***ФГБОУ ВО «Кубанский государственный аграрный университет имени И.Т. Трубилина»,***

***Российская Федерация***

***Ватамов Д.Н. ПИ2104***

**156 Теория информации, данные и знания. Практическая 4. Инструкция по разработке собственного приложения 2020-10-23**

## Заголовок

Разработка облачных Эйдос-приложений: Инструкция, поиск данных (UCI, Kaggle) и публикация

## Резюме текста (1 страница)

**1. Введение и контекст занятия**
Практическое занятие №4 для группы ИТ-2001 по дисциплине "Теория информации, данные, знания" состоялось 23 октября 2020 года. Тема занятия: инструкция по разработке собственных интеллектуальных облачных Эйдос-приложений. Преподаватель подчеркнул важность взаимного приветствия и отметил присутствующих.

**2. Постановка задачи и её важность**
Основная задача студентов – не просто ознакомиться с инструкцией по разработке Эйдос-приложения, размещенной на сайте преподавателя, а приступить к её непосредственному выполнению. Это задание является ключевым для получения оценки на экзамене (вплоть до "самоэкзамена" при полном выполнении всех шагов). Лекции дают знания, а практические занятия – навыки их применения через выполнение конкретных заданий.

**3. Проблема: Поиск исходных данных**
Выяснилось, что студенты не знают, где искать наборы данных (датасеты) для машинного обучения, необходимые для разработки своих приложений.

**4. Источники данных и требования к теме (Пункт 4 Инструкции)**
Преподаватель подробно разбирает пункт 4 инструкции, касающийся выбора темы и поиска данных:

* **Требования к теме:**
	+ Уникальность: Тема не должна повторять уже существующие приложения (список доступен по ссылке на сайте).
	+ Соответствие нормам: Тема не должна противоречить морально-этическим нормам и законодательству РФ.
* **Источники данных:**
	+ **UCI Machine Learning Repository:** Старый, авторитетный репозиторий с огромным количеством датасетов для различных задач (в основном классификации). Данные представлены в разных форматах (категориальные, целочисленные, "действительные"), часто требуют предварительной обработки и преобразования для использования в системе "Эйдос". Преподаватель имеет опыт работы с UCI и статьи на его основе.
	+ **Kaggle:** Более современная платформа, известная своими конкурсами по машинному обучению с призовыми фондами. Данные часто представлены в удобном формате CSV, который легко импортируется в "Эйдос". Рекомендуется выбирать датасеты умеренного размера (мегабайты, а не гигабайты) и, по возможности, с разделением на обучающую и тестовую выборки. Следует избегать сложных многофайловых наборов данных, требующих программирования для объединения.
	+ **Статистические данные:** Росстат, криминальная статистика и т.д. Позволяют исследовать реальные зависимости (например, влияние промышленности на экологию и качество жизни).
* **Система "Эйдос":** Способна обрабатывать большие объемы данных (протестировано до 4 млн записей) и различные форматы ввода (текстовый, табличный Excel/DBF, CSV, графический).

**5. Регистрация на научных платформах и публикация результатов**
Для полноценного выполнения задания и получения "самоэкзамена" необходимо:

* **Зарегистрироваться на ResearchGate:** Требуется корпоративная почта (типа @edu.kubsau.ru). Студентам рекомендуется добиваться получения таких адресов через куратора или ЦИТ. Платформа позволяет публиковать препринты, отчеты, статьи с присвоением DOI.
* **Зарегистрироваться в РИНЦ и Science Index:** Не требует специальной почты. Необходимо для получения SPIN-кода.
* **Получить SPIN-код:** Дает право заключать договор с РИНЦ на размещение непериодических изданий (монографий, пособий, отчетов, препринтов).
* **Получить DOI (Digital Object Identifier):** Цифровой идентификатор публикации, присваиваемый через ResearchGate (бесплатно) или другие сервисы (платно). Обеспечивает уникальность и цитируемость работы в базах данных, включая РИНЦ.

**6. Процесс выполнения задания и мотивация**
Студенты должны выбрать тему, найти данные, разработать приложение в системе "Эйдос", подготовить описание работы по шаблону (используя статьи преподавателя как пример) и разместить его на ResearchGate (с DOI) и в РИНЦ (через договор). Это не только задание для экзамена, но и возможность получить ценные навыки, пополнить портфолио и повысить шансы на стипендию.

**7. Заключение**
Преподаватель призывает не откладывать выполнение задания и регистрации на платформах, так как это требует времени. Предлагает обращаться с вопросами по почте.

## Детальная расшифровка текста

**1. Введение и приветствие**

Здравствуйте, ребята.
Вы должны поздороваться, ребят, с вами здороваются.
Здравствуйте.
Да, здравствуйте.
Вот когда-то раньше, когда в аудиториях мы встречались, преподаватель здоровался, и студенты с ним здоровались. И сейчас тоже так же надо. С вами преподаватель здоровается, вы тоже с ним здороваетесь.

**2. Контекст занятия**

Сегодня 23 октября 2020 года. Четвертая пара. 13:50-15:20.
Четвертое практическое занятие с группой ИТ-2001 по дисциплине Теория информации, данные, знания.

**3. Тема занятия: Инструкция по разработке приложений**

На этом занятии у нас учебный вопрос: инструкция по разработке собственных приложений, собственных интеллектуальных облачных Эйдос-приложений.

**4. Организационные моменты: Посещаемость и доступ**

Теперь я сейчас отмечаю, что вас много здесь присутствует, хотя на самом деле совсем мало.
И отмечаю, что я присутствую.
В трех системах отметился. Теперь можно вести занятие. Причем я там отметил ту информацию, которая есть в Вебексе. Сколько вас присутствует, отметил. На 10 человек больше отметил, чем на самом деле. И себя отметил. И всё это известно, вся эта информация есть в Вебексе. Я еще в двух системах разместил эту информацию.

Так, ребята, теперь, значит, я вам открываю доступ к экрану. И мы переходим к изучению задания.

**5. Обзор задания-инструкции на сайте**

Заходим на мой сайт, ребята. Видите мой сайт, да?
Да.
И второй пункт я выбрал на сайте. Здесь вот есть, где картинка, справа от неё чуть пониже написано: Задание-инструкция для учащихся по разработке собственного интеллектуального облачного Эйдос-приложения.

**6. Проверка понимания и выполнения задания**

Мы с вами эту инструкцию изучали уже или нет? Вы отвечаете.
Изучали.
На каком пункте остановились?
Извините...
Вот.

