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

***Российская Федерация***

**164 Теория информации, данные и знания. Практическая 3. Инструкция по разработке собственного приложения 2020-10-27**

## Заголовок

Практическое занятие по разработке приложений Эйдос: Регистрация, выбор темы, новый CSV-конвертер и работа с данными

## Резюме текста

**1. Введение и детали записи**
Онлайн-занятие для группы Т20-03 проводится 27 октября 2020 года (11:30-13:00). Это третье практическое занятие по дисциплине "Теория информации, данные, знания". Занятие ведут профессора Луценко Е.В. и Аршинов Г.А. Цель – продолжить изучение инструкции по разработке собственных приложений в системе Эйдос-Х++.

**2. Административные вопросы и вовлеченность студентов**
Проводится проверка посещаемости. Подчеркивается важность выполнения заданий для получения оценки. Выражается обеспокоенность низкой вовлеченностью некоторых студентов, пропускающих занятия, что затруднит сдачу курсовых работ и экзаменов. Обсуждается необходимость получения корпоративной почты через ЦИТ (Центр информационных технологий) университета, так как она требуется для регистрации в некоторых системах (например, ResearchGate) и доступа к ресурсам. Отмечается, что КГУ предоставляет такие почты (на базе Google Workspace) своим студентам.

**3. Основные задачи и требования к проектам**
Главная задача студентов – разработать собственное приложение Эйдос. Для этого необходимо:

* Зарегистрироваться в РИНЦ и ResearchGate (регистрация в РИНЦ может занять до 2 месяцев из-за оформления договора).
* Выбрать тему для своей работы/приложения.
* Найти и подготовить исходные данные для анализа.

**4. Обновление системы Эйдос: Конвертер CSV-DBF**
Анонсирована новая версия системы Эйдос от 27 октября 2020 года. Ключевое нововведение – встроенный конвертер для преобразования данных из формата CSV в DBF (старый формат Excel).

* **Назначение:** Позволяет импортировать данные из внешних источников (например, Kaggle, официальная статистика) в систему Эйдос.
* **Преимущества:**
	+ Преодолевает ограничения Excel на количество строк (более 1 млн).
	+ Избегает проблем с автоматическим форматированием данных в Excel (например, преобразование чисел в даты).
	+ Обеспечивает высокую скорость конвертации.
	+ Автоматически определяет структуру CSV (имена полей, типы данных, длины) и создает оптимальную структуру DBF файла.
	+ Работает в два прохода: первый – анализ структуры, второй – перенос данных.
* **Использование:** Требуется указать номера колонок для классификационных и описательных шкал. Структуру CSV можно определить, просмотрев файл заголовков.

**5. Обработка данных: Источники, подготовка, ограничения**
Обсуждаются источники данных (Kaggle, сайты госорганов, например, Генпрокуратуры по статистике преступности). Подчеркивается необходимость подготовки данных перед импортом (например, транспонирование таблиц, удаление ненужных строк/столбцов). Важное ограничение: при размещении приложения в Эйдос-облаке (хостинг Byethost) размер файла исходных данных (CSV) не должен превышать 10 МБ.

**6. Пример работы с данными о преступности**
В качестве примера рассматривается работа с данными о преступности. Демонстрируется процесс:

* Анализ структуры CSV файла (просмотр заголовков, определение разделителей).
* Подготовка данных в Excel (транспонирование, удаление лишних строк/столбцов).
* Использование нового конвертера для импорта данных в Эйдос.
* Проверка результата (созданные шкалы, градации).

**7. Заключение и дальнейшие шаги**
Студентам напоминают о необходимости зарегистрироваться в РИНЦ и ResearchGate, определиться с темой проекта и начать работу с данными, используя возможности системы Эйдос, включая новый конвертер. Подчеркивается важность активного участия в занятиях для успешного освоения материала и сдачи итоговых работ.

## Детальная расшифровка текста

**1. Введение и детали записи**

* **Приветствие и начало записи**
Ой, боже мой. Наклейки мне. Выключи.
Так, ребята, запись включена. Здравствуйте.
Здравствуйте.
Здравствуйте, здравствуйте. Здравствуйте, Георгий Александрович.
Добрый день.
Добрый. Да.
* **Информация о занятии**
Значит, сейчас у нас, ребята, 27 октября 2020 года. Четвертая пара. Ой, извините. Третья пара. 11:30-13:00.
Третье практическое занятие группы Т20-03, на котором мы продолжаем изучать инструкцию по разработке собственных Эйдос-приложений, чтобы вы могли сделать свое приложение и сдать его и получить соответствующую оценку.

