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

***Российская Федерация***

***Ватамов Д.Н. ПИ2104***

**154 Теория информации, данные и знания. Практическая 4. Инструкция по разработке собственного приложения 2020-10-22**

## Заголовок

Разработка собственных интеллектуальных приложений в системе Aidos: Руководство и оценка

## Резюме текста

Практическое занятие №4 для группы ИТ-2002 по дисциплине "Теория информации, данные, знания" проводит профессор Луценко Е.В. 2 октября 2020 года. Тема занятия – инструкция по разработке собственных интеллектуальных облачных приложений с использованием системы Aidos.

**Основная цель для студентов:** Разработать собственное приложение, описать его и разместить на FTP-сервере системы Aidos, а также опубликовать описание в ResearchGate и РИНЦ. Успешное выполнение этого задания гарантирует оценку "отлично" по дисциплине без сдачи экзамена (самоэкзамен).

**Процесс разработки:**

1. **Выбор темы:** Тема может быть любой, но не должна противоречить морально-этическим нормам и законодательству РФ, а также не должна полностью повторять уже существующие в системе Aidos приложения (допускается развитие старых тем на новом уровне). Возможна работа в соавторстве (2 студента).
2. **Поиск исходных данных:** Данные можно найти на специализированных ресурсах (UCI Machine Learning Repository, Kaggle), сайтах статистики (Росстат и др.), или использовать примеры с сайта профессора. Рекомендуется использовать данные в формате CSV или Excel. Следует выбирать задачи с умеренным объемом данных для начала.
3. **Создание модели:** Используя систему Aidos, импортировать данные, настроить параметры анализа (классификационные и описательные шкалы), синтезировать и верифицировать модель. Система поддерживает различные форматы данных (CSV, Excel, DBF, текстовые, изображения).
4. **Описание и публикация:** Подготовить описание работы по предложенному шаблону и разместить приложение на FTP-сервере Aidos, а описание – в ResearchGate и РИНЦ (eLibrary).

**Источники данных и рекомендации:**

* **UCI:** Большой репозиторий, но данные могут быть в специфических форматах и без разделения на обучающую/тестовую выборки.
* **Kaggle:** Современная платформа с конкурсами, часто предоставляет данные в формате CSV, разделенные на обучающую (train) и тестовую (test) выборки, что удобно для оценки модели.
* **Сайт профессора:** Содержит примеры ранее выполненных работ студентов, которые можно использовать как образец или для выбора/развития темы.
* **Рекомендации:** Начинать с не очень больших наборов данных (не гигабайты), предпочтительно в формате CSV. Проверять достоверность получаемой модели.

**Поддержка:** При возникновении вопросов можно обращаться к профессору по электронной почте, указав группу, ФИО и суть проблемы, приложив исходные данные.

**Административные моменты:** Отмечается избыточность систем учета посещаемости и некоторые технические сложности с MS Teams.

## Детальная расшифровка текста

### 1. Введение и детали сессии

Здравствуйте, ребята.
Здравствуйте. Здравствуйте. Здравствуйте. Здравствуйте. Здравствуйте.
Сейчас у нас 2 октября 2020 года.
Пятая пара.
15:35-17:05.
Практическое занятие номер четыре
с группой ИТ-2002
по дисциплине Теория информации, данные, знания.
Занятие проводит профессор Луценко Евгений Вениаминович.

### 2. Основная тема: Разработка пользовательских приложений

Тема сегодняшнего занятия у нас -
это инструкция по разработке собственных приложений,
собственного, собственного приложения.
И я эту инструкцию подробно вам комментирую, рассказываю по ней,
как
создать собственное приложение.

*(Пауза, технические действия)*

Сейчас вас отмечу и потом продолжу.
Потому что я в нескольких системах отмечаю, что это занятие проводится.
Себя отмечаю, вас отмечаю.
Потом в третьей системе веду занятие.
А в этой третьей системе, в Вебексе,
есть информация о том, кто на занятии. А в другой системе я отмечаю, кто на занятии.
Хотя эту информацию можно взять из Вебекса.
А в третьей системе я отмечаю, что я веду занятие, хотя в Вебексе есть об этом информация, что я сейчас веду занятие в своей
комнате, в своей группе.
То есть получается, что всё, что я там вношу,
вся информация обо всём этом есть, но надо её вносить вручную.
Что, вообще-то,
очень странно,
на мой взгляд, как программиста
с
очень длительным стажем.
Я с семьдесят шестого года работаю на компьютерах,
то есть ещё начиная с дипломной работы.
То есть это
там
24 года и здесь
20 лет. То есть я 44 года имею стаж работы на компьютерах.
И что-то в этом соображаю немножко.
И работал всё время как разработчик и как руководитель
подразделений.
И как преподаватель, как учёный.

*(Пауза, технические действия с MS Teams)*
Так, что-то здесь непонятно, что такое?
Почему какая-то возникает проблема?

Так, ребята. Значит, теперь напомните, пожалуйста, мы уже с вами изучали эту тему, связанную с разработкой собственных приложений?
Или, может быть, я на чём-то там остановился, и надо продолжить? Инструкция, значит, группа 2002, смотрим.
Нет, вроде ещё не изучали.
Не изучали. Да, у нас был сценарный АСК-анализ предыдущее занятие.
Был сценарный АСК-анализ.
Поэтому сейчас, значит, начнём эту тему.