**7. Цели и задачи практических занятий**

Значит, вы изучаете дисциплину. У нас есть лабораторные и практические занятия. На лекционных занятиях вы должны осваивать знания. То есть я вам рассказываю теоретический материал, который вы должны как бы понять просто, знать про это. А вот на лабораторных и практических занятиях вы должны получать навыки применения этих знаний, тренировать умение, то есть как-то один раз я вам показываю, как это делается, а потом должны навыки формировать.

**8. Необходимость выполнения задания**

И это делается путем выполнения заданий. Вот вам это задание, здесь описано. Какое задание? Вам надо непосредственно приступить к его исполнению, прямо вот сейчас этого задания. Потому что я так чувствую, что вы как на Марсе там живете где-то. Вот скажите мне сейчас, пожалуйста, кто из вас выполнял мое задание уже, которое я вам дал? Возьмите сейчас в чате, поставьте плюсики. Кто из вас начал выполнение задания, которое я вам дал?
Если вы видите, что я это...
Какое задание?
А я ж вас сейчас спросил, вы видели это задание или нет? Вы сказали, что видели. Эту инструкцию, задание-инструкцию вы видели, я спрашиваю, или нет?
Да.
Видели. Это задание-инструкция, ребята. Вы должны начать его выполнять, это задание. Я вас спрашиваю, кто начал его выполнять это задание? Если вы его видели. Те, кто его видел, сразу начинают его выполнять, потому что это задание, по которому будет выставляться оценка на экзамене. По результатам выполнения которого будет выставляться оценка. То есть это не такое задание, которое надо на экране посмотреть и забыть, понимаете? А это задание, которое надо выполнить.

Поставьте плюсики, кто начал выполнять это задание в чате.
(Студент ставит плюс)
Лада, а что извините? За что извините? Вы можете микрофоном пользоваться, пожалуйста, пользуйтесь микрофоном. Лада, за что вас извинить?
Мм. Вот, видите, Артём уже определяется с темой. Молодец.
(Студент: Я пока с темой определяюсь, пока что мне... Да, найти для машинного обучения.)

**9. Уточнение пройденного материала и проблема поиска данных**

А я ж вам рассказывал, где их найти? Или еще не успел?
(Студент: Пока нет, пока не успел.)
Не успел. Поэтому я вас и спрашиваю, на каком пункте мы остановились.

Значит, ребята, тогда, я так понял из того, что вот вы сейчас мне сказали, и из того, что вот я сам, так сказать, помню, я записываю просто в самом расписании, записываю... Ну, видимо, надо еще пункт записывать. Вот я записал себе инструкция по разработке собственного приложения. А какой пункт, вы не знаете. Вот я сейчас вас поэтому и пытал.

Значит, ребята, вот судя по тому, что вы сейчас сказали, значит, такое впечатление возникает, что вы не знаете, где искать данные для собственного Эйдос-приложения. Значит, пожалуйста, мне сейчас вот ответьте плюсиками, кто знает, где искать данные для облачного Эйдос-приложения? А минусиками, кто не знает. Вот давайте, отмечайте. Если бы мы были в аудитории, я бы просто спросил, кто знает? Вы бы ответили: там я знаю, я знаю, там, или руку подняли бы. А здесь такой возможности у нас нет. Хотя можете сделать, чтобы я видел ваши лица, видеокамеры включить и посмотреть. То есть я тогда посмотрю на вас.

Ну, никто не знает, короче говоря, никто не знает, где брать тему.
А?
(Студент: Там в пункте всё сказано, где искать их.)
Я не разобрал.
(Студент: В четвертом пункте задания-инструкции сказано, где искать данные для машинного обучения.)
Конечно, конечно. Но я просто всё это комментирую, объясняю как бы подробнее. Ну я так понял, что никто не знает, короче говоря. Ну да, вы верно сказали.

**10. Разбор Пункта 4 Инструкции: Требования к теме и источники данных**

 \*\*10.1. Начало разбора пункта 4\*\*

Значит, смотри. Короче так, вот у нас смотрим четвертый пункт. Начинаем с четвертого пункта изучать.

\*\*10.2. Требования к теме приложения\*\*

Значит, к теме, к теме предъявляются определенные требования. Значит, эти требования здесь все написаны.

\* \*\*Уникальность:\*\* Первое требование - эта тема не должна повторяться с теми, которые уже выполнены. А какие выполнены, у нас написано вот здесь вот, вот ссылочка сразу же на те приложения, которые уже существуют. Мы видим, что здесь довольно много есть приложений. Вот, их на самом деле 208, а здесь что-то в кэше застряло и показывает 207. А 208 не показывает. Хотя вот даже F5 нажимаешь, она не переписывает кэш.

 \* \*Примечание преподавателя:\* (Преподаватель замечает ошибку в отображении описания приложения №191 на сайте, находит его локально и исправляет/загружает на сервер). Отсутствует описание нормальное... Смотрим здесь у меня... Вот у меня все эти приложения... 191-е приложение... Readme... Всё здесь в порядке... Значит, я сейчас это описание записываю на FTP-сервер... Вообще-то эта система делает сама, но иногда бывает интернет плохо работает, и могут быть всякие там проблемы... Вот... Ну да, видно, что повреждено... (Ожидание загрузки)... Вот. И смотрим... Перезагрузить... Вот сейчас только что мы его закачали, теперь оно там всё есть.

 \* Значит, вот у каждого, ну не у всех, если честно, не у всех приложений, но у многих есть описание. Видите, тут кое-где нет описания, пустые строчки... А так есть вот ссылочки на все... Лучше брать, конечно, если как смотреть как прототип, то брать лучше новое, как они сделаны. Но прототипы там есть, описание, их называю шаблонное описание.

\* \*\*Соответствие нормам:\*\* Так вот, ребята, тема приложения должна удовлетворять следующим требованиям. Первое - её не должно быть вот здесь среди уже созданных приложений. То есть такое же точно брать не надо. Второе. Значит, эта тема приложения не должна противоречить общепринятым в России морально-этическим нормам и действующему законодательству. То есть это не должны быть темы, которые могут нарушать там общественный порядок, призывать там к беспорядкам или там описывать, как производить марихуану на балконе. Ну понятно, да? То есть, короче говоря, это такое, должны быть разумные, нормальные темы.

\*\*10.3. Источник данных: UCI Machine Learning Repository\*\*

Вот. Теперь, где и брать эти данные для, э-э, данные для обучения, для создания приложения? Значит, есть замечательные сайты. Вот есть сайт UCI. Вот, который содержит очень большое количество различных наборов данных для машинного обучения. Это называется датасет. Данные для машинного обучения датасет. А когда их много, то это называется датасетс. То есть огромное количество баз данных. Я даже и не знаю, сколько их, не считал, но их количество постоянно возрастает. И здесь дана их характеристика. Можно, используя Google Chrome, правой клавишей клацнуть, перевести на русский.

