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

***Российская Федерация***

**165 Теория информации, данные и знания. Лабораторная 7. спектральный аск-анализ 2020-10-27**

## Заголовок

АСК-анализ для ИТ2002: Задачи регистрации, выбор темы, работа с данными и демонстрация ПО Луценко

## Резюме текста

**1. Введение и организационные моменты:**
Занятие (лабораторная/практическая работа №7/№5) для группы ИТ2002 проводится 27 октября 2020 года профессором Луценко Е.В. по дисциплине "Теория информации, данные, знания". Тема: Спектральный АСК-анализ и инструкция по разработке собственных приложений. Уточняется статус занятия (лабораторная работа, несмотря на запись "практическое" в системе профессора).

**2. Постановка задач для студентов:**
Напоминается о необходимости регистрации в РИНЦ (включая Science Index для получения спин-кода и заключения договора) и ResearchGate. Студентам нужно выбрать тему для исследования, найти исходные данные (объемом не более 10 МБ) и представить тему на утверждение. Подчеркивается важность активной работы для выполнения 99 часов самостоятельной работы и сдачи экзамена.

**3. Обсуждение проблем с регистрацией:**
Обсуждаются трудности с регистрацией в ResearchGate, связанные с политикой университета по предоставлению корпоративной электронной почты (одна общая почта вместо индивидуальных). Профессор рекомендует старосте или куратору обратиться в ЦИТ для решения проблемы, так как регистрация необходима для учебного процесса. Отмечается низкая активность студентов в выполнении задачи по регистрации. Предлагается решение: включение профессора в соавторы для размещения публикаций тех, кто не смог зарегистрироваться. Упоминается необходимость заключения договора с РИНЦ для физических лиц для размещения публикаций, дается ссылка и пояснения по процедуре.

**4. Демонстрация работы системы (на примере данных Александра):**
Студент Александр предоставляет ссылку на данные (статистика оружия из компьютерной игры). Профессор демонстрирует процесс работы с данными в системе АСК-анализа:

* **Загрузка и подготовка данных:** Импорт данных, очистка (удаление знаков доллара), преобразование формата в числовой. Определение классификационных (текстовых) и описательных (числовых) шкал.
* **Настройка параметров анализа:** Установка параметров шкал (2-я классификационная, 3-16 описательные), использование адаптивных интервалов.
* **Запуск анализа (режим 2 3 2 2):** Формирование шкал, градаций, обучающей выборки.
* **Интерпретация результатов:** Анализ достоверности модели (очень высокая, >0.93-0.95 по критерию L1), визуализация результатов (классическая кривая, показывающая закономерность).
* **Утверждение темы:** Тема Александра утверждается для дальнейшей работы и описания по шаблону.

**5. Демонстрация работы системы (дополнительные возможности):**

* **Работа с данными (CSV, DBF):** Демонстрируется режим 2 3 2 10 для преобразования CSV в DBF. Подчеркиваются преимущества встроенного конвертера перед Excel (обработка файлов >1 млн строк, корректное распознавание данных без искажений типа дат/чисел). Объясняется структура файлов (indata.csv в папке indata).
* **Генератор случайных чисел:** Рассказывается об улучшении генератора случайных чисел (использование времени как внешнего источника энтропии для избежания циклов повторения).
* **Обновления системы:** Показывается, где найти информацию об обновлениях системы (файл Sheet Change) и как скачать обновления.

**6. Заключительные указания и важность самостоятельной работы:**
Профессор призывает студентов активно выбирать темы и данные к следующему занятию для их проверки и утверждения. Повторно подчеркивается важность выполнения заданий как формы отчетности по самостоятельной работе и для подготовки к экзамену, особенно в условиях дистанционного обучения.

## Детальная расшифровка текста

### 1. Введение и организационные моменты

Здравствуйте, ребята.
Здравствуйте. Здравствуйте.
Да, здравствуйте.

Сегодня 27 октября 2020 года. Четвёртая пара. 13:50-15:20.

Лабораторная работа номер семь с группой ИТ2002, на которой мы изучаем такой вопрос: спектральный АСК-анализ по дисциплине Теория информации, данные, знания. Занятие ведёт профессор Луценко Евгений Вениаминович.

Так, ребят, сейчас я уточню, что у нас за занятие. Отмечусь тут везде. Отмечу вас.

Практическое занятие номер пять.

Что-что?
Это лабораторное занятие.
Да, но у меня написано практическое.
Здравствуйте, Кирилл Станиславович, ещё раз.
Приветствую вас.
Нет, ну я сейчас вот посмотрел в этом, где мы отмечаем занятия. Вот, здесь. Здесь лабораторные все пройдены, вверху у нас есть только практическое занятие.

Что-что?
Если лабораторное ставится... практическое...
Ну, не знаю. У меня... Сейчас... Лабораторное, вот у меня там практическое стояло. А у вас, так сказать, сейчас, секундочку.
Ну, обычно, если в расписании стоит две фамилии, то значит лабораторная.
Мм, понятно. Ну тогда, значит, я неправильно отмечал когда-то в своё время. То есть я отмечал сначала лабораторные занятия, а сейчас вот практические остались. Лабораторные все у меня отмечены уже, что они уже пройдены. Ну, значит, я неправильно отмечал в этой системе, что там лабораторное, что практическое. Ну ладно, значит, у нас лабораторная работа.

