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

***Российская Федерация***

**225. Практическое занятие №9. По дисциплине "Теория информации, данные и знания". Решение задач.**

**Резюме**

На практическом занятии студентам оказана помощь в решении задач по дисциплине "Теория информации, данные и знания". Обсуждены вопросы корпоративной почты, размещения приложений в облаке и самостоятельного изучения материала. Выявлены студенты, успешно разместившие свои приложения и проверено описание к ним.

**1. Приветствие и начало занятия**

Здравствуйте, Георгий Александрович, здравствуйте, ребята. Добрый день. Евгений Вениаминович, добрый день всем. Подключившиеся. Да. Их немного. Немного, но есть. Да. Сейчас ещё будут. Э, так, ребята, сегодня 24 ноября 2020 года, прям странно даже говорить уже. Вот, третья пара, 11:30 13:00. Практическое занятие номер девять с группой ИТ 2003. По дисциплине теория информации, данные знания. Занятие ведут профессор Луценко Евгений Вениаминович и профессор Аршинов Георгий Александрович.

**2. Разбор заданий и успехи студентов**

На этом занятии, ребята, мы продолжаем э-э решать, помогать вам решать задачи, которые даны в качестве задания для вас. На основе которого будут определяться уровень ваших учебных достижений, ваши знания, умения, навыки и компетенции, насколько вы их освоили, те, которые положено по учебному рабочему плану, по рабочей программе. Вот. Поэтому значит, у кого есть вопросы по поводу этих собственных работ, заданий, которые вам даны, то сейчас вы можете их задать, и мы вам поможем. Я могу вам сказать сразу две вещи. Но я не знаю, не помню, честно сказать, говорил я это или нет. Но вот сейчас вы мне напомните. Говорил ли я про то, что можно получить уже совершенно законно э корпоративные адреса электронной почты. Я про это говорил? Нет, ребят? Уже совершенно спокойно можно это сделать.

**3. Вопросы и информация о передовиках**

Так, Волков здесь, да, староста? Староста, вы это помните, чтобы я про это говорил? Вы мне отвечаете, когда я спрашиваю. Помните вы или нет. Вы что-нибудь отвечаете, помню, не помню. Я не староста, меня здесь нет. Или что-нибудь отвечаете, короче. У меня с сейчас проблема. Я не слышу вас. А, понял. Сейчас я зайду через другой аккаунт, другого компьютера. Давайте. Ещё ещё другой вариант оказался, чем то, что я перечислил, да? Пока он там входит, я сейчас вам другую ещё информацию дам. у нас уже есть свои передовики, те, кто уже разместил свои приложения в облаке, в Эдас облаке. И кто же это? А вот смотрите, ребята. Вот, видите? У нас 209, 210, 211 приложения. Они размещены вот сейчас не так давно, восьмого, двенадцатого, восемнадцатого, 22 декабря. И их разместили студенты. Минченко, Баженов, Величко, Кожухарь и Никита Шевило. Вот. И они представили мне исходные данные, источники данных, и описание своих приложений. Я эти описания проверил, сделал замечания им. Они их исправили эти замечания, после чего эти приложения были размещены в облаке. И теперь они могут их размещать с полным основанием, совершенно спокойно в Research Gate и. Вот такие вот красивенькие у них формы получились тут. Всё, я перезошёл. Зато спрашиваю. Да, да, конечно. Я вот что спрашивал. Значит, вы помните, чтобы я говорил о том, что можно э совершенно на законных основаниях получить корпоративную почту. Да, вы говорили, и я собираюсь в среду пойти, я вот написал документ, сейчас витель дописываю, и в среду пойду. Хорошо. Не забывайте, что вот это как раз является очень важным для студентов. А напомните, какой кабинет? Ну, номер кабинета я не помню, но если ты вот войдёшь в первое общежитие, там фое такое квадратное. Это знаю. Коридор прямо идёт, и коридор налево. А в углу сидят вахтёры. Вот идёшь налево, тебя спрашивают: куда, куда