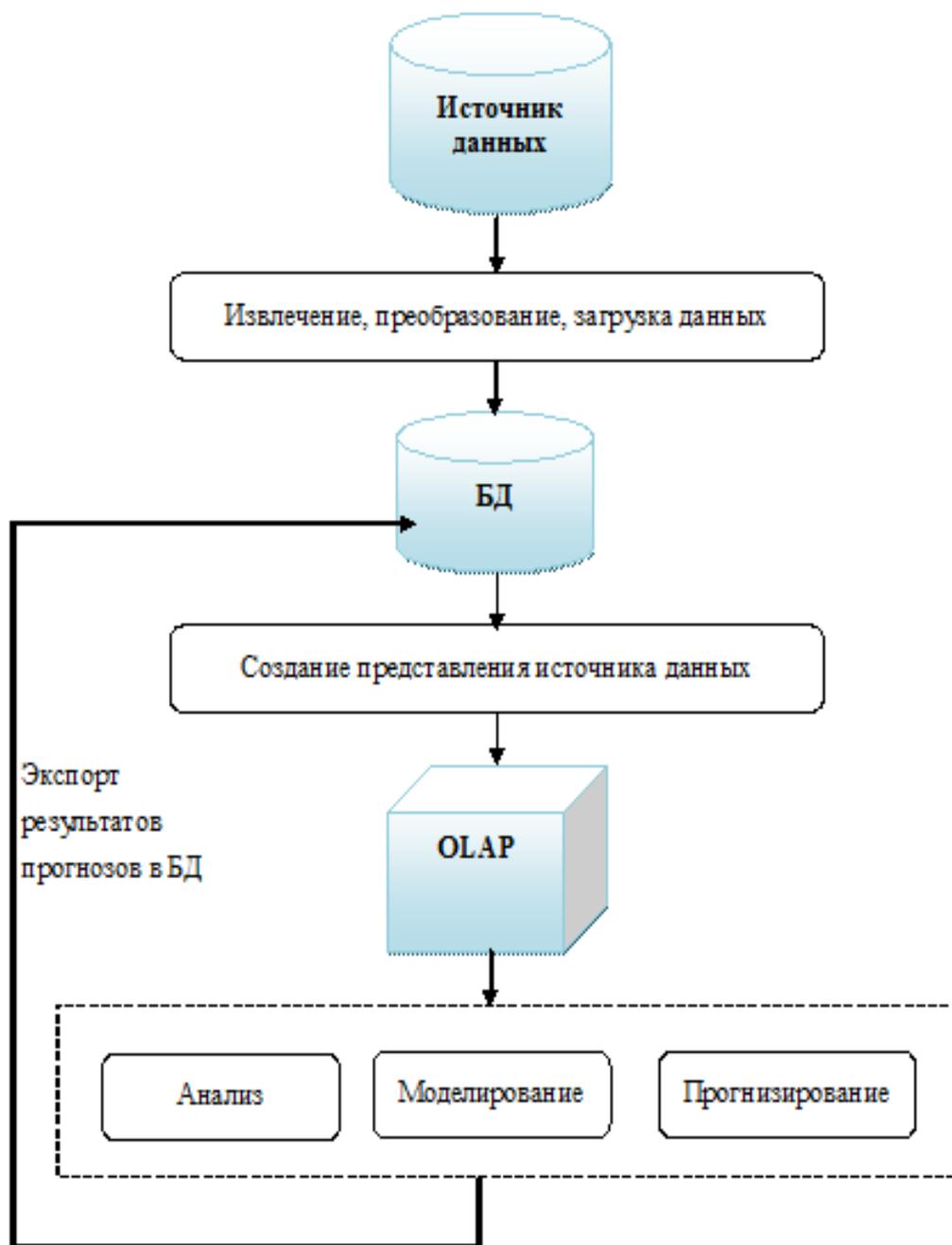


Рисунок 3.1 – Архитектура аналитической системы

Данные, подлежащие анализу, содержатся в хранилище данных, представляющее собой определённое количество таблиц базы данных. Данные обрабатываются в хранилище данных из внешнего источника после их извлечения и преобразования.

На основе данных, извлеченных из хранилища, построен ONLINE Online Analytical Processing (оперативная аналитическая обработка данных) - куб. На

уровне Online Analytical Processing (оперативная аналитическая обработка данных) - куба во время процессинга выполняется агрегация показаний приборов учета по времени, услугам и адресам и строятся модели Data Mining для прогнозирования. Разрабатываемый модуль включает в себя две модели интеллектуального анализа данных, одна из которых использует алгоритм дерева принятия решений (Microsoft Decision Trees), а другая модель – алгоритм временных рядов (Microsoft Time Series). Построение моделей реализуется в среде Business Intelligence Development Studio. Модели выполняют вычисление прогнозных значений показаний приборов учета.

