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

***Российская Федерация***

***Потемин Илья ПИ2103***

**128 Теория информации, данные и знания. Лабораторная 6. работа 3.0.4 2020-10-09**

Заголовок: Лабораторная работа: Оценка подержанных авто с помощью системы Эйдос

Резюме

Лекция посвящена выполнению лабораторной работы №6 (внутренний номер 3.04) по дисциплине "Теория информации, данные, знания" для группы ИТ-2001. Тема работы – "Оценка автомобилей с пробегом по их характеристикам" с использованием интеллектуальной системы Эйдос.

1. Введение и подготовка:

Профессор Е.В. Луценко начинает занятие с организационных моментов и представления темы. Обсуждается необходимость обновления системы Эйдос, особенно для пользователей Windows 10, из-за возможных проблем со стандартным файлом запуска. Предлагается скачать специальный файл обновления и установить его путем распаковки в папку системы с заменой файлов. Подтверждается, что обновленная версия датирована 08.10.2020.

2. Настройка и запуск лабораторной работы:

Для начала работы необходимо очистить предыдущие приложения (меню 1.11) и добавить новую лабораторную работу 3.04 через диспетчер приложений (1.3). Студентам рекомендуется ознакомиться со статьей (2013 г.), на которой основана работа, доступной по ссылке в интерфейсе.

3. Работа с данными в Эйдос:

Система загружает исходные данные об автомобилях из файла Excel старого формата. Настраиваются параметры импорта: диапазоны классификационных (цена) и описательных шкал (характеристики авто), обработка пропусков. Система автоматически создает шкалы, градации и обучающую выборку (401 автомобиль).

4. Обзор исходных данных:

Данные включают марку, модель, ценовую категорию (классификационная шкала), тип кузова, год выпуска, объем двигателя, пробег, цвет, тип КПП, отделку салона.

5. Методы оценки автомобилей:

Обсуждаются методы оценки активов:

\* Затратный (восстановительный): Стоимость создания аналога сейчас (часто нерелевантно для рыночной цены).

\* Сравнительный (аналоги): Анализ цен на похожие объекты на рынке (используется в ЛР).

\* Доходный: Оценка потенциальной прибыли от объекта (обычно для бизнеса).

Приводится аналогия с экспертной оценкой риелторами или автоспециалистами, которые интуитивно комбинируют эти подходы, учитывая состояние, возраст, комплектацию и рыночную конъюнктуру.

6. Моделирование и анализ в Эйдос:

Проводится синтез моделей в системе Эйдос для выявления зависимостей между характеристиками авто и их ценовой категорией. Выполняется эксперимент с параметрами модели (количество градаций, адаптивные интервалы) для повышения достоверности. Анализируются результаты идентификации: система определяет ценовую категорию конкретных авто из выборки.

7. Анализ результатов:

Исследуется надежность распознавания разных ценовых категорий. Анализируется влияние отдельных характеристик (марка, модель, объем двигателя, год выпуска, тип кузова, отделка) на принадлежность к той или иной ценовой категории. Подтверждается, что марка и модель являются ключевыми факторами, но и другие параметры вносят свой вклад, что соответствует экспертным знаниям. Рассматривается ценность признаков для решения задачи.

8. Заключение и дальнейшие шаги:

Обсуждается задание к лабораторной работе. Анонсируется следующая тема – работа с графическими объектами в Эйдос.

Детальная расшифровка текста

1. Введение и подготовка

1.1. Сессионные детали

Здравствуйте.

Здравствуйте.

Да, здравствуйте.

Сегодня 9 октября 2020 года.

Четвертая пара идет с 13:50 до 15:20.

Лабораторная работа номер шесть у группы ИТ-2001 по дисциплине Теория информации, данные, знания.

Занятие ведет профессор Луценко Евгений Вениаминович.

Вот, так что еще раз здравствуйте, ребята. Еще раз здравствуйте.

1.2. Обзор и Введение в новую лабораторную работу

Так, значит, теперь что у нас по плану? На прошлой работе лабораторной мы закончили рассмотрение всех вопросов. И теперь мы начинаем новую лабораторную работу 3.04.

2. Инструкции по обновлению системы

2.1. Проверка установки

Ребят, значит, сейчас вы в чате напишите мне, у кого установлена система и кто на своих компьютерах повторяет то, что я делаю. Установлено, повторяю, установлено, повторяю. Или там не установлено, не повторяю.

Угу.

Хорошо. Владислав, хорошо. Это достижение современного уровня технологий, что можно ехать домой и проводить в это время лабораторную работу, да? Ну смотреть, по крайней мере.

Да. Ну хорошо. Понятно. Нормально, ребята, хорошо.

2.2. Процедура обновления (Загружаемый файл)

Теперь обращаемся к сайту моему. Я сейчас вам отправлю ссылочку на обновление системы. Я так понял, что довольно проблематично бывает обновить ее, потому что старт Eidos он выдает ошибку у многих, да? Вот под десятый Windows. Поэтому вы можете вот что сделать, ребят. Вы скачаете вот этот файлик downloads.exe. 10 МБ, он недолго будет скачиваться, малюсенький. Вот, и не запуская системы, разверните его в папочке системы, этот файл, с заменой всех файлов. Это файл обновлений. То есть вы будете вместо старт Eidos работать. Вот. И тоже в чате напишите: обновил, обновил.