Значит, какие тут типы данных есть? Какая задача по умолчанию? Какие типы атрибутов? Сколько примеров обучающей выборки? Сколько атрибутов? И год. И вот мы видим, видите, классификация, классификация там задачи. Ну, в основном задачи классификации. Это что это за задача классификации? Это задача отнесения объекта конкретного к какому-то обобщенному классу, которые должны быть сформированы в результате использования обучающей выборки. Ну, примерно, как я вам показывал.

Какие типы данных, смотрите, существуют: категориальные, целочисленные и действительные. Категориальные - это значит символьные, текстовые переменные или логические переменные. То есть 0, 1 там, или да, нет, что-нибудь такое, или текстом может быть написано. Вот. Целочисленные и со знаками после запятой. Вот насчет действительных, они, конечно, маху дали. Потому что в школе учат, что действительные числа - это числа, вообще говоря, с бесконечным числом знаков после запятой, включая иррациональные. Ну, то есть это вот целые и иррациональные числа. То есть это, в частности, такие числа, как число Пи, например. Вы же понимаете, что число Пи нельзя ввести в компьютер. Даже его нельзя ввести во все компьютеры Вселенной, не хватит нашей галактики, соседних галактик, не хватит этих всех носителей для того, чтобы это сделать, потому что в нем бесконечное количество знаков после запятой.

Вот. Система Эйдос обеспечивает обработку и текстовых переменных, и целочисленных, и с дробной частью, то есть с какими-то знаками после запятой, она это обеспечивает.

Значит, вот это вот количество объектов обучающей выборки - 48.000. Для неё это вообще семечки. Я сейчас апробировал систему на 4 млн записей обучающей выборки. То есть она с 4 млн записей работает. Для неё вот это 100.000 записей - это не так много. Или там 129.000 там. А вот здесь вот, смотрите, есть пример анонимные веб-данные там какого-то сайта. Там 989.818 записей. А вот есть, смотрите, КДП, КУАП, данные. А там похоже, что 4 млн записей, видите? Вот я вот такую вот задачу, где 4 млн записей, сейчас э-э, ввёл эти данные в систему, создал модели и стал их анализировать. То есть для того, чтобы проверить, как система работает с большим числом объектов обучающей выборки. Она работает с таким числом. Миллион - это даже, видите, в четыре раза меньше, чем то, что я проверял просто вот недавно. Правда, надо признать, что она работает медленно, конечно. Здесь вот тоже около миллиона. А вот здесь вот у нас аж 8 млн. Вот такое тоже она, по идее, должна обрабатывать, и 8 млн. Но здесь у нас 100.000 чего? 100.000 шкал, что ли? Атрибутов. Ну это тоже много, но она такое тоже может обрабатывать. В общем, было бы интересно.

\*\*10.4. Связь UCI с работами преподавателя\*\*

Значит, это сайт древний. Значит, я сейчас вам дам ссылочку на своем сайте, покажу вам, где у меня есть статьи, в названии которых встречаются сочетание символов UCI. Вот смотрите, ребята. Сейчас я вам даю ссылочку. Ссылочка вам знакома. Я про неё вам говорил, когда рассказывал, где скачать систему Эйдос. Ага. Ясно. Вот смотрите, эта ссылочка, где можно скачать систему Эйдос, да? А вот здесь вот, смотрите, если мы зададим сейчас поиск UCI, то мы увидим, что тут там сколько статей можно найти? Ребят. У нас практическое занятие, вы должны отвечать на вопросы. Сколько статей можно найти, если задать поиск текстов, в которых есть UCI на этой страничке? Давайте, делайте этот запрос и ищите. Сколько текстов можно найти? Отвечайте мне быстренько, быстренько. Правильно. Марк, молодец. Ты что, разглядел на моем экране это что ли, или сам сделал запрос? Разглядел. Ну, тоже всё равно молодец.

Значит, ребята, смотрите. Значит, у меня есть семь статей, в которых я решаю задачи какие-то на основе наборов данных, которые вот на этом сайте UCI. Самая первая статья, как вы видите, 2003 год. То есть уже 17 лет назад этот сайт существовал, я его нашел. Как я его нашел? Я его нашел просто, когда искал данные для машинного обучения. Нашел этот сайт. Вот. И мне понравился он, в принципе. Но там единственная проблема - там данные в такой форме, что надо их преобразовывать как-то самому, чтобы ввести в систему.

\*\*10.5. Форматы ввода данных в систему "Эйдос" (CSV)\*\*

Значит, в системе есть интерфейсы ввода данных из внешних источников: текстовый, табличный (табличных даже два, и текстовых два). Вот, ну это обобщенно говорю: текстовый, табличный и графический. То есть можно изображения вводить. Вот. И, значит, э-э, еще есть формат ввода dBase-файлы, табличные, значит, это экселевские файлы, старый Excel, новый Excel и DBF-файлы, которые мало кому знакомы, но, может быть, кто-то слышал, те, кто с 1С работал. Вот. И есть еще CSV-файлы. CSV - это форматированный, э-э, стандартный формат, то есть форматированные текстовые файлы, в которых поля есть точно такие же, как в DBF или в Экселе. В Экселе - это колонки, в DBF-файле - это поля. Такие же точно поля есть, тоже разных типов данных, разного размера, ну, текстового и числового типа данных. И категориальные, то есть может быть и логического типа, но они все в виде текста представлены. Вот. Там может быть да/нет написано или 0/1. Так вот, они разделены запятой обычно. Бывает иногда изредка, что табуляцией или точкой с запятой, но я такого не встречал, обычно запятой.

И сразу вам могу сказать, что система Эйдос имеет интерфейс ввода CSV-файлов, и это позволяет вводить информацию в систему значительно большего объема, чем можно ввести из экселевских файлов. Экселевский файл, старый Excel позволяет 65.535 записей ввести. Файл, а новый Excel позволяет около миллиона, чуть побольше миллиона записей. И всё. А вот если задача там у нас в UCI стоит, э-э, в этом, э-э, в базе, да, датасет, а там написано 4 млн, 6 млн, 8 млн. Ну как их ввести в систему эти данные? С помощью, это можно сделать с помощью интерфейса 2.3.2.2. Там есть режим ввода данных из CSV-файлов. Вот тогда это можно сделать, именно из текстовых файлов можно ввести такой объем информации.

\*\*10.6. Использование статей преподавателя как примера\*\*

Теперь смотрите, ребята, я не буду вам делать статьи на все ссылки, то есть на все ссылки на все свои статьи, где есть сочетание слов UCI. Вы их легко можете найти. Для чего я вам говорю о том, что вы можете найти эти статьи? Их можно рассматривать как пример написания статьи, проведения исследования и написания статьи на основе данных этого репозитория UCI. Вот самая новая статья. То есть самая старая - 2003 год, а самая новая - 2018 год. То есть 2 года назад. А вот в прошлом году и в этом году я не писал статьи на основе данных этого репозитория UCI. Вот. Ну, можно глянуть, как она выглядит эта статья.