ты идёшь? А ты говоришь: "Я иду к начальнику Центра информационных технологий, Алексей Семёновичу Кремеру". Вот, и его дверь вторая направо, с правой стороны. Заходишь туда, он обычно там, но если его там нет, он бывает на совещаниях различных, он не проректор, но как бы в ректорат входит. То есть, когда ректорат работает, он там тоже сидит в уголке где-то там сзади. Вот. То есть он, ну, как не то, что член ректората, но его там могут спросить что-то, поэтому он там присутствует. То есть есть э руководители как высокого уровня, которые не члены ректората, но э как бы, скажем так, являются руководителями такого уровня, что члены ректората могут к ним в любой момент обратиться. Чтобы это было удобнее, они там присутствуют на ректорате прямо. Начальник отдела кадров там присутствует. В общем, ряд сотрудников руководителей. И вот он тоже там может присутствовать. А так он у себя, в общем, практически в кабинете сидит. Если там его не будет, тогда можно в первую дверь зайти, и там есть Крепошнов Дмитрий Александрович и попросить его, чтобы он передал эту служебную. Сказать, что Алексей Семёнович сказал, что можно в такой форме делать заявку на получение этих адресов почты корпоративной, и что вы так вот и сделали. Можно сослаться на меня, что он не это сказал. Он действительно действительно мне это и сказал. Написал по почте. Так что вот Никита сделал приложение, уже размещено у него. То есть он уже вот сейчас пришёл на занятие э, ну как, наверное, чтобы ещё больше повысить свою рудицию, но вообще-то он уже получил самоэкзамен отлично. То есть те, кто вот разместил, они уже молодцы, понимаете. Ну, остальным тоже надо к этому стремиться. Вот. Никита, как только появится адрес корпоративной электронной почты, моментально в Research Gate регистрируешься, и там тоже размещаешь своё это вот описание. Можно даже и с исходными данными. Там есть и возможность данные разместить. И ну, в принципе, это нет необходимости, потому что ты можешь просто у тебя ж ссылка указана на исходные данные. Где ты их взял? Вот. И в Ринс то же самое. То есть мы должны прямо вот регистрироваться там, потому что тут много времени. Многие там уже попытались получить эту спинку, но не получается как просто. Вообще уходит длительное время на это. Там бывает месяц уходит. Хотя они говорят, что неделя, но как-то вот они говорят, а практически получается, несколько раз надо просить их. Просить их можно вот в этом э в их э кабинете заявление. Вас, как только вы зарегистрируетесь, если ещё спин-кода нет, а в Ринс уже зарегистрировались, вы сразу можете э уже войти по этому вот, по этой ссылке и начинать их там трамбовать, чтобы они вам дали, ну, вежливо. Если вы будете трамбовать невежливо, вас пошлют. Ну, не пошлют, но просто вообще не будут ничего делать. Надо как-то так вежливо просить, что вот нам надо там, пожалуйста, там. Ну, так вот примерно. Понятно, да? Вот. Теперь, э, ребята, кто кто из вас присылал мне свои э материалы? Вот сейчас вот, чтобы я их смотрел там, проверял. Если присылал, то напомните мне, я их посмотрю или, может быть, вы сейчас повторно пришлёте их. Потому что там почты очень много приходит. И я вижу там иногда эти студенты пишут, но не могу проверить, потому что времени нет. Вот. пишет мне всё время. Ну, понятно. С какого перепуха. Просто, ребята, с вашей группы здесь нет никого. Полина, Нела, Вячеславна. Нет, да? просто спам попадает. Ну, Кал, видимо, отслеживает, что я туда захожу иногда, и мне это знает. А в общий доступ подключи. А, всё. Ну демонстрация. Что, чтобы чтобы видели, чтобы видели экран, да? Ну, да-да-да. Хорошо. У тебя сейчас ничего такого там не было, чтобы смотреть на него, особенно с интересом каким-то. Ну, ребят, я даю я всё-таки, у кого темы нет, будем искать, у кого