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

***Российская Федерация***

***Воронин Г.Г. ПИ2104 КубГАУ mailto:glebvr17@gmail.com***

***вашсайт***

**206 Лабораторная работа №11. По дисциплине Инженерия знаний и интеллектуальные системы. 2020-11-16**

**Заголовок:** Использование системы Aidos для анализа рейтинга ВУЗов: Практическое занятие

**Резюме:**

1. **Введение и Информация о Занятии:**
	* Занятие (Лабораторная работа №11) состоялось 16 ноября 2020 г. для группы B1701 по дисциплине "Инженерия знаний и интеллектуальные системы".
	* Проводят профессора Луценко Е.В. и Аршинов Г.А.
	* Тема: Разработка приложений, выполнение индивидуального задания.
2. **Цель и Важность Задания:**
	* Задание направлено на подтверждение усвоения студентами знаний, умений и навыков по дисциплине.
	* Выполнение задания является обязательным для получения положительной оценки (4 или 5) и особенно актуально в условиях дистанционного обучения из-за пандемии.
	* Задание размещено на ResearchGate (предоставлена ссылка на новую версию).
3. **Ход Занятия и Взаимодействие:**
	* Основная цель сессии – разбор вопросов и проблем студентов при выполнении задания.
	* Проведена проверка присутствия с использованием чата (+/- в зависимости от наличия темы). Отмечена низкая явка (5 из 12 студентов).
	* Профессор просит передать отсутствующим информацию о важности выполнения задания для итоговой оценки.
	* Произошло короткое прерывание на телефонный звонок.
4. **Практическая Демонстрация (На примере задачи студентки Екатерины):**
	* **Подготовка данных:** Показан процесс работы с исходными данными (рейтинги ВУЗов). Обсуждается необходимость обновления системы Aidos (указана старая версия от 15 ноября) и процедура обновления через патч. Исходный файл данных в формате DBF преобразуется в Excel с помощью онлайн-конвертера (Convertio), так как встроенный конвертер может не справиться. Данные в Excel структурируются: столбец рейтинга перемещается, лишние столбцы (например, "Год") удаляются, пропущенные значения в числовых столбцах заменяются нулями.
	* **Работа в Aidos:** Демонстрируется ввод данных и настройка параметров в системе Aidos. Указываются классификационные (рейтинг) и описательные (прочие характеристики ВУЗов, страна) шкалы. Устанавливается количество градаций для шкал.
	* **Синтез Модели:** Запускается процесс автоматического построения моделей в системе Aidos. Отмечается высокая достоверность полученных моделей (F-критерий около 0.87).
	* **Интерпретация Результатов:** Демонстрируются и кратко поясняются различные формы представления результатов: статистические модели, системно-когнитивные модели (включая графики информативности признаков, когнитивные функции, дендрограммы сходства объектов и признаков). Анализ дендрограмм подтверждает логичность кластеризации ВУЗов по рейтингам.
	* **Решение Задач:** Показано решение задачи идентификации (прогнозирование рейтинга ВУЗа по его характеристикам).
	* **Передача Результатов:** Профессор отправляет обработанные файлы данных, скриншоты параметров и результатов анализа студентке по почте.
5. **Требования к Отчетности и Публикации:**
	* Студентам необходимо подготовить описание своей работы (статью/отчет), используя предоставленные шаблоны (рекомендован шаблон №5).
	* Обязательна регистрация на платформах ResearchGate, РИНЦ (Science Index).
	* Нужно заключить договор на размещение публикаций и разместить итоговую работу.
	* В работе необходимо указать источник данных (Kaggle) и используемую систему (Aidos).
6. **Завершение:**
	* Даны ответы на вопросы (о получении корпоративной почты через ЦИТ).
	* Еще раз подчеркнута важность качественного оформления работы перед публикацией.
	* Занятие завершено.

**Детальная расшифровка текста:**

**1. Введение и Информация о Занятии**

Здравствуйте, ребята.
Здравствуйте. Здравствуйте. Здравствуйте. Здравствуйте. Здравствуйте.
Да, здравствуйте, здравствуйте.
Сегодня у нас 16 ноября 2020 года.
Шестая пара, 17:20-18:50.
Лабораторная работа номер 11 с группой B1701 по дисциплине инженерия знаний и интеллектуальные системы.
Занятия ведут профессор Луценко Евгений Вениаминович и профессор Аршинов Георгий Александрович.