**2. Административные вопросы и вовлеченность студентов**

* **Дисциплина и преподаватели**
По дисциплине, по дисциплине занятия: теория информации, данные, знания с группой, я сказал, ИТ-2003.
Занятия ведут профессор Луценко Евгений Вениаминович и профессор Аршинов Георгий Александрович.
* **Посещаемость и отчетность**
Теперь позвольте, я сейчас вас тут всех поотмечаю. Это 2003. Практическое занятие номер три.
У нас такая система, что мы проводим занятия и сразу же должны везде отметиться. И сами, что мы ведем занятия, и вас, что вы присутствуете.
Значит, это у нас третья пара.
А здесь присутствуют восемь. Присутствуют. И вот здесь мы должны...
Я им пишу, сейчас они зайдут.
* **Обеспокоенность вовлеченностью и последствиями**
Ну, как сказать, хозяин – барин.
Я-то уже сдал экзамены и Георгий Александрович, понимаете, и курсовые мы все сдали, защитили кандидатские, докторские. А вы еще не сдали, ребята. Вот в этом разница, понимаете? Вы будете нам сдавать их. И как вы будете их сдавать – это, в общем-то, совершенно непонятно, потому что вы не ходите на занятия даже.
Я, правда, ругаю тех, кто ходит. А тех, кто не ходит, я их не ругаю, потому что их же нету. Поэтому они об этом и не знают, что я ругаю, да?
Вот. Ну, я даже не ругаю, я просто хочу сказать, что это ваши проблемы, конечно, но хотелось бы все-таки, чтобы вы что-то соображали, умели и получили нормальные оценки, заслуженные.
* **Необходимость корпоративной почты и регистраций**
Какую почту, Никит? Мою почту, что ли? Или чью почту? О, господи.
Корпоративную почту? В ЦИТе. Центр информационных технологий. Подойти в ЦИТ. Ну, наверное, не вам лично, а, наверное, там куратору или кто там у вас, старосте. И просить, чтобы дали вам корпоративную почту, потому что уже сейчас это совершенно необходимо. И давить на то, что вот, скажем, КГУ – у всех есть корпоративная почта. Вот смотрите, вот я беру студента. Вот у него корпоративная почта на домене edu.kubsuru. Это домен, на самом деле, домен Google. Гугловская почта. Там есть возможность организации создавать свою почту со своим доменом, учебной организации. Они могут там сделать хоть миллион студентов, им совершенно все равно. Это Гугловская почта, там совершенно без проблем можно сколько угодно дать этих адресов. Почему они с этим как-то там вот... Э-э, вот другой адрес. Это студенческие адреса, буковка S. А у преподавателя такая же, только буковка P и номер какой-то. Вот.
А в РИНЦ вы должны уже начать регистрацию сами, потому что там ничего не требуется, кроме вашего усилия определенного. То есть надо сесть, заполнить карточку, послать эту карточку на регистрацию. Это совершенно стандартная процедура, никаких проблем нет в этом.

