МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ И РОССИЙСКОЙ ФЕДЕРАЦИИ

федеральное государственное бюджетное образовательное учреждение

высшего образования

**«УЛЬЯНОВСКИЙ ГОСУДАРСТВЕННЫЙ ТЕХНИЧЕСКИЙ УНИВЕРСИТЕТ»**

А.А. РОМАНОВ, В.В. ШЕРКУНОВ

**ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА ИНФОРМАЦИОННЫХ СИСТЕМ ДЛЯ ПРИНЯТИЯ КРЕДИТНЫХ РЕШЕНИЙ**

**лабораторный практикум**

Ульяновск

УлГТУ

2022

**УДК 004.9 : 338(076)**

**ББК 32.973.26 – 018.2я73**

|  |  |
| --- | --- |
| К69  К69 | Рекомендовано научно-методической комиссией факультета информационных систем и технологий в качестве практикума.  **Романов, Антон Алексеевич**   Проектирование и разработка информационных систем для принятия кредитных решений: лабораторный практикум / А. А. Романов, В. В. Шеркунов. – Ульяновск : УлГТУ, 2022. – 38  с.  Практикум адресован студентам для выполнения и оформления лабораторных работ по дисциплине «Методы искусственного интеллекта для принятия кредитных решений». Предоставлены задания, рекомендации и требования к лабораторным работам, разработанные в соответствии с рабочей программой дисциплины.  Предназначен для студентов, обучающихся по направлению: 38.04.01 «Экономика» (программа Искусственный интеллект и анализ больших данных в банковской сфере)  Работа подготовлена на кафедре «Информационные системы». |

**УДК 004.9 : 338(076)**

**ББК 32.973.26 – 018.2я73**

© Романов А.А., Шеркунов В.В. 2022

© Оформление. УлГТУ, 2022

СОДЕРЖАНИЕ

[Введение 4](#_Toc113537502)

[Собеседование по лабораторным работам 4](#_Toc113537503)

[Экзамен 5](#_Toc113537504)

[Лабораторная работа №1 6](#_Toc113537505)

[Лабораторная работа №2 14](#_Toc113537506)

[Лабораторная работа №3 19](#_Toc113537507)

[Лабораторная работа №4 26](#_Toc113537508)

[Лабораторная работа №5 31](#_Toc113537509)

[Список литературы 36](#_Toc113537510)

# Введение

Целью освоения дисциплины «Методы искусственного интеллекта для принятия кредитных решений» является освоение интеллектуальных методов, технологий, инструментов и систем при автоматизации операций принятия кредитных решений.

Задачами освоения дисциплины (модуля) являются:

* формирование навыков управления рисками с применением систем искусственного интеллекта;
* изучение способов и инструментов проведения верификации потенциальных клиентов;
* формирование набора критериев оценки кредитных решений;
* изучение способов формирования интеллектуальных моделей для формирования рекомендаций по кредитным решениям.

Кроме того, в результате изучения дисциплины «Методы искусственного интеллекта для принятия кредитных решений» обучающиеся на основе приобретенных знаний, умений и навыков достигают освоения компетенций на определенном уровне.

Тематический план дисциплины

Раздел 1. Кредитные операции

1.1. Механизмы кредитования.

1.2. Кредитная документация.

1.3. Методы анализа кредитоспособности.

1.4. Кредитная политика

Раздел 2. Системы знаний в принятии кредитных решений

2.1. Модели знаний.

2.2. Методы извлечения знаний.

2.3. Логический вывод.

2.4. Интеллектуальные модели клиентов.

Раздел 3. Интеллектуальные инструменты в принятии кредитных решений

3.1. Источники данных для анализа.

3.2. Методы обработки и интеграции данных.

3.3. Интеллектуальные системы поддержки принятия кредитных решений.

## Собеседование по лабораторным работам

Собеседование по выполнению лабораторных работ осуществляется с целью проверки уровня знаний, умений, владений, понимания студентом основных методов и методик проектирования, разработки и поддержки информационных систем в задачах автоматизации бизнес-процессов, умения применять на практике полученных знаний. Каждое лабораторное занятие студент выполняет объемную задачу по конкретной теме с возможностью внесения доработок и изменений. Шкала оценивания имеет вид (таблица 1)

Таблица 1

|  |  |
| --- | --- |
| Оценка | Критерии |
| Сдано | Студент демонстрирует знания теоретического и практического материала по теме лабораторной работы, дает правильный алгоритм решения, в конце занятия студент выдает законченную и полностью функционирующую разработку. |
| Не сдано | Студент в конце занятия выдает не законченную и/или не полностью функционирующую разработку, некорректно отвечает на дополнительные вопросы. |

## Экзамен

Экзамен по дисциплине проводится в форме ответа на теоретические вопросы билета. Билет содержит два теоретических вопроса, обеспечивающих контроль освоения умений и навыков всех запланированных в ходе изучений дисциплины компетенций. Билет формируется таким образом, чтобы в него комплексные вопросы, контролирующие уровень усвоения всех заявленных дисциплинарных компетенций.

Шкала оценивания имеет вид (таблица 2)

Таблица 2

|  |  |
| --- | --- |
| Оценка | Критерии |
| Отлично | Выставляется обучающемуся, если студент в полном объеме ответил на оба теоретических вопроса и дал пояснения на ряд дополнительных вопросов |
| Хорошо | Выставляется обучающемуся, если студент в полном объеме ответил на один теоретический вопрос, а на второй дал частично верный, но не исчерпывающий ответ и дал пояснения на ряд дополнительных вопросов |
| Удовлетворительно | Выставляется обучающемуся, если студент в полном объеме ответил на один из двух теоретических вопросов или дал два частично верных не исчерпывающих ответа. И дал пояснения на ряд дополнительных вопросов. |
| Неудовлетворительно | Выставляется обучающемуся, если студент не смог ответить ни на один из предложенных вопросов |

# Лабораторная работа №1

Тема: Выбор и реализация способа принятия кредитных решений.

Сам по себе процесс выдачи кредита потребителям всегда кажется наиболее простым. Клиент приходит к кредитному специалисту, заполняет кредитную документацию, например анкеты, заявления и ждёт решения. После того, как ему представляется решение, если оно положительное, заключает кредитный договор, и клиент получает сумму кредита и может ей распоряжаться по своему усмотрению. С виду всё это кажется довольно простым. Но на самом деле внутри банка происходит работа крупной системы от оценки кредитоспособности по документам, которые были предоставлены и составления всей документации, которая будет сопровождать этот кредит, до выдачи и полного его погашения.

Если говорить про определение кредита, то это ссуда в денежной или материальной форме, предоставляемая кредитором заемщику на условиях возврата. Сам по себе, непрерывный процесс по выдаче кредиты и последующему его погашению представляет собой механизм кредитования. Механизм кредитования – это структурированная методика использования приемов и способов реализации кредитных отношений, которые расположены в определенной последовательности и приняты в каждом конкретном банке. Механизм кредитования в банке является частью кредитного процесса. Сам по себе кредитный процесс представляет собой комплекс организационно-технических операций, при помощи которых осуществляется непосредственно само кредитование в банке, в который включаются процедуры предоставления и возврата банковской ссуды.

В отличие от механизма кредитования, кредитный процесс начинается ещё до выдачи ссуды, где происходит значительная работа, осуществляемая банком и заемщиком, то есть кредитный процесс включает в себя ещё подготовительные работы. Помимо выдачи самого кредита, кредитный процесс включает четыре этапа. Это планирование, представление, использование и возврат ссуды.

Соответственно, на 1 стадии «планирования» происходит разработка кредитной политики, то есть набора документов, приказов и прочей документации, которые определяют основные условия кредитования. Стадия «предоставление» включает в себя оценку кредитоспособности клиента, где определяется, выдавать ему кредит или нет, а также сюда входит само заключение кредитного договора и предоставление денежных средств. Следующая стадия «исполнения» подразумевает контроль за своевременным внесением платежей по кредиту, то есть на этой стадии происходит контроль того, все ли платежи в полном объёме и вовремя поступают на кредитный банковский счет в тот период, когда они должны там быть по графику платежей. И последняя стадия – это «возврат». Она подразумевает под собой возврат полной стоимости кредита банку, включая проценты, не ежемесячных платежи, которые покрывает какую-то долю, а полное погашение суммы кредита. Это либо досрочное погашение полной стоимости, либо внесения последнего платежа, после которого кредитный договор закрывается и все обязанности клиента перед банком считаются выполненными.

В рамках процесса кредитования реализуется непосредственно сам механизм кредитования, который включает в себя все действия, которые необходимо выполнить от момента, когда клиент только обращается в банк, до полного погашения кредита. Механизм кредитования представляет собой последовательную реализацию стадий кредитного процесса и включает 4 этапа. «Определение кредитоспособности клиента» – этап, где происходит оценка клиента по каким-то параметрам, на соответствие того предоставлять или нет ему этот кредит. После того, как вынесено положительное решение, происходит переход на этап «заключения кредитного договора» и выдачу кредита. То есть на этом этапе, формируется и подписывается договор и происходит перечисление денежных средств. После чего наступает процесс «контроля», когда происходит отслеживание, того поступили ли денежные средства в нужный момент по графику платежей или нет? И в конце переходим к этапу «закрытия кредита» после его полного погашения.

Механизм кредитования, также описывает то, как должна работать банковская система или комплекс банковских систем по предоставлению и обслуживанию кредита в рамках кредитного процесса.

Для того, чтобы получить кредит, заключается и составляется определённая кредитная документация. Под кредитной документацией понимаются составляемые клиентом и банком документы, которые сопровождают кредитную сделку с момента обращения клиента в банк и до погашения ссуды. Главным и ключевым документом между клиентом и банком, который носит двусторонний характер, является кредитный договор, который регулирует весь комплекс взаимоотношений банка с клиентом.

К документам, которые составляют банки, можно отнести заключение на кредитную заявку клиента, ведомости выполнения клиентом условий договора, это некоторые документы, где указывается, что все платежи внесены вовремя или нет, и кредитные дела клиента, к которым могут относиться анкеты и другие документы.

Можно выделить документы, которые оформляется как между банком, так и юридическими и физическими лицами. К таким документам относятся: заявка и заявление на получение кредита. Для физических лиц дополнительно могут запрашиваться справки о доходах, какие-то персональные документы, например паспорт, ИНН, трудовая книжка или справка с места работы, а также анкета заёмщика, где может указываться: семейное положение, наличие детей и так далее.