Так, теперь что касается темы 202. У нас последнее занятие было — это инструкция по разработке собственных приложений. Правильно? Которое было 2 октября. Поэтому у нас и сейчас тоже продолжение этой же темы. Инструкция по разработке собственных приложений. Ну вот, разобрались.

### 2. Постановка задач для студентов

Так, ребята, ну и, значит, с чего мы начнём? С того, что... с того, что я на прошлом занятии, если у нас тема соответствует, там, она у нас соответствует, то на предыдущем занятии я вам рассказывал о том, что нужно регистрироваться в РИНЦ, ResearchGate и нужно выбрать тему, найти данные исходные, выбрать тему и предложить её до утверждения.

Вот. Значит, давайте мне информируйте, как у вас это получилось.

Ну, наверное, знаете как, давайте в чате, наверное, напишите там, где вы зарегистрировались, где вы там ещё не зарегистрировались. Ну, допустим, РГ - это ResearchGate, да, РИНЦ - да, или там ResearchGate - да, РИНЦ - нет, там. Вот такое вот пишите, значит, короче, в чате сейчас.

Вот. Ну Андрей вообще крутой. Ну не совсем, потому что тему не нашёл, да? И даже ты, Андрей, и в Science Index зарегистрировался, да? ResearchGate и да, РИНЦ да, РГ нет.
Угу.
То есть в РИНЦ имеется в виду, что и не только в самом РИНЦ, но и в Science Index, чтобы спин-код получить.

Ну хорошо. Вот Александр тоже написал, что где-то зарегистрировался, где-то нет. А вот что касается остальных? У нас тут всего сейчас 18 человек здесь на занятии.

Как остальные? Чем занимались? Только в РИНЦе.

Значит, нужно выбирать те, ребята, темы, которые по исходным данным не больше 10 МБ. Видите? Почему? Потому что 10 МБ - это максимальный объём, который можно разместить на FTP-сервере системы. Ну и вторая причина, что их легче обрабатывать такой объём данных.

### 3. Обсуждение проблем с регистрацией

Ну, в общем, я так понимаю, что регистрация в ResearchGate проблема, да? Вот, так что нужно, видимо, со старостой вашим... Кто у вас староста?
Там просто нужно разобраться с темой.
Герман.
Герман.
Ну да, Герман. Герман у нас староста.
Вот. Да, это понятно. Ну как с ней разобраться? Значит, нужно, видимо, чтобы ваш староста, а может быть, лучше куратор, обратился с этим вопросом к руководству ЦИТа, что нам это необходимо для учебного процесса.

Ну и всё равно очень мало здесь ответили мне. Ответили шесть-семь человек из 18.
А остальные вообще, что, не слушали, что я спрашивал, что ли? Или вы нигде не зарегистрировались, ребята? Ни в РИНЦе, ни в ResearchGate, да?
Мы со старостой сегодня были у секретаря. А он сказал, что у нас будет одна корпоративная почта на всех. У нас типа индивидуальной не будет.
Ну, с одной вас вряд ли дастся зарегистрироваться, потому что вам ResearchGate напишет, что эта почта уже использовалась и всё.
Вот поэтому и проблема.
Ладно. Будем исходить из того, что есть.

Ну что, ребята, я вам посылал вот эту страничку, адрес странички с заданием. Вот, по-видимому, придётся вам включать меня в соавторы, чтобы я вас там ваши публикации размещал.

При этом я могу сказать, что угрызения совести у меня по этому поводу не будет, потому что я разработал инструментарий, который вы используете для решения задач, учу вас этим инструментарием пользоваться и учу описывать результаты исследования, моделирования, и учу ещё и размещать, но размещать, видите, не очень получается, потому что у нас почему-то вам не дают адреса электронной почты. В отличие от, скажем, КГУ, где у всех студентов есть адреса электронной почты корпоративные, абсолютно 100% у всех. Я не понимаю, в чём проблема. Дело в том, что эта почта гугловская, она никакой нагрузки на наш сервер не оказывает. То есть это то же самое, что вы сейчас возьмёте в Гугле зарегистрируетесь, понимаете? То есть абсолютно никакой роли это не играет для нашего университета. В чём проблема, я не знаю.

Так, ребята, значит, с этим мы разобрались. Тогда, значит, будете включать в соавторы те, кто не смог зарегистрироваться в ResearchGate и РИНЦ. То есть и там и там. Потому что, если, допустим, вы в РИНЦ зарегистрировались, а ResearchGate нет, то всё равно придётся мне размещать в ResearchGate. И в РИНЦ вы можете сами, если там уже зарегистрировались, Science Index тоже, и спин-код получили, договор заключили, то можете там тоже разместить. Но я так чувствую, что пока договор никто не заключил с РИНЦем, да? Ну хотя бы напишите, у кого спин-код уже есть.

Есть, есть. Ну человека два-три, может быть, и напишут сейчас. Ну напишите. Есть спин-код РИНЦовский получен.

