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

***Российская Федерация***

**176 Лабораторная работа № 9. По дисциплине "Инженерия знаний и интеллектуальные системы". 2020-11-02З**

**аголовок:** Инженерия знаний: Лаб. №9 - Работа с данными, обсуждение заданий и демонстрации

**Резюме текста:**

Видеозапись представляет собой онлайн-занятие (лабораторную работу №9) по дисциплине "Инженерия знаний и интеллектуальные системы" для группы БИ17-01, проведенное 2 ноября 2020 года.

1. **Введение и Организационные моменты:**
   * Занятие начинается с приветствий и шуток по поводу низкой явки студентов (изначально присутствует один студент Олег, позже присоединяются Виктория и Марина).
   * Преподаватель (Георгий Семёнович) объявляет тему занятия, дату и номер лабораторной работы. Упоминается необходимость отметить присутствующих.
   * Обсуждается предстоящая аттестация и важность выполнения индивидуальных заданий.
2. **Обсуждение Заданий и Административные Вопросы:**
   * Преподаватель просит старосту Марину напомнить, на чем остановились на прошлом занятии, но она испытывает технические трудности с микрофоном.
   * Основная задача студентов – продолжение работы над собственными приложениями/заданиями.
   * Возникает вопрос о сдаче ведомости текущей успеваемости. Марина объясняет, что пропустила напоминание из-за технических проблем с телефоном, но обещает подготовить и прислать ведомость.
3. **Работа с Данными (Пример Кристины):**
   * Студентка Кристина прислала свое задание. Преподаватель анализирует его данные.
   * Обсуждается размер датасета Кристины (4000 изображений, 1028 МБ), который признается слишком большим для обработки в облаке или текущими средствами.
   * Кристина предлагает альтернативный датасет (144 МБ), который преподаватель считает более подходящим.
4. **Демонстрация Преподавателя:**
   * **Концепции Инженерии Знаний:** Преподаватель кратко касается теоретических аспектов дисциплины, упоминая инженерию знаний и интеллектуальные системы.
   * **Обработка Изображений (Розы):** Демонстрируется пример работы с небольшим набором изображений роз, включая изменение их размера (до 200 пикселей) как этап предобработки.
   * **Конвертация Данных (CSV в DBF):** Показывается процесс конвертации большого CSV-файла (данные о транзакциях для задачи обнаружения мошенничества) в формат DBF с использованием специализированного конвертера. Отмечаются потенциальные проблемы с форматированием данных (точки вместо запятых) и большой объем данных (284 тыс. записей).
   * **Анализ Данных в AIDOS:** Преподаватель настраивает систему AIDOS для анализа данных о транзакциях, определяет входные и выходные (классификационные) признаки. Отмечается неравномерное распределение данных и необходимость фильтрации недостоверных результатов.
5. **Технические Трудности:**
   * В ходе демонстрации возникают проблемы с конвертацией данных из-за их форматирования и большого размера, а также с работой некоторых инструментов (онлайн-конвертеры, системный конвертер AIDOS).
6. **Завершение:**
   * Преподаватель напоминает о необходимости сдачи работ в рамках аттестации.
   * Занятие завершается прощанием.

**Детальная расшифровка текста:**

**1. Введение и приветствие**

* Спасибо. Желаю наилучшего здоровья на веки вечные, долгих лет жизни.
* Спасибо большое.

**2. Обсуждение посещаемости и административных вопросов**

* **2.1. Комментарии о количестве студентов**
  + А это вот у нас Олег, это что? Это у нас аншлаг, да, Олег?
  + Совершенно верно.
  + У нас профессора, а на одного студента – это, по-моему, перебор какой-то. Ну было бы хотя бы два, там, я понимаю, ещё как-то.
* **2.2. Ожидание других студентов**
  + О, Виктория пришла.
  + Сейчас их много будет, они сейчас подтянутся.
  + Понятно, понятно.
  + Да. Вот так вот.
* **2.3. Формальное начало занятия**
  + Ну что, давайте тогда начнём занятие.
  + Сегодня 2 ноября 2020 года.
  + Шестая пара. 17:20–18:50.
  + Лабораторная работа номер девять.
  + У группы БИ17-01 по дисциплине Инженерия знаний и интеллектуальные системы.
  + Вот такие дела у нас. Сейчас будем отмечаться везде сначала.
* **2.4. Проверка посещаемости и административные заметки**
  + Лёсь, не очень долго.
  + Вот. Отмечаться будем.
  + Вот. И у нас тут уже такая прям компания.
* **2.5. Студент из другой группы**
  + И даже у нас пришёл Андрей из группы ЭТ20-02.
  + О, ушёл.
  + Заблудился.
  + Потому что из другой группы. Видели, да, Георгий Семёнович? Появлялся.
  + Что скажешь?
  + А?
  + Он вернётся.
  + Нет, так это ж из другой группы.
  + Я понимаю, но, наверное, там ему неинтересно.
  + Тут здесь интереснее, да, наверное?
  + Ну ладно, не знаю. Вот Марина. Здравствуй, Марина.