3.3 Проектирование Online Analytical Processing (оперативная аналитическая обработка данных) - куба

При проектировании Online Analytical Processing (оперативная аналитическая обработка данных) - куба в качестве источника данных (Data Source) использовалась созданная ранее реляционная база данных на SQL сервере, соединение с которой организовано с помощью диспетчера соединений OLE DB. Online Analytical Processing (оперативная аналитическая обработка данных) - куб построен поверх созданного представления источника данных (Data Source View). Основная идея Online Analytical Processing (оперативная аналитическая обработка данных) заключается в создании многомерных кубов, которые будут доступны для пользовательских запросов.

Типичная структура хранилища данных существенно отличается от структуры традиционной реляционной базы данных. Как правило, эта структура денормализуется (это позволяет увеличить скорость выполнения запроса), поэтому она может позволить избыточность данных.

Online Analytical Processing – куб содержит основные данные и информацию о размерах. Куб потенциально содержит всю информацию, которая может потребоваться для ответа на любые запросы.

Online Analytical Processing система автоматический преобразуется в SQL-запрос к реляционным данным.

Схема представления источника данных представлена на рисунок 3.2:

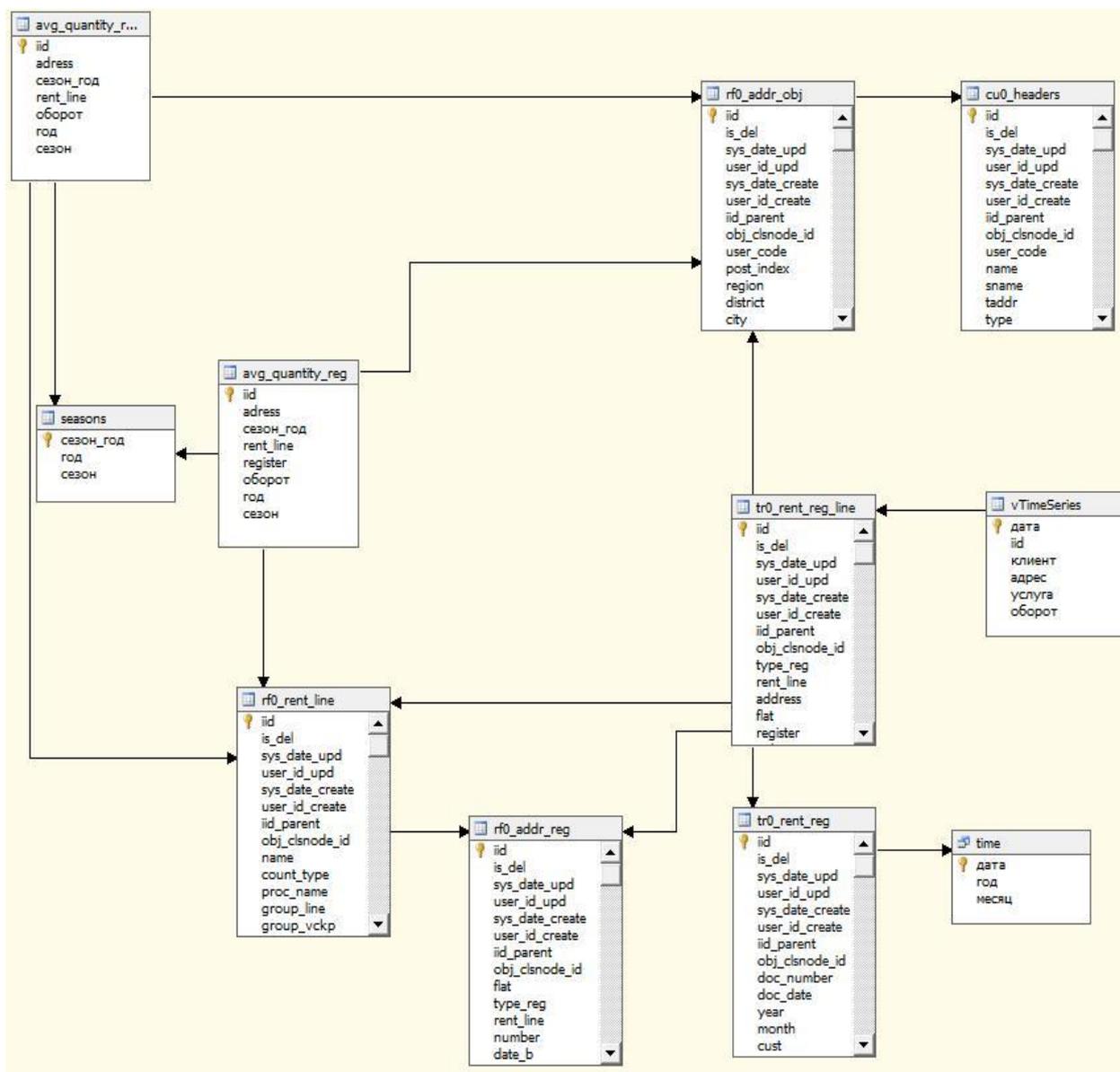


Рисунок 3.2 – Представление источника данных