Ну это в лучшем случае, Александр, так говорят. О, Андрей написал: есть. Молодец. Значит, теперь тогда бери вот по этой ссылочке, которую я указал здесь в инструкции. Переходи по этой ссылочке. Вот. Один человек. Вот, видите? Надо заключить договор теперь на физическое лицо на размещение непериодических изданий в РИНЦе. Вот ссылочка я дал. И показал, откуда эта ссылочка взялась. Вот теперь туда выходите и начинаете заключение договора на физическое лицо. И выполняете там этапы, которые там предусмотрены. Там много разных этапов. Сначала свои данные заполняете, посылаете на проверку, потом они там подписывают со своей стороны, потом вы подписываете, делаете PDF-файл со своей подписью. Можно любым способом, можно взять Adobe Acrobat Professional, там вставить подпись, а можно, значит, просто распечатать, подписать, отсканировать в PDF и послать им. Понятно, да? Вот. И, в общем, другими тоже вариантами, какими там придумаете, можно. Пин-ридером отсканировать либо в Word, подписать, обратно в PDF им послать. Ну тогда может исказиться там что-нибудь по форматированию. Вот. Ну, в общем, короче говоря, делаете договор со своей подписью, отсылаете им из их из их договора, который уже подписан. И, в общем, у вас получается, что вы заключили договор. После этого вы можете там размещать публикации. Те, которые, допустим, размещены в ResearchGate, например, вами же. То есть замечательная такая возможность, которую очень желательно вам освоить эту возможность.

Ну, у кого не получится, значит, тогда придётся мне помогать. Я прошлые годы помогал студентам это делать. Ну как помогал? Ну просто брал, да и делал.

### 4. Демонстрация работы системы (на примере данных Александра)

Так, теперь давайте с темами разбираться. Значит, у кого есть какие-то варианты уже выбранных тем?
Кто-нибудь выбрал тему или нет?

Вот Александр прислал исходные данные. Сейчас мы попробуем с ними сделать задачу.
Александр, расскажи, в чём суть этой задачи?
Ты микрофон есть?

Ага. А какой компьютерной игры?
Ну напишешь. Здесь дело в том, что здесь такой странный какой-то чат в этой Вебексе, что оттуда ничего блоком не вытащишь. Хотя вроде отмечается блок, но... Что это такое вообще? Может... Может она конвертировать? Сейчас я покажу, ребята. О, господи. Ну в общем, не буду я сейчас это пытаться вставить, потому что это как-то тяжело. Конкретно.

Вот. Теперь смотрим, какие у нас колоночки. Значит, у нас вторая колоночка классификационная, я Shift держу и вправо двигаю курсором. А всего у нас 16 колоночек. То есть у нас классификационные шкалы со второй по вторую, а с третьей по шестнадцатую — описательные шкалы.
И запускаем систему и смотрим, что у нас получается. Удаляем все приложения. Это приложение формируем, то есть 2 3 2 2. Указываем новый Excel. Со второй по вторую, с третьей по шестнадцатую. Я вам советую, ребята, всегда ставить адаптивные интервалы, то есть разные интервалы, сравнимые с числом наблюдений. И вводим данные. То есть мы вводим прямо в этом режиме 2 3 2 2, без использования то есть не в лабораторных работах это делаем, а прямо делаем тут.
Теперь смотрим, ребята, а что у нас там такое? Классификационная шкала у нас — это текстовая шкала, потом идут числовые шкалы.
Вот. Числовых шкал, а можем задать побольше интервальных значений. Ну сначала давайте посмотрим, что там при таких получается. Значит, вот у нас в классификационных шкалах четыре градации. И по 11 градаций в описательных шкалах. Это нормально.
Выходим на создание модели.
У нас формировались классификационно-описательные шкалы и градации. Одна классификационная шкала и сколько здесь? 14 описательных.
Вот. И обучающая выборка, состоящая из небольшого числа 34 примера.
Выходим на создание модели. Все параметры по умолчанию.

Вот у нас проходит 10 секунд. Все 10 моделей посчитаны и проверены на достоверность.
Смотрим, какая достоверность модели. Она оказалась очень высокой. Значит, по разным интегральным критериям, в разных моделях 0,931, 0,955. То есть это исключительно высокая достоверность модели.
Вот. Смотрим достоверность модели, Александра. Вот выбираем модель, которая наилучшая является по критерию L1. У неё 0,955 достоверность по критерию L1. Это очень высокий уровень достоверности.
Получаем замечательную классическую картиночку. Вы смотрите, нет сейчас, Александр?
То есть у вас очень хорошо получается всё. Модель очень высокая достоверности. И чётко совершенно видно закономерность, что ложные решения есть при очень малых значениях, потом их становится гораздо меньше при увеличении значения уровня сходства, а истинных решений становится гораздо больше. То есть это классическая кривулечка, замечательная на вид, которая показывает, что выше 30% уровня сходства у вас все решения истинные, практически так. То есть вы получаете критерий. Дальше вы можете спокойно решать задачу, описывать решение задачи по шаблону.

Слышишь, Александр? Тема у вас утверждена, можете начинать описание её по шаблону.
Хорошо.

### 5. Демонстрация работы системы (дополнительные возможности)

Вот. Теперь вопрос возник. Может ли система преобразовывать, да?

