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

***Российская Федерация***

**186 Лабораторная работа №8. По дисциплине "Интеллектуальные информационные системы и технологии". Приложения Эйдос 2020-11-06**

Заголовок: Eidos System Lab: Демонстрация навыков через анализ данных (грибы, диабет) и решение проблем с посещаемостью

Резюме текста:

1. Введение и посещаемость:

Сессия начинается с приветствия профессорами (Луценко и Аршинов) присутствующих студентов (Георгий, Елизавета, Дима и др.). Отмечается крайне низкая явка (8 из 17/13). Обсуждается разница между сложностью посещения очных занятий (вечер, ехать далеко) и легкостью подключения онлайн. Подчеркивается важность присутствия, так как идет запись и возможны проверки из учебного управления. Старосте (Никита) поручается сообщить остальным о необходимости явки.

2. Цель лабораторной работы №8:

Занятие посвящено лабораторной работе №8 для группы ИТ18-21 по дисциплине "Интеллектуальные информационные системы и технологии". Цель – продемонстрировать приобретенные компетенции (знания, умения, навыки) через создание приложения в системе Eidos. Студенты должны выбрать тему (набор данных), обработать данные и построить модель. Это подтвердит их освоение учебной программы.

3. Работа с системой Eidos и данными:

Профессор объясняет процесс работы:

Выбор данных: Использовать небольшие наборы данных (не гигабайты/мегабайты) с сайтов UCI или Kaggle.

Преобразование данных: Данные часто представлены в символьном виде (буквы, слова). Их нужно преобразовать в числовой или понятный системе формат. Показан пример импорта CSV в Excel и дальнейшей обработки. Демонстрируется использование конвертера CSV-DBF в Eidos и ручное преобразование данных в Excel (замена букв/слов на числа или стандартизированные строки).

Создание модели: Ввод данных в систему Eidos, указание классификационных и описательных признаков, запуск синтеза и верификации модели. Установка параметра тестовой выборки (например, 30%).

4. Примеры задач:

Классификация грибов: На примере набора данных о грибах (съедобный/ядовитый) демонстрируется полный цикл: загрузка данных, преобразование символьных признаков (форма шляпки, цвет и т.д.) в понятные системе строки, создание и верификация модели. Модель показывает высокую достоверность (0.98).

Диагностика диабета: Аналогично разбирается задача диагностики диабета на основе медицинских показателей у индейцев Пима. Модель также показывает хорошую достоверность.

5. Взаимодействие со студентами и выбор тем:

Профессор просит студентов отметить в чате, кто уже работает над темой (+), а кто нет (-). Обсуждается тема, выбранная Никитой (Мировой рейтинг университетов). Выясняется, что ссылка на похожую работу не работает (ошибка 404), но сама тема подходит. Профессор подчеркивает, что данные нужно будет обработать. Студентке Виктории Комардиной предлагается взять тему диагностики диабета, так как модель уже успешно построена. Студентке Полине предлагается взять тему классификации грибов.

6. Заключение и предупреждения:

Подчеркивается, что будущее за студентами, но оно зависит от их усилий. Исследования показывают, что зарплаты у выпускников дистанционного обучения могут быть ниже. Профессора готовы помогать, но студенты должны проявлять активность. Невыполнение заданий приведет к низким оценкам.

Детальная расшифровка текста:

Раздел 1: Введение и обсуждение посещаемости

Подраздел 1.1: Приветствие и проверка присутствующих

Ну, надо ещё раз поздороваться, потому что они не здороваются. А это идёт запись уже.

Ну мы, главное, знаете, как в Карнавальной ночи, помните?

Что-то помню там, но что-то не помню.

Здравствуйте!

Да.

Здравствуйте.

Здравствуйте, Георгий Александрович, здравствуйте, Елизавета, Дима.

Здравствуйте, все здравствуйте.

И другие студенты, которые...

Здравствуйте.

И другие студенты, которых нету.

Да.

Подраздел 1.2: Проблема низкой явки

Утром, может быть, мало.

Ну, утром они ещё спят, а вечером уже спят.

Вообще их должно быть 13.

Они должны быть все.

Да. Вообще-то их тут 17 должно быть, но в моей подгруппе 13.

Но они должны быть все.

Ну да. Что-то как-то я бы не сказал, что они все здесь. Не верится как-то.

В моей подгруппе только я.

Так, ладно. Смотрим. Кто у вас староста, ребят? После Елизаветы и Димы, кто у вас староста?

Ну, Никитин.

Никитин. Ну вы, пожалуйста, как-нибудь ему сообщите, чтобы он сюда загнал народа побольше. Потому что сейчас придёт проверка учебного управления, и все получат тык хороший.