**3. Основные задачи и требования к проектам**

* **Цель - разработка приложения Эйдос**
Значит, теперь, значит, у меня по расписанию я отмечаю, чем мы занимались на прошлом занятии. И на прошлом занятии мы с вашей группой начали изучать инструкцию по разработке собственных приложений. И сейчас у нас продолжение. Продолжение этого занятия. То есть задачу я ставлю какую сейчас? Значит, я в прошлый раз на занятии поставил вам задачу регистрироваться в ResearchGate, ResearchGate и РИНЦ и выбирать тему работы. Было дело такое, нет? Отвечайте микрофоном. Да, нет? Было дело, не было? Было, да? Или вы уже забыли всё?
Ребят, давайте реагируйте на то, что я говорю. Я вам сказал, на прошлом занятии мы изучали инструкцию по разработке собственных приложений. Эта инструкция является и одновременно заданием.
* **Регистрация в РИНЦ и ResearchGate**
Значит, это первые два вопроса: регистрация в РИНЦ, регистрация в ResearchGate. Причем они друг с другом особо не связаны. То есть это можно делать независимо. То есть можно и то, и то делать сразу, понимаете? В РИНЦ вообще надо сразу делать, потому что там, чтобы заключить договор, проходит, наверное, месяца два, наверное. Вот пока вы сейчас зарегистрируетесь в РИНЦ, получите Science Index, пока получите спин-код, пока оформите договор, пока он станет действующим у вас, пройдет месяца два, ребята. Понимаете? Уже октябрь, конец октября. Нужно иметь в виду, что вы просто не успеете.
Вот. Ну а в ResearchGate там регистрация недолго происходит. Там где-то буквально там несколько минут, там 5 минут, 10 минут. Но там проблема в том, что нужно вот адрес корпоративный иметь. Ну, я так понимаю, что он решаем этот вопрос. Это зависит от ЦИТа.
Можно также использовать адрес какой-нибудь преподавательницы там или кафедры, методиста. Ну то есть какой-нибудь адрес сотрудника университета, любого сотрудника.
Значит, вопрос первый – это регистрация. Второй вопрос – это выбор темы.
* **Выбор темы и типы работ**
Мы с вами прошли все лабораторные работы, в том числе и, значит, ну, трех видов. Трех видов работы.
Первого вида – это обработка табличной информации, числовой и текстовой. Второе – это обработка, интеллектуальная обработка текстов. Третьего типа – это обработка, интеллектуальная обработка изображений, в частности, спектральный СК-анализ.
Вот. И потом мы сейчас подошли к вопросу выбора темы. Теперь напомните мне, пожалуйста, мы уже рассматривали этот вопрос выбор тем или еще нет?
Выбор темы. Начали рассматривать, вы сказали нам подумать над темами.
Вот. Ну и как вы?
Кто-то подумал, кто-то, не знаю. Я подумал.
Ну и что ты придумал?
Я вам показывал, где искать данные?
Я хочу посмотреть этот, количество преступлений, совершённых в городе Краснодаре. Ну, взять как тему эту.
Угу. Данные составить. Понятно, понятно. Вы ж, я же вам показывал, где данные искать, да? UCI, Kaggle.
Да, да, про Kaggle вы говорили.
То есть это всё я рассказывал, да?

