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

***Российская Федерация***

***Апасов Максим Вадимович, ПИ2102***

***macsim.apasov@yandex.ru***

**140 Интеллектуальные информационные системы и технологии. Лабораторная 5. 2020-10-15**

**Заголовок:** Aidos System Lab 3.04: Риэлторская оценка методом аналогии, использование ПО и анализ признаков

**Резюме текста:**

Лекция посвящена лабораторной работе № 3.04 по дисциплине "Интеллектуальные информационные системы и технологии", проводимой профессорами Луценко Е.В. и Аршиновым Г.А. 15 октября 2020 года. Работа фокусируется на риэлторской оценке объектов (на примере автомобилей) с использованием метода аналогии и системы Aidos.

**1. Обзор лабораторной работы 3.04:**
Лабораторная работа 3.04 использует Excel-файл как источник данных и представляет собой реалистичный пример применения системы, в отличие от учебной работы 3.03. Цель – разработать методику риэлторской оценки методом аналогии.

**2. Процедуры обновления ПО:**
Даны инструкции по обновлению системы Aidos. Пользователям рекомендуется запустить StartAidos.exe для автоматического обновления или, если это не работает (особенно под Windows 10), скачать файл обновления (downloads.exe) вручную с сайта, поместить его в папку с программой и запустить для замены старых файлов новыми.

**3. Методологии риэлторской оценки:**
Кратко рассмотрены три метода оценки:

* **Доходный:** Оценка на основе ожидаемой прибыли от объекта.
* **Затратный (Восстановительная стоимость):** Оценка на основе стоимости создания аналогичного объекта. Критикуется за неучет времени, рисков и рыночной конъюнктуры.
* **Сравнительный (Метод аналогии):** Оценка на основе цен на аналогичные объекты на рынке. Считается наиболее обоснованным и является основным методом в данной работе.

**4. Метод аналогии и система Aidos:**
Метод аналогии требует нахождения похожих объектов. Система Aidos помогает в этом, решая задачу распознавания образов (идентификации) для поиска аналогов по заданным характеристикам или классифицируя объект по ценовым категориям. Подчеркивается важность экспертной оценки деталей (состояние конструкций, фундамента), которую система не заменяет.

**5. Демонстрация системы Aidos:**
Показан процесс работы с системой:

* Навигация по меню (режим 5-14 для доступа к справке и статьям).
* Загрузка данных для лабораторной работы 3.04 (тип 3, работа 4).
* Объяснение параметров: использование классификационной шкалы (ценовая категория) и описательных шкал (характеристики автомобиля). Упоминается использование адаптивных интервалов для числовых шкал, основанное на теореме Котельникова (Найквиста-Шеннона), что позволяет эффективно кодировать данные с разной степенью изменчивости.

**6. Анализ данных и интерпретация:**

* Продемонстрирован Excel-файл с исходными данными, отмечены возможные несоответствия и ошибки ввода (разные написания типов кузова, пропуски данных).
* Показаны результаты работы моделей Aidos: матрицы абсолютных и относительных частот, информативности (Inf3, Хи-квадрат).
* Проведено сравнение достоверности моделей (Inf4 и L1 показали хорошие результаты).
* Интерпретированы результаты: выявлены характеристики, наиболее сильно влияющие на принадлежность к той или иной ценовой категории (объем двигателя, отделка салона, тип коробки передач, год выпуска, пробег, тип кузова). Использованы когнитивные функции для визуализации этих зависимостей.
* Проведен анализ ценности признаков, показывающий, что небольшое количество ключевых характеристик определяет основную часть ценовой информации.

**Заключение:**
Система Aidos позволяет автоматизировать часть процесса риэлторской оценки методом аналогии, формализуя поиск похожих объектов и выявляя значимые характеристики, влияющие на цену.

**Детальная расшифровка текста:**

**1. Введение и детали сессии**

Ну что, начинаем наше занятие. Сегодня у нас 15 октября 2020 года. Первая пара 8:00-9:30. Лабораторная номер пять, группа группой 418-21. Микрофоны, ребят, выключите или поздоровайтесь или выключите. Вот. По дисциплине, по дисциплине. Интеллектуальные информационные системы и технологии. Занятие ведёт профессор Луценко Евгений Вениаминович и профессор Аршинов Георгий Александрович.

Ребят, мы с вами поздоровались, поэтому тоже поздоровайтесь с нами.
Здравствуйте.
Да, здравствуйте. Доброе утро.

**2. Обзор лабораторной работы 3.04**

Вот. Значит, у нас по плану на этой лабораторной работе мы начинаем, начинаем изучать лабораторную работу 3.04. Что это за лабораторная работа? Сейчас мы, я поделюсь экраном.

Это лабораторная работа, аналогичная работе 3.03. То есть там в качестве источника исходных данных используется экселевский файл. Но, в отличие от работы 3.04, является полной, так сказать, такой реалистичной работой. То есть она отражает прямо вот реально, как это может выглядеть на самом деле на практике. Работа 3.03, она имеет чисто учебный характер, для она сделана для ознакомления с основными понятиями. И там я подробно-подробно всё объяснял.

А вот работа 3.04, она демонстрирует, как это может быть применено в реальных условиях. Поэтому давайте эту работу и будем сейчас осваивать.

**3. Процедуры обновления ПО (StartAidos и ручное обновление)**

Значит, у кого есть, ребята, у кого запускается StartAidos, запускайте StartAidos, будет обновление. А у кого оно не пусть он не запускается, я знаю, что у многих он не запускается почему-то под десятое Windows, я сейчас вам скажу, как скачать обновление и сделать обновление системе. То есть поработать вместо файла StartAidos, короче говоря.

Вот. И в будущем вы можете, если захотите, всегда скачать обновление. Но StartAidos он не всегда его скачивает, а только тогда, когда она более новая, чем исполнимый модуль, который у вас там в папочке находится.