Я пишу статьи по определенному, так сказать, принципу. Сначала рассказываю про какая проблема решается, потом какие методы можно для этого применить, потом описываю новый метод, который для этого можно применить - это системно-когнитивный анализ. А потом идет разработка и потом идет применение, описание применения. Вот данные в таком виде - это типичный вид для UCI. Вот. Ну тоже такие данные можно ввести в Excel. И здесь описываю, как это сделать. Я всегда ссылаюсь на источник данных в статье. То есть я пишу статьи таким образом, чтобы можно было повторить всё, что здесь написано. И ссылаюсь всегда на авторов этих источников данных и на, если у них есть, то на их публикации. Обязательно. Вот. Вот я эти данные тоже ввёл и тут уже показан экселевский файл. И дальше уже описывается решение задачи.

Ну, поскольку это было 2 года назад, то сейчас возможности больше у системы, потому что я непрерывно над ней работаю. И появляются всё время новые возможности у неё.

Вот. Ну, статья более-менее новая, поэтому там более-менее возможности освещены, какие у неё есть. То есть вы можете вот так вот познакомиться с тем, как писать статьи и по форме, и по содержанию. Такие статьи можно рассматривать как шаблон описания.

\*\*10.7. Источник данных: Kaggle\*\*

Теперь смотрим на следующее, на следующие источники данных, на другие. Значит, мы видим также здесь сайт Kaggle есть, Kaggle. Сайт Kaggle, он имеет более позднее происхождение, чем UCI. UCI, вы видели, самые первые наборы данных - 95-й год, 96-й. Я их довольно быстро заметил, не прошло там и 10 лет, и я их заметил. Но сначала у них было довольно мало данных там. Вот, собственно, и все вот на 2003 год, вот если посмотреть, то данных не так много. А вот даже здесь уже 2006. А сейчас они очень интенсивно пополняют свои базы этих вот датасетс, то есть данных для обучения машинного. Потому что сейчас это очень популярно, интенсивно развивается, запрос, так сказать, есть запрос общества на такие данные.

Значит, сайт Kaggle, он возник гораздо позже, чем UCI, но он обладает по сравнению с ним очень интересной особенностью. Значит, которая заключается в том, ребята, слушайте внимательно. На сайте UCI просто, просто наборы данных. А на сайте Kaggle - конкурсы, конкурсы. И эти конкурсы, они есть действующие, текущие в настоящее время конкурсы, а также есть уже закрытые конкурсы, прошедшие. То есть конкурсы выполняются в течение определенного времени. Ну, обычно от одного до трех месяцев. Вот. То есть если вы обнаруживаете, что появляется новый конкурс, который будет длиться 3 месяца, то вы можете повозиться, у вас есть возможность по времени. Вот. То есть вы можете попробовать эту задачу решить, попробовать свои силы.

И на сайте Kaggle есть призовой фонд для тех, кто занимает призовые места. У них на некоторых конкурсах три призовых места, а на некоторых - пять. И распределение вот этой суммы призового фонда между призерами происходит. По тем правилам, которые там описаны в этом, в этом конкурсе. Вот. И я вам могу показать. Вот смотрите, видите, довольно приличные призовые фонды - 100.000 долларов. Вот. То есть вообще-то так довольно-таки много, я бы сказал. Даже если на пять человек их разделить, то это всё впечатляет.

