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

***Российская Федерация***

**170 Интеллектуальные информационные системы и технологии. Лабораторная 7. Инструкция по разработке собственного приложения 2020-10-29**

**Заголовок:** Выбор темы и данных для студенческого проекта по разработке интеллектуальных информационных систем

**Резюме:**

**I. Введение**
Занятие начинается с приветствия и краткого обмена информацией между преподавателями и студентами.

**II. Обсуждение публикации в Scopus**
Преподаватель сообщает об успешной индексации статьи студента в базе данных Scopus и пересылает соответствующее письмо от сотрудницы, занимающейся этим вопросом.

**III. Контекст занятия**
Уточняется дата (29 октября 2020 г.), время (8:00 - 9:30), группа (ИТ-18-21) и дисциплина (Интеллектуальные информационные системы и технологии). Занятие представляет собой лабораторную работу №7.

**IV. Основная задача: Разработка приложения**
Основной темой занятия является продолжение рассмотрения инструкции и задания на разработку собственного программного приложения в рамках изучаемой дисциплины.

**V. Подготовка к демонстрации**
Преподаватель настраивает демонстрацию экрана, предоставляя доступ студентам и готовясь отвечать на вопросы.

**VI. Выбор темы для приложения**
Центральная часть занятия посвящена выбору темы для разрабатываемого приложения. Преподаватель предлагает помощь в поиске темы или оценке уже выбранной.

* **A. Ограничения при выборе темы:**
	+ Тема не должна дублировать уже существующие и размещенные работы в EDS-облаке, за исключением случаев, когда старая работа устарела или недостаточно проработана.
	+ Размер исходных данных (одного файла) не должен превышать 10 МБ из-за ограничений FTP-сервера (ByHost). Файлы большего размера автоматически удаляются.
	+ Тема не должна противоречить морально-этическим нормам и законодательству РФ.
* **B. Источники данных и тем:**
	+ Предлагается изучить темы предыдущих работ в EDS-облаке.
	+ Рекомендуются репозитории данных для машинного обучения, такие как UCI Machine Learning Repository и Kaggle, где можно найти готовые наборы данных. Упоминается, что на UCI данные раньше были в специфическом формате, но сейчас часто используется CSV.
* **C. Требования к структуре данных:**
	+ Данные должны быть представлены в формате Excel (.xls/.xlsx) или аналогичном (CSV, DBF).
	+ Первая колонка должна содержать информацию об источнике данных или идентификатор наблюдения.
	+ Далее следуют колонки с классификационными признаками (категории, группы).
	+ Затем идут колонки с характеристиками (свойствами, факторами) – числовыми или текстовыми. В числовых колонках не должно быть текста (использовать 0 при отсутствии данных).

**VII. Обсуждение прогресса студентов**
Преподаватель интересуется, выбрали ли студенты темы, и напоминает о предыдущих заданиях: регистрация в РИНЦ (Science Index) и ResearchGate, выбор темы. Проводится опрос в чате для оценки прогресса.

**VIII. Пример обсуждения темы (Системы впрыска)**
Один из студентов предлагает тему "АСК-анализ используемых систем впрыска топливной смеси автомобилей". Преподаватель уточняет наличие данных и их структуру, советует начать с небольшого объема (10 строк) для проверки концепции и структуры данных перед сбором полного массива вручную.

**IX. Представление преподавателей**
Преподаватель уточняет, что занятие ведут два профессора: Луценко Евгений Вениаминович и Аршинов Георгий Александрович.

**X. Напоминание о заданиях и сроках**
Подчеркивается необходимость активизировать работу над проектом (регистрации, выбор темы, сбор данных, разработка приложения и описания по шаблону), так как приближается сессия, и оценка будет выставляться по результатам выполнения этих заданий.

**(XI. Технические моменты и отвлеченная демонстрация - *не входит в основное резюме*)**
В ходе занятия возникают технические трудности с демонстрацией и обновлением файлов на FTP-сервере. Преподаватель также показывает небольшой видеоролик об изготовлении поддельного вина ("Меланж Инферналь") во время решения технических проблем.