Значит, что нужно сделать, чтобы чтобы обновить систему? Нужно, во-первых, чтобы она не была запущена, потому что файл исполнимый будет заменён на другой. Кроме того, там есть ещё файлы, которые тоже будут заменены в папочке системы. 10 МБ обновление. То есть вы его скачиваете этот архивчик самораспаковывающийся, помещаете в папочку с исполнимым модулем. Там есть такой файл, но он старый. Вот, заменяете его. То есть там есть файл downloads.exe в папочке с системой. Вот вместо него записываете тот, который скачали и разархивируете, заменяя все файлы. После этого у вас установлена новая версия системы. Обновлённая.

Евгений Вениаминович, извиняюсь, можно вопрос?
Да, конечно. Лабораторная.
А если не запускается StartAidos, а запускается просто AidosX.exe? Тоже окно авторизации есть. Это не то, да? Нам именно StartAidos, да?
Нет, нет. StartAidos он делает то, что я сейчас вам объясняю. Он проверяет наличие этого обновления. Если там дата этого обновления на моём сайте более новая, чем дата исполнимого модуля, тогда StartAidos скачивает. А сейчас я просто вам говорю: скачивайте, разворачивайте этот вот файл обновлений. И ещё там кое-что делает, то, что не очень важно, если честно. Вот. Значит, вы можете просто, не запуская этот StartAidos, сами это обновление скачать и развернуть. То есть выполнить ту работу, которую выполняет StartAidos. Но дело в том, что он-то её делает только тогда, когда это нужно. То есть когда этот файл обновления более новый, чем исполнимый модуль. А я вам сейчас говорю, просто скачайте и обновите, потому что я его заменил там вчера вечером. Ну то есть я вчера вечером выставил новую версию. Вот. Всё целиком её скачивать нет смысла, а вот этот файлик скачать и развернуть в папочке системы, и он заменит исполнимый модуль, тот, который вот вы назвали, AidosX.exe. Он его заменит и другие файлы заменит, которые в этой папке с системой. И у вас будет новая версия системы установлена. Понятно, нет?
Да-да, сейчас скачаю всё.
Угу. Вот. Ну да, я знаю, что почему-то он у многих не запускается. Ничего он такого не делает там. Он что? Он проверяет просто обновление скачивает и проверяет, что этот исполнимый модуль не повреждён. То есть подсчитывает его контрольную сумму, проверяет, совпадает она или нет, выдаёт сообщения там всякие соответствующие. То есть вот такие дела. То есть он ничего такого странного, экстраординарного не делает. Я даже не понимаю, почему там ошибка возникает. Теперь смотрите, ребята, он работает по FTP. Ну и сам этот сама система тоже работает по FTP, ничего не происходит страшного. А там почему-то происходит. Значит, ну это надо разбираться, надо десятку ставить, под ней ставить всё инструментальное обеспечение и посмотреть, что там где вылетает. Я могу, конечно, это исправить, но для этого нужна десятка на компьютере рядышком.

**4. Методологии риэлторской оценки (Аналогия, Затраты, Доход)**

Значит, существует три метода риэлторской оценки. Один метод называется метод аналогии как раз, вот или сравнительных продаж. Другой метод называется восстановительная стоимость, метод восстановительной стоимости. И третий - по ожидаемой прибыли.

Значит, по ожидаемой прибыли. Объект, который мы хотим купить, стоит тем дороже, чем больше он даёт, чем больше ожидается, так сказать, от него прибыли, и чем больше он даёт прибыли по факту за предыдущий период. Тогда, когда его покупают вот сейчас вот. Сколько он дал там за последний месяц, за последний год прибыли его владельцам. Вот исходя из этого определяется цена. При этом никакой роли не играет, сколько было затрат на создание этого объекта, на его поддержку. Просто вот сколько он даёт прибыли. Вот даёт прибыли, вот столько. Значит, вот он стоит столько. Это один способ.

Другой способ называется восстановительная стоимость. Здесь имеется в виду, что объект стоит столько, сколько стоит его создать, этот объект. Вот. Ну, здесь у меня есть некоторые соображения по этому поводу. Ну, допустим, мы хотим купить дом. Видим, стоит дом в том месте, которое нам нравится. Подходим, спрашиваем, сколько этот дом стоит. Нам говорят, ну, продаётся. Телефон указан. Звоним, спрашиваем, говорят 4,5 млн. Мы говорим: "Что это так как-то дороговато?" Вот. Они говорят: "Вот так нам он обошёлся". Вот. Говорю: "Ну что-то как-то не верится". Приглашаю я специалистов, строителей. Они говорят: "Да мы такой за 2 млн сделаем". Я говорю: "Вы знаете, вообще-то такой дом создать вот, ну, построить вот в этом месте сейчас стоит 2 млн". Он говорит: "А, ну, хотите, пожалуйста, стройте". Значит, рядышком там покупайте участок, стройте. Значит, вы потратите 2 года, во-первых, на это. Во-вторых, э стройка - это дело, вообще-то, сложное и рискованное. Вот. Возможно, вам и не удастся его построить. А этот дом, он уже стоит. То есть я о чём хочу сказать? Во-первых, этот полтора или два года, которые надо его потратить, чтобы его построить, это время, которое вы могли бы этим домом уже пользоваться, правильно? Вот. И, э, значит, это время стоит денег тоже. То есть время - деньги. Ну то есть получается что? Что когда вы говорите о том, что этот дом стоит столько, сколько стоит его создать, то это не совсем правильно. Более таким разумным подходом является к этой сумме вот себестоимости его добавить ещё э некую сумму, которая будет э компенсировать временные затраты на его строительство и различные риски, связанные с этим процессом. Вот. То есть вы платите не столько, сколько он стоит, сколько его стоит построить вот сейчас, а ещё немножко добавляете, потому что вы будете 2 года сэкономите, и уже у вас дом вот он готовый, а так вот через 2 года, кто его знает, что там произойдёт. Может быть, там евро станет 1000 руб., и вы ничего не построите. Ну это к примеру говорю.