есть, я помогу, там что-то проверю, посмотрю. У кого там уже определилось интересы кого выделились, уже понятно, что вы хотите посмотреть. Поэтому давайте сейчас вот мне информацию для того, чтобы мы могли заниматься дальше, конкретно содержательно, помогать вам. Или вы полностью самостоятельно всё сделаете, в чём я сильно сомневаюсь. Такое чувство, что вы вообще там не делаете ничего. Вот возьмите сейчас, пожалуйста, напишите в чат плюсики, кто занимается описанием своего приложения. У кого утверждено уже тема утверждена, и кто занимается описанием. Мм. Ну да. Восемь человек. Вот Максим только что отправил. Молодец. Молодец. Сейчас работаю с системой, и у меня там есть приложение, которое я не хочу его портить, поэтому я в другой папочке развернул минимальную инсталляцию. У меня будет достаточно. Так, что это такое? Расскажи нам. Максим. Смысл этой задачи в чём заключается? Цена, определение цены по всем характеристикам, да? Максим, ты как-то ответь голосом. Потому что я смотрю сейчас на экран. Суворов. Анализ ликвидности автомобилей. Ну, цены их фактически. Да, давайте цену переставим в начало. Вот. Или в конец можно, но мне удобно в начале, чтобы было. Как-то привыкли к этому. И не так много у тебя тут. Так, ребята, значит, сейчас сразу же тут какие вопросы. Вот смотрите, вот я беру, ставлю э в клеточку А1 курсор, нажав Shift, N и Shift Ctrl и N, а нет. э Ctrl Home. Попадает курсор в клеточку А1. Потом я держу клавишу Shift, нажимаю Ctrl и end. У меня, видите, выскакивает куда. У нас есть много, то есть две колонки, и много строк после таблицы. Вот это очень нехорошо. Это значит, что если мы сейчас это в систему будем загонять, начнёт нам высказываться там, что она думает про нас, что там колоночки есть без наименований, э, без данных, без вариабельности значений. В общем, начнёт ругаться. Поэтому давайте сделаем так. Мы возьмём, сейчас выделим область данных в этой таблице. Возьмём другой лист, и туда вставим эти данные. Вот таким образом. А тот лист предыдущий, его можно вообще убрать. Потому что там э область данных некорректная, шире, чем реально там данные. Вот. Потом следующий момент, ребята. Как правило, хорошего тона, является указывать в колоночке одинаковое число знаков после запятой, во всех значениях. Вот у нас объём двигателя 1,6, а потом два, видите, а потом три. Ну, надо тогда написать 2,0, например, понимаете. То есть нужно сделать одинаковое число знаков после запятой. Почему? Потому что э, вот это число знаков, оно указывает на точность, с которой вам известно это число. Вот, допустим, мы вы пишете 1,6, значит, вам известно десятое, да? А вот два, когда вы пишете, вам известно десятый или нет? Может быть, там 2,1, 2,2, а вы не пишете этого, понимаете. Ну, чтобы было видно, что вы знаете, что там находится, на десятой десятые доли. Ну, так написать, что там находится. Вот вы, допустим, написали ноль. Это что означает? Это значит, что вы знаете, что там ноль, понимаете, не 2,1, не 2,7, а именно 2,0. Вот. А если вы этого не знаете, тогда получается, если вы это или не пишете, тогда может такое возникнуть впечатление, что у вас разные числа известны с разной точностью. Вот там, где у вас 2-3, у вас точность была до целых, а там, где 1,6, там точность была до десятых. Но если вы будете эту информацию обрабатывать, то точность у вас будет до целых, потому что э э погрешность самая большая, значит, становится погрешностью всей э колонки. Вот. Ну, и получается, что вы могли бы это шесть и не писать, а просто округлить там как-то. Или просто один написать и всё. Понимаете, о чём речь идёт? Это нужно обязательно. Второе. Значит, нужно обязательно написать единицы измерения. Максим, ты знаешь, в каких единицах здесь единица, то есть цена? Я тут смотрю, я думаю, то