**5.1 Работа с данными (CSV, DBF)**

Значит, я сейчас вам покажу, как это делается. Смотрите, ребят. Берём CSV-файл, значит, ну вот сначала я я закрываю этот файл, который вот прислал Александр.
Александр, я вам советую сделать вот так, как я сделал. То есть убрать этот доллары эти, и категории, значит, это классификационную шкалу выделить жёлтым фоном, а также категории можно поставить вертикально. Тогда у вас таблица сузится по размеру, будет очень узенькая. Вот, и вы можете её поместить прямо в статью целиком эту эту таблицу в таком вот виде.
Видите, да, ребята? Александр, видишь, нет?
В таком виде, Александр, помещаешь статью.

Так, теперь смотрим, что касается CSV. Значит, я как раз недавно поработал так довольно-таки плотно, сделал свой конвертер, который намного превосходит Excel по возможностям. В чём именно он его превосходит? Вот я сейчас беру и CSV-файл копирую в папочку indata.
Значит, имя файла должно быть indata, in\_data.csv. Он должен находиться в папочке indata. Путь у него должен быть вот такой.
Я вам показываю вам.
Вот, то есть вот этот CSV-файл должен быть в этой папочке, ну там, где система на вашем компьютере, потом indata и вот, и потом он должен иметь имя этот файл indata.csv.
В чём превосходит система Excel в плане преобразования CSV-файла и его использования? Значит, Excel не может больше тысячи, то есть больше миллиона строк использовать. Я проводил отладку на 1000 строк вот на этом файле. Вот этот файл небольшой, 1000 строк. Но потом, когда провёл отладку, то, значит, я экспериментировал, пробовал обрабатывать, преобразовывать файлы размером 4 млн записей. То есть в четыре раза больше, чем позволяет Excel.

Теперь я показываю, как можно это делать в системе. Значит, вот мы в системе находимся, я удаляю приложения все, по которому Александру мы сейчас посмотрели. Заходим в режим 2 3 2 10, ребят, смотрите, 2 3 2 10. Этот режим обеспечивает, ну здесь когда вот мы курсор ставим на этот режим, то под этим ниспадающим меню появляется Help, такой Help плавающий, подсказчик. Что этот режим делает? Он преобразует вот такой файл, там полностью путь указан, в другой файл. Этот файл не является экселевским файлом, это файл DBF, база данных. Но Excel 2003 его спокойно открывает, 2010 не может. Значит, но система с такими файлами работает. То есть его можно и в Экселе не преобразовывать.
Вот. Значит, смотрим, значит, теперь да, онлайн-конвертеры есть DBF в Excel. То есть можете онлайн-конвертер использовать, он вас конвертирует.
Но чем этот хороший этот конвертер, который в системе? Когда Excel преобразует CSV-файлы в экселевские, это делается в новом Экселе, начиная там с 2007, 2010, начиная с этих версий. Это делается во вкладочке Данные, ввод из текстов. Данные, ввод из текстов.
Значит, при этом что у нас получается? Сейчас вот я запущу Excel новый и покажу вам, как это делается.
Вот смотрите. Значит, я беру вкладочку Данные, из текста и показываю, где этот текст находится. Он находится в системе в папочке Data. Вот. Спрашивает Excel, а как мы этот файл где там у нас колоночки, как мы будем форматировать? Я перевожу дальше, следующее и указываю, что разделитель колоночек у нас запятая. И готово. У нас получается тут формула и получается экселевский файл. Смотрим, что у нас тут получилось. Смотрите, ребята, в девятой строчке у нас написано август 80. Видите? А в строчке двадцать второй написано июль 80. А, значит, в строчке тридцать шестой написано ноябрь 00. И так далее. Я даже не буду всё это показывать. Нас это не устраивает, понимаете? То есть этот такое преобразование нам не нужно.
Поэтому я и сделал этот вот конвертер. Это первая причина.
Вторая причина такая, что он может преобразовывать CSV-файлы, в которых, ну, примерно до 12 млн строк. Ограничение связано с чем? С тем, что DBF-файл, который получается в результате, он должен быть меньше 2 ГБ. Я проводил эксперимент вот с этим 4-миллионным CSV-файлом, он полчаса преобразуется. Вот. Но там при том числе полей, которое там было, там получился файл DBF-файл DBF 680 МБ. Ну я из этого заключил, что можно в три раза больше примерно сделать при таком числе полей. Ну это означает 12 млн строк можно преобразовать.
Вот. Как это сделать? Давайте, значит, заходим в режим 2 3 2 10. И просто нажимаем, и всё, вот оно преобразуется. Все дела. Вот вот 1000 строк преобразовались за 1 секунду. Смотрим, что у нас там получилось, ребят, в папочке, в папочке, где был исходный файл. Там появился DBF-файл. И ещё input name файл. Input name он показывает, какие там имена у нас полей. Вот. А DBF-файл просто открывается в Экселе. Я обращаю ваше внимание, ребята, DBF-файл открывается в Экселе. Но не в любом Экселе, а только в третьем. И более ранних, более поздних почему-то они посчитали, что это не нужно. Меня это несколько удивляет, что они это посчитали. Вот. Ну ладно. Как посчитали, так уж и посчитали. Мы-то на них не можем повлиять.
Вот. Вот у нас эти поля. Здесь никаких в девятых строчках, никаких дат нету, видите? Написано нормальное число, как и в остальных местах. Вот. Теперь этот файл спокойно можно записать как экселевский, если есть желание. Но в системе нет необходимости использовать только экселевские файлы для ввода информации.