Представление источника данных позволяет изменять структуру исходной базы данных, выбирать только требуемые для решения задачи таблицы и представления, определять связи между ними, добавлять именованные запросы и при этом не вносить изменения в источник данных. Оно содержит информацию, накопленную в процессе сбора показаний по приборам учета. Измерения составляют таблицы, содержащие информацию о клиентах и прикрепленным к ним объектам недвижимости, приборах учета и временное измерение.

Описание измерений спроектированного куба приведено в таблице 1.

Таблица 1 Описание измерений многомерного куба

Название измерения	Иерархии	Тип измерения
Rf0_addr_obj	Клиент-район-улица-дом-корпус	Обычное (regular) измерение
Rf0_addr_reg	нет	Обычное (regular) измерение
Rf0_rent_line	нет	Обычное (regular) измерение
Tr0_rent_reg	нет	Обычное (regular) измерение
Time	Год - месяц	Временное измерение

Схема построенного Online Analytical Processing (оперативная аналитическая обработка данных) - куба приведена на рисунке 3.3.

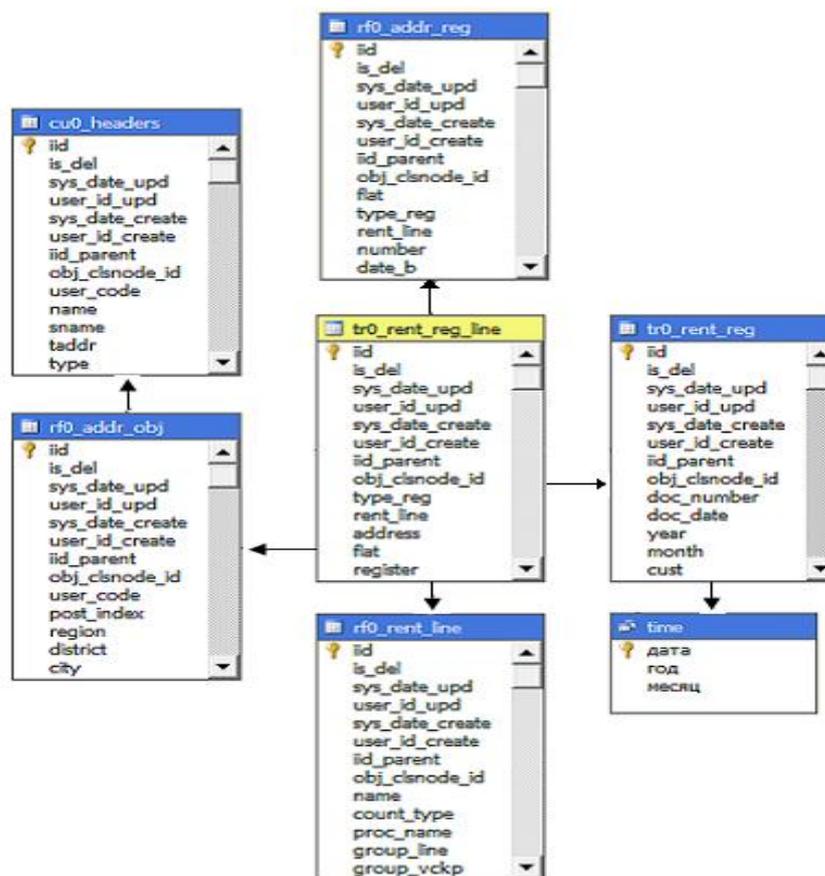


Рисунок 3.3 – Схема данных Online Analytical Processing (оперативная аналитическая обработка данных) - куба

После развертывания измерений и Online Analytical Processing (оперативная аналитическая обработка данных) - куба на сервере было

протестировано его функционирование. Для этого был выполнен тестовый срез куба. Для среза были использованы измерения «Times», «Rf0_rent_line» и иерархия по клиентам и прикрепленным к ним объектам недвижимости. Кроме того, для среза был использован показатель оборота таблицы фактов. Одним из способов получения среза Online Analytical Processing (оперативная аналитическая обработка данных) - куба является язык MDX, который специально предназначен для создания запросов к многомерным кубам.