есть в рублях, наверное, там, в рублях. В рублях. Второе. Пробег. Ну, похоже, что в километрах он. Год выпуска понятен. Коробка, объём двигателя. В литрах. Надо написать кубических. Ну, вот так напишем. Литры кубические. Мощность в лошадиных сил. Вот. Отлично. Теперь. Если пробег ноль, это что означает? Новая машина, что ли? Что они не вообще не ездила, да? Ну, хорошо. Теперь смотрим, сколько здесь у нас различных моделей. В принципе, мы можем сами эти модели тоже сравнить друг с другом при желании. То есть мы можем добавить туда колонку и туда сами эти модели скопировать. Марка и модель. Ну, что-то я не пойму, как её вставить. Совсем прямо уже тяжело. Так, значит, вот марка модель. Тогда мы можем и проанализировать эти марки. Но если их окажется слишком много, и для нас это окажется трудоёмким этот анализ, ну, не будем тогда его делать. Тогда не будем. И модели. Значит, теперь давайте смотреть. У нас вторая, третья колоночки классификационные, а четвёртая по десятую описательные. И мы ставим задачу, это сейчас мы выполняем первый, не автоматизированный этап системно-когнитивного анализа. Это единственный не автоматизированный этап. То есть мы решаем на основе чего и что мы хотим узнать. Вот мы хотим узнать на основе характеристик машин, хотим узнать, какая марка, модель и цена. Примеров маловато у тебя, я тебе скажу. А где эти примеры взял, Максим? А? Признавайся. У тебя микрофон не работает, да? Ну, тогда надо написала система, что локализация произведена. Почему? Потому что я её записал в новой папочке. И она до этого не запускалась. Вот. И смотрим, какой там у нас этот файл исходных данных. Новое Excel. Со второй по третьей классификационный, с четвёртой по десятой, описательный. стало. Я советую всегда вот этот указывать адаптивные интервалы использовать, и нули мы будем считать отсутствием данных, или будем считать данными. В данном случае, вот в этой, в этой задаче, наверное, надо считать их данными, потому что там новая машина имеет пробег ноль. Но я вам скажу так, что, ребят, вы, я не знаю, знаете про это или нет, в новых машин никогда не бывает пробег ноль. У них пробег бывает, ну, там, скажем, 18 км, там 8 км. Вот такие вот у них пробеги. Вот. 7 км может быть, но не ноль. Ноль не бывает, это точно я вам скажу. Вот. Так. Что мы здесь увидим сейчас? Я вам сейчас скажу. Вот смотрите, ребята. Классификационная шкала, марка-модель. Их 51. Ну, это такое количество, которое можно провести кластерный анализ без проблем. То есть это будет интересно. Цена. От 33.300 330.000 минимальная и аж тут что-то такое 9.390.000 максимальная цена. И смотрите, у нас пробег от нуля до 208.000. Новая да, нет. Год выпуска. У нас год выпуска только вот 2 года выпуска. Это такое подозрение, что вот год выпуска можно убрать совсем, потому что у нас либо новая, либо не новая. Там нет, э, почему-то нет машин э девятнадцатого года, восемнадцатого года. Не знаешь почему? Тип до коробок, объём от 2,4 до 5,7. Кузовы мощность от 87 до 575. И можно попробовать сдать модель и посмотреть, что получится. Ну, видимо, год можно и убрать было бы. Потому что это дублирует совершенно одно и то же там. Дублируется с колоночкой новой и не новой. Ну, смотрим. Что получается? Ну, зря я зря запустил вообще, но сейчас посмотрим. Она считает полторы минуты, 2 минуты, пишет, что будет, значит. Это приемлемое время. Ну вот. 2 с половиной минуты, да, получится, даже меньше. Ну вот, смотрим. Какая у нас получилась достоверность? Сразу же. Ну сейчас посмотрим, на как это выглядит. Вообще не не поймёшь, как оно выглядит. Ну, этот график, он показывает, что доля истинных решений, начиная с 23%, там 22% уровня сходства, становится больше, чем доля ложных решений. И где-то, начиная с 60%, ложных решений нет. И, в общем, есть