**4. Обновление системы Эйдос: Конвертер CSV-DBF**

* **Анонс новой версии**
Значит, ребят, я сейчас хочу вам такую новость сообщить. Значит, ну, приятную. Значит, я сегодня ночью, ну именно сегодня, то есть это было там в начале первого, разместил новую версию. С 27 числа эта версия системы Эйдос. То есть если ее запустить, то она показывает 27 число. На названии окна главного написана дата версии – 27.10. Видите, да? Я же вам доступ открыл. Вот.
* **Представление CSV-DBF конвертера**
В этой версии есть в этом, в автоматизированном программном интерфейсе API 2322, есть опция CSV-DBF конвертер. И там есть описание стандарта CSV, как должно, как должен быть, какое должно быть наименование файла, где он должен находиться исходных данных. Короче говоря, можно вводить данные из CSV-файлов.
* **Использование конвертера и определение структуры CSV**
Но для этого надо знать, какая у них структура, чтобы указать номера полей классификационных шкал, описательных шкал. Вот. И для этого, чтобы узнать структуру, есть два способа. Один способ – это сам CSV-файл открыть, взять первую строчку. Вот. Ну сейчас я вам покажу вам, чтобы не отходя от, как говорится, далеко, чтобы мы понимали, как это делается.
Значит, вот сейчас я вам покажу. Вот берем CSV-файл. Вот он такой вид имеет. То есть у него поля разделены запятой. Выделяю строку заголовков, помещаю в буфер обмена. Захожу в Word. Вот. Сюда эту строку заголовков копирую. Потом беру, заменяю запятую на абзац, конец абзаца. Заменяю запятую на конец абзаца. Заменить всё. Для чего я это делаю? Для того, чтобы не считать самому номера колонок этих рукой, как говорят, а просто вот автоматически это сделать. И я вижу, смотрите, ребята, что у меня 29 колонок. По смыслу этих колонок, вот эта последняя колонка является классификационной шкалой. Вот. Ну, если разбираться, перевести и разбираться в этом приложении, а там взято, этот файл взят с определенного источника в Kaggle, где написано, что это такое. В общем, ну я сейчас вам не буду детализировать, но я вам просто вам показал, что вот у нас описательные шкалы со второй по двадцать восьмую, а классификационные с двадцать девятой по двадцать девятую. Вот этот файл я должен скопировать согласно input data, то есть согласно хелпу, в папочку input data. Вот. И показываю вам в чате путь на этот файл. Вот там он должен находиться. Он должен так называться: data.csv.
* **Пример конвертации**
После этого я могу вводить информацию из него. Вот. Указываю классификационные шкалы с 29 по 29. А это значит, с второй по 28. Ну лучше не со второй, с четвертой. Потому что мы если посмотрим на этот CSV файл, то здесь у нас первые три колонки у нас, видите, идут уникальные текстовые идентификаторы. Но это слишком уже. То есть первая, вторая, третья. Вот с четвертой мы начинаем ввод данных, последняя – это у нас классификационная шкала, а это источник данных. То есть там идентификаторы в трех видах. И в общем, нажимаем Окей. Значит, собственно, всё преобразование уже произошло. Файл получился меньше 2 ГБ. Полностью завершено преобразование. Потом передаётся управление на выполнение следующей операции – ввод данных из DBF файла, который создан после преобразования, конвертации CSV в DBF. DBF – это старинный формат Excel, когда еще был второй, всего лишь навсего, Word. Не сейчас вот 2000 там 19-й там, да? А когда-то было второй Excel, второй Word. Вот это вот формат второго Ворда, второго Экселя. DBF – это формат второго Экселя. Вот. И у нас получились уже градации классификационной шкалы, описательные шкалы и градации. Всё у нас получилось. То есть мы ввели данные из CSV файла.
* **Альтернативный режим конвертера (23210)**
И есть еще режим, я сделал 2322, который то же самое, 23210, извините, 23210, который тоже представляет собой тот же самый конвертер. Но для чего я его сделал? Для того, чтобы мы могли, если мы не можем взять эту строку заголовков и посмотреть, какие там заголовки, вот так вот их пронумеровать. Тоже нужно все-таки чуть-чуть соображать, чтобы это сделать. Это кажется, может быть, тяжело кому-то. Тогда, значит, можно просто преобразовать в DBF файл, экселевский файл. И потом можно зайти в папочку, где вот он образовался этот DBF файл. Вот он он. И просто экселем его открыть и посмотреть, что здесь такое. Вот. Пожалуйста, можно посмотреть.