**2. Цель и Важность Задания**

И у нас сегодня темой нашего занятия является разработка приложений.
То есть выполнение вами задания, которое вами вам было дано.
Что это за задание, ребята?
Это задание, оно направлено на подтверждение вами получения вами знаний, умений и навыков, предусмотренных при изучении этой дисциплины.
Вот.
Я разместил это задание в ResearchGate.
В чате даю ссылочку.
Это новая версия этого задания, размещённая в ResearchGate.
Вот. И мы с вами сейчас должны решать те вопросы, которые у вас возникают или проблемы в связи с выполнением этого задания.
Значит, я подчёркиваю, это задание является способом, который мы придумали профессора, преподаватели, в условиях пандемии, в условиях дистанционного обучения, для того чтобы проверить факт освоения вами необходимых, предусмотренной предусмотренной дисциплиной знаний, умений и навыков и получения вами соответствующих компетенций. То есть, по тому, как вы это задание выполните, вам будет выставляться оценка, о чём вам говорилось с самого начала преподавания этой дисциплины, на первом занятии.

**3. Ход Занятия и Взаимодействие**

Так, теперь, пожалуйста, в чате напишите мне, ну, во-первых, вас там в группе-то побольше, пожалуй, чем семь человек. У вас там семь человек - это вместе с нами, с двумя профессорами. То есть вас сейчас на занятии пять человек.
Вот. А по перекличке вас должно быть 12 человек.
Есть некая разница, да?
Так вот, тем, кто отсутствует, ребята, пожалуйста, передайте, что вот то, что я сейчас говорил, это я говорил вполне серьёзно.
То есть я, конечно, двойки вам не поставлю, не выгоню из университета за то, что вы не изучали эту дисциплину, но и четвёрки не поставлю, и пятёрки, если вы не решите задачи соответствующие. Не продемонстрируете, что вы освоили лабораторные работы, которые мы проходили по изучению табличной информации, числовой и текстовой, и по изучению технологии обработки, интеллектуальной обработки изображений, в частности спектральной, интеллектуальной спектральной обработки изображений.
Вот. Теперь, пожалуйста, сейчас в чате мне напишите, у кого есть тема и кто работает над её выполнением, плюсики поставьте. У кого нет темы, поставьте минусики.
Ребят, вы обязаны поставить либо плюс, либо минус. Я должен их посчитать потом, и у меня должно получиться пять. Пять плюсиков и пять минусиков. В смысле, всего плюсов и минусов должно быть пять в сумме.
То есть пока что я вижу пять, то есть три плюсика.
Вот, это Анастасия, Олег и Виктория. А минусиков не вижу.
Из этого я могу сделать вывод, что эти ребята, которые не поставили минусики, что они отсутствуют.
Так. Ну, я не знаю, у кого-нибудь ещё звук пропадает, то, что я говорю, или нет?
Вот тут...

*(Телефонный звонок)*
Алло. Леонид Петрович, я сейчас занятие веду.
Когда вы освободитесь? Ну, сразу после занятия.
А я не знаю, когда они закончатся. 18:50. 18:50 закончатся. Ага. Всё, спасибо. В 6:00 часов звони. Нет, в 7:00 часов. 18:50 – это ж 7:00 часов, а не 6:00.