Значит, то есть, что можно сказать? Что сайт Kaggle создал дополнительную мотивацию у специалистов со всего мира для решения задач в области искусственного интеллекта. То есть вы не просто свои силы пробуете или там сдаете какой-то экзамен, как вот у нас. Вот сейчас вот я вас как мотивирую? Я вас мотивирую следующим образом: вы получите самоэкзамен в случае, если разработаете приложение, мы утвердим его тему, вы сделаете описание по шаблону, который здесь вот есть, и разместите это описание в РИНЦ, ResearchGate, в РИНЦ и в Эйдос-облаке. А в Эйдос-облаке еще исходные данные разместите. Ну это со мной делается. Вот. И тогда вы получаете самоэкзамен, ребята. Но чтобы получить самоэкзамен, нужно хорошо работать на лекциях и лабораторных работах, и практических занятиях. А что значит хорошо работать? Это значит вот это задание и выполнять. Потому что я вас остальному уже научил. Мы уже, я уже показал вам, как выполняются исследовательские работы на основе табличных исходных данных, внешних текстовых исходных данных, даже табличные показал реальную работу - методику риэлторской оценки. Да? Ребят, помните методику риэлторской оценки или не помните? Скажите мне. Оценку автомобилей с пробегом. Помните, да? И даже показал... Помните, да? Да. Хорошо, молодцы. Дело в том, что слышно очень неразборчиво. Я просто не знаю почему, но вот крайне неразборчиво. То есть шум слышно, а вот что вы говорите, не очень понятно. Вот. И я вам даже показал работу по обработке графических изображений, а именно спектральный анализ. Было дело, да? Показывал? Вроде было. Было. Ну, слава богу. Я уже испугался, думал, что может я перепутал что-то. Ну, в общем, вот это всё мы прошли. То есть мы прошли второй пункт, третий пункт - это просто выход на мою страничку на ResearchGate. Вот, и ознакомление с этой страничкой. То есть здесь вы можете почитать мои работы. Тут довольно много - 582 работы. Вот. И, э-э, я очень вам советую, прямо настоятельно рекомендую, не откладывая, ребята, не откладывая, зарегистрироваться в ResearchGate. Для этого вам потребуется корпоративный адрес электронной почты типа edu.kubsau.ru, на домене edu.kubsau.ru. Это гугловская почта. У вас его нет. Значит, староста должен идти к куратору, кто там у вас руководитель этой... Ну да, куратору группы. Вот. И этот вопрос ставить. Или вы можете пойти в ЦИТ и спросить, как вот мне получить такой адрес электронной почты корпоративный. У всех преподавателей есть. И в других вузах, ребята, вот берем другой вуз - это Кубанский государственный университет. Я там тоже веду занятия. И там вот есть студентка Лада Муравьева, видите? А вот у неё есть электронная почта edu.kubsou.ru. КубГУ, Кубанский госуниверситет. Видите, да? А начинается с буковки S. А у меня есть такая почта, только она начинается с буковки П. Вот. Вот у меня почта, видите, начинается с буковки П. Это преподавательская почта. То есть в КГУ, который у нас Ставропольский, у всех студентов и у всех преподавателей есть корпоративная электронная почта. А в нашем университете аграрном даже не у всех преподавателей она есть. Вот. Надо добиваться, чтобы тоже вам дали. Значит, я знаю, что если пойти туда и попросить, то они могут просто подтвердить, что вы являетесь студентом университета этому ResearchGate. Просто они шлют официальное сообщение, что да, вы действительно являетесь. И происходит регистрация. То есть этого достаточно. Вот. То есть я не раз уже в этом убеждался, что они могут там зарегистрировать любого преподавателя, студента. Потому что наши преподаватели тоже не могли там зарегистрироваться, тоже у них почты нету корпоративной, я имею в виду. Сейчас-то уже у всех есть. А когда это было, не было. Вот. Добились того, что присвоили. То есть вы тоже должны этого добиваться. Значит, я здесь голоса никакого не имею, потому что я просто преподаватель. Я даже никакой не руководитель, никакого уровня, понимаете? То есть у меня вообще никакого, никакой административной должности нет. То есть чтобы кто-то мне подчинялся в чем-то, такого вообще нет. Лаборантки подчиняются заведующему кафедрой. Вот. Если я что-то их прошу, то это просто они мне делают услугу, навстречу идут. Значит, вот здесь есть мои работы, ребята, 582, можете на них посмотреть. Вот. А вот здесь вот есть лаборатория. Видите, вот корпус лаборатории, видно? Ну, это, конечно, не весь, не вся лаборатория, только главный корпус. А вот здесь вот видите, написано название этой лаборатории. И вы, значит, можете нажать на это название этой лаборатории, и начинает, появляется доступ ко всем моим работам, размещенным в ResearchGate. Это что за работы? Это статьи, опубликованные в различных изданиях, монографии, у меня их 38, учебные пособия, у меня их 27, патенты, у меня их 30, а всего вместе 635 примерно публикаций, из них 280 там, 290 сейчас уже в изданиях, входящих в перечень ВАК. И некоторые размещены просто как препринты, отчеты о научной работе здесь тоже есть. Многие в соавторстве, но многие без соавторства. Разные. Значит, вам это надо обязательно сделать, ребят, зарегистрироваться в ResearchGate. Теперь следующее. Вам надо обязательно зарегистрироваться в РИНЦе. Для того чтобы зарегистрироваться в РИНЦе, не требуется никаких корпоративных адресов, ничего. Просто регистрируйтесь в РИНЦе и в системе Science Index обязательно. Почему? Потому что если вы в этой системе зарегистрируетесь, в системе Science Index, то вы тогда получите SPIN-код. Science Personal Index, э-э, это, ну, по-русски говорят SPIN-код. Вот. Так вот, значит, вы его получите. И после этого, кстати, я вам показывал, что вот кибернетика - 5.000 ученых в России в этой области пишут статьи, книжки, и среди них у меня первый рейтинг. Вот. И вот у меня есть SPIN-код, видите? То есть когда-то давно я его получил. Получение SPIN-кода дает право на заключение договора на размещение непериодических изданий. Нажимаем автором на титульной странице, ищем здесь разделчик размещение непериодических изданий. Имеется в виду вот этих вот отчетов научных, препринтов, а также монографий, пособий. И попадаете вы в систему заключения договоров отсюда. Но я вам скажу, что они, ребята, с юмором эти вот специалисты этого РИНЦа. Они закопали так это вот, настолько глубоко там закопано это всё, что, в общем-то, найти практически невозможно. То есть если кто-то не подскажет, не покажет, как это делается, то и не найдете, понимаете? Я вам подсказываю, показываю, то есть вы прямо вот дорожку вам выстилаю, так сказать, чтобы вы шли по этой дорожке, не было у вас проблем. Вот здесь заключаете договор на физическое лицо обязательно, не на юридическое, на физическое. После этого вы можете размещать в РИНЦе публикации, которые сделали в ResearchGate, если вы делали с присвоением DOI. А DOI - это Digital Object Identification - это цифровой идентификатор статьи, объекта, который сохраняется у неё при всех её размещениях в любых библиографических базах данных мира, включая РИНЦ, Российский индекс научного цитирования. И вот, э-э, это считается достаточно престижным. И обычно стоит денег, где-то около 200 руб. за страницу, но бывает дешевле иногда. Вот, тогда возникают сомнения. Вот здесь я сделал небольшую инструкцию, как это всё делается. Вот ResearchGate.doc. Её можно скачать и просто это. А здесь ссылочки на все эти вот есть: на ResearchGate, на eLibrary, на заключение договоров в РИНЦе. Все ссылки я дал вам. То есть прямо начинаете регистрацию. Почему? Потому что это требует времени. Потому что чтобы в РИНЦ получить право заключать договор, проходит недели две, а то и три. А еще заключение договора тоже тратится время на это. То есть вы можете просто не успеть. Если вы не успеваете это сделать, но сделали само приложение и описание, то тогда я могу его разместить. Но я могу разместить только в том случае, если я буду соавтором. У меня есть моральное право быть соавтором, потому что я создал эту технологию, научил вас этой технологии и помогаю вам это всё сделать, и само приложение, и описание этого приложения. То есть я всю эту инфраструктуру создал. Но, значит, я на это не претендую, и у меня нет никакого особого желания быть соавтором ваших статей, потому что у меня эти статьи, я не знаю, куда их девать уже, понимаете, там. Ну, уже 640, скоро... У меня там летом вышло 13 статей, надо вам сказать, летом, понимаете? Ну куда их? Ну, ну что это такое? То есть это уже такой явный перебор. Но я пишу те статьи, которые от меня требуют. Доктор, профессор, член совета диссертационного. От меня требуют, я делаю то, что от меня требуют. Получается очень большое количество различных работ: методических, и научных, и таких вот статей, и