Подраздел 1.3: Сравнение очного и дистанционного формата

Ну я понимаю, там в аудиторию трудно было прийти, там ехать надо, вечер уже, темно, куда-то ехать, там в какую-то аудиторию, там... э-э... Да? Там уже пятница вечером. Я понимаю.

Но на телефончике-то при дистанционном обучении можно было бы ссылочку кликнуть один раз и дальше заниматься своими делами, в принципе. Что, это даже трудно уже, что ли? Тогда я не знаю, что ещё для вас сделать. Прямо в постель, так сказать, вам что ли кидать этот какой-нибудь там гаджет, чтобы он вам прямо читал вам, всё рассказывал. А вы там лежали себе там, пили кофе. Не знаю там.

Может, в будущем так и будет, да, Георгий Александрович?

Ну они может вместо кофе...

Ладно. Ладно, ребята.

Раздел 2: Контекст и цель лабораторной работы

Подраздел 2.1: Важность задачи для студентов

Значит, э-э... Вам... вам всё простительно, ребята, потому что за вами всё равно будущее, понимаете? Куда ни как ни кинь, всё равно за вами будущее. Просто оно будет разным, в зависимости от того, будете вы учиться или не будете. У вас, и у всей страны, и у всех вообще разное оно будет.

Подраздел 2.2: Формальное объявление темы и цели занятия

Ладно. Значит, теперь давайте ближе к нашим баранам. У нас сегодняшнее занятие... э-э, значит, я уже сказал, да, про него? Или не сказал?

Георгий Александрович, я сказал, что занятия ведут профессор Луценко? Нет, по-моему, ж не говорил.

Я думаю, не лишним будет сказать.

Это надо сказать, потому что мы там похихикали, а запись идёт, и непонятно, о чём запись.

А запись вот о чём: шестое, сегодня шестое ноября 2020 года, пятница, шестая пара, которая идёт 17:20-18:50. Лабораторная работа номер восемь у группы ИТ-1821 по дисциплине "Интеллектуальные информационные системы и технологии".

Занятие ведут профессор Луценко Евгений Венеминович и профессор Аршинов Георгий Александрович.

Вот, так вот.

Подраздел 2.3: Связь с компетенциями и учебным планом

Теперь тема занятия нашего сегодняшнего - это ваши Eidos-приложения, которые подтверждают, что вы приобрели во время обучения планируемые рабочей программой, фондами оценочных средств, э-э, учебным рабочим планом компетенции. Компетенции в учебном рабочем плане – первоисточник, потом они попадают в рабочие программы и проверяются в фондах оценочных средств.

То есть вы должны приобрести, ребята, во время обучения знания, умения и навыки. Знания вы на лекциях приобретали. Умения и навыки – на лабораторных работах и на практических занятиях. Сейчас у нас как раз лабораторная работа номер восемь. Вот. На которой мы должны освоить определённые компетенции, которые вы должны подтвердить тем, что выполнили... выполнили задание. Причём это задание, оно не просто задание, а задание с инструкцией по выполнению.

Подраздел 2.4: Роль преподавателей и ожидаемая активность студентов

И мы с вами сидим здесь вот с Георгием Александровичем для того, чтобы вам ещё и объяснять к тому же, как это задание выполняется. Как его выполнять. Вот. То есть вообще почти что кофе в постель уже, понимаете? Полный сервис. Вот. Полный, полный сервис.

Вот это задание. Вы должны к нему отнестись серьёзно, к этому заданию. Это задание реальное на проверку ваших знаний, умений, навыков и приобретённых компетенций.

Раздел 3: Проверка прогресса студентов и выбор тем

Подраздел 3.1: Запрос обратной связи о работе над темами

Теперь, значит, пожалуйста, мне напомните, ребята. Мне необходима обратная связь определённая. Значит, э-э, напомните мне, кто из вас, э-э, уже разрабатывает какие-то темы? Пожалуйста, в чате отметьте плюсиками, кто разрабатывает, а минусиком – те, кто ещё не разрабатывает. Мы ждём. Давайте быстренько.

Так. Значит, должно быть столько плюсиков и минусиков, сколько вас присутствует сейчас. Вот, присутствует вас восемь человек. Поэтому, значит, должно быть восемь плюсиков и минусиков. Или сообщений ваших. Раз, два, три, четыре, пять, шесть, семь. И одного нету сообщения ещё, не хватает. О!

Ну давайте думать. Для этого мы здесь и собрались, для того, чтобы думать.