Вот, то есть скачиваете этот файлик и разворачиваете, не запуская систему Eidos в папочке, где находится исполнимый модуль. Ну и так вот изредка, не знаю, там раз в месяц, если пользоваться, так примерно вот это делать. Ну или запускать файл Start Eidos.

2.3. Проверка обновления

Если вы обновите систему, то у вас вверху, вот здесь там, где имя окна, будет написано, что система от 8.10.2020. То есть вчера я ее эту версию сделал и разместил, выставил ее на своем сайте.

3. Лабораторная работа 3.04: Оценка подержанных автомобилей

3.1. Настройка и первоначальная загрузка данных

И теперь давайте устанавливать работу 3.04. Но сначала нажимаем, выбираем пункт меню 1.11. Это пункт, который удаляет все ранее созданные приложения. То есть пустая система, чистенькая.

Заходим в диспетчер приложений 1.3. Добавить лабораторную работу третьего типа 3.04. Работа, ребята, называется Оценка автомобилей с пробегом по их характеристикам. И справа здесь есть, видите, теория по лабораторной работе 3.04.

3.2. Обзор исходных данных (Статья и файл)

Можно прямо и клацнуть даже. Если это сделать, то скачивается, сразу же открывается статья, по которой сделана эта лабораторная работа. Статья немножко так староватенькая, 2013 года, но она не устарела в том смысле, что вполне можно ей пользоваться для тех целей, для которых мы собираемся.

Что мы собираемся сделать, ребят? Мы хотим уже не учебный пример рассмотреть, а такой, который похож на реальный, более-менее.

Выбрали эту работу, открыли статью, нажали О'кей. Появилось сообщение, что сейчас все файлы в этой папочке будут удалены. Ну и хорошо. Значит, давайте О'кей нажимаем.

3.3. Интерфейс ввода данных и параметры

Появляется окошко вот такое. Это окошко мы уже видели в лабораторных работах, где мы обрабатывали таблицы, учились обрабатывать таблицы, учились обрабатывать тексты. Вот, это окошко программного интерфейса 2.3.2.2 табличного интерфейса, которое позволяет вводить в систему Eidos данные из таблиц, как числовые, так и текстовые. И здесь мы видим, что мы используем старый Excel, исходные данные в файле старого Excel, стандарт. Эту птичку можно убрать, она нам не нужна. Нули и пробелы будем считать отсутствием данных. Классификационные шкалы у нас со второй по вторую, а описательные шкалы с третьей по одиннадцатую. И нажимаем О'кей. Все эти параметры они заданы правильно по умолчанию при установке лабораторной работы, эти параметры все устанавливаются сразу же.

3.4. Процесс создания модели

Теперь смотрите, ребята, когда появляется вот этот вот калькулятор внутренний, я так называю, это внутренний калькулятор интерфейса ввода данных. Что он делает? Он подсчитывает, сколько каких шкал, какого типа, классификационных и описательных есть в этой модели, которая сейчас будет создаваться. Здесь мы видим, что три классификационных шкалы. Вот количество градаций, суммарное количество градаций мы видим. В классификационных шкалах три, а в описательных 272, 272 градации. И у нас числовая шкала одна, классификационная. В ней три градации. А описательных шкал, смотрите, ребята, три числовых шкалы и шесть текстовых.

Значит, здесь ничего мы не меняем. Если если будем менять, тогда надо пересчитать шкалы, чтобы поменялось вот это вот число вверху, шкалы и градации. Вот. Ну сейчас мы просто выходим на создание модели.

И что происходит? Как и раньше и в других лабораторных работах, мы видим, что система создала классификационные описательные шкалы и градации на основе файла исходных данных, а потом с их помощью закодировала файл исходных данных и получилась обучающая выборка в результате. Потом всё это переиндексировалось, и теперь работа готова к изучению. И для студентов сообщение здесь, которое студенты обычно не читают, но я вам советую посмотреть, что здесь написано. Здесь написано: Открыть файл исходных данных по такому пути. Этот путь реальный. То есть это прямо тот путь реальный, где э на вашем компьютере находится система Eidos, и там э файлы исходных данных.

4. Исследование исходных данных

4.1. Структура и содержание файла

Вот этот файл исходных данных, ребята, вот смотрите. Что мы здесь видим? Что у нас первая колонка - это марка и модель автомобиля. Вторая колонка, классификационная шкала - это ценовая категория. Ну, то есть, грубо говоря, так: дешёвая машина, средняя она по цене или дорогая. Потом марка идёт, потом марка и модель, потом какой кузов, год выпуска, объём двигателя, пробег какой, объём двигателя, цвет, тип коробки передач: автомат, механика. Вот. Даже есть вариатор, робот. Вот. И какая отделка салона. Вот. Ткань, комбинированная отделка. Вот. Кожа. Велюр даже вон там есть. Разные варианты отделки салона.