### 3. Мотивация и система оценивания

Значит, вы сейчас должны... сейчас я поделюсь изображением.
Увидели, да?
Вот. Значит, выходим на мой сайт.
Второй пункт.
Картинка слева, титульная видеограмма досовской версии системы Aidos.
И вот здесь внизу есть ссылочка: Задание, инструкция для обучающихся по разработке собственного интеллектуального
облачного Aidos-приложения.
Шлю ссылочку вам
в чат.
Начинаю с самого главного.
Самым главным является мотивация, ребята.
Как вас заинтересовать в том, чтобы вы это сделали? Я попытаюсь это сделать разными способами, но у преподавателя всегда есть в запасе один такой, немножко, не знаю, насколько сказать, ну, немножко, может быть, подленький такой способ, но надёжный.
Как вас мотивировать.
Вы видите табличку в пункте девять, да, ребята? Все ж видят?
Значит, смотрите, что здесь написано. Что вы должны
по результатам изучения дисциплины...
Некоторые преподаватели, я знаю, на каждом лабораторном занятии или на каждом практическом занятии дают задания, вы их решаете, потом они их собирают, проверяют, как вы их решили и так далее, и так далее. У нас же здесь немножко по-другому поставлено. Значит, я вам даю теоретический материал на лекциях, на лабораторных мы приобретаем умения и навыки
по их применению. На практических я даю вам консультации по выполнению практических работ, то есть таких, э, которые вы будете сами выполнять
в той технологии,
которую мы изучаем.
И вы разрабатываете в результате собственное приложение. Я называю это интеллектуальное Aidos-приложение. Размещаем его на FTP-сервере системы Aidos, её же собственными средствами. Эти средства входят в состав системы.
Размещаем это же приложение в ResearchGate и РИНЦ. Но не само это приложение, а все... если в системе Aidos размещается именно приложение с описанием в облаке, то в ResearchGate и РИНЦ только одно описание размещается.
И если всё это осуществляется, ребят, то вы получаете самоэкзамен отлично.
Что означает самоэкзамен? Это означает, что вы билет не тянете.
Понятно, да? То есть нет такой процедуры вот, чтобы вы прямо вот брали билет, отвечали на вопросы. Почему? Потому что всё вот это вместе взятое подтверждает, что вы уже приобрели знания, умения и навыки, которые, те компетенции, которые предполагалось вам дать на этой дисциплине.

### 4. Руководство по разработке приложений

#### 4.1. Выбор темы и ограничения

Теперь смотрим на вопрос, как выбрать тему приложения и как дальше, где найти исходные данные для приложения. Это вопрос, который рассматривается в инструкции, четвёртый пункт.
Четвёртый пункт инструкции как раз этот вопрос раскрывает.
Что мы здесь видим?
Ребят, вы эту инструкцию видите, нет?
На экране было видно?
Да, было. Видно, видно, было видно.
Видно, видно, да? Ну хорошо. Ссылочку я вам послал, да? В чат. Послал.
Значит, как выбирать тему приложения?
Значит, тема может быть абсолютно любой, но есть три ограничения.
Первое ограничение: тема не должна противоречить общепринятым в России морально-этическим нормам и действующему законодательству Российской Федерации.
Вот. Второе.
Тема не должна повторяться с уже имеющимися. Вот есть ссылочка здесь в четвёртом пункте на уже ранее созданные облачные Aidos-приложения, уже размещённые в облаке. Вы можете с ними ознакомиться, посмотреть на темы этих приложений. Для чего? Для того, чтобы у вас возникло понимание того, какие могут быть темы.
Вот. То есть здесь многие приложения сделаны мной, а некоторые сделаны студентами, и некоторые сделаны соавторами.
Вот. Значит, здесь вот справа есть есть описание этих приложений.
Значит, мы видим, какая, какой номер приложения. Вот, допустим, вас интересует Rocket Lego, да, например, 196-е приложение. Это игра, да? Вот мы берём вправо двигаемся, ищем это описание этого приложения и видим, что написали студенты.
Автоматизированная система когнитивного анализа матчей в игре Rocket, да? Ракеты какие-то. И вот здесь вот у нас оглавление есть, и потом идёт краткое описание метода, потом описание исходных данных.
Вот, откуда эти исходные данные взяты.
Вот. Что мы вообще, в чём смысл задачи? Что мы в этой задаче рассматриваем как причины, а что как следствия, как факторы, а что как результаты влияния этих факторов. И вот мы это фрагмент таблицы исходных данных. А может и все даже. Как-то здесь странно разбито. И потом показано, как это всё решается в системе Aidos. Решения приводятся.
Это вот картинка показывает, что модель выявления зависимости между факторами и результатами их влияния. То есть система вывела эти зависимости. И результаты решения задачи.
И, в общем, так довольно полно представлена решение. Даже значимость факторов и степень детерминированности классов тоже здесь есть.
Значит, это опубликовано, ребята, работа. Сейчас мы её ищем, чтобы, не отходя от кассы, как говорится, чтобы показать вам, что она есть эта работа. То есть у этого студента появилась в результате публикация, в результате выполнения этой работы.