Вот. Так вот, поэтому метод затратный, он такой несколько сомнительный, сомнительный. Потому что с точки зрения покупателя, он, конечно, выгодный, себестоимость. Но с точки зрения продавца, он, как сказать, э понятно, что он не учитывает некоторые таких вот важных моментов. Вот. Если их учесть, тогда получается не совсем уже затратный метод, а вот уже какое-то его развитие.

**5. Метод аналогии в деталях (Использование Aidos)**

И очень хороший метод, на мой взгляд, обоснованный - это метод аналогий. То есть объект стоит столько, ну, тот же самый дом или автомобиль. Здесь у нас на примере автомобиля это статья описана на примере автомобиля и лабораторная. Значит, он стоит столько, сколько стоят такие же вот сейчас на рынке. Ну если дом, то сколько стоит аналогичный дом в этом же примерно месте, с этими же примерно коммуникациями там, ну, то есть похожий. Но раз похожий, тогда возникает вопрос: а как определить, что он похожий?

Значит, ребят, сейчас давайте вот что. Вот в чате отметьте, кто обновил систему. Ну, во-первых, я увижу, у кого она установлена и у кого она обновлена. Вот прямо плюсики поставьте там, да и всё, и можете написать словами. Вот. Угу, ну хорошо. Давайте. Значит, имейте в виду, что надо её обновить, там кое-какие есть довольно ценные изменения. Я когда веду занятия, я замечаю различные неточности в том, что я рассказываю, когда показываю это в системе, там, в пособиях. И обычно сразу либо записываю, если длительно их изменять, то это я записываю, потом исправляю это, уточняю там, улучшаю, либо сразу даже по ходу, бывает, там что-то напишу там в этом пособии, фразу, слово.

Вот. Так вот, э, ребята, для того, чтобы оценить тот же самый дом по методу аналогии, нужно найти аналогичный. А что значит аналогичный? Значит, мы должны задать его какие-то параметры этого, характеристики этого дома. И потом решить задачу идентификации или распознавания. То есть мы должны найти э похожий дом по таким же вот характеристикам. Для этого нужно иметь систему. Нет, ну можно это делать с помощью естественного интеллекта. Вот опытный риэлтор, он просто посмотрит, скажет: "А". То есть у него там в голове всё это уже, вся эта информация есть, он ориентируется в этой ситуации. И он примерно может сказать, сколько это может стоить, сразу же причём. То есть он так посмотрел, походил. Ну, надо вовнутрь войти. Я могу вам сказать, вот из своего опыта, что надо войти вовнутрь, посмотреть на потолки. Почему? Потому что бывает такой дом, что вот он прямо вот итальянский кирпич, там евро, монолит, там всё такое. Вот заходишь, смотришь, а потолки провисшие. То есть там не плита, а балки. И эти балки, они старые. А вот этот кирпич, который там снаружи, он просто его облепили этим кирпичом, а внутри там ткнуть ножом между кирпичами, ну, если удастся, конечно, или штырём каким-нибудь, а там может там вообще саман там или что угодно может быть. Понимаете? То есть это люди сейчас делают запросто. То есть внешне очень красиво, внутри там трухлятина всё это, там уже там всё может завалиться в любой момент этот перекрытие там и так далее. Вот. А бывает, оно вообще не горизонтальное. Вот заходишь, так вроде всё нормально, смотришь, а там наклон идёт, понимаете? Причём такой весьма заметный. Если воду разольёшь нечаянно, то она в угол стечёт там вся. В комнате, понимаете? Это что, нормально, что ли? То есть он уже там на угол этот дом падает. Я не знаю там что, или фундамент, или он так построен прямо. Ну едва ли он так построен, вы же понимаете, что люди же не сумасшедшие, чтобы так строить. Хотя, кто его знает, да? Короче говоря, надо смотреть. Вот если так вот походить, посмотреть, то можно составить представление, на фундамент посмотреть, на трещинки там, где там они. И всё будет понятно, в общем-то. Вот. Я, например, видел дом такой, где сток воды сделан под на угол. Ну просто вот с крыши течёт вода на угол дома. И этот угол стал проседать. Вы представляете? То есть он там подмыл что-то, я не знаю, там за длительное время. Трещины пошли на этом здании, на самом стенке, на фундаменте. Вот в этом месте, куда вода льётся на углу. Угол дома отваливается, так, грубо говоря. Ну так надо же чуть-чуть соображать же, ну что ж туда воду лить-то под фундамент? Понимаете? Ну, ну, разный народ есть, так сказать, ну, в общем, короче говоря, что здесь говорить.

Так вот, по сути дела, идёт речь о решении задачи распознавания. То есть если у нас создана модель, в которой есть много разных объектов, то надо просто решить такую задачу: найти похожий объект. Ну так это задача нормальная, классическая задача системы распознавания образом. То есть это что означает? Что программная система, которая эту задачу реализует, она подходит для того, чтобы найти аналог и определить цену. Можно считать, что цена является такой же, как у аналога. Вот. А можно решить эту задачу в другом варианте. Можно взять и сформировать обобщённые образы э различных ценовых категорий. Как выглядят э дешёвые, очень дешёвые машины? Как выглядят дешёвые машины? Какие у них параметры? Как выглядят машины средней ценовой категории? И как выглядят машины дорогие? И как очень дорогие?

