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

***Российская Федерация***

**195 Лекция №11. По дисциплине "Теория информации, данные и знания". 2020-11-10**

**Лекция по теории информации: Меры, энтропия, энергия, тестирование и разработка ЭДС**

**Резюме**

**I. Введение и детали сессии:**  
Лекция №11 по дисциплине "Теория информации, данные, знания" состоялась 10 ноября 2020 года. Ведущий – профессор Луценко Евгений Вениаминович. Занятие предназначено для групп ИТ-2001, ИТ-2002, ИТ-2003.

**II. Обзор предыдущей лекции и введение в текущую тему:**  
Продолжается изучение темы 1.11.1 "Взаимосвязь материи, энергии, информации в природе и обществе". Ранее были рассмотрены меры информации Больцмана, Хартли, Шеннона и их взаимосвязь.

**III. Объявление о предстоящем тестировании и обсуждение:**

* **A. Первоначальное упоминание:** Преподаватель напоминает о предстоящем тестировании.
* **B. Запрос студента и уточнение:** Возникает вопрос о том, какие курсы будут проходить тестирование по информатике. Студенты сообщают, что им говорили только о 2-4 курсах. Преподаватель уточняет, что получил информацию от Лаптева С.В. о тестировании первого курса по данной дисциплине и что тесты (50 вопросов) доступны в Moodle и на ResearchGate. Он предоставляет ссылку на ResearchGate.

**IV. Основная тема: Взаимосвязь энергии, информации и материи:**

* **A. Обзор мер информации:** Кратко повторяются меры Больцмана, Хартли, Шеннона. Отмечается, что формула Шеннона использует формулу Хартли, а не является ее обобщением.
* **B. Мера Харкевича и системное обобщение:** Упоминается мера Харкевича (вероятность достижения цели) и системное обобщение формулы Хартли, предложенное лектором.
* **C. Недвоичные единицы информации:** Обсуждаются ниты (натуральный логарифм) и диты (десятичный логарифм) как альтернативные единицы измерения информации, аналогично разным единицам измерения веса (кг, фунты). Приводятся коэффициенты пересчета между битами, нитами и дитами.
* **D. Связь энтропии, энергии и информации:** Энтропийная мера Больцмана связывает информацию с энтропией, а термодинамика связывает энтропию с энергией. Приводится пример с таянием льда: сообщение энергии увеличивает хаотичность (энтропию) и стирает информацию о структуре. Обратный процесс: сообщение информации (структурирование) уменьшает энтропию и выделяет энергию (вода остывает, превращаясь в лед). Ставится вопрос об аналоге формулы E=mc² для связи информации (I) и энергии (E), возможном виде I=f(E).
* **E. Информация, труд и структурирование:** Труд рассматривается как информационный процесс – придание структуры объекту. Пример с изготовлением вазы: информация, вложенная гончаром, уменьшает энтропию глины. Упоминаются разработанные лектором информационная теория времени и стоимости.
* **F. Пределы точности:** Приводятся примеры, иллюстрирующие ограниченную точность измерений и оценок сложных систем: игральный кубик (информация зависит от того, знаем ли мы конкретное число или только четность/нечетность), рейтинг университетов (сложность точного ранжирования тысяч вузов, модель лучше работает с меньшим числом градаций), дегустация вина сомелье (субъективность и ограниченная точность 100-балльной шкалы по сравнению с 5-балльной).

**V. Фокус практического занятия: Разработка ЭДС-приложения:**

* **A. Упоминание задачи:** Текущее занятие – практика, посвященная разработке собственных ЭДС-приложений.
* **B. Доступность шаблонов/примеров:** Преподаватель разместил на ResearchGate пример-шаблон (работу студентов группы 2002), который можно использовать как образец описания. Файлы доступны для скачивания.

**VI. Заключение и вопросы/ответы:**  
Преподаватель призывает задавать вопросы по разработке приложений. Происходит короткий телефонный разговор с Сергеем Владимировичем Лаптевым об отборе вопросов для завтрашнего теста. Преподаватель подтверждает, что Лаптев сам выберет 50 вопросов из предложенного большего набора. Снова предлагается задавать вопросы по практической работе.