выступлений на конференциях. Очень большой объем работ у профессора. У других нету, тех, кто имеет другие звания пониже, у них таких нагрузок нет. Это именно нагрузка профессора. Вот. Значит, таким образом вы получаете самоэкзамен. Если бы где-то чего-то там не удается вам сделать, в РИНЦ, например, не можете, и я там у вас не соавтор, тогда у вас появляется такая такой вариант, что, может быть, хорошо вы получите. Если нигде вы не размещаете, ну, в ResearchGate тоже я могу разместить. Но только это может сделать автор, понимаете? Ну и соавтор. Вот. То тогда, значит, соответственно, удовлетворительно. И я категорически не люблю ставить двоек, но получается, что если никто ничего не, кто-то ничего не делал, ну, вообще ничего, понимаете? То есть не выполнил никакого объема работ не выполнил по заданию, а это задание и по всей дисциплине, по итогам работы, освоения вообще дисциплины, понимаете? То есть здесь, если вы это задание делаете, то значит, вы подтверждаете, что вы получили необходимые знания, умения, навыки и получили соответствующие необходимые компетенции. И эти компетенции касаются не только того, чтобы проводить само исследование, вот, и его описывать правильно, методически, но и разместить его в библиографических базах данных. И заработать себе на стипендию и на портфолио. То есть это я всё думаю в ваших интересах, самым непосредственным образом связано с вашими интересами. И к тому же это очень ценно. Ценно уметь бесплатно писать публикации когда угодно. То есть вы в системе Эйдос сможете эти публикации каждый день делать, грубо говоря. Вот, ну каждый день не получится, каждую неделю. Вот. Взяли задачку, скачали данные, взяли, сделали задачку, описали по шаблону, разместили там, разместили там. Всё, понимаете? У вас за год, может быть, десятки публикаций сделать можно, если есть желание. Значит, теперь смотрим дальше на сайт Kaggle. Они э-э, заинтересовывают тем, что есть призовые фонды, как вы видите. Вот. Но, правда, туда попасть трудно в призовой фонд. Там есть такие монстры, которые, э-э, их называют гроссмейстеры, мастера, которые прямо вот этим занимаются. То есть если там я, я сейчас о себе скажу. Я взял и там зарегистрировался на Кагле. Вот. И можно мой профиль посмотреть. Значит, я принял участие в двух соревнованиях. В одном я получил результат, который лучше, чем 85% других участников, а в другом соревновании получил результат, который лучше, чем 91% других участников. Вот. Но это не призовое место, ребята. Это вот если взять всех участников, разделить там на 10 категорий по рейтингу, то я попадаю там, скажем, в первую, вторую десятку, понимаете? Вот. Не десятку этих вот пользователей, а десятку градацию по рейтингу десятую. Верхнюю, первую или вторую. Ну, поскольку я участие принял только в двух соревнованиях, то, значит, статус новичок. То есть я, ну, новичок, вы знаете, это такое средство, которое вот если его один раз нюхнешь, то всё. Вот его там нюхнули эти скакуны, и готово, да? Вот. Ну это я так подшучиваю. Так вот, теперь смотрите, ребята. Значит, есть, ну я их не знаю, как даже сказать. Ну, действительно мастера, понимаете, без всяких вопросов. Вот мастера и гроссмейстера, даже великие мастера есть. Гроссмейстер - это великий мастер. Действительно есть великие мастера. Вот здесь вот эта линия здесь вокруг, она показывает уровень достижений этого человека. Вот смотрите, девушка из Индии, из Мумбаи, Махараштра, Индия. Гроссмейстер, понимаете? Гроссмейстер. Высший ранг 71, текущее место 72 из 166.000 участников. Ранг восьмой из 143.000 участников. Ну то есть это вообще-то круто очень. Вот она приняла участие в восьми соревнованиях эта девушка. И 13% сама, 88% в команде. Вот. И у неё есть исследования по Титанику, по ценам на жилье. Вот. И э-э, первых мест нету у неё, по-моему. Я просто видел гроссмейстеров, у которых, допустим, два места, два раза она заняла первое место в чемпионате мира по искусственному интеллекту. В одной из тем, правда. Вот. Соревнования. Некоторых она в соревнованиях принимала участие без рейтинга. То есть это песочница называется. То есть, ну, просто пробуешь свои силы, смотришь, что получится. Вот. Ну вот какой-то тут у нас есть из Риги, из Латвии. Гроссмейстер. Тоже первое место 3 года назад занял, третье место, восьмое место, девятое место. В соревнованиях. Понимаете? А если там есть, э-э, вот 116 соревнований. То есть принял участие в 116 соревнованиях. То есть это люди очень опытные. У них есть и знания, и есть инструментарий, позволяющий решать сложные задачи. Значит, да, кстати, бывают довольно много русских. Смотрите, Юрий Матвиенко, да? Вот. Потом что тут у нас? Афанасьев Дмитрий, Сергей Силаев, Татьяна и Андрей, Артём Меренов. Вот, это я русских ищу. Вот. Ну, китайцы там, индусы, то есть, в общем, народ со всего света. Вот. Ну интересно, что вот Юрий, например, этот, он, видите, э-э, высокие, высокий рейтинг, эксперт. А, он живет White Plains, New York, США. То есть он вообще не наш товарищ уже. Ну вот так вот. Ну, видимо, вот тут он там... Да, кстати, тем, кто занимает призовые места, предлагают работу. Не так давно, где-то года два назад, сайт Kaggle, который был раньше создан одноименной фирмой, в общем, то есть это фирма занималась этим самостоятельно. Значит, ну, которая просто частная фирма. Её купила, купил Google эту. То есть Kaggle, то есть Google купил Kaggle. Поняли, да, ребята? Теперь, где еще есть наборы данных? Смотрите. Значит, вот здесь вот делаем ссылочку. Кликаем по ссылочке. Здесь мои методические работы кое-какие есть, изданные, не совсем изданные. А вот здесь вот есть шаблонное описание, источники данных и 50 наилучших баз для наборов данных для машинного обучения. А также, смотрите, ссылочки на федеральную статистику, на криминальную статистику федеральную. Значит, что это такое вообще? Значит, вы знаете, что система Эйдос позволяет выявлять зависимость между факторами и результатом их влияния. Если вы возьмете, допустим, федеральную статистику, то вы можете в качестве факторов рассматривать, скажем, развитие тяжелой промышленности, а в качестве результата влияния этих факторов можете рассматривать экологию. Вот как вы считаете, промышленность тяжелая улучшает экологию или ухудшает? Вот вы думаете... Ухудшает. Ну, очевидно, но как она её ухудшает, это не очень понятно. Так вот, ну, в смысле, в деталях. Нет модели, которая детально отражает, как она ухудшает экологию. Значит, выходим на мой сайт. На вот эту страничку, где система Эйдос. И ищем, как вы думаете, что мы ищем? Экология ищем. И мы находим 10 публикаций по экологии. Видите, да? И вот мы можем их все посмотреть. Так вот экология, ребята, она влияет, ну там чуть поменьше публикаций, потому что некоторые слова встречаются два раза, как я сейчас заметил. Вот. Сейчас вот. Ага. Эколого-социально-экономические системы, а исследование влияния экологических... То есть подзаголовок тоже включает в себя слово экология. И есть статьи новые. Вот, допустим, вот эта статья, она достаточно новая. Она написана в каком году? В 2017 году. Это тоже 2017. Вот. Но это не самое лучшее, если честно. Вот. А вот сейчас тут я вам большую статью найду. Вот. Вот если большая статья, то это вот четыре, видите, четыре условных листа - это большая статья. Вот. Сейчас мы я вам ссылочку на неё пошлю. Здесь идея в чем заключается? В том, что мы можем посмотреть по регионам России, областям, краям, округам, можем посмотреть, как там развита тяжелая промышленность и какое там экологическое состояние окружающей среды. И можем сопоставить это одно с другим. А дальше что мы можем сделать? А дальше мы можем взять и узнать, как экология влияет на качество жизни. И она влияет, ребята, чётко совершенно. То есть если у нас экология повреждена, то и качество жизни, продолжительность жизни, причины смерти тоже там совершенно явно от этого зависят. Вот. Есть места в России, где отвратительнейшая экология. Отвратительная. Вот Магнитогорск, Норильск, Никель там. Вот. Просто ужасно, что там происходит. Вот. Ну это не совсем та статья, которую я хотел показать. Сейчас я попробую найти ту, которую я хотел показать. Вот она тоже. Вот эта со Стрельниковым, профессор Стрельников, он зав кафедрой экологии Кубанского госуниверситета. Вот. Это вот моя статья, которая вышла раньше, чем статья с профессором Стрельниковым. И вот тут я как раз использовал статистические данные с Росстата. А там данные обычно по регионам России. И вот можно считать, что регион России - это наблюдение. И по этому региону России можно вытащить информацию о том, какое там развитие тяжелой промышленности и какая там экология, и сопоставить это одно с другим. Вот.

