***ФГБОУ ВО «Кубанский государственный аграрный университет имени***

***И.Т. Трубилина»,***

***Российская Федерация***

***Колесников Роман Юрьевич, ПИ2102***

***roman563412@gmail.com***

**Лабораторная работа.**

**Заголовок**

Важность и методы проверки достоверности аналитических моделей данных

**Резюме**

I. Введение и Постановка Проблемы

A. Начальная Настройка и Приветствия: Обсуждение начинается с приветствий и настройки демонстрации экрана, показывающей материалы лабораторной работы.

B. Ограничения Матриц Абсолютных Частот: Анализ необработанных матриц частот (абсолютных значений) может вводить в заблуждение из-за разных размеров выборок по категориям. Прямые выводы на основе таких матриц неосновательны.

II. Относительные Частоты и Начальный Анализ

A. Использование Относительных Частот для Сравнения: Переход к относительным частотам (процентам или долям) позволяет проводить более корректные сравнения между различными категориями объектов.

B. Условные и Безусловные Частоты: Различаются условные (внутри конкретной категории) и безусловные (по всей выборке) относительные частоты. Сравнение этих частот дает понимание характерности признака для класса.

C. Необходимость Автоматизированного Анализа Больших Матриц: Визуальный анализ больших матриц частот (сотни признаков, десятки/сотни классов) неэффективен и требует автоматизированных методов сравнения.

III. Методы Автоматизированного Сравнения

A. Хи-квадрат (Пирсон): Мера, разработанная Пирсоном, сравнивает фактическую и теоретическую (ожидаемую) частоту признака в классе путем вычитания. Хорошо показывает наличие связи, но не её силу в терминах кратности.

B. Мера Информации (Фартиевич) и ROI: Альтернативные подходы, основанные на делении фактической частоты на теоретическую. Мера Фартиевича использует логарифм отношения (количество информации), а ROI (коэффициент возврата инвестиций) вычитает единицу из отношения. Эти меры показывают, во сколько раз признак более (или менее) характерен для класса, чем ожидалось.

C. Сравнение Методов: Методы на основе деления (информация, ROI) нормированы относительно отсутствия связи (0 для информации, 1 для ROI без вычитания единицы) и позволяют оценивать кратность связи.

IV. Ключевая Роль Проверки Достоверности Модели

A. Определение Вероятности и Практические Ограничения: Вероятность – это предел относительной частоты при неограниченном увеличении выборки. На практике выборки конечны, поэтому использование термина "вероятность" вместо "частота" вносит погрешность, уменьшающуюся с ростом выборки. Точность стоит денег (сбор данных).

B. Цель Моделей: Модели создаются для решения задач: идентификации (распознавания, классификации, диагностики), прогнозирования, принятия решений (управления) и исследования предметной области.

C. Опасности Непроверенных Моделей: Использование моделей без проверки достоверности крайне рискованно. Применение таких моделей для выработки рекомендаций или принятия решений может привести к серьезным ошибкам и убыткам (пример с диссертацией и фирмой).

D. Оценка Достоверности Модели: Достоверность модели проверяется путем решения задач (например, идентификации или прогнозирования) на ретроспективных данных и сравнения результатов с известными фактическими данными.

V. Обработка Многоклассовых Объектов и Специфика Системы

A. Многоклассовость Объектов: Реальные объекты часто относятся ко многим классам одновременно. Некоторые методы