Значит, теперь хочу вам вот что сказать, что, конечно, эти задачи все решаются. Но решаются они обычно экономистами и экспертами в определённой предметной области, где вот эта продажа осуществляется, там или недвижимость, или автомобили. И принцип какой? Значит, я вот сейчас вам показал в этом в этой статье есть список литературы. И в списке литературы указано указаны сайты, описывающие, так сказать, методику, э то есть подходы, подходы к разработке методики риэлторской оценки. Риэлтор - это как раз вот и есть стоимость, стоимость объекта. Вот. Так вот, э я вам могу сказать, что если так вот посмотреть просто на эту методику, вот, глазами, то становится, мне лично становится так тоскливо как-то так, знаете, это вызывает досаду какую-то определённую. То есть сначала так вроде ничего там ПВД = S умножить на CK там или А. Вот, а потом вот так вот ды-ды-ды-ды-ды-ды-ды-ды-ды-ды-ды. И вот так вот идёт, идёт, какие-то там формулы, там, какие-то коэффициенты там, вообще, понимаете? То есть вообще довольно заумная э технология разработки методики риэлторской оценки. Но если очень упрощённо рассказывать, то берётся базовая цена. Ну, допустим, вот если фирма э продаёт машины и сотрудничает с автосалоном, и вот напротив через дорогу там на Аэропортовской салоны стоят, а через дорогу стоит эта фирма, которая трейд-ин выполняет эти, покупает машины тех, кто хочет купить новую, бывшие уже в употреблении машины с пробегом. Вот. И, значит, берём мы базовую цену машины такой же точно, которая в салоне. Ну, берём её в сопоставимых единицах, то есть не в рублях, а в евро там, потому что они покупаются за евро, за доллары эти машины. Если они не локализованы, не российской сборки. Ну тогда там комплектующие покупаются. То есть она всё равно цена вырастет у неё, даже вот из-за инфляции. Вот. Ну, короче говоря, берётся базовая цена. Вот. А потом различные слагаемые умноженные, значит, э то есть различные коэффициенты добавляются, которые идут как слагаемые, как вычитаемые к этой базовой цене. Ну обычно слагаемые, то есть базовая цена, она растёт, да? Вот, или нет, она уменьшается. С течением времени она уменьшается. То есть эти коэффициенты, они уменьшают эту цену. Ну, например, вот выехал, взял, купил машину, выехал за автосалон. Просто вот э шлагбаум открылся, и вы выехали. Всё, можете, так сказать, 100.000 или там 200.000 вычесть из стоимости. Сразу же. Можете обратно заехать, сказать: "Вот я что-то передумал, можете её взять обратно?" Они говорят: "Можем, только как трейд-ин уже там всё, она будет стоить уже не миллион, а уже будет стоить 800.000 ваша машина". Говорите: "Как так? Я же только что вот". А вот так. Вот так, говорю, эта машина уже бывшая в употреблении. Да, она с малым пробегом, она не битая. Вот. Ну вот всё, это уже машина не новая, понимаете? У неё есть владелец. Всё, она стоит дешевле. Значит, теперь следующее - это сразу просто в уме можно так вычитать. Потом следующее: прошёл год - 10% вычли стоимости. Вот. Домотала там, ну, скажем, 100.000 км, ещё 10% вычли. Вот. Ну и так вот стукнули её. Если чуть-чуть, так сказать, кузовной ремонт, тогда, значит, ещё там 10%. Если сильно, там вытягивали раму там всё, ну тогда, значит, 20% там. Ну, в общем, короче говоря, можно посчитать. Сделали тюнинг там какой-то, э, красиво. Сделали, значит, что-то с салоном, взяли там, заменили дермантин на кожу или что-то такое. Ну всё, тогда дороже. Тогда добавляются там определённые коэффициенты. Получается некая формула. Ну так, в общем, понятно, что да, понятно, что её параметры машины как-то те, которые накручиваются в процессе эксплуатации. То есть время идёт и различные истории у машины может быть. Пробег наматывается, то есть и могут быть ДТП ещё. Вот, собственно, что с ней происходит. Ну ещё тюнинг, рестайлинг, там что какие-то там с ней работы могли проводиться. Ну, интересно, конечно, было ли регулярное э гарантийное и просто за деньги техобслуживание, которое положено ежегодное. Вот, в салоне там или в сервисе где-то, обычно в салоне, где покупали. И, в общем, вот всё это влияет на цену тоже. Если было обслуживание, то цена растёт. Если не было обслуживания, цена падает. Вот. То есть опытный специалист этой фирмы Trade-in, он зайдёт, обойдёт вокруг вашей машины с таким приборчиком, который меряет толщину покраски, и скажет: "А, у вас тут и крыло-то вообще-то-то уже не родное, там или вообще или родное, но оно там зашпаклёванное, покрашенное, понимаете?" Вот. То есть всё сразу выясняется это быстро. И он говорит: "Так, мы за вашу машину дадим вам 350.000". Вы говорите: "Как 350.000?" Я вот сейчас вот смотрел в интернете, вот такая машина, точно такая же, 400 стоит. Говорит: "Ну хорошо, идите там в этот Авито там или где хотите, продавайте за 400". Да, там есть такая машина, есть. Или действительно, они тоже смотрят туда. Но дело в том, что у нас же фирма, понимаете? Фирма будет выполнять услугу. Какую услугу? Продажи вашей машины. Естественно, мы хотим на этом заработать. Мы хотим на этом заработать 50.000. Не бог весть сколько. Значит, ну вот, э, а это там продаёт сам продавец, сам хозяин, собственник продаёт. Поэтому он просто продаёт эту сумму. А мы вашу машину тоже продадим за эту же самую сумму, тоже за 400.000. Вы понимаете, что если мы дадим ей 450, то купят не у нас, а у того, кто там в интернете выставил свою машину на продажу за такую цену. Такими же параметрами, характеристиками. Вот. Говорит: "Да, понимаю". Ну тогда вот сами выбираете, или будете сами продавать, и все головные боли с этим связанные, будете сами, так сказать, это. Либо вот сейчас можете её отдать нам ключи там и всё, и видите, идите себе, а мы вам, значит, деньги на карточку перечисляем. Вот. Ну, короче говоря, обычно люди соглашаются всё-таки с фирмой иметь дело. Вот. Так вот так примерно определяется цена, когда работает э сотрудник фирмы, опытный человек, который знает рынок, знает, у него там стоит много машин, он знает очень хорошо, за какую цену они уйдут, за какую цену они там будут навечно на стоянке у него там прирастут там к асфальту. Короче говоря, всё это они прекрасно понимают. Значит, для чего нужна э какая-то автоматизированная методика? Для того, чтобы можно было э избежать грубых ошибок, чтобы видеть, э какие параметры машины, как влияют на ценовую категорию, к которой она относится. Значит, для того, чтобы молодые сотрудники э не допустили грубых ошибок, я повторяю, а опытные, так сказать, тоже застраховались, так сказать, от всяких неточностей. Вот. И, значит, в принципе, эта система позволяет, так сказать, повысить достоверность принятия решений в таких случаях, адекватность этих решений.