Если мы говорим про юридических лиц, то в их случае процесс кредитование и наличие кредитной документации является более расширенным и включает в себя финансовый отчёт о деятельности предприятия, отчет о движении кассовых поступлений, внутренний финансовый отчет, могут запрашиваться налоговые декларации, которые подаются в налоговые органы, бизнес-планы, если кредит требуется на развитие, или каких-то бизнес-направлений. Если имущество находится в залоге перед другими кредиторами, то могут запрашиваться сведения по заложенному имуществу.

Когда мы говорили про механизм кредитования, то 1 этапом, и одним из самых важных, был оценка кредитоспособности клиента. На данном этапе производится его оценка, чтобы снизить финансовые риски. Это связано с тем, что клиенты часто не возвращают полученный кредит. В связи с чем выделяют актуальной является проблема определения платежеспособности клиента. Для выполнения такой оценки применяются следующие методы анализа кредитоспособности физических лиц: это скоринговая оценка, изучение кредитной истории, оценка по финансовым показателям платёжеспособности и андеррайтинг.

Скоринговая оценка кредитоспособности – это система оценки клиентов-заёмщиков, которая базирована на статистических способах. Скоринг реализуется банковской информационной системой, куда вводится анкета соискателя по кредиту, а в ответ выводится оперативное заключение о том, давать заёмщику кредитные средства или нет. Соответственно, когда специалист кредитного отдела загружает анкету в систему, то система, используя статистические методы оценивает вероятность невозврата кредита по показателям, которые были переданы в анкете, путем сравнения с другими анкетами, которые хранятся в базе данных банка, и на основе этого происходит решение о том предоставлять кредит конкретному лицу или нет.

Для оценки параметров применяются баллы в пределах установленного банком максимума. И сама по себе оценка баллов, происходит в 2 этапа. Сперва оцениваются данные отдельных критериев заявки заёмщика, для получения предварительной оценки вероятности выдачи кредита, а на 2 этапе оцениваются риски с учетом таких вспомогательных характеристик, как например предыдущие возвраты или невозвраты кредитов. Как правило, скоринг-системы применяются в экспресс кредитовании, к которому можно отнести, например, покупку товаров и услуг, когда нужно принять решение в кратчайшие сроки.

К плюсам таких систем можно отнести сокращение уровня безнадежного долга, количество и объективность принятия решений. То есть в данном случае исключается зависимость от человека, а всю оценку выполняет информационная система, что позволяет исключить человеческий фактор. Если говорить про эффективное управление кредитным портфелем и уменьшение времени подготовки сотрудников кредитного отдела, то это больше плюс для банка, которому не нужно содержать целый отдел, куда приходит большой объем кредитных заявок, поскольку всё делает одна система в более короткие сроки, что и подразумевает проведение экспресс-анализ заявки на кредит в присутствии клиента. Для клиента же плюс в том, что ему не нужно, уходить на несколько дней домой, а потом опять приезжать для получения своего решения. Соответственно если говорить о минусах, то это то, что оценка параметров производится только на базе информации клиентов, которые уже повторно обращаются в банк. Если клиент обратился впервые, то об этом клиенте отсутствует информация, но новые ИС сейчас работают так, что они сравнивают одного клиента с рядом других, либо запрашивают необходимую информацию у других банков. Второй недостаток связан с тем, что из-за накопления исторической информации, система может давать некорректный результат. Допустим, размеры зарплат и уровень цен 10 лет назад были другими, и изменились с уровнем инфляции, поэтому возникает необходимость корректировки таких данных.

Второй вид оценки — это оценка на основе кредитной истории. Кредитная история – это информация, которая определена Федеральным законом «О кредитных историях», и которая описывает выполнение заёмщиком принятых на себя обещаний по договорам кредитования и хранится в бюро кредитных историй. Соответственно, бюро хранят следующие данные:

* Социально-демографические характеристики, это может быть информация о наличии детей, семейном положении и так далее.
* Юридические заключения. Были ли какие-то судебные разбирательства или нет.
* Информация о банкротствах клиента, если такая процедура у него проводилась.
* Данные об персональных заёмщиках, приходящих по принципу обмена данными. В бюро кредитных историй находится информация, не только какого-то отдельного конкретного банка, а различных банков, а также происходит обмен данными между и различными бюро кредитных историй, чтобы оценка по кредитам была более точной.

Следующий механизм оценки — это оценка по финансовым показателям платёжеспособности. В основе таких показателей лежат данные о доходах заёмщика и степени риска потери этих доходов. При выдаче кредита рассчитывается платёжеспособность заёмщика, на базе данных о среднемесячном доходе за предшествующие 6 месяцев, который определяется по справке о заработной плате или по налоговой декларации, например справке 2-НДФЛ. Доход уменьшается на обязательные платежи по кредиту и по каким-то другим обязательным платежам, например, услуг ЖКХ, кредитам в других банках и корректируется на коэффициент, который дифференцируются зависимости от величины дохода от 0,3 до 0,6. Соответственно, чем больше доход, тем больше корректировка этого коэффициента, поскольку наибольший доход, проще потерять, чем, скажем, минимальный, например, вместо миллиона получить 800000 наиболее вероятно, чем вместо 15000 получить 10000. Поэтому применяются такие различные коэффициенты. Такой подход позволяет минимизировать риски невозврата в случае высокой долговой нагрузки заёмщика.

И последний способ это андеррайтинг. Андеррайтинг – это тестирование клиента для определения его кредитоспособности. Андеррайтинг является современным способом ограничения кредитного риска. С помощью него происходит анализ и оценка платежеспособности заёмщиков и принятия решения о выдаче кредита.

В процессе андеррайтинга выявляется возможность своевременного внесения платежей. Для этого анализируются данные о доходах и расходах клиента, например, по базе данных банка о том, что покупает клиент, какие доходы к нему поступают, а также оценивается достаточность для обеспечения имущества залога. Соответственно, для этого требуются справка о доходах, длительность неизменной работы, кредитная история и справка о залоге имущества. Представленная методика предполагает применение к конкретному клиенту индивидуального подхода, определение характеристик необходимых конкретно для него. А недостатком является трудоемкость такого анализа и требования к высокой квалификации персонала, что существенно увеличивает расходы данного метода. Но с развитием технологии такой подход становится более актуальным, когда можно загрузить полностью все данные по клиенту в ИС. Разработать сложные механизмы и правила оценки соответствия клиента требованиям к заёмщикам, для выдачи каких-то крупных сумм кредита, что позволяет снизить расходы банка в случае невозврата таких сумм.

Непосредственно сам процесс кредитования и порядок выдачи кредитов регулируется кредитной политикой. Сама по себе кредитная политика это составная часть экономической политики нашего государства, которая представляет собой систему мер в области кредитования народного хозяйства. Это если говорить о политике с точки зрения государственного управления. Но если говорить о кредитной политики с точки зрения банка, то это стратегия и тактика банка в области кредитных операций, которые принимаются в конкретном банке. Соответственно, кредитная политика в части стратегии вбирает в себя приоритеты, принципы и содержательные цели конкретного банка на кредитном рынке, а в части тактики - финансовый и иной инструментарий, используемый данным банком, для реализации его цели при осуществлении кредитных сделок, правила совершения и порядок организации кредитного процесса.

Данный документ показывает, какую стратегию выбрал банк, какие основные направления кредитного процесса он охватывает, что там непосредственно должно происходить, как должен быть обеспечен процесс организации кредитования и непосредственно сам контроль за реализацией этой политики. Кредитная политика дает необходимые общие предпосылки эффективной работы персонала кредитного подразделения, объединяет и организует усилия персонала и уменьшает вероятность ошибок, поскольку кредитная политика является общим документом для всего банка.

Кредитная политика регулирует все этапы кредитования. Если говорить про 1 этап, где происходит предварительная работа по предоставлению кредита, то здесь определяется состав возможных заёмщиков, например, это могут быть юридические или физические лица; так же определяются виды кредитов, которые могут быть предоставлены, например потребительские или агрокредиты; какие-то количественные пределы кредитования, например максимальные сроки кредитования, суммы кредита, процентные ставки; также определяются методы обеспечения возвратности кредита и контроль за соблюдением процедуры подготовки выдачи кредита. На 2 этапе, где происходит оформление кредита регламентируются формы документов, технологии и процедуры выдачи кредита, контроль за правильностью оформления кредита. В кредитной политике соответственно прописываются все эти пункты. На 3 этапе происходит управление кредитом и то, как должен происходить контроль за исполнением кредитных договоров, условия продления или возобновлении просроченных кредитов, порядок и покрытие убытков, а также контроль за управлением кредита. Здесь как раз отписываются все условия возврата, приостановления и возобновления кредита.

В пределах нормативных ограничений, которые устанавливает Центральный Банк, коммерческий банк самостоятельно определяет круг потенциальных заёмщиков, виды кредитов, формирует кредитный портфель и устанавливает процентные ставки, из своих внутренних соображений, и финансовых показателей, на которых ему выгодно работать. Кредитная политика является одним из наиболее важных документов банка и на основе неё как раз строится процесс выдачи кредитов.

В процессе выполнения лабораторной работы необходимо:

1. Выбрать способ принятия кредитных решений.
2. Разработать анкету для получения необходимых данных клиента.
3. Провести анализ анкеты по личным данным.
4. Оценить результат анализа анкеты по выбранному способу принятия решения.

Контрольные вопросы к лабораторной работе

1. Что такое механизм кредитования?
2. Перечислите основные способы принятия кредитных решений.
3. Дайте понятие кредитного скоринга.
4. Дайте понятие кредитной политики.
5. Перечислите основные кредитные документы.

# Лабораторная работа №2

Тема: Анализ источников данных.

При оценке кредитоспособности клиента учитывается множество параметров, которые хранятся в различных источниках информации.

К таким источникам можно отнести:

* Банковские базы данных;
* Бюро кредитных историй;
* Социальные сети;
* Кредитную и иную документацию;
* Открытые базы знаний.

Основным источником данных, для проведения интеллектуального анализа кредитоспособности клиентов служат базы данных банков, которые к настоящему времени содержат огромные массивы информации о клиентах по различным направлениям:

* это и личные данные;
* и информация о действующих кредитах и выплатах по ним;
* информация по закрытым кредитам и история платежей;
* информация об операциях клиента.

С технической точки зрения, среди таких баз данных, которые применяются в банковском секторе, используются следующие виды:

* Реляционные (PostgreSQL, MySQL);
* Ключ-значение (Redis);
* Документоориентированные (MongoDB);
* Графовые (Neo4j);
* И столбцовые (ClickHouse).

Следующим источником могут служить базы данных бюро кредитных историй.

Кредитная история – это документ, который характеризует платежную дисциплину человека или организации.