Значит, здесь у нас, я вам скажу, у нас описательная шкала, все описательные шкалы со второй, даже по с четвёртой по двадцать восьмую. А классификационная шкала у нас только одна, двадцать девятая. Вот. Значит, почему со второй, не со второй, а с четвёртой? Потому что вот эти две колоночки, они нам не нужны, они нам не интересны. Значит, здесь уникальные значения, идентификаторы пользователей. Вот. Поэтому что мы делаем? Мы заходим, значит, я оставляю один CSV-файл в папочке исходной. Да, мы можем вот так посмотреть, для чего этот конвертер используется 2 3 2 10. Чтобы посмотреть на этот файл и понять, где там у нас классификационные, где описательные шкалы. Ну это можно делать и без преобразования. Вот мы CSV-файл открываем, берём все колоночки вот так выделяем блоком, заходим в Word и создаём туда переносим через буфер обмена. Потом берём, заменяем запятую, запятую, заменяем на конец знак абзаца. Получаем все поля, которые там есть в этом CSV-файле. Нам понятно, что вот это поле двадцать девятое — это классификационное поле, шкала. А остальные вот, начиная с четвёртой — это описательные. То есть можно не преобразовывать DBF в Excel, чтобы узнать вот это, какие поля у нас, какие колонки будут у нас классификационные, какие будут описательные. Но можно преобразовать, чтобы лучше видно было.

Теперь дальше что мы делаем? Как мы вводим информацию в систему? Вот мы узнали, что у нас классификационное поле шкала — это двадцать девятая. Теперь мы заходим в 2 3 2 2 режим. И там у нас, обратите внимание, есть у нас типы данных какие? Excel старый, Excel новый, Dbase есть. То есть мы можем Dbase преобразовать вот этот CSV-файл, и его использовать как входной. А можем использовать CSV-конвертер, CSV Dbase-конвертер. Здесь есть помощь, здесь описано вот то, что я сейчас вам рассказываю, что этот файл должен иметь какое-то имя, он должен находиться в такой-то папочке.
Иногда бывают файлы CSV, которые формально являются CSV-файлами, но фактически испорченными являются CSV-файлами. То есть они испорчены, формат нарушен. В чём это может заключаться? Значит, я сталкивался с тем, что там, допустим, текст идёт некий, и в этом тексте есть запятые. И этот текст, по идее, он должен быть значением какого-то поля определённого, ну, скажем, второго поля. Я смотрю, там сдвиг идёт по полям, данные сдвинуты. Я думаю, ну что такое? Потом смотрю, а в этом первом поле CSV-файла две запятых есть. На основе этого поля создано две две колонки. То есть оно помещено часть первая часть до первой запятой помещена в правильное поле. А вторые две вот этих вот фрагмента текстовых помещены в третье поле и в четвёртое, во второе, третье, четвёртое. А потом всё дальше тоже идёт правильно, только оно сдвинуто на две колонки, понимаете? Это испорченный CSV-файл. Я такие советую просто не использовать, это, ну, это глупость полнейшая.
Вот. Значит, что мы сейчас делаем? Смотрите, я указываю классификационная шкала двадцать девятая, двадцать девятая, потому что мы посмотрели это в самом э-э строке наименования полей и э-э посмотрели в DBF-файле, когда его преобразовали вот в этом режиме 2 3 2 10. Вот. Ну а здесь что мы делаем? Мы просто нажимаем О'кей. Вот я указал CSV-формат, классификационные шкалы, описательные шкалы, нажал О'кей. Произошло преобразование файла CSV в DBF. Оно завершено полностью. Размер файла меньше 2 ГБ. Если бы файл получался больше 2 ГБ, то об этом было бы сообщение, что процесс преобразования прерван, э-э потому что файл получился больше 2 ГБ. Вот. И дальше мы смотрим. Э-э дальше происходит передача управления, ребята, на э-э режим 2 3 2 2 с параметром входные данные из DBF-файла. Считываем. Смотрим здесь, как получается. То есть, как нам нравится, не нравится. Вот. Потом выходим на создание модели.