**6. Демонстрация системы Aidos (Навигация, Загрузка данных, Параметры)**

Ну теперь давайте в системе зайдём в режим 1.11. Вы должны видеть экран сейчас. Я, по-моему, поделился, да? Вот. Для чего? Для того, чтобы стереть все приложения, которые там есть. Потом заходим в режим 1.3, добавить лабораторную работу третьего типа, четвёртую работу добавляем. Нажимаем о'кей. У нас выдаётся стандартное сообщение, что сейчас там в папочке для исходных данных всё будет стёрто. Мы говорим: "Да, пусть так и будет стёрто. Мы так и хотим". Вот. И появляется уже окошко, которое вы уже знаете. Где заданы в качестве классификационной шкалы вторая колонка в экселевском файле, а в качестве описательных шкал, характеризующих машины, заданы колоночки с третьей по одиннадцатую. И обратите внимание, заданы адаптивные интервалы, то есть разного размера, но с одинаковым числом наблюдений плюс один или минус один. Значит, я вам могу сказать, что здесь же наблюдение - это же целое число. Поэтому сделать совершенно точно невозможно, а можно сделать так, чтобы они отличались на единицу число наблюдений в интервалах числовых.

Ну а теперь посмотрим на сам файл, чтобы понять вот эту вот, э, что здесь, какие здесь параметры. Надо открыть сам файл. Он уже находится в папочке Indata, этот файл. Он скопирован туда, как вы, как я уже объяснял, из папочки LabWorks. И в данном случае это у нас работа третьего типа, четвёртая. То есть это двадцать четвёртая работа. И вот здесь у нас этот файл есть. Вот он он. И мы видим название работы. У неё, так сказать, сейчас, извините. Название у неё оценка автомобилей с пробегом по их характеристикам. То есть это эта работа. А вот этот файлик - это файлик параметров, которые мы видим вот здесь в экранной форме программного интерфейса API 2322. Это стандартный, очень развитый интерфейс ввода данных из экселевских таблиц, в которых могут быть и тексты, и числа.

И вот мы, значит, открываем этот файл в папочке прямо входных данных. И я не знаю, насколько вам хорошо видно это. Видно файл или файлик, ребят?
Видно.
Значит, где я его взял? Значит, я порылся в интернете. Значит, в самой статье вот в этой, там есть ссылочка на сайт восемь, видите, сайт восемь. В списке литературы ищем восьмой источник. Вот. И там у нас уже нет этой таблички. Вот. Но это фирма, продающая автомобили. Короче, я просто порылся в интернете и нашёл фирму. Кстати, она называется Ton Auto, да, Trade-in, видите? Вот. Ну это путь, которого уже не существует, уже 404 выдаётся, но переадресуется на то, что есть. Вот здесь я нашёл прайс-лист прямо в виде экселевского файла. Значит, ну, что, в общем-то, мне и надо было. И пришлось его немножко скорректировать. Значит, что, значит, там цена указана, да, это было в 2013 году. Вот. То есть давно. И цены там соответствующие. И новыми машинами считаются машины 2013 года. Здесь более поздних нету машин.

Вот. Так вот, что здесь я изменил? Там в объёме двигателя ещё было написано буковки Д, ДТ там, ТБ там и такое всякое. Я их поубирал. Ну я понимаю, что это турбодизель, например, там, компрессор там и такое всякое. Дизель, не дизель, бензин. Вот, но я их убрал. Можно было сделать отдельную колонку, но если честно, я просто поленился её делать. Думаю, оно, в принципе, для тех целей, которые мы вот сейчас вот ставим перед собой, ознакомиться с методикой решения подобных задач, этого вполне достаточно то, что сейчас нам есть. Вот. В принципе, можно было, конечно, добавить колоночки ещё вид топлива и наличие, допустим, компрессора. Вот. Коробка здесь, да, ещё здесь есть неточности. Ну как в любом материале. Здесь, например, кое-где написано В, но я так думаю, это вариатор коробка. Механика, автомат - это стандартная, А, М там. А ещё нам бывает написано робот, например. Вот. А ещё бывает написано ВАР, вариатор, да? То есть это не автомат классический, но вот тоже похож по функционально на автомат. А ещё бывает просто В написано, понимаете? То есть понятно, что это одно и то же. Ну надо было как-то одинаково и написать, иначе нарушается целостность данных. И при нормализации у нас получится, что это два разных значения по смыслу. Ну, короче говоря, вы понимаете, о чём я говорю. Вот. Может быть написано робот, а может буковка Р быть, например. Ну здесь такого нет буковки Р, а вот робот есть. Ну то есть это, так скажем, не идеальный, э, не идеальный прайс-лист в этом смысле, но оно тем даже и лучше, мы можем понять, какие могут быть неточности. Вот вот, допустим, белый цвет, да, а может быть написано белая, понимаете? Или может быть тёмно-красное написано, вот тёмно-красное, а может быть написано Т красы. Ну что такое цвет Т красы? Я так думаю, что девочка писала какая-то и написала Т красный. Но буковку Н пропустила, а вместо тёмного написала буковку Т. То теперь у нас в результате тёмно-красный и Т красы - это разные цвета, понимаете? А это один и тот же цвет. То есть здесь можно вот таких вот понаходить кучу. Кроме того, бывает, что нет данных в каких-то колонках. Вот не было данных в колонке кузов. Пусто просто было, понимаете? Вообще ничего не написано было. Ну я взял, написал, что неизвестного типа кузов НЛО. Вот. Такое обтекаемое. Значит, ну я так подшутил немножко, но, в общем, это мелочи. Вот. Ну так вот, в общем-то.