Кредитные истории формируют специальные организации – бюро кредитных историй (БКИ) на основании информации о заемщике, которую банки туда направляют. Также в бюро может передаваться информация от организаций, в пользу которых вынесены судебные решения о взыскании долгов за ЖКХ, услуги связи, или от судебных приставов, например, по неисполненным алиментным обязательствам.

Бюро кредитных историй – это организация, занимающаяся оказанием услуг по формированию, обработке и хранению кредитных историй, а также по предоставлению кредитных отчётов.

С 1 октября 2019 по кредитной истории рассчитывается Показатель долговой нагрузки (ПДН) для физических лиц. Если этот показатель у гражданина слишком высокий, то есть его платежи по кредитам «съедают» слишком большую часть его дохода, то банк может отказать в выдаче нового займа.

Кредитная история охватывает 10 лет, этот срок отсчитывается с момента любых последних изменений в кредитной истории (например, изменения паспортных данных и т. д.).

Запрашивать кредитные истории в БКИ могут как сами заемщики, так и банки, микрофинансовые организации и другие юридические лица.

Кредитные истории и юридических, и физических лиц включают информацию об их кредитах с суммами и сроками их погашения, сведениями о текущей и просроченной задолженности, одним словом — платежной дисциплине, а также процедурах банкротства. Эта информация передается только с согласия субъекта кредитной истории. Поэтому, сейчас перед подачей документов для получения кредита берется согласие клиента.

Кредитные истории физических лиц (включая индивидуальных предпринимателей) имеют также информационную часть, которую банк или МФО может получить без согласия человека, если тот обратился за получением кредита. Информационная часть содержит сведения о выданных займах или отказах в их предоставлении, сведения о договорах поручительства, а также сведения о просрочке в выплате кредита (к которой можно отнести пропуск двух и более платежей подряд в течение 120 дней).

Как правило, кредитная история хранится в нескольких бюро. При этом банк или МФО самостоятельно решают, в какое бюро (одно или несколько) направлять информацию о заемщике.

В свою очередь, бюро кредитных историй предлагают банкам следующие виды услуг:

* получение кредитных отчётов в режиме онлайн;
* получение информации из Центрального каталога кредитных историй;
* получение информации из Федеральной миграционной службы;
* мониторинг финансовой дисциплины заёмщиков;
* скоринг, построенный по данным бюро, т. е. бюро уже предварительно самостоятельно строит скоринговую оценку;
* разработка индивидуальных скоринговых карт с использованием своей базы данных;
* дисциплинирование заёмщиков банка путём рассылки им писем от имени бюро или банка при нарушении финансовой дисциплины или увеличении долговой нагрузки;
* системы борьбы с мошенничеством;
* системы обработки кредитных заявок;
* системы интеграции к нескольким бюро кредитных историй, что позволяет получать сводную информацию от различных кредитных организаций.

К открытым источникам данных для получения социального портрета клиента можно отнести социальные сети. Они позволяют оценить активность клиента в цифровом пространстве. На основе данных из социальных сетей можно узнать о клиенте следующую информацию:

* Его интересы (на основе лайков, репостов и комментариев);
* Популярность (посты, просмотры, частота публикаций);
* Социальный круг (друзья, подписки);
* Частоту использования (время, проводимое в социальных сетях);
* Посещаемые места (геотеги, отметки и др.);
* И даже эмоциональный портрет (окраска тональности сообщений).

Например, компания Doble Data разработала программный продукт, который находит и собирает сотни дополнительных скоринговых факторов из социальных сетей и других открытых интернет-источников. Из-за чего ВКонтакте подавала иск в суд, с требованием запретить такой сбор данных и проиграла его, что говорит о том, что данные социальных сетей могут быть легко доступны для такого рода анализа.

В России наиболее популярными являются следующие социальные сети:

* Instagram;
* Вконтакте;
* ТикТок;
* Одноклассники;
* Facebook.

Каждая социальная сеть предоставляет специальный инструмент, позволяющий получать доступ к открытым данным – называемый API. API (Application Programming Interface или программный интерфейс приложений) – это совокупность инструментов и функций в виде интерфейса для создания новых приложений, благодаря которому одна программа будет взаимодействовать с другой. Инструкцию по работе с API Вконтакте можно найти по адресу https://dev.vk.com/reference

Используя API можно получить доступ к следующей информации:

* список друзей;
* опубликованные сообщения:
  + текст;
  + местоположения;
  + тэги;
* список групп и пабликов;
* и др.

До появления централизованных хранилищ данных, таких как базы данных, вся информация о клиентах хранилась в текстовом формате. Кроме того, нормативная документация банковских учреждений, также хранится в текстовом формате, в таких документах как устав, кредитная политика, нормативно-справочная информация и др.

Соответственно, для получения информации о клиентах, данные о которых не были занесены в базы данных, применяются методы интеллектуального анализа текстовой информации для получения необходимых сведений.

Также в банки могут поступать внешние документы как от клиентов, так и от различных структур: ИФНС, ПФР, суды и эта информация также должна попадать в цифровые хранилища данных.

Помимо социальных сетей, среди открытых источников данных можно выделить открытые базы знаний. К таким базам знаний можно отнести WikiData.

WikiData – это совместно редактируемая база знаний, созданная Фондом Викимедиа. Используется для обеспечения централизованного хранения данных, которые могут содержаться в статьях Википедии.

WikiData является одной из крупнейших в мире баз знаний, содержащей более 15 млн. понятий из различных областей знаний с описанием на различных языках.

В WikiData каждое понятие имеет свой идентификатор в виде Q324254 (соответствует понятию «Онтология»), а каждое свойство в виде P279 (соответствует свойству «Подкласс от»), например предложение «Онтология является подклассом от Базы знаний» в WikiData будет записано как утверждение вида «Q324254 P279 Q593744».

WikiData может представлять дополнительную информацию и служить источником для обогащения существующей базы знаний общими наборами данных, такими как понятия из банковской сферы и пр.

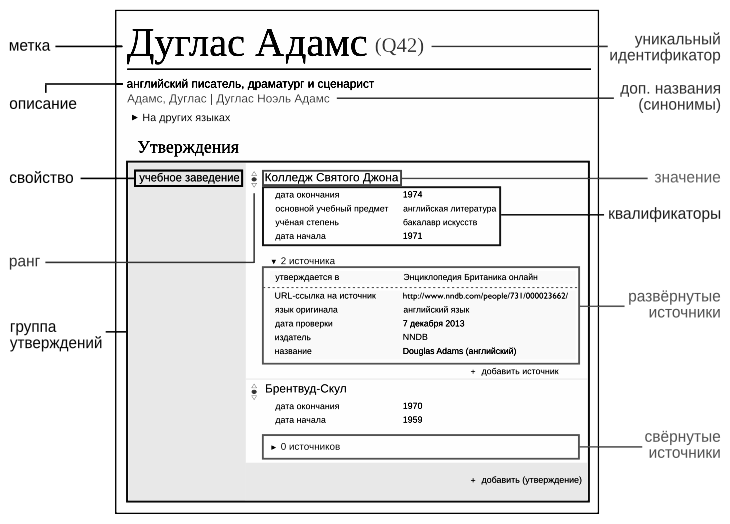


Рисунок 1. Пример записи в WikiData.

В процессе выполнения работы необходимо:

1. Осуществить поиск наборов данных с информацией по заёмщикам в открытых источниках.
2. Провести анализ найденных данных.
3. Охарактеризовать полученные результаты.
4. Предложить вариант хранения найденных данных в банковской ИС.

Контрольные вопросы к лабораторной работе

1. Перечислите основные виды источников данных.
2. Перечислите основные виды представления данных.
3. Перечислите виды СУБД.
4. Перечислите форматы и способы хранения данных.
5. Охарактеризуйте инфраструктуру хранения данных в банковском секторе.

# Лабораторная работа №3

Тема: Формирование базы знаний. Разработка правил и системы знаний для принятия кредитных решений.

Процесс оперирования знаниями для человека кажется обычным делом, поскольку в течение жизни он сталкивается с различными объектами, устанавливает связи между ними и делает различные умозаключения. В отличие от человека, машины лишены такой возможности, и чтобы представить обычные данные в виде знаний в компьютерных системах, были разработаны формальные модели знаний. Но сначала нужно понять отличие знаний от данных.

Данные – это предоставление информации в формальном виде, пригодном для передачи, интерпретации или обработки людьми, или компьютерами (ISO/IEC 2382:2015).

Другое определение данных. Это формы представления информации, с которой имеют дело информационные системы и их пользователи (ISO/IEC 10746-2:1996).

Данные хранятся в базах данных. База данных – это организованная в соответствии с определёнными правилами и хранящаяся в памяти компьютера совокупность данных, характеризующая актуальное состояние некоторой предметной области и используемая для удовлетворения информационных потребностей пользователей.

Далее рассмотрим определение знания. Знания – это набор правил, законов и закономерностей, полученных в результате профессиональной деятельности человека в рамках определённой предметной области.

Как видно из определений, данные представляют какой-то организованный набор информации в то время, как знания представляют закономерности, правила и связи между данными, что позволяет получать новые знания и принимать решения.

Знания, в компьютерных системах, по аналогии с данными, хранятся в базах знаний.

База знаний – это хранилище знаний, содержащее правила вывода и информацию об опыте и знаниях в некоторой предметной области, которые позволяют устанавливать связи между вводимой и выводимой информацией.

Другое понятие базы знаний – это хранилище знаний, которое позволяет формировать новые знания на основе существующих в результате выполнения логического вывода.

Область науки и техники, которая занимается вопросами сбора и хранения знаний, называется инженерия знаний.

Инженерия знаний – это раздел инженерии, который направлен на внедрение знаний в компьютерные системы для решения сложных задач, обычно требующих богатого человеческого опыта.

Для того, чтобы организовать хранение знаний в компьютерных системах, их нужно структурировать и представлять в каком-то формальном виде. Для решения этих задач были разработаны модели представления знаний.

Модель представления знаний – это способ представления конкретных и обобщённых знаний, сведений и фактов для накопления и обработки информации в ЭВМ.

Главная задача моделей представления знаний в области искусственного интеллекта – научиться хранить знания таким образом, чтобы компьютерные программы могли осмысленно обрабатывать их и тем самым достигнуть подобия человеческого интеллекта.

Под термином «представление знаний» чаще всего подразумевают способы представления знаний, которые ориентированы на автоматическую обработку современными компьютерами, и, в частности, представления, состоящие из явных объектов и из суждений или утверждений о них.

Можно выделить следующие модели знаний, которые применяются в компьютерных системах:

* Продукционную;
* Фреймовую;
* Логическую;
* Семантическую сеть;
* И Онтологию.