4.2. Размер данных и область применения

И у нас здесь, ребят, уже довольно приличное количество объектов обучающей выборки. 401 объект обучающей выборки. То есть пример, примеры, 401 пример машин с пробегом. Вот. Давайте посмотрим на статью. Я вам немножко в двух словах расскажу об этой задаче.

5. Аналогия с экспертной оценкой

5.1. Как специалисты оценивают автомобили (толщина краски, повреждения)

Значит, ребята, значит, когда специалист оценивает машину с пробегом, как он это делает? Кто-нибудь видел, нет? Я видел. У него такой есть приборчик, который измеряет толщину краски, покрытия. Ну, то есть расстояние до металла, грубо говоря. И он этим приборчиком пробует на дверях, на крыльях машины, обходит вокруг машины и так вот касается ее, и этот прибор измеряет толщину покрытия. Вот. Я могу вам сказать, что если машина была в аварии, то там будет уже шпаклевка какая-то, вот, и еще там много слоев, и потом краска. Вы про это знаете. И обязательно толщина будет отличаться от заводской, обязательно. Вот. И сразу моментально он это увидит, что дверь-то там уже другая на машине. Или крыло-то вот это вот левое-то, оно ведь стукнутое вообще-то.

5.2. Учет возраста и амортизации

Выровнено, всё покрашено профессионально, на глаз не видно. То есть если так вот приглядываться даже, смотреть, специалист нечуттер не увидит, даже э который вообще в этой области разбирается хорошо, потому что сделано профессионально, в хорошем сервисе, подобрана краска профессионально. То есть там всё, как говорится, как положено сделано. Но этот приборчик обнаружит, что машина битая. Ну скажет: "Так, ну тогда у вас эта машина-то битая, получается, что э цена-то будет теперь ниже на какую-то сумму там или на какой-то процент". Потом говорит: "Так, а какого года машина у вас?" И вы даёте, значит, ему техпаспорт. Он: "А, такое 3 года уже или 4 года. Ну тогда, значит, минус 30%. То есть каждый год 10% цены сгорает в огне времени".

5.3. Знание рынка и стратегия ценообразования

Так, потом смотрит ещё что там такое. И вот так вот ходит, ходит по ней, смотрит или так рестайлинг есть, да? Это плюс, это увеличивает стоимость на 5%. В общем, короче говоря, что он делает? Он берёт какую-то базовую цену этой машины и к ней добавляет или удаляет, высчитывает из неё какие-то суммы, и э приходит к цене, которая обоснованная в настоящее время рыночная цена.

Ну, могу вам сказать, что, в принципе, хороший специалист, он может достаточно быстро это сделать. Он просто понимает и знает э рынок, знает конъюнктуру, знает спрос и предложения и знает, что такую машину вот за такую цену он продаст за 400.000. А вот за 450 уже едва ли. Она будет стоять там до скончания века, её никто не возьмёт за 450. Он это знает прекрасно.

5.4. Перспектива продавца и риэлтора

И говорит э покупа- продавцу, хозяину говорит: "Я вот могу вашу машину э взять за 400.000 на продажу, но вам будет 350, потому что фирма должна заработать". Вот. А хозяин говорит: "Ну ничего себе, я вот в интернете смотрел, машина такой даты выпуска, вот этой модели, марки, не битые, в хорошем состоянии, обслуживалась машина, техобслуживание проходила. Вот, с таким вот пробегом, как у неё, она стоит там 450-500". Говорит: "Я вот вас не понял, вы пришли продавать машину или хотите торговаться, как на рынке? Ну давайте, хорошо, выставляйте её на Авито, ставьте там 500 и ждите", - говорит, "до скончания века, когда кто-нибудь обратится". "Я могу вас заверить", - говорит этот риэлтор, - "что никто не обратится, не спросят вашу машину за такую цену". Вот. Это реальная картина.

6. Методы оценки активов

6.1. Метод восстановительной стоимости (затратный)

Значит, теперь, как можно, э какие существуют методики оценки машин? И не только машин, а вообще объектов э продаж и объектов недвижимости, здания, сооружения и объектов, которые двигаются, то есть машины.

Значит, есть всего три метода. Один метод называется восстановительный метод. То есть смысл этого метода в том, сколько стоит создать такой объект вот сейчас, в этом месте. Вот взяли там где-то, нашли чистое поле, там купили там шесть соток в этом чистом поле, где-нибудь там в предгорье там, допустим. И взяли там и построили дом за 4,5 млн руб. То есть взяли, купили все материалы, вырыли яму, сделали фундамент, всё, как положено. Э-э, потратили на это дело 4,5 млн. И выставили его на продажу и думаете, что кто-то его купит и вы заработаете. И поставили цену 5 млн ему или там 5,5. То есть дом хороший, очень хороший. Вот. Рядом там нет никаких коммуникаций. Вот. Ну не в этом дело. Значит, э-э, люди смотрят на этот, на это объявление и удивляются, думают: "Вот надо же было в таком месте построить. Вот, вот этот дом был бы вот он стоял бы, вот он в Динской там где-нибудь или там хотя бы там, э, ну, в общем, в каком-то населённом пункте, цены бы ему не было, понимаете?" А вот там, ну он никому там не нужен, понимаете? Вот. И говорят: "Ну мы мы взяли бы так, пожалуй, там за грибами там приехать, чтобы было где остановиться". Но дело в том, что э-э, дом за 4,5 млн за грибами, как-то оно хороший дом, понимаете, жилой. Вот. И люди говорят: "Нет, ну это цена завышенная, не будем мы его брать". Это же я потратил 4,5 млн, чтобы его построить. Говорят: "Ну, дурак, что потратил", - так типа такого тебе, ну так не скажут. Ну говорят: "Ну это ваше дело", - говорят, конечно, говорят, сколько тратить. Ну видно вам девать некуда было, вы их, можно сказать, что выкинули. Потому что этот дом в этом месте не будет столько стоить никогда. Вот.

