**АЗЕРБАЙДЖАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСТИТЕТ НЕФТИ И ПРОМЫШЛЕННОСТИ**

**На правах рукописи**

Гусейнов Рустам Ровшан оглы

**“Разработка и исследование моделей прогнозирования основанных на технологи нейронных сетей для информационных систем электронного правительства”**

Специализация: 060632.1 – Информационные технологии и системная инженерия

**МАГИСТЕРСКАЯ ДИССЕРТАЦИЯ**

**Научный руководитель: доц. Салимов В.Г.**

**БАКУ – 2024**

Реферат

**Тема диссертации:** «Разработка и исследование моделей прогнозирования основанных на технологи нейронных сетей для информационных систем электронного правительства».

**Объектом исследования** являются информационные системы e-government.

**Цель этой работы** – разработка и исследование моделей прогнозирования основанных на технологи нейронных сетей.

**Результаты** проведенных исследований представлены в параграфе «Вопросы практической реализации».

В информационных системах e-government существует две основные проблемы: управление большим объемом данных и необходимость быстрого принятия решений в условиях неопределенности. Для решения этих проблем могут быть применены нейронные сети и основанные на них системы интеллектуального анализа данных. Кроме того, применение нейронных сетей может помочь в улучшении безопасности информационных систем e-government. Например, нейронные сети могут использоваться для обнаружения и предотвращения кибератак.

В данной работе будут исследоваться, в том числе, способы приложения нейросетей к ИС электронного правительства в разрезе прогнозирования событий, например, для оценки числа потенциальных пользователей в будущем или изменения содержания запросов, поступающих от пользователей.

Применение методов прогнозирования, основанных на нейронных сетях, имеет большую актуальность в информационных системах e-government, поскольку это может улучшить эффективность и точность прогнозирования, ускорить принятие решений и повысить безопасность этих систем.

**ОГЛАВЛЕНИЕ**

[ВВЕДЕНИЕ 3](#_Toc133663945)

[ГЛАВА 1. ИССЛЕДОВАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ 4](#_Toc133663946)

[1.1 Обзор систем электронного правительства 4](#_Toc133663947)

[Глава 2 33](#_Toc133663948)

[2.1 Обзор технологий data mining 33](#_Toc133663949)

[2.2 Основы вычислительных нейронных сетей 49](#_Toc133663950)

[Глава 3 60](#_Toc133663951)

[3.1 Обзор программного обеспечения data mining 60](#_Toc133663952)

[ЛИТЕРАТУРА 72](#_Toc133663953)

# **ВВЕДЕНИЕ**

Исследования электронного правительства страдают от неопределенности концепции электронного правительства, чрезмерного упрощения процессов развития электронного правительства в сложных политических и институциональных средах и различных методологических ограничений. Для решения этих вопросов рассматриваются ограничения в литературе по электронному правительству и предлагаются пути продвижения вперед. В данном параграфе критически анализируется развитие и различные определения концепции электронного правительства. После обсуждения ограничений концепции в заключительной части анализа приводятся методологические и концептуальные меры, такие как (i) лучшее изучение и объяснение процессов и моделей участия в проектах электронного правительства в сложных политических условиях, (ii) решение проблемы недостаточной конкретизации в соответствующей литературе путем производства более обоснованных эмпирических исследований, которые создадут новые теоретические аргументы и предоставят новые концепции и категории, чтобы улучшить наше понимание политических процессов и действующих лиц электронного правительства, и (iii) прочное связывание предмета электронного правительства с основными исследованиями государственного управления.

# **ГЛАВА 1. ИССЛЕДОВАНИЕ ПРЕДМЕТНОЙ ОБЛАСТИ**

# **Обзор систем электронного правительства**

Цель этого параграфа состоит в том, чтобы рассмотреть ограничения в литературе по электронному правительству и описать предложения относительно того, как преодолеть эти ограничения, а также методологические и актуальные предложения, которые позволят продвинуть эту область дальше в области инновационных исследований. Часть проблемы, о которой идет речь, возникает из-за нечеткости концепции электронного правительства (Олдрич, Бертрот, МакКлюр, 2002, стр. 351; Хуанг, Чой, Мйонг, 1999, стр. 277–278). Чего также не хватает при рассмотрении предмета, так это более глубокого анализа политической природы процессов развития электронного правительства и более глубокого признания сложной политической и институциональной среды. Однако современные исследования электронного правительства по большей части ограничиваются изучением результатов и результатов проектов электронного правительства. Таким образом, понимание политических процессов, лежащих в основе развития электронного правительства, имеет жизненно важное значение для преодоления как определений, так и аналитических ограничений. Такие усилия требуют исторического понимания взаимосвязи между технологиями и управлением.

Первые исследователи, изучавшие технологию, рассматривали технологические вопросы в правительстве скорее, как второстепенную заботу, чем как основную управленческую функцию. Технология рассматривалась как средство преодоления ограничений ограниченной рациональности и обеспечения инфраструктуры для более эффективного принятия решений (Саймон, 1976, стр. 286). Другими словами, до появления Интернета и широкого распространения персональных компьютеров основными целями использования технологий в правительстве были повышение управленческой эффективности государственных администраторов при одновременном повышении производительность государственного управления. До тех пор основным применением технологий в правительственных организациях была автоматизация массовых транзакций, таких как финансовые, с использованием мейнфреймов (Шелин, 2003, стр. 121). Это была эпоха, когда большинство правительственных учреждений создавали и эксплуатировали свои компьютерные системы независимо друг от друга, по принципу "дымохода" (Олдрич и др., 2002, стр. 349). Технология была изолирована от ядра, чтобы справиться с неопределенностью. Это было необходимо, поскольку технологии и окружающая среда были воспринимаются как два основных источника неопределенности, которые бросают вызов рациональности при принятии организационных решений (Томпсон, 1967, стр. 1). Кроме того, поскольку информационные технологии (ИТ) использовались для автоматизации закулисных операций и повышения эффективности канцелярской деятельности (Зубов, 1988), государственные ИТ-специалисты были изолированы от функционального и исполнительного надзора (Холден, 2003, стр. 56). Перроу (1967) отличился, утверждая, что технология является важным фактором, определяющим структуру и стратегию организаций, которые ее используют. Распространение персональных компьютеров в 1980-х годах обеспечило каждого государственного администратора персональной системой информационных технологий и, таким образом, открыло новый период использования информационных технологий в правительстве. На этом этапе управление технологиями в государственных учреждениях начало децентрализовываться. Вместе с децентрализацией пришло осознание того, что вопросы информационных технологий должны быть интегрированы в основные функции правительства. Три важных события ознаменовали движение к интеграции технологий и связанных с ними вопросов в государственное управление в Соединенных Штатах.

1) Первый — это проект "Городские информационные системы" (URBIS), который проводился с 1973 по 1978 год в Калифорнийском университете в Ирвине многопрофильной командой. Это было “первое крупное, систематическое, эмпирическое исследование, в котором особое внимание уделялось политике и результатам, связанным с использованием компьютеров в организациях комплексного обслуживания” (Кинг, 2004, стр. 97). Это выявило “продолжающиеся социальные и политические процессы, в которых технология ограничена – в некоторой степени контролируется и формируется – окружающей средой" (Данзигер, Даттон, Клинг, Крамер, 1982, стр. 7). Эти исследователи придерживались точки зрения теории открытых систем на технологию и ее окружение и подчеркивали непрерывное взаимодействие между правительственными организациями и их внутренней и внешней средой (Данзигер и др., 1982, стр. 8). Они утверждали, что “вычислительная техника усилит мощь и влияние тех субъектов и групп, которые уже обладают наибольшими ресурсами и властью в организации” (Данзигер и др., 1982, стр. 18).

2) Во-вторых, в 1985 году комитет Национальной ассоциации школ общественных отношений и администрации (NASPAA) рекомендовал, чтобы вычислительная техника была основным навыком, преподаваемым в программах MPA (Нортроп, 2003, стр. 2). В итоговом отчете NASPAA представлены четыре важные рекомендации для программ магистратуры государственного управления (MPA) (Крамер и Нортроп, 1989). Эти рекомендации включали в себя обязательный курс изучения компьютера для всех студентов, обязательный курс "компьютерные приложения для управления" для всех студентов, специализацию по управлению информацией в нескольких университетах и интеграцию компьютерных навыков и знаний в основные курсы государственного управления.

3) В-третьих, Боузман и Бретшнайдер (1986) написали основополагающую статью в журнале Public Administration Review, в которой они утверждали, что технологии трансформируют правительство и этой области следует уделять больше академического внимания. Тем не менее, как уже упоминалось выше, для появления полноценной концепции электронного правительства нужно было дождаться широкого использования Интернета и паутины в целом. До этого ИТ-технологии использовались в правительстве в основном для внутренних целей и управления (Ho, 2002). Вместе с появлением Всемирной паутины 1990-е годы также стали свидетелями включения информационных технологий в государственную реформу благодаря отчету о национальном обзоре эффективности (NPR) в 1993 году и последовавшему за этим движению ‘переосмысление правительства’. Одним из важных результатов движения NPR является создание универсального правительственного портала, который в настоящее время называется ‘firstgov’ (Олдрич и др., 2002, стр. 350). Принятие некоторых очень важных законодательных актов в течение этого десятилетия поддержало движение за реформы и их использование в правительстве (Шелин, 2003, стр. 122-123). Поправка 1995 года к Закону о сокращении бумажной волокиты 1980 года (PRA) содержала руководящие принципы для государственных инвестиций в ИТ и поощряла более широкий межведомственный обмен информацией. Закон об электронной свободе информации 1996 года (EFOIA) разъяснил правила выдачи государственных электронных документов и публичного доступа к ним. Закон о согласовании личной ответственности и возможностей трудоустройства 1996 года (PRWORA) обязал агентства социальных служб проверить перспективность приложений электронного правительства на местах на межправительственном уровне. Закон о реформе управления информационными технологиями 1996 года, известный как Закон Клингера–Коэна, учредил должность директора по информационным технологиям (CIO) в каждом агентстве (Рельеа, 2000, стр. 382) и поощрял измерение эффективности как показатель отдачи от государственных инвестиций в технологии и интеграцию ИТ в процесс стратегического планирования (Уэстербэк, 2000), хотя звучат призывы к критической оценке последствий создания такой должности (МакКлюр и Бертрот, 2000). Кульминацией всех этих законодательных усилий стало принятие Закона об электронном правительстве 2001 года, который обеспечил как организационную, так и финансовую инфраструктуру для широкого распространения приложений электронного правительства (Шелин, 2003, стр. 124).

Трагические события 11 сентября 2001 года вызвали серьезный сдвиг в восприятии электронного правительства: от инструмента повышения удобства предоставления государственных услуг, содействия административной реформе и продвижения демократического участия к инструменту защиты от террористических угроз (Холчин, 2004, стр. 406-407, 416; Зайферт & Рельеа, 2004, стр. 400-401, 404). Среди изменений, вызванных событиями после 11 сентября, - желание правительства содействовать обмену информацией между ведомствами (Холчин, 2004, стр. 409-410; Рельеа, 2004; Зайферт & Рельеа, 2004, стр. 402), слияние и/или совместное использование правительственных баз данных (Холчин, 2004, стр. 410), повышение защищенности правительственных информационных систем от возможных террористических атак (Холчин, 2004, с. 411), оценка и, при необходимости, сокрытие и/или устранение содержимого правительственных веб-сайтов, которое могло бы поставить под угрозу безопасность, практика, известная как "очистка веб-страниц" (Feinberg, 2004, с.. 445; Холчин, 2004, стр. 412-416; Зайферт & Рельеа, 2004, стр. 404), расширение количества и сферы применения методов анализа фактических данных и интеллектуального анализа данных (Файнберг, 2004, стр. 451; Зайферт, 2004; Зайферт & Рельеа, 2004, стр. 402-403), сопровождающееся некоторыми негативные внешние факторы, такие как "ползучесть миссии" (Геллман, 2004, стр. 499-500; Зайферт, 2004, стр. 467), снижающие гарантии против сбора, интеграции и межведомственного обмена частной личной информацией, даже в том числе из частного сектора (Рейган, 2004; Зайферт & Рельеа, 2004, стр.. 402-403), создание новых категорий классификации информации, таких как "чувствительная, но не конфиденциальная", "информация о критической инфраструктуре" (Feinberg, 2004, стр. 443-444), и, таким образом, создание тревожной тенденции к секретности в правительстве и поднятие вопросов конфиденциальности и законного использования информации (Feinberg, 2004, стр. 451, 454; Зайферт & Рельеа, 2004, стр. 402-403), среди прочих. В дополнение ко всем этим изменениям сама система электронного правительства и ее инфраструктура стали потенциальной мишенью терроризма (Холчин, 2004, с. 410-411). Время покажет, поставит ли это серьезное смещение акцента под угрозу потенциальные административные и политические выгоды электронного правительства и его дальнейшее развитие.

С одной стороны, с изменением акцента на электронное правительство тесно связана неотъемлемая несовместимость между восприятием электронного правительства, ориентированного на безопасность, и, по крайней мере, тремя первоначальными основополагающими принципами феномена электронного правительства, а именно быстрым и легким доступом к правительственной информации, открытым правительством, правом людей к знанию, прозрачности и отзывчивости (Доти & Эрделез, 2002, с. 370; Холчин, 2004, с. 417; Хернон, 1998). С другой стороны, независимо от изменения акцента в усилиях по электронному правительству, несколько критиков предупредили общественность о возможных ловушках феномена электронного правительства. Йагер (2002), например, указал, что широкое сотрудничество и обмен информацией между ведомствами могут поставить под угрозу некоторые конституционные принципы, такие как разделение властей, а также распределение и баланс полномочий между федеральным правительством, правительствами штатов и местными органами власти (Доти & Эрделез,2002).

Было признано, что информационно-коммуникационные технологии (ИКТ) обладают огромным административным “потенциалом” (обсуждение ограничений и неудач ИКТ в оказании помощи правительствам в предоставлении информации и услуг (см. Хикс, 1999a, 2001a). Например, ИКТ могли бы помочь создать сетевую структуру для обеспечения взаимосвязанности (МакКлюр, 2000), предоставления услуг (Беккерс и Зуридис, 1999), эффективности (Хикс, 2001b), интерактивности (ДиКатерино и Пардо, 1996), децентрализации, прозрачности (Ла Порт, Де Йонг, Демчак, 1999) и подотчетность (Гере & Йанг, 1998; Хикс, 1998, 1999b; МакГрегор, 2001). Электронное правительство, или e-government, стало популярным словосочетанием в государственном управлении, охватывающим все эти функции. Какого-либо общепринятого определения концепции электронного правительства не существует (Холчин, 2004, с. 407). Чтобы в достаточной степени охватить разнообразие применений и нюансы, ниже представлено несколько определений. Электронное правительство определяется как “использование Интернета и Всемирной паутины для предоставления государственной информации и услуг гражданам” (UN & ASPA, 2002). Это может также включать использование других ИКТ в дополнение к Интернету и паутине, таких как “базы данных, создание сетей, поддержка дискуссий, мультимедиа, автоматизация, отслеживание и прослеживаемость, а также технологии идентификации личности” (Йагер, 2003, стр. 323). Фонтейн (2001) предпочитает называть это явление "цифровым правительством" или "виртуальным государством’ вместо электронного правительства.