**Детальная расшифровка текста**

**I. Введение и детали сессии**

Добрый вечер, ребята.  
Слышно вас? Отлично. Ага, отлично.  
Здравствуйте.  
Да, здравствуйте, здравствуйте.

Здравствуйте.  
Здравствуйте, здравствуйте.

Сегодня 10 ноября 2020 года. Шестая пара.  
Идет она в 17:20-18:50.  
Лекция 11. Сейчас мы еще уточним это. По дисциплине: теория информации, данные, знания.  
Занятие ведет профессор Луценко Евгений Вениаминович.

Вот.  
Отмечаемся. Мы всё ведем.  
И с группой, с группой.  
Со всеми группами. Это лекция. ИТ-2001, 2002, 2003.

Лекция 11.

Шестая пара.

Да. Ну, конечно, вас не столько, сколько здесь, но хотя вы такие приличные.  
Ну ладно.

**II. Обзор предыдущей лекции и введение в текущую тему**

Вот. По теме лекций.  
У нас, мы на прошлом занятии начали тему 1.11.1.  
Взаимосвязь материи, энергии, информации в природе и в обществе.  
Вот давай эту тему сейчас найдем в пособии.  
И начнем изучать эту тему.

**III. Объявление о предстоящем тестировании и обсуждение**

* **A. Первоначальное упоминание**

Вы знаете, что у вас завтра тестирование будет, ребята?  
М?

Вас предупредили уже, нет?

* **B. Студентский запрос и уточнение (Кто тестирует, доступность тестов)**

Что такое? Марк? Что ты имел в виду? Я уже не понял. Что ты имел в виду?  
А. Ну да, завтра тестирование будет у вас.  
Какое? Обычное тестирование.  
То есть будете сидеть в аудитории, отвечать на вопросы, выбирать варианты ответа.

Вы об этом не знали, да?  
Ну вот мне сейчас вот сегодня сказали, что и вы будете тестироваться. Причем это несколько странно, потому что у вас первый курс.  
Поэтому это странно.  
Ну факт. Мне взяли тесты. Вот. Тесты находятся в Moodle.  
И есть в ResearchGate.

Сейчас я вам дам ссылочку на ResearchGate.  
Ну и в Moodle эта ссылочка есть.

Евгений Вениаминович. Меня слышно?  
Я вас слышу.  
Я вот в чате написал. Нам говорили сегодня, что будет только второй, третий и четвертый курс писать по информатике. На первой паре говорили.  
Значит, мне сейчас во время занятия звонил Лаптев Сергей Владимирович и просил тесты. И сказал, что будут вас тестировать завтра по этой дисциплине.  
А на какой паре не сказали?  
Нет. И очень обрадовался, когда оказалось, что у меня там есть тест в этом пособии.  
Вот это учебник. И вот смотрите. Вот видите, тест по дисциплине. 236.  
И вот…  
А ответы есть?  
Ну, сейчас идет запись, ребята.  
Я думаю, вы сами посмотрите.

Вот. Значит, теперь, ребят, пожалуйста, мне напомните сейчас.

**IV. Основная тема: Взаимосвязь энергии, информации и материи**

* **A. Обзор мер информации (Больцман, Хартли, Шеннон)**

Я вам рассказывал про меру Больцмана, меру Хартли, про взаимосвязь меры Хартли и Больцмана, про меру Шеннона. Это уже у нас было.

И даже это было, по-моему. Скажите мне, пожалуйста, значит, что мы на прошлом, на прошлом занятии закончили, на чём? Лекция 11. Да, мы на этом, на следующем занятии взаимосвязь материи 1.11.16.

Хм. Ребят, конечно, примучены, но я тоже примученый.

Это я вам рассказывал.

Вот. Значит, теперь, э-э, ребят, пожалуйста, мне напомните сейчас. Я вам рассказывал про то, как при управлении какими-то системами у нас уровень системности повышается этих систем, выделяется энергия из них. И работа становится более эффективная.

Формула Шеннона переходит в формулу Хартли, как и формула Больцмана. Но дело в том, что формула Хартли используется в формуле Шеннона без изменений. То есть формула Шеннона является не обобщением на неравновероятные случаи формулы Хартли, а просто её использованием для этого случая. Вот. Это вот моё мнение по этому поводу. Оно довольно такое оригинальное это мнение. Но оно основано на том, что реально, как мы видим, эти формулы взаимосвязаны.