*(Поиск публикации)*

Что-то как-то...

#### 4.2. Соавторство

Допустимо ли, чтобы два студента были авторами одной работы? Допустимо, ребята. Вот, видите, Ладинский, Смирнова там, Захаров, Милега. Вот. Пяткин, Грумет, пожалуйста. Допустимо.

### 5. Источники данных: UCI и Kaggle

#### 5.1. Общие сведения и форматы данных

Где брать исходные данные для работы? Где брать исходные данные? Когда-то давно, лет, скажем так, 15-20 назад, с этим была проблема определённая. Ну, тогда можно было взять данные статистики. Сейчас тоже можно, конечно. Я вам покажу сейчас сайты статистики различных ведомств федеральных. И вы там увидите, что там полным-полно данных различных для анализа. Вот. Ну, а раньше, когда-то я делал вот такой запрос: наборы данных для машинного обучения, и ничего не получал. Сейчас можно зайти на любой из этих сайтов, вот, и попробовать там найти данные для своей работы.
Значит, система Aidos имеет интерфейсы с внешними с внешними источниками данных, которые поддерживают возможность ввода в систему текстовых данных, табличных данных экселевских и CSV файлов. CSV файл - это стандартный, то есть стандарт есть, форматированный текст. То есть досовский текст, где тоже есть колонки, так же как вот в экселевской таблице, но эти колонки разделены через запятую.
То есть вполне можно вводить все эти типы данных в систему. И можно вводить также изображения. То есть табличные данные, текстовые файлы и изображения. Табличные данные можно вводить в стандарте Excel, в стандарте dBase, DBF файлы и в стандарте CSV файла. То есть довольно большое число различных интерфейсов у системы есть, позволяющих вводить данные.

#### 5.2. Использование репозитория UCI

Ну, сейчас, значит, я вам скажу, что сейчас вот есть огромное количество сайтов, где уже есть наборы данных для машинного обучения. Но когда-то давно было только вот, по существу дела, был один сайт такой развитый. Это сайт UCI. Я не ссылочки не шлю, потому что я послал ссылку на эту инструкцию, а там эти ссылки есть. Значит, здесь огромное количество исходных данных. Такие данные исходные для машинного обучения называются датасет.
Вот. А сама страничка называется Datasets.
То есть огромное количество данных для машинного обучения, намного больше, чем реализовано в виде задач в системе Aidos. То есть вы можете здесь эти данные находить, которые вас интересуют, и пытаться, значит, можно, если вы находитесь в Google Chrome, используете, то можно перевести на русский это всё. Вот, можно попытаться решить какие-то задачи, посмотреть, как это получится.
Вот здесь вы видите описание, информация об атрибутах, признаках. Вот.

#### 5.3. Рекомендации по выбору данных (формат, размер, структура)

Значит, какие у меня рекомендации?
Рекомендация у меня использовать CSV файлы для ввода данных, потому что это удобнее, чем какие-то форматы своеобразные такие, уникальные, которые есть на UCI. Иногда там бывают и CSV, а очень часто бывают какие-то странные форматы, там непонятные.
Вот. Значит, и чтобы было два файла желательно. Один файл - это обучающая выборка или тренировочная её называют. И другой файл - это тестовый файл. Ну вот на UCI тестового файла нет, там только есть исходный файл исходных данных для формирования моделей. И проверка достоверности этих моделей осуществляется на том же самом файле исходных данных.