**5.2 Генератор случайных чисел**

Что сейчас происходит? Сейчас созданы классификационно-описательные шкалы и градации, с их помощью закодированы исходные данные, получилась обучающая выборка. Это заняло 15 секунд. То есть мы ввели данные из CSV-файла и создали прямо сразу модель. Вот шкалы создали. Вот, пожалуйста, они. И обучающую выборку. Всё есть.
Теперь можем дальше создавать модель и смотреть на её достоверность. 999 объектов распознаваемой выборки. Ну это вот уже нужно только для того, чтобы посмотреть, какая достоверность, и дальше потом анализ проводить. Вот. А само преобразование я вам показал уже, ребят.
Вот формирование массива случайных кодов объектов выборки без повторов из N элементов. Значит, вот do while, я на смотрю, э-э, ну, выбралось нужное количество элементов или нет? И сравнивает это с нулём массива. Значит, ну в массив этот запоминаются коды тех вот уникальных объектов обучающей выборки. Если, значит, ещё меньше он, тогда крутится, ещё находит. Вот. И суммирует постоянно счётчик тут, значит, и в, я сейчас уже не понял, зачем он нужен. По-моемому, он уже и не нужен, можно убрать. Вот. Так вот, значит, что я сделал? Смотрите, ребята, интересный момент. Я взял э-э время, э-э, значит, в сотых долях секунды, начиная с начала суток. Это компьютерная, то есть это команда языка программирования, которая возвращает время в виде числа, возвращает время, прошедшее с начала суток в долях секунды, в сотых долях секунды. Я умножил на 100, чтобы целое было число. Преобразовал в текст, убрал пробелы, нашёл там э-э в этом тексте э-э точку, вернулся на один назад. А там получается э-э формат этого числа всё равно получается точка 00 теперь, после того, как я на 100 умножил. Там просто в конце будет два два нуля, но всё равно точка есть. Вот. И я эту точку нахожу и возвращаю э-э фрагмент этого текста до точки. Вот. Потом преобразую э-э, а дальше слушайте внимательно. Беру три последних знака. То есть длина этого текста с числом сотых долей секунды. Беру, возвращаюсь на два знака от конца назад. Получается три знака последних. Если вы проверите вот на бумажке, допишите там, посмотрите, получается три знака. Эти три знака я преобразую, ребята, вытаскиваю из этого текста, преобразую в число. Мне получается число э-э, значит, которое э-э постоянно меняется, потому что там сотые доли секунды, прошедшие с начала суток, оно постоянно разное это число, понимаете? Каждое следующее обращение - это другое число. Вот. Это число, а вот это было раньше вот у меня. Э-э то есть я просто брал и вычислял случайное число вот это вот рэндом от нуля до единицы, там оно так напишешь, будет до 999. Э-э равномерно распределённое. Так вот этот генератор оказался такой слабоватенький. Я значит, что сделал? Смотрите, я взял и прибавил вот это вот время, э-э которое тоже меняется от нуля до 999, три знака могут изменяться. Умножил вот на этот, на эту случайную величину и разделил на их квадрат, на квадрат 999. Вот это 998 01 - это квадрат 999. И получилось, получился некий коэффициент, который меньше э-э единицы всегда. И может быть, максимум равно единице. И умножил его на число объектов обучающей выборки. Вот этого числа объектов обучающей выборки. Потом проверил всё это, что всё это действительно вот считается правильно, получается. А потом проверил, э-э есть ли вот в этом массиве э-э, который сначала пустой этот массив, он в начале там опустошается. Вот же много разных способов выборки. Вот. Ну, обнуляю, сбрасывается этот размер этого массива нулевой. Добавляется туда, значит, проверяется, есть ли там этот код, который обнаружен таким образом? Ну, поскольку здесь может быть э-э начинаться с нуля вот эта величина, то я единицу добавил. Вот. И если там он есть, ребята, тогда э-э, значит, если там его нет, вернее, вот если ноль - это значит, если там не найден этот вот код в массиве, тогда он добавляется к массиву. Это динамический массив, у него э-э размерность меняется. Добавляется к нему этот элемент. И потом конец цикла. В результате она э-э раньше она висла там на 1.000, например, на на 5.000. А сейчас она на 10.000, она раз выскакивает и дальше работает всё, понимаете? Значит, при этом у нас, я называю это так. Это генератор случайных чисел э-э с внешним источником энтропии. Работает великолепно, ребята. Это я сделал вот двадцать пятого числа. Это достаточно такое новое решение, ребят, интересное. Нетривиальное. Вот. Имейте в виду, что если вам надо будет хороший генератор случайных чисел, то можете использовать эту идею. Генератор Луценко.
Вот. А ещё я, значит, э-э пробовал генератор фон Неймана. Ну это, конечно, оригинальный парень этот Нейман. Вот. Смотрите, что он придумал. Генератор фон Неймана. Значит, ну, я его немножко усовершенствовал. Он берёт, э-э, что делает? Он берёт, э-э, умножает какое-то число, э-э, или там возводит его в квадрат. Вот. А потом берёт четыре знака из этого числа из э-э серединки. И использует его, значит, как случайное число. Зацикливание в случае, если очередное число будет равно нулю. Я проверяю, равно оно нулю? Нет, не равно. Ну тогда я его беру. Если равно, то я То есть я обошёл этот этот недостаток генератора фон Неймана. Вот. Но я сравнивал свой генератор с генератором фон Неймана, мой гораздо лучше. Значит, я использовал его э-э фон Неймана, э-э первое число, которое там вот задаётся, вот. S random. В качестве него я использовал время. Вот такие тут фокусы. Вот. Но потом я посмотрел, м-м что-то как-то оно э-э, ну, в общем, идея хорошего Неймана, но у меня всё равно лучше. У меня вообще идеально работает. И быстро, и никогда вообще, в принципе, не повторяется. То есть там нет циклов этих вот повтора, понимаете? Их там быть не может, в принципе, этих циклов.