* **B. Мера Харкевича и системное обобщение**

Ребят, теперь вы, пожалуйста, мне напомните, я вам рассказывал про меру Харкевича или нет? Вероятность достижения цели при сообщении информации, вероятность достижения цели без сообщения информации.  
Ребят, вы как-то мне всё-таки отвечаете, реагируете на то, что я спрашиваю.  
Да, по-моему, рассказывали.  
Вот в этом-то всё и дело. Значит, давайте смотреть. А системное обобщение формулы Хартли? Меры системности, вот это вот я рассказывал, нет? То, что я там предлагал когда-то.

Модель-то я вам рассказывал, конечно.

* **C. Недвоичные единицы информации (Ниты, Диты)**

А вот про недвоичные меры и количество информации рассказывал? Диты, ниты.  
Про это, по-моему, нет.  
Ну, в общем, если вы возьмете, ребята, килограмм пшеницы, например, да, и измерите этот килограмм пшеницы в фунтах, или в пудах, то у вас получатся разные числа. Вот. Но при этом этот килограмм пшеницы так и останется килограммом пшеницы. То есть он будет иметь тот же самый вес, а разница в числах этих связана с тем, что у нас будут разные единицы измерения этого веса. Пуды там, допустим, там эти килограммы, фунты, да? Разные единицы измерения веса. Вот точно так же мы можем и использовать разные единицы измерения количества информации. Но мы знаем традиционно, мы привыкли к единицам, основанным на степенях двойки. Вот, на двоичной системе счисления. Но тот же самый подход пригоден и для других оснований логарифма. То есть если мы возьмем логарифм не по основанию 2, а по основанию E, где E - это основание натурального логарифма, число Эйлера, 2 целых там 718 и так далее, это иррациональное число, то мы получим тогда количество информации в нитах. Причем, если у нас N будет каких-то равновероятных состояний, то логарифм этого N по основанию 2 - это будет количество информации в битах по Хартли. А если мы возьмем натуральный логарифм от этого N, то будет это количество информации в нитах. Оно будет равно. То есть логарифм двух, логарифм по основанию 2 от N бит равно логарифм по основанию E N нит и равно тоже логарифм по основанию 10, то есть десятичный логарифм дит. Вот. Соответственно, получается, что мы можем... По-моему, я это тоже рассказывал, ребята. Я вот это помню визуально. Получается, что если мы посчитаем, сколько бит содержится в одном ните, в одном дите, то мы можем это посчитать. Вот. И получится у нас, что 1 бит равен 0,623 тысячных нита, 0,301 тысячная дита. Ну и так вот у нас это получится всё. Вот эта табличка, соответственно.

То есть если мы возьмем два равновероятных состояния, то в битах мы получаем 1 бит. А в нитах получаем 0,693 тысячных, а в дитах 0,301 тысячная. А если возьмем E равновероятных состояний, 2,718 и так далее, то получаем 1,443 и так далее бита, 1 нит и 0,43 с хвостиком дита. И точно так же, если мы возьмем 10 равновероятных состояний, мы получаем 1 дит. А в нитах это будет 2,3, а в битах 3,2. Вот. Но эти единицы информации очень редко используются не десятичные, не двоичные. Это такая экзотика, в общем-то. Ну ясно, что можно брать по любому основанию логарифм от того же числа состояний. А количество информации определяется числом состояний. А основанием логарифма определяется только единица измерения этой информации. И этих единиц измерения можно, так сказать, использовать сколько угодно: натуральные, десятичные. Ну, наиболее употребительные, конечно, двоичные единицы измерения.

* **D. Связь энтропии, энергии и информации (Больцман, Лед Пример)**

Так, смотрим. 1.11, 1.11. Где же оно здесь? 1.11.

Ребята, я вам рассказывал уже о взаимосвязи материи, энергии, информации в природе. Вот. А теперь взаимосвязь энергии, информации в обществе. Общество, энергия, информация в обществе. Информационное общество, общество, основанное на знаниях. И существует ли аналог выражения E = mc² для взаимосвязи энергии и информации?