Продукционная модель – это модель представления знаний, основанная на множестве правил (или как их ещё называют, продукций) следующего вида:

IF <условие>, THEN <действие>

Продукция состоит из следующих частей:

* это условие или антецедент;
* действие или консеквент;
* и условия, с помощью которых можно сочетать логические функции AND и OR.

Т.е. продукционные модели представляют собой наборы правил, на основе которых решается, будет выполнено некоторое действие или нет.

Действия можно разделить на два типа:

* Промежуточные, выступающие далее, как условия и передающиеся в следующее условие;
* И терминальные или целевые, которые завершают работу или действие.

Системы обработки знаний, использующие продукционную модель, получили название «продукционных систем». В состав экспертных систем продукционного типа входит база правил (и знаний), рабочая память (которая представляет собой временное хранилище фактов или гипотез, которые являются промежуточными решениями или результатами общения системы с внешней средой), и также в состав входит интерпретатор правил (или решатель), реализующий определенный механизм логического вывода.

Вывод, построенный на таких базах знаний, бывает двух типов:

* прямой – от данных к поиску цели;
* и обратный, т.е. от цели для её подтверждения - к данным.

Продукционная модель представления знаний нашла широкое применение в автоматизированных системах управления технологическим процессом, поскольку позволяет регулировать определенные производственные процессы от определенных условий, например температуры охлаждающей жидкости, плотности материала и т.д.

*Пример продукции:*

*IF* Температура = 39 *AND* Кашель = Есть *AND* Давление = 110-130 *THEN* Бронхит



Рисунок 1. Пример продукционной модели позволяющей условно определять диагноз по ряду симптомов.

Следующая модель, про которую мы поговорим – это логическая.

Основная идея подхода при построении логических моделей представления знаний заключается в том, что вся информация, необходимая для решения прикладных задач, рассматривается как совокупность фактов и утверждений, которые представляются как формулы в некоторой логике. Данная модель, в частности, строится на основе логики первого порядка.

Знания в такой модели отображаются совокупностью формул, а получение новых знаний сводится к реализации процедур логического вывода.

В основе же логических моделей представления знаний лежит понятие формальной теории, задаваемое следующим кортежем:

* счетное множество базовых символов (или алфавит). Это некие переменные из окружающей среды;
* множество, называемое формулами, в которых как раз используются переменные;
* выделенное подмножество истинных формул (аксиом);
* и конечное множество отношений между формулами, называемое правилами вывода.

Основными достоинствами логических моделей знаний является то, что:

* в качестве «фундамента» здесь используется классический аппарат математической логики, методы которой достаточно хорошо изучены и формально обоснованы;
* и то, что в таких базах знаний можно хранить лишь множество аксиом, а все остальные знания получать из них по правилам вывода.

В логических моделях знаний слова, описывающие сущности предметной области, называются термами (это константы, переменные, функции), а слова, описывающие отношения между сущностями – предикатами. Соответственно говоря простым языком, у нас есть набор термов, связанных предикатами.

Пример записи логической модели в виде формул и получение необходимых знаний через логический вывод представлен ниже:

ВЫБОР(ФИО, Должность) ЕСЛИ ПРЕТЕНДЕНТ(ФИО, Образование, Стаж), ВАКАНСИЯ(Должность, Образование, Опыт), Стаж >= Опыт.

ПРЕТЕНДЕНТ(«Петров», «Высшее», 12),

ВАКАНСИЯ(«Менеджер», «Высшее», 10).

ВЫБОР(ФИО, Должность) –> ФИО = «Петров», Должность = «Менеджер»

Далее перейдем к рассмотрению семантической модели. Собственно говоря, для того, чтобы понять почему модель называется семантической, рассмотрим понятие семантики.

Семантика – это наука, устанавливающая отношения между символами и объектами, которые они обозначают.

Соответственно, семантическая сеть – это информационная модель предметной области, которая представлена в виде ориентированного графа.

Вершины графа определяют объекты предметной области, а дуги (или рёбра) – отношения между ними. Объектами могут быть: понятия, события, свойства, процессы предметной области. Таким образом, семантическая сеть отражает семантику предметной области в виде понятий и отношений.

Для отображения связей, в семантических сетях обязательным является наличие следующих типов отношений:

* класс – подкласс;
* и свойство – значение;

В зависимости от используемых отношений, можно выделить следующие классификации семантических сетей:

* по количеству типов отношений:
  + это однородные (это с единственным типом отношений);
  + и неоднородные (с различными типами отношений).
* и по типам отношений:
  + бинарные (в которых отношения связывают два объекта);
  + и N-арные (в которых есть отношения, связывающие более двух объектов);

Например, семантическая сеть может быть неоднородной и N-арной, когда различные типы отношений связывают более двух объектов.

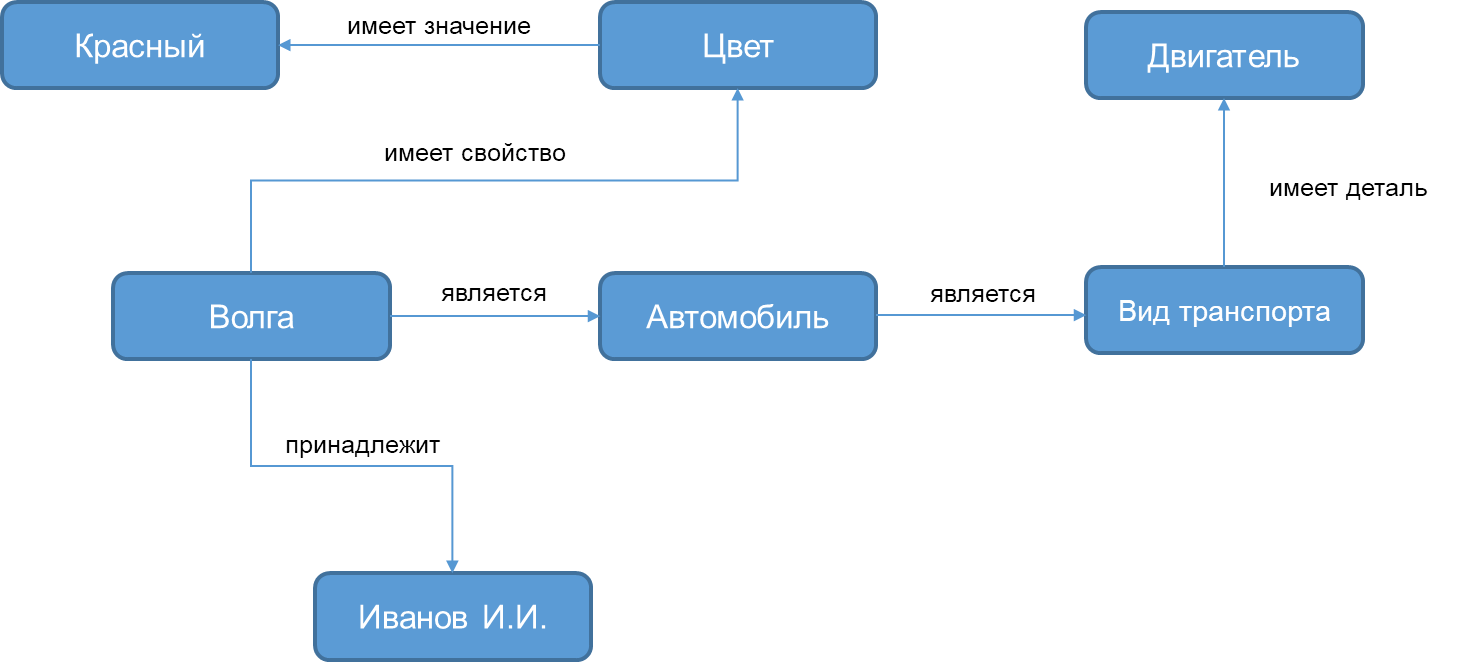


Рисунок 2. Пример описания понятия «Автомобиль» в виде семантической сети.

Перейдем теперь к фреймовой модели.

Фреймовая модель представляет собой систематизированную психологическую модель памяти человека и его сознания.

Фреймовая модель основана на понятии фрейма.

Фрейм – это модель абстрактного образа, минимально возможное описание сущности какого-либо объекта, явления, события, ситуации, процесса.

Другое определение фреймов, это структуры данных, в которой в определенном порядке представлены сведения о свойствах объекта. Информация, которая относится к фрейму, содержится в составляющих его слотах.

В свою очередь, слоты – это незаполненный набор атрибутов фрейма, после заполнения которых конкретными данными, фрейм будет представлять ту или иную сущность (или объект), явление, событие, ситуацию или процесс предметной области.

Слот имеет следующие параметры:

* имя слота – это идентификатор, присваиваемый слоту. Слот должен иметь уникальное имя во фрейме, к которому он принадлежит;
* значение слота – значение, соответствующее типу данных слота и удовлетворяющее условиям наследования;
* указатель атрибутов – указатель типа данных слота. Может быть, например INTEGER (целое число), REAL (вещественное число), BOOL (булево значение истинности) и т.д.;
* так же слот содержит такой параметр, как демон – это процедура, автоматически запускаемая при выполнении некоторого условия;
* и указатель наследования – только для фреймовых моделей иерархического типа; они показывают, какую информацию об атрибутах слотов во фрейме верхнего уровня наследуют слоты с такими же именами во фрейме нижнего уровня. Что показывает то, что фреймы могут наследоваться от частного к общему.

При этом, имя фрейма должно быть уникальным в пределах одной фреймовой модели. А имя слота должно быть уникальным в пределах одного фрейма.

Слот может также ссылаться на другой фрейм, тогда фреймовую модель можно рассматривать как некую семантическую сеть.

Пример записи объекта «Клиент» в виде фреймовой модели представлен на таблице ниже.

Если говорить про современные реализации моделей знаний, то на сегодняшний день одной из самых популярных моделей является онтологическая модель знаний. Она очень схожа с семантической сетью, поскольку является её развитием и имеет более богатый арсенал функций при моделировании.

Онтологическая модель – это попытка подробной формализации предметной области с помощью концептуальной схемы.

Для формализации представления знаний применяются различные формальные языки описания понятий и отношений между ними – так называемые дескрипционные логики.

Дескрипционная логика – это язык представления знаний, позволяющий описывать понятия предметной области в недвусмысленном, формализованном виде, организованный по типу языков математической логики.

Знания о предметной области, представленные в базе знаний дескрипционной логики, подразделяются на:

* T-Box (terminological knowledge) – это общие знания о множестве классов понятий предметной области, свойствах и отношениях между ними;
* и A-Box (assertional knowledge) – знания об конкретных объектах (экземплярах класса), их состояниях и связях с другими объектами.

В базе знаний T-Box и A-Box взаимосвязаны и позволяют наиболее полно представить особенности предметной области. Более конкретно мы рассмотрим эти понятия на следующих лекциях.

