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

***Российская Федерация***

**193 Практическое занятие. По дисциплине "Теория информации, данные и знания". 2020-11-10**

## Заголовок

Практическое занятие по системе Эйдос: Разработка приложений и анализ рейтинга ВУЗов

## Резюме текста

**1. Введение**  
10 ноября 2020 года состоялось практическое занятие №6 по дисциплине "Теория информации, данные, знания" для группы ИТ-20-03. Занятие проводили профессора Е.В. Луценко и Г.А. Аршинов. Время проведения: 11:30 - 13:00. Цель занятия – разработка студентами собственных приложений в системе Эйдос или выбор темы для них.

**2. Обзор студенческих работ и инструментов**  
Профессор Луценко отметил получение работы от студентов Ващенко и Боженова (группа ИТ-20-02), размещенной в облаке Эйдос, что свидетельствует о начале активной фазы работы. Также была упомянута другая полученная работа, по которой даны замечания и отправлена на доработку. Были внесены изменения в инструкцию по разработке приложений: рекомендовано использовать меньшие объемы данных (Кб вместо Мб) и добавлен 5-й шаблон описания работы, основанный на примере Ващенко/Боженова. Ссылка на шаблон и рекомендованный онлайн-конвертер файлов были предоставлены студентам.

**3. Проверка прогресса студентов**  
Проведена проверка наличия утвержденных тем у студентов группы ИТ-20-03 с использованием чата (+/-). Большинство студентов подтвердили наличие тем. Студентам без утвержденных тем (София, Илья) было предложено использовать время занятия для обсуждения и доработки своих предложений.

**4. Пример анализа данных: Рейтинг ВУЗов**  
В качестве демонстрации возможностей системы Эйдос был проведен анализ данных по рейтингам университетов мира.

* **Подготовка данных:** Из исходного файла были выбраны шкалы для классификации (мировой и национальный рейтинг) и описательные признаки (название ВУЗа, страна, качество обучения, число публикаций, цитирований, патентов и др.). Год был исключен как нехарактерный признак.
* **Настройка модели:** В системе Эйдос был задан тип файла (XLS), определены классификационные и описательные столбцы, установлено 10 градаций для рейтинговых шкал с использованием адаптивных интервалов для равномерного распределения наблюдений.
* **Синтез и анализ модели:** Были синтезированы модели, проведена их проверка на достоверность на обучающей и распознаваемой выборках. Результаты (дендрограмма, таблицы важности признаков) были продемонстрированы.
* **Интерпретация результатов:** Анализ показал логичную кластеризацию ВУЗов по уровням рейтинга (низкий, средний, высокий). Наиболее значимыми факторами, влияющими на рейтинг, оказались количество публикаций и цитирований. Другие показатели, такие как плата за обучение, международное влияние, качество персонала, имели меньшую ценность для модели.

**5. Заключение и рекомендации**  
Студентам (София, Илья) были переданы результаты анализа и скриншоты настроек модели в качестве примера. Было рекомендовано использовать полученную модель и шаблон №5 для описания своей работы. Обсуждалась возможность использования другого набора данных (кредитный рейтинг) и его анализ. Занятие завершилось.

## Детальная расшифровка текста

**1. Введение**

Здравствуйте, Евгений Вениаминович.  
Да, здравствуйте, Юрий Александрович. Здравствуйте, ребята.  
Здравствуйте, ребята.  
Здравствуйте.  
Здравствуйте.  
У нас сегодня  
10 ноября 2020 года.  
Третья пара.  
11:30 - 13:00.  
Практическое занятие  
по дисциплине Теория информации, данные, знания.  
Занятие ведут  
профессор Луценко Евгений Вениаминович  
и профессор Аршинов Георгий Александрович.  
Сейчас, ребята, я вас всех поотмечаю  
и сам поотмечаюсь, и потом начнём занятие.

ПИН-2003.

Практическое занятие номер шесть.  
Занятия, как и  
предыдущие,  
посвящены разработке собственных приложений  
и  
выбору темы, если темы ещё нет.

Практическое занятие номер шесть.