«Цифровое правительство — это правительство, которое все чаще организуется с точки зрения виртуальных агентств, межведомственных и государственно–частных сетей, структура и возможности которых зависят от Интернета. Виртуальное агентство, следуя модели веб-портала, используемой в экономике, организуется клиентом»

Минс и Шнайдер (2000) определяют электронное правительство как отношения между правительствами, их клиентами (предприятиями, другими правительствами и гражданами) и их поставщиками (опять же, предприятиями, другими правительствами и гражданами) с использованием электронных средств. Аналогичным образом, для Хернона (в Даффи, 2000) электронное правительство — это

«Просто использование информационных технологий для предоставления государственных услуг непосредственно клиенту в режиме 24/7. Клиентом может быть гражданин, предприятие или даже другое государственное учреждение»

Браун и Брадни (2001, стр. 1) определяют электронное правительство как использование технологий, особенно веб-приложений, для расширения доступа к правительственной информации и услугам и их эффективного предоставления.

В этот список можно включить две дополнительные категории: взаимодействие правительства с организациями гражданского общества (G2CS) и гражданина с гражданином (C2C), если взаимодействие между гражданами связано с тремя другими категориями электронного правительства. Эти категории вместе с их характеристиками, определением и примерами приведены таблице 1.1:



Таб. 1.1 «Категории связей e-government»

Эта таблица показывает, что концепцию электронного правительства можно воспринимать совершенно по-разному в зависимости от того, на чем сосредоточен человек. Электронное правительство также воспринимается по-разному в связи с его теоретической подоплекой. Согласно Гарсону (1999), существует четыре теоретические основы, в рамках которых концептуализируется электронное правительство:

* Первая структура предполагает использование потенциала информационных технологий в области децентрализации и демократизации.
* Вторая нормативная/антиутопическая структура подчеркивает ограничения и противоречия технологии.
* Третья концепция предполагает социотехнический системный подход подчеркивает непрерывное и двустороннее взаимодействие технологии и организационно–институциональной среды.
* Четвертая концепция помещает электронное правительство в рамки теорий глобальной интеграции.

Развитие электронного правительства изучается путем построения моделей его этапов. Первая модель (представленная Лейном и Ли, 2001, стр. 124) утверждает, что проекты электронного правительства проходят четыре стадии разработки по мере увеличения их интеграции и технологической и организационной сложности. Первый этап — это каталогизация, предоставление правительственной информации путем создания веб-сайтов правительственных учреждений. На данном этапе возможна только односторонняя коммуникация между правительством и управляемыми. Второй этап — это транзакция. Агентства на этом этапе могут осуществлять онлайн-транзакции с государственными учреждениями. Это делает возможной двустороннюю связь. Этапы каталогизации и транзакций сосредоточены на создании электронного интерфейса для правительственной информации и услуг. Третий этап — это интеграция правительственных операций в рамках функциональных областей государственного управления. Агентства, работающие в одной и той же функциональной области, интегрируют свои онлайн-операции. Например, совместное использование баз данных ФБР, ЦРУ и АНБ. Заключительный этап - горизонтальная интеграция. Различные функциональные области интегрированы в одну и ту же электронную систему и доступны для использования через центральный портал. Последние два этапа сосредоточены на интеграции обеспечения деятельности электронного правительства в рамках существующей правительственной структуры.

Вторая модель развития электронного правительства была представлена в исследовании, проведенном Организацией Объединенных Наций и Американским обществом государственного управления (UN & ASPA 2002, стр. 2). В нем была предложена пятиэтапная модель развития. Первый этап — это этап "становления", на котором устанавливается официальное присутствие правительства в Интернете. Во-вторых, количество правительственных сайтов увеличивается и становится более динамичным на этой ‘расширенной’ стадии. Третий ‘интерактивный’ этап позволяет пользователям загружать формы и взаимодействовать с должностными лицами через Интернет. На четвертом "транзакционном" этапе пользователи получают возможность осуществлять онлайн-платежи за транзакции. Заключительный ‘бесшовный’ этап делает возможной интеграцию электронных услуг между государственными учреждениями.

Модель ASPA-UN очень похожа на модель Лейна и Ли (2001, с. 124). Стадии ASPAUN "становления " и " расширенной" примерно соответствуют стадии каталогизации Лейна и Ли. ‘Интерактивные’ и ‘транзакционные’ стадии сравнимы со стадией ‘транзакции’ Лейна и Ли. Этап ‘бесшовности’ охватывает как вертикальную, так и горизонтальную интеграцию. Признавая совпадение, (Шелин 2003, стр. 129) организовали типологию электронного правительства, используя обе модели (см. таблицу 1.2).



Таб. 1.2 «Типология e-government»

Можно утверждать, что как модели Лейна и Ли (2001), так и ООН и ASPA (2002), а также типология Шелина (2003) являются чрезмерными упрощениями. Другими словами, поэтапный подход к электронному правительству неудовлетворителен. Этапы развития электронного правительства не обязательно четко следуют друг за другом в хронологическом или линейном порядке. Более того, такие модели могут быть неприменимы к развитию электронного правительства в развивающихся странах, поскольку у этих стран есть шанс извлечь уроки из успехов и неудач электронного правительства развитых стран. Можно утверждать, что в развивающихся странах процесс обучения проходит гораздо быстрее; они могут выполнять требования всех этапов практически одновременно.

**Взгляд на внедрение технологий электронного правительства**

В дополнение к этапам разработки электронного правительства, представленным выше, Фонтейн (2001) представила рамки внедрения технологий. Эта структура состоит из трех основных элементов (см. рис.1.1).

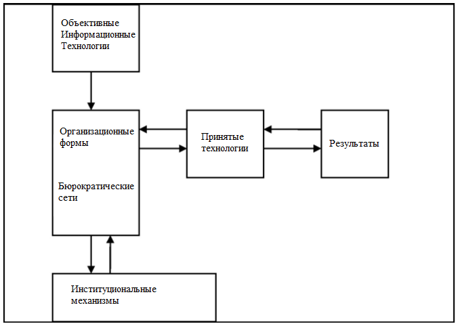


Рис. 1.1 «Схема внедрения по Фонтейн»

Во-первых, применение ИТ в организации изменяет объективную форму этой технологии из-за ее адаптации к организационной форме. Во-вторых, существует двустороннее взаимодействие между существующими институциональными механизмами и организационными формами. В-третьих, первые два элемента, то есть процессы принятия и внедрения, преобразуют объективную форму технологии в ее принятую форму (Бретшнайдер, 2003, с. 738). Другими словами, технология адаптируется к потребностям и среде конкретной организации в процессе внедрения. Система внедрения технологий Фонтейн подвергается критике по трем причинам. Во-первых, Бретшнайдер (2003) утверждает, что структура настолько абстрактна и обобщена, что ее трудно использовать для прогнозирования. Он также подчеркивает отсутствие многочисленных проверяемых гипотез в работе Фонтейн. Норрис (2003) соглашается с этим аргументом и утверждает, что в этих случаях не представлено достаточных доказательств для проверки теории принятия закона. Например, оба ученых считают, что Фонтейн преувеличивает свои аргументы в пользу межорганизационных альтернатив бюрократии (Бретшнайдер, 2003; Норрис, 2003). Норрис (2003) даже утверждает, что книга Фонтейн посвящена не информационным технологиям и институциональным изменениям, а организационной политике. Гарсон (2003) пересматривает ведущее дело Фонтейн и представляет доказательства того, что дело развивалось совсем не так, как оно представлено в книге. Янг (2003) утверждает, что использование Фонтейн неоинституционалистской теории не объясняет, как агенты преодолевают институциональные барьеры. Данцигер (2004), с другой стороны, защищает рамки принятия технологии. Он утверждает, что разделение объективных и встроенных технологий, более глубокое изучение социальных сетей и более детальное обсуждение институциональной теории полезны. Вторая критика заключается в том, что исследовательская программа Фонтейн недостаточно хорошо увязана с предыдущей литературой как по государственному управлению, так и по информационным технологиям (Бретшнайдер, 2003; Норрис, 2003). Норрис (2003) утверждает, что электронное правительство не так уж сильно отличается от использования других технологий в государственном управлении. Таким образом, предыдущие уроки применимы и к электронному правительству. Графтон (2003) согласен с Бретшнайдером в том, что основная часть работы Фонтейн представляет собой “полезную переформулировку общепринятых идей”. Он также не согласен с утверждениями Фонтейн об оригинальности ее фреймворка. Третья критика связана с ограниченной направленностью Фонтейн, поскольку в книге используются примеры только из федерального правительства США. В то время как Бретшнайдер (2003) видит в этом проблему, Графтон (2003) с этим не согласен, заявляя, что обсуждение применимо к правительствам штатов и местным органам власти в Соединенных Штатах и других странах развитого мира. Существует более общая критика в связи с тем, что большая часть литературы по электронному правительству сосредоточена на примерах внедрения электронного правительства в США на национальном уровне (Стоуэрс, 2003); и многому можно научиться на региональном и местном уровнях электронного правительства (Йагер, 2003, с. 327) и из других национальных контекстов.

В свете представленных выше обсуждений концепция электронного правительства ограничена четырьмя аспектами. Первое ограничение электронного правительства заключается в том, что до сих пор не существует стандартного определения этого понятия. Другими словами, трудно определить, что именно представляет собой электронное правительство. Эта трудность проистекает из нескольких причин: во-первых, электронное правительство — это концепция, определяемая целью деятельности (передача правительственной информации и услуг между правительствами, их клиентами и поставщиками), а не конкретной используемой технологией, поставщиком услуги/информации или четкими действиями связанных с ними субъектов. Следовательно, многие определения электронного правительства довольно расплывчаты и замалчивают множество значений, которые электронное правительство может иметь в зависимости от конкретного контекста, нормативно-правовой среды, доминирования группы субъектов в данной ситуации, различных приоритетов в государственных стратегиях (Торрес, Пина, & Ачерете, 2005) и т.д. Хикс (2003) приводит противоположный аргумент в переписке по электронной почте с автором этой статьи:

*...Существует очень важная скрытая дискуссия / трудность ... в том, что некоторые рассматривают электронное правительство как цель, некоторые рассматривают его как инструмент для достижения других, более широких целей реформы государственного сектора.*

*Я бы также добавил свое собственное основное различие в определениях — некоторые рассматривают электронное правительство как применение Интернета в государственном управлении (и, следовательно, как нечто новое и непохожее); другие - включая нас здесь, в Манчестере – рассматривают электронное правительство как применение цифровых ИКТ в государственном секторе (и, следовательно, как то, что продолжается уже много десятилетий, хотя раньше мы не называли это электронным правительством).*

Это заставляет задуматься, разумно ли обсуждать электронное правительство с точки зрения технологий, используемых при его создании. По сути, технологии приходят и уходят. Технология — это всего лишь средство для достижения электронного правительства, которое представляет собой фундаментальное изменение в том, как правительства ведут бизнес с заинтересованными сторонами в сфере правительственной информации и услуг. Определенные технологии принципиально не определяют, что такое электронное правительство и каким оно будет. Понимание процессов, посредством которых определяются конечные продукты электронного правительства (правительственная информация и услуги), избавляет нас от излишнего внимания к артефактам (содержанию веб-сайтов, использованию определенных технологий). Детальное понимание процессов также помогает нам распознать ключевых игроков в разработке политики электронного правительства и последствия включения и/или исключения определенных политических субъектов в процессе разработки политики электронного правительства (Джонас, 2000). Если мы просто сосредоточимся на технологии или технологических артефактах, то невозможно оценить эволюционирующий характер концепции электронного правительства (Хуанг и др., 1999) и понимать, что независимо от технологии, используемой для предоставления услуг электронного правительства, главная проблема заключается в том, чтобы заставить правительство работать лучше, быстрее, удобнее для заинтересованных сторон и обеспечить административные и демократические каналы, которые было невозможно открыть с помощью старых технологических инструментов.

Во-вторых, электронное правительство — это одна из тех концепций, которые означают много разных вещей для множества разных групп (Грант & Чау, 2005). Например, Перри (2001) определяет различные части электронного правительства как предоставление электронных услуг, электронную демократию и электронное управление. Быстрые технологические изменения также затрудняют “полное понимание смысла, возможностей и пределов концепции” (Принс, 2001). Таким образом, существует множество альтернативных определений, каждое из которых подчеркивает определенный аспект этих взаимоотношений, например, те, которые относятся к вопросам подотчетности, прозрачности, интерактивности, участия, экономической эффективности и т.д. Например, в своем исследовании распространения электронной информации Министерством энергетики США Уитсон и Дэвис (2001, с. 79) определил электронное правительство как “внедрение экономически эффективных моделей для граждан, отраслевых федеральных служащих и других заинтересованных сторон для проведения деловых операций онлайн". Концепция объединяет стратегию, процесс, организацию и технологию.” Такое, казалось бы, ограниченное определение электронного правительства вполне приемлемо, поскольку это определение отражает характеристики определенного контекста и применения.

В-третьих, как будто недостаточно, чтобы реальное содержание концепции было двусмысленным, плохо определенным и / или зависящим от контекста, электронное правительство также содержит много шумихи и рекламных усилий / литературы, аналогичных концепциям “управления знаниями” (Лев, 2000; Лиссак, 2000) или “управление по целям” (Миллер & Хортуик, 2002). Шумиху часто обвиняют в том, что она завышает ожидания выше реалистичных уровней, не позволяет людям увидеть, что происходит не так в той или иной области, и таким образом откладывает корректирующие действия. Однако шумиха не всегда неэффективна. Это может быть функционально, если мобилизует интерес и дает людям общее (хотя иногда и довольно искаженное) видение, в соответствии с которым можно действовать. Наконец, кто-то может спросить, насколько существенные изменения требуются, чтобы соответствовать критериям для того, чтобы правительственный технологический проект назывался проектом электронного правительства. Например, достаточно ли статичных веб-сайтов или адресов электронной почты государственных менеджеров или требуется какой-то уровень взаимодействия? Лейн и Ли (Layne and Lee, 2001) отвечают на этот вопрос с помощью своей модели поэтапного развития электронного правительства. Проекты на любом из этих этапов можно было бы определить как проекты электронного правительства. Однако необходимо обеспечить более высокий уровень концептуальной ясности.