Онтологии также позволяют выполнять логический вывод, в том числе на основе логических правил.

А графически онтологическая модель представляется аналогично семантической сети в виде графа.

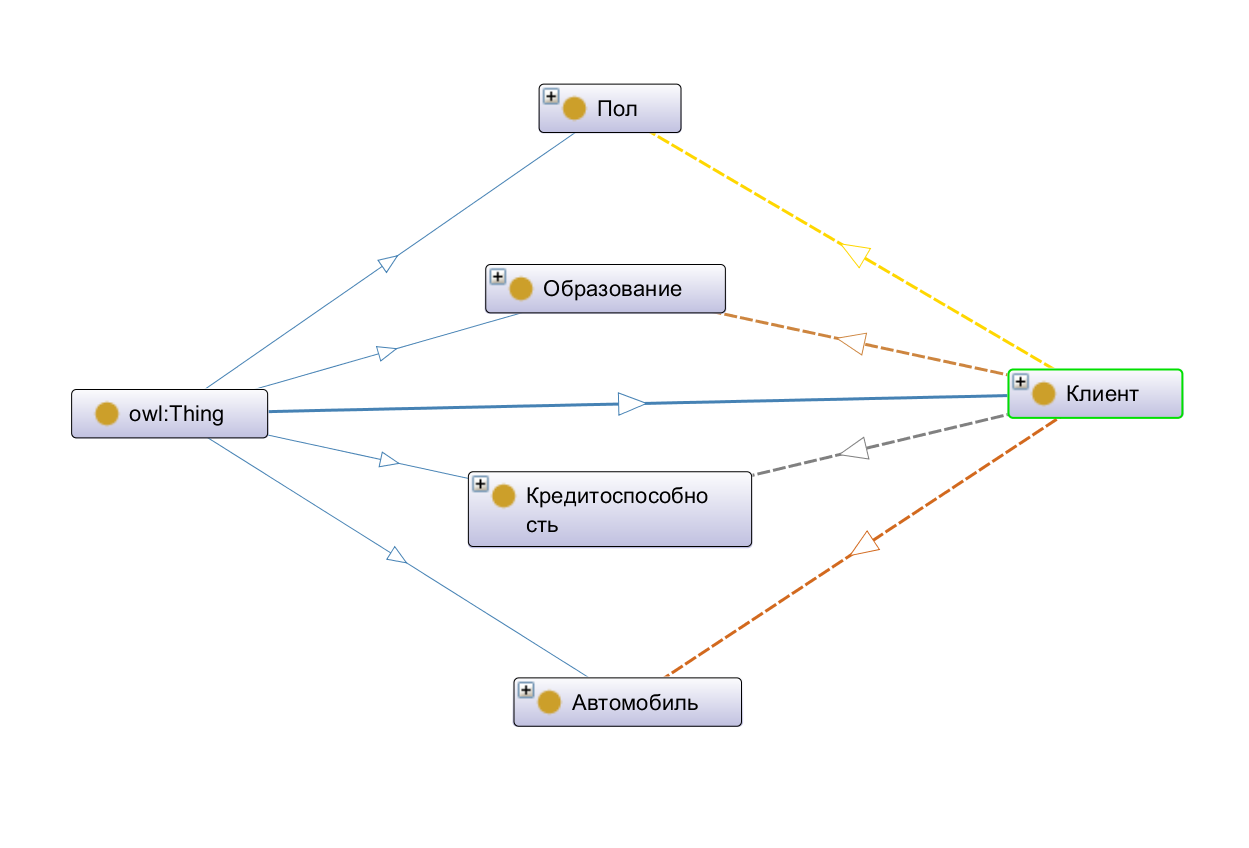


Рисунок 3. Простой пример реализации онтологической модели такого объекта как «Клиент».

Для выполнения работы необходимо:

1. Разработать базу знаний, реализующую систему знаний о заёмщиках.
2. Разработать систему правил для оценки кредитоспособности заёмщиков.
3. Внести данные анкеты из ЛР1 в базу знаний.
4. Осуществить логический вывод на введённых данных.

Контрольные вопросы к лабораторной работе

1. Что такое база знаний?
2. Перечислите основные модели знаний.
3. Назовите основные способы извлечения знаний.
4. Перечислите инструменты для построения баз знаний.
5. Дайте определение понятию логический вывод.
6. Перечислите инструменты, реализующие логический вывод.

# Лабораторная работа №4

Тема: Создание модели клиента. Построение модели клиента для возможности применения методов классификации и кластеризации.

Поскольку каждый клиент характеризуется набором определенных параметров и атрибутов, то можно построить для него формальную модель, с которой могли бы работать информационные системы.

Что же такое, модель клиента? Это набор параметров, характеризующий каждого клиента банка и позволяющий проводить интеллектуальный анализ для получения необходимой информации в рамках поставленной задачи.

Можно выделить следующие модели клиента, которые позволяют делать оценку кредитоспособности:

* табличную;
* онтологическую;
* и векторную.

Используя множество параметров, можно, например построить цифровой образ клиента, который будет включать:

* Личную информацию, включающую:
  + Пол;
  + Возраст;
  + Образование;
  + Профессия;
  + Место жительства;
  + Интересы;
  + Доход и т. д.
* Социальный профиль – это:
  + Активность в соц. сетях – например, средняя;
  + Активность в комментариях - низкая;
  + Активность репостов - низкая;
  + Эмоциональная окраска сообщений – нейтральная и т. Д.
* И Финансовый профиль (по оценки банка):
  + Средний объем трат – например 30’000 руб./мес.;
  + Тенденция трат – периодическая;
  + Предпочтения – одежда, автомобиль;
  + Поступления – постоянные и другие.

Так, табличная модель предполагает, хранение данных клиентов в виде таблицы и извлечение их по требованию. Такой формат позволяет извлекать необходимые параметры и использовать их в машинах логического вывода основанных на продукционной или фреймовой модели знаний.

Табличная модель поддерживается многими современными системами управления базами данных, такими как:

* PostgreSQL;
* MySQL;
* MS SQL Server и др.



Рисунок 1. Пример табличной модели клиента.

Также некоторые значения характеристик можно представить в нечетком виде.

Нечеткие параметры клиента – это параметры, которые подразумевают использование лингвистических переменных для описания набора свойств клиентов с определенной степенью уверенности их задания на числовом множестве.

Нечеткие параметры клиента позволяют охарактеризовать четкие (числовые) параметры каждого клиента в виде лингвистических переменных, например выразить понятие кредитоспособности как «Высокая» учитывая доход клиента указав его функцию принадлежности как MF(100’000)=0.80.

Используя онтологическую модель знаний, можно построить онтологическую модель клиента.

Онтологическая модель клиента – это модель позволяющая представить знания о клиенте в виде набора классов (концептов) и связей между ними (ролей) для дальнейшего применения механизмов рассуждений и получения логического вывода для оценки его кредитоспособности.

Данная модель позволяет создавать логические правила, и применяя логический вывод получать кредитную оценку клиента.

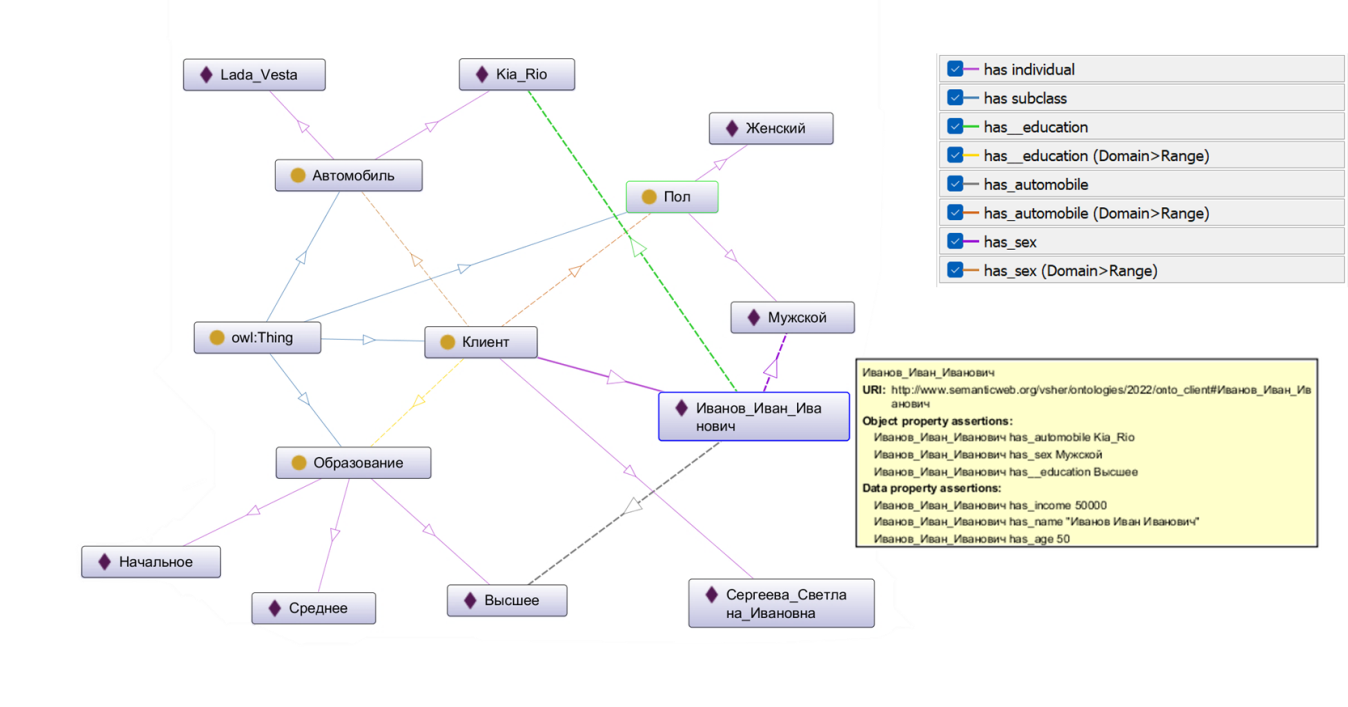


Рисунок 2. Пример онтологической модели клиента

А пример записи SWRL правила и осуществления логического вывода с его использованием в редакторе онтологий Protégé представлен так:



Рисунок 3. Пример записи SWRL правила

Ещё к одному из вариантов представления данных о клиенте, можно отнести векторное представление. Поскольку современные системы интеллектуального анализа данных, такие как нейронные сети, используют векторное представление данных, то зачастую табличную или онтологическую модель преобразуют в векторное пространство, получая векторное представление клиента.