**4. Практическая Демонстрация (На примере задачи студентки Екатерины)**

**4.1. Исходные Данные и Подготовка**

Так. Екатерина, направь мне ещё раз, потому что писем много приходит. Я на, ну, скажем так, вот это вот занятие, они для того и это время отведено, чтобы я разбирался с вашими этими приложениями.
Так, смотрим.
Ну так я же вам ответил. Я вам даже модель, по-моему, сделал.
Катерина, я же, по-моему, вам модель сделал. Я же всё ж посчитал вам там и отправил ответ. Сразу же причём, быстро.
Катерина, почему вы не отвечаете? Я вам сказал, я вам отправил уже.
Ну, тогда, возможно, я вас с кем-то путаю, потому что я кому-то эту задачу решал. Я её решал уже эту задачу.
Ну хорошо, посмотрим сейчас.
Вот вам экселевский файл. Тут он какой-то вообще не особо. Вроде как и не экселевский. А вот экселевский файл.
2.000 ВУЗов.
Вот. Ну что? Значит, дальше я предлагаю понять, ребята, что у нас является классификационной шкалой, и что описательными шкалами.
Екатерина, вы меня слышите, нет? Напишите в чате, слышите или нет, Екатерина?
Да, слышите. Ну, отлично. И видите, что я показываю, да?
Значит, смотрите, ребята, нам нужно определить рейтинг университета по его характеристикам. Это означает, что рейтинг университета будет у нас классификационной шкалой. У нас рейтинг вот где находится, на первой колонке. Мы его переставляем на третью колонку, а первую удаляем.
У нас получается первая колонка - название ВУЗа, а потом дальше идут названия, то есть рейтинг ВУЗа идёт.
Вот. И идёт название страны.
И всё, собственно говоря.
Смотрим, нет ли тут пустых мест где-нибудь.
Сейчас мы это быстро разберёмся.

**4.2. Настройка Параметров в Aidos**

Смотрите, у нас классификационной шкалой является вторая. А третья, четвёртая, пятая и и тринадцатая являются описательными шкалами.
Год, год мы не будем рассматривать, ребята, это не информативный показатель. Вот. У нас он особо не интересует.
Всё. Соответственно, значит, Екатерина, пришлите мне письмо, чтобы я мог вам в ответ послать результаты этого исследования.

Значит, мы вводим данные в систему 2 3 2 2 интерфейс. Указываем во второй шкале, это классификационная шкала, третья - тринадцатая - описательные шкалы. И задаём здесь разные интервалы.
Кать, вы послали мне письмо? Давайте.
Ребята, обратите внимание, система от 15 ноября. То есть дата системы 15 ноября.
Поэтому надо обновить систему, ребята. Как её обновить, помните, нет? Её обновить очень просто. Скачать патч маленький, 10 МБ, и развернуть его в папочке системы. При этом система не должна быть запущена. Файл исходных данных должен находиться по этому пути, который я вот сейчас показываю.
Значит, я советую, раз у нас конвертер данных, которые в системе Aidos не преобразовал этот файл DBF в Excel, то это значит, что нужно использовать онлайн-конвертер. Я рекомендую использовать онлайн-конвертер. Он и в системе там в хелпе рекомендуется. И сейчас я вам рекомендую Convertio. Convertio. Он наиболее хорошо работает из тех, что я вот видел пока что.