**Детальная расшифровка текста:**

**I. Введение**

Доброе утро. Георгий Александрович, доброе утро, ребята.
Доброе утро.

**II. Обсуждение публикации в Scopus**

Георгий Александрович, видели, я вам письмо написал?
Читал, посмотрю.
Видели, да?
Сейчас открою.
Там, короче говоря, это вот сотрудница, которая занимается тем, чтобы статьи в Scopus, Web of Science, с которой Валерий Иванович работал.
Да.
Она мне вчера написала письмо, и вот я вам там переслал часть информации.
Я прочту.
Ага.
Ну, что прописалась статья в Скопусе ваша.
Хорошо.
Угу.

**III. Контекст занятия**

Так, ребята, значит, сегодня у нас 29 октября 2020 года. Первая пара, 8:00 - 9:30. Лабораторная работа номер семь с группой ИТ-18-21 по дисциплине "Интеллектуальные информационные системы и технологии".

**IV. Основная задача: Разработка приложения**

По плану у нас мы сейчас продолжаем рассматривать инструкцию, задание на разработку собственного приложения.

**V. Подготовка к демонстрации**

Поэтому я открываю доступ к экрану и задаю вам вопрос.
Да, еще вот что. Еще нужно нажать Окей здесь. И вот здесь тоже нажать Окей надо. Несмотря на то, что там эта информация есть, но надо все равно нажимать Окей.
Ребята, ссылка на инструкцию в чате.

**VI. Выбор темы для приложения**

Вот. И теперь я готов вам помогать или найти тему, или разобраться с тем, можно ли ее принять. То есть создать приложение и посмотреть, будет ли достаточно интересная модель, можно ли ей пользоваться.

Так что давайте, ребята, сообщайте, у кого уже есть тема? Кто кто выбрал себе тему?

*Ну, пока что не выбирали. Но, наверное, можно, да, посмотреть, какие у других людей темы, сейчас тогда плюс-минус примериться, тогда чуть позже сказать.*

А я на прошлом занятии не говорил вам, что надо это сделать, нет?

*Да нет, вы сказали, что просто потихоньку будем рассматривать инструкцию. Не говорили.*

Ясно. Ну давай тогда так и будем делать, потихоньку будем рассматривать инструкцию. Тогда я сейчас вам покажу, как находить темы.
Вот. Ну, наверное, мы с чего начнем? Посмотрим, значит, ваша группа 18-21. 18-21.
Вот, ну пункт четвертый. Пункт четвертый - это как раз вот выбор темы.

Ну, ребята, э-э, есть, ну, я над инструкцией этой работаю, над заданием, то есть как-то здесь ее модифицирую, развиваю.
Значит, короче говоря, так, выбор темы. Вы можете выбрать любую тему, какая вам интересна, чтобы вам было приятно делать эту работу.

* **A. Ограничения при выборе темы:**

Вот, но есть некоторые ограничения.
Значит, вот слушай, какие ограничения есть. Первое - этого, эта работа не должно быть уже сделанной и размещенной в EDS-облаке. Вот. То есть вот среди этих работ быть не должно этой работы.
Вот, но есть исключение из правил, знаете? Значит, какое исключение? Если эта работа была давно проведена и описание у нее слабенькое, ну, вот, скажем, 34-я работа, смотрим сюда вправо, тридцать четвертая. Вот, находим здесь тридцать четвертую и смотрим, она вот так выглядит, а ее там и нету вообще. Ух ты! Я даже и не знал, что такое бывает. Не думал, не думал, что такое может быть. Ага, а ну давай посмотрим.
*(Технические сложности с отображением файла, проверка на FTP-сервере)*
Ну вот, в общем, смотрим мы описание. Если мы видим, что это описание сильно устаревшее, слабенькое, то, в принципе, можно взять и сделать эту работу на новом уровне и сделать описание на новом уровне. Потому что здесь работы есть семнадцатого года. За это время в системе много чего появилось, и просто не могли устареть работы.
Ну да, это я знаю, что они не очень-то... Вот. Так что вполне можно взять.
Вот, но всё-таки лучше, конечно, новое приложение.