Значит, если мы возьмем энтропийную меру информации, ребята, Больцмана, то мы обнаружим, что она связана с энергией, потому что сам Больцман связывал информацию с энтропией, а энтропия связана с энергией, по крайней мере, в термодинамике. Вот это вот, по-моему, на этом мы остановились. Вот это, где мы должны были бы продолжить. И отсюда вытекают очень интересные, интересные выводы, что если мы возьмем, допустим, примеры некоторые, применение энтропийной меры информации на каких-то реальных физических объектах. Вот самое простое такое физическая система, на основе которой мы можем посмотреть это, это кубик игральный, конечно. То есть если мы возьмем шесть граней этого кубика, да, которые выпадают равновероятно, N = N первое = 6, а N второе может быть равно либо единице, а может быть равно и не единице, сейчас посмотрим. Значит, то есть начальное состояние их шесть равновероятных. А конечных состояний может быть точно, если известно, какая грань выпала, тогда единица. А если не очень точно известно, а только известны какие-то вещи об этом конечном состоянии, какая-то информация только известна о нем, то тогда мы сейчас посчитаем, какую количество информации мы получаем из этой, узнав там что-то об этом, что там выпало. Значит, если у нас выпало какое-то конкретное состояние, тогда логарифм N1 минус логарифм N2, мера Больцмана - это логарифм 6 минус логарифм единицы. Логарифм единицы равен нулю. Получается 2,58, там бесконечное число знаков после запятой, количество информации, примерно 2,6 бита. А если мы узнали лишь то, что выпало чётное, вот, то мы тогда понимаем, что выпало 2, 4 или 6. То есть это означает, что у нас осталась неопределенность, если в первом случае полностью неопределенность была снята, потому что точно было известно, какое конечное состояние, то во втором случае эта неопределенность осталась, значит, некоторая.

Вот. Значит, если у нас осталась неизвестным, какое из одно из трёх состояний выпало, а только известно, что чётное, тогда N2 = 3. Тогда мы получаем очень интересный вариант такой: логарифм 6 минус логарифм 3 - это разность логарифмов - это логарифм отношения. Логарифм 6 / 3 - это логарифм 2. А это равно единице, если логарифм по основанию 2. То есть получаем замечательную такую вариант, формулу для количества информации, которая учитывает и какая была начальная неопределенность, и какая конечная. И насколько изменилась неопределенность наших представлений.

* **E. Информация, труд и структурирование (Пример с вазой)**

Вот что замечательно, что если мы рассматриваем кубик как систему, у которой есть только чётные и только нечётные грани, то по сути дела, он представляет собой систему из двух равновероятных состояний, эквивалентную с этой точки зрения монетки. Когда вот мы кидаем орёл-решка, да, орёл или решка. Ну по сути дела то же самое, мы имеем систему из двух состояний. Вот. То есть играет роль, как мы рассматриваем систему, насколько уменьшается неопределенность наших представлений о состоянии системы. Количество информации можно измерять степенью уменьшения этой неопределенности.