#### 5.4. Использование платформы Kaggle

А в Kaggle у них есть отдельный файл для теста, для того, чтобы проверить эффективность модели.
Ну здесь вот у нас есть какая-то характеристика этих задач. Год, во-первых, есть, вот 2020, и старенькие есть задачи, вон начиная с девяносто четвёртого года, с девяносто второго даже, кошмар, с девяносто пятого года.
Вот. А последние, видите, вот эти 2020 года. Ну я советую, конечно, использовать данные 2020 года или девятнадцатого. Теперь смотрим, сколько здесь у нас набор обучающей выборки, сколько у нас там примеров приведено. И сколько у нас есть признаков в задаче. Значит, если у нас, допустим, приведено 37.000 примеров, а 294 признака, то это задача большой размерности, ребята. Она может тяжеловато обрабатываться на компьютере. Но система Aidos, сейчас вот я её тестирую на обработке данных, где 4 млн примеров обучающей выборки. То есть в обучающей выборке 4 млн примеров. Ну это, ну не знаю, в сотни раз больше, чем вот здесь вот у нас на UCI приведено. Хотя вот мы видим вот 581.000 примеров, видите, вот.
Теперь смотрим, какие типы данных там, смотрите. Категориальные, целочисленные, действительные. Категориальные - это имеется в виду текстовые. То есть там текстовые шкалы, в которых значения в ячейках таблицы текстовые. Целочисленные - это не требует действительно какого-то комментария, так понятно, это числовые шкалы. Вот. Значит, если есть, значит, система Aidos позволяет обрабатывать и категориальные, целочисленные, действительные данные.
Вот. Но я советую вам брать задачи, где есть численные шкалы, числовые шкалы. Почему? Потому что там, где вот только категориальные, они могут быть очень большой размерности эти задачи. Там могут быть уникальные текстовые наименования, которых может быть большое количество, и может быть с этим, ну, проблемы при обработке, скажем так. Большой объём данных исходных.
Вот. Теперь смотрим на сайт Kaggle.
Чем интересен сайт Kaggle? Сначала когда-то давно этот сайт был, скажем так, создан некой фирмой, вот, которая просто была владельцем этого портала, на котором тоже размещались материалы для машинного обучения, решения различных задач. Но специалисты этой фирмы, они проявили творческий подход, и они не просто разместили данные, а создали дополнительно мотивацию. Ну, когда просто данные есть, то это, конечно, интересно взять эти данные, посмотреть, получится ли у нас задача, решение задачи, насколько там достоверно всё это получится. Это всё интересно. Значит, я могу вам сейчас сказать, ребята, что вы можете вот выйти на мою страничку, сейчас я дам ссылочку. Ну, ту, где мы скачиваем систему Aidos, короче говоря, вот на эту страничку. А потом сделать запрос по этой страничке UCI, то есть нажать Ctrl+F и задаёте запрос вот этот вот запрос. Вот сейчас, пожалуйста, сделайте это, чтобы я не переписывал туда в чат все эти статьи. Просто я уже их несколько раз уже туда переписывал на разных занятиях, разных группах. И уже сейчас мне больше не хочется.
Попробуйте это сделать и в чате поставьте плюсики, те, кто это сделал.
То есть нашёл мои статьи, в которых, в названии которых есть UCI символы.
Я вам дал задание: вот эту ссылочку нажать и произвести поиск статей, где есть название UCI. И скажите мне, сколько таких статей и какой год самой первой статьи?
Да, 2003. То есть это довольно давно было, 2017 год. То есть, Господи, что я говорю? Я уже устал. Значит, э-э, 17 лет назад я написал первую статью.
Вот. А последняя статья в каком году написана, посмотрите? Самая новая статья, где есть UCI? 2017 или восемнадцатый? Восемнадцатый. Вот это я и хотел сказать. 2018 год. Ну, сейчас мне этот UCI стал меньше нравиться, чем раньше. Я уже объяснил почему? Потому что там просто размещены данные. А на сайте Kaggle, там, в отличие от UCI, создана дополнительная мотивация, которая заключается в том, что есть конкурсы. Значит, есть конкурсы, которые сейчас являются действующими, их называют открытые конкурсы. Есть уже завершённые конкурсы.
И вы данные всех этих конкурсов вам доступны. То есть вы можете взять данные по завершённым конкурсам и по сейчас, которые сейчас активны, в активной фазе, то есть сейчас вот можно прямо решать задачи по этим конкурсам и размещать их в Kaggle.
Как это делается? Значит, скачиваются данные, читается описание задачи, скачиваются данные этой задачи и решается эта задача.
Кубок больших данных.
Вот тут описан смысл этой задачи, которая предложена решать. И есть тут данные.
Видите, данные можно скачать. Значит, я обращаю ваше внимание вот на что здесь. Смотрите, объём данных 2 ГБ с хвостиком. Недели тут есть, я так смотрю. Игроки. То есть здесь довольно много CSV файлов. Видите, да? Вот. Ну, может быть, их можно объединить в один, я не знаю. Но надо смотреть эту задачу. Но я вот такие задачи вам не советую брать, где 2 ГБ исходных данных. Это означает, что там гигантский объём исходных данных, который будет сложно обработать по времени просто. То есть система будет долго эти данные вводить, долго всё будет считать, вот такое вот. Вот, допустим, сутки вводил объём данных 4 млн записей, вот сутки примерно. Сейчас до завтрашнего утра задал на подмножестве определить достоверность модели. Там 4 млн записей. Я посчитал на 1000 примеров, на 1000 примеров, сколько это займёт времени верификация модели. Оказалось 12 минут. Я взял тогда посчитал, сколько ж тогда, если у нас 1000 примеров заняло 12 минут, то сколько нужно примеров взять, чтобы получилось сутки, 1440 минут. Получилось 120.000 примеров. Я сейчас запустил расчёт достоверности модели на 120.000 случайным образом выбранных наблюдений из 4 млн. То есть 4 млн наблюдений, из них случайным образом выбрано 120.000. И на них решается задача.