6.2. Сравнительный метод (аналоги)

Значит, какой применён метод в данном случае? Применён метод э сравнения. То есть, э и метод э восстановительной стоимости. Да, для того, чтобы восстановить этот дом, построить его, нужно потратить такие деньги. Но за эти деньги его там не купят.

6.3. Доходный метод

А какой метод был применён для того, чтобы определить, э за какие деньги его купят? А для этого нужно аналоги найти, аналоги. То есть нужно найти аналоги и посмотреть, сколько они стоят на рынке. И вот эти аналоги, они, если они действительно похожи на этот дом, который продаётся, то можно по их цене судить о том, сколько он может стоить. И существует третий способ. То есть первый способ - это восстановительная стоимость, сколько его стоит построить просто этот дом в этом месте. Второе - это восстановительная стоимость может быть завышенной, может быть заниженной. Это его как бы себестоимость, затраты на такое создание такого дома. Э-э, метод аналогий, сравнительных продаж, он показывает, сколько стоят такие дома в таком месте. И вы можете опираться на это, чтобы понять, сколько может стоить ваш дом. И третий способ оценки - это обычно касается уже м-м каких-то фирм, когда фирмы оценивают, какую прибыль это может дать. Вот чем больше это может дать прибыли этот объект, тем он будет стоить дороже. Даже если он сам по себе дешёвый, например, этот объект, а он прибыль большую даёт. Ну тогда он может стоить дорого. Вот, допустим, этот дом стоит в таком месте, где очень много туристов. Вот они туда приходят, там отдыхают там, э-э, ну, в общем, занимаются своими делами. А потом у них возникает такой вопрос: где бы вот, значит, ещё можно было бы переночевать? Хотя у них там обычно палатки есть, еда. Ну, в общем-то, вот если дом в таком месте стоит, где э как гостиницу его использовать и как столовую, то, может быть, он даст прибыль тогда. И вот тогда возникает некий интерес к этому уже, как к объекту, который может приносить прибыль. Вот три способа: восстановительная стоимость, аналогии, метод аналогии и э метод э-э получения, оценки потенциальной прибыли от этого объекта.

7. Применение системы Эйдос (сравнительный подход)

7.1. Обзор методологии (сложность оценки риэлторами)

И вот э мы рассмотрим метод аналогии. Если мы посмотрим на то, как э разрабатываются методики риэлторской оценки и используются. Вот значит, смотрите, значит, я вам показываю оценка бизнеса сайт. Значит, вот здесь у нас э пример методики э оценки. Как разрабатывается методика оценки? Я могу вам сказать, что, в общем-то, это дело непростое. То есть берётся, ну я вам сказал, в принципе, общий принцип такой. Но я специально вам показываю это для чего? Для того, чтобы вы представили себе, что, в общем-то, это работа такая серьёзная. Я бы сказал, даже заумная, наверное, в какой-то степени. То есть здесь прямо много, понимаете, всего. Вот когда много возникает всяких вот таких моментов, то это уже сложно. То есть высокая сложность разработки такой методики. И суть этой методики заключается в том, суть разработки методики, что берётся некая базовая цена и к ней добавляются, удаляются какие-то коэффициенты. Разработка такой методики может стоить определённых денег, требует определённых затрат.

7.2. Поиск аналогов (задача распознавания образов)

Ну и когда мы ищем аналог, то возникает вопрос: а что это за задача поиска аналога? Вот, допустим, у нас есть база данных, в которой описаны объекты недвижимости, те же самые дома. И нам нужно найти аналог. Значит, тогда мы должны применить технологию, похожую на ту, которая применяется в системах распознавания образов. То есть мы должны создать образы этих объектов, вот, на основе каждого из этих объектов создать свой класс, не с одного категорию класса, а именно вот образ этого конкретного объекта. И сравнивать эти объекты друг с другом, проводя кластерный анализ этих классов, созданных на каждый из которых создан на основе одного примера. И мы тогда найдём аналог. То есть найдём наиболее похожие объекты на наш этот объект. И посмотрим, сколько они стоят. Это задача распознавания образов.