Второе ограничение. Файл исходных данных не должен быть больше 10 МБ. Потому что вот на этом FTP-сервере, ByHost, замечательном и быстром, и неограниченном по размеру, и без рекламы, то есть у него ряд достоинств есть. На нём всё-таки есть недостатки тоже. Значит, файлы больше 10 МБ просто стираются молча сразу же и всё. То есть туда не записывается, кстати. То есть можно записать файл, скажем, 20 МБ. Он запишется, он там будет там секунду, наверное, не знаю, сколько там, виден, а потом раз, и его не будет, исчезнет он. Поэтому файлы исходных данных, которые у нас input data CSV, input data XLS, XLSX, DBF, не должны быть больше 10 МБ. Один файл не должен быть больше 10 МБ. Все вместе там неограниченное пространство.
То второе ограничение.

Третье ограничение. Темы не должны противоречить общепринятым в России морально-этическим нормам и действующему законодательству Российской Федерации.
В остальном у вас полная свобода.

* **B. Источники данных и тем:**

Вы можете посмотреть на темы, которые брали студенты в прошлые времена. Там довольно много есть таких тем, которые, может быть, вам будут интересны. Ну, можете аналогичные выбрать. Я имею в виду, что там есть темы, связанные с анализом игр, пресцен или матчей, или оружия, ну, разные там есть варианты. Ну, в общем, вы можете посмотреть.

Вот. Ну вот здесь вот у меня, видите, есть семь статей с использованием именно данных с портала UCI. Самая новая вот эта вот. Я вам эти ссылки давать не буду, потому что они все даны вот здесь. Вот ссылки на мои работы. То есть просто сюда клацаете, Ctrl+F, UCI, вам находятся семь статей, где про это говорится. Вот, первая статья, которая об этом упоминается, относится к 2003 году. То есть 17 лет назад я написал статью про использование репозитария UCI для оценки качества моделей систем искусственного интеллекта. Вот выходим на этот портал UCI, на Datasets. Dataset - это набор данных для машинного обучения, так называется. А S, Datasets - это наборы данных для машинного обучения. А вот это вот ML, видите, UCI, там и всё такое, архив, а потом написано ML - это машинное обучение, Machine Learning.
И вот здесь вот мы видим огромное количество, вот я сейчас на русский язык переведу, Google на лету переводит, огромное количество различных наборов данных. Причём действительно огромное, понимаете? То есть тут вполне вам хватит найти что-то. И можно не собирать вот так данные. Ну, конечно, если у вас э такой интерес есть, но можете собрать. Может быть, конечно, а может быть, вам это будет не очень сложно, может быть, у вас есть какие-то там данные в таком виде, что можно их объединить, я не знаю.
Значит, смотрите, с каких лет это всё идёт? Значит, идёт с девяносто пятого года у них архив. И вот здесь вот справа, видите, здесь написано, какого типа данные там: категориальные, целочисленные, действительные. И тип задачи: классификация, рекомендательные системы. Значит, категориальные - это имеется в виду текстовые данные, ребят. Целочисленные, действительные не требуют комментариев. Вот здесь вот написано, сколько у нас э примеров обучающей выборки: 4.048, 1.798, 37.000, 6.000. И сколько атрибутов, то есть сколько свойств. Вот, а каждое свойство имеет значение ещё.