### 6. Практический пример: Задача классификации листьев

То есть достаточно длительный процесс преобразования идёт. У меня сейчас компьютер перегружен, может поэтому.
Не будем это сейчас делать. Вот. Ну я вам показал. Эти данные, по-моему, я вам послал, да? То есть эту ссылочку на это приложение я вам послал. Что-то не вижу тут. Так, сейчас. Да.
Вот так вы можете найти приложение, интеллектуальные данные к нему, апробировать его на достоверность. Если будут возникать вопросы по ходу этого процесса, то, значит, тогда будете меня спрашивать, что делать. Вот. Ну я вам сейчас могу сказать, что, значит, сейчас мы откроем опять эту инструкцию. Я с неё убежал с инструкции.
Вот. То есть порядок такой: вы выбираете тему, которая вам нравится, которой нет среди уже выполненных тем. Или, может быть, в качестве исключения, что она там и есть, но давно выполнена, и вы можете сделать гораздо лучше это сейчас. Выбираете данные, скачиваете их, делаете модель. Если модель получает достаточно высокую достоверность, то тогда вы можете прислать это мне на утверждение. Я проверяю, если это действительно так и есть, то тогда мы утверждаем тему, и вы её описываете по шаблону. Который здесь у нас четыре шаблона есть разных. Вот. Ну они несколько со временем устаревают, но здесь очень старых нету, 2019, тут 2020 год. Где ещё можно найти исходные данные? Ну я сказал, вот UCI, Kaggle, прямым поиском по набор, поиск наборов данных для машинного обучения. А также на моём сайте есть страничка, на которую есть ссылочка тоже в этой инструкции. Это страница P14 есть ссылочка тут. И здесь есть разные методические работы, учебно-методические, а потом идёт ниже информация о том, какие статьи могут быть выбраны в качестве шаблона описания. Вот. А потом ссылочки тоже на UCI и Kaggle даются. А потом ещё на 50 данных, на 50 лучших баз данных для машинного обучения.
Посмотрим, работает эта ссылочка или нет. Может, она ещё и не работает.
Как-то не очень быстро, да. Что-то не работает. Ну можно вот этот вопрос сделать вот и в поисковой системе найти. Вот. А также, смотрите, что можно сделать? Можно найти статистические базы данных. Вот здесь перечисляются разные варианты. Федеральная статистика, криминальная статистика. Вот. То есть здесь ещё есть указаны разные варианты, как это можно сделать.
Вот. То есть примерно вот таким образом я даю вам много разных вариантов поиска исходных данных. Ищете исходные данные, пробуете сделать модель и апробировать её, определить её достоверность. Как я вот вам показывал в лабораторных работах. И потом утверждаем её и делаем описание по шаблону, размещаем это и вы получаете пятёрку. Если что-то нам не удаётся сделать, ну тогда, значит, не пятёрку, а там четвёрку, тройку получаете. Вот. Или просто сдаёте экзамен, берёте билет и отвечаете. Ну это я, честно сказать, не знаю, какая будет процедура сдачи экзамена онлайн, дистанционного обучения. Я не очень себе представляю, как это может быть. Но если вы такую работу сделаете, опишете, разместите, то это значит, вы разбираетесь уже во всём этом.
Это уже подтверждает, подтверждает, что вы имеете соответствующие необходимые знания, умения, навыки и необходимые для этого компетенции. Есть такой вариант, ребята, который был раньше, который вот прошлые году реализовался этот вариант. Сейчас я не хотел бы его... То есть я вас ориентирую на то, чтобы вы сами зарегистрировались в РИНЦе, в ResearchGate и сами всё это описывали. Ну а я буду вам помогать методически с системой, с описанием. Но чтобы вы сами описание делали и размещали в ResearchGate и РИНЦ, а в облаке я размещу. А теперь смотрите, ребят, вот значит, в прошлый период довольно много у меня есть работ в прошлый период. Вот они жёлтеньким выделены. Прямо подряд идут вот так, видите? Это работы со студентами, которые в основном как раз вот учились по этой дисциплине, которую сейчас мы изучаем. И сделали свои работы. Очень многие сделали свои работы. Ну я так оцениваю, ну я не считал прямо точно там до десятой доли после знака после запятой, но примерно процентов, наверное, 70-80 студентов сделали свои работы, получили пятёрки. Это даже вызвало вопросы в учебном управлении. То есть почему так много пятёрок? Вот. Даже те, кто не сделал работу, тоже они хорошо освоили материал и получили пятёрки. А я им говорю этим сотрудникам управления учебного: а если студент по результатам обучения дисциплины опубликовал статью с присвоением DOI, это как оценивать вообще? Тройка что ли, или двойка там, или четвёрка? Я считаю, что это большое достижение, только пятёркой можно это оценивать. То есть это подтверждает, что он уже приобрёл как навыки исследования, так и навыки описания результатов этого исследования, так и навыки публикации, то есть размещения этого исследования в международных библиографических базах данных. Всё это очень ценно. И для вас ценно прежде всего. То есть вы получается что, что вы заканчиваете обучение по дисциплине, вы готовы уже проводить самостоятельные исследования в контексте, в том, который изучался на этой дисциплине. И писать, делать описание этого исследования и размещать его в международных площадках и в нашей российской системе научного цитирования, индекс научного цитирования. Это, по-моему, достойно всяческой похвалы.
Вот я бы хотел этого от вас сейчас тоже добиться такого же результата. И стараюсь вам сориентировать вас на то, что вам необходимо это сделать будет. Вот. Но для того, чтобы это сделать, нужно, как я вам сказал, решить, найти тему, утвердить её со мной, согласовать со мной. И зарегистрировать, описание сделать по шаблону, который я привёл. Вот здесь вот шаблоны приведены. Причём шаблоны приведены в вордовские файлы. Я специально не ПДФ файлы сделал, а вордовские, чтобы вы могли их просто взять и корректировать, что проще, чем писать с нуля. Вот. И потом разместить их в, как я сказал, в облаке, ResearchGate и РИНЦе. Чтобы разместить в РИНЦе, ребята, вот это Elibrary имеет, так сказать, это общее название, есть некое подмножество Elibrary, которое более престижные журналы включает, которые входят в РИНЦ, Российский индекс научного цитирования. А ещё более престижное есть подмножество, которое называется Science Index. И вот мы сейчас, я вам сейчас дам ссылочки на саму эту систему Elibrary и на Science Index. Ну в Science Index он там при регистрации вы просто птичку поставите, что регистрироваться не только в Elibrary, но и в системе Science Index. И там будет ещё заполнение идти разных пунктов. Значит, эти пункты, карточки регистрационные, есть обязательные, есть необязательные. Вот эти вот необязательные, вы там не заморачивайтесь, не заполняйте, да и всё.
Вот. Потом дальше смотрите внимательно сейчас на экран. Когда вы получаете регистрацию, это проходит недели две, как минимум, я так думаю. Быстрее вы не получите. Дай бог, чтоб к концу года вы это сделали. Значит, смотрите, мы значит, я вам показываю, где мы заключаем договора на размещение непериодических публикаций. Автором прокручиваем, размещение непериодических изданий, прокручиваем система заключения договоров. Шлю вам ссылочку. Эти ссылочки прямо копируете, сохраняете и прямо ими пользуетесь. Но я вам могу сказать, что вот в этой инструкции здесь есть файлик такой, вот смотрите, ResearchGate Doc, видите, да? Вот где примерно вот описано то, что я сейчас вам рассказываю, ну, не так подробно, но суть описана, что как вот там зарегистрироваться и там, и там. И для чего? Здесь это написано вот в этом файлике. Вы можете его скачать и почитать. Если мы его скачаем, откроем, то здесь вот, в общем-то, мы видим некое описание несложное. И кое-какая информация здесь есть. Для чего это нужно и что это даёт.
Ну если об этом сказать словами, что это даёт, то это даёт вам возможность, ребята, в любой момент опубликовать работу, которая будет иметь статус работы с DOI (Digital Object Identification), то есть это цифровой идентификатор объекта, международный идентификатор статьи, который будет у неё сохраняться, где бы она, в какой бы базе библиографической она ни была, в Elibrary, в РИНЦе или там где-то в других базах данных, международных, этот идентификатор он является тем же самым, остаётся. Вот. Но у вас появляется публикация, которая входит в портфолио. Вы демонстрируете то, что являетесь уже начинающим учёным. То есть вы можете поставить исследование, провести исследование, описать его в виде научного материала какого-то, типа статьи или отчёта научного, и опубликовать его можете. И при этом вы ни от кого абсолютно не зависите, и всё это абсолютно бесплатно. А вообще-то мне приходит реклама о том, что, пожалуйста, можно опубликовать статью с присвоением DOI, и стоимость всего 183 руб. за страницу, к примеру. Вот. Ну и статья, допустим, 10 там страниц, ну значит, около 2.000 руб. платишь и получаешь эту публикацию. А здесь вы ничего никому не платите и не ждёте, когда выйдет журнал или там конференция произойдёт, а просто у вас появляется публикация в списке ваших работ, в портфолио. По-моему, это играет роль для получения стипендии также, мне кажется. И прямую роль играет для получения хорошей оценки по дисциплине, отличной оценки.