7.3. Моделирование зависимости цены от характеристик

То же самое касается определения цены. Мы можем найти не просто похожие объекты, а найти, сколько они стоят, похожие объекты. И предположить, что наш тоже, наверное, примерно столько же и должен стоить. Потому что если он будет стоить намного больше или намного меньше, то есть будет заметно отличаться, то люди не поймут, они скажут: "А почему это ваш этот вот стоит намного дороже, чем вот такой же точно вот тут продаётся?" Сейчас же информация у людей есть. Они посмотрят, скажут: "Ну как-то необоснованно вы просите столько". Ну, ну, просить-то можно, но нужно понимать, что могут тебе отказать, скорее всего, и скорее всего откажут. Так вот, ребята, мы решаем задачу сейчас сравнения э нашего конкретного автомобиля с конкретными свойствами, определёнными, имеющим конкретную стоимость, с другими автомобилями и формирование обобщённых образов различных ценовых категорий. После этого мы увидим модель, мы получим модель, которая будет отражена зависимость стоимости машины от различных её характеристик.

7.4. Источник данных для лабораторной работы

Значит, э ребята, когда я тут, значит, где я взял эту таблицу вообще? Значит, я взял её на сайте, на котором ссылочка есть, но сейчас там уже несколько другое там находится, наверное. Сейчас посмотрим. Где этот сайт? Вот он. Ну всё, уже его нету. Слишком много прошло времени. Хотя, в общем-то, фирма осталась, но Там, значит, я искал сайт, ребята, где продаются машины, но такой, где был бы прямо прайс приведён, прайс-лист. И нашёл. То есть мне нужно, нужен был прайс-лист. Я нашёл такой сайт, этот прайс-лист скачал оттуда и сделал это исследование провёл.

8. Генерация и анализ модели в Эйдос

8.1. Запуск модели (начальные параметры)

Значит, э смотрим теперь, что у нас получилось. Мы ввели эту информацию в систему, эту таблицу. И у нас сформировались ценовые категории, это класс, классы. То есть мы всего лишь три ценовые категории выбрали, то есть это дешёвые машины, средние, дорогие. И э есть у нас описание этих машин: марка, марка плюс модель, какой кузов, год выпуска - числовые диапазоны, 12 диапазонов, пробег какой - тоже 12 диапазонов, диапазонов объём двигателя - это всё числовые шкалы. Цвет какой, цветов очень много оказалось. То есть там палитра широкая, гамма цветовая. Какая коробка? Автомат, механика, вариатор, робот. И В. Но В, наверное, это тоже вариатор, скорее всего. И какая отделка салона? Вот такие характеристики машин. Там были ещё характеристики, я там под таблицей написал, что ещё было, ещё был вид топлива там, бензин, дизель, и есть ли там компрессор или нет. Но я не стал с этим возиться, потому что там всё в одной колонке было, и числовые, и текстовые наименования, добавлены к числовым. То есть нужно было повозиться немножко, чтобы эту информацию оттуда вытащить. Ну надо было сделать колоночки ещё, вид топлива и наличие компрессора и туда перенести информацию из этой колонки по 400 там строкам вручную. Вот. И в этой колонке, где это было, убрать буковки. Вот. В общем, короче, работать надо было. Ну я не стал, я просто взял, удалил эти колонки и всё. Теперь вот насчёт кузова, здесь вот есть такой кузов 175, видите, вот. Это там, так не было написано, это я написал, когда там не было информации о том, какой кузов, то я взял, написал неизвестный тип кузова. Пошутил немножко. Вот. И у нас выборка обучающая, смотрите. Уже там в этой выборке 400 машин. 401 машина.

8.2. Проверка надежности (начальная)

Теперь мы что делаем? Создаём модели и проверяем их на достоверность. Можно это делать на графическом процессоре, потому что я исправил неточность, которая была при расчёте на графическом процессоре модели PRC2. Считаются эта модели, эти модели считаются и ещё 10 выходных форм. И у нас это всё проходит сравнительно быстро, там около минуты это займёт, наверное, даже меньше, 45 секунд. Мы можем посмотреть, какая достоверность моделей. Достоверность моделей где-то примерно на три балла достоверность. Но есть модели и неплохие. Вот 920 сейчас посмотрим. Вот. Ну, видно, что зависимости были выявлены между наличием свойств определённых машины и её стоимостью. Видно, что достоверных решений гораздо больше, чем ложных, истинных решений больше, чем ложных.

8.3. Эксперименты с параметрами (адаптивные интервалы)

Теперь давайте, знаете что, сделаем, проведём эксперимент. Значит, поскольку эта задача уже не совсем лабораторная работа, а похожа уже на реальную задачу, значит, я вам показываю впервые, ребята, наверное. Вы слушайте внимательно сейчас, я вам впервые показываю, как вводить информацию из собственных каких-то м исходных данных, которые мы где-то нашли, представили в нужном стандарте и записали и и хотим сделать приложение. Вот смотрите, значит, у нас есть папочка с системой, папочка с базами данных этой системы. И там есть папочка InpData. Ну, можно догадаться, что это входные данные. И в этой папочке у нас, видите, файл InpData находится, который мы, значит, э где-то нашли и записали сюда. Ну, возможно, немножко пришлось его переделать. Вот мы этот файл хотим ввести, но для этого мы только лабораторные работы устанавливали, используя вот эти кнопочки, да, вот добавить работу. А сейчас я вам показываю, что можно просто этот файл туда записать и запустить режим 2.3.2.2. Вот тот самый программный интерфейс, который мы используем. И вот здесь вот, смотрите, мы можем задать, что интервалы у нас не одинаковые с разным числом наблюдений, а разные по размеру интервалы с одинаковым числом наблюдений, адаптивные интервалы. При адаптивных интервалах получается достоверность модели выше.