Смотрим сайт Kaggle, ребята. Kaggle, я считаю, предпочтительным вариантом по сравнению с UCI, потому что на Kaggle стандартная форма представления данных - это CSV. А CSV - это форматированный текст, через запятую поля отмечены, он легко вводится в систему с помощью DBF CSV DBF конвертера. Вот в этом режиме 2322. То есть я прямо это сделал специально, чтобы это можно было использовать. А также, в общем-то, есть и онлайн-конвертеры, и Excel тоже это позволяет сделать, преобразовать, но только старый Excel. К сожалению, в новых экселях эту эту функцию убрали почему-то, посчитали, что незачем, не знаю я, как-то. Но новый Excel он может позволяет CSV файлы в экселевский файл преобразовывать, но дело в том, что экселевский файл имеет ограниченное количество строк, не более миллиона. А я вот этот конвертер, когда делал, я его недавно сделал, это сам конвертер, а ещё есть в самой системе, я вам показывал, в системе в режиме 2322, просто можно этот формат использовать CSV. И тогда он преобразуется в DBF, а DBF, а потом сразу передаётся управление на DBF, вот этот вот пункт. И, в общем, дальше идёт уже ввод данных.

И вот здесь вот у нас есть требования к файлу исходных данных. Требования очень простые: строчки - это наблюдения, первая колонка - источник данных, откуда взято это наблюдение (дата там какая-то или там поле или не знаю, в общем, откуда это вот эта строчка именно взята). Вот, можно просто пронумеровать 1, 2, 3 там наблюдения. Вот. И вот здесь вот у нас зелёненькие - это каждая колоночка такая - это свойство, а в клеточке - это значение этого свойства. Вот, или фактор и его значение. А это группировки, что это наблюдение, вот объекты с такими свойствами, э и объект относится к такой-то категории обобщающей, или классу. Вот колоночка - это категория, а клеточка конкретно - это уже значение этой категории, это класс.
Ну то есть здесь, допустим, может быть написано: фирма-производитель, а здесь будет написано: Mercedes, там, BMW, там, Volvo, там, MAN. Понятно? Да, всё понятно.
Вот и всё. Значит, могут быть числовые данные и текстовые в колоночках. Если есть хотя бы одно текстовое значение какое-то нечисловое, в смысле, я считаю, что числа - это тоже тексты, но не все тексты являются числами. Вот. То есть если там есть текст не число, то тогда остальные все числа будут тоже рассматриваться как текст. Вот. Такая вот деталь. Ну это понятно, наверное, потому что это же база данных, и если мы хотим данные представить здесь, которые приведены, то другого варианта просто нет. То есть поле либо числовое, либо текстовое. Значит, если там появился текст какой-то, то эти данные, если мы хотим ввести в систему, ну то мы их введём, сделаем поле текстового типа, но тогда и числа будут как тексты представлены. Вот. То есть если мы хотим, чтобы там числа как числа рассматривались, тогда там текстов быть не должно, ни чисел. Вот. Ну это понятно, да, наверное?

* **C. Требования к структуре данных:** (Продолжение из предыдущего раздела)

Вот. Ну и они обрабатываются немножко по-разному эти шкалы, числовые и текстовые. Значит, числовые, значит, текстовые просто производится сортировка с признаком Unique, то есть уникальные значения выделяются, формируется справочник этих значений. Ну, то есть для нормализации, для того, чтобы это значение один раз встречалось в базах данных и на него ссылки шли из других мест. А если числовая колонка, тогда сортируются тоже значения, находятся минимальное и максимальное значение, именно путём сортировки, потому что это гораздо быстрее происходит, чем другим способом. Одна команда языка программирования работающая оптимально. Вот. И потом этот диапазон изменения величины соответствующей делится на заданное пользователем количество числовых диапазонов, интервальных значений. Эти интервальные значения могут быть равные по величине с разным числом наблюдений или одинаковые по величине с примерно одинаковым числом наблюдений. Почему примерно? Потому что наблюдение всегда целое число. И там может быть, э, скажем так, в диапазоне или 30 наблюдений, или 31, а вот там 30,3 не может быть наблюдений, поэтому там на единицу может отличаться. Ну и всё, вот, собственно говоря, это делается. Потом вот в этом режиме мы указываем диапазон классификационных шкал, группировок, диапазон описательных шкал, указываем тип числовых диапазонов и указываем тип исходных данных, какой Excel, старый, новый, DBF файл или CSV. Если CSV, то тоже система преобразует, но бывает иногда изредка, что не получается, потому что эти вот CSV-файлы, они довольно разнообразные есть. И бывают файлы, где внутри поля есть запятые. Вот, то есть стандартом это поля отделены запятой друг от друга, поля, значения полей. Но иногда бывает, что там э внутри поля тоже есть запятые. И тогда это поле разбивается на два, происходит смещение, ну, в общем, короче, данные записываются не в то поле фактически, вот так получается. Вот. Ну, в общем, когда есть смысл использовать онлайн-конвертер. Вот так вот.