Подраздел 3.2: Источники данных и требования к ним

Вот у нас есть, э-э, два источника данных основных – это сайт UCI и сайт Kaggle. Значит, обращаю ваше внимание: исходные данные должны быть небольшой размерности. Не гигабайты ни в коем случае, и даже не мегабайты. То есть если вы сделаете там, скажем, 1-2 МБ, ещё можно обработать. Но если 10, то уже будет очень проблематично. А больше 10 МБ вообще стираются на FTP-сервере информация автоматически. Поэтому это вообще нельзя.

Подраздел 3.3: Призыв к выбору темы и обработке данных

Вот. Ну теперь давайте смотреть темы и выбирать темы для вас.

То есть, поди, никто ещё не делает темы. Ребят, вы думаете, что это что-то такое сверх эти темы, что ли? Это не сверх, это просто вы подтверждаете, что вы учились, понимаете? Мы же не были у вас на лабораторных работах рядом с вами с Георгием Александровичем и не видели, вы делаете вы что-нибудь на них или нет.

Вот возьмите сейчас, пожалуйста, мне в чате напишите: мы там или там делали на лабораторных работах то, что я показывал на экране на своём компьютере, вы это делали. Или мы этого не делали. Вот давайте, напишите быстренько. Чтобы я хоть знал, вообще вы лабораторные проходили или нет? Может вы там сидели, смотрели телевизор в это время?

Конечно, делали. Когда была возможность. А возможность когда была, Никита? Часто она была такая возможность или редко? М-м. Угу. Понятно. Не всегда вы были за компьютерами, не всегда у вас был интернет.

Раздел 4: Важность обучения и последствия дистанционного формата

Подраздел 4.1: Исследование о влиянии дистанционного обучения на зарплату

Ребят, США провели исследование, э-э, в котором постарались определить, э-э, как будут, какая будет зарплата у тех, кто учился дистанционно. У них получилось, ребята, совершенно сногсшибательная информация, на мой взгляд. У них получилось, что у тех, кто учился дистанционно, зарплата будет 12-15% ниже, чем у тех, кто учился очно.

Раздел 5: Демонстрация работы с данными в Eidos (Примеры)

Подраздел 5.1: Пример 1 - Классификация грибов

Значит, я смотрю темы, ребята, какие можно было бы выбрать.

Смотрим, что там у нас скачалось. 1 МБ.

Сейчас, ребята, я буду делать то, что вы должны были бы сделать сами уже. Потому что мы с вами всё это проходили, изучали, рассматривали. И вы должны были бы это сами всё сделать. А сейчас вот смотрите, пожалуйста, что я делаю, и готовьтесь к тому, что вы должны будете всё-таки сами найти себе тему. Сейчас, возможно, мы для кого-то найдём тему.

Классификация грибов. Попробуем. У нас тут два гриба есть: бледная поганка и не бледная. Ну, тут вообще не поймёшь. Что оно такое. Ладно. Попробуем.

Ребят, э-э, пожалуйста, обновите систему. Если у кого-то возникают вопросы о том, как это сделать, спросите. Я вам несколько раз это говорил, как это делать.

Значит, смотрите внимательно, ребята. Здесь вот в системе Eidos появился э-э CSV-DBF конвертер. Режим 2 3 2 10. Который и не хочет работать ещё к тому же. Вот. Смотрим, что он там преобразовал нам.

Что ж ему не понравилось? Ладно. Тогда давайте так сделаем. Значит, запускаем просто Excel. Попробуем в Экселе это сделать. Данные из текста. Вот. И указываем источник данных. CSV файл. Видим, что не очень-то она там их преобразовала эти данные. Указываем разделитель – знак табуляции. Получилось гораздо лучше. Вот. Поворачиваем это всё вертикально, потому что в ячейках одноразрядное значение. И делаем поуже.

Теперь смотрите, ребята, у нас тут класс грибов прямо он указан как класс, видите? Сейчас его и будем идентифицировать. Этот класс. Просто протягиваем номер. Здесь оказывается довольно много. 8.000 примеров. Отлично. Вот. И сохраняем как Excel. В папочку системы. С именем indata.

Читаем, ребята, колонки. Значит, вторая колонка – классификационная, с третьей по двадцать четвёртую – описательная. 2 3 24 описательная. Вот.

Теперь в системе самой уже пытаемся ввести информацию. Для этого указываем новый тип Экселя, со второй по вторую колоночку. Вот. И с третьей по двадцать четвёртую. И, собственно говоря, и всё. Значит, смотрим, как вводится информация, что у нас получается в результате.

Получается у нас две колоночки: ядовитые, не ядовитые. И 117 строчек. То есть признаков 117.