* **Проблемы Excel при конвертации и преимущества нового конвертера**
Значит, вопрос возникает такой: а что в Экселе не преобразовывать CSV в DBF? Есть две причины, ребята. Есть две причины. Первая причина: Excel, он сверхинтеллектуальный, и иногда он настолько интеллектуальный, что даже его когда не просят, а он все равно интеллектуальный, понимаете? Знаете, что значит? Ну есть такое выражение, что слишком хорошо тоже нехорошо, или всё хорошо в меру. Слышали такое, нет? Кто слышал? Да слышали, слышали. Слышали, да? Ну вот. Так вот, он настолько интеллектуальный, что иногда это превосходит всякие, так сказать, разумные пределы. То есть он берет и числа такого типа, вот сегодня какое число? 27.10.2020, да? А можете себе представить, что в текстовом файле, вот в этом CSV, там встретилось число такое: 27 10 20 20, без точек, без подряд вот просто вот 27 10 20 20. Может такое произойти? Ну редко, но бывает. То есть бывают числа, которые похожи на даты, понимаете? Может быть, не 27.10.20, а может быть, там, допустим, 15.03.2020 там. То есть вот только без точек, просто вот подряд. Он, гад, берет этот Excel и пишет там 27 октября, например, там. Прямо вот в колоночке, понимаете? Вот от этого избавиться довольно сложно. Значит, приходится лезть в интернет и искать онлайн-конвертеры и преобразовывать с помощью онлайн-конвертеров. Некоторые онлайн-конвертеры тоже так делают, а некоторые нормально. Значит, я сделал конвертер на этом языке, на котором написана система Эйдос, который хорошо всё это обрабатывает, понимаете? То есть он не дурит с этим делом там, не пытается там даты подпихивать там. То есть там числа – так числа, текст – так тексты. Там два прохода, если вы заметили. Сейчас я вам покажу еще раз этот режим, чтоб показать уже 23210. Два прохода. В первом проход определяются имена полей, длин и длинные типы данных, число знаков после запятой. А второй проход создается DBF файл с таким, с такой структурой, именно вот с такими именами, с такими длинами полей, с такими типами данных. А потом туда переносится информация, это второй проход, из CSV файла. Получается оптимальная структура DBF файла, то есть он все поля минимального достаточного размера, понимаете? Вот. Это очень ценно. Значит, второе качество этого конвертера, который вот я сейчас включил в систему, заключается в том, что Excel, даже новые версии, они обеспечивают ну больше миллиона записей в файле. А если мы хотим обработать побольше, ну на Kaggle, когда мы заходим на Kaggle, то мы часто обнаруживаем там файлы, которые 4 млн записей. Даже я видел там 12 млн записей. Значит, вот этого, у этого CSV конвертера нет никаких принципиальных ограничений на число записей. Таких жестких каких-то ограничений нет. Я отладку вел на маленьком примере, где 1000 записей вот этот вот. Ну это маленький пример. Он мгновенно преобразуется, вы заметили, да, очень быстро. Вот. Кстати, его скорость работы – тоже это его достоинство. То есть там определенные решения технические, которые обеспечивают очень высокую скорость его работы этого конвертера. Гораздо выше, чем если просто на языке программирования написать в лоб, в тупую, как говорится, написать, то он будет работать там в десятки раз медленнее, а то и в сотни раз. Значит, здесь очень хорошие решения приняты технические. И этот конвертер, я отладку вел сначала на маленьком, чтобы все получилось, а потом на большом. И вот в этом большом было 4 млн записей. Это уже всё, это Excel уже это не способен преобразовать. Значит, теперь, в новом Экселе, где миллион записей, нет CSV-DBF конвертера. То есть он позволяет только в Excel преобразовать исходные тексты, значит, исходные, то есть исходные данные взять из текста. Это как раз из CSV имеется в виду. И секундочку, это издательство... Это издательство.