### 7. Поддержка и коммуникация

Вот. Ну, так что я вам, в общем, описал, как это всё делается. Вот, как искать исходные данные, как их проверять, насколько хорошо получается модель с этими данными. И как утвердить тему и всё остальное я вам коротко рассказал. Теперь, когда у студентов возникали вопросы по ходу реализации этой программы, выполнения этого задания, я на них отвечал на эти вопросы. И вот в это, в эту инструкцию я записывал в краткой форме ответы на эти вопросы. Поэтому эта инструкция, она уже такая отработанная и содержит ответы, которые, на которые, ответы на вопросы, которые возникали у студентов, когда они выполняли это задание. Ну, не на все, конечно. Вот, допустим, здесь не написано, что можно вдвоём выполнять задания, втроём нельзя. Да, вот ещё что. Значит, если вам не удастся зарегистрироваться в Elibrary и Science Index или в ResearchGate, то тогда я могу разместить там ваши статьи, но тогда мне придётся быть соавтором ваших статей. Вот. Почему? Потому что только соавтор может размещать материалы в этих системах. Значит, для того, чтобы зарегистрироваться в ResearchGate, нужно иметь корпоративный адрес электронной почты. Для того, чтобы зарегистрироваться в РИНЦ, ничего не надо иметь, кроме, так сказать, головы там и рук, компьютера. Садишься и просто регистрируешься и всё, и ждёшь. Определённое время придётся подождать, потому что там не очень быстро присваиваются эти все дела. Ну так вот, вкратце, ребят. Примерно я вам описал, о чём идёт речь.