**3. Обсуждение Заданий и Административные Вопросы (Продолжение)**

* **3.1. Переход к теме занятия**
  + Ну, давай уже это самое, ближе к нашему, нашей теме.
* **3.2. Запрос к старосте о предыдущем занятии**
  + Марина, давай, напоминаю, ты же у нас староста. Напоминаю, на чём мы остановились на прошлом занятии.
  + (Пауза, Марина не отвечает)
  + Не можешь вспомнить, да? Или даже не слышишь, что я говорю, наверное.
  + (Длинная пауза)
* **3.3. Уточнение текущей задачи**
  + Ну, у нас продолжается работа над собственными приложениями. Это у нас уже до самой сессии будет, потому что сейчас уже аттестация начинается. Сегодня первый день, когда аттестация уже идёт.
* **3.4. Обсуждение ведомости успеваемости с Мариной**
  + Кстати, Марина прислала-то, да, эту ведомость?
  + Сейчас посмотрим.
  + Марина пока не прислала ведомость.
  + Марин? Так, смотрим. Мы уже по темам своим работам говорили. Угу. А ты мне должна прислать, Марина, ещё ведомость о текущей успеваемости студентов.
  + (Пауза)
  + Здравствуйте, извините, я микрофон только подключила. А я по ведомости поняла, по текущей успеваемости. Это точно? Просто может, вчера на прошлом паре минут за 10 у меня телефон выключился, и я, наверное, прослушала.
  + А я ж тебе ещё письмо написал.
  + А, я не видела.
  + Странно.
  + Странно, странно.
  + (Пауза)
  + Что, нашла?
  + Нашла, у меня просто уведомления на почту не приходят, я не видела, что пришло.
  + Ясно, ясно. Ну хорошо, ну тогда когда сделаешь, пришли пришли. Может там сегодня вечером, может завтра.
  + Ну я сейчас постараюсь сделать её.
  + На почту пришли мне.
  + Угу.
  + Давайте.

**4. Работа с Данными (Пример Кристины)**

* **4.1. Анализ данных Кристины**
  + (Обращается к чату или другому источнику) Кристина, Кристина, посмотрим сейчас.
  + Нет, ещё твою не успел посмотреть. Посмотрю после занятия, Кристина. Или, вообще-то, вы знаете, что можно даже сейчас на занятии есть смысл. Давай сейчас на занятии посмотрю. Тогда я сейчас сделаю экран, чтобы вы видели.
  + И посмотрим.
  + Да, это очень серьёзно. Четыре 4.000 изображений – это круто.
  + Значит, Кристина. Значит, 1028 МБ – это слишком много. Это точно, 100%.
  + Ну, меньше гигабайт на таких датасетах вообще нет.
  + Да ладно, полно.
  + Ну ладно.
  + А 144 подойдёт?
  + Чего 144?
  + Ещё, ещё одну отправлю, посмотрите, пожалуйста, её объём.
  + Давай.