**6. Пример работы с данными о преступности (Продолжение)**

* **Ограничения Excel и возможности DBF/CSV**
Извините, это руководство, начальство, поэтому я не могу им не ответить. Вот. Значит, на чем мы остановились, ребят? Я что-то вообще... Там связано с защитой, ну короче говоря... Напомните, на чем я остановился, пожалуйста.
Ну вы говорили для Экселя очень много 4 млн данных. Excel не может преобразовать такой объем данных. Вот. И вот этот вот конвертер данных, он есть только в третьем Экселе, 2003. А в более поздних версиях его убрали. Получается, что они вынуждают использовать экселевский формат. Система Эйдос воспринимает экселевский формат, но экселевский формат – это не более миллиона записей. А если использовать только Word, то есть, ой, извините, что я, я уже устал, наверное. Значит, если использовать только Excel для конвертации, сам вот Excel использовать, то тогда это 2003 Excel, а в нем 65536 строк в экселевском файле максимальная. И всё, понимаете? Вот. И всё. То есть получается, что такая некая сложная ситуация. Но выход из нее только с использованием онлайн-конвертеров, которые не всегда работают, в общем-то, тоже надо их искать и так далее. Вот. Так что вот я сделал конвертер, который хорошо работает на больших объемах данных и готовит файлы, которые может обрабатывать система Эйдос. А в системе Эйдос есть режим, вот этот, который я сказал, 2322. Он позволяет вводить информацию из экселевских файлов разных стандартов Экселя, старого и нового. Вот. Из DBF файлов позволяет и CSV файлов. Так вот, старый Excel 65.000 строк ограничение, новый миллион строк. DBF вообще нет ограничения принципиального, и CSV нет ограничения принципиального.