**VII. Обсуждение прогресса студентов**

Сейчас я небольшой опрос проведу, ребят. Отметьте, пожалуйста, в чате плюсиками, кто занимался выбором темы.
Вот, я вам давал, собственно, три задания: регистрироваться в РИНЦ, Science Index, заключать договор - это всё, что связано с РИНЦ. Первое задание. Второе задание, какое я давал? ResearchGate регистрироваться. И третье - выбирать тему и проверять, насколько хорошая модель получается. Если модель хорошая, посылать мне, я утверждаю. Давайте, плюсики ставьте, кто это делал. Можете написать: РИНЦ - да, там, ResearchGate - нет, тема - да, там, или что-нибудь такое. Давайте, быстренько, полный отчёт. Все, кто сейчас присутствует, кроме Георгия Александровича.
И даже не видно никаких попыток с вашей стороны даже написать там плюсики.
Так, тогда давайте, Елизавета Антоненко. Отвечайте мне на эти вопросы. Удалось вам что-нибудь делать в этой области?
*М-м.*
Понятно. Ну надо начинать, начинать надо, ребят.

**VIII. Пример обсуждения темы (Системы впрыска)**

*А можно спросить? Вот, ну...*
Конечно, мы у нас занятие именно этому и посвящено, ребят, чтобы спрашивать. Ну давай, давай, спрашивай.
*Ну вот, допустим, если взять такую тему, то есть АСК-анализ используемых систем впрыска топливной смеси автомобилей крупных концернов, выпущенных там за последние десятилетия, допустим, для самых основных там.*
А у вас данные есть по этой теме?
*Э, ну, конечно.*
То есть, ну, там, во-первых, куча видов самих впрысков, учитывая там центральный, наддроссельный...
Значит, всё это нужно попробовать. То есть если вот сейчас вы в чат или на почту мою пришлёте файл исходных данных. Значит, лучше или ссылку пришлёте в чат на источник данных, либо на почту мою сам файл. Сейчас я почту дам.
*Просто их нужно будет собрать сейчас исходные данные.*
А они где находятся? В каком виде? Разрозненном?
*Да, ну, вручную придётся повыбирать.*
Ну, ну давайте, выбирайте. Ну я тогда советую вам сделать небольшой объём там сначала, ну, строчек 10. Потом дать показать мне это. Вот. Это должен быть экселевский файл, в котором первая колонка - источник данных, ссылка на откуда взяты данные. Второй, вторая и последующие колонки там, несколько колонок могут быть классификационные шкалы, вот, которые, ну, группировки, скажем так, категории наблюдений. И потом колонки, которые описывают характеристики. Вот. То есть что вы хотите в этой задаче решить, какую какую, вернее, не задачу, в этом приложении, какую задачу вы хотите решить в этом приложении? Это нужно сориентироваться в этом плане, понять задачу. То есть я вот пока из того, что вы сказали, я, честно сказать, не совсем уловил, в чём суть задачи будет.
*Э, ну, то есть суть в том, что вот, допустим, у нас есть куча разных автомобилей, да, выпущенных за последние десятилетия. И каждый из них в своём двигателе использует разный тип впрыска топливной смеси.*
Ну да, и поэтому, потому какие типы впрыска, да? Что будем определять?
*Э, ну, допустим, можно, какие концерны, каких стран предпочитают разные системы, там, может быть. Вполне возможно, что так и есть.*