Теперь обсудим возможные решения четырех проблем, описанных выше в литературе по электронному правительству. Первое предложение состоит в том, чтобы изучить и лучше объяснить процессы и формы участия в проектах электронного правительства. Полученные результаты могут помочь государственным планировщикам и лицам, принимающим решения, разрабатывать более эффективные планы и решения, касающиеся электронного правительства. Кроме того, студенты, изучающие государственное управление, могли бы лучше понимать процессы внутри "черного ящика" разработки политики электронного правительства.

Феномен электронного правительства многими рассматривается как последний эпизод использования технологий в правительстве для проведения государственных реформ (Бретшнайдер, 2003, с. 739). Такой аргумент предполагает, что люди, которые реформируют правительство, будут использовать технологии для дальнейшей рационализации правительственных процессов. Чтобы прояснить, что делается во имя электронного правительства, будущие исследования должны проверить утверждения о растущей рациональности на эмпирических данных проектов электронного правительства и представить альтернативные объяснения, когда это применимо. Например, проект может быть инициирован отчасти потому, что другие организации готовят аналогичные проекты, или потому, что какая-то фирма-поставщик убеждает руководителей высшего звена государственных учреждений в том, что им “необходимо” немедленно реализовать такой проект. Кроме того, различные сделки, побочные процессы и переговоры, происходящие внутри субъектов политики электронного правительства и между ними, могут создать среду “полурациональных” или мусорных моделей принятия решений в процессах электронного правительства. Браун и Брадни (2003, с. 34) объясните этот нерациональный способ поведения неопределенностью организационной среды:

*Многие эмпирические данные свидетельствуют о том, что, сталкиваясь с плохо структурированной средой (высокая неопределенность и изменчивость задач), лица, принимающие решения, часто предпочитают полагаться на интуитивный, символический, политический подход к принятию решений, а не один из них основан на систематизированных данных и эвристике, заложенных в информации и технологиях.*

Второе предложение заключается в решении проблемы недостаточной конкретизации в литературе по электронному правительству. Эта литература все еще находится в зачаточном состоянии (Олдрич и др., 2002, стр. 354), с небольшим количеством обоснованных эмпирических работ (Данзигер & Андерсен, 2002), а то, что измеряется в основном, - это развертывание веб-сайта (Норрис & Мун, 2005). Для изучения эволюции электронного правительства необходимы лонгитюдные исследования, такие как Норрис и Мун (2005). Эмпирические данные, полученные в результате будущих исследований, также могут внести свой вклад в литературу, создав новые теоретические аргументы и предоставив новые концепции и категории, которые улучшат наше понимание политических процессов и действующих лиц в области электронного правительства. Отличными примерами создания таких новых концепций и категорий являются концепции "силового поля конкурирующих сил" и "давления поставщиков", используемые Йылдыз (2004) для лучшего объяснения процесса разработки политики электронного правительства в Турции. Главный аргумент в пользу концепции "силового поля конкурирующих сил" заключается в том, что в данном проекте существует множество конкурирующих сил, которые инициируют проект и помогают ему попасть в повестку дня принятия решений. Несколько факторов, выявленных Йылдыз в ходе полевых исследований, - это актуальные потребности, то есть проблемы, для которых проекты электронного правительства являются подлинными решениями, потребности в государственной реформе и административном контроле, различные виды изоморфного давления (нормативное, имитационное, принудительное), давление поставщиков (фирмы-поставщики ИТ используют своих сотрудников в сетях ИТ-политики для влиять на процесс принятия решений, который приводит к созданию и/или формированию проекта электронного правительства), и символические действия, которые являются "ритуальными аспектами административного поведения, поскольку оно включает повторение, ролевую игру, стилизацию, порядок, постановка и создание смысла’ (Гудселл, 1997).

Третье предложение заключается в объяснении процессов разработки политики в проектах электронного правительства в сложной политической среде. Проблемные области, в которых действуют правительства, плохо структурированы. Государственные администраторы пытаются решать трудноразрешимые и порочные проблемы, которые затрагивают все учреждения по вертикали и горизонтали (Браун & Брадни, 2003). Проблемы, связанные с электронным правительством, не являются исключением. Гил-Гарсия и Пардо (2005) утверждают, что практики недостаточно подготовлены к решению проблем, связанных с технологиями, поскольку они не могут использовать большую часть исследований в этой области. Это усложняет процессы планирования и принятия решений в правительстве. Отношение лиц, принимающих решения в правительстве, также ограничивает эти процессы (Браун & Брадни, 2003). Лучшее понимание этих установок могло бы помочь сделать сложность более управляемой. Этого можно достичь, используя опыт электронного правительства, накопленный ключевыми политиками, поскольку их восприятие составляет важный компонент их действий в отношении определения повестки дня и формулирования политики (Хикс, 2003). Только понимая эти процессы, можно проверить наличие или отсутствие системы внедрения технологий (Фонтейн, 2001) в электронном правительстве и использовать эту систему для оптимизации процессов принятия правительственных решений и планирования по вопросам ИКТ.

Последнее предложение состоит в том, чтобы тесно связать тему электронного правительства с основными исследованиями в области государственного управления. Электронное правительство не ограничивается только использованием существующих и появляющихся технологий в деятельности правительства. Это также связано со многими старыми и новыми основными проблемами государственного управления, такими как дихотомия политика–администрация (например, выборные должностные лица или назначенные доминируют в процессе развития электронного правительства?), межправительственные отношения (например, как национальная политика электронного правительства влияет на местное электронное правительство?), сети (например, какова роль социальных сетей в развитии электронного правительства?), стороннее правительство (например, при каких условиях следует передавать деятельность электронного правительства на аутсорсинг и как следует отбирать и контролировать эти фирмы?) и управление (например,, какова надлежащая роль граждан, организаций гражданского общества и частных фирм в развитии электронного правительства?), и это лишь некоторые из них. Результаты будущих исследований могут укрепить связь между электронным правительством и традиционными проблемами государственного управления.

Теперь объясним, почему изучение политической природы и политических процессов проектов электронного правительства, как было предложено выше, повышает ценность текущих исследований в области электронного правительства. Главный аргумент заключается в том, что в дополнение к исследованиям, ориентированным на результат, которые доминируют в литературе, необходимы исследования, объясняющие процесс и политическую среду проектов электронного правительства, исследования, в которых используются первичные данные, собранные в ходе обширной полевой работы, и подходы с использованием нескольких площадок и нескольких методов.

Электронное правительство – относительно новый предмет академического интереса в области государственного управления. Хотя более ранние версии Интернета и сопутствующие ему субтехнологии (электронная почта, передача файлов и т.д.) были доступны в течение последних трех-четырех десятилетий, только в последнее десятилетие 20-го века, особенно с появлением Всемирной паутины, Интернет стал использоваться в правительствах и вокруг них становится все более популярным среди граждан, а также среди правительств различных уровней, их поставщиков и клиентов, как подробно объяснялось выше.

В первое десятилетие академических исследований электронного правительства доминировали исследования, которые были сосредоточены на наблюдении и оценке результатов инициатив электронного правительства. В этих исследованиях было проанализировано содержание веб-сайтов местных, государственных и федеральных правительственных учреждений, перечислены виды онлайн-услуг, предоставляемых государственными учреждениями, и представлены примеры наилучшей практики для сравнительного анализа (Коэн & Эймик, 2001). Хотя исследования были полезны с практической точки зрения, они носили в основном описательный характер (Джонас, 2000, стр. 45).

Однако процессы развития электронного правительства привлекли ограниченное внимание членов исследовательского сообщества в области государственного управления. В одном из передовых исследований рассматривается использование технологий в местных органах власти США и объясняется, как технологии использовались для укрепления существующей структуры власти, путем изучения влияния политических и организационных факторов на применение компьютерных технологий в местных органах власти США (Данзигер и др., 1982). В дополнение к таким ранним исследованиям необходимы будущие исследования, направленные на изучение процессов, которые формируют управление проектами электронного правительства, а не процессов политических и организационных изменений.

Книга Фонтейн "Построение виртуального государства" (2001) важна как методологически, так и в других отношениях, поскольку в ней подробно рассматриваются процессы развития электронного правительства путем проведения многочисленных интервью и трех углубленных тематических исследований. Есть еще три причины методологической важности работы Фонтейн. Во-первых, она собрала свои собственные данные с помощью интервью и тематических исследований, а не путем анализа вторичных данных. Во-вторых, она провела теоретическую выборку (Фонтейн, 2001, стр. 16), отбирая случаи для углубленного изучения. Наконец, она использовала методологию обоснованной теории, поскольку стремилась “проверить утверждения на основе эмпирических данных для создания теории, а не для оценки прогностической теории” (Фонтейн, 2001, стр. 16). Хотя некоторые исследователи критикуют ее работу по разным причинам, как обсуждалось выше, ее методологический подход по-прежнему является скорее исключением, чем правилом. В большинстве исследований электронного правительства по-прежнему рассматриваются результаты электронного правительства, а не сами процессы. Они используют вторичные, а не первичные данные и редко проверяют предыдущие теории, не говоря уже о дополнении к этим теориям и/или создании новой теории.

Критика в адрес атеистичности не нова для сферы государственного управления. Например, Боузман утверждает, что подход государственной политики к государственному управлению в значительной степени атеоретичен, ориентирован на основанные на практике предписывающие правила. Предпочтительным методом исследования является также предпочтительный метод обучения — погружение в тематические исследования. Все те же аргументы применимы и к исследованиям в области электронного правительства. Большинство исследований электронного правительства носят либо описательный, либо рекламный характер с использованием таких методов, как контент-анализ правительственных веб-сайтов или маркетинговые усилия поставщиков. Более того, тематические исследования на сегодняшний день являются излюбленным исследовательским (если не обучающим) методом.

В настоящем исследовании выдвигаются два методологических предложения по изучению электронного правительства. Во-первых, в то время как изучение результатов усилий электронного правительства составляет большую часть литературы по электронному правительству, изучение и объяснение процессов развития электронного правительства также необходимо. Тщательный анализ политических процессов электронного правительства необходим для обеспечения того, чтобы ресурсы расходовались на проекты с наивысшим приоритетом, чтобы предотвратить повторение предыдущих ошибок и максимизировать выгоды при минимизации затрат ресурсов. В таблице 1.3 обобщен этот аргумент и, ввиду этого, категории, представленные там, не являются взаимоисключающими.



Таб. 1.3 «Методология исследований электронного правительства»

Как следует из названий, исследования электронного правительства, ориентированные на конечный результат, фокусируются на определенном этапе разработки проектов электронного правительства. Они изучают результаты усилий электронного правительства, артефакты, такие как веб-сайты и онлайн-правительственные услуги. С другой стороны, исследования, ориентированные на конечный результат, объясняют, какой показатель эффективности государственного управления (т.е. затраты, прозрачность, эффективность) улучшается в результате конкретных усилий электронного правительства. Ограниченная направленность этих двух типов исследований, как правило, сопровождается целью определения наилучшей практики для сравнительного анализа. Они рассматривают усилия электронного правительства глазами научного наблюдателя, который изучает продукты усилий электронного правительства (например, веб-сайты, онлайн-сервисы) извне и, как правило, с помощью вторичных данных. Цель состоит в том, чтобы найти успешные примеры для подражания. Такое внешнее обследование является дедуктивным подходом к электронному правительству (Агранофф, 2004). Эти исследования, по большому счету, носят ознакомительный и описательный характер. Они не говорят нам, что происходит внутри черного ящика электронного правительства.

Как указывалось выше, изучение процессов электронного правительства не является новым видом исследований. Проект городских информационных систем (URBIS), проведенный в 1970-х годах в Калифорнийском университете в Ирвине, привел к некоторым ранним исследованиям, ориентированным на процессы. Совсем недавно Беллами и Тейлор (1998) и Фонтейн (2001) провели исследования, ориентированные на процесс. Однако, вообще говоря, исследования, ориентированные на результат, доминировали в литературе по электронному правительству.

В исследованиях, ориентированных на процесс, обычно используются первичные данные, полученные в результате обширных полевых исследований. Они используют такие методы сбора данных, как интервью, наблюдение за участниками и архивный анализ, чтобы понять и объяснить процессы электронного правительства и выработать теорию. Один из способов создания этой теории заключается в том, что люди, работающие в проектах электронного правительства, рассказывают свои истории исследователю. Исследователь преобразует их слова в абстрактную мысль в результате систематического сбора данных, анализа данных и триангуляции. Это индуктивный, вывернутый наизнанку процесс, “...шпионская стратегия проникновения внутрь организации... чтобы понять, как оперативные силы преобразуют “входные данные” в “выходные данные”” (Агранофф, 2004).

Второе предположение заключается в том, что определенные методологические подходы, такие как использование первичных данных, триангуляция результатов и концепций с аналитическим подходом для создания теории, редко встречаются в литературе по электронному правительству. Использование таких стратегий значительно повышает ценность, поскольку они иллюстрируют более строгий методологический выбор, наблюдаемый лишь в небольшой части текущих исследований в области электронного правительства. Многие исследователи электронного правительства ограничивают свой анализ рамками государственного учреждения, использующего приложение электронного правительства. Триангуляция данных и выводов с использованием различных источников данных, методов, исследователей и типов данных, которые собираются на нескольких объектах, повышает строгость исследований. Кроме того, как государственные учреждения, так и их окружение должны анализироваться одновременно путем проведения периферийной выборки. Наконец, попытки тестирования или построения теории — это лишь малая часть текущих исследований в области электронного правительства. Исследования, использующие обоснованную теорию, не только строят теорию, переходя от эмпирических данных к гипотезам индуктивным способом; но они также выявляют новые переменные и концепции с помощью систематического анализа и тем самым создают дополнительные возможности для исследований по мере выявления новых переменных и концепций с помощью систематического анализа.

**1.2 Постановка задачи исследования**

# **Глава 2**

# **2.1 Обзор технологий data mining**

С развитием технологий, особенно за последние три десятилетия, огромное количество информации было переведено в цифровую форму, что привело к формированию огромных хранилищ данных. По мере накопления информации в этих хранилищах сохранялась проблема извлечения из нее значимых знаний. Для решения этой ситуации был использован интеллектуальный анализ данных в качестве инструмента. Интеллектуальный анализ данных, рассматриваемый как ступень к процедуре обнаружения знаний в базах данных, представляет собой процедуру извлечения скрытой информации из огромных наборов баз данных для выявления шаблонов и правил. Интеллектуальный анализ данных (data mining) в настоящее время стал незаменимым компонентом практически во всех сферах человеческой жизни. В данном параграфе представлен анализ доступной литературы по интеллектуальному анализу данных. Кратко излагается концепция интеллектуального анализа данных, а также его различные методологии. Также были проиллюстрированы некоторые приложения, задачи и проблемы, связанные с ИТ.