**5. Демонстрация Преподавателя**

* **5.1. Введение в Демонстрацию (Концепции Инженерии Знаний)**
  + Ребята, ещё давайте я вам расскажу, э-э, представление определённое дам вам. Ну это на лекции вообще-то э-э надо делать, но сейчас тоже уместно на практическом занятии. Вполне на лабораторной работе. Почему? Потому что у нас дисциплина называется Инженерия знаний и интеллектуальные системы. Поэтому этот вопрос такой, один из центральных. То есть я вам объяснял, что такое интеллектуальные системы, какие виды систем и так далее, и так далее. Вот. Сейчас посмотрим.
* **5.2. Демонстрация Обработки Изображений (Розы)**
  + Вот розы. Сейчас мы их откроем. То есть я взял небольшое количество. И смотрим, смотрим. Вот. Э-э, мы сейчас сделаем этим всем изображениям размер 240. А нет, 200, 200 пикселей по ширине. Resize 200 пикселей. Вот, все изображения. И, в общем, можем попробовать решить задачу на небольшом объёме вот этих вот изображений.
  + (Пауза, возможно, идёт обработка)
* **5.3. Демонстрация Конвертации Данных (CSV в DBF)**
  + Сейчас попробуем с транзакциями.
  + (Пауза)
  + Что-то не то с изображениями.
  + Ну ладно, давай попробуем с этим.
  + (Пауза)
  + Что-то я этот CSV файл записал. Ты видела, да, Кристина? И запускаю сразу конвертер. Вот, он сообщает, что там в папочке должен быть CSV файл с именем input data. Надеюсь, что он преобразует этот файл. Этот конвертер. А, здесь ещё к тому же несколько версий системы запущено. Странно, она обычно это обнаруживает и сообщает. Повторный запуск. А сейчас не обнаружила.
  + Вы видите, да, экран, ребята?
  + Да, видим.
  + Вот. Вот сейчас э-э конвертер работает, преобразование вот этого CSV файла в DBF файл. Пишет, что 4 минуты будет работать. 240.000 записей. Это такое солидное количество записей. Немножко быстрее идёт процесс, чем в начале оценивалось системой. Ну, уточняет постоянно прогноз, с учётом текущей загрузки компьютера, вот так я бы сказал. Ну пишет, полторы минуты осталось. То есть получается, что где-то около 3 минут должно это всё продолжаться. Но дело в том, что здесь два процесса разных идёт. Сначала файл просматривается целиком и определяются размеры и типы данных всех полей. А для числовых полей ещё число знаков после запятой. А потом создаётся база данных с такими параметрами, чтобы эти данные с CSV файла э-э поместились вот в этот DBF файл. Вот. Структура его уже определена. То есть известны типы данных и размеры полей и число десятичных знаков. Создаётся база соответствующая, и туда уже данные заносятся. То есть это разные операции, они идут с разной скоростью. Ну, так, в принципе, вот она прогнозирует, что в 18:40 закончит.
  + (Пауза, идёт процесс конвертации)
  + Всё, преобразование завершено. Теперь смотрим на тот файл, который получился в результате. Получился вот этот вот DBF файл, который даже не поместится в этом Экселе 2003.
* **5.4. Анализ Данных о Транзакциях**
  + И сейчас давай посмотрим на описание задачи.
  + Значит, функция время содержит секунды, функция сумма – суммы транзакций. Функция может использоваться для обучения с учётом затрат. Функция класс – это переменная ответа, которая принимает значение ноль в случае... один в случае мошенничества и ноль в противном случае. Вот. Довольно интересная задача.
  + Вот. Значит, что мы делаем? Посмотрим сейчас. Сумма. Что это у нас за колонки? Посмотрим сейчас. Я запишу такой файл, но мы будем не его использовать, а DBF файл будем использовать. И посмотрим. Значит, у нас э-э колонки со второй по... двадцать девятую у нас это параметры. А с тридцатой по тридцать первую – это классы: сумма и сам факт, совершено, не совершено это вот нарушение.
* **5.5. Настройка Системы AIDOS для Анализа**
  + И смотрим, ребята, значит, берём э-э создаём... входим в режим 2 3 2 2 и создаём модель. Значит, указываем DBF файл, потому что система, этот конвертер, он создал DBF файл. То есть мы будем его использовать для входа данных, ввода данных. Размером 174 МБ. И у нас э-э колонки классификационные. Колонки. Давай так сделаем. Со второй по тридцатую. Для нумерации нужно. Чтобы не запутаться. Нумерация. Вот. Со второй по двадцать девятую... по тридцатую – это описательные шкалы. А с тридцать первой по тридцать первую... классификационные шкалы. Вот. Значит, мы здесь задаём. Классификационные шкалы с тридцать первой по тридцать первую – это сам факт нарушения. А со второй по тридцатую – это у нас различные там параметры транзакций, которые они сказали, что они нам не раскроют их.
* **5.6. Обработка и Анализ Данных (Продолжение)**
  + Вот. И сейчас этот файл будет вводиться в систему и формироваться... сформировалась уже классификационно-описательные шкалы и градации. Сейчас формирует обучающую выборку. 10.000 записей, 13 колонок, 13 этих шкал. То есть, я бы сказал так, модель не самая большая, но довольно большая. Вот. Тут мы видим категории и рейтинги. Описательных шкал довольно побольше. Вот это вот у нас эти значения. И обучающая выборка. И синтез и верификация. Здесь бы я вот так бы сделал бы.
  + Синтез модели, смотрите, на графическом процессоре застряла. Видите? 10.000 примеров и размерность большая матрицы: 683 на 9.000. Время идёт. Да. Артур, я сделаю, пришлю тебе. Виктория, давай.
  + Ну, короче говоря, вот эта задача, она тяжёлая. Вот распознавание 10.000 объектов на шестистах классах, на десяти... на 9.000 признаках. Признаков. Даже для графического процессора, видите, он встал. Вот. Ну, обычно там мгновенно проскакивает всё. Вот. Но сейчас вот прогноз исполнения, вы увидите, что тут не 9 минут, а гораздо больше будет расчёт идти.