Ну вот, короче говоря, видите структуру файла какая? Строчки - это наблюдения, а колонки справа - это характеристики, свойства и значения, свойства и значения. Может быть числовое, может быть текстовое. Но в колонке должны быть, если числа, то текстов там быть не должно. То есть если данных нет, то ноль ставить. Вот. Ну, а текстовых данных если нет, то отсутствие данных, пробел, ну, даже не пробел, а даже отсутствие пробела. А классификационные шкалы, вы можете, допустим, взять год, например, год и э-э концерн. Ну, даже можно и марку, и модель указать машины. А дальше здесь вот уже характеристики э-э топливной системы. Там же не только будет тип впрыска, там же будет ещё что-то, наверное, ж, да? Вид топлива, например, там, может быть. Наверное, да. Потому что одной колонки это маловато как-то характеристик. Вот. Ну и здесь вот у вас в первой колонке будет написано э-э марка-модель, а потом будет написано, допустим, там, что опять та же самая, название концерна, производителя, да? И тоже, то же самое, может быть, марка и модель машины. А потом дальше вот эти характеристики топливной системы: вид топлива там, и что там оно, инжектор, карбюратор, да, вот такое что-то, да? И там у них каждого много есть разных ещё, видимо, особенностей каких-то, которые вы укажете. И вот собрать, ну, что я говорю, штук 10 строчек таких, прислать мне. Я посмотрю, в принципе, это разумно или нет, вот такая структура исходных данных. Если она разумна, я скажу: "Ну давайте теперь набирайте побольше таких данных". Если вы будете это делать вручную, то это, конечно, дело трудоёмкое, наверное. Но, может быть, там есть в одном источнике одни данные, в другом - другие. Как-то вы их объедините, не знаю. Вот. Но я просто о чём говорю? О том, что, в принципе, вот, может быть, проще взять и найти данные на каких-то вот источниках данных, наборах данных для машинного обучения. Вот, допустим, UCI, там, Kaggle. Вы смотрели эти сайты, нет? Может, там и есть что-то такое, что вам... А? *Пока что не смотрели.* Вот. Ну давайте смотреть.

**IX. Представление преподавателей**

Да, ребят, я когда вот говорил, что у нас такое занятие, то я упустил сказать, что занятие ведут два профессора: профессор Луценко Евгений Вениаминович и профессор Аршинов Георгий Александрович. Это тоже нужно говорить в самом начале, чтобы было понятно, что вот занятие такого-то числа, на такую-то тему, с такой-то группой.

**X. Напоминание о заданиях и сроках**

А теперь давайте мне отчёт. Тема, тема да, там, или тема нет. ResearchGate или РГ там можно писать, РГ да, нет. РИНЦ да, нет. Я буду вас спрашивать, а почему? Почему нет?
Потому что сейчас вот через 2 дня тут ноябрь будет, и потом будет сессия, понимаете? А сессия будет проводиться, то есть оценка ваша будет ставиться по тому, как вы сделали приложение, описание, где разместили. Вот на основе этого будет ставиться оценка. Пункт девять. То есть надо к тому времени быть зарегистрированным в РИНЦ, в ResearchGate, сделать работу и сделать описание по шаблону, ребята. То есть я вам не говорю, чтобы вы прямо сели там и выдумывали, что писать. Берёте шаблон и по этому шаблону пишете, но на свою тему. Я привёл вам ссылки на свои работы, которые можно использовать в качестве шаблонов, и на подборки работ по анализу текстов и анализу изображений. То есть любую можно брать шаблон. Ну вот эти вот первые здесь четыре приведены, они такие довольно новые, неплохие. Можно их брать за основу и пользоваться, делать.