Теперь очень интересный пример и такой многозначительный, очень серьёзный пример. Это лёд и вода. Значит, под энтропией можно понимать степень неопределённости хаотичности состояния системы. Так в термодинамике и понимают, которую разработал Больцман. Хаотичность противоположна структурированности. Если мы сообщаем энергию куску льда, то он начинает таять и превращается в воду. Лёд - это кристалл, то есть высокоупорядоченная структура. В воде молекулы движутся намного более хаотично, чем в кристалле льда. И энтропия воды гораздо выше, чем энтропия льда. Получается, что если сообщить системе определённое количество информации, то её энтропия уменьшается, из неё выделяется энергия. Вода остывает, и при температуре уменьшится, вода структурируется, то есть превращается в лёд. То есть получается, что мы можем связать энтропию с энергией, а энтропию можем связать с информацией. При этом получается, что мы, по сути дела, информацию можем связать с энергией. То есть если мы сообщаем информацию некоторой системе, энтропия её уменьшается, определённость, степень структурированности, определённости увеличивается, и выделяется из системы энергия. Значит, для вот этого конкретного объекта, вода, например, да, это можно даже посчитать соотношение какое между сообщённым количеством информации и выделившейся энергией. И если мы наоборот сообщаем энергию льду, то он превращается обратно в воду, тает, да, энтропия его возрастает, а при этом информация стирается в его структуре. Представляется вполне возможным сделать конкретный расчёт, ну это относится к физике, наверное, получить конкретное выражение, отражающее взаимосвязь между переданным в воде объёмом информации и количеством выделившейся в результате этого энергии, которое может энергия Е может иметь вид такой: И = функция от Е, которое очень похоже на выражение Е = mc². То есть Е = функция от М, от массы. Здесь очень важно отметить, что человек может быть и источником, и приёмником информации. Получение информации человеком от некоторого источника - это познание этого источника. А передача информации человеком некоторому приёмнику - это труд, придание определённой структуры этому приёмнику информации. Причём в результате получения этой информации у него повышается у этого приёмника информации, получается, повышается уровень системности. Представление о том, что труд - это информационный процесс, сегодня уже никого не удивляет. Особенно, значит, в условиях вот дистанционного обучения, дистанционной работы. Сейчас нас как бы поставили перед фактом, так сказать, очень резко показали нам, что мы можем на самом деле работать дистанционно, если работа имеет информационный характер. Но я предложил эти идеи 40 лет назад. Я предложил 40 лет назад представление о том, что труд имеет информационную природу, что информационная сущность процесса труда. На основе этого информационную теорию времени и информационную теорию стоимости. Тогда это было совершенно не так очевидно, как сегодня.

* **F. Пределы точности (Куб, Рейтинг университетов, Примеры сомелье)**

Очень интересный пример с вазой, ребята. Значит, вы мне, пожалуйста, если я начинаю повторяться, говорите, что вы это уже рассказывали, потому что у меня ощущение, что я это рассказывал. Вот вам именно я рассказывал это или нет, ребята? Пожалуйста, скажите мне.  
Ну, студенты, они такой народ, что что-то из них вытянуть довольно-таки сложно.