MDX запрос:

```
SELECT NON EMPTY { [Measures].[Quantity] } ON COLUMNS, NON EMPTY { ([Tr0
Rent Reg Line].[Клиент].[Клиент].ALLMEMBERS * [Tr0 Rent Reg
Line].[Улица].[Улица].ALLMEMBERS * [Tr0 Rent Reg Line].[Дом].[Дом].ALLMEMBERS *
[Tr0 Rent Reg Line].[Корпус].[Корпус].ALLMEMBERS * [Tr0 Rent Reg
Line].[Услуга].[Услуга].ALLMEMBERS * [Time].[Год].[Год].ALLMEMBERS *
[Time].[Месяц].[Месяц].ALLMEMBERS ) } DIMENSION PROPERTIES
MEMBER_CAPTION, MEMBER_UNIQUE_NAME ON ROWS FROM [Umka1 1] CELL
PROPERTIES VALUE, BACK_COLOR, FORE_COLOR, FORMATTED_VALUE,
FORMAT_STRING, FONT_NAME, FONT_SIZE, FONT_FLAGS
```

Результат выполнения запроса представлен на рисунок 3.4.

Quantity		Названия столбцов											
		2009											
Названия строк	1	2	3	4	5	6	7	8	9	10	11	12	
Озерки	25058,58	22882,93	22174,36	18613,86	20333,87	18533,4	17951,66	17342,16	19467,35	56318,55	31932,42	37502,73	
ул. Есенина	25058,58	22882,93	22174,36	18613,86	20333,87	18533,4	17951,66	17342,16	19467,35	56318,55	31932,42	37502,73	
12	25058,58	22882,93	22174,36	18613,86	20333,87	18533,4	17951,66	17342,16	19467,35	56318,55	31932,42	37502,73	
1	25058,58	22882,93	22174,36	18613,86	20333,87	18533,4	17951,66	17342,16	19467,35	56318,55	31932,42	37502,73	
Водоотведение ОД ГВС	5302,94	5136,02	4654,06	5091,75	4631,45	3849,41	2845,67	1507,38	3976,86	4429,83	4234,21	4594,44	
Водоотведение ОД ХВС	9432	8349	8340	6365	7690	7235	7475	7880	7650	8240	8140	8560	
ГВС										13509	545	634	
ГВС общедомовое	400,61	346,68	303,96	323,78	287,85	213,99	155,99	74,78	190,49	285,73	293,03	337,46	
Отопление ОД	491,03	702,23	536,34	468,33	34,57	0	0	0	0	317,99	500,18	794,83	
ХВС										21296	721	814	
ХВС общедомовое	9432	8349	8340	6365	7690	7235	7475	7880	7650	8240	8140	8560	
Электроэнергия дневная ОД	0	0	0	0	0	0	0	0	0	0	6218	8941	
Электроэнергия ночная ОД	0	0	0	0	0	0	0	0	0	0	3141	4267	
Общий итог	25058,58	22882,93	22174,36	18613,86	20333,87	18533,4	17951,66	17342,16	19467,35	56318,55	31932,42	37502,73	

Рисунок 3.4 - Срез куба

3.4 Построение модели прогнозирования на основе алгоритма временных рядов

Результаты анализа с помощью Online Analytical Processing (оперативная аналитическая обработка данных) - куба позволили выявить сезонную периодичность исходных данных. Полученные исторические данные о показаниях приборов учета были использованы для обучения модели Microsoft Time Series, предназначенной для прогнозирования временных рядов. Точность прогнозирования можно оценить путем сравнения прогнозируемых значений с историческими данными. Для этого в исторических данных были выделены данные для обучения модели и для тестирования полученных результатов прогноза.

Временной ряд для обучения модели формируется с помощью представления vTimeSeries, который содержит результаты измерений показаний приборов учета по месяцам. Полученный временной ряд является чередующимся рядом, т.к. созданное представление включает в себя несколько записей, которые соответствуют одной дате, но содержат показания приборов для разных объектов недвижимости и услуг. Для того, чтобы выделить один элемент чередующегося временного ряда, были заданы ключевые атрибуты, включающие отметку времени и услугу. Таким образом, для каждой услуги создана отдельная модель временных рядов. Кроме того, были установлены фильтры для модели, позволяющие анализировать данные по конкретному объекту недвижимости.

С помощью мастера Mining Model Wizard была создана структура Data Mining и модель. При этом также основывается первая модель, основанная на алгоритме Microsoft Time Series.

Аналитик будет использовать алгоритм Microsoft Time Series для выполнения своей задачи.

Структура Data Mining созданная модель представлена на рисунок 4.4. Она содержит единственную модель, определённую вами в мастере Mining Model Wizard, ключевые столбцы и прогнозируемый столбец.