Вот, 46 секунд вводилась информация. И теперь запускаем синтез и верификацию модели. Здесь вот задаём 30, ребята, вместо 100%, задаём 30.

Вот. И смотрим, самое главное. Ну что, сразу видно, что модель очень хорошая получилась. Достоверность очень высокая. Даже, это мягко сказать, великолепная получилась модель. Вот смотрите, сумма знаний, у неё э-э даже F-критерий Вальд-Ресбергена 0,925. Ну, ребята, это это не просто отлично, это великолепно, понимаете, то есть просто великолепно.

Смотрим теперь на распределение. Смотрите, ребята, э-э истинных решений положительных всегда значительно больше, чем отрицательных. Всегда. А отрицательные решения есть, есть э-э случаи, когда ложных решений больше, но это только при очень низких уровнях сходства. При уровне... вернее, уровне различия. При уровне различия выше 35% ложных решений нет отрицательных. А положительные решения вообще практически все истинные, кроме тех, кто очень с низким уровнем сходства. То есть эта задача решилась очень хорошо.

Подраздел 5.2: Предложение темы студентам (Грибы)

И я задаю вам такой вот вопрос, ребята, риторический: кто берёт эту задачу? Кто будет её описывать? То есть сейчас я нашёл в Кагле хорошую задачу, которую можно реализовать.

Теперь смотрим, что здесь у нас про описание этой задачи. Здесь должны быть описания вот этих вот обозначений. Вот смотрите, ребят. Кто будет эту задачу делать, скажите мне, во-первых? Кто-нибудь будет её брать? То есть я нашёл вам хорошую модель, которая точно получается хорошо, работает.

Вот смотрите, здесь, видите, есть расшифровка вот этих вот буковок, которые там есть. Съедобный Е или И, ядовитый П. Раструб Б, конический С, выпуклый Х, плоский F, видите, написано. Вот нужно в этой таблице всё заменить. Синяки Т, нет Ж. То есть нужно все вот эти вот э-э элементы в этой таблице исходных данных заменить.

П – съедобные, то есть или И – съедобные, а Е – несъедобные. Вот. Колпачок бывает, смотрите, какой: Б – это раструб, С – конический, Х – выпуклый. Вот я беру и заменяю раструб. То есть Б заменяю на раструб. Колпачок. Ctrl+H Б раструб. Потому что все эти данные, они не рассчитаны на такие системы, как система Eidos, которая может обрабатывать и числовую, и текстовую информацию. Понимаете, да? Очень удобно для этого. Значит, теперь смотрим конический колпачок у неё. Конический. Это у нас С. Ctrl+H С. Надо напишешь письмо, скажи мне там как там микрофоном. Заменить конический. Вот. Следующий вариант – выпуклый. Выпуклый. Это Х. Выпуклый. Вот. Нужно ещё сделать. Тогда будет вообще классная, всё это будет выглядеть.

Подраздел 5.3: Предложение темы студентам (Диабет) и назначение

Ядовитый. Нужно ещё сделать. Тогда будет вообще классная, всё это будет выглядеть.

Евгений Венеминович. Да. Там Леонтьева согласилась помочь с грибами. А, уговорил? Ну то пряником, да?

Так, Никита, ты... То есть ты эту тему возьмёшь, да, Никит?

Да, да, я возьму. Сейчас я найду на этом сайте.

А, ну так я что её искать? Она найдена. Сейчас я вам ссылочку кину, да и всё.

А, ну давайте тогда легче.

Значит, э-э, тогда, Никита, вот что: пиши мне письмо на почту, на мою. Для чего? Чтобы я тебе прислал эту таблицу. Я ж уже её преобразовал и немножко над ней поработал. А тебе надо будет это продолжить и потом уже делать модель. Смотри, что я делаю. То есть я, видишь, в этой графе класс я заменил э-э вот эти буковки Е и И, так сказать, и П, которые там были. Я их заменил на русские слова: ядовитый, съедобный класс.

Так, теперь, значит, что мы делаем? Записываем это в виде экселевского файла. Вот. Ну здесь всё очевидно как бы, всё понятно. Поэтому даже не требуется каких-то особых пояснений. Значит, сейчас мы просто решим эту задачу и всё. Посмотрим, получается или нет. Надо посмотреть, что из них означает 0,1. Вот. Посмотрим. 0,1. Значит, если ноль, э-э... возраст, родословная, индекс массы, инсулин... Не не удостоились даже написать переменные какие. Как называются. Ну ладно. Мы и так, в общем, понятно. Ну ноль – это нет, а да – есть. То есть один – есть. Давай так и напишем. Нет. Есть. Заменяем. Заменяем. Вот. И нам этого достаточно.