Наличие большого объема данных практически в каждой области и желание извлекать из них полезную информацию и знания были главной мотивацией, которая в недавнем прошлом привлекала внимание исследователей к интеллектуальному анализу данных. Извлеченные информация и знания могут оказаться чрезвычайно полезными для различных приложений, начиная от управления малым бизнесом и заканчивая сложным инженерным проектированием и научными исследованиями. Интеллектуальный анализ данных - это анализ и тщательное изучение огромных массивов данных с целью выявления важных закономерностей и правил, которые ранее были неизвестны. Основная цель - использовать возможности компьютера по обработке данных в сочетании со способностью человека воспринимать шаблоны (Хан и Кэмбер 2001). Эпоха приложений интеллектуального анализа данных началась в 1980 году преимущественно исследовательскими инструментами, ориентированными на выполнение рутинной работы в одиночку (Пятецкий-Шапиро 2000). В последнее время интеллектуальный анализ данных становится доминирующим среди статистиков и аналитиков данных. В 1989 году Пятецкий-Шапиро ввел в обиход KDD (Knowledge Discovery in Database / Поиск Знаний в Базах Данных). Признание интеллектуального анализа данных и KDD не должно вызывать удивления, учитывая масштаб данных, собранных из различных доступных источников, собранные данные проверяются вручную, и часто автоматический анализ данных, поддерживаемый классической статистикой и машинным обучением, может вызвать проблемы, если процедура сложная, а собранные знания включают в себя проблемные сущности. Огромный объем данных, собранных из многочисленных источников и хранящихся в обширных и разнообразных хранилищах. Таким образом, сбор данных превышает возможности человека для анализа без мощного инструмента анализа, в результате чего эти хранилища становятся "хранилищами данных", которые посещаются нечасто. Поскольку лицам, принимающим решения, не хватает инструментов для извлечения ценных знаний, содержащихся в огромном объеме данных, следовательно, при принятии жизненно важных решений не используются богатые информацией данные. Инструменты интеллектуального анализа данных выполняют анализ данных и определяют важные закономерности, которые ранее оставались не замечены. Поскольку каждая сфера человеческой жизни стала насыщенной данными, это привело к тому, что интеллектуальный анализ данных стал незаменимым компонентом. KDD можно рассматривать как всеобъемлющую процедуру извлечения полезных знаний из данных, в то время как интеллектуальный анализ данных можно рассматривать как ядро KDD, которое включает алгоритмы, которые исследуют данные, строят модели и обнаруживают неизвестные закономерности.

Файяд и др. в своей статье “От интеллектуального анализа данных к обнаружению знаний в базах данных” описали KDD как “нетривиальный\* процесс распознавания действительных, новых, потенциально полезных и, наконец, понятных шаблонов в данных”. При разработке определения данными были любые достоверные факты, доступные в электронной форме. Шаблоны - это модели, выражаемые некоторым языком в виде подмножества данных. Шаблоны должны быть допустимыми, чтобы они были истинными и могли быть смоделированы для любых новых данных. Процесс включает в себя множество этапов от подготовки данных до расширения знаний, которые используются повторно до тех пор, пока не будут достигнуты желаемые результаты. Нетривиальный\* указывает на то, что должно существовать своего рода вычисление вывода, чтобы отличать его от традиционного вычисления значений. Файяд и Столорц (1997) в своей статье описали KDD как “обобщенную процедуру извлечения ценных знаний из данных, при этом интеллектуальный анализ является одним из других этапов этого процесса, который использует некоторые алгоритмы для процесса извлечения знаний”. Чарльз и др. (1998) предложили интеллектуальный анализ данных в качестве эффективного инструмента прямого маркетинга, чтобы улучшить маркетинг продукции в век, когда традиционные средства маркетинга, такие как массовый маркетинг, демонстрируют тенденцию к упадку. Используя интеллектуальный анализ данных, можно определить модели поведения покупателей, чтобы выделить потенциальных покупателей из списка клиентов. Было продемонстрировано, что интеллектуальный анализ данных как инструмент прямого маркетинга может приносить больше прибыли, чем традиционные средства массового маркетинга, поскольку он ориентирован только на потенциальных покупателей. Майкл Гебель и др. (1999) в своей статье “Обзор инструментов интеллектуального анализа данных и обнаружения знаний” представили обобщенное представление об общих задачах обнаружения знаний и различных методологиях их решения. Была предложена схема классификации признаков, которая использовалась для изучения знаний и программного обеспечения для интеллектуального анализа данных. Они указали некоторые важные функции, которые следует считать важными для программного обеспечения для обнаружения знаний, чтобы оно могло эффективно использоваться и решало больше проблем, которые были недостаточно изучены. Многие организации в современном мире располагают огромными базами данных, которые не имеют никаких ограничений на рост. Новые данные добавляются в эти БД со скоростью миллионов записей в день. Эти типы баз данных создают новую проблему и уникальные возможности для обработки этих потоков данных. Отличая data mining от информации Хенд и др. (2001) определили интеллектуальный анализ данных как “анализ огромных наборов данных с целью обнаружения неожиданных взаимосвязей и более логичного анализа данных, чтобы это привело к желаемым результатам”. По словам Ригельского и др. (2002), технология data mining добавила новое измерение в CRM. Возможности интеллектуального анализа данных по извлечению прогнозируемой неизвестной информации из обширных наборов данных нашли свое применение в CRM для выявления и оценки ценных клиентов, прогнозирования покупательского поведения клиентов, что помогает поставщикам принимать упреждающие решения, основанные на знаниях. Имон Киф и др. (2004), обсуждался “интеллектуальный анализ данных без параметров”, поскольку алгоритмы, нагруженные параметрами, могут завышать или занижать оценку определенных параметров, что приведет к шаблонам, которые могут быть не совсем точными. Майнинг без параметров может помешать нам применять наши собственные предположения или предубеждения. Они предложили парадигму обработки данных, основанную на сжатии. Потоковый анализ данных считается сложной задачей при обнаружении знаний в базах данных, традиционные подходы интеллектуального анализа данных неосуществимы для решения этой задачи, поскольку данные поступают в виде множества, непрерывных и изменяющихся во времени потоков данных. Альхаммади и др. (2007) представили необычную методологию интеллектуального анализа эволюционирующих шаблонов потоковых данных, которая обладает большей сложностью интеллектуального анализа и точностью классификации, что было доказано экспериментально, поскольку большинство современных методологий интеллектуального анализа данных исследуют знания в единой таблице данных. Но в последнее время большинство из этих методологий применяются к реляционным случаям. Реляционный data mining предполагает применение подхода интеллектуального анализа данных к данным из нескольких таблиц для абстрагирования содержащихся в них знаний (Сасо Дзероски 2010). Венкатадри и др. (2011), обсуждают, какие соответствующие методы и методологии необходимы в будущем для удовлетворения потребностей области интеллектуального анализа данных, поскольку она исследует все более сложные области, чтобы мы могли исследовать такие сложные ситуации, когда данные огромны и полны скрытой информации. Типаван Сильваттананусарн и др. (2012) в своей работе исследовали применение методов интеллектуального анализа данных для поддержки процесса управления знаниями. Ими было показано, что интеллектуальный анализ данных может быть интегрирован и внедрен в систему управления знаниями. Дивья Томар и др. (2013) представили интеллектуальный анализ данных как наиболее динамичную и привлекательную область исследований, которая набирает популярность в медицинской сфере. Интеллектуальный анализ данных дает ряд преимуществ в области здравоохранения. Это повышает эффективность медицинских услуг с точки зрения затрат. Ананд Сауркар и др. (2014) определили интеллектуальный анализ данных как “междисциплинарную область, которая состоит из интегрированных баз данных, искусственного интеллекта, машинного обучения, статистики и т.д.”. Они определили интеллектуальный анализ данных как многоступенчатый процесс, который включает подготовку данных для интеллектуального анализа, алгоритмы интеллектуального анализа данных, анализ и интерпретацию результатов. Способность интеллектуального анализа данных глубоко копаться в данных и извлекать из них скрытую информацию и знания привлекла огромное внимание бизнес-профессионалов, чтобы генерировать модели, связанные с поведением клиентов, и прогнозировать будущие продажи и тенденции, а также помогать политикам в принятии решений с целью увеличения прибыли (Шраддха Сони, 2015).

Теперь затронем тему архитектуры Data Mining. Она представлена на рисунке 2.1.

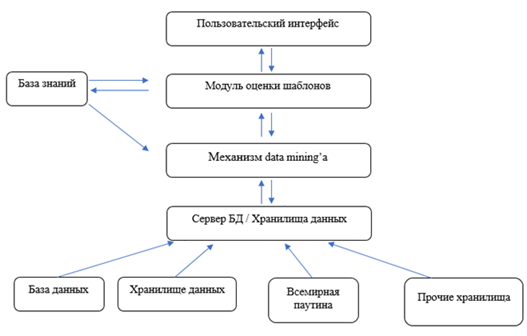


Рис. 2.1 «Архитектура data mining»

**База знаний:** служит отправной точкой для всего процесса интеллектуального анализа данных. Она действует как руководство для поиска или оценки интересности полученных паттернов.

**Механизм data mining’a:** является основным компонентом системы интеллектуального анализа данных, он состоит из всех необходимых модулей, таких как характеристика, прогнозирование, кластерный анализ, анализ выбросов и эволюционный анализ для выполнения задач интеллектуального анализа данных.

**Модуль оценки шаблонов:** Этот модуль обычно связан с показателями интересности. Он постоянно взаимодействует с механизмом интеллектуального анализа данных для поиска закономерностей. Часто использует пороговые значения для отсеивания обнаруженного шаблона или может использовать модуль оценки шаблона, интегрированный с модулем интеллектуального анализа данных, в зависимости от используемого метода интеллектуального анализа данных.

**Пользовательский интерфейс:** Модуль служит связующим звеном между пользователями и системой интеллектуального анализа данных. Это облегчает взаимодействие пользователей с системой простым и эффективным способом, не беспокоя пользователя о сложностях, стоящих за процессом.

**Источники данных** (www, хранилище данных, база данных, другие хранилища). Организации обычно хранят данные в базах данных или хранилищах данных. Иногда в хранилище данных содержится более одной базы данных, текстовых файлов или электронных таблиц.

**База данных или сервер хранилища данных:** Они содержит данные, которые настроены для извлечения. Получение данных по запросу пользователей является их ключевым назначением.

**Другие обработки:** перед передачей данных на сервер хранилища данных данные необходимо очистить и интегрировать, поскольку данные собираются из разных источников и в разных форматах, поэтому их нельзя использовать непосредственно для процесса интеллектуального анализа данных. Данные должны быть очищены, интегрированы, и только надежные данные должны быть отобраны и переданы на сервер хранилища данных. Для этого процесса может потребоваться ряд методов очистки, интеграции и выбора.

Поговорим теперь о самом процессе data mining’a. Файяд и др. (1996) определили “интеллектуальный анализ данных как один из нескольких этапов в процессе обнаружения знаний, он включает в себя применение анализа данных и обнаружение алгоритмов, которые дают точное перечисление шаблонов по данным при любой приемлемой эффективности вычислений”. Эта процедура является совместной и повторяющейся и включает в себя множество шагов с решениями, принимаемыми пользователем, с попытками, предпринимаемыми на каждом шаге для выполнения конкретной задачи обнаружения, каждая из которых выполняется с помощью применения метода обнаружения. Интеллектуальный анализ данных, синонимично используемый некоторыми для процесса обнаружения знаний в базах данных (KDD), и наоборот, многие рассматривают его как важный этап KDD, который приводит к созданию полезных шаблонов или моделей для данных. Различные этапы data mining показаны на рисунке 2.2.

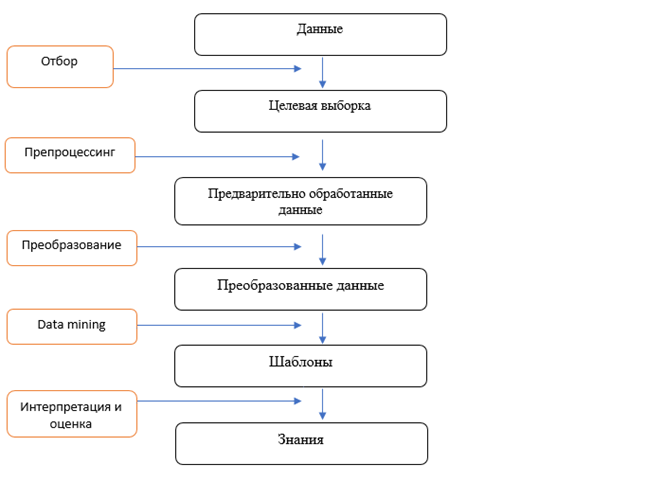


Рис. 2.2 «Этапы data mining»

**Отбор**: Отбор соответствующих данных из различных источников для процесса интеллектуального анализа данных.

**Предварительная обработка (препроцессинг):** Поскольку данные собираются из разных источников, они содержат несоответствия, для устранения которых на этом этапе выполняются различные действия, исправляются или удаляются дефектные данные, устраняются шумы и расхождения и данные из разных источников объединяются.

**Преобразование:** Здесь данные передаются в соответствующую форму для интеллектуального анализа данных. Могут использоваться отбор признаков, выборка, агрегирование.

**Интеллектуальный анализ данных:** это важный шаг, при котором выбирается алгоритм интеллектуального анализа данных, соответствующий шаблону в данных. Также осуществляется извлечение шаблонов данных.

**Интерпретация и оценка:** распознавать и преобразовывать результаты интеллектуального анализа данных или закономерности в знания путем устранения избыточности и нерелевантных закономерностей. Здесь используется множество стратегий визуализации и графического интерфейса для преобразования выгодных шаблонов в понятные человеку термины.

**Задачи data mining**

Задачи data mining сгруппированы в две основные категории:

1. Прогнозирующие

2. Описательные

Определяют следующие основные функции майнинга данных:

1. Классификация

2. Регрессия

3. Кластеризация

4. Моделирование зависимостей

5. Обнаружение отклонений

6. Обобщение

Классификация, регрессия и обнаружение отклонений отнесены к прогностической категории, в то время как кластеризация, моделирование зависимостей отнесены к описательной категории. Прогностическая модель составляет прогнозы, используя некоторую переменную в наборе данных, чтобы предсказать неизвестные значения другой релевантной переменной, в то время как описательная модель классифицирует закономерности или взаимосвязи и охватывает понятные человеку закономерности и тенденции в данных (Флорин Горунеску 2011).