**5. Обработка данных: Источники, подготовка, ограничения (Продолжение)**

* **Ограничение размера DBF файла (2ГБ) и хостинга (10МБ)**
Значит, я могу сказать, что есть некое ограничение, связанное с вообще с максимальным размером DBF файла. Этот размер он не больше 2 ГБ. Вот. Я когда проводил эксперименты с четырехмиллионным CSV файлом, вот где 4 млн записей, я оценил, какого размера получается DBF файл. И сколько можно было бы еще, так сказать, добавить, чтобы было 2 ГБ. Чтобы DBF файл был 2 ГБ. Получилось у меня где-то 10-12 млн записей, в зависимости от числа полей. То есть, если вот в этом CSV файле много полей, вот, то тогда, может быть, там, ну, скажем, 10 млн записей получится. Если вот если столько, сколько здесь, то может быть, там и 20 млн получится, понимаете? То есть это зависит от числа полей в CSV файле. Может быть, там и 200-300 полей. Ну тогда получится там 5 млн записей при 2 ГБ уже DBF файла. В любом случае, DBF файл имеет значительно большие возможности, чем экселевский файл по размеру и, значит, по объему информации, которая там может быть. Он превосходит в разы там, ну, скажем так, вот от пяти раз, от четырех-пяти раз до десяти там, пятнадцати раз вот может он превосходить возможности Экселя современного. Я когда появлялся, появился новый Excel, ожидал, что там будет больше миллиона. Ну будет, допустим, там 20 млн записей. Нет, они так миллионы оставили. Это, вообще-то, печально немножко, потому что на сайте Kaggle мы там видим данные, которые, ну, скажем так, Excel просто не позволяет вообще их обрабатывать. То есть не позволяет их считывать, обрабатывать, использовать для решения задач. А система Эйдос позволяет. Я вот эти этот файл 4 млн записей ввёл в систему Эйдос, создал модели, то есть я его обрабатывал. Вот. Но надо признать, что это происходит, надо, ну так, довольно-таки медленно. Потому что такой объем обучающей выборки, размерности, значит, уже настолько велики, что это всё ворочается даже вот шкалы формируются там, вот данных формируется. То есть это уже это идет часами, сутками. Вот. В этой связи возникает вопрос о том, как это всё ускорить.
* **Способы ускорения обработки больших данных**
Значит, ну есть два способа, как это всё ускорить, или три. Один способ – это написать на каком-нибудь эффективном языке программирования, на C, например. Есть версия вот этого языка программирования, который я использую для системы Эйдос, которые на самом деле не являются отдельным языком, а являются, скажем так, неким интерфейсом между синтаксисом вот этого языка, который я использую, и C. Ну то есть всё фактически преобразуется в C и потом компилятор C используется и так далее. Тогда будет это называется Harbor язык. Язык Harbor – это язык XBS++, то есть тот, на котором написана система Эйдос, но по сути дела, это C. Вообще C просто, то есть это фактически C, потому что там компилятор C используется. Вот. И второй вариант есть – это использовать увеличенные аппаратные возможности компьютера. Здесь есть два пути. Один вариант – это использование скоростных SSD дисков, которые высокой скоростью обладают обмена. И другой вариант – это использовать графическую карту, графический процессор для неграфических вычислений. В системе Эйдос используется графический процессор для расчета моделей и распознавания. А вот эта карточка, вот мне недавно это прислали. Обратите внимание, сейчас я вам пришлю ссылочку на неё. Я вам еще не присылал эту ссылочку, нет? Вот. Вот эта ссылочка, ребят. Обратите внимание: 330 МБ/с на запись, 530 МБ/с на считывание. Значит, у меня товарищ поставил на какой-то старенький компьютер. Ну, тогда должен иметь разъём М2. Вот я не в курсе, на каких материнках, с какого, начиная с какого года есть разъём М2. Он прямо вставляется в материнку. Вот. Вид имеет, по сути дела, флешки какой-то. Вот похоже на флешку, мне кажется. Видите, такое устройство? Вот у него 122 ГБ. Вот это устройство вставляется прямо в разъём на материнке, разъём М2. И получается у вас 530 МБ/с скорость обмена с Винчестером. Он поставил на старенький там какой-то свой компьютер, которому там 10 лет. Значит, оно, видимо, давно этот разъём существует уже. И, значит, поставил туда операционную систему на него, Windows, на эту, на это, на этот SSD диск. И там же стал считать на системе Эйдос. То есть там же сделал папочку с системой Эйдос, стал считать. У него получилось так, что, значит, компьютер вообще ожил, вторую жизнь получил. Вот, стал работать гораздо шустрее, намного просто шустрее. Вот. А система Эйдос, он такой пример привёл, что у него синтез, верификация модели шли 20 минут, вот до того, как он эту штучку воткнул в разъём М2. А когда он её воткнул, он на ней сделал диск, то есть на ней сделал папочку, вернее, на этой, на этом устройстве, на SSD, на этом устройстве. И, значит, туда он поместил систему Эйдос, и расчёт стал идти полторы минуты. Ну то есть больше, чем в 10 раз быстрее. Ну он говорит, там полторы-две минуты, то есть в 10 раз быстрее практически. Ну то есть это очень серьёзно, по-моему. То есть если вы, допустим, 4 млн записей обрабатываете, она у меня обрабатывала там чуть ли не сутки эту, этот объём. Ну будет там час, например, обрабатывать там, полтора. Ну то есть это совершенно, так сказать, меняет дело, понимаете, уже. То есть я вам сообщаю вот такую информацию. Я считаю, что это довольно такая важная информация, полезная. Вот. И мне было бы интересно, если кто-то из вас это сделает, чтобы мне прямо вот сказали, что вот мы это сделали, получилось. Действительно вот хорошо работает там. Вот. И заодно я не занимался этим. Я вот даже мне самому захотелось на компьютер поставить такое устройство.