**5.3 Обновления системы**

Значит, э-э есть список изменений. Значит, когда вы устанавливаете новое обновление, скачиваете обновление, вот сейчас я вам покажу, вот файл обновлений. Значит, вот э-э вот этот файл обновления скачивается, когда вы старт запускаете. Вот этот файл скачивается, если его дата более новая, чем дата исполняемого модуля. И запускается на разархивирование. Вы получаете вот эти файлики заменяют те файлы, которые там у вас есть. Вот я, допустим, изменил э-э этот алгоритм, добавил здесь кое-что, исправил, вот сюда стрелочку поставил. Я его изменил. И вы, когда скачиваете обновление, то, соответственно, у вас появляется новый этот алгоритм. То же самое касается вот э-э преобразования вот этой схемы. Видите, я вот здесь вот это добавил снизу сопоставление. Здесь названия немножко поменял. Здесь поглубже названия. Очень интересный смысл там заложен. Э-э я вот здесь вот добавил стрелочки вот здесь и нарисовал, написал, чем эти модели э-э как они соотносятся эти модели. Вот чем, например, сходны, отличаются идентификация, прогнозирование, прогнозирование, принятие решений. То есть я нарисовал связи между этими вот задачами, которые здесь. И объяснил, почему необходим вот этот развитый алгоритм. Потом сами вот эти вот файлы, исполняемый модуль, э-э исходные тексты я заменил. А вот этот файлик Sheet Change, Sheet Change. Вот этот файлик, он содержит все обновления, начиная с момента создания системы. Значит, э-э на 31 августа 2012 года, э-э, значит, э-э я сделал вот это. И вот дальше потом пошло, пошло, пошло, пошло, пошло, пошло, пошло. Все э-э существенные изменения я записываю. Есть иногда мелкие изменения, которые я не отражаю. Ну там, допустим, в каком-то слове буква пропущена, вот такое вот. Я иногда замечаю, что там не всегда бывает э-э идеально с точки зрения грамматики, орфографии и грамматики, понимаете? Вот это жёлтым - это выделен этап очень существенный, когда начали использоваться расчёты на графическом процессоре. Вот это этапы, вот эти шестнадцатый год, э-э когда э-э, значит, а, я использовал мультиязычную поддержку, э-э, значит, э-э ну, реализовал там, использовал сначала Google, потом он стал платным, я перестал его использовать. Вот смотрите, ребята, вот это сегодняшнее изменение, вчерашнее, вернее. Я довёл до ума CSV-конвертер. Я начал его делать э-э где-то здесь ошибка, по-моему. Конечно, ошибка. Вот. Я начал делать, делать девятнадцатого октября. И вот, в общем-то, в черне сделал его, но потом находил там неточности. Даже тут ошибки орфографические есть в самом файле этих изменений. Потом просто его улучшал, доводил до ума. И кое-что ещё делал. Хелпы там скорректировал кое-где там, э-э лучше сделал визуализацию э-э прогноза и времени исполнения улучшил. Раньше я когда-то, когда делал эту визуализацию, то там суть была в том, что э-э заданное число э-э изменений изображения было. Ну, скажем, 1.000 изменений изображения. Я предусматривал 1.000 изменений изображения. И этого, в общем-то, как правило, было так более-менее. Но потом я увидел, что иногда вообще оно не меняется, иногда меняется там через какое-то время, через 5 минут, например. И скачкообразно так меняется. Короче, в зависимости от объёма данных, это получалось э-э не совсем всегда. Обычно так более-менее, но не всегда, понимаете? В некоторых случаях вообще-то ничего не ползёт там, некоторых э-э, ну, в общем, короче говоря, я потом это всё переосмыслил, переделал. И сейчас, независимо от объёма данных, изображение меняется 10 раз в секунду. Вот эти прогресс-бары, всё-всё, вот все числа, которые там прогнозы, они все меняются 10 раз в секунду. Независимо от того, какой объём данных. Стало работать намного приятнее всё это выглядит. То есть он так ну видно, что работает, видно, что всё делается, так сказать. Но я могу вам сказать, что процессы вот эти вот визуализации э-э, скажем так, уровня процесса исполнения и прогноза времени исполнения, они где-то процентов, ну, чтобы не соврать, ну где-то на 15-20% замедляют работу системы. Вот. Но я этим пожертвовал. Почему? Значит, там получалось как? Если я сделаю 1/100 секунды, то она в два раза увеличила время работы, понимаете? Если сделаю одну раз в секунду визуализация изображения меняется изображение, то тогда получалось, что практически не замедляется работа, но дёргается всё, понимаете? Короче говоря, вот я остановился на 1/10. Оно немного замедляет, но как бы не принципиально. То есть там было, скажем, полчаса, стало полчаса там 6 минут, например. Ну, замедляет, но в пределах разумного. Зато хорошо очень понятно, когда это всё закончится там и так далее.
Вот. И этот файл, э-э, Sheet Change, он есть и его можно и в системе тоже посмотреть. Я когда-то делал так, чтобы можно его и с системы посмотреть. Вот, по-моему, где-то здесь вот. Патенты, монографии, учебные пособия, статьи, архив, текущие обновления. Вот можно обновления скачивать всегда вот здесь. Ссылки здесь действующие. То есть они реально э-э работают.
И информация об обновлениях системы. Вот есть, смотрите. Тут написано, что он 560 КБ, но я в этом не уверен. Сейчас мы посмотрим, сколько он там килобайт. 700 КБ. Надо было измерить его размер, здесь написать. И вот мы этот файл тоже можем скачать и открыть. Что-то как-то она не особо-то открывает. Ну открыла. Вот этот файл. Ну здесь видно, что так достаточно систематически я работал над ней, добавлял там что-то совершенствовал. Довольно-таки много всяких неточностей. Иногда вроде как сделано правильно, но потом я догадывался, что можно гораздо лучше сделать.