закономерно растёт доля истинных решений при увеличении уровня сходства. Что, в общем-то, означает, что этот уровень сходства является адекватной мерой достоверности решения. Вот. Ну, сказать, что модель получилась какая-то очень хорошая, я б не сказал, хотя вот здесь вот по второму критерию вообще высокий критерий, 939. Здесь этого не просто по количеству. Здесь не скажешь, что это хорошо. Вот. Ну, давайте сейчас вот что сделаем. Возьмём сейчас эту таблицу, скорректируем. Уберём из неё год, как бесполезный. Вот. Остальное всё вполне разумно оставим. И создадим новую модель. И сейчас уже будет у нас вот здесь вот девять тогда. И сейчас тогда Максиму буду на почту это всё скриншоты эти отправлять. Чтобы у тебя были все эти параметры, какие надо задать, при которых надо расчёт делать. Вот. Можно сделать ценовых категорий побольше. Вот. Ну, я не знаю там, пять, например, или 10. У нас получится. У нас получается марки и модели и ценовые категории. Значит, а вот это вот у нас очень странное. Пробег 00, 20. По двадцати одному примеру. С нулевым пробегом. Почему-то два интервала с нулевым пробегом. Я не очень понял. И здесь вот. Ну ладно. Короче. Так как это, в общем-то, с пробегом там что-то не совсем хорошо. Ну, ладно. Вот. Что-то не совсем понятно. Так, и теперь смотрим, какая у нас обучающая выборка. Ну, когда вот синтез модели будем проводить, вот здесь вот оставим 10% наиболее достоверных результатов идентификации. Потому что мы знаем, что уровень сходства является адекватным критерием достоверности, то зачем нам нужны недостоверные результаты? Давай оставим только самые достоверные. Используя этот вот адекватный критерий достоверности. Расчёты будут быстрее идти, потому что размерность модели уменьшилась, я там шкалу в одну удалил и уменьшил э число результатов распознавания, убрав шум. Получилось 43 секунды, место двух минут. Вот и смотрим, значит, у нас как-то назвал задачу свою, назвал её анализ ликвидности. Ну, ладно, не буду я это смотреть. Вот тут чат этот сделан кошмарно, вы видите, да? Ой. Ну вот так. Собачки зарабатывают на колбаску. Такая вот собачкой. На. Ценовые диапазоны. 10. На, на. Ну ладно. Что-то тут не совсем понятно. Так, и теперь э смотрим, какая у нас обучающая выборка. По и тогда вот с индификацией моделей, тут здесь вот оставим 10% наиболее достоверных результатов идентификации, потому что мы знаем, что уровень сходства является адекватным критериям достоверности, то зачем нам нужны недостоверные результаты? Давай оставим только самые достоверные. Используя этот вот адекватный критерий достоверности. Расчёты будут быстрее идти, потому что размерность модели уменьшилась, я там шкалу в одну удалил и уменьшил э число результатов распознавания, убрав шум. Получилось 43 секунды, место двух минут. И смотрим, значит, у нас как-то назвал задачу свою, назвал её анализ ликвидности. Ну, ладно, не буду я это смотреть. Вот тут чат этот сделан кошмарно, вы видите, да? Ну, совсем другое дело. Закономерность она есть, только они зашумлены. Как шум уберёшь, сразу становится красиво всё. Это вот основание для того, чтобы сделать очень важный вывод, что всё это было обосновано. Исключение шума, что уровень сходства является адекватной мерой достоверности. Ну и дальше уже можно сам анализ проводить. Вот можно посмотреть, например, а что такое самые дорогие машины? И у вас будет информация, что это э огромные мощности у машины до 570 сил с автоматической коробкой передач, с огромным объёмом в литрах, 5,7 от 2,2 до 5,7 с практически отсутствующим пробегом, новая. Что-то не дорожник или пикап. И ни в коем случае не объём двигателя 1,4, 1,6, ни в коем случае не механика и ни в коем случае не старая машина и не седан и не маломощная. И так далее. То есть вы можете сейчас