**7. Анализ данных и интерпретация (Excel-данные, результаты моделей, важность признаков)**

Теперь смотрите, значит, у нас здесь уже 401 пример. Машины, цены указаны, ну, в тысячах, можно сказать. То есть здесь можно было вот эти три нуля во всех этих ценах убрать спокойно, потому что там никогда нет никакой никаких цифр, кроме нуля. То есть здесь ясно, что единица измерения - тысячи. Но они не написали там в колонке, что единица измерения тысячи, а просто написали в рублях стоимость. Ну, это как бы такие мелочи, я бы сказал. То есть особо, так сказать, не будем придираться к этому. То же самое касается пробега. Вот 146.000. Нет, чтобы написать 146.750 там метров там, километров там пробег, да? То есть здесь тоже до тысяч округлён он. Вот так смотрю. Ну здесь вот, правда, написано 6.400, например. Ну то есть ясно, что точность здесь так себе.

Вот. Ну, параметры понятно, что есть классификационная шкала, ценовая категория, и есть описательные шкалы - это вот эти характеристики машин. Ну, значит, э нажимаем о'кей. Происходит что? Конвертация этого файла во внутренний формат данных. Найдена одна числовая шкала классификационная - это ценовая категория. Найдено три числовых шкалы описательных суммарным числом градаций девять и шесть найдено текстовых суммарным числом градаций 236. Мы, значит, берём, что делаем? Значит, в шкале классификационной, где ценовая категория, пять градаций делаем. А в описательных шкалах делаем три градации. Смотрим, вот когда у нас адаптивные интервалы, вот здесь появляется кнопочка, ребята, которой нету, когда одинаковые интервалы. Смотрим на, что здесь у нас по этой кнопочке. Мы здесь видим, сколько у нас наблюдений на каждое, на каждый числовой диапазон. И видим размеры числовых диапазонов. Мы видим, что они эти размеры разные числовых диапазонов. От минимального до максимального в каждом диапазоне, если их вычесть, то размер интервала разный. Но у них одинаковое число наблюдений. Это соответствует теореме Котельникова об отсчётах, или как её называют на Западе, теорема Найквиста-Шеннона. Котельников пытался опубликовать 15 лет статью, где описал эту теорему, ему не удалось в Советском Союзе. Рецензенты, доктора наук, профессора писали, что какая-то белиберда, вообще ничего не поймёшь. Ну естественно, потому что это было цифровая математика. То есть математика систем передачи цифровой информации. Вот. И эти учёные, которые рецензировали эту статью, а раньше без рецензирования ничего издать было невозможно, и сейчас тоже к этому нас призывают вышестоящие структуры различные, разного рода. Так вот, в результате эта статья так и не вышла. А через 15 лет эту теорему доказал Шеннон и Клод Шеннон, который разработал теорию передачи данных по каналам связи, которую почему-то называют теорией информации, с чем он сам был не согласен, он против этого протестовал названия. И вот её теперь называют теоремой Шеннона, а не теоремой Котельникова. Но в российской литературе, советской, её называют теоремой Котельникова. Но надо сказать, что, как сказать, после драки что там? Ну, в общем, не совсем это правильно, это теорема Шеннона. Потому что мир узнал её из работ Шеннона, из работ Котельникова он её не узнал. Работу Котельникова ему не дали ему опубликовать, понимаете? Он их опубликовал в преклонном возрасте там уже, ну просто вот, а я тоже вообще-то это понимал там. Ну да, понимал, но не опубликовал же. Не смог он. Вот если был бы интернет, то сейчас что сделал Гоша Перельман? Он взял, опубликовал просто текстовый файл со своими доказательствами этой проблемы Гильберта, за которую миллион давали долларов. И всё. И там написано внизу было Гоша Перельман. Я видел этот текст. Ничего там не понял, но понял, что какой-то Гоша, а потом узнал уже про это всё. Вот так бы Котельников опубликовал бы где-то, он бы, конечно, уже зафиксировал бы своё авторство.

Так вот, ребята, смотрите, значит, в числовых диапазонах разного размера одинаковое число наблюдений - 80. Это такое достойное число наблюдений. То есть можно считать, что достаточно оно убедительное. И вот здесь вот мы видим марка, модель, э, кузов, какой. Причём кузов тоже есть универсал, а есть унив. Есть лифт, а есть лифтбэк. Есть хэтчб, а есть хэтчбек. Ну я так думаю, вы догадываетесь, что это одно и то же. Вот унив, универсал - это одно и то же. Вот. Понятно, да? Хэтчб и хэтч, хэтчб и хэтчбек - это тоже одно и то же. То есть можно было бы, конечно, подшаманить эту таблицу, потому что она не нарушена целостность данных. То есть там одно и то же называется по-разному. Как в самых таких отстойных э системах работающих с базами данных, где где нарушена нормализация, нет нормализации. Ну взяли бы, сделали справочник вот этих вот кузовов, и тогда бы такого уже никогда бы не было.

Значит, теперь смотрим на описательные шкалы. Там мы задали три градации. Мы видим 133, 134 наблюдения. Видите, да? Отличается на единицу. По какой причине? А вот по той причине, что нет полтора землякопа. Если мы сейчас вот В - вариатор, например, видите, там. Значит, велюр и ткань - это вообще-то одно и то же, по-моему. Но я не уверен в этом. Вот. Если здесь вот будет мало у нас наблюдений, смотрите, вот здесь, да, недостаточно, тогда можно увеличивать число, то есть уменьшать число интервалов, увеличивать размер интервала.