**XI. Технические моменты и отвлеченная демонстрация**
*(Начало во время объяснения ограничений при выборе темы)*
Ух ты! Я даже и не знал, что такое бывает. Не думал, не думал, что такое может быть. Ага, а ну давай посмотрим.
*(Манипуляции с файлами, попытки открыть и проверить на FTP)*
Так вот, когда вам показываешь, и обнаруживаешь. Потому что я же всё это не в состоянии просмотреть, пересмотреть. Когда делаешь, то всегда какие-то есть нюансы.
*(Продолжение манипуляций с FTP, копирование файлов, проверка)*
Это сейчас вы видите FTP-сервер.
Вот, теперь обновляем это дело и смотрим. Тридцать четвёртая. Не обновилось. Вот такая особенность на этом Гугле. Почему-то он не особо-то и обновляется. Надо... Вот, смотрим здесь.
Что-то не очень получилось, ребята, честно сказать. Ну не очень я понимаю, почему, потому что я её заменил, обновил. Ну вот. Ну ладно, всё ж всякое бывает. Потом разберёмся, может быть, когда будет время.

*(Начало во время объяснения источников данных)*
Раньше у них был не очень удобный формат данных, такой свой специфический. Надо было определённые усилия предпринять, чтобы эти данные преобразовать в такую форму, которая является стандартной для программного интерфейса системы Эйдос. Эта форма несложная, ну экселевский файл просто, строчки - это наблюдения. Но тем не менее, нужно было определённые провести мероприятия, чтобы это сделать. Вот такого типа, видите, вот оригинальный вид исходных данных. Ну, можно используя Excel, вводить такие данные, потом точки заменять на запятую, вот эти единички, двоечки заменять на соответствующие тексты. Вот, и получить такой вид данных, который уже, вот я как раз это всё описываю довольно подробно, как это всё делается, вот в Экселе, видите? Разделители, получается вот такая таблица. Потом я какую-то программку сделал небольшую, чтобы руками не ковыряться, и значит, эта программка ввела данные, должна и получилась вот такая таблица исходных данных. Но я вам скажу, что очень сильно отличается от исходной по виду. То есть там были единички, нули, а здесь у нас тексты. Почему я это сделал? Потому что система обеспечивает обработку текстов. То есть в выходных формах будет написано прямо конкретно, а не будут там циферки какие-то, которые ещё непонятно, что значат, понимаете? Вот. И этот файл можно скачивать. Вот раз, вот он скачался. Сейчас попробуем даже открыть. Открылся. Вот файл исходных данных, видите? Так я же его преобразовал из того, который в UCI был, я сделал такой, который уже более разумный выглядит для системы. И во всех выходных формах вот эти все наименования будут. Два раза уже открылся, гарантированно.

*(Вставка видеоролика во время объяснения источников данных)*
Это вот прикольный клип про то, что как сделали бурду.
*(Воспроизведение видеоролика "Меланж Инферналь" про подделку вина)*
Ну как, ребят, понравилось? Вы это видели ж, да, наверное, так я понимаю? *Да, интересное видео.* Короче говоря, вот оценка Самелье и физико-химические свойства, они совпадают, ну, примерно там, скажем так, на три-четыре балла. То есть они откровенно плохое, откровенно хорошее и среднее отличают. А вот точнее им сказать сложно уже. Ну это вот видно из этой статьи, которую я вам показал своей. Оказывается, зависимость есть. Но то есть могло быть так, что и вообще её нет, да? Это один вариант, что там от фонаря эти Самелье оценивают. Другой вариант, что они очень точно оценивают, прямо вот как вот деталь, как вот физико-химические свойства меняются, они это сразу улавливают. И этот вариант тоже не осуществился, ребят. Вот видите, UCI ссылочка есть 2-3. Ну то есть здесь вот описано, как это всё делалось. То есть можете почитать и вмиг это у вас полное представление о том, как это делается. Здесь всё подробно описано, в каких режимах, каких, что как делается. Вот, допустим, режим ввода данных, вот видите, здесь все параметры указаны ввода данных. Всё, нули, пробелы считали значениями данных. Вот. Получилось у меня результаты такие, ребят, в этой статье, что самелье оценки зависят от свойств вина реальных физико-