Значит, когда мы говорим о высокой точности определения рейтинга, то можно сказать так, что рейтинг практически не связан с этими параметрами университетов, с их характеристиками. Ну эта ситуация мне напоминает ситуацию с сомелье. Кто такие сомелье, знаете? Это мужики там и женщины, которые пьют вино по своим служебным обязанностям. То есть обычно люди за это бабки платят, а они просто вот пьют, понимаете, это их работа. Такая у них работа, короче. Ну правда, они там, ну да, они понемножку там нюхают там, как-то его там лизнут там чуть-чуть, часть там пол ложечки. Ну, в общем, я немножко подшучиваю над ними. Так вот, они оценивают качество вина в стобалльной системе. Они говорят: 87 баллов у вина. 93 балла. А другой говорит: "Да я бы дал 94". Ну вот, ну давай 94. В общем, они голосуют там, получается какое-то среднее количество баллов. Вот. И вот я попытался понять, связаны ли эти баллы, которые они присваивают, с физико-химическими свойствами этого вина объективными. И оказалось, что связано, но связано, знаете как? Слабо. То есть, если сомелье оценивает качество вина, то ему хватило бы пятибалльной шкалы. Практически вот он оценивает в пятибалльной шкале. Вот это всё, что там они выдумывают 87, то есть эта стобалльная шкала - это для них слишком высокая точность. Они так не оценивают, понимаете? Они не могут оценить 87 или 86 баллов там. Они могут оценить там, ну, грубо говоря, плохое вино, среднее или хорошее. Это они довольно-таки нормально, ну, неплохо делают. Или можно сказать так: очень плохое, плохое, среднее, хорошее, очень хорошее. Вот это они ещё способны сказать. При этом они иногда начинают путаться вот с этим средним там, плохим вином и средним, средним и хорошим, вот такие вот. Понимаете? Они очень хорошо определяют очень хорошее, очень плохое и среднее. Вот это они как-то ещё более-менее. Вот то же самое получилось и с этими университетами. Когда мы сделали 100 градаций в шкале рейтинга, вот, то получилось, что мало связан показатель работы университета с этим его рейтингом. А когда я говорю: "Давай сделаем 10". Не 100, а 10, то есть снизим точность. То есть сами эти интервалы, размер интервалов увеличим для оценки рейтинга. Как только мы это сделали, сразу получилась хорошая модель. И они стали её описывать. Ну то есть не нужно слишком многого требовать от модели, короче говоря. Если она не может оценить там из тысячи интервальных значений относится рейтинг. Можно было сделать от одного до тысячи. Она не может это определить, понимаете? То есть там какой какой-то определённой взаимосвязи нет. Из 100 тоже она плохо, ну, практически не определяла. Ну там что-то такое чуть-чуть, но не очень. А вот когда 10, она довольно-таки неплохо определила. Вот. Но сами-то эти интервалы большие стали. Уже точность модели ниже, но зато ложных решений мало. То есть если она говорит, что вот такой диапазон, значит он такой практически и есть этот диапазон числовой рейтинг. То есть получилось как с сомелье, что рейтинг зависит от характеристик вуза, но вот эта тысяча - это большой перебор. Тысячи градаций рейтинга. Там хватило бы 10 градаций им. Вот. То есть 10 хватило бы вполне. Ну, было бы много вузов с одинаковым рейтингом. Ну и что? Это нормально, так бывает, это есть. А у них там, допустим, идёт рейтинг там 386, 387. И вот я вправе спросить: а вы когда вот ставили этому на 386, а этому на 387, вы чем руководствовались? Вот этими показателями, которые там сбоку? Они скажут: "Да". Я говорю: "Не надо мне лапшу на уши вешать, ребята". Покажите мне модель математическую, которая на основе этих показателей сможет с такой точностью, с точностью до единицы при 1000 градаций определить это значение. Я могу это сделать. Я могу взять и не не про вопросы я вам говорю. Я могу взять и определить по показателям рейтинг, могу. Но я могу это сделать приближённо, понимаете? Не с такой высокой точностью, где 237, 238 там. А вот там, грубо говоря, в первой десятке, второй десятке, там, третьей десятке и в десятой десятке, и всё. Вот это я могу вполне обоснованно показать вам, продемонстрировать.