Структура ▲	v Time Series
	Microsoft_Time_Series
Адрес	Пропустить
Дата	Key
Клиент	Пропустить
Оборот	Predict
Услуга	Key

Рисунок 3.5 - Модель Data Mining на основе алгоритма временных рядов

Для корректной работы модели был установлен ряд параметров алгоритма временных рядов:

- Точное указание известных периодичностей PERIODICITY_HINT. В данном случае, параметр равен 12 и позволяет учитывать ежегодную периодичность.
- Используемый алгоритм прогнозирования FORECAST_METHOD. В данном случае используется
 - ARIMA модель, так как нас интересуют долгосрочные прогнозы.
 - Параметр для заполнения пропусков в предыстории MISSING_VALUE_SUBSTITUTION. В данном случае выбрано среднее значение за сезон.

После обучения модели итоговый прогноз извлекается с помощью DMX-запросов. В данном случае при написании инструкций DMX для получения требуемого прогнозного значения указывается сдвиг от текущего месяца и количество месяцев.

Результатом работы алгоритма временных рядов является график (рисунке 4.5), часть которого строится по имеющемуся временному ряду, а другая часть графика по прогнозным значениям и представляет собой пунктирную линию. Полученные прогнозные значения сохраняются в отдельную таблицу, чтобы иметь возможность получить результаты в источнике данных.

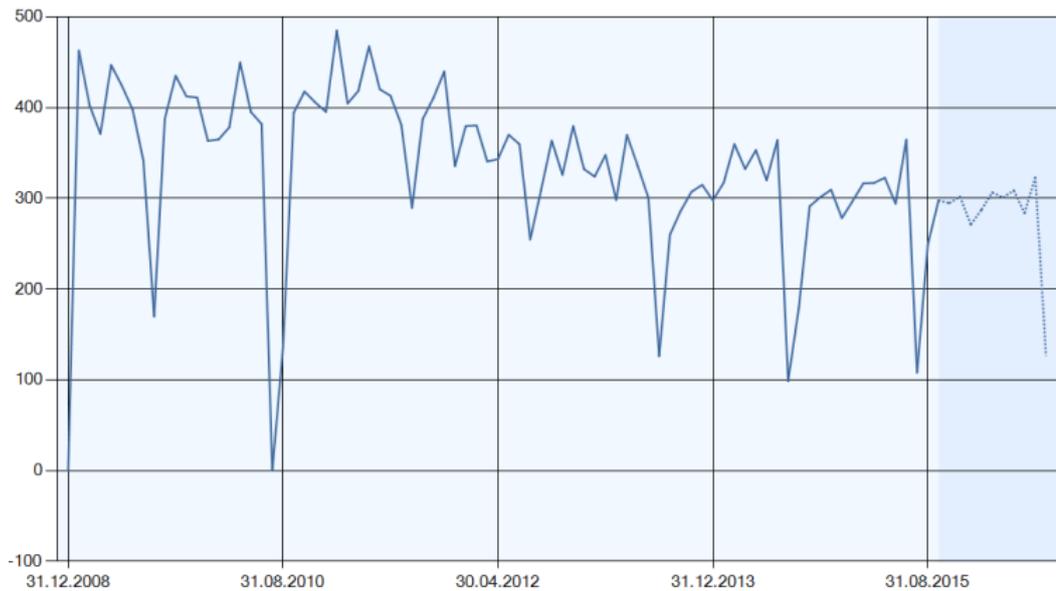


Рисунок 3.6 - Результат работы алгоритма временных рядов

3.5 Построение модели прогнозирования на основе алгоритма дерева принятия решений

Созданная модель на основе алгоритма дерева принятия решений представлена на рис. 3.7. Она содержит ключевой столбец, входные столбцы и один прогнозируемый столбец.

Структура	Time Series2
	Microsoft_Decision_Trees
Key	Key
Date	Input
Month	Input
Quantity	PredictOnly
Quarter	Input
Season Name	Input
Season	Input
Year	Input

Рисунок 3.7 - Модель Data Mining на основе алгоритма дерева принятия решений

Для корректной работы модели необходимо установить ряд параметров алгоритма:

- Параметр `COMPLEXITY_PENALTY`, который управляет ростом дерева решений. В данном случае значение параметра установлено согласно данным о количестве входных атрибутов (для атрибутов с 1 до 9) и равно 0.5.

- Параметр `FORCE_REGRESSOR`, с помощью которого устанавливаются регрессоры. В данном случае в качестве регрессоров были выбраны столбцы `Month` и `Quarter`, содержащие информацию о месяце и квартале года.