8.4. Проверка надежности (второй запуск)

И когда у нас появится калькулятор, сейчас преобразование данных произошло, то мы вот здесь вот можем посмотреть, сколько у нас примеров по каждой шкале. Вот здесь вот у нас по каждой градации, по каждому числовому диапазону 133-134 примера. А я сказал, что разные по величине интервалы. Здесь вот 32 размер интервала, 34, а здесь аж 3 млн размер интервала. Значит, почему? А потому что таких вот данных с высокой ценовой категорией намного меньше, чем вот этих. И по нему тоже 134 примера. А почему 134, 133, а не какое-то одинаковое количество? По той же простой причине, по которой, если три землекопа копают две ямы, то на одну яму приходится не полтора землекопа, а приходится на одну яму один землекоп, а на другую два землекопа. Вот по этой же причине и здесь. Значит, э целое, дело в том, что эта задача решается в целых числах. То есть это что значит? Что у нас не может быть дробного числа наблюдений, оно всегда целое. И здесь это целое число наблюдений и показано. Значит, в текстовых шкалах не написано, сколько наблюдений. Хотя, в принципе, и можно было написать, посчитать. А вот в числовых шкалах, смотрите, значит, э написано 33 наблюдения. Ну это вообще-то достаточное количество для того, чтобы ну видно уменьшить, то есть здесь, значит, описательные шкалы, 34 наблюдения. Ну можно уменьшить даже число интервальных значений. Вот. Вполне оправдано будет. А вот э ценовую категорию можно увеличить число значений. Давайте мы это и сделаем. Здесь напишем, допустим, 15 или вот 10 напишем здесь, 10, 10 ценовых категорий. А здесь напишем пять, например, интервальных значений в числовых шкалах. И, обратите внимание, здесь у нас матрица э три колоночки, три класса. Вот, но 272 э строчки - это значение свойств. Мы пересчитываем, у нас получается 10 колоночек, 251 э строчка, 251 строчка. И выходим на создание модели. И смотрим, какая получилась достоверность модели у нас по оценке. То есть сейчас я вам показываю, что можно, по идее, э пытаться задавать какие-то параметры. Да, создаём модели и смотрим, какие у них достоверности, конечно, после создания, ясное дело. Значит, вот сейчас э посчитаются модели, и мы можем сравнить достоверность э при тех параметрах создания модели и при этих. И вот так вот, если создаётся реальное приложение, которое мы собираемся реально вот использовать для научных целей или для практических целей, то есть смысл попытаться несколько раз вот так вот задавать разные параметры и понять, какая зависимость достоверности модели от этих параметров и выбрать, постараться выбрать наилучшие параметры, при которых модель наиболее достоверна. Когда-то я делал э программы, которые это автоматизируют этот процесс. Но они имели специальное назначение, не универсальное. Хотя идея универсальная, но я её в универсальной форме не не реализовал в системе. А в специальном виде я пробовал это делать, получается. То есть там для конкретных задач получается неплохо. То есть там за определённое время было просмотрено большое число моделей с различными параметрами. Они автоматически генерировались и автоматически проверялись на достоверность. И, в общем, потом была выбрана наиболее достоверная, и мне были решены задачи. Достоверность модели меняется от этих параметров. Кстати, длительность расчёта немножко выросла. То было там сколько, 45 секунд, а сейчас минута 26 секунд. И смотрим, что у нас получилось с достоверностью. Вот. Ну, во-первых, изменилась достоверность. Вот. И, скажем так, Выросла она или не выросла? Сейчас видно, что вот эти вот распределения числа истинных и ложных положительных и отрицательных решений очень напоминают нормальные распределения, сдвинутые относительно друг друга.

8.5. Выбор лучшей модели (INF5)

Ну, выберем модель, у которой наибольшая достоверность. Вот, это модель либо INF4 по критерию L1, либо L5 с э с интегральным критерием резонанс знаний, коэффициент возврата инвестиций, кстати, коэффициент.

8.6. Решение задачи идентификации

И будем в этой модели, этой моделью пользоваться для решения задач. Будем считать, что мы используем модель INF5, наиболее достоверную модель. Или, сейчас посмотрим ещё. Ну вообще-то можно INF4. Вот, даже вот её, пожалуй. Вот мы видим, что начиная с 10%, достоверность растёт, уже становится положительной достоверность. И, в общем, по критерию L1 уже модель имеет достоверность 0,8, что, в общем-то, неплохо. Вот. И мы эту модель можем и выбрать. А INF5 интересно, она по двум критериям является наилучшей: по L2 и по L1. L2, L1, а это по L2. Вот. Ну, давайте INF5 возьмём текущей моделью. Заходим в режим 5-6, выбираем INF5, коэффициент возврата инвестиций. И в этой модели решаем задачу идентификации.