**V. Фокус практического занятия: Разработка ЭДС-приложения**

* **A. Упоминание задачи**

Так, ребята, у вас какие-то вопросы есть по разработке собственных приложений? М? Вы этим занимаетесь вообще?  
Занимаемся.  
Молодцы.

* **B. Доступность шаблонов/примеров (ResearchGate)**

Значит, я как раз проверяю вот компетенции, ребята, именно те, которые я пытаюсь со всей силы вам дать. Провести исследование, описать его и опубликовать. Вот именно этим сейчас все интересуются. Понимаете? Что интересно. То есть я прямо вот в десятку глядел. То есть я прямо вот усилия совершал для того, чтобы научить вас тому, что сейчас будет проверяться. И достиг огромного результата, ребята, великолепного. Все мои усилия закончились тем, что у нас один студент, вернее, пара студентов, два Александра из группы 2002 разместили всё-таки приложение. А у вас группа 2003.

Так это так вообще всё. Вот. Так что, видите, одно уже есть. Значит, и я в инструкции, ребята, соответствующей, которую вы знаете, сослался теперь на эту работу, как на шаблон описания. Вот она пятая эта работа. Это теперь шаблон описания. Почему так? Потому что она довольно-таки неплохо сделана эта работа. Вот это моя инструкция. Это шаблон описания пятый. Можете брать. Почему? Потому что он-то сделан на основе, они сделали, вернее, на основе вот этих шаблонов, которые уже там были стареньких. А я просмотрел и немножко обновил информацию. И систему ИДС там описание обновил немножко, и как вот сейчас вот оно на сайте выглядит, вставил. И рисунок обновил, более корректный рисунок, более интересный вставил. А откуда вы можете брать некоторые пункты? Из шаблона? Ну, можете, если у вас они не считаются. Если у вас, допустим, там 300 этих самых классов, ну тогда вы не сможете кластерный анализ провести. Ну как-то так, разумно подходить к этому всему. То, что вот вы там исследуете, то вот, собственно говоря, и нужно делать. Ну иногда, ребята, бывает, не особо получается. Но один из основных моментов таких, который позволяет получить хороший результат, является такой вот момент. Вот слушайте внимательно, я поделюсь опытом. Значит, если вот сейчас недавно вот сегодня решали задачу, и решили, и студенты будут её описывать. На занятии решили. Значит, там суть в чём состояла? В том, что мы должны были оценить рейтинг университета по его характеристикам. То есть масса характеристик различных. Ну как масса? Не очень, не такая уж масса, ну, 10 характеристик там, грубо говоря. И рейтинг этих 1000, 2000 университетов. И этот рейтинг меняется от единицы до 1000. Вот в этом, в этом, то есть эти 2000 университетов им присвоен рейтинг от одного до 1000. Первым не с таким присвоен, вторым не с таким присвоен. Понятно, да? И вот вопрос возникал такой: а можно ли вот по этим характеристикам определить, какой рейтинг? Или они как-то вообще не связаны? Или они от фонаря там этот рейтинг вставили, а характеристики сами по себе, а рейтинг сам по себе? Или они взаимосвязаны? Ну, по идее, они, конечно, должны были рейтинг ставить на основе вот этих характеристик, так по уму. Но фактически так оно или нет? А кто его знает? Ну давай проверим. Ну давай, значит, вот взяли мы, проверили. Значит, когда я взял, сделал 100 градаций рейтинга, он от одного до 1000 меняется. Ну получилась модель плохенькая, ребята. То есть когда мы говорим о высокой точности определения рейтинга, то можно сказать так, что рейтинг практически не связан с этими параметрами университетов, с их характеристиками.

Мне бы хотелось, чтобы вы просто почувствовали, что можно вот это, можно вот это. Чтобы это было как бы не в нагрузку вам, что там вас заставляют, обязывают, а чтобы вам интересно было. Вы же понимаете, что точно так же можно сделать и любую предметную область, и там, где вам это интересно. Вот. Вы же, собственно, я не заставлял вас брать темы неинтересные. То есть вы можете взять тему, которая вам интересна, и с ней поработать.

Сейчас, ребята, я вам покажу, как размещать приложение в облаке. Вот я взял два файлика, сделал вордовский и PDF-файл. И начинаю размещение. Вот сейчас смотрите, как это будет выглядеть. Смотрите, нет?  
Да-да.  
Вот. ResearchGate. Добавить preprint. Добавить файл. Сначала я размещаю вордовский файл. Он спрашивает: "Вы в своём уме вообще? Вы думаете, что вы размещаете?" Я говорю: "Да, я в своём уме, я просмотрел, проверил". Я подтверждаю, что я имею на это право. Вот. Значит, дальше вопрос возникает: а какое наименование этого файлика? Я вам советую всё писать на английском языке. Потому что читать будут, ну, люди, которые владеют по-английски, в общем, общаются. Вот, авторов указываем. До и я здесь убираю, а в другом месте скажу, что да, надо присваивать.