**Классификация.** Основная идея алгоритмов классификации состоит в том, чтобы предсказать целевой класс путем анализа обучающего набора данных, а именно классифицировать данные по заданному количеству классов. Методы классификации используются для прогнозирования принадлежности к группе или классу (отсюда и название методов классификации) индивидов (данных) для предопределенного членства в группах, а также для описания того, какие характеристики индивидов могут предсказать их принадлежность к группе. В качестве примера такого алгоритма можно привести метод логистической регрессии.

*Логистическая регрессия* (Logistic Regression) - метод, используемый для моделирования вероятности отнесения события к одному из двух классов (на самом деле, классов может быть больше. В таком случае используется множественная логистическая регрессия). Первый класс назовем A, а второй B. Вероятности попадания некоторого значения в эти классы будут и соответственно. Идея метода заключается в том, что пространство исходных значений можно разделить линейной границей на две, соответствующие двум классам, области, как на рисунке 2.3.

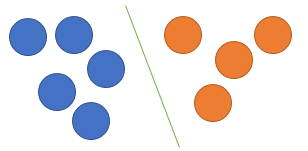


Рис. 2.3 «Множество А (слева), линейная граница и множество В (справа)»

Рассмотрим двумерный случай, как на рисунке. Тогда функция, характеризующая линию разделения, будет иметь вид:

При подстановке некоторого значения \ точки (a,b) получим

Тут возможны три варианта:

1. (a,b) лежит в области, ограниченной точками класса A. Тогда  положим положительной. С математической точки зрения, чем больше величина этого значения, тем больше расстояние между точкой и границей. А это означает большую вероятность того, что (a,b) принадлежит классу А. Следовательно, будет находиться в пределах (0,5, 1].
2. Аналогичным образом,  находится в интервале [0, 0.5), если (a,b) лежит в области, ограниченной точками класса B.
3. (a,b) лежит на самой границе. В этом случае,  и  = 0,5

Теперь, с помощью функции отношения шансов

И навешиванием над ней натурального логарифма, приходим к

, имеющей областью значений интервал (-). Теперь, если знать граничную функцию, можно вычислить вероятность попадания некоторого значения х = (a,b) в класс А. Для этого проделываем следующие этапы:

Метод обучается путем максимизации среднего значения функции

**Регрессия:** Это один из методов интеллектуального анализа данных, который определяет связь между зависимыми и независимыми переменными. Прогнозирование осуществляется с помощью поддержки регрессий. Статистически регрессия - это математическая модель, которая устанавливает связь между значениями зависимой переменной и значениями другого предиктора или независимой переменной. При регрессии прогнозируемая переменная может быть непрерывной переменной. При регрессии вещественнозначные прогнозирующие переменные сопоставляются с элементами обучающей функции. В качестве примеров таких методов так же можно привести логистическую регрессию и линейную регрессию.

**Кластеризация:** это метод интеллектуального анализа данных, который группирует физические или абстрактные объекты в классы похожих объектов . Кластеризация - это метод разделения набора данных (записей / кортежей / объектов/выборок) на несколько групп (кластеров) на основе предопределенного сходства. Основной целью кластеризации является нахождение групп (кластеров) объектов на основе сходства таким образом, чтобы внутри отдельного кластера было большое сходство друг с другом, в то время как кластеры достаточно отличаются друг от друга. В терминологии машинного обучения кластеризация - это форма неконтролируемого обучения.

**Моделирование зависимостей** (анализ ассоциативных правил): это один из лучших признанных методов интеллектуального анализа данных, который относится к неконтролируемому методу интеллектуального анализа данных, который направлен на поиск связей между элементами или записями, принадлежащими к большому набору данных, и выявляет существенные зависимости между переменными. Интеллектуальный анализ правил ассоциации является следствием формы X → Y, где x и y являются различными элементами или наборами элементов, содержащими инструкции if-then относительно значений атрибутов. При анализе рыночной корзины это правило широко используется, оно пытается проанализировать покупателей, покупающих определенные товары, и дает представление о сочетаниях, которые клиенты часто покупают вместе.

**Обнаружение аномалий:** синонимом его названия является обнаружение наиболее существенных изменений или отклонений от стандартного поведения.

**Обобщение:** Хотя это и не относится к методам интеллектуального анализа данных, но является результатом этих методов и имеет дело с определением компактного описания для подмножества данных, синонимично называемых обобщением или описанием.

**Последовательные шаблоны**: Обнаружение последовательности - это метод интеллектуального анализа данных, который используется для определения последовательных шаблонов, ассоциаций или регулярных событий / тенденций между переменными полями данных в течение бизнес-периода.

Интеллектуальный анализ данных хорошо развит, но он все еще сталкивается с целым рядом проблем при его практической реализации. Некоторые из них приведены ниже:

**Проблемы безопасности:** Безопасность является наиболее важной проблемой, касающейся любого процесса обработки данных, учитывая чрезвычайно конфиденциальный характер данных, потенциальный незаконный доступ к знаниям должен быть предотвращен, и тайна должна соблюдаться.

**Проблемы методологии интеллектуального анализа данных:** поскольку разные пользователи проявляют интерес к различным видам знаний, интеллектуальный анализ данных должен охватывать широкий спектр задач анализа данных и обнаружения знаний, которые могут по-разному использовать одну и ту же базу данных и требуют разработки многочисленных методов интеллектуального анализа данных.

**Проблемы с пользовательским интерфейсом**: Обнаружение знаний с помощью инструментов майнинга данных выгодно и выразительно только до тех пор, пока они представлены явно и привлекательно для пользователя. Поскольку трудно понять, что может быть обнаружено в базе данных, процесс интеллектуального анализа данных должен быть интерактивным, данные должны быть представлены на языке высокого уровня, в визуальных представлениях или других графических формах выражения, чтобы пользователь мог понять и интерпретировать их и использовать по мере необходимости.

**Обработка зашумленных и неполных данных:** Данные, хранящиеся в базах данных, могут отличаться, поскольку различные проблемы связаны с источниками данных, данные могут быть неполными, данные могут содержать случаи, которые могут вызывать исключения. Интеллектуальный анализ данных с такими нарушениями приводит к возникновению неоднозначностей в процессе, в результате чего построенная модель знаний перегружает данные и снижает точность результирующих знаний, поэтому требуются методы интеллектуального анализа данных, которые могут справиться с этими несоответствиями.

**Проблемы с производительностью:** Для обработки данных эффективность и масштабируемость являются стратегическими факторами при внедрении data mining’a баз данных. Информация должна эффективно и квалифицированно извлекаться из баз данных, поскольку их объем достаточно велик. Используемые алгоритмы должны быть масштабируемыми, время их выполнения должно быть предсказуемым и приемлемым для больших баз данных.

Ниже приведем примеры применения технологий data minin:

**Применение в здравоохранении:** data mining / интеллектуальный анализ данных может быть существенно выгоден в системе здравоохранения, но его успех зависит от наличия чистых данных. В здравоохранении он используется для диагностики и прогноза заболевания, также может быть установлена взаимосвязь между заболеваниями. Врачи могут определить эффективные методы лечения для каждого пациента. Поскольку огромное количество медицинских данных является сложным и обширным для обработки и анализа, data mining предоставляет методологию и инструменты для преобразования данных в информацию для принятия эффективных решений.

**Применение в образовательных системах:** Интеллектуальный анализ данных в образовательных системах - развивающаяся область, к которой исследователи проявляют большой интерес. Методы интеллектуального анализа данных могут помочь восполнить пробел в знаниях в системе образования путем выявления скрытых закономерностей, коннотаций и различий. Это помогает заинтересованным сторонам повысить эффективность принятия решений, что приводит к совершенствованию образовательной системы.

**Применение в CRM**: Интеллектуальный анализ данных в CRM в настоящее время является наиболее обсуждаемой темой исследований в промышленности и академических кругах с целью предоставления резюме исследований по использованию методов интеллектуального анализа данных в области CRM.

**Применение при анализе рыночной корзины (MBA**): Для анализа рыночной корзины используются различные методы интеллектуального анализа данных -MBA, метод, который помогает выявить связь между различными товарами, которые покупатель кладет в свою корзину во время покупок, он отслеживает покупательские привычки покупателей. Торговые дома могут использовать методы сбора данных для выявления моделей покупок и поведения клиента, на основе которых клиенту может быть представлен широкий выбор в соответствии с его привычкой к покупкам.

**Применение в спортивных данных:** Методы интеллектуального анализа данных также проникли в сферу спорта. Проводится огромное количество игр, и каждый вид спорта генерирует огромное количество статистических данных. Эти массивные данные необходимо поддерживать в отношении планирования событий и статистики отдельного игрока в этих событиях. Интеллектуальный анализ данных может использоваться для прогнозирования и анализа производительности, а также для стратегического планирования.

# **2.2 Основы вычислительных нейронных сетей**

Для начала введем определение вычислительных нейронных сетей. Вычислительные нейронные сети — это модель машинного обучения, вдохновленная структурой и функциями человеческого мозга. Они широко используются в различных областях, таких как компьютерное зрение, обработка естественного языка и распознавание речи, для выполнения сложных задач, таких как классификация изображений, языковой перевод и распознавание речи. В этом параграфе показаны основные принципы вычислительных нейронных сетей, включая их архитектуру, процесс обучения и различные типы нейронных сетей, обычно используемые на практике. Понимание этих основ необходимо для создания эффективных моделей нейронных сетей для широкого спектра приложений.

Рассмотрим подробнее архитектуру нейронных сетей. Перед этим вкратце будет описано устройство биологической нейронной сети. Биологическая нейронная сеть состоит из огромного количества взаимосвязанных клеток, называемых нейронами, которые взаимодействуют друг с другом через специализированные структуры, называемые синапсами. Каждый нейрон обычно имеет тело клетки, дендриты и аксон. Дендриты получают электрические импульсы от других нейронов, а аксон передает этот импульс дальше другим нейронам. Эта связь между нейронами позволяет обрабатывать и передавать информацию по всему мозгу и нервной системе. Сложность и разнообразие нейронных связей и сетей позволяют мозгу выполнять широкий спектр функций, включая восприятие, память и поведение.

Аналогично биологической нейронной сети, искусственная нейронная сеть состоит из большого количества взаимосвязанных узлов, называемых искусственными нейронами, которые обрабатывают входные данные и вырабатывают выходные сигналы.

Базовая архитектура нейронной сети состоит из трех типов слоев: входного слоя, скрытого слоя и выходного слоя. Входной уровень является первым уровнем в нейронной сети и отвечает за прием входных данных, которые необходимо обработать. Количество нейронов во входном слое соответствует количеству входных объектов в данных. Например, в задаче классификации изображений входной слой будет иметь нейрон для каждого пикселя изображения. Пример с классификацией изображений будет раскрываться позднее.

Скрытые слои нейронной сети — это то место, где происходит большая часть вычислений. Каждый скрытый слой состоит из набора нейронов, которые получают входные данные от предыдущего слоя и применяют ряд математических операций для преобразования входных данных. Эти преобразования определяются весами и смещениями, связанными с каждым нейроном и соединением. Количество скрытых слоев и нейронов в каждом слое зависит от сложности задачи и объема данных, доступных для обучения.

Скрытые уровни называются "скрытыми", потому что их выходные данные непосредственно не наблюдаемы, в отличие от входов и выходов сети. Выходные данные скрытых слоев передаются следующему слою до тех пор, пока не будет создан последний слой, который называется выходным. Выходной слой создает конечный результат работы сети, который может быть меткой класса, распределением вероятностей или числовым значением, в зависимости от задачи.

Количество нейронов в выходном слое зависит от типа выходных данных, необходимых для выполнения задачи. Например, в задаче двоичной классификации выходной уровень будет иметь один нейрон, который выдает значение от 0 до 1, указывающее на вероятность принадлежности входных данных к классу. В задаче многоклассовой классификации выходной слой будет содержать несколько нейронов, каждый из которых соответствует отдельному классу. Например, в задаче распознавания цифр десятичной системы счисления, на выходном слое будет 10 нейронов: по одному на каждую цифру.

Преобразования, которые происходят в скрытых слоях, имеют решающее значение для того, чтобы сеть научилась распознавать закономерности во входных данных и производить точные прогнозы. Связи между нейронами в нейронной сети моделируются по образцу синапсов в биологической нейронной сети, где электрические импульсы передаются между нейронами. В нейронной сети каждое соединение между нейронами имеет связанный с ним вес, который определяет прочность соединения (в некоторых источниках насчет этого проводят параллели с биологической нейронной сетью: нейрон передает импульс связанным с ними нейронам с разной силой. Эту силу как раз и характеризует вес). Веса корректируются в процессе обучения, чтобы повысить производительность сети. Процесс корректировки весов называется обучением нейронной сети.

Теперь опишем понятие функции активации. Функция активации каждого нейрона определяет, будет ли он "срабатывать" и выдавать выходной сигнал, основываясь на получаемых им входных данных. Функции активации применяются к выходному сигналу каждого нейрона после того, как он был умножен на соответствующий ему вес и суммирован с членом смещения. Результирующее значение затем передается через функцию активации, которая выдает конечный выходной сигнал нейрона.

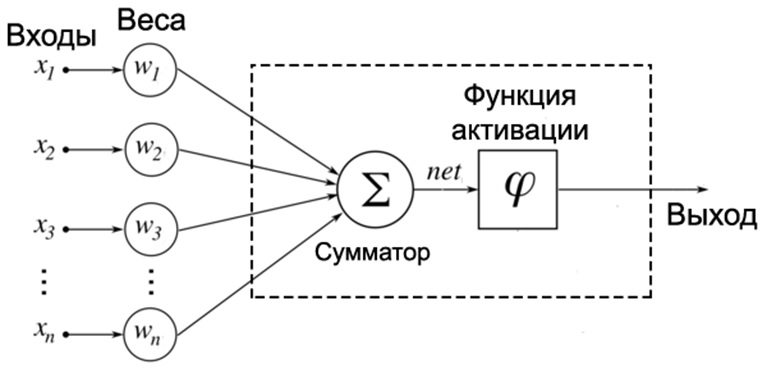
Схематично, описываемая математическая модель нейрона выглядит так: 

Рис. 2.4 «Математическая модель нейрона»

Приведем пример простой нейронной сети. Она будет состоять из одного входного слоя, одного выходного слоя и одного скрытого слоя. Раньше такая модель называлась бы трехслойной, однако сейчас принято считать только число скрытых слоев. Итак, имеем однослойную нейронную сеть. Пусть на входном слое 6 нейронов: 5 “классических” и 1 нейрон смещения (константа, как правило, равная 1). На скрытом слое, положим, 3 нейрона: 2 обычных и 1 нейрон смещения. На выходном слое 2 нейрона. Ниже схематически показана подобная сеть, где у каждого нейрона есть пара индексов: нижний показывает условный порядковый номер нейрона, а верхний – номер слоя, которому данный нейрон принадлежит. Каждый нейрон следующего слоя связан со всеми нейронами предыдущего слоя. Каждая такая связь имеет свой вес, который условно обозначим где i отвечает за номер исходящего слоя, j – за номер исходящего нейрона, k – за номер входящего нейрона.