9. Анализ результатов

9.1. Обзор результатов идентификации

Значит, э-э, мы идём полностью, следуем вот этой логике преобразования данных в информацию, а её в знания, которые мы уже рассматривали на предыдущих работах. То есть мы формализовали предметную область, осуществили синтез и верификацию модели, причём несколько раз попробовали разные модели с разными параметрами, выбрали наиболее достоверную модель, решили задачу идентификации уже, но ещё не смотрели результаты этой, решения этой задачи. Потом будем остальные задачи решать.

Значит, смотрим результаты решения задачи идентификации. Ну вот у нас есть автомобили, и они отнесены к ценовым категориям. Вот. И внимание, видите, вот ценовая категория, и это соответствует действительности. Но иногда не соответствует. Вот. Мы видели, какова достоверность решения этой задачи.

9.2. Надежность идентификации классов

Мы можем посмотреть на форму интересную, это здесь форма 4.1.3.2. Здесь мы видим ценовые категории, а здесь видим машины, соответствующие этой ценовой категории. И видим, что некоторые машины, хотя и соответствуют, птичка стоит, и системой определены как соответствующие, но они как-то не очень высокий уровень сходства. И мы видим, что есть и ложные отрицательные решения. То есть на самом деле машина вот эта вот Ford Focus 2 относится к этой ценовой категории, а система не отнесла, видите, показала, что она с минусом имеет сходство с этой ценовой категорией. Вот. Ну и так по каждой ценовой категории есть у нас результаты. И вот мы можем посмотреть, как вообще классы, какие классы хорошо идентифицируются, а какие не очень хорошо. Можем рассортировать классы в порядке убывания достоверности идентификации. И мы видим, что высокая ценовая категория хорошо идентифицируется, низкая, средние вот эти вот, то есть тоже довольно дорогие машины 9 10, 8 10, 6 10. А вот эти вот, они похуже. И хуже всего вот вторая, пятая градация из десяти. И видим, что всё равно модель INF4 является наилучшей в большинстве случаев.

9.3. Влияние признаков на ценовые категории

Ну, теперь мы можем провести некий анализ. Можем посмотреть, а что, как влияет на принадлежность к той или иной ценовой категории? Ну что влияет, что обуславливает принадлежность э к э наиболее дешёвым машинам, первым из десяти цен этих вот э первая градация, первый диапазон из э десяти. И мы видим, что э скорее всего, это машина Lifan, Chevrolet Lacetti, Ford Fusion, Hyundai Getz, Land Rover, Lifan, Renault, Renault. Понимаете? То есть мы видим машины этой ценовой категории, как они связаны, то есть что обуславливает принадлежность к этой ценовой категории. Мы видим, что в основном на это влияет марка и модель. А вот машины дорогие взять, если, тогда что на них влияет? Тоже марка-модель в основном. Марка-модель: BMW X5, X6, Infiniti FX37S, Lexus, звучит как оно, да? Хорошо. То есть мы видим, что тоже обуславливается в основном маркой и моделью. И другие вещи тоже влияют, кроме марки и модели. Вот. А вот как влияет объём двигателя? Посмотрим фильтр по объёму двигателя устанавливаем и получаем такую информацию, что очень дорогие машины имеют огромный объём двигателя. Либо максимальный, либо либо очень большой. Ну если пять градаций, то это так можно сказать: очень маленький, маленький, средний, большой, очень большой, да, пять градаций. То есть мы имеем сейчас вот большой и очень большой объём двигателя до 5,7, 3,2, 5,7, 2,2, 3,2. Вот. И для них что не характерно для этих машин? Для очень дорогих машин. Для них не характерно, чтобы это был седан. То есть седан - это дешёвые машины, наверное. Ну, не самые дорогие. Э это не старые машины. То есть вот этот год выпуска 1.5 - это самые старые машины, девяносто шестого года. Вот. И отделка салона у них не ткань у этих вот дорогих машин. И пробег у них не 120.000 км, ребята. И цвет не красный. И пробег у них не 460.000 км. Понятно? Вот. То есть мы видим картину, что обуславливает принадлежность к этой ценовой категории. Прежде всего по объёму двигателя - это очень большой объём двигателя. Мы здесь фильтр сделали по объёму двигателя. И не вот эти параметры. Но я могу вам сказать, что когда вот мы читаем про эти параметры, что это не будет очень дорогая машина, то для нас это вполне очевидно. То есть это очень дорогая машина не может быть у неё там отделка ткань, не может быть какой-то сумасшедший пробег там около 500.000 км.