Векторная модель клиента – это модель, которая представляет параметры клиента в формальном виде, заменяя конкретные лингвистические значения, на числовой аргумент, для дальнейшего использования в алгоритмах машинного обучения.

Такое представление позволяет выполнять ряд математических операций и преобразований, в результате которых можно, например использовать статистические методы для поиска похожих клиентов и производить оценку способности выплаты клиента на основе исторических данных.

Пример векторной записи параметров клиента представлено ниже:

[id, образование, пол, наличие автомобиля, оценка БКИ, доход, возраст, просроченные кредиты]

[1, 3, 1, 1, 630, 60000, 35, 0]

Используя такие модели как табличную или векторную, и имея данные о кредитной истории клиента, можно строить графические представления данных о влиянии ряда показателей на возвращаемость кредита.

Так на рисунке 4 представлена зависимость невозвратов кредитов от возраста клиента, на рисунке 5 от дохода и на рисунке 6 от оценки бюро кредитных историй. Если посмотреть на графики, то можно заметить, что наибольшее влияние на вероятность невозврата может указывать возраст и оценка БКИ.

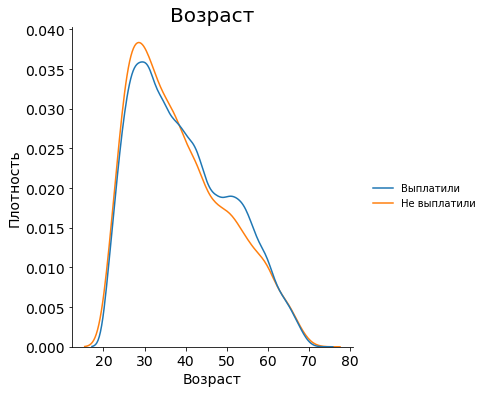


Рисунок 4. Пример влияния параметров на кредитную оценку.

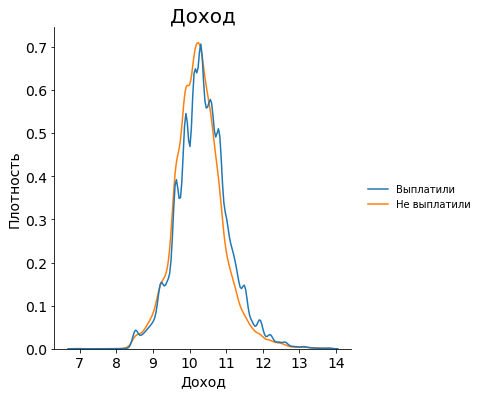


Рисунок 5. Пример влияния параметров на кредитную оценку.

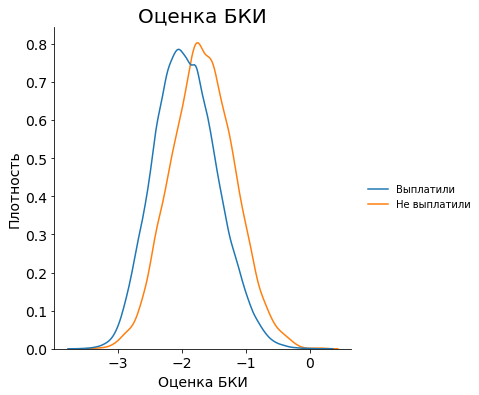


Рисунок 6. Пример влияния параметров на кредитную оценку.

Для выполнения работы необходимо:

1. Выделить показатели клиентов, по которым можно будет классифицировать возможность кредитоспособности.
2. Выделить показатели, по которым можно будет произвести кластеризацию для определения максимальной суммы кредита.
3. Обобщить выделенные показатели, и создать общую модель клиента.
4. Перевести анкету из ЛР1 в разработанную модель, при необходимости скорректировать анкету и дополнить данные.
5. Разработать базу данных для хранения информации по заданной модели клиента.

Контрольные вопросы к лабораторной работе

1. Что такое модель?
2. Дате определение машинного обучения.
3. Дайте определение классификации.
4. Дайте определение кластеризации.
5. Дайте оценку разработанной модели.

# Лабораторная работа №5

Тема: Реализация инструментальной поддержки принятия кредитных решений. Разработка программного решения анализа кредитоспособности клиентов.

Система поддержки принятия решений (или сокращенно СППР) (а по англ. Decision Support System, DSS) – это компьютерная автоматизированная система, целью которой является помощь людям, принимающим решение в сложных условиях для полного и объективного анализа предметной деятельности. Это означает, что она выдаёт информацию (в печатной форме, на экране монитора, или звуком), основываясь на входных данных, помогающую людям быстро и точно оценить ситуацию и принять решение. СППР возникли в результате слияния управленческих информационных систем и систем управления базами данных.

Другое определение системы поддержки принятия решений – это класс информационных систем, которые обеспечивают руководителей различных уровней знаниями и информацией, позволяющими принимать более обоснованные и правильные управленческие решения в различных сферах деятельности. Что говорит о том, что основная цель данных систем, это помощь в принятии решений на основе анализа данных о предметной области.

Информационные системы, поддерживающие принятие решений, являются промежуточным звеном в управлении бизнес-процессами любых предприятий и организаций. В частности, это продукт, помогающий лицам, принимающим решения (или сокращенно ЛПР) выстраивать возможные варианты развития событий, исходя из заданных алгоритмов и загруженной базы данных. В результате менеджеры среднего и высшего звена принимают более взвешенные решения, даже в условиях динамично изменяющейся обстановки.

Источниками СППР становятся факторы и бизнес-процессы, касающиеся деятельности определенного объекта. Дополнительно учитывается опыт сотрудников в предметной области. На выходе пользователь получает гибкий анализ данных и имитационную модель событий, необходимые для принятия взвешенного решения в сложившейся ситуации, с учетом всех первостепенных и второстепенных факторов.

Можно выделить следующие классы СППР:

* Пассивные, которые поддерживают процесс принятия решений, но сами эти решения не формулируют и не предлагают, а только предоставляют информацию и отчеты: решения человек принимает сам;
* Активные – не только предоставляют информацию, но и сами могут предлагать решения и альтернативы;
* И комбинированные, которые обеспечивают итеративный процесс взаимодействия между ЛПР и СППР для достижения консолидированного решения: ЛПР (или его консультант) может изменять, дополнить или уточнить предложения СППР и отправлять их обратно в систему для проверки. Данный процесс продолжается до тех пор, пока не будет получено оптимальное решение.

По соотношению данные\модели (по методике Стивена Альтера) выделяют также следующие виды СППР:

* Файловые системы. Этот тип СППР в первую очередь обеспечивает доступ к хранилищам данных/элементам связанных данных.
* Системы анализа данных. Такой тип СППР поддерживает манипулирование данными с использованием конкретных или общих компьютеризированных настроек или инструментов.
* Аналитические информационные системы. Данный тип СППР обеспечивает доступ к наборам баз данных, ориентированных на принятие решений, и простым небольшим моделям.
* Учет и финансовые модели. Такая СППР может обрабатывать вопросы «что делать, после анализа» и вычислять результаты различных путей решений.
* Репрезентативные модели. Этот тип СППР также обрабатывает вопросы «что делать, после анализа» и вычислять результаты различных путей принятия решений на основе созданных моделей.
* Модели оптимизации. Данная СППР предоставляет решения с использованием оптимизационных моделей, которые имеют математические решения.
* И модели предложений. Этот тип СППР работает, когда решение, которое необходимо принять, основано на хорошо структурированных задачах.

Классификацию по модели Альтера можно разделить на 5 типов компьютерных СППР:

* Коммуникативные. Ориентированы на одновременную работу нескольких специалистов, занятых одной общей задачей.
* Информационные. Сосредоточены на сборе и обработке данных, преимущественно анализе временных рядов, функционируют, как СУБД в пределах одной компании.
* Документальные. Предназначены для обработки и анализа документов разного формата со структурированными и не структурированными данными.
* Интеллектуальные. Содержат данные о решениях аналогичных задач, нормы и правила, на основании которых они принимались, предлагают готовые алгоритмы, исходя из накопленного опыта.
* Моделируемые. Подбирают модели бизнес-процессов по заданным условиям (к ним относятся статистические, финансовые, аналитические СППР).

Современные виды систем принятия решений наделены следующими возможностями:

* Формирование статистики и ее проверка;
* Составление трендовых прогнозов;
* Планирование и контроль качества;
* Финансовый анализ и прогнозирование;
* Анализ рисков и скрытых закономерностей;
* Отслеживание поведения человека и формирование клиентских групп;
* Управление активами;
* Перераспределение задач между сотрудниками и пр. задачи, где может быть применим какой-либо алгоритм для помощи в поддержки принятия решения.

СППР реализующие методы интеллектуального анализа данных называются интеллектуальными. Рисунок 1 иллюстрирует множество методов, которые могут быть использованы для поддержки принятия решений в кредитном скоринге.