\*\*10.8. Рекомендации по выбору данных для учебных целей\*\*

То есть для ваших целей вполне достаточно исследовать один год. Вам нет необходимости там за много лет данные эти обрабатывать. Для целей нашего обучения. То есть взяли данные, вот, посмотрели эти данные, вот. Что это за данные? Вот. Взяли эти данные, ввели в систему Эйдос. Как это можно сделать? Значит, удаляем всё в папочке indata, куда файлик этот переносим, меняем его название. У системы Эйдос предопределенное, ну, как сказать это, предназначенное, ну, предопределенное имя файла исходных данных и место, где он находится - input data. Но расширения могут быть xlsx, xls, и CSV, и DBF расширения могут быть. Теперь мы, значит, что делаем? Открываем Excel. И в Экселе вводим этот файл. Для чего? Для того, чтобы посмотреть на его, ну, можно даже так вот сделать, взять в любом редакторе открыть. Поля вот эти нам нужно понять, что они означают. Это поля. И через запятую вот здесь вот у нас значения находятся. Значит, проще всего вот так сделать: взять переводчик и с английского на русский перевести. Общий рейтинг, страна или регион, оценка ВВП на душу населения, социальная поддержка, ожидаемая продолжительность здоровой жизни, свобода делать жизненный выбор, щедрость, восприятие коррупции. Вот. Значит, нам нужно первую колоночку взять - общий рейтинг, да? Правильно? То есть общий рейтинг - это у нас будет классификационная шкала. Всё остальное - это будут у нас факторы, которые влияют на общий рейтинг. Вот. То есть первая колоночка - это у нас классификационная шкала, остальные описательные шкалы. Значит, мы что делаем? Берём, указываем это, значит, ну тут нам сейчас этот файл не совсем такой, как нам надо, поэтому будем, значит, надо хотя бы знать, что их девять. Вот. Значит, будем указывать со второй, а здесь с третьей по девятую. И посмотрим. У нас сейчас этот файл преобразовался, он маленький. И сразу пошло процесс ввода его в систему. Вот. Сразу что-то получилось не то. Но нам сейчас в данном случае это не так важно, надо будет посмотреть, что там такое. Значит, иногда бывают нестандартные. Значит, создан вот на основе ввода информации, которую мы сейчас вводили в этом режиме, создан такой файл DBF. Здесь вот у нас нарушено. Страна оказалась у нас в числовой колонке. Ну это бывает, знаете когда? Когда там в файле CSV-файле, э-э, в самих этих полях есть некие тексты, которые с запятой. То есть бывает иногда такое вообще там... То есть он вроде как CSV, но с нарушением стандарта. Вот. Значит, я сейчас ввожу с помощью Экселя этот файл. Указываю, что у него разделитель запятая. Готово, готово. Вот у нас этот файлик есть. Теперь нам надо вот эту колоночку страна скопировать в самую левую колоночку. Вот. То есть мы добавляем здесь колоночку, вставить. Почему? Потому что в системе Эйдос первая колонка - это информация о том, откуда взята информация. Ранг у нас будет... А, здесь это порядковый номер просто. Вот, тогда нам это вообще он не нужен. А что тогда мы будем смотреть? Будем, какая страна... Ну не знаю, вот этот, вот эта колонка скоре, скоре является оценкой уровня качества жизни. То есть берём вот эту колонку скоре тоже в качестве классификационной шкалы. И записываем её в папочку уже для ввода в систему с именем input data xlsx. И смотрим, ребята, смотрите, сколько у нас колоночек? Значит, в клеточку А1 ставим, нажимаем Shift кнопочку и двигаем вправо. Вторая, третья у нас это классификационные шкалы, а потом до девятой описательные шкалы. Со второй по третью классификационные, с четвертой по девятую описательные. И мы эту информацию вводим в систему. Ну да, здесь у нас надо прервать это. Какая-то была ерунда, что-то не то получилось. И попробуем ввести сейчас из этого экселевского файла. То есть нам нужно немножко проанализировать этот CSV-файл, чтобы нормально его ввести, заголовки проанализировать и понять, что они означают. Вот. Значит, мы вводим из этого нового Экселя со второй по третью и с четвертой по девятую. Вот. Вводим её. Вот. И смотрим, какие у нас получились классификационные шкалы и описательные. Смотрите, у нас шкала, которая должна быть числовой, она у нас, видите, вроде как там числа, но они там все уникальные значения, этих чисел там у нас прямо приведены. И то же самое, по-видимому, будет и с этими описательными шкалами. Почему это, ребята? Почему так вышло? Потому что в CSV-файлах запятые - это разделители между полями. А сами, сами числа - десятичная точка используется. А в Экселе наоборот. Значит, поля там и так по-другому там сделаны. А, значит, числа у них десятичная запятая. Поэтому мы заменяем точку на запятую, и у нас все вот эти вот значения становятся числовыми. Теперь мы еще раз это вводим, но уже теперь там будут числа. Параметры те же самые все. Ну можно сделать разные интервалы, тогда будет достоверность выше обычно. Вот. Можно посмотреть, сколько у нас значений, наблюдений есть на интервальное значение. Мы видим там 31, 50 там. То есть вполне достаточно наблюдений. Можно разные делать эти вот число интервалов здесь, но можно вот взять, оставить то, что сейчас. И посмотреть, что получилось. Смотрите. Вот у нас страны так они и есть страны, как были, да? А еще у нас появился рейтинг страны. Там пять значений рейтинга: очень низкий, низкий, средний, высокий, очень высокий. По оценке индекса счастья. Вот. И описательные шкалы, они у нас все числовые и по три числовых диапазона. Теперь создаем модель и проверяем её на достоверность. Эта модель составляет 5% от того, что может система Эйдос обработать. Значит, вот сейчас эти модели создаются и проверяются на достоверность. Они уже созданы, все модели. Они создаются на графическом процессоре быстро. И распознавание тоже происходит на графическом процессоре. А вот формирование форм выходных, 10 штук, они это происходит на центральном процессоре. У нас там есть 156 наблюдений, то есть разные страны, я так понимаю. Вот. Смотрим, что у нас получается. Ну сейчас придется подождать немножко, полторы минуты. И мы посмотрим, получим информацию о достоверности модели. Если достоверность модели хорошая, модель разумная, даже не очень хорошая достоверность, но модель разумная, видно, что она отражает зависимости между факторами и результатами, то тогда можно это взять за соответствующую тему выбрать и, пожалуйста, можно её сделать, разработать приложение под это, вот то, что сейчас я делаю, и описать по шаблону и разместить. И у вас сразу появляется много зайцев, так сказать, вы убиваете. Вот. То есть вы и получаете самоэкзамен пятерку, и получается у вас стипендия, для стипендии это играет роль, и для портфолио тоже это играет роль. Кроме того, вы просто получаете нужные знания, умения, навыки, которые вам пригодятся. То есть вы сможете проводить исследования с использованием этой технологии, описывать его это описывать это исследование методически грамотно и публиковать бесплатно с присвоением DOI в РИНЦ. То есть всё это очень полезные для вас навыки, знания, умения, навыки. То есть это для вас довольно ценно. Вот я хочу вас всему этому и научить. По сути дела, я вот это и делаю, показываю вам, как это всё делается. Ну сейчас придется немножко подождать. О, уже 2 минуты она считает. Ну тогда получается 3 минуты уйдет на это. Ну если модель большая, там много наблюдений, там десятки тысяч, ну тогда уйдет час, два часа, понимаете? То есть это уже более такие тяжелые задачки для обработки. Теперь, кластерный анализ реально работает, ну где-то там до 150 классов, 150 значений факторов. То есть если их больше, тогда, возможно, не удастся кластерный анализ провести. Почему? Может долго просто быть это, трудоемко. Вот, уже у нас 2 минуты. Да, уже получается 5 минут прогноз. 16:05. Ну еще 2 минуты и должно закончиться. Значит, дело в том, что прогноз постоянно уточняется с учетом всех текущих операций. И сначала там он основывался на той части расчета, где модели создавались. А теперь учитывается также, что формы создаются и проводится распознавание в этих моделях. Поэтому время уточняется, но оно постепенно стабилизируется, этот прогноз стабилизируется и становится точным.