**4.3. Синтез Модели и Интерпретация**

Значит, преобразуется экселевский файл во внутренний стандарт данных. Мы видим, что у нас очень большая размерность модели по числу строк. Мы резко уменьшаем размерность. А рейтинг можем сделать побольше градаций.
Вот. Смотрим, насколько у нас представлены у нас эти градации. Вот у нас получилось, смотрите, по 220 наблюдений значений на один на одно значение рейтинга и по 800 наблюдений на описательные шкалы. А одна описательная шкала у нас одна, по-моему, пока что, она у нас получилась текстовая, не числовой. Это девятая шкала, Broad Impact.
Значит, это непорядок, ребята. Мы отсюда тогда выходим из этого режима и смотрим на на файл исходных данных. Что там не так? Broad Impact. Здесь очень много нулей, то есть очень много данных отсутствует. Мы эти данные, которые отсутствуют, заменим на нули. А в системе там задаётся, что нули считаем мы отсутствием данных.
Вот. Сейчас посмотрим, как это выглядит.
Значит, сейчас вы, ребята, должны все на своих компьютерах повторять то, что я делаю, чтобы вы могли научиться это делать всё. Если у вас возникают какие-то вопросы, вы должны их задавать.
Вот совсем другое дело. Видите, получилось 10 на 89. Это вполне такая нормальная, обозримая модель.
Созданы классификационные описательные шкалы и градации. Мы на них смотрим. Смотрим мы в режиме 2.1 на них. Это 10 значений рейтинга.
Смотрим на описательные шкалы и градации. У нас их 11. Все девять стран участвуют в рейтинге всемирном.
И смотрим обучающую выборку.
Вот. И также можем и другую форму посмотреть обучающей выборки. Режим 2.4, 2.3.1 и 2.4.
Вот теперь вам форма исходных данных посылаю. И заодно файл, откуда она получилась с помощью этого конвертера онлайн.
Вот. Теперь мы создаём модели. Модели мы создаём при вот таких параметрах. Это параметры, которые оставляют 10% наиболее достоверных результатов идентификации. Остальные все результаты игнорируются как шум, в общем-то, зашумлённые.
При этом нужно отметить, что 2.200 объектов распознаваемой выборки - это довольно приличное количество. Это всего столько-то представлено ВУЗов за всё время, за весь период. А период там составляет с 2012 года до 2015, за 3 года. Ну, то есть можно посчитать, сколько получится за год.
Пишет, что в 18:54 закончат. Эта операция синтеза моделей и проверки их на достоверность.
Вот. Заняло около 2 минут операция синтеза моделей и проверки их на достоверность.
Смотрим, какие у нас получились модели. Это мы делаем в режиме 5.5. Видим здесь модели статистические. И системно-когнитивные.
Вот сколько показывает статистически. Вот модель количество информации. Какие признаки несут информацию о том, какой будет рейтинг. Рейтинг от 1 до 10 колоночки, а строки - это в данном случае страны, а дальше там это значения каких-то свойств: национальный рейтинг, например, и качество обучения, качество факультетов, число публикаций, число ссылок. В общем, короче говоря, плата, вот сколько патентов, то есть много разных показателей. Вот эти все показатели, они каким-то образом непонятным связаны с рейтингом всемирным рейтингом. И я могу вам сказать, что не всегда очевидным образом они связаны. Бывает иногда непонятно, почему так. Вот. Ну, в общем, насколько я знаю, всё-таки оказывается, что связано.

Значит, АСКА-анализ всемирного рейтинга ВУЗов. Тему вот такую вот так формулируете. И смотрим мы на результаты этого анализа. У вас получается довольно высокий уровень достоверности моделей по F-критерию Квандта-Редсбергена – это 0,867, по его обобщённому варианту нечёткому мультиклассовому обобщению – 0,870, вот, 0,879. Сейчас мы посмотрим. Ну это можно сказать, что модель у вас получилась. Сейчас посмотрим, где картинки. Для этого она несколько странный вид имеет. Ну вот это ещё куда ни шло.
Ну я могу сказать, что то, что такой странный вид эта кривулька имеет, это может говорить о том, что у них там какие-то неформализуемые пути принятия решений существуют. Но если это перевести на русский язык, то это означает, что они от фонаря принимают решения или исходя из таких параметров, информация о которых не содержится в этих базах данных, которые размещены на Kaggle. Ну то есть они каким-то непонятным образом принимают решение о принадлежности ВУЗов к тем или иным категориям по рейтингу международному.
Что, в общем, вполне естественно.
Ну вот, эта форма о том, как признаки связаны по смыслу. Ну здесь эти наименования признаков лепятся, видите, одно на другое. Есть очень простой способ этого избежать. Для этого вот здесь вот в системе в параметрах задаём по вертикали больше пикселей. И перерисовываем без перерасчёта саму эту дендрограмму.
Вот. Одна немножко покарёженная, да? Мне не очень не нравится. Здесь есть ошибки построения самой дендрограммы. Верхняя часть - это всё разумно, а в нижней как бы есть не совсем корректные вещи. Вот сюда вот она должна бы была бы попасть. А давай попробуем сейчас другую модель. Достаточно быстро формируется, поэтому не проблема.
Вот, сейчас поприличнее. Сейчас она более корректная эта форма.
Сейчас посмотрим на неё в просмотрщике. Видите, все названия, всё прекрасно. Ничего не наползает. Значит, я вам посылаю эту дендрограмму и график межкластерных расстояний посылаю вам.
Вот. Теперь следующий момент интересный. Это определение того, какой признак, какой имеет смысл. А, давайте сначала ещё Россию посмотрим. У нас Россия есть. Что характерно для России?
Между прочим, неплохо характерно. Шесть ВУЗов российских попадает в международный рейтинг ВУЗов. Шесть ВУЗов попадает российских. Для России характерно характерна вторая десятка, четвёртая десятка, десятая десятка. И в меньшей степени характерна первая десятка, пятая и восьмая. То есть вообще так, в общем-то, уровень довольно высокий.
Вот, ну и теперь можно решать задачи. Берём, делаем модель INF3 самой достоверной. Вот. И с помощью этой модели решаем задачу определения рейтинга ВУЗа. В режиме 4.1.2. Заняло 7 секунд. Вот то, что я сейчас делаю, всё это описано в хелпе режима 1.3. Вот то, что я делаю один к одному.
И вы можете теперь решать задачи 4.1.3.1. Это результат решения задачи определения рейтинга ВУЗа Харвардского университета. Другие ВУЗы. То есть иногда она ошибается, но незначительно модель.
Вот это результат решения задачи идентификации, то есть определения рейтинга. То есть мы можем говорить о том, что по параметрам ВУЗа мы можем определить его рейтинг.
Первый рейтинг, что предполагает? Высокую оплату, много мало патентов, низкое качество, ну тут очень надо переводить. Что это всё значит? Лучше перевести.