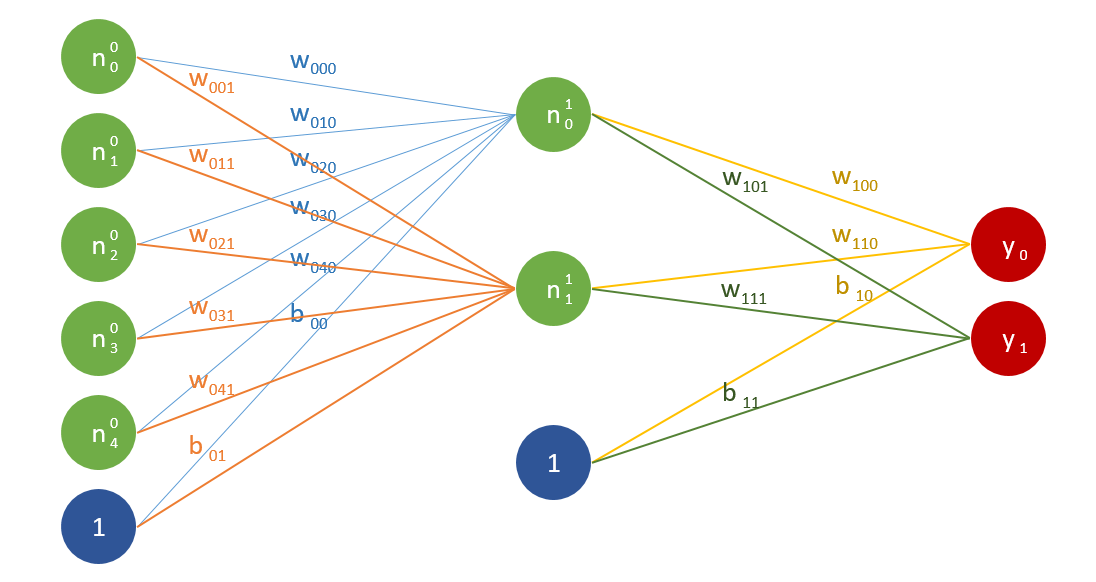
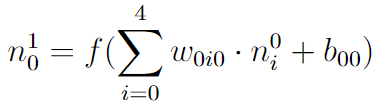
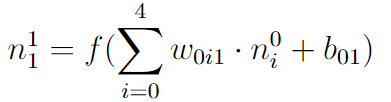


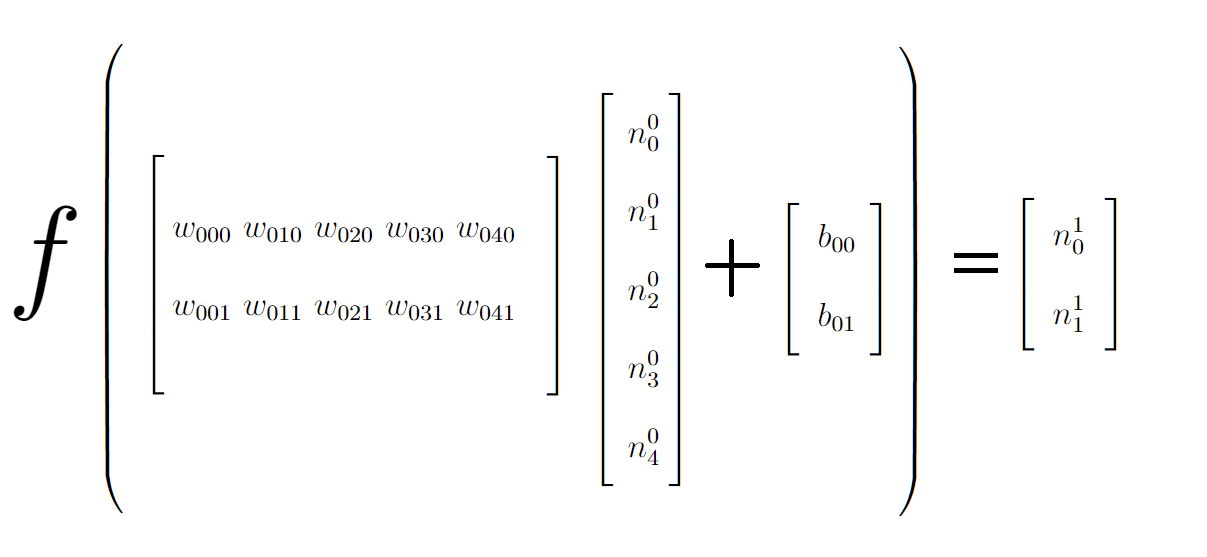
Рис. 2.5 «Математическая модель нейронной сети»

Математическая запись нейронов второго уровня будет иметь следующий вид:

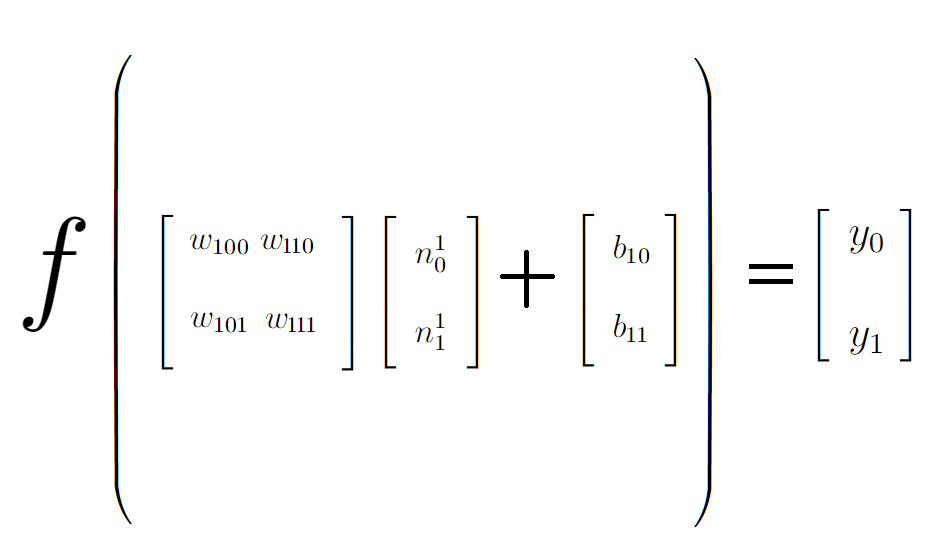




В терминах линейной алгебры запись выглядит так:



Аналогично для нейронов выходного слоя:



Вышеописанные математические выкладки называются прямой трансляцией или feedforward или алгоритмом прямого распространения.

Теперь посмотрим, что будет, если взять в качестве функции f активации линейную функцию. В таком случае, очевидно, получим суперпозицию линейных преобразований между входным слоем и выходным, что, в свою очередь, лишает смысла наличие скрытого слоя, ведь можно обойтись только упомянутыми двумя. Поэтому в качестве функции активации принято использовать нелинейные функции, такие как следующие:

**Сигмоид**: Сигмоидальная функция выдает выходные данные в диапазоне от 0 до 1, что полезно для задач двоичной классификации. Функция определяется как

**ReLU**: Функция ReLU выдает значение 0 для отрицательных входных значений и само входное значение для положительных входных значений. Функция определяется как

**Softmax**: Функция определяется как

где i - индекс текущего нейрона, а K – количество всех нейронов в выходном слое.

Теперь ответим на вопрос “как же обучаются нейронные сети?”. Обучение нейронной сети заключается в минимизации отклонения выходных результатов от «правильных» результатов. Если нейросеть обучается на наборе из входных и заведомо известных «правильных» выходных данных, такой процесс обучения называют обучение с учителем или контролируемый процесс обучения. Помимо этого, существуют другие подходы в обучении нейросетей. Например:

* Неконтролируемое обучение: при неконтролируемом обучении нейронная сеть обучается на данных без каких-либо явных меток или указаний. Вместо этого сеть пытается самостоятельно найти закономерности и структуру во входных данных.
* Обучение с подкреплением: при обучении с подкреплением нейронная сеть обучается, получая обратную связь в виде вознаграждений или наказаний за определенные действия, предпринятые в окружающей среде. Сеть пытается выработать стратегию, которая максимизирует ожидаемое вознаграждение с течением времени.
* Трансфертное обучение: это метод, при котором предварительно обученная сеть настраивается под новую задачу с меньшим объемом данных, что может ускорить обучение.

Отклонение полученных результатов от идеальных характеризует функция ошибки.

Она имеет следующий вид:

Где n – это число выходных нейронов – 1.

Задача обучения нейронной сети сводится к задаче поиска минимума функции ошибки, то есть к задаче вида:

Эта задача может быть решена с помощью метода градиентного спуска

Градиент – это вектор частных производных:



Его геометрический смысл состоит в том, что он направлен в сторону скорейшего возрастания функции. Метод градиентного спуска опишем на упрощенном примере, когда вес всего один. В таком случае, пусть график зависимости функции ошибки от выбранного веса будет иметь следующий вид:

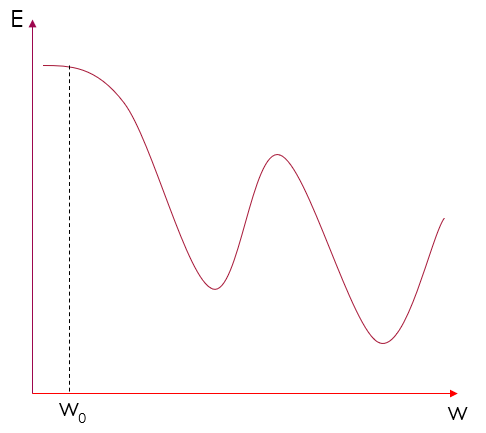
Пусть будет начальной инициализацией веса.

Рис. 2.6 «График зависимости функции ошибки от веса»

Цель – достигнуть наинижайшей точки графика.

Для этого воспользуемся итерационной формулой следующего вида:

То есть, новое значение веса получается вычитанием некоторой константы помноженной на значение градиента в точке из . Константа характеризует скорость «скачков». С слишком большим альфа можно «проскочить» искомую точку, а со слишком малым метод будет работать долго. Метод остановится, когда мы попадем в -окрестность нуля, где —заранее заданная точность. В многомерном случае, этот алгоритм проделывается для каждого веса.

В терминах исследований нейронных сетей, три рассмотренных процесса (прямое распространение, вычисление ошибки, обратное распространение) составляют алгоритм backpropagation.

**Типы нейронных сетей**

Существует множество различных типов нейронных сетей, каждая из которых предназначена для решения определенных задач. В данном параграфе мы рассмотрим основные типы нейронных сетей, их особенности и области применения. Основными типами нейронных сетей считают следующие:

* Сети прямого распространения (Feedforward Neural Networks)
* Сети обратного распространения (Backpropagation Neural Networks)
* Рекуррентные нейронные сети (Recurrent Neural Networks)
* Сверточные нейронные сети (Convolutional Neural Networks)

и др.

**Сети прямого распространения –** это самый простой вид нейронных сетей: сигнал передается только в одном направлении - от входного слоя к выходному, поэтому такие сети также называются однонаправленными нейронными сетями. Несмотря на название, сети прямого распространения также обучаются алгоритмом обратного распространения ошибки. Сети прямого распространения широко применяются для решения задач классификации и регрессии, в том числе для распознавания рукописных цифр, анализа текстов, распознавания образов, предсказания цен на акции и много других. Схематично они выглядят следующим образом:

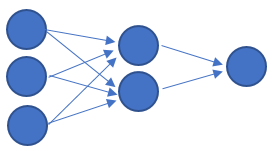
****

Рис. 2.7 «ИНС прямого распространения»

**Сети обратного распространения**, наоборот, являются двунаправленными сетями, где данные передаются как вперед, так и назад. Сети обратного распространения имеют широкий спектр применений, включая распознавание образов, прогнозирование временных рядов, классификацию и регрессионный анализ. Одним из главных преимуществ этого типа нейронных сетей является их способность обучаться на больших объемах данных и быстро адаптироваться к изменениям в данных. Обобщенная схема их устройства представлена ниже:

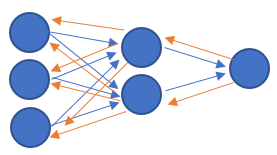


Рис. 2.8 «ИНС обратного распространения»

**Рекуррентные нейронные сети** являются классом нейронных сетей, которые используются для анализа последовательностей данных, таких как тексты, аудио и видео. Для этого в сетях имеются циклические связи, позволяющие сохранять информацию о предыдущих состояниях сети. Процесс обучения рекуррентных нейронных сетей может быть осуществлен методом Backpropagation Through Time (BPTT) - метод обратного распространения ошибки во времени. Рекуррентные нейронные сети широко применяются в различных областях, таких как обработка естественного языка, распознавание речи, машинный перевод, генерация текста, прогнозирование временных рядов и др. Одним из наиболее популярных типов рекуррентных нейронных сетей являются LSTM (Long Short-Term Memory) сети, которые позволяют эффективно управлять и сохранять информацию в циклических связях. Схематично они выглядят так:

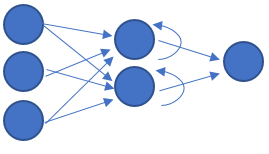


Рис. 2.9 «Рекуррентная ИНС»

**Сверточные нейронные сети**. Основная идея СНС заключается в том, чтобы использовать свертку для обнаружения локальных признаков во входных данных. Свертка — это операция, при которой ядро (набор весов) перемещается по входным данным, выполняя умножение каждого элемента ядра на соответствующий элемент входных данных и суммируя полученные произведения. Это позволяет выделить локальные признаки, такие как границы или углы, которые могут быть использованы для классификации. Связывая несколько сверточных слоев, сеть может обнаруживать более абстрактные и сложные признаки, построенные на базе локальных признаков. После сверточных слоев обычно следуют слои объединения, которые уменьшают размерность данных, сохраняя при этом информацию о наиболее важных признаках. Далее следуют полносвязные слои, которые используются для классификации или выполнения других задач. Полносвязные слои связывают выходы предыдущих слоев с каждым нейроном в следующем слое, что позволяет вычислять вероятности различных. Сверточные нейронные сети применяются в распознавании изображений, классификации текста, распознавании речи, обработке естественного языка и др. Схематически, СНС отличается от сетей прямого распространения наличием сверточных слоев.