10. Дальнейшие варианты анализа (Семантическая сеть признаков)

Значит, давайте ещё раз посмотрим, значит, просто пролистнём вот так. И посмотрим. Здесь марки, модели. А что для них характерно для этих машин? Ну, кроме марки? Марка Porsche там, это вот мы видим, что больше всего BMW X5, Infiniti, Lexus, наиболее Porsche Panamera, да? Вот. Land Rover Sport, Toyota LC. Вот, Люксовские, да? Вот. А потом мы смотрим, что дальше? Э, объём двигателя очень большой. Вот. Кузов непонятно какой вообще тип этого кузова. Вот. Пробег. Давайте на на год выпуска или на пробег посмотрим. Значит, э статья написана в 2013 году. А здесь написано год выпуска 5.5, 2010-2013. То есть это машины новые. Новые машины. Вот. Год выпуска фильтр. То есть по году выпуска это у нас либо максимальный год выпуска, либо большой год выпуска. То есть выпущены недавно, которым, ну, грубо говоря, там не очень много, если в тринадцатом году статья написана, то это, значит, здесь получается, ну, не больше 5 лет этим машинам. А здесь посмотрим год выпуска. А здесь год выпуска машины очень старые, девяносто шестого - 2007 и до 2007 до 2008. А здесь начиная с 2008 года машины. То есть э год выпуска влияет совершенно однозначно на цену. То есть новые машины стоят дорого, старые - дёшево. Вот. И вот так вот мы можем по всем параметрам смотреть.

10.1. Кластеризация признаков (Сходный смысл)

Но мы можем посмотреть теперь, ребята, и на влияние тех или иных свойств машины. Вот возьмём, допустим, для риэлтора, для них это задача определить, как влияет, допустим, там, э если машина такой-то марки модели, вот, допустим, хэтчбек, как это влияет на то, какая машина? Это влияет так, что цена будет у неё очень низкая у этой машины. Вот. То есть на ценовую категорию это влияет так вот, что цена низкая. Вот. А если вот будет седан, тогда как это влияет? Тогда это более сложно влияет. Ну тоже не самая дорогая машина. То есть седан дорогой не будет. Очень дорогой не будет седан. Вот. То есть мы можем вот так вот посмотреть. Теперь давайте ещё сравним, мы сравнили сами, то есть мы сравнивали конкретные машины с обобщёнными образами классов, а также э сравнивали, то есть определяли, какие характеристики у этих обобщённых образов различных ценовых категорий. А посмотрим теперь на то, какие свойства машины имеют сходный смысл. О, это у нас долго будет. Свойств много, 250 свойств. Вот здесь вот у нас приведённые экселевские формы, где они находятся в папке приложения, в которых это не в графическом виде представлено эта информация, а табличным. Причём эти таблицы все открываются в Экселе, ребята. Я когда проектировал систему, я использовал базы данных, которые открываются в Экселе. Вот. Ну и правда, ещё я записал с расширением экселевским, чтобы было проще пользователям понять. Значит, смотрите, ребята. Значит, вот у нас признак такой, э отделка салона ткань содержит 4,5% суммарной значимости всех признаков. Объём двигателя от 1 до 1,6 л, ну маленький, минимальный объём двигателя, 1.5 градация, содержит 4,4 суммарной значимости. Отделка салона кожа тоже содержит 4,35. Джип тоже содержит очень много, 4% значимости. Автомат, коробка, объём двигателя бешеный совершенно 5,7 л, содержит 3,3 э десятых суммарной значимости. То есть мы видим вот здесь вот, какие признаки наиболее ценные. А здесь в конце признаки идут, которые, смотрите, содержат 9% суммарной значимости от того, какая суммарная значимость всех признаков. 9. Понимаете? Вот, допустим, э кузов лифт написанный, не знаю, что это такое. Значит, э там какие-то шутники видно таблицу эту заполняли этот прайс-лист. Значит, вот, допустим, бежевый металлик - это содержит 9% суммарной значимости. Вот. Иногда марка-модель тоже не очень ценной является. И вот мы вопрос возникает такой: как убрать из модели признаки, которые не очень ценными являются? А вот здесь вот нарастающим итогом это суммарная зна- это значимость признака нарастающим итогом показана. И вот смотрите, вот у нас всего лишь э 8% наиболее значимых признаков обеспечивают 50% суммарной значимости. Здесь вот 8% мы видим, а здесь 50% суммарной значимости. То есть всего 8% признаков половину суммарной значимости обеспечивают. А вот 50% наиболее значимых признаков, смотрите, здесь вот 50% вот. Они обеспечивают 87% суммарной значимости. То есть половина наиболее ценных признаков обеспечивает почти 90% суммарной значимости. Ну, в разных моделях это по-разному выглядит. Ну вот, здесь вот в модели N6 даже ещё больше. 50% обеспечивает 88% суммарной значимости. Вот. То есть это вот и есть вот эта кривулечка, которую я там показал вам, рисовал её графически. Вот она. Узнаёте, да?

11. Заключение и последующие шаги

11.1. Обзор задания

Так, это уже конец занятия, что ли, да? Ребят, э значит, на следующем занятии мы будем рассматривать модель э графически, графического анализа, модель, которая позволяет э графические объекты обобщить, изображения этих объектов, проанализировать, идентифицировать. То есть работа с графическими объектами.

11.2. Следующая сессия: Графический анализ

Ребята, значит, и вы можете мне напомнить, чтобы я вот это тоже дальше вам рассказал. Но это, видимо, уже на следующей лабораторной, на примере графических объектов расскажу вам. Вот. Ну, на этом мы занятие заканчивается, ребята. До свидания. Всего самого-самого хорошего вам.

11.3. Прощание

До свидания.

До свидания. До свидания. До свидания. До свидания.