### 8. Вопросы и ответы

Теперь мне нужно, чтобы вы задали какие-то вопросы. Какие у вас возникли вопросы в этой связи?
Ну и что означает тишина?
Вопросов нет.
Всё понятно, да? Ну тогда за работу, товарищи. Что это значит? Это значит, сейчас давайте начинайте процесс регистрации, вот, процесс поиска темы. Вот, и всё остальное продолжайте делать. Значит, если у вас какие-то возникнут вопросы по ходу выполнения этих работ, вот, то вы можете мне написать на почту.
Вот. Пишите мне на почту. Спрашивайте. Пишите, значит, в таком стиле, что я такой-то, такой-то, студент такой-то группы. Вот я выбрал такую тему предварительно и хотел бы, так сказать, её утвердить. То есть пишите, так сказать, сразу, что вы хотите. Или у меня там что-то не получается, не получается ввести данные в систему, не получается создать модель, возникают какие-то ошибки. И обязательно в этой, в этом письме указывайте источник данных, откуда вы взяли данные. И было бы неплохо, если б сами данные привели. Вот. Ну, в принципе, мне несложно их оттуда взять и посмотреть, что там у вас такое получается.
А можно вопрос?
Конечно, не можно, а нужно.
А в РИНЦе требуется год окончания вуза и максимум можно выбрать 2020?
Значит, вы, значит, не пишите просто, там это поле не является обязательным, насколько я понимаю.
Оно обязательно нарисовано.
Звёздочка стоит, да?
Да.
Ну тогда поставьте вот этот год, да и всё, какой можно, самый поздний. 2020. Вы должны понимать, что эта информация особой роли не играет, которая там будет. Там главное, чтобы было правильно фамилия, имя, отчество ваше. И обратите, внизу там есть птичка такая, где, если вы её поставите, то ещё добавится много вопросов. Это регистрация в системе Science Index. Вот тоже эту регистрацию обязательно заполняйте.