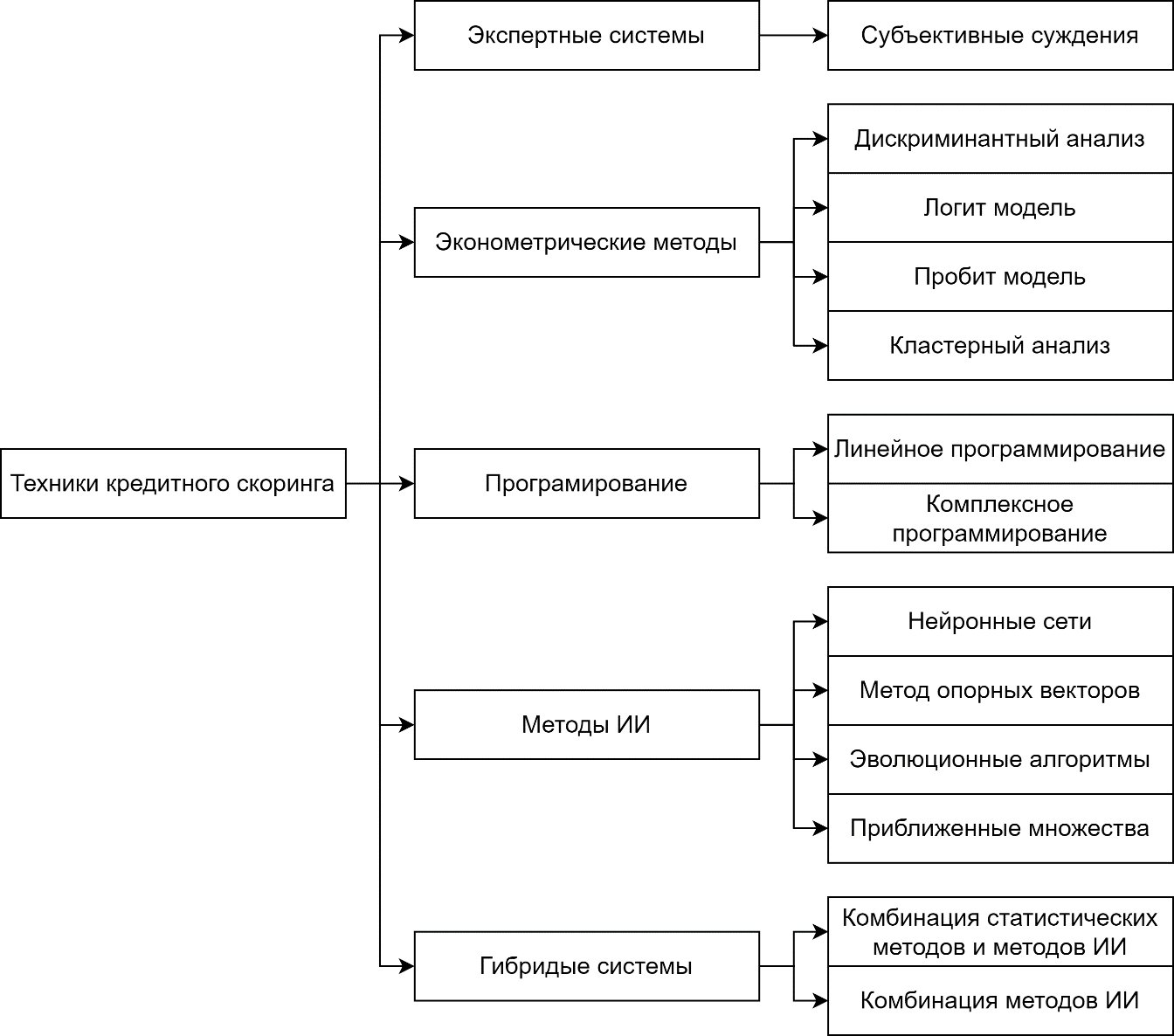


Рисунок 1. Методы кредитного скоринга.

Экспертные системы могут быть самым основным методом, использующим субъективный анализ для оценки кредитного риска. В экспертных системах индекс оценки является преобладающим фактором, определяющим окончательные результаты оценки. Например, самой известной системой индексов оценки корпоративного кредита является система 5С (Альтман, 1968), в которой такие атрибуты, как характер, мощность, капитал, обеспечение и состояние (character, capacity, capital, collateral, condition) оцениваются субъективно (т. е. экспертным суждением). Однако подход экспертной системы сильно зависит от субъективного суждения; следовательно, результаты оценки кредитного риска могут быть предвзятыми или дискриминационными по сравнению с другими количественными методами, основанными на исторических данных.

Эконометрические подходы являются наиболее популярными количественными методами в областях анализа данных и прогнозирования, основанных на математике, статистике и информатике. Наиболее популярными подходами для кредитного скоринга являются дискриминантный анализ, логит (или пробит) модели, кластерный анализ и метод k-ближайших соседей.

Логит-модель представляет собой типичную вероятностно-статистическую модель классификации. В исследованиях было показано, что интеграция факторного анализа и логит-оценки показала более лучшие результаты. Пробит-модель также является важным методом классификации в предположении, что кумулятивное распределение вероятностей должно быть стандартизированным нормальным распределением. По сравнению с другими подходами пробит-модель использовалась в относительно небольшом количестве исследований по кредитному скорингу.

Метод k-ближайших соседей — это стандартный непараметрический подход к классификации, учитывающий только k-наиболее похожих экземпляров данных. Часто используется для индивидуальной кредитной оценки и показывает хорошие результаты.

В 2002 г. была построена модель квадратичного программирования для получения индивидуального кредита, которая включала экспертные оценки. Численные эксперименты показали, что модель оценки, включающая экспертные сведения, может повысить качество. В 2005 г. была предложена модель классификации, использующая многокритериальное линейное программирование для выявления моделей поведения заявителей на получение кредитных карт.

Искусственные нейронные сети (ИНС) представляет собой модель обработки информации, напоминающую структуру связей в синапсах. Они состоят из большого количества узлов (также называемых нейронами или единицами), соединенных связями. Нейронная сеть прямого распространения с обратным распространением ошибки широко используется для кредитного скоринга, когда нейроны получают сигналы от предварительного слоя и выводят их на следующие слои без обратной связи. На рисунке 1 показана стандартная структура сети с прямой связью, которая включает входной слой, скрытый слой и выходной слой.

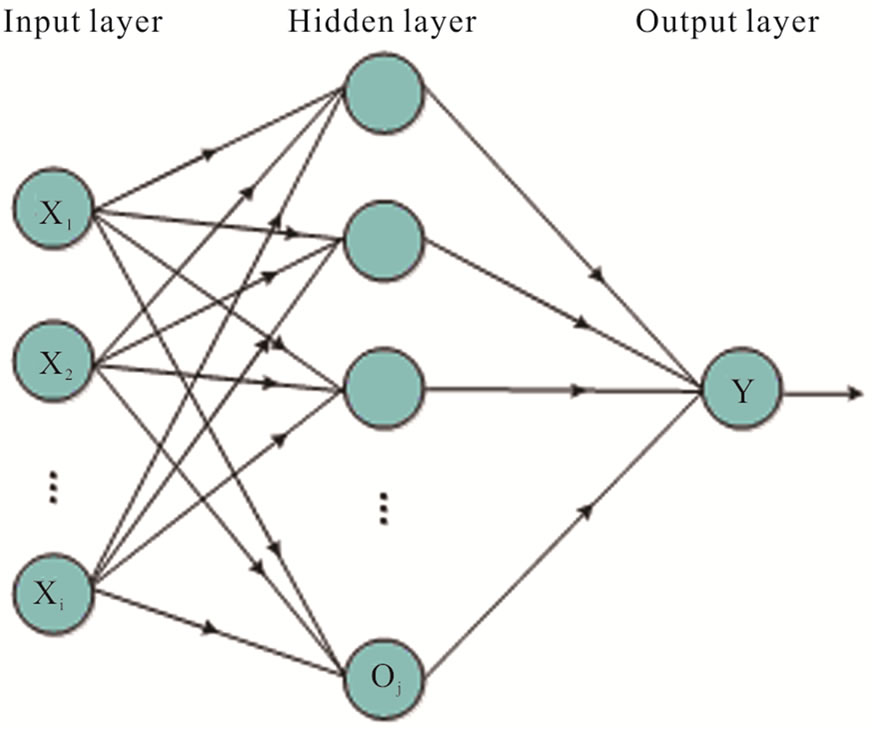


Рисунок 1. Пример нейронной сети

Узлы входного слоя получают значения атрибутов каждой обучающей выборки и передают взвешенные выходные данные скрытому слою. Взвешенные выходные данные скрытого слоя вводятся в модуль, составляющий выходной слой, который выдает прогноз для заданных выборок. Обратное распространение обучается путем итеративной обработки набора обучающих выборок, сравнивая предсказание сети для каждой выборки с фактически известным классом. Для каждой обучающей выборки веса изменяются таким образом, чтобы минимизировать среднеквадратичную ошибку между прогнозом сети и фактическим классом. Модификации выполняются в «обратном» направлении, поэтому мы называем сеть обратным распространением. На скрытом уровне могут применяться различные функции активации. Преимущества нейронных сетей включают их сильную способность к обучению и отсутствие предположений о взаимосвязи между входными переменными. Однако он также имеет некоторые недостатки. Главный недостаток нейронных сетей заключается в их плохой возможности объяснения результатов. Из-за природы «черного ящика» для нейронной сети очень сложно обеспечить представление знаний. Вторая проблема заключается в том, как спроектировать и оптимизировать топологию сети. Это представляет собой очень сложный экспериментальный процесс. Очевидно, что количество единиц и слоев в скрытом слое, различные функции активации и начальные значения веса могут повлиять на окончательный результат классификации. Кроме того, ИНС требует большого количества обучающих выборок и длительного времени обучения.

В настоящее время популярной темой исследований стали гибридные модели, использующие преимущества различных методов. Однако нет четких инструкций, как классифицировать гибридные модели. Как правило, классификация используется в соответствии с другим методом, используемым на этапах выбора признаков и классификации. Было выделено четыре типа: кластеризация + классификация, классификация + классификация, кластеризация + кластеризация и классификация + кластеризация.

Интеллектуальные системы поддержки принятия решений (ИСППР), представляют собой автоматизированную компьютерную сеть, обеспечивающую объективную аналитику данных с построением математической модели предполагаемого развития событий. Задача таких продуктов – помощь людям, принимающим сложные управленческие решения в сложных условиях.

ИСППР состоит из следующих компонентов:

* база данных (или база знаний);
* модели данных;
* интерфейса пользователя.