### 6. Заключительные указания и важность самостоятельной работы

Ребят, значит, я сейчас открыл рабочую программу по вашей дисциплине. И могу вам сказать, что у вас экзамен. Понимаете? Экзамен по дисциплине. И самостоятельная работа 99 часов. Как вы собираетесь эту самостоятельную работу закрывать? Скажите мне. Такой объём 99 часов. Как вы её закроете? Чем?

Ребят, ну хорошо, что кто-то отвечает. Вот Глеб ответил, что пока в процессе выбора. Время-то идёт. Вот сейчас занятие, на котором я мог бы помочь вам посмотреть модель, понимаете? Если кто-то из вас выбрал бы сейчас вот тему, я бы помог вам её посмотреть. Вы такой возможностью не используете. Потом у вас этой возможности не будет, просто не будет времени. Вас же там много. Понимаете?

Значит, давайте сейчас тогда ищите тему. Давайте сейчас.

Ребят, что вы там, заснули, что ли, все? Давайте ещё темы выбирайте. Будем проверять.

Чтоб вы имели это в виду. И потом не говорили, что мы вам, вас не предупреждали. И ещё вам хочу сказать, ребята, что сейчас в этих условиях дистанционного обучения, мы не видим, как вы работаете. Вы сейчас по идее должны сидеть в аудитории и работать, понимаете? Мы этого не видим, как вы работаете. Работаете вы сейчас на своих компьютерах или нет? У меня есть очень большое подозрение, что у вас их рядом нет компьютеров, что вы на них не работаете. Понимаете? А должны. На лабораторных и практических занятиях вы должны работать на компьютерах, выполнять задания. Тогда было очное обучение, мы ходили с Геосанычем, мы видели, как вы работаете. И, в принципе, мне этого было достаточно для того, чтобы поставить вам аттестацию. Я понимал, кто из вас хорошо занимается, кто, так сказать, не особенно. Вот. Сейчас же мы этого не видим. И поэтому для того, чтобы оценить, какие у вас знания, умения и навыки, получили ли вы нужные компетенции, вы должны выполнить задание. Понимаете? То есть вам даётся задание для того, чтобы продемонстрировали, что вы это время не профуфукали, так сказать, впустую, а сидели, что-то делали, занимались, учились. Огромное количество времени отводятся на лекции, лабораторные, на самостоятельную работу 99 часов, ребята. Попробуйте 99 часов посидеть, поработать. Это много времени. Вы 100 раз это всё сделаете.

Так, ребята, пожалуйста, ещё вопросы.
А что это за хостинг?
Byhost? Американский хостинг, в смысле, да, этот? Сейчас. Сейчас, сейчас, я попробую найти. Да, по-моему, так и называется, Byhost. Его перевести, что это значит. Вроде бы как это значит драться там, ударить, вот такое что-то. Ну я зарегистрировался, взял тариф бесплатный и на SSD дисках пишут. Сейчас я вам ссылочку дам. Я разные видел хостинги, и когда-то там один был лучше, когда-то другой, когда-то третий. Времена идут, всё меняется. Раньше когда-то Яндекс был самый хороший хостинг бесплатный. Потом что-то там у них в голове там произошло, стали заставлять нас делать сайты с их инструментами. Кому они нужны, я не знаю. Совсем надо быть сумасшедшим, чтобы их инструментами пользоваться. Ну, короче говоря, потом они перестали разрешать делать, размещать файлы, на которые нет ссылок. Ну, в общем, стали выпендриваться, так скажем, по-простому. Ну я взял и бросил его просто, перестал пользоваться. Ну сейчас вот этот Byhost, он, по-моему, такой один из лучших. У него говорю, таких явных нет. Единственный недостаток, что вот не больше 10 МБ файл. Вот если бы они убрали, ну сделают не больше 100 МБ, например, ну было б тогда хорошо. Вообще б тогда никаких вопросов не возникало бы. Но они вот что-то как-то вот пожмотились чуть-чуть. Хотя пространство не ограничено.

Ну что, на этом, ребята, занятие заканчивается. Я очень прошу вас всё-таки заняться регистрацией плотно и получением спин-кода, заключением договора и выбором темы. Очень желательно, чтобы на следующее занятие, когда придёте лабораторное или практическое, чтобы вы имели вариант темы. Мы посмотрим. Если будет хорошо получаться, то тему утвердим, и вы будете описывать дальше уже. И это у вас ваше будет задание, выполнив которое вы получаете самоэкзамен. А не выполните, ну тогда тяните билет. А можно ещё один? Можно. А можно ещё один? Можно. А можно ещё один? А можно уже и не тянуть, потому что каждый раз на один балл снижается, понимаете, когда вы тянете билет. Ладно, всё, не буду вас пугать. До свидания, ребята, всего самого хорошего вам.
До свидания. До свидания.
До свидания.
До свидания. До свидания.