**2. Обзор студенческих работ и инструментов**

Вот, ребята.  
Вчера произошло интересное событие.  
Прислали мне  
работу  
сделанную,  
описание.  
Я разместил её эту работу в системе Эйдос.  
Выдаст в облаке.  
То есть лёд тронулся.  
Процесс пошёл.  
Вот смотрите.  
Смотрим. Диспетчер приложений.  
Скачать приложение из облака.  
И вот видите,  
Ващенко, Боженов  
восьмого числа, восьмого...  
Это позавчера.  
Вот.  
Кликаем по этой строчке.  
Открывается  
описание.  
Группа ИТ 2002.  
А у вас 2003.  
И сегодня ещё пришло одно  
описание с исходными данными, но тоже  
для проверки и размещения. Но я его,  
в общем, прочитал, сделал ряд замечаний и отправил  
обратно  
на исправление замечаний.  
И в инструкцию по разработке собственных приложений внёс небольшие изменения.  
Совет такой: брать данные, которые измеряются не в мегабайтах, а в килобайтах,  
поменьше объёма данных.  
И добавил сюда...  
Сейчас я вам пошлю в чат.

Саму инструкцию.  
Добавил пятый шаблон,  
который как раз представляет собой статью этих студентов.  
И советую вам этот шаблон использовать,  
потому что это новый шаблон. То есть там обновлена информация,  
рисунки обновлены по сравнению с теми, которые были до этого.  
И ссылка дана прямо на сам вордовский файл  
в Эйдос облаке.

**3. Проверка прогресса студентов**

Ребята, если какие-то вопросы у вас есть, пожалуйста, задавайте.

У нас занятие посвящено  
разработке собственных приложений.  
Поэтому, значит, возникает вопрос:  
у вас с этим есть какие-то проблемы с разработкой собственных приложений? Вы работаете над этим или нет?  
Работаем, но сейчас сложно пока статью написать, ну, описание сделать.  
А в чём проблема?  
Во времени?  
Ну, проблема... просто слова подобрать нужно.  
Поэтому думаем, как сделать.  
Ясно.  
Ну вот берите шаблон этот описания, который я вам послал сейчас.  
О'кей, спасибо.  
Вот. И по нему  
пишите,  
подбирайте  
слова по своей тематике, конечно.

Так, теперь мне в чате, пожалуйста, напишите, у кого нет тем.  
Кто не делает приложение?

Ну можете так написать: если есть тема, вы работаете, утверждённая уже,  
то плюсик, а если нету, тогда минус.

И у нас сейчас здесь 14 студентов присутствует.  
Поэтому должно быть плюсов и минусов 14  
вместе.  
Пока шесть плюсов есть.  
Это неплохо, но вот седьмой Максим.

Восьмой Егор.

У таких, у которых нет темы утверждённой, таких нет студентов у нас?

Так.  
Ну не все написали.  
У нас сейчас присутствует, ребята, уже 16 студентов.  
Вот. Хотя, может, кто-нибудь из них и не студент, а проверяющий, например.  
Похоже, что все студенты  
присутствуют.  
16, 16.  
Значит, этих вот плюсиков и минусиков в совокупности должно быть 16.

Вот. Значит,  
София и Илья.  
Вот сейчас как раз вот та пара,  
на которой мы можем посмотреть.  
Это я имел в виду вот такое занятие,  
как сейчас.  
Поэтому давайте мне  
ещё раз посылайте на почту, чтобы я там не искал особенно,  
а чтобы это вверху у меня было сразу же.  
И то, что вы уже присылали.

Пока не вижу вашего письма.

**4. Пример анализа данных: Рейтинг ВУЗов**

**4.1. Подготовка данных**

Посмотрите, пожалуйста, я вам отправила.  
Спасибо.

Сейчас мне... сейчас я  
пошлю  
тому, кто мне всё требует  
материал.

И сейчас посмотрим.

Ну не прислали ссылочку на  
источник данных, ребята.

Ладно. Значит,  
вот здесь вот смотрите, ребят, значит, нам эта первая колоночка не нужна практически.  
Или это ранг  
мировой ранг?  
Ну тут сейчас может как раз она и нужна. Сейчас посмотрим.