**5. Требования к Отчетности и Публикации**

Вот, Екатерина, получается, что мы тебе почти все скриншоты записали. Ещё несколько штук сама сделаешь. Ну, если хочешь, могу послать саму модель. Сейчас давай минимизирую систему. То есть я её освободил от всего лишнего, того, что сейчас не нужно для этого приложения. Архивирую. И если ты развернёшь на диске C её, то у тебя всё вот это будет точно так же, как у меня на компьютере выглядеть и работать.
Вот. Всё тебе послал сейчас письмо со всеми этими скриншотами, со исходными данными. Там все параметры, всё есть. То есть можно брать описание вот это и по нему делать описание по шаблону, который там приведён. Здесь приведено у нас пять шаблонов на данный момент. Это пункт шесть. Здесь у нас есть шаблон, наиболее новый – это пятый шаблон. Вот его берите и прямо по нему делайте описание. Там везде ссылочки на вордовские файлы.
Вот. Потом вам надо обязательно зарегистрироваться в ResearchGate, РИНЦ, Science Index, заключить договор на размещение публикаций и разместить публикацию. И обязательно указать источник данных – это Kaggle, я там указал в письме. А метод – это система Aidos, и на неё тоже надо сослаться.

**6. Завершение**

Ну что, ребята, на этом наше занятие заканчивается. Какие у вас вопросы есть?
Ну вот Екатерина уже, я так понял, она написала спасибо. Надеюсь, теперь тебе будет легче сделать описание уже. Но всё-таки нужно установить систему на своём компьютере и всё это вот на своём компьютере делать. Скриншоты – Alt+Print Screen, появляются у вас скриншот в буфере обмена. После этого вы можете его вставить в статью.
Давайте, Виктория и Олег, какие у вас вопросы? У вас тема уже есть, да? Данные есть или нет? Вы уже проводите, создаёте модели, проверяете их на достоверность или нет?
Пока я нахожусь на стадии регистрации, просто не знаю, откуда корпоративную почту достать.
Ну её даёт ЦИТ, Центр информационных технологий. То есть надо к ним обращаться, чтобы они дали эту почту. Центр информационных технологий. Начальник центра Кремер Алексей Семёнович. Почту даёт ЦИТ, Центр информационных технологий. А описание вы сделали, нет?
Ну, пока нет, но вы сказали то, что модели разумная, утвердили её. В общем, хорошая модель получается.
Хорошо. Тогда надо делать описание, регистрироваться и размещать её. Но сначала прислать мне описание. И только после того, как я скажу, что нормально, и размещу в системе Aidos в облаке, после этого в ResearchGate и РИНЦ можно размещать. Это связано с тем, что я просто не хотел бы, чтобы низкого качества материалы размещались. Иногда очень небрежное форматирование, такое тяп-ляп называется, которое вызывает неприятное впечатление вообще, понимаете, о том, как человек относится к своему делу. Это ж публикация, это статья в международные, понимаете, это ваше лицо и в какой-то степени и моё лицо, потому что вы ж по моим технологиям это всё делаете и ссылаетесь на мои работы. И мне неприятно, если это будет работа низкого качества по оформлению. По содержанию я уже посмотрел, что она будет нормальная. Вот, поэтому вы мне её покажете, и я вам какие-то, может быть, сделаю замечания, исправления. Если не очень большие, то я сам внесу эти исправления. Ну я вам могу сказать, что вот пятый шаблон берите за основу, он так более-менее приличный, более-менее соответствует действительности, текущему моменту. Хотя тоже немножко устарел уже. Именно устарел в чём? Здесь 210 работ. Ну, в