Так вот по теме Котельникова, ребята, значит, если мы имеем некоторую кривулечку, сейчас я вам популярно объясню. Значит, э сейчас я вам покажу, как это может выглядеть. Вот берём мы, допустим, какую-то кривулечку рисуем. Ну я вот такую нарисую вот так, а потом вот так. Вот. Эту кривулечку нужно оцифровать. То есть ввести, измерить её значение и ввести в систему. Вот оси. Ну я так нарисую отстойно немножко, ну вот так. Вы же понимаете, да? Вот. Теперь смотрите, значит, если мы возьмём точки на равных расстояниях, вот возьмём вот так вот отсчёты померяем, то у нас получится что? Что, значит, э в результате у нас получится, что вот эти точки мы установим. Вот эта точка, вот, вот эта точка, вот эта точка, вот будет у нас. Вот. И вот так вот точки мы померяем, чему равны значения функции в этих точках. И узнаем про это, какие значения функции в этих точках. Вот. Узнаем, какие значения. И что у нас в результате получится? Что если мы возьмём точки вот эти, вот, вот эту точку, вот эту точку, вот эту точку, и занесём всё это в таблицу значения этих соответствующих координат и координаты Y для этих точек, вот для этих иксов. Иксы через равное равное расстояние. То есть я вот здесь вот я считаю по оси X равное расстояние. То потом из этой таблицы, если мы нарисуем эту функцию, то она будет вот такая получится. Здесь вот прямая линия. Вот. Вот здесь вот получится у нас прямая линия. Вот здесь прямая линия. Вот здесь прямая линия. Вот здесь прямая линия. Вот здесь. И вот так она пойдёт. Вот сюда, потом сюда. Ну и так далее. Что это значит, ребята? Что мы потеряем полностью, потеряем информацию о высокочастотной компоненте, вот об этой. Об этом и говорится в теореме Котельникова, что если мы разложим эту функцию в ряд Фурье, получим её спектр, где представлены различные амплитуды различных её гармоник этой функции, то мы увидим, что там представлены и низкие частоты, и высокие частоты. А теперь смотрите, если мы берём точки, ставим на некотором расстоянии друг от друга, то все частоты высокие мы теряем. Вот об этом и говорится в теореме Котельникова или Шеннона. Найквиста-Шеннона. Если мы хотим не потерять информацию об этой вот высокочастотной компоненте, то здесь надо точки намного чаще ставить, где мы измеряем значение функции. Тогда мы не потеряем вот этих вот вещей. Мы тогда вот тут вот всё это проследим, понимаете, как она идёт. Вот. Но если мы везде вот так сделаем с такой частотой, то это неоправдано. Почему? Потому что здесь-то она почти прямая эта линия. То здесь хорошо аппроксимируется вот такими вот редкими отсчётами. А вот здесь нужно часто ставить отсчёты. Что это значит? Это значит, что э если мы берём равное по величине интервала числовые, то они неплохо аппроксимируют, когда функция гладкая. А если она содержит высокочастотные компоненты, тогда нужно чаще ставить отсчёты. Но здесь это неоправдано на гладкой части. А вот только нужно делать вот на этой части. Так вот я где есть высокочастотные компоненты. То есть мы можем взять, разделить эту функцию на какие-то диапазоны и посчитать спектры в этих вот диапазонах какие. И мы увидим, что здесь вот у нас спектр такой, что нет высокочастотных гармоник, вот здесь и здесь. А вот в этих, в этом диапазоне вот, и в этом диапазоне, и в этом вот, у нас есть, особенно вот здесь, у нас есть высокочастотные компоненты. Что в этом случае я предложил? Я значит, предложил и реализовал в системе Aidos. Я предложил измерение вот этой вот э, скажем так, степени динамичности функции. То есть насколько быстро она меняется. Вот. И выбор такого интервала вот этого э для аппроксимации этой функции, который позволяет не потерять данных. Вот. То есть вот здесь вот система Aidos определяет, что здесь можно редко брать данные, здесь большие будут у нас величины этих интервалов адаптивных. А вот здесь они будут маленькие, потому что если взять их здесь большие, то мы потеряем информацию. И система Aidos это определяет. Она использует адаптивный размер интервала, то есть он разный. Здесь вот они будут широкие интервалы, а здесь будут маленькие. А потом опять они будут большие. Вот она что делает. В результате мы не теряем данных, но при этом у нас число точек минимальное. То есть мы очень разумно используем отсчёты. То есть там, где функция плавно меняется, у нас отсчётов мало, а там, где она высокочастотные компоненты есть, то там отсчётов побольше мы берём. То есть там, где надо, мы берём побольше, там, где не надо, не берём, короче. Вот смысл такой. Это вот и есть адаптивные интервалы. Я вам показал, как это выглядит. И почему у нас получаются разного размера эти интервалы. Вот так вот. То есть здесь вот у нас э плавно меняется функция, поэтому не нужно тут маленьких. А здесь она меняется. И, соответственно, у нас получается вот так.

Теперь, почему 80-81, а не везде 80 или 80,3 там, допустим, и ещё 2 сотых? Потому что, если взять две ямы и трёх трёх землякопов, то одну яму копает не полтора землякопа, то есть 3 / 2, а одну яму копает один землякоп, а другую - два землякопа. А полземлякопа, они вообще ничего копать не могут. Вот по такой же причине. То есть у нас число наблюдений вот в этом экселевском файле, оно целым числом является. И если мы разделим на какое-то число интервалов, то у нас может получиться, что будет на единицу отличаться число наблюдений в разных интервалах.