**7. Заключение и дальнейшие шаги**

Вот. Так что, э-э, в системе Эйдос сейчас есть улучшенные возможности для использования данных в формате CSV с сайта Kaggle. Там в основном именно в таком формате и есть. Но я вам хочу сказать, что вот сейчас тут здесь вот я написал недавно, это написал, что не следует брать исходные файлы, вот эти CSV, больше 10 МБ размером. То есть если там у вас вы находите на Kaggle какое-то приложение, и там, допустим, 4 ГБ надо скачивать, то даже можно не смотреть, что там такое находится. Почему? Потому что мы хотим разместить это приложение с исходными данными и описанием в Эйдос-облаке на хостинге Byethost. Вот этот хостинг здесь немножко видно, вот здесь вот Eidosbyhost.com – это FTP сервер системы Эйдос. Этот хостинг, он обладает, ну это американский хостинг, скоростной, 1000 Мбит/с. Вот, у него для пользователей предоставляется. Вот. Там кластеры стоят компьютерные, сервера мощные, там тридцатидвухядерные сервера. В общем, короче, там у них круто всё. Объём памяти неограниченный предоставляется пользователю. Нет никакой рекламы вообще. Всё это бесплатно. Вот. Не знаю, почему, честно сказать. Там есть и платные тарифы, но, значит, с услугами, но там, может быть, какие-то у них преимущества. Но меня это интересует бесплатное. Значит, вот это, что я сейчас говорю, бесплатно, без рекламы, неограниченная память, очень высокая скорость взаимодействия, поддержка всех языков: PHP, JavaScript, там SQL, всё это у них там поддерживается и другие возможности, всё, что есть. Вот. Единственный недостаток у них есть. Вы уже, наверное, догадываетесь. Файлы больше 10 МБ размером, они молча стирают просто вообще. То есть вот запишешь туда 19 МБ, они просто сразу же его... То есть ты увидишь, что он там появился при FTP. Вот. Сейчас я покажу вам FTP доступ к нему. Вот, тут он появился этот файл вот здесь или здесь, вот где-нибудь. Вот здесь вот у меня файлы, папки приложений. Ну сейчас 208 приложений. Вот сюда можно записать файл любого размера. Обычно, э-э, обычно вот в приложениях вот здесь я записываю файлы описания. Вот, допустим, описание этого приложения 4 МБ занимает. Вот. А были случаи, когда вот, ну и файл исходных данных. Этот файл всего 766 КБ. То есть он меньше мегабайта. Вот. Так вот, представьте себе, что здесь будет файл 10 МБ исходных данных. Тогда он еще будет здесь находиться и будет скачиваться, когда мы в системе Эйдос запустим режим соответствующий. Сейчас я вам покажу, как это выглядит. Вот в режиме 1.3 скачать приложение из облака. Вот, пожалуйста. Вот. Мы здесь видим и ссылочки на сами эти описания этих приложений и установка приложения и каталог обсуждений. Вот я беру, допустим, обсуждение и беру, получаю гиперссылки на ссылки на файлы этого приложения. И вижу здесь, что, что есть файлы размером, скажем, там 2 МБ, например. Файлы исходных данных, видите? 2 МБ. Я могу любую из этих ссылочек взять, поместить её в строку адреса и этот файл скачать. Но если я сейчас закрою это, то, значит, этот файлик уже будет измененный. Этот файл обсуждений по вот этому конкретному приложению. Вот. То есть мы можем его обсуждать это приложение. Можем точно так же другие приложения обсуждать. То есть каждое приложение является темой форума фактически. То есть оно можно использовать его, то есть эти файлы для обсуждения. Есть каталог обсуждений. Мы можем посмотреть, что там с этими приложениями кто делал, какие там, когда это происходило, дата и время. Тут есть и скачивание приложений, и сообщения в приложениях есть. То есть мы можем сейчас вот увидеть, что у нас сегодня, вот, допустим, появилось сообщение вот в этом приложении 207. Вот. Так вот, все эти дела, они возможны только тогда, когда файлы там меньше 10 МБ находятся. Вот. Если больше, просто сразу стираются и всё. Вот такое ограничение. В этой связи вот я и не рекомендую большего размера файлы использовать. То есть можно создать приложение большего размера, но вы его тогда не разместите в Эйдос-облака. А это очень желательно. Вот. А в ResearchGate и РИНЦ вы размещаете только описание, поэтому там без разницы, какого размера исходные данные. Ну я вам сказал, что если размеры очень большие, то обработка будет затянута по времени. Долго будет идти.
Так, ну теперь, ребята, мне бы хотелось, чтобы кто-то из вас назвал тему свою. То есть вы вообще что должны сделать сейчас? Вы должны зарегистрироваться в РИНЦ и ResearchGate, выбрать тему, и эту тему мы должны проверить, насколько она пригодна для того, чтобы делать по ней работу. Критерии пригодности такие: мы должны создать приложение с исходными данными по этой теме и посмотреть, какова достоверность этого приложения, насколько оно получилось хорошее. Вот вы сейчас на лабораторных занятиях, на практических занятиях уже должны этим и заниматься. То есть вы сейчас должны сказать: "Вот я выбрал такую тему. Вот на неё ссылочка в чате на эту тему". Мы по этой ссылочке идём, скачиваем CSV файл и смотрим. Я показываю вам, и все смотрят на то, как это можно делать. И тоже начинают искать там приложения и делать то же самое, смотреть, какие получаются модели и так далее. Вот. Значит, ищем мы данные для приложений, ищем вот на этих сайтах UCI, Kaggle, на моём сайте вот здесь и ищем просто путём поиска данных для машинного обучения. Если взять мой сайт вот этот, то здесь ссылочка прямо на страничку, где как раз об этом и говорится. Раньше это было не видео такой инструкции, а раньше было просто вот на страничке написано. И здесь где есть база данных для машинного обучения. Кстати, ссылка уже не рабочая эта. Вот. И тут указаны кое-какие сайты. Вот, скажем, сайт указан Федеральной статистики. Сайт указан Федеральной прокуратуры, статистика по криминалистике. Вот. Как раз вот то, что, может быть, заинтересовало товарища по преступлениям. Вот. И здесь есть сборники. Они нам как бы не очень подходят. Вот архивы сборников. Вот. Ну, в общем, можно здесь найти данные о состоянии преступности в России. И, пожалуйста, значит, можете здесь искать таблицы для обработки в системе Эйдос. Если тут они есть. Где-то мелькнул Краснодарский край, но это всё в графической форме. Вот. Нет, чтобы сделать возможность скачивания, да? Они там придумали. Может быть, прокручивание сделали что ли? Когда блоком выделяешь. Угу, всё-таки она хоть и видно не было, но ползла. То есть можно до конца. Вот. Ну это что-то такое уже напоминающее.
Ну вот из такой таблицы у меня возникает такое впечатление, вот когда на неё смотришь, что надо её транспонировать, чтобы показатели вот эти вот были колонками, а года были наблюдениями. Понимаете, нет? Да-да-да. Вот. То есть её надо вытащить в Excel. Как это сделать, тут мне совсем понятно. Скачивания таблицы нету такой прямо вот чтобы скачать. Не видно. И я бы, значит, убрал бы вот эти вот место. Я не знаю, что за место. Не понял. Рейтинг какой-то, что ли? Не поймёшь. И вот она не выделяется у нас блоком. Угу. Вот. В общем, я эти вот колонки, первые две колонки склеить, то есть показатель 20, значение, показатель 10 место, показатель 20 значение. Почему 20 20? Что один и тот же показатель? Я вообще их... Вот. Ну что ж, пошлём. Я думаю, что, может быть, надо это место, вот где место поубирать строчки. Вот.