Дерево решений, построенное в результате работы алгоритма, приведено на рисунок 3.8.

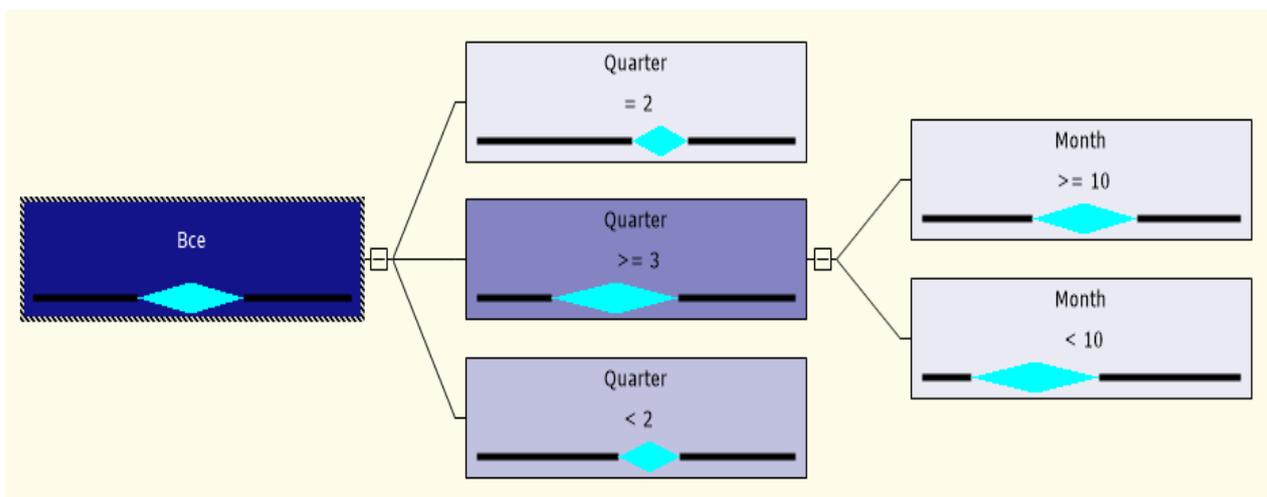


Рисунок 3.8 - Дерево решений

Полученная иерархия построена на основе классификации данных по правилу «Если..., то...», при этом насыщенность цвета ветвей отражает объем исходных данных, попавших в указанную ветвь.

Численные результаты работы алгоритма, которые были получены с помощью прогнозирующих DMX-запросов, представлены на рис. 3.9.

Полученные результаты сохраняются в отдельную таблицу исходной база данных, чтобы иметь возможность получить к результатам при помощи MSSQL Management Studio.

Quantity	date	rent_name	client	adress
67996,71875	31.12.2008...	Водоотвед...	Озерки	Есенина д.14 к.1
27316,347...	31.01.2009...	Водоотвед...	Озерки	Есенина д.14 к.1
27316,347...	28.02.2009...	Водоотвед...	Озерки	Есенина д.14 к.1
27316,347...	31.03.2009...	Водоотвед...	Озерки	Есенина д.14 к.1
27316,347...	30.04.2009...	Водоотвед...	Озерки	Есенина д.14 к.1
67996,71875	31.05.2009...	Водоотвед...	Озерки	Есенина д.14 к.1
67996,71875	30.06.2009...	Водоотвед...	Озерки	Есенина д.14 к.1
67996,71875	31.07.2009...	Водоотвед...	Озерки	Есенина д.14 к.1
67996,71875	31.08.2009...	Водоотвед...	Озерки	Есенина д.14 к.1
67996,71875	30.09.2009...	Водоотвед...	Озерки	Есенина д.14 к.1
67996,71875	31.10.2009...	Водоотвед...	Озерки	Есенина д.14 к.1
67996,71875	30.11.2009...	Водоотвед...	Озерки	Есенина д.14 к.1
67996,71875	31.12.2009...	Водоотвед...	Озерки	Есенина д.14 к.1
27316,347...	31.01.2010...	Водоотвед...	Озерки	Есенина д.14 к.1
27316,347...	28.02.2010...	Водоотвед...	Озерки	Есенина д.14 к.1
27316,347...	31.03.2010...	Водоотвед...	Озерки	Есенина д.14 к.1

Рисунок 3.9 - Результат работы алгоритма дерева принятия решений

3.6 Анализ результатов реализации модели - алгоритм дерева и алгоритм времени

Имеющаяся информация по показаниям приборов учета была разделена на тренировочные и тестовые данные.

На рисунок 3.10 приведены графики сравнения результатов прогнозирования моделей с реальными показаниями.