Ну всё, теперь, значит, если мы здесь поменяем число диапазонов, что вполне можно сделать, мы можем пересчитать тогда матрицы. Вот, допустим, здесь вот 10 сделаю я. А здесь сюда смотрите наверх. Размерность матрицы модели и пересчитаю. И у нас получается теперь здесь 10, видите? 10 числовых диапазонов, ценовых диапазонов будет. Значит, здесь есть фраза такая: суммарное количество градаций классификационных и описательных шкал. То есть сначала идёт число классификационных градаций классификационных шкал, то есть классов. А потом число градаций описательных шкал, то есть признаков. Ну, давайте, значит, я не буду ничего здесь менять, а так вот сделаю, как было. Значит, но с этим можно экспериментировать и пытаться сделать так вот подобрать эти интервалы, чтобы модель была более достоверная. При этом, чем больше у нас данных исходных, тем точнее можно сделать модель, то есть тем меньше можно делать размер интервалов. А их больше.

Значит, теперь смотрите, что у нас произошло? У нас произошло э формализация предметной области осуществлена была. То есть у нас созданы классификационные описательные шкалы и градации, а потом с их помощью исходные данные закодированы и получена обучающая выборка. Потом здесь вот написано, что теперь надо делать. Вот это, кстати, я исправил. Вот, немножко этот хелп. Здесь сейчас режимы, название режимов соответствует текущей версии системы Aidos. То есть мы должны посмотреть сам файл, что мы уже сделали по пути. Этот путь будет у вас указывать всегда на то место, где у вас реально находится этот файл. А потом посмотреть классификационные описательные шкалы и градации, посмотреть обучающую выборку, создать модели, посмотреть на сами эти модели, потом посмотреть на их достоверность и так далее. Вот у нас эта работа установлена. Выполняем то, что там написано в этом файлике. Вот у нас ценовая категория, и здесь у нас пять интервальных значений разной величины, но таких, в которых одинаковое число наблюдений - 80. А вот описательные шкалы. Эти шкалы есть и номинальные, и порядковые, и, значит, и числовые. Ну, порядковых здесь нету. Вот марка, например, машины - это номинальная шкала. Здесь просто по алфавиту они рассортированы. Марка плюс модель, кузов, год выпуска. Вот год выпуска - это числовая шкала. Здесь у нас три диапазона есть. Ну это, грубо говоря, так, старые машины, среднего возраста и новые машины. Вот 2008-2013 - это новые машины на тот момент, когда писалась эта статья. Пробег 1.800 км. То есть совсем, это она только приехала в автосалон, и вот у неё получилось. И или 460.000 км. Это уже какой-то такой запредельный пробег. Объём двигателя от единички литра до 5,7. Цвета, коробка передач, где мы видим В и ВАР, вот. Отделка салона. Вот. И видно, сколько каких э значений здесь есть. Вот это первое из 137. Вот их 137 у нас и есть марка модель. А вот марка у нас 31 значение марки у нас, 31 марка машины здесь у нас в прайс-листе.

Теперь мы создаём модели. Теперь, значит, я не помню, говорил вам или нет, мы можем это делать на графическом процессоре, потому что я исправил неточность, которая там была при расчёте модели ПРЦ-2. То есть сейчас он работает совершенно идеально даёт результат. Ну тот, который должен быть. Осталось 3 минуты, но по времени показывает, что 1 минуту осталось считать. Я вчера интересную информацию сообщили, что есть вот такой Винчестер, SSD накопитель с интерфейсом М2, который очень ускоряет. А Елизавета, что там с видео у тебя? Очень ускоряет работу. У него скорость обмена где-то тут я читал, 530 МБ в секунду, ребята. От 530 до 330 МБ в секунду на чтение запись. То есть ставите такой вот SSD накопитель, ставите на него систему, и у вас компьютер начинает работать, как минимум, в 10 раз быстрее. И мне сказал товарищ мой, что он проводил расчёт в системе Aidos, формировал модели там и всё прочее, где-то минут 15-20. Поставил эту штуковину, она стала за полторы минуты считаться всё. То есть система Aidos тормозит не процессор, а скорость обращения к Винчестеру. Я знаю. То есть этот Винчестер, когда устанавливаешь SSD, именно с интерфейсом М2. У меня тоже SSD Винчестер, на нём система и всё, но но интерфейс IDE. То есть обычный стандартный интерфейс для Винчестера. И вот он работает, ну, быстрее, конечно, там раза в полтора-два, чем Винчестер, но не так существенно. А вот этот - это уже качественно меняет ситуацию. То есть или 10 часов считать, или там, допустим, полчаса считать. Это ж есть разница, правильно? Вот примерно такая разница и будет. Это проверено, ребята, на практике. То есть даже если ставить на ноутбук какой-то тормозной, вот это ставишь, то он становится очень крутой, не тормозной. Да. Ну, похоже, что не успеет она посчитать такие такие модели, потому что уже до конца занятия, потому что уже конец занятия. Ну вы видите, что раз размерность модели больше, ну тогда и расчёты тоже дольше идут, естественно.

Ребята, какие э вопросы у вас? Есть вопросы, нет? Ну я не все выходные формы показал, но уже модель большей размерности, уже это потребует большего времени. Ну так суть вы поняли, что можно с помощью системы Aidos разработать методику, позволяющую определять ценовой диапазон различных товаров: машин, зданий, телефонов, видеокарт там, чего угодно, понимаете, планшетов. Вот. И вы не забывайте, что вам нужно сделать приложение собственное. И, соответственно, вы получите оценку потом в результате. Вот. И вот вам задание, я вам это показывал. Задание и инструкция, как его выполнять. А что у вас всё хорошо, Елизавета? Уже уже видео нормально работает, да? Ну я рад. Ну вот, это задание, имейте в виду, что нужно это задание выполнить будет. Поэтому вы вот, учитывая эту работу, которую я сейчас показал, можете выбрать себе тему, которой ещё нет. Ну и так далее, тут посмотреть. Всё. А сейчас у нас конец занятия. Всё самого-самого хорошего вам.
До свидания.
До свидания.
До свидания.
До свидания.