# **Глава 3**

# **3.1 Обзор программного обеспечения data mining**

В современном мире, основанном на данных, предприятия и организации постоянно собирают и анализируют огромные объемы данных, чтобы получить представление и принять обоснованные решения. Интеллектуальный анализ данных (data mining) - это процесс выявления закономерностей и взаимосвязей в больших наборах данных, который может помочь компаниям выявлять скрытые идеи и делать прогнозы. В связи с растущим спросом на аналитику, основанную на данных, появилось множество программных средств интеллектуального анализа данных, которые помогают аналитикам данных, специалистам по обработке данных и исследователям в их анализе. В этом параграфе будут рассмотрены некоторые из наиболее популярных программных средств интеллектуального анализа данных, доступных на сегодняшний день, их функции, преимущества и недостатки.

Вот некоторые из наиболее популярных программных инструментов для data mining:

* RapidMiner
* SAS
* IBM SPSS Modeler
* Oracle Data Mining
* Knime
* Microsoft SQL Server Analysis Services
* Weka
* Orange
* Tableau
* Alteryx

Каждый из этих инструментов обладает своими уникальными функциями и возможностями, и выбор подходящего инструмента зависит от конкретных потребностей в анализе данных. Среди программных инструментов, упомянутых в списке, RapidMiner, SAS, IBM SPSS Modeler, Oracle Data Mining, Knime, Weka и Alteryx предоставляют множество алгоритмов и методов для предварительной обработки и анализа данных, построения прогностических моделей и выполнения других задач интеллектуального анализа данных.

С другой стороны, Tableau и Orange — это программные инструменты, которые обеспечивают расширенные возможности визуализации и создания информационных панелей. Эти инструменты позволяют пользователям создавать интерактивные визуализации и отчеты, которые помогают передавать информацию из данных в простом для понимания формате. Их также можно использовать для исследовательского анализа данных, очистки данных и других задач обработки данных, но их основное внимание уделяется визуализации и обмену данными. Благодаря этому обзору вы сможете получить представление о функциях и преимуществах каждого инструмента и принять обоснованное решение о том, какой из них лучше всего подходит для конкретных задач интеллектуального анализа данных. Рассмотрим некоторые из них.

**RapidMiner** - популярное программное обеспечение для интеллектуального анализа данных с открытым исходным кодом, предоставляющее ряд функциональных возможностей для исследования данных, предварительной обработки, моделирования, оценки и развертывания. Он предназначен для того, чтобы помочь пользователям извлекать полезную информацию и прогнозные модели из больших наборов данных, даже если у них нет опыта работы с данными.

Удобный интерфейс RapidMiner позволяет пользователям создавать рабочие процессы обработки данных с помощью интерфейса перетаскивания. Он поддерживает различные форматы данных и предоставляет широкий спектр инструментов предварительной обработки данных и моделирования, включая классификацию, кластеризацию, регрессию, изучение ассоциативных правил и интеллектуальный анализ текста.

Одной из ключевых особенностей RapidMiner является его способность интегрироваться с другими инструментами и языками обработки данных, включая Python, R и SQL. Это позволяет пользователям использовать сильные стороны различных инструментов и языков и создавать более сложные рабочие процессы и модели.

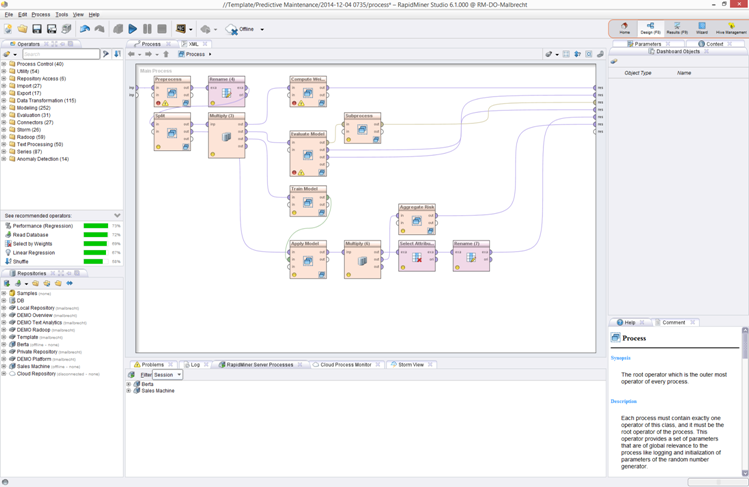


Рис.3.1 «GUI RapidMiner»

Теперь перейдем непосредственно к преимуществам и недостаткам этого ПО.

**Преимущества:**

* **Удобный интерфейс:** RapidMiner имеет простой GUI (Graphic User Interface), который позволяет пользователям легко создавать рабочие процессы и модели обработки данных, не требуя навыков программирования.
* **Открытый исходный код:** RapidMiner – это инструмент с открытым исходным кодом, что означает, что он находится в свободном доступе и может быть изменен в соответствии с конкретными потребностями.
* **Интеграции:** RapidMiner хорошо интегрируется с другими инструментами и языками data science, включая Python, R и SQL, что позволяет пользователям использовать сильные стороны различных инструментов и языков.
* **Обширная функциональность:** RapidMiner предоставляет широкий спектр инструментов предварительной обработки данных и моделирования, включая классификацию, кластеризацию, регрессию, изучение ассоциативных правил и интеллектуальный анализ текста.
* **Варианты развертывания:** RapidMiner предоставляет несколько вариантов развертывания, включая возможность развертывания моделей в виде веб-служб, пакетных процессов или потоков реального времени.

**Недостатки:**

* **Ограниченная поддержка:** бесплатная версия RapidMiner довольно ограничена.
* **Ограниченная документация:** Документация для RapidMiner может быть скудной, что затрудняет пользователям освоение некоторых более продвинутых функций.
* **Ограниченная масштабируемость:** RapidMiner с трудом справляется с очень большими наборами данных или моделями высокой сложности, что ограничивает его масштабируемость.
* **Ограниченная визуализация:** хотя RapidMiner предоставляет базовые параметры визуализации, он не обладает расширенными возможностями визуализации других инструментов, таких как, например, упомянутые ранее Tableau или Orange.
* **Ограниченный контроль:** Интерфейс «перетаскивания» может быть ограничен для пользователей, которым требуется детальный контроль над рабочими процессами обработки данных и моделирования.

**SAS** (Statistical Analysis System) – это ПО для интеллектуального анализа данных, который широко используется в таких отраслях, как финансы, здравоохранение и государственное управление. Одной из ключевых особенностей SAS является его способность обрабатывать структурированные и неструктурированные данные из различных источников, включая базы данных, электронные таблицы и текстовые файлы. Он предоставляет широкий спектр инструментов для обработки и подготовки данных, таких как очистка, преобразование и агрегирование данных, а также поддерживает различные статистические методы и методы машинного обучения для моделирования и анализа.

SAS также предоставляет гибкий язык программирования SAS Language, который позволяет пользователям писать пользовательские скрипты и автоматизировать обработку данных и моделирование рабочих процессов. Это может быть особенно полезно для продвинутых пользователей, которым требуется детальный контроль над рабочими процессами обработки данных и моделирования.

Помимо этого, SAS, так же, как и RapidMiner, имеет способность интегрироваться с другими инструментами и языками data science, включая Python и R. Это позволяет пользователям использовать сильные стороны различных инструментов и языков и создавать более сложные рабочие процессы и модели.

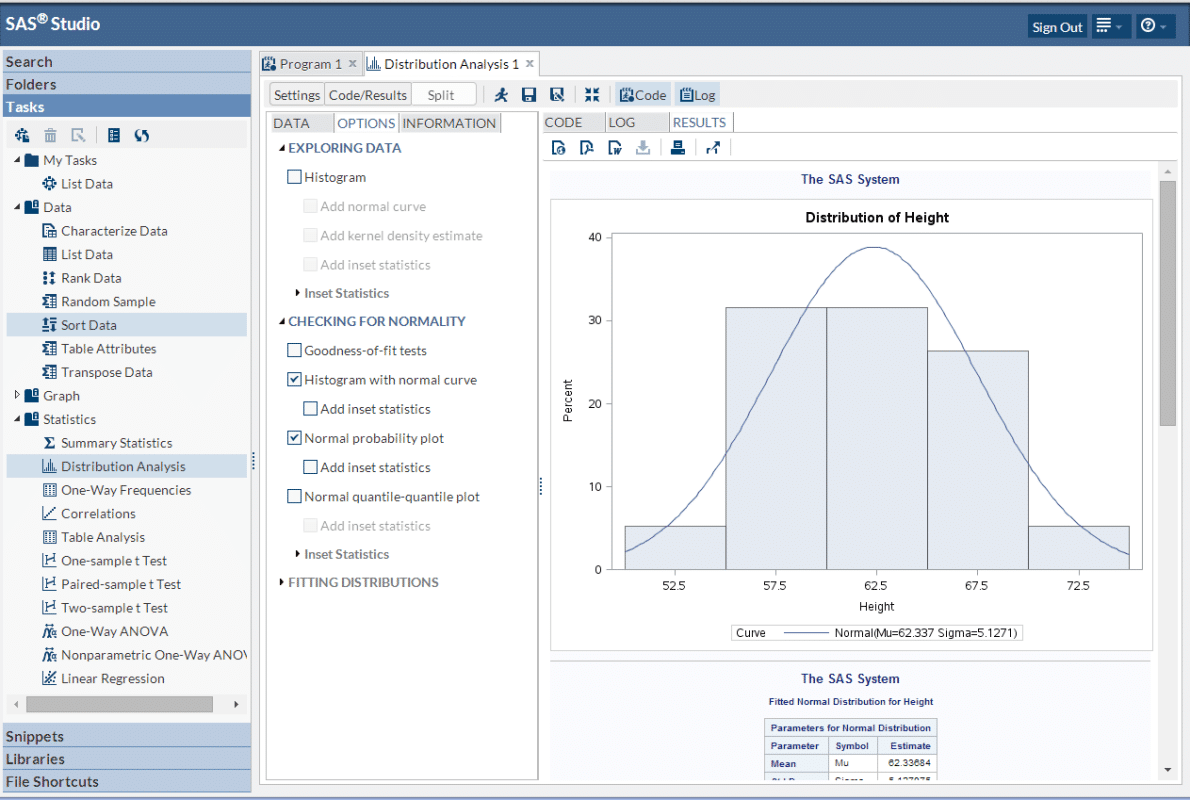


Рис. 3.2 «GUI SAS»

**Преимущества**

* **Всесторонняя функциональность:** SAS предоставляет широкий спектр функциональных возможностей для обработки, анализа и моделирования данных, включая поддержку структурированных и неструктурированных данных.
* **Гибкость:** SAS предоставляет гибкий язык программирования SAS Programming Language, который позволяет пользователям писать пользовательские сценарии и автоматизировать процессы обработки данных и моделирования.
* **Интеграция:** SAS хорошо интегрируется с другими инструментами и языками data science, включая Python и R, что позволяет пользователям использовать сильные стороны различных инструментов и языков.
* **Большое сообщество пользователей:** у SAS есть большое и активное сообщество пользователей, которое предоставляет пользователям поддержку и ресурсы.
* **Высокая производительность:** SAS предназначен для обработки больших и сложных наборов данных и оптимизирован для высокопроизводительных вычислительных сред.

**Недостатки:**

* **Сложность освоения**: особенно для пользователей, которые не знакомы с программированием или статистическим анализом.
* **Ограниченная настройка:** хотя SAS предоставляет широкий спектр функциональных возможностей, может быть сложно настроить рабочие процессы или модели за пределами предоставленной функциональности.
* **Ограниченная визуализация:** SAS предоставляет базовые параметры визуализации, но не обладает расширенными возможностями визуализации других инструментов, таких как Tableau или Orange.
* **Ограниченная масштабируемость:** SAS может работать с очень большими наборами данных или моделями высокой сложности, что может ограничить его масштабируемость.
* **Проприетарность:** SAS является проприетарным инструментом, что означает, что пользователи ограничены функциями, предоставляемыми программным обеспечением, и не могут изменять или расширять его функциональность.

**IBM SPSS Modeler** широко используется в таких отраслях, как финансы, здравоохранение и маркетинг. Главным преимуществом IBM SPSS Modeler, как и RapidMiner, является его интуитивно понятный интерфейс, который позволяет пользователям легко создавать и изменять рабочие процессы обработки данных и моделирования, не требуя навыков программирования. Так же как и интерфейс RapidMiner, интерфейс IBM SPSS Modeler предоставляет интерфейс перетаскивания, который позволяет пользователям создавать сложные рабочие процессы простым перетаскиванием готовых узлов моделирования на рабочее пространство. IBM SPSS Modeler также предоставляет широкий спектр инструментов для обработки и подготовки данных, таких как очистка, преобразование и агрегирование данных, а также поддерживает различные статистические методы и методы машинного обучения для моделирования и анализа. Как и вышеупомянутые инструменты, SPSS Modeler так же поддерживает интеграцию с R и Python для расширения диапазона доступных алгоритмов. Еще одной примечательной особенностью IBM SPSS Modeler является его способность развертывать модели в виде веб-служб или пакетных процессов, что позволяет пользователям интегрировать свои модели в более крупные бизнес-процессы или приложения. Из главных недостатков можно отметить достаточную дороговизну этого инструмента и ограниченные возможности масштабируемости.