взять, э, Максим, запомни. Берёшь пятый шаблон и описываешь. Ну, в принципе, всё есть. И сейчас мы посмотрим, как влияют различные характеристики машин на их цену. Значит, здесь, видите, на ползает, да? Поэтому надо, видимо, в настройках задать по высоте побольше. Вот и посмотреть. Станет ли приемлемо? Если недостаточно, то можно ещё побольше сделать. Ну вот сейчас совершенно достаточно. Ну, правда, на экране не видно всего, но это всё записывается всё равно. В виде файлов. И эти файлы мы сейчас сделаем и вставим их в текст, сначала в письмо. Здесь видно, как влияют различные характеристики машин на принадлежность этой машины с такими характеристиками к определённой марке и модели и к определённой ценовой категории. Значит, э ближе к красному это большое количество информации, ближе к синему это отрицательное количество информации, тоже большое. Точки соединены белой линией с максимальным количеством информации, белой с минимальным. Ну, то есть белая линия это то, что будет при таких показателях, а красная чёрная линия это то, что не будет. То есть мы видим, как влияет на ценовую категорию и на марку модель различные свойства машин. Выборка очень неравномерная. Это видно по числовым диапазонам. видите, здесь есть узенькие и есть широкие. Вот эти широкие, они мало представлены, поэтому они такие широкие. Чтобы набрать там эти 40 наблюдений, нужно сделать шире интервал. Вот, то есть очень дорогих машин очень мало. В основном не очень дорогие машины. Ну, и дальше я не буду, Максим, я сейчас просто прикреплю этот вот файл исходных данных. Вот. Вот, который считали мы сейчас. И прикреплю сейчас все вот эти вот когнитивные диаграммы. Что на что влияет. И если ты будешь делать описание по шаблону номер пять, то у тебя получится хорошее описание. Старайся аккуратненько оформлять всё, форматировать аккуратно, потому что мне работы подавали очень неаккуратно наформатировано, я их отклонил. Ну, так по-простому сказать, тяп-ляп. Просто вот сделали: "Ну, на, только отстань". Ну, я отстану, но оценки хорошей не будет за такую работу. Это ж сделанная просто по-хамски работа, понимаете. Я и не стал их оценивать никак. Я просто взял их, отклонил да и всё. Ну, здесь вот мы что видим, что, допустим, СВТ коробка, она характерна для Nissan X. Вот. Автоматическая походу для Hyundai Santa Fe. Вот. Э, механика для Патриота ВАЗ Патриота характерно. И также мы видим для ценовой категории, как кузов какой характерен от. Вот, внедорожник ВАЗ Патриот характерно. Ну, то есть мы видим характеристику машины, грубо говоря. Мощность. Вот мощность очень низкая, низкая, средняя, высокая. В каких моделях это мы тоже видно. Вот. Пробег, как пробег влияет на ценовую категорию. Нулевой пробег характерен для очень высокой ценовой категории. Но здесь практически машин нету. Нет, наоборот, очень много машин. Что интервал узенький, то есть там много наблюдений. Вот. Для средней ценовой категории э для средний средний пробег характерен для э очень дешёвых машин. Большой пробег для просто дешёвых. И очень большой пробег тоже для дешёвых машин характерен. А вообще нулевой для очень дорогих, а очень маленький пробег характерен для дорогих машин. Ну, в общем, мы получаем информацию, которая понятно, что витор понимает это всё. Ну, мы это узнали, что седан - это дешёвая машина, внедорожник - это дорогая машина. То есть мы всё это вот узнали, увидели. Так, ну, что теперь? Теперь я послал вам это. Максим, послал вам эту все выходные формы с параметрами. Короче, берёшь просто пятый шаблон. Где он там, вот здесь. Вот. И по нему прямо всё делаешь. Как там он такой довольно неплохо сделанный. Он немножко похуже был, я его так немножко под подшаманил там и форматировал. Не пересчитывал, вот так как там было, так и сделал. Ну, всё.