этом плане нормально всё. И он даже более-менее соответствует сейчас. Я его расширил несколько. Мне поразило, что запрещали студентам устанавливать систему Aidos дома, потому что это лицензионное программное обеспечение, и они не имеют права дома устанавливать. А я написал, здесь вот добавил, что это открытое программное обеспечение, которое кто угодно имеет право устанавливать где угодно. И что эти права им делегировал всем пользователям, желающим использовать систему, делегировал её разработчик и первичный правообладатель – это я.
Ну хорошо. Хорошо, что хорошо получается, что разумная модель. Вот описание вот здесь, шаблон описания.
У нас, по-моему, не так много времени осталось уже, во всех смыслах, и до сессии, и до конца занятия. Ещё полчаса до конца занятия. В эти полчаса нужно чем-то занять, ребята. Мы же не будем сидеть молча просто, правильно?
Ну тогда давайте, я сейчас буду, если так вот больше ни у кого вопросов нет по их работам, ну тогда я Екатерине буду дальше её модели считать дальше. Ты слышишь, Екатерина, да?
Значит, во-первых, возникает такой вопрос, какие у нас параметры наиболее характерны для ВУЗов, имеющих тот или иной рейтинг. Вот для наиболее продвинутых ВУЗов характерны вот такие значения параметров слева, а справа - те, которые не характерны.
Смотрим, как это выглядит в форме нейрона локального. Для ВУЗов с самым высоким рейтингом, как это выглядит? Ну здесь, я вам скажу честно, мало что понятно, потому что наползают друг на друга вот эти изображения рецепторов. Поэтому я их сейчас… Ну такая форма тоже, она имеет право на жизнь. Её можно показать и сказать, что вот наползает, поэтому нужно уменьшить число рецепторов. Из опыта я знаю, что это число рецепторов должно быть, ну, 32 вообще-то отлично получается видно всё. А 48 уже плохо видно.
Ну вот эти параметры, которые наибольшей силой обуславливают принадлежность ВУЗа первому рейтингу в мире. А это те, которые препятствуют этому.
И можем посмотреть на нейронную сеть, на один слой нейронной сети. Екатерина, формы, которые у вас должны в работе быть, в описании. Вы там увидите, где они там находятся, их место.
И очень интересная форма – когнитивные функции. Как влияют каждое значение параметра на рейтинг. Вот из этих показателей, которые там в рейтинге учитываются, всего лишь несколько играют роль существенную для того, чтобы оценить рейтинг ВУЗа. Остальные играют крайне слабую роль в этом. Вот эти формы в табличной форме. То есть в графической форме мы видим как бы сам вид картины, а в табличных формах мы видим уже наименования этих показателей, как они у нас расположены.
Национальный рейтинг находится в соответствии с международным. Качество обучения тоже. Ну у них что странно, у нас другая система оценки. У нас максимальная оценка – это хорошо, а минимальная – это плохо. А у них наоборот. Вот у них, допустим, самый высокий рейтинг – это первый из тысячи. Оценка самая высокая – это единица, а самая низкая – пятёрка там, ну так, грубо говоря.
Да. Пожалуй, здесь надо, конечно, не так немножко делать, а типа вот такого что-то делать. Это когда с экселем. Вот. Всё симпатичнее. И даже аналогично и здесь посимпатичнее можно сделать.
Ну что ж, ребята, есть ещё вопросы? Какие-то вот такие молчаливые, ничего не реагируете никак на мои вопросы. Всё, ребята. На этом наше занятие заканчивается. До свидания. Всего самого хорошего.
До свидания. До свидания, Георгий Александрович.