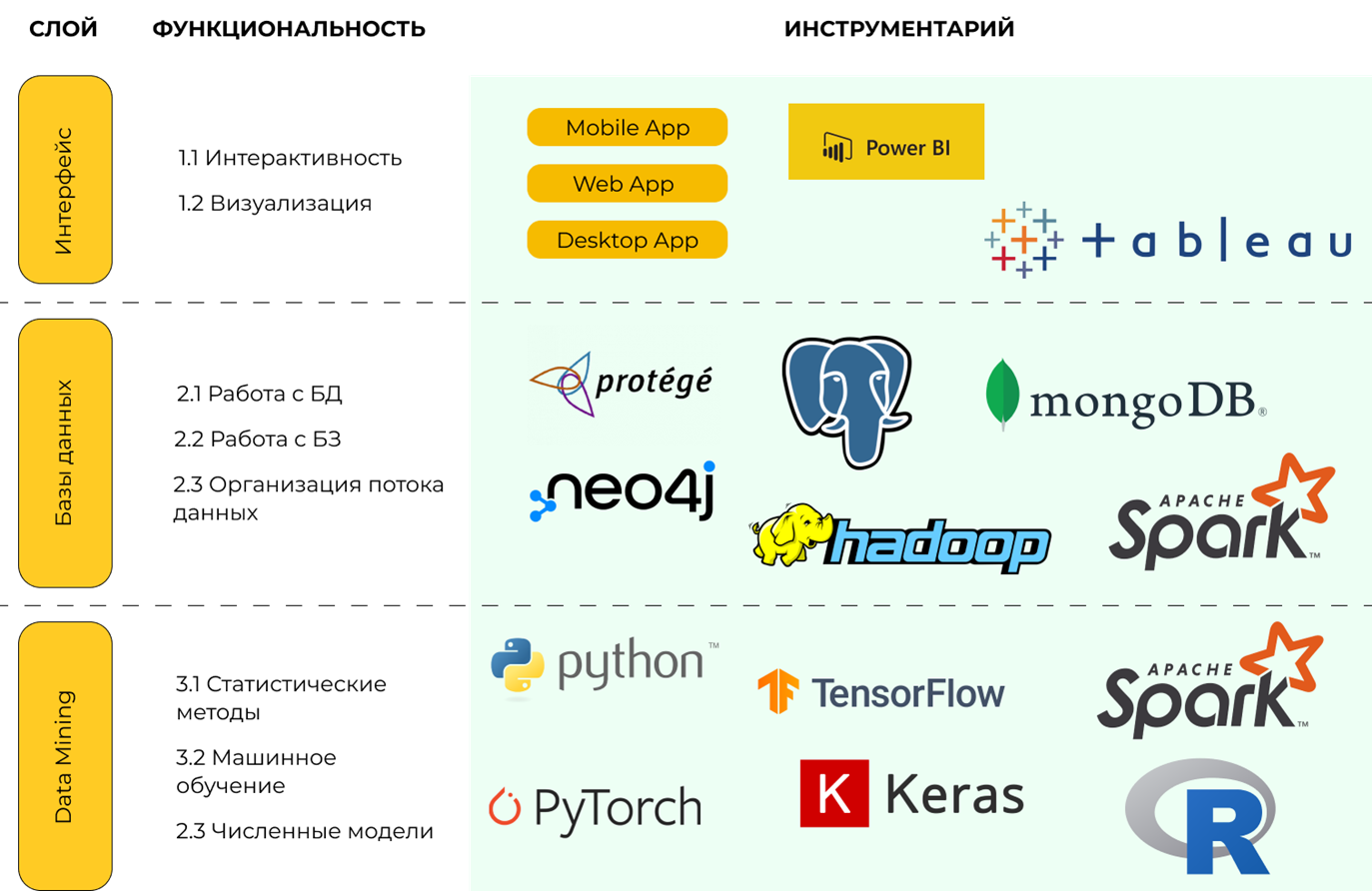


Рисунок 1. Структура ИСППР с основными инструментами, используемыми на разных уровнях архитектуры.

Минэкономразвития России выпустило приказ №392 от 29.06.2021 «Об утверждении критериев определения принадлежности проектов к проектам в сфере искусственного интеллекта», где отражаются основные задачи и критерии в сфере искусственного интеллекта, в который вошли также основные задачи, которые должны решать, в том числе, ИСППР.

В соответствии с этим приказом выделяют следующие задачи, которые должны решаться ИСППР:

1. Предиктивный и прескриптивный анализ, позволяющий предсказывать развитие ситуации на основе анализа данных и автоматизировать принятие решений в режиме реального времени (включая создание методов и моделей).

2. Оценка качества моделей машинного обучения без тестирования в реальной среде, в том числе в рекомендательных системах, тестируемых без участия пользователя.

3. Подготовка решений на основе открытых источников данных и неструктурированной информации, в том числе для использования в интеллектуальных системах поддержки принятия решений для решения стратегических вопросов и (или) адаптивного динамического управления сложными объектами.

4. Интеллектуальное имитационное моделирование поведения участников рынка товаров, работ и услуг на основе транзакционных данных и моделей машинного обучения.

5. Управление и обучение персонала, и построение персонализированных карьерных или образовательных траекторий.

6. Обеспечение поддержки принятия решений на основе многолетних данных, в том числе для расчета нормирования в отраслях экономики.

Поскольку целью данного курса, является банковская сфера, то отдельно стоит выделить задачи, решаемые СППР в банковском секторе:

* Это выдача и отслеживание использования банковских карт;
* Анализ риска невозврата кредитов;
* Борьба с мошенниками (по банковским картам, и в сфере кредитования);
* Анализ поведения потенциальных клиентов;
* Проектирование новых услуг;
* Формирование индивидуальных предложений;
* Отслеживание расчетов по займам;
* Своевременное информирование о сроке платежа и просрочках;
* И взыскание задолженности.

В сфере кредитования эффективны те типы систем поддержки принятия решений, которые ориентированы на анализ поведения потребителей и их классификацию по группам со схожими поведенческими признаками. В зависимости от них кредитное учреждение формирует целевые предложения: жилищные кредиты молодым семьям, программы overdraft для вип-клиентов, кредитование пенсионеров, начинающих предпринимателей и т. д.

Внедрение СППР повышает безопасность сделок. Обрабатывая анкету заемщика, менеджеру придется переносить данные из нее в систему. По мере заполнения граф анкеты, автоматически сформируется ответ на тематический вопрос – относится ли клиент к нужной категории заемщиков. Например, попадает ли он в категорию граждан или компаний, для которых разработаны льготные программы кредитования. Последовательное заполнение граф анкеты «отдает» команду системе на проверку введенных данных. В результате оценка значимости клиента производится не субъективно, а при поддержке авторитетного ПО, что исключает человеческий фактор.

Применение СППР в кредитовании позволяет проверять клиента на предмет совершения мошеннических действий. В той же программе, которая проверяла клиента на соответствие условий для получения займа, можно проверить его чистоту – со своим ли паспортом обратился клиент, активен ли номер телефона и т. д.

В узкоспециализированных кредитных DSS-системах реализованы технологии, позволяющие отлеживать погашение займов: предупреждать о приближении срока платежа посредством СМС, push-уведомлений, звонков из колл-центра, фиксировать просрочку, работать с ней, начислять штрафные санкции и распределять всю эту работу между менеджерами, отвечающими за погашение и взыскание.

Всё это, делает СППР важной частью в работе современных банков, позволяя за счет использования таких средств сокращать риски и издержки, связанные с кредитованием, а также повышать качество и точность в обслуживании клиентов.

Для выполнения работы необходимо:

1. Для модели клиента из ЛР4 разработать модель машинного обучения для оценки кредитоспособности клиентов.
2. Провести обучение модели на данных, полученных в ЛР2.
3. Разработать тестовый набор данных и проверить на нём работу модели.
4. В случае необходимости скорректировать модель клиента и модель оценки, и произвести повторное обучение до момента получения корректных данных.
5. Оценит полученную модель по уровню точности и уровню потерь.

Контрольные вопросы к лабораторной работе

1. Назовите основные инструменты для реализации машинного обучения.
2. Какие параметры характеризуют качество модели машинного обучения?
3. Перечислите основные этапы машинного обучения.
4. Назовите основные способы машинного обучения.
5. Назовите наборы данных, используемые при обучении моделей.

# Список литературы

1. Советов, Б. Я. Моделирование систем: учебник для студентов вузов, обучающихся по инженерно-техническим направлениям и специальностям / Советов Б. Я., Яковлев С. А.; С. - Петерб. гос. электротехн. ун-т "ПЭТИ им. В. И. Ульянова-Ленина". - 7-е изд. - Москва: Юрайт, 2014. - 343 с.: ил. Гриф: УМО РФ (10 экз).
2. Федотова, М. Ю. Организация деятельности коммерческого банка и современные банковские технологии : учебное пособие / М. Ю. Федотова. — Пенза : ПГАУ, 2020. — 262 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/170974 (дата обращения: 09.10.2021). — Режим доступа: для авториз. пользователей.
3. Козлова, Г. Г. Информационные системы и технологии банковского дела : учебное пособие / Г. Г. Козлова. — Москва : РТУ МИРЭА, 2020. — 112 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: https://e.lanbook.com/book/163893 (дата обращения: 09.10.2021). — Режим доступа: для авториз. пользователей.
4. Советов, Б. Я. Моделирование систем. Практикум: учебник для студентов вузов, обучающихся по направлениям "Информатика и вычислительная техника" и "Информационные системы" / Советов Б. Я., Яковлев С. А.; С. - Петерб. гос. электротехн. ун-т. - 4-е изд., перераб. и доп. - Москва: Юрайт, 2013. 295 с.: ил. Гриф: МО РФ (10 экз).
5. Афанасьева, Т. В., Афанасьев А.Н. Введение в проектирование систем интеллектуального анализа данных: учебное пособие. – Ульяновск : УлГТУ, 2017. 64 с. http://venec.ulstu.ru/lib/disk/2017/172.pdf
6. Прогнозирование временных рядов: нечеткие модели / Т. В. Афанасьева, А. М. Наместников, И. Г. Перфильева, А. А. Романов, Н. Г. Ярушкина; под науч. ред. Н.Г. Ярушкиной. – Ульяновск : УлГТУ, 2014. – 145 с. http://venec.ulstu.ru/lib/disk/2015/192.pdf
7. Афанасьева, Т. В., Ярушкина Н.Г. Моделирование в задачах анализа свойств систем : учебное пособие. – Ульяновск : УлГТУ, 2019. 114 с.
8. Клячкин, Владимир Николаевич. Сборник заданий по статистическим методам анализа данных : учебное пособие / В. Н. Клячкин, Ю. Е. Кувайскова, В. А. Алексеева. – Ульяновск : УлГТУ, 2016. – 123 с. http://venec.ulstu.ru/lib/disk/2016/207.pdf
9. Эконометрика : учебное пособие / Н. И. Шанченко. – Ульяновск : УлГТУ, 2018. – 136 с. http://venec.ulstu.ru/lib/disk/2017/374.pdf
10. Афанасьева Т. В. Применение методов интеллектуального анализа данных и процессов: Лабораторный практикум. [Электронный ресурс]: – Ульяновск : УлГТУ, 2017. https://virtual.ulstu.ru/extranet/workgroups/group/3813/
11. Воронина В. В. Теория и практика машинного обучения: учебное пособие /В. В. Воронина и др.. – Ульяновск: УлГТУ, 2017. – 290 с. <http://venec.ulstu.ru/lib/disk/2017/191.pdf>
12. Николенко С., Кадурин А., Архангельская Е. Глубокое обучение. СПб.: Питер, 2018. 480 с.: ил. ISBN 978-5-496-02536-2.
13. Пойтнер Я. Программируем с PyTorch. Создание приложений глубокого обучения. СПб.: Питер, 2020. 256 с.
14. Шолле Ф. Глубокое обучение на Python. – СПб.: Питер, 2018.

ПРОЕКТИРОВАНИЕ И РАЗРАБОТКА ИНФОРМАЦИОННЫХ СИСТЕМ ДЛЯ ПРИНЯТИЯ КРЕДИТНЫХ РЕШЕНИЙ

Лабораторный практикум

Авторы РОМАНОВ Антон Алексеевич,   
ШЕРКУНОВ Вячеслав Владимирович

Усл. печ. л. 4

УлГТУ, 432027, г. Ульяновск, ул. Сев. Венец, д. 32.