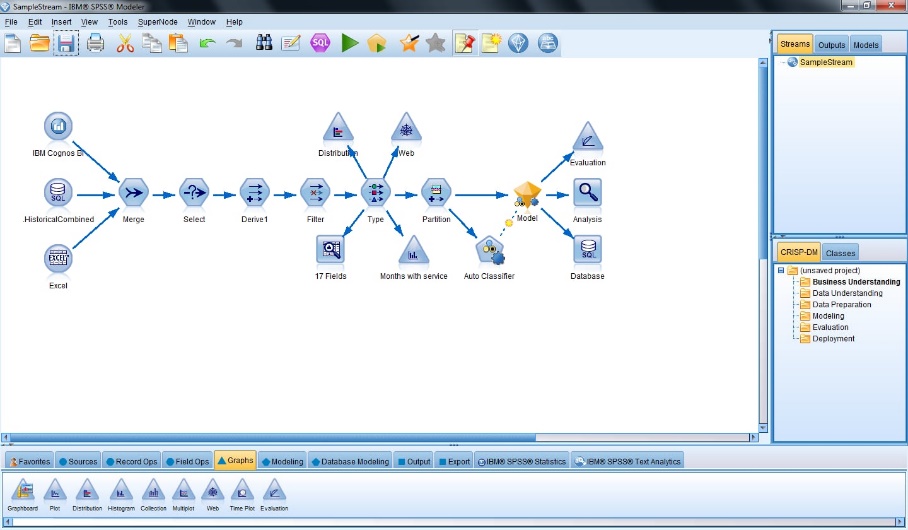


Рис. 3.3 «GUI IBM SPSS Modeler»

**Преимущества:**

* **Удобный интерфейс:** IBM SPSS Modeler предоставляет удобный интерфейс, который позволяет пользователям легко создавать и изменять рабочие процессы обработки данных и моделирования, не требуя навыков программирования.
* **Всесторонняя функциональность:** IBM SPSS Modeler предоставляет широкий спектр функциональных возможностей для обработки, анализа и моделирования данных, включая поддержку структурированных и неструктурированных данных.
* **Интеграция:** IBM SPSS Modeler хорошо интегрируется с другими инструментами и языками data science, включая R и Python, что позволяет пользователям использовать сильные стороны различных инструментов и языков.
* **Варианты развертывания:** IBM SPSS Modeler предоставляет варианты развертывания моделей в виде веб-служб или пакетных процессов, что позволяет пользователям интегрировать свои модели в более крупные бизнес-процессы или приложения.
* **Большое сообщество пользователей**: IBM SPSS Modeler располагает большим и активным сообществом пользователей, которое предоставляет пользователям поддержку и ресурсы.

**Недостатки:**

* **Стоимость**: IBM SPSS Modeler довольно дорогой, что ограничивает его доступность для небольших организаций или отдельных лиц.
* **Ограниченная масштабируемость:** IBM SPSS Modeler, так же как и SAS, может работать с очень большими наборами данных или моделями высокой сложности, что может ограничить его масштабируемость.
* **Ограниченная настройка**: хотя IBM SPSS Modeler предоставляет широкий спектр функциональных возможностей, может быть сложно настроить рабочие процессы или модели за пределами предоставленной функциональности.
* **Ограниченная визуализация:** IBM SPSS Modeler предоставляет базовые параметры визуализации, но они довольно скудны в своей функциональности.
* **Проприетарность:** IBM SPSS Modeler является проприетарным инструментом, в том же смысле, что и SAS.

**Oracle Data Mining** - это ПО для data mining, который является частью инструмента Oracle Advanced Analytics для Oracle Database. Oracle Data Mining имеет возможность интеграции с Oracle Database, которая позволяет пользователям получать доступ к большим и сложным наборам данных и анализировать их непосредственно из базы данных. Эта интеграция также предоставляет ряд функций подготовки и обработки данных, таких как очистка, преобразование и агрегирование данных. Oracle Data Mining поддерживает как контролируемые, так и неконтролируемые методы обучения и предоставляет инструменты для выбора и оценки моделей. Еще одной примечательной особенностью Oracle Data Mining является ее способность развертывать модели в виде функций SQL, что позволяет пользователям интегрировать свои модели в существующие приложения баз данных и рабочие процессы. Графический интерфейс, тем не менее, не является user-friendly, что требует достаточно времени и опыта от пользователей. Помимо этого, функциональность Oracle Data Mining довольно негибкая, что ограничивает пользователей в настройках под конкретные задачи.

**Преимущества:**

* **Интеграция с Oracle Database**: Oracle Data Mining предназначена для бесперебойной работы с Oracle Database.
* **Эффективная обработка:** Oracle Data Mining оптимизирован для крупномасштабного интеллектуального анализа данных и аналитики и предназначен для эффективной работы на многоядерных процессорах и кластерах.
* **Варианты развертывания моделей:** Oracle Data Mining предоставляет варианты развертывания моделей в виде функций SQL, что позволяет пользователям интегрировать свои модели в существующие приложения баз данных и рабочие процессы.
* **Поддержка и ресурсы:** Oracle предоставляет всестороннюю поддержку и ресурсы для Oracle Data Mining, включая документацию, учебные пособия и форумы.

**Недостатки:**

* **Сложность:** Oracle Data Mining может быть сложным в использовании для пользователей, которые не знакомы с Oracle Database или SQL, и для его настройки могут потребоваться дополнительные ресурсы и опыт.
* **Стоимость**: Oracle Data Mining является частью опции Oracle Advanced Analytics, которая стоит достаточно дорого и может потребовать дополнительных лицензионных сборов.
* **Ограниченная интеграция:** Хотя Oracle Data Mining разработан для бесперебойной работы с Oracle Database, интеграция с другими источниками данных или инструментами за пределами экосистемы Oracle может оказаться сложной.

Также к минусам можно отнести проприетарность ПО и ограниченность визуализации.

# **3.2 Решение практической задачи**

Прогнозирование роста регистрации автомобилей

# **ЛИТЕРАТУРА**

1. Brown, M. M., & Brudney, J. L. (2003). Learning organizations in the public sector? A study of police agencies employing information and technology to advance knowledge. Public Administration Review, 63(1), 30−43.
2. Gant, J. P., & Gant, D. B. (2002). Web Portal Functionality and State Government E-Services, Proceedings of the 35th Hawaii International Conference on Systems Sciences, 2002.
3. Bellamy, C., & Taylor, A. J. (1998). Governing in the information age. Buckingham: Open Univ. Press.
4. Danziger, J. N., Dutton, W. H., Kling, R., & Kraemer, K. L. (1982). Computers and politics: High technology in American local governments. New York: Columbia Univ. Press.
5. Doty, P., & Erdelez, S. (2002). Information micro-practices in Texas rural courts: Methods and issues for Egovernment. Government Information Quarterly, 19, 369−387.
6. Danziger, J. N. (2004). Innovation in innovation: The technology enactment framework. Social Science Computer Review, 22(1), 100−110.
7. Bauer, C., & Scharl, A. (2000). Quantitative evaluation of web site content and structure. Internet Research: Electronic Networking Applications and Policy, 10(1), 31−43.
8. Bekkers, V. J. J. M., & Zouridis, S. (1999). Electronic service delivery in public administration: Some trends and issues. International Review of Administrative Sciences, 65(2), 183−196.
9. Cohen, S., & Eimicke, W. (2001). The use of Internet in government service delivery. In M. Abramson & G. E. Means (Eds.), E-Government 2001. The Pricewaterhouse-Coopers endowment for the business of government (pp. 9−43). Oxford: Rowman & Littlefield Publishers, Inc.
10. Aldrich, J., Bertot, J. C., & McClure, C. R. (2002). E-government: initiatives, developments, and issues. Government Information Quarterly, 19, 349−355.
11. Bozeman, B. (1993). Introduction. In Barry Bozeman (Ed.), Public Management: State of the Art (pp. 275−293). San Francisco: Jossey-Bass.
12. Bretschneider, S. (2003). Information technology, e-government and institutional change. Public Administration Review, 63(6), 738−741.
13. Cullen, R., & Houghton, C. (2000). Democracy online: An assessment of New Zealand government Web sites. Government Information Quarterly, 17(3), 243−267.
14. Garson, G. D. (1999). Information systems, politics, and government: Leading theoretical perspectives. In G. D. Garson (Ed.), Handbook of public information systems (pp. 591−605). New York: Marcel Dekker.
15. Bozeman, B., & Bretschneider, S. (1986). Public management information systems: Theory and prescription. Public Administration Review, 46, 475−487.
16. DiCaterino, A., & Pardo, T. A. (1996). The World Wide Web as a universal interface to government services. Available at:
17. Heeks, R. (2003, January). Personal e-mail correspondence.
18. Ho, A. T. (2002). Reinventing local governments and the e-government initiative. Public Administration Review, 62(4), 434−441.
19. Kraemer, K. L., & Northop, A. (1989). Curriculum recommendations for public management education in computing: An update. Public Administration Review, 49(5), 447−453.
20. Grafton, C. (2003). ‘Shadow theories’ in Fountain's theory of technology enactment. Social Science Computer Review, 21(4), 411−416.
21. Hwang, S. D., Choi, Y., & Myeong, S. H. (1999). Electronic government in South Korea: Conceptual problems. Government Information Quarterly, 16(3), 277−285.
22. Holden, S. H. (2003). The evolution of information technology management at the federal level: Implications for public administration. In G. D. Garson (Ed.), Public information technology: Policy and management issues (pp. 53−73). Hershey, PA: Idea Group Publishing.
23. Gil-Garcia, J. R., & Pardo, T. A. (2005). E-government success factors: Mapping practical tools to theoretical foundations. Government Information Quarterly, 22, 187−216.
24. Jaeger, P. T. (2002). Constitutional principles and e-government: An opinion about possible effects of Federalism and separation of powers on e-government policies. Government Information Quarterly, 19, 357−368.
25. Jonas, D. K. (2000). Building state information highways: Lessons for public and private sector leaders. Government Information Quarterly, 17(1), 43−67.
26. La Porte, T. M., De Jong, M., & Demchak, C. C. (1999). Public organizations on the World Wide Web: Empirical correlates of administrative openness. Available at: http://www.cyprg.arizona.edu/publications/correlat.rtf. Accessed June 17, 2001.
27. Heeks, R. (1999b). Public sector accountability: Can IT deliver? Available at: http://idpm.man.ac.uk/idpm/egovacc.htm. Accessed March 29, 2003.
28. Goodsell, C. T. (1997). Administration as ritual. International Journal of Public Administration, 20(4−5), 939−961.
29. Grant, G., & Chau, D. (2005). Developing a generic framework for e-government. Journal of Global Information Management, 13(1), 1−30.
30. Halchin, L. E. (2004). Electronic government: Government capability and terrorist resource. Government Information Quarterly, 21, 406−419.
31. Jaeger, P. T. (2003). The endless wire: E-Government as a global phenomenon. Government Information Quarterly, 20(4), 323−331.
32. Heeks, R. (1999a). Information technology, government and development: Workshop report. Available at: http://idpm.man.ac.uk/idpm/itgovsem.htm. Accessed March 12, 2003.
33. Ghere, R. K., & Young, B. A. (1998). The cyber-management environment: Where technology and ingenuity meet public purpose and accountability. Public Administration and Management: An Interactive Journal, 3(1).
34. Relyea, H. C. (2004). Homeland security and information sharing: Federal policy considerations. Government Information Quarterly, 21, 420−438.
35. Norris, D. F. (2003). Building the virtual state … or not: A critical appraisal. Social Science Computer Review, 21(4), 417−424.
36. Lev, B. (2000). Knowledge management: Fad or need. Research Technology Management, 43(5), 9−10.
37. Lissack, M. R. (2000). Knowledge management redux: Reframing a consulting fad into a practical tool. Emergence, 2(3), 78−89.
38. McClure, C. R., & Bertot, J. C. (2000). The Chief Information Officer (CIO): Assessing its impact. Government Information Quarterly, 17, 7−12.
39. Miles, M. B., & Huberman, M. (1994). Qualitative data analysis: An expanded sourcebook. Thousand Oaks, CA: Sage.
40. Miller, D., & Hartwick, J. (2002). Spotting management fads. Harvard Business Review, 80(10), 26−27.
41. Seifert, J. W. (2004). Data mining and the search for security: Challenges for connecting the dots and databases. Government Information Quarterly, 21, 461−480.
42. Simon, H. A. (1976). Administrative behavior (3rd ed.). New York, NY: The Free Press.
43. Stowers, G. N. L. (1999). Becoming cyberactive: State and local governments on the World Wide Web. Government Information Quarterly, 16(2), 111−127.
44. Layne, K., & Lee, J. (2001). Developing fully functional e-government: A four stage model. Government Information Quarterly, 18, 122−136.
45. Regan, P. M. (2004). Old issues, new contexts: Privacy, information collection and homeland security. Government Information Quarterly, 21, 481−497.
46. Seifert, J. W., & Relyea, H. C. (2004). Do you know where your information is in the homeland security era. Government Information Quarterly, 21, 399−405.
47. McGregor, E. B., Jr. (2001). Web page accountability: The case of public schools. Paper presented at the National Public Management Research Conference, Bloomington, IN.
48. Means, G., & Schneider, D. (2000). Meta-capitalism: The e-business revolution and the design of 21st century companies and markets. New York: John Wiley & Sons Inc.
49. McClure, D. L. (2000). Government online: Strategies and challenges. Available at: http://www.house.gov/reform/gmit/hearings/2000hearings/000522.egovt/000522dm.htm. Accessed June 3, 2001.
50. Northrop, A. (2003). Information technology and public administration: The view from the profession. In G. D. Garson (Ed.), Public information technology: Policy and management issues (pp. 1−19). Hershey, PAe: Idea Group Publishing.
51. Mahmood, R. (2004). Can information and communication technology help reduce corruption: How so and why not? Two case studies from Southeast Asia. Perspectives on Global Development and Technology, 3(3), 347−373.
52. West, D. (2003a). Urban E-Government, 2003.
53. Yildiz, M. (2003). A general evaluation of the theory and practice of e-government (In Turkish). In M. Acar & H. Ozgur (Eds.), Cagdas Kamu Yonetimi-1 (pp. 305−328). Istanbul: Nobel Publications.
54. Thompson, J. D. (1967). Organizations in action: Social science bases of administrative theory. New York: McGraw Hill.
55. Westerback, L. K. (2000). Toward best practices for strategic information technology management. Government Information Quarterly, 17(1), 27−41.
56. Torres, L., Pina, V., & Acerete, B. (2005). E-government developments on delivering public services among EU cities. Government Information Quarterly, 22, 217−238.
57. United Nations, & American Society for Public Administration (ASPA). (2002). Benchmarking e-government: A global perspective. New York, NY: U.N. Publications.
58. Whitson, T. L., & Davis, L. (2001). Best practices in electronic government: Comprehensive electronic information dissemination for science and technology. Government Information Quarterly, 18, 7−21.
59. Zuboff, S. (1988). In the age of the smart machine: The future of work and power. New York: Basic Books.
60. Stowers, G. N. L. (2003). Information technology and its social dimensions. Public Administration Review, 63(2), 243−246.
61. Yildiz, M. (2004). Peeking into the black box of e-government policy-making: The case of Turkey. Unpublished Ph. D. Dissertation, Public Affairs Program, Indiana University-Bloomington.
62. Yang, K. (2003). Neoinstitutionalism and e-Government: Beyond Jane Fountain. Social Science Computer Review, 21(4), 432−442.