\*\*10.9. Выбор данных: Резюме\*\*

Значит, теперь, когда вы смотрите в Кагле данные, рассматриваете, брать, не брать эти данные для исследования, то вы чем должны руководствоваться? Тем, чтобы эти данные были небольшие по объему и составляли два файла основных: один файл - обучающей выборки, а другой файл - тестовой выборки. Вот. То есть вообще-то так довольно-таки... Ну, это я вам могу сказать, что не исключено, что некоторые там виды пестицидов, которые применяются для того, чтобы травить там жучков каких-нибудь, что они влияют на людей так, что у них там крыша едет, они начинают что-то там вытворять, понимаете? Нет, кроме крышу, так, то есть действительно реально химия может влиять на состояние здоровья, на заболеваемость. Мы это знаем прекрасно, что влияет она. Вот. А и на психические процессы тоже, в принципе, почему бы и нет? Тоже может влиять. Вот вы можете посмотреть, что здесь у нас по играм, видите, Dota, Tanks там, World of Tanks, Warface, Warcraft там был где-то. Вот. То есть вы можете выбрать, если вас какая-то игра затянула там в свое время или интересовались какими-то, приобрели какой-то опыт, достижения. Ну, пожалуйста, можете проанализировать там сценарии там или там достижения. World of Warcraft, вот игра есть. Вот тут Kaggle есть, видите? То есть много тут есть работ, выполненных по данным Kaggle. И по данным UCI тоже много есть работ. То есть то, что я вам рекомендую, студенты это и делали. И получились хорошие работы, интересные у них. То есть мы рассмотрели подробно вопрос о поиске исходных данных, наборов данных для создания интеллектуального приложения.

**11. Заключение и вопросы**

Так, ребят, теперь у нас что у нас по времени? У нас занятие закончится через одну минуту. Поэтому вы сейчас давайте задавайте мне вопросы, какие у вас возникли, пока не поздно, пока занятие не закончилось.

То есть порядок такой: вы ищете там данные, тему, в общем, ищете тему, которая вас может заинтересовать, смотрите, какие там данные, насколько они там удобоваримы, насколько мы можем их ввести, обработать, чтобы там не 5 ГБ было, а там, ну, в пределах там мегабайтов там, двух-трех там максимум. Вот. И чтобы разумные пределы были количество примеров, размерность там, число этих атрибутов, вот такое, чтобы всё это было в разумных пределах. Пробуйте это всё скачать и сделать модель по вот этой схеме, которой я вас учил. Вот. Определить, какая получается достоверность. Не получается - присылаете мне ссылочку даже просто на источник данных и спрашиваете: как вы считаете, можно ли эту тему взять? Я сам это всё делаю, то, что вам сейчас рассказывал: скачиваю, делаю модель, смотрю достоверность и говорю: "О, это хорошая тема. Давайте". Вот. И утверждаю эту тему. После этого вы начинаете её описывать. И потом после описания даете мне этот текст, я на него смотрю, делаю замечания и так далее. Значит, я пишу, что мы рассмотрели четыре пункта инструкции и начинаем рассматривать пятый пункт инструкции.

Всё, ребята, значит, конец занятия. Вопросы писать, значит, не стесняйтесь. Находите тему, пишите. Регистрацию начинаете прямо сейчас. Вот когда будет у вас возможность, прямо начинаете регистрироваться. Получаете SPIN-код, начинаете заключать договор. Ну, в общем, всё тут надо делать. Прямо начинать исполнять. Ну это не совсем лекция, это практическое занятие было. То есть я вам подробно рассказывал, как выполнить задание. Там в этой инструкции всё это есть со ссылочками, всё там написано.
Всего самого хорошего, ребята. До свидания.
(Студенты: До свидания.)
До свидания. Вопросы задавать не стесняйтесь, пишите на почту.
До свидания.
(Студент: До свидания.)