Вот. Ну вот здесь вот я уже несколько раз показывал, как я его размещаю. Там нет такого. В общем, всё дело. Там такого нет. Поэтому я пишу, что нету. Описание. Описание я то же самое пишу, как разработать инструкцию. Прошёл. Лицензия самая примитивная. Нет этой опубликованной версии. Добавить. Создать DOI. Вот эта вот волшебная кнопочка, смотрите. Создать DOI. Какой проект? Ну я указываю свой проект. Добавить проект. Хочу ли я, чтобы там все ахали-охали? Нет, не хочу. Вот. И, значит, указываю ResearchGate опять, смотрю, что ж получилось. Получилось, что у меня добавилась работа, ребята. Вот. Смотрим эту работу. Было 582 работы, стало 583 работы. Смотрим на эту работу. Видите, вот, вот она. Она входит в мой проект. Никого я не хочу об этом уведомлять. Есть DOI, видите? Вот этот текст вордовский. Чтобы, чтобы текст сразу появлялся на экране, для этого нужно здесь вот сделать так вот. Вот. И теперь вот этот файл уже я делаю как PDF его. Где уже указан DOI. Вот в этом и состоял смысл записи вордовского файла. То есть я его записал и сразу же получил DOI и указал его в PDF-файле. А теперь смотрите, вот здесь вот я добавляю файл. И уже добавляю PDF-ку. А PDF-файлы, они сразу показываются на экране, в отличие от вордовских, которые просто можно скачать. Никого я не хочу про это говорить. И всё отлично, successfull. Смотрим, что у нас получилось. У нас получилось задание инструкция, как разработать собственное EDS-приложение, application, да? Название на английском языке, проекта, в который это добавлен, что-то сделано в ноябре двадцатого года, DOI с лицензией. Пока никто не читал. Как это сделать? Вот сейчас вы попробуйте прочитать. Я вам ссылочку дал. Давайте читайте это то, что там размещено. Вот здесь появится сейчас, сейчас появятся те, кто прочитал. А вот здесь вот уже идёт текст, видите? Какой вполне, так сказать, приемлемый в виде. Значит, если там будут какие-то погрешности, я могу их исправлять вот таким образом. Я могу здесь вот нажать, удалить или сделать приватным файл, задать. И любой файл могу удалить. Этот 18-19 размещён, а этот 18-17. Этот, если вот этот файл первый я удалю, то тогда DOI изменится вот для этого файла. Будет там DOI/1, DOI/2. Поэтому первый файл никогда не удаляете. А вот последующие их без проблем можно удалять. Поэтому я и сделал первый файл тот, который не будет просматриваться фактически. И вот у нас уже этот, эта инструкция уже размещена в ResearchGate. То есть это вообще совершенно элементарно делается. Значит, давайте сейчас попробуете, почитаете, поклацаете там. То есть кликните вот по этой ссылочке, которую я вам дал на ResearchGate. Последнее, которое в чате. Ну давайте, кликайте, выходите на неё, на эту инструкцию в чате уже в ResearchGate. Вот, смотрим. Никто пока не сделал этого, то, что я сказал. Или сделали?  
Сделали.  
Сделали, да? Ну, может, она не потом не отмечает вас, потому что вы... А, нет, вы уже смотрите. О! Вот видите, появилось 11 прочтений, ребята. То есть этот материал уже посмотрели 11 человек. То есть это вы. Значит, он резко набирает популярность. Ну так вот примерно. То есть вы можете вот так вот спокойно абсолютно размещать свои публикации. Это даже никакая не статья, ничего. Видите, задание, инструкция. Очень скромно это всё сделал, вообще простейшим образом.

**VI. Заключение и вопросы/ответы**

Вот. Ну, и когда вы, значит, вот Никита, вы когда с сайта скачаете исходные данные, то эти данные будут в формате CSV. Этот формат - это упакованный... А, а что ж ты тогда спрашивал насчёт таблицы? Ну так естественно, они должны отличаться. Они могут быть различные. И данные в них могут быть различные, сами таблицы могут быть различные. Они не должны быть одинаковы. Это ж понимаешь же, тема же разная. И данные разные по ним. Они содержатся в разных таблицах. Всё нормально. Но задачи вот решаются сходные. Исходные данные - это прямо вот то, что вы скачиваете с этого Кагла или UCI. А обучающая выборка - это уже закодированная информация. Уже после работы системы, она формирует справочники, классификационные описательные шкалы и градации и кодирует исходные данные. Получается обучающая выборка. Сейчас я покажу тебе. Не займёт много времени. Вот. Теперь смотрим. Смотрим исходные данные. Вот исходные данные. Смотри. Видно, да? Вот это. А вот, смотри, обучающая выборка. Ну, вопрос возникает такой: похоже на исходные данные? Ну, вообще-то, довольно-таки похоже. Вот. Явно похоже. Вот. Но в чём разница? В том, что здесь у нас тексты и числа, а здесь коды текстов и чисел. А эти коды откуда взялись? Они взялись из справочника классов. И вот эти коды сбоку. Из справочника признаков. И вот здесь вот коды. Вот они.

Всё, ребята, у нас занятие закончилось. Всего самого хорошего вам.  
До свидания.  
До свидания.  
До свидания.  
До свидания. Будут вопросы, пожалуйста, спрашивайте. Удачи вам на тестировании.  
Да. Угу. Спасибо за помощь. До свидания.  
Да, счастливо. До свидания.