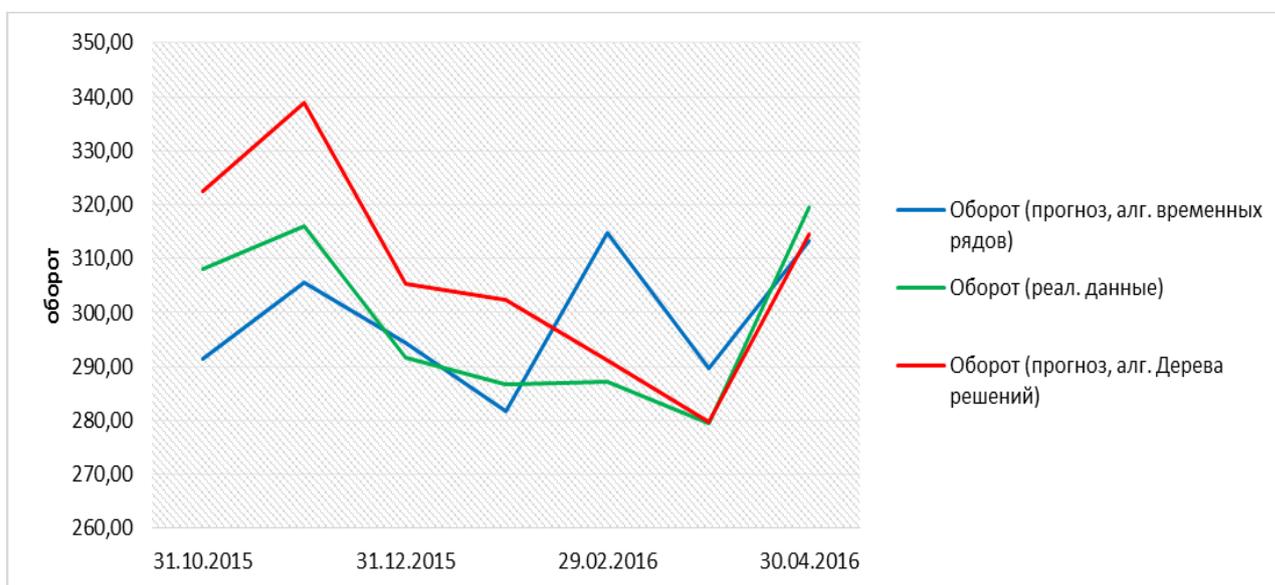


Рисунок 3.10 – График сравнения исходных данных с прогнозом на основе модели с алгоритмом Деревя принятия решений и алгоритмом Временных рядов

В ходе обработки результатов были вычислены оценки показателей качества моделей прогнозирования. Для анализа результатов были выбраны следующие показатели:

- средняя абсолютная ошибка (MAE):

$$MAE = \frac{1}{2} \sum_{t=1}^n |y_t - y_t^*| \quad (3.1)$$

где y_t – реальное значение в момент времени t ; y_t^* – прогнозное значение в момент времени t , полученное в результате работы модели прогнозирования; n – количество ретроспективных наблюдений.

Данный показатель позволяет количественно определить величину расхождения между реальными данными и значениями прогноза.

- коэффициент несоответствия КТ, равный нулю в случае совершенного прогноза (данный коэффициент не имеет верхней конечной границы):

$$K_T = \sqrt{\frac{\sum_{t=1}^n (y_t - y_t^*)^2}{\sum_{t=1}^n y_t^2 + \sum_{t=1}^n y_t^{*2}}} \quad (3.2)$$

- время работы алгоритма.

В таблице 2 представлены оценки рассчитанных показателей качества моделей прогнозирования.

Таблица 2 Оценки показателей качества моделей прогнозирования

	MAE	КТ	T(с)
Алгоритм временных рядов	11,29	0,034	13
Алгоритм дерева принятия решений	15,15	0,049	24

В результате сравнения реализованных моделей прогнозирования наилучшие показатели качества наблюдаются при работе модели на основе алгоритма временных рядов. Это можно объяснить тем, что у исходных данных присутствует ярко выраженная периодичность, которая может быть учтена при работе данной модели.

3.7 Вывод по главе 3

В этой главе приведена реализация моделей data mining для чего производится выбор средств реализации Data Mining. Спроектирован Online Analytical Processing (оперативная аналитическая обработка данных) – куб, после чего построен модель прогнозирования на основе алгоритма временных рядов и на основе на основе алгоритма дерева принятия решений. И в конце главы производится анализ результатов реализации модели - алгоритм дерева и алгоритм времени. Извлечение научных гипотез (data mining) и их последующий анализ (анализ данных) - два сложных неразрывно связанных процесса. Они следуют стандартной схеме установления физических законов: сбор экспериментальных данных, их организация в виде таблиц и поиск метода обработки, который позволит найти в исходных данных новые знания об анализируемом процессе. В то же время должно быть четкое понимание того, что это знание, как всегда, для любого сложного явления, остается в некоторой степени приблизительным: чем глубже анализируется реальная сложная система, тем менее очевидны наши суждения о ее поведении.