**6. Завершение и Дополнительные Комментарии**

* **6.1. Рекомендации по выбору задач**
  + Ну, в общем, эта задача, я не рекомендую её решать. Её интересно решать всё-таки задачи, но для этого надо иметь время. А у нас его нет этого времени. Ну, так как вы уже, когда вы там попробуете маленькие задачи порешаете, тогда потом можно браться вот за такие задачи, типа кредитных этих карт. Задача интересна вот эта по кредитным картам. Это вот я не очень понимаю, но, в принципе, можно выявить зависимости, попытаться. Ну, поскольку там данные очень специфические, такие, видели, там в основном тексты какие-то странные, то это получится такое, не совсем... проще, когда там числа, тогда хоть как-то понятно, о чём речь идёт.
* **6.2. Итоги занятия и прощание**
  + Вот. Теперь я так понимаю, что у нас конец занятия сейчас, да, ребята? Правильно? У вас ещё есть занятия?
  + Занятие есть ещё?
  + У меня есть, восьмая пара ещё.
  + Надо отменять.
  + Почему?
  + Ну так вы ж понимаете, почему. Уже надо проводить праздник.
  + Праздники нам только снятся.
  + Достархан накрыт?
  + Ага. Ну, в общем, я это, э-э сейчас скажу. Вот Артур, я тебе э-э посмотрю, что там ты прислал, и пришлю тебе ответ. Хорошо?
  + Да, а можно, ну, я кое-что поясню. Там, а-а, в датабее, а, в датасете, а-а, ну, который сайт я скинул, там шесть Excel-таблиц. Я пока спарсил только одну и вам показал, да? Но у меня вот тема, которую я предложил, там, ну, по датам ещё есть, да, и как бы, ну, если что, можете зайти посмотреть. Там шесть разных. По датам, динамика, всё там есть, в общем. Ну всё брать, как вариант вам, ну, прислал.
  + Понятно, понятно. Ну лучше брать попроще что-то. Я не вижу там необходимости всё это объединять и глубоко анализировать. Сейчас у нас такой задачи не стоит. Вот можно взять просто какую-то таблицу, ну ту, которая вот как обучающая выборка, можно её рассматривать. И вот её попробовать реализовать.
  + Ну всё, хорошо.
  + Вот. То есть, ну ты видишь, что там есть задачи такие, что вообще крутые. Но они требуют времени. Вот даже вот сейчас вот я запустил там на счёт её, и она встала колом, видишь? На графическом процессоре, можешь себе представить, что это такое? Если что, там я выберу датасет с небольшими количествами данных.
  + Ну да, вот, то есть видел, сейчас мы вот э-э с Кристиной там всё несколько задач посмотрели. Вот я думал, что вы примерно так это смотрите их, в общем-то, если честно. Вот я вам показываю на занятии, как это можно делать. Посмотрели там, ну как, э-э, попытались укусить, а оно слишком большое, не кусается. Наметишь что-то такое, его можно укусить, понимаешь, схватить там. Вот. А такое большое, ну... Ну это, конечно, очень престижно, но это уже потом. Если вы почувствуете к этому вкус, будет у вас получаться, ну тогда можно такие задачи. Если будет время для этого, понимаете, и будет какой-то стимул, интерес к этому.
  + Ну ладно, ребята, на этом мы занятие заканчиваем. Наверное, вам было полезно тоже посмотреть, что мы там поковырялись, попробовали, что-то там получилось, что-то не получилось. Это очень полезно, потому что, в принципе, вы вот так вот можете выбирать задачи, смотреть, можно ли их решить. И в будущем мы примерно так же можете выбрать себе задачи, описать их. У вас появятся публикации. То, как вот я вам это всё объяснял, рассказывал. Что, в общем-то, уже какой-то продукт реальный от того, что от ваших усилий что-то получится из этих.
  + Ну а сейчас пока что, ребята, до свидания. Тех писем, которые ко мне пришли, я их обработаю, конечно, пришлю ответы. Всего самого хорошего вам, ребята.
  + До свидания.
  + До свидания.
  + До свидания.
  + Ну а праздник всё ж стол готов, Евгений?
  + Да, стол, стол там, мелкая палка.
  + Как казахи говорят, дастархан надо накрывать.
  + Ну у нас, у нас был перерыв между парами. Во время перерыва у меня был там стол.
  + Да это ж невозможно за такое короткое время.
  + Да. Да, конечно. Ну, всего самого хорошего. До свидания.
  + До свидания, всего доброго. Ещё раз поздравляю.
  + Спасибо большое.