### 9. Заключение и дальнейшие шаги

Ну, ладно, ребят. Значит, сейчас тогда я думаю, что ж мне сделать? Тут как-то не совсем понятно, что делать, потому что студенты ведут занятия везде.
Такая вот ситуация.
Ну, давайте тогда сейчас я вам приведу пример, как можно найти данные и попробовать их проверить на то, насколько они являются приемлемыми, насколько они приемлемы для того, чтобы сделать приложение. Вот, значит, выходим на сайт Kaggle.
Вот, и смотрим здесь, конкурсы идут.
Можем, используя Google Chrome, можем перейти на русский регистр, то есть русифицировать вообще всё полностью этот интерфейс. Видно всё по-русски.
Ну смотреть, конечно, что там за данные. Вот, скажем, сейчас смотрим эти данные. 80 злаков.
Download. Всего лишь 5 КБ. Мизерный объём данных.
Вот.
Берём, открываем Excel. Помещаем этот набор данных куда-нибудь в какую-нибудь папочку, которая будет легко доступна из Excel. Вот, открываем Excel, ну чтобы легче искать, не закопанную где-нибудь там глубоко. И пытаемся её ввести. Значит, я ещё раз вам показываю, что CSV файлы вводятся следующим образом: выбираем в горизонтальном меню, выбираем пункт Данные, ввод из текста и указываем путь на этот файлик. Она здесь нам показывает, видите, там не поймёшь что, да? Мы, значит, двигаемся дальше, выбираем разделитель. И тогда у нас данные все по колоночкам распределяются. И мы видим, что это нам достаточно. Получаем... Сейчас она введёт эти данные, конвертирует их в Excel, в экселевскую таблицу. Вот эту операцию конвертирования, её система Aidos выполняет у неё в режиме 2.3.2.2 есть конвертер.
Вот. И мы эти параметры и показываем системе Aidos. Сначала удаляем все приложения в системе, которые там были. Вот. Потом переходим в интерфейс программный 2.3.2.2. И мы здесь вот задаём, что у нас классификационная шкала шестнадцатая, а описательные шкалы у нас идут со второй по пятнадцатую. И нажимаем О'кей. Этот файл вводится в систему. Вот сейчас мы видим уже, что он введён. И теперь происходит переход на ввод уже dBase файла, то есть он преобразован из CSV в dBase. И дальше идёт так, как будто бы мы этот dBase файл или экселевский файл указали системе. То есть дальше идёт обычный порядок действий. То есть мы указываем, сколько у нас градаций в классификационных шкалах, сколько в описательных шкалах. Вот, ну, допустим, здесь как бы семь укажем градаций. Вот. И пересчитываем их, пересчитываем. Здесь вот давайте пять укажем, чтоб поменьше было. Вот. И выходим на создание модели.
Вот. Теперь, ребята, вы видим, что в двух колоночках нет вариабельности данных. Они просто не будут использоваться. Создаём модель в режиме 3.5. Поскольку у нас выборка... А, ну давай посмотрим на объём выборки. На классификационные шкалы, описательные шкалы. Одна классификационная шкала, семь градаций у неё. И сколько у нас тут? 12 описательных шкал. И в последней 60 градаций, код. Пять градаций в них. Они все у нас числовые, да? Все числовые шкалы. И там дат, ничего, то есть корректно преобразовалось. Смотрим обучающую выборку. Смотрим, сколько здесь у нас объектов? Всего 77 объектов в обучающей выборке. Это очень мало. Это составляет одну тысячную процента от того, что может обрабатывать система Aidos. Нажимаем О'кей. И создаются модели, а потом проверяются на достоверность. И мы как раз доходим до того пунктика, который нас и интересует. Нас интересует достоверность. Если получилось у нас неплохая достоверность, то можно брать эту тему для собственной работы, для описания её по шаблону, размещения. И вы осваиваете все знания, умения и навыки, компетенции, которые нужно по этой дисциплине положено освоить. И заодно зарабатываете себе стипендию и для портфолио получаете статью. Всё. Смотрим достоверность. Ну здесь вот надо, конечно, указать название приложения. Сейчас мы посмотрим на Kaggle. Вот, тут у нас есть название.
Мы его можно, кстати, ничего не переводить на русский, потому что вполне нас устраивает и чтоб были английские названия всех колонок. Ну приложение. И смотрим достоверность приложения. То есть удалось ли нам выявить какие-то зависимости между значениями признаков этого зерна и его качеством. И можно сказать так, что это удалось сделать. То есть мы видим по модели N3 сумма знаний, у нас критерий, F-критерий L1, который я предложил, это мультиклассовое нечёткое обобщение F-критерия Ван Рисбергена, 0,705. Ну это, скажем так, не идеальная модель, но вполне, так сказать, это разумная, но отражающая зависимости между признаками и принадлежностью объектов с этими признаками к каким-то классам. То есть зависимости выявлены. Вот. И мы можем теперь их исследовать, какие выявлены зависимости. То есть такие задачи уже можно решать, можно брать их для решения. Здесь, конечно, вот не очень хороший такой есть момент, провал такой, видите? При уровнях сходства 80-80%, ложных решений больше, чем истинных. А до этого всё шло хорошо. Повышался уровень сходства и решений, достоверных решений тоже становилось больше. А при критерии, при модели N1... Да, этот провал есть во всех моделях, к сожалению. А если мы возьмём адаптивные интервалы, там же все шкалы числовые, тогда мы можем взять адаптивные интервалы. Теперь мы можем уже вводить данные не из CSV файла, а из dBase файла. У нас все параметры сохраняются, dBase файл там есть. Поэтому мы можем уже этот конвертер не использовать CSV, он нам не нужен. А здесь мы видим, ребята, сколько у нас наблюдений в тех или иных шкалах. По 16 наблюдений. А рейтинг у нас что-то вообще... Надо опять указать, может быть. Странно, странно. Да, это здесь что-то странное.
Вот. Ну и выйдем на создание модели, посмотрим. Вот. И также осуществляем синтез модели и проверки её, проверку её на достоверность. Ну я вам что могу сказать, что адаптивные интервалы обычно дают лучшие модели, чем равные интервалы. То есть вот мы сейчас как раз пробуем выкрутиться, получить модель более высокого качества.
Ссылочку на это приложение даю вам. Может, кто-нибудь его и возьмёт сейчас себе, кто успеет.
Вот. Ну это даже получше. По классическому критерию Ван Рисбергена даже получше. То есть нормальная модель получилась, которая позволяет определять рейтинги этой пшеницы. Всё, ребят, получилось у нас. Максимальный рейтинг, модель N1, диаграмма. Сколько там калорий, протеина, сахара. То есть максимальный рейтинг, когда сахара минимальное количество, калорий минимальное, клетчатки там максимальное, протеина максимальное. То есть дальше можно работать с моделью уже. То есть она получилась.
Вот такой путь мы проходим для того, чтобы определить, можно ли браться за это приложение или нет. Если вот эти кривульки получаются разумные, при оценке достоверности, вот эти вот, которые я вам показал, вот эта картина, она разумная. То есть при повышении уровня сходства доля истинных решений становится больше, чем доля ложных решений. Ложные решения есть, но они все с низким уровнем сходства. То есть уровень сходства, который даёт система, он является адекватным критерием достоверности решения.
Так. Значит, теперь, ребята, в аграрном университете. Значит, э, эту тему можно брать для описания. Я сейчас вам показал с начала до конца, как мы её нашли, проверили её на пригодность для, ну, в общем, короче, мы определили, насколько система Aidos может решить эту задачу выявления взаимосвязи между этими признаками и принадлежностью к этим категориям. Получилось, что позволяет. Ну раз позволяет, то можно брать её за основу для описания. Дальше делать описание. Ну и регистрироваться одновременно везде во всех этих системах. В ResearchGate, чтобы зарегистрироваться, нужен корпоративный адрес электронной почты иметь или обратиться в ЦИТ, Центр информационных технологий, чтобы мне подтвердили, что вы являетесь, ну, что учитесь в университете. И всё. И больше от них ничего не требуется. Всё, ребят, до свидания.
До свидания. До свидания. До свидания. До свидания. До свидания. До свидания. Начинайте регистрироваться. На следующем занятии мы продолжим. Если будут вопросы, я буду помогать вам решать задачи конкретные. То есть вы будете озвучивать, я буду помогать на них ответы.