ЗАКЛЮЧЕНИЕ

В результате работы был создан прототип модуля прогнозирования с использованием технологии интеллектуального анализа данных.

В рамках данной работы было выполнено следующее:

- рассмотрен процесс планирования и прогнозирования денежной наличности;
- сформированы требования к разрабатываемому модулю;
- изучены алгоритмы интеллектуального анализа данных и выбраны соответствующие алгоритмы для построения моделей прогнозирования;
- разработаны многомерное аналитическое хранилище данных, модели прогнозирования на основе алгоритма дерева принятия решений и алгоритма временных рядов с использованием инструмента MS SQL Server Analysis Services;
- произведена оценка полученных результатов.

СПИСОК ИСПОЛЬЗУЕМОЙ ЛИТЕРАТУРЫ

1. Барсегян А.А. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, ONLINE ANALYTICAL PROCESSING (ОПЕРАТИВНАЯ АНАЛИТИЧЕСКАЯ ОБРАБОТКА ДАННЫХ) / А.А. Барсегян, М.С. Куприянов, В.В. Степаненко, И.И. Холод. — 2-е изд., перераб. и доп. — СПб.: БХВ-Петербург, 2017
2. Кураков. Л.П. Современные банковские системы: учебное пособие / Л.П. Кураков. 3-е изд., перераб, и доп. - М: Гелиос АРВ, 2014-320с.
3. Карпова И. П. Базы данных: модели, разработка, реализация: учеб. пособие / И. П. Карпова. — СПб.: Питер, 2017
4. Макаренко С. И. Интеллектуальные информационные системы: учеб. пособие / С. И. Макаренко – Ставрополь: СФ МГГУ им. М. А. Шолохова, 2016
5. Вдовин, В. М. Информационные технологии в финансовобанковской сфере: Учебное пособие / В. М. Вдовин, Л. Е. Суркова. - М.: Дашков и К, 2013. - 304 с.
6. Макленнен Дж. Microsoft SQL Server Data Mining – интеллектуальный анализ данных: Пер. с англ./ Дж. Макленнен, Чж. Танг, Б. Криват. – СПб.: БХВ – Петербург, 2013.
7. Лямин, Л. В. Применение технологий электронного банкинга: риск-ориентированный подход / Л. В. Лямин. - М.: КНОРУС: ЦИПСИР, 2015. - 336 с.
8. Гришин В.Н, Панфилова Е.Е Информационные технологии в профессиональной деятельности: Учебник / В.Н. Гришин, Е.Е. Панфилова. - М.: ИД ФОРУМ: ИНФРА-М, 2015. - 416 с.
9. Ершов М. В. Банковская система и развитие российской экономики / М. В. Мирова экономика и международные отношения. – 2015. - N 3. - С. 28-34.
10. Смирнов В. Базельский вызов / В. Смирнов Банковское дело. № 2. 2014. С. 27 - 28.
11. Царегородцев. Ю. Теоретические основы разработки и моделирования систем автоматизации / Ю. Царегородцев. - Издательство «Форум», 2014

12. Миронов Г. А. Автоматизированные информационные системы; Главная редакция физико-математической литературы издательства "Наука" / Миронов Г. А. – 2014
13. Бахтегараева В. Российская банковская система может стать одной из самых эффективных в мире / В. Бахтегараева - Аналитический банковский журнал. – 2015. - N 4. - С. 28-30.
14. Мезенцев К. Н. Автоматизированные информационные системы / К. Н. Мезенцев - Академия - М., 2014. - 176 с.
15. S. Manthorpe. S. madel of video codecs in an ATM ntwork, ITC-13 / S. Manthorpe, 2015. - 382 p.
16. Riley V. What avionics engineers should know about pilots and automation / V. Riley - IEEE Systems Magazine. - 2014.-May. - С.3-8.
17. Endsley M. Situation awareness, automation, and decision support: designing for the future / M. Endsley Human - Systems IAC GATEWAY. 2013
18. Ritter R.D. Pilot error in automated systems shown by altitude deviation reports / R. D. Ritter - 12th AIAA / IEE digital avionics systems conference, Fort Worth. 2016. – 144p
19. Author A.A. Title of book: Subtitle. Location: Publisher / A.A. Author 2013. pp. 33–135.
20. Daniel, A. Begun Amazing Android Apps For Dummies / Daniel A. Begun., 2016. - 256 с.