Вот. Ну давай введём ещё раз. Тут там ещё есть параметры ввода. Ну там ничего ты не задашь, поэтому это роли не играет. Там можно изменить только в том случае, когда есть числовые поля. Ну у тебя вот так вот получится. Вот. Ладно, тогда мы не будем это делать. Вот, вот с такими параметрами ввода делаешь. Вот этот файл переделываешь и делаешь с такими параметрами. У тебя получается хорошая модель. Так что ты начинаешь уже идти потихоньку на самоэкзамен.

Так. Хорошо, спасибо, да. На хорошую оценку. Вот, в отличие от других.

Раздел 6: Заключение и дальнейшие шаги

Подраздел 6.1: Обсуждение темы "Рейтинг университетов" (Никита)

Евгений Венеминович, можно вот спросить? Я смотрел здесь по темам, э-э, нашёл вроде бы интересную по поводу мирового рейтинга университетов. То есть здесь, в принципе, страна, название, э-э, рейтинг публикаций, количество студентов и так далее. Из тем, что были, я здесь нашёл только какой-то университетский рейтинг Гарвард в системе Eidos, и то он сейчас недоступен, ошибка 404. Вот. Это, получается, можно будет использовать эту тему?

Сейчас, сейчас посмотрим. А, в смысле, вот в этих уже сделанных работах, да?

Да, она прямо так и называется: университетский рейтинг Гарвард в системе Eidos. А у меня просто мировой рейтинг университетов.

Угу. А какой номер там? Ты говоришь, она недоступна.

Сейчас найду. Шестьдесят девятый - это мир.

Ага, шестьдесят девятый. Да, точно. Ух ты! Вообще не пойму, почему мышка не движется ниже. Она должна двигаться ниже. Что такое вообще? Так. Шестьдесят девятый. И там действительно нету её. Теперь смотрим на этот... Ладно. Смотрим на... Всё. Исправили. Спасибо. Значит, теперь ответ на вопрос на твой. Ну, спасибо за то, что наметил, заметил.

Подраздел 6.2: Итоговые замечания и предупреждения

Значит, э-э... Нет, ну то, что в пятницу вечером едут домой, это я знаю, ребят. Но они ж едут с телефонами домой, наверное, да? Вот. Ну ясно, ладно. Интернет там какой-то там не дают там, ну, в общем, понятно всё с вами.

Значит, ребята, э-э, чтобы определить, годится ли тема для вашей разработки, для этого нужно сделать модель и посмотреть, что получилось, понимаете? Если получилось нормально, то почему бы и нет. Вот 5 минут 51 секунду, ну 6 минут, грубо говоря, делалась модель эта. Смотрим теперь, вот здесь вот в режиме 1.3 нужно задать название приложения. РСК анализ и классификация грибов. Что-нибудь такое. Вот. И смотрим, что у нас получилось. У нас получилось исключительно хорошая модель. 980, видите, 0,980. Причём все все модели очень хорошие. Вот берём модель сумма знаний.

Ну, ребят, ну со своим азартом залетите. Я вас предупреждаю, ребят. Я поставлю вам тройки, если вы не сделаете этого приложения. Ну как хотите. Мне это начинает надоедать, понимаете? То есть я мы вам с вами сидим тут занимаемся, всё вам показываем, рассказываем, разжёвываем, можно сказать, понимаете? Разжёвываем. А вы ещё и носом воротите.

Так, ребята, ну что, э-э, у вас нет соображений по поводу темы? Да, кто-то говорил там по поводу темы. Извините, ребят.

Так, ну что, в общем, мы сейчас, конечно, немножко поразжились. Но если бы были немножко поактивнее, тогда могли бы и сами тоже нечто подобное сделать. Но сейчас вы видели, я несколько раз я э-э смотрел разные модели и отказывался от них. От некоторых отказывался, от некоторых брал. Принцип какой? Если я понимаю, как её сделать, вижу, что это не очень трудоёмко будет, я беру. Если не очень понятно, что с ней делать, то я не беру. Ну, вам, конечно, сложнее понять. Вот. Но тоже, я так думаю, и вам тоже что-то понятно или непонятно. То есть вы уже так вот видите, как я это делаю. Ну, примерно так же и вы можете делать.

Так, осталось 7 минут, ребят. Но я думаю, что больше мы не будем, наверное, смотреть приложений. А эти 7 минут э-э потратим на то, что вы вы мне сейчас э-э, ну как, можете задать вопросы, короче говоря. Тут получили одно сообщение. Где я получил одно сообщение? А. Понятно. Пожалуйста, пожалуйста.