Значит, смотрите, ребята, мы можем взять мировой ранг, да?  
Рейтинг. И  
национальный рейтинг.  
Потом само название университета возьмём просто  
скопируем первую колонку, а здесь уберём.  
Это просто у нас будет  
информация о том,  
что за университет.  
А классификационные шкалы у нас будут мировой рейтинг и национальный рейтинг.  
Вот, ну дальше у нас идут строчки.

Сейчас здесь Россия тут есть вообще или нет?  
А, есть.  
Ломоносов.  
90 рейтинг в мире.  
48-й рейтинг в мире.  
Санкт-Петербургский.  
Новосибирский.  
Ну, в общем, кое-где тут Россия проскальзывает.

**4.2. Настройка модели**

Значит, смотрите, ребята, мы определяем, какие шкалы и градации.  
Классификационные шкалы и описательные шкалы. Классификационная вторая шкала у нас.  
Третья по  
тринадцатую – это описательные шкалы.  
Год мы, наверное, не будем брать, потому что он не характеризует университет. А вот эти показатели характеризуют.

Сбрасываем все приложения.  
Заходим в режим 2 3 2 2.  
Здесь задаём  
какой тип файла смотрим.  
Старый XLS.  
Старый Excel.  
Вторая колоночка – это у нас рейтинг  
всемирный. А с третьей по тринадцатую – это у нас  
описательные шкалы.  
И задаём разные интервальные значения,  
разные интервалы примерно одинаковым числом наблюдений.  
Вполне нормально.  
То есть вот такие задаём параметры.

**4.3. Синтез и анализ модели**

Смотрим, сколько у нас на одно  
на диапазон получилось  
наблюдений.  
Данные распределены неравномерно, поэтому адаптивные интервалы очень кстати.

Задача достаточно большой размерности.

Что у нас получилось? У нас получилось 100 градаций мирового рейтинга.  
Вот. И  
показатели этих вузов.  
Национальный рейтинг, страна.  
Дальше числовые показатели.  
Дальше мы осуществляем синтез моделей  
и оставляем из них  
10%  
наиболее достоверных результатов идентификации.

Время займёт у нас. Ну,  
может, поменьше чуть-чуть, может, побольше.  
Сейчас уточнится прогноз.

Вот, то было 7 минут, а сейчас стало около 2 минут.  
Меньше, чем 2 минуты.

Сейчас она создала все модели  
и все их проверяет на достоверность на примере 2.000  
2.000 вузов.  
2.201 вуз.

**4.4. Интерпретация результатов**

Ну что, получилось вполне нормально.  
То есть у них рейтинг зависит от всех этих показателей, ребят.

Вот смотрите. Значит, здесь вот у нас рейтинги, видите? Вот девятый из десяти и восьмой из десяти.  
И десятый из десяти. Они сходны. Вот восьмой и девятый сходны.  
К ним примыкает десятый. Это правильно. Восьмой, девятый, десятый – они все сходны по своим  
показателям вузы. Пятый и шестой сходны, видите? Это средний уровень.  
И седьмой к ним примыкает тоже. Вот. Тоже они сходны.  
И низкий рейтинг вузов. Первый и второй, они очень похожи, видите?  
На самом высоком уровне сходства.  
К ним примыкает третий, четвёртый, тоже низкий рейтинг.  
То есть можно сказать так: у нас конструкт сформирован  
рейтинга вузов.  
Низкий рейтинг  
от первого до четырёх.  
И высокий –  
все остальные.  
Ну, можно, по-видимому, выделить и средний какой-то.  
Вот.  
Пятый, шестой, седьмой.  
Вот с первого по четвёртый – это низкий.  
Шестой, с пятого по седьмой – средний.  
И с восьмого по десятый – это высокий.  
Вот примерно так.  
Это очень логично всё, ребята. Это означает, что эта шкала рейтинга работает  
при таком числе градаций  
небольшом, 10. Она вполне, так сказать,  
разумные даёт результаты.  
А когда было большое число градаций, 100, то она давала безумные результаты.  
Очень странная была  
форма по зависимости достоверности, то есть числа решений истинных и ложных в зависимости от достоверности.

Это как формы, ребята, которые показывают, как влияют различные показатели  
вузов  
на их рейтинг всемирный.  
Национальный рейтинг находится в соответствии с международным.  
Страны, как влияют. Но страны не являются  
ранговой шкалой.  
Качество обучения.  
Число публикаций явно зависит от уровня мирового.  
Число ссылок явно зависит.  
Вот. То есть некоторые показатели, они прямо вот линейно связаны с рейтингом.  
А некоторые обратно пропорционально присвоены, связаны или каким-то непонятным, сложным образом.

Вот. Я посылаю это всё Софии и Илье.

Смотрим теперь,  
создаём модель.  
Смотрим, какая модель, насколько она достоверная.

Вот, теперь пишем название. София, обрати внимание.  
В режиме 1.3  
пишем название.  
АСК-анализ рейтинга вузов.  
Вот. И смотрим, насколько у нас получилось всё.  
Ну, такое впечатление, что вроде как получилось более-менее.  
Так себе.

Одна не получилась.  
Странное дело.  
Вот это получше.

То есть мы попросили у системы  
очень высокую достоверность.  
Но всё равно такой провал есть.  
София, вы видите, нет, что я показываю?  
Видите, здесь очень интересное есть такое место, где при 60 до 80 уровня сходства, там иногда ложные решения больше  
доля больше, чем доля истинных решений.

Ну вот, ребят. Теперь давайте ещё раз смотрим какую-нибудь задачу. Ну вот София и Илья уже я много вам форм набросал там.  
То есть у вас вполне, так сказать, интересный получается анализ.  
Если будут вопросы, спрашивайте там на почту пишите.  
Хотя вот вам пятый шаблон, он позволит вам аккуратно всё сделать.  
Особо так по поводу объёма листах описания особо не стесняйтесь, ну там может по 50 получиться. Потому что это вы говорили, скажите ещё раз.  
Что-то сказали, я не слышал, что вы сказали. Я говорил в это время.

**5. Заключение и рекомендации**

Ну вот, ребята, можно узнать, как зависит кредитный рейтинг  
от параметров  
человека:  
возраста, пола,  
закона страны  
и так далее. Оказалось, вполне зависит.  
То есть когда приходит человек занимать деньги в банк  
на физическое лицо,  
кредит брать на физическое лицо, то они могут посмотреть,  
какой у него кредитный рейтинг,  
просто введя его данные.  
Если есть желание, можете взять эту задачу кто-нибудь.  
Ну то я сейчас сотру этот экселевский файл и всё.  
Я послал ссылочку на источник данных.

Так, ребята, у меня вообще слышно, нет?  
Там вы ещё не там?  
Я говорю, давайте ещё раз...  
Да, слышно.  
Молодцы. У вас вопросов вообще нет, что ли? Давайте рассмотрим какие-то ещё работы, какие-то пробы, вопросы возникли.

Ребята, сейчас 40 минут занятия, понимаете?

Ну что, ребята, может быть, есть посмотреть, есть смысл посмотреть и  
побольше.  
100, например.  
100 интервальных значений.  
И смотрим. У нас получается  
по 22 наблюдения на одно интервальное значение.  
Вполне нормально.

Ну вот, я ещё одну задачу стал смотреть, можно ли её решить.

Ну что, ребята, давайте теперь кто-нибудь ещё.  
Кто-нибудь ещё может, пожалуйста, сейчас можно ещё раз смотреть какую-нибудь работу.

Ну что, ребята, давайте ещё что-то рассмотрим.  
У нас ещё времени достаточно.

София, тогда я вам ещё кое-какие формы пришлю сейчас.  
Раз там все думают,  
надо или не надо. Ну, значит,  
будем вам делать.  
Значит, София, вы видите эту форму, дендрограмму?  
Вот смотрите.  
Ну на там, где вот источник данных вы взяли, там тоже есть  
автор этого этих данных. У них, может быть, есть какие-то статьи. На них тоже надо сослаться.  
Ну и на источник данных обязательно. Там, где вот таблица приводится в описании.  
Кстати, используйте шаблон пять  
для  
описания  
вашей работы.  
В общем, при таких параметрах эту модель можно описывать.  
Шаблон описания присылаю.  
В общем, при таких параметрах давайте делайте и описывайте.  
Я вам послал какие параметры создания модели,  
какие формы  
получаются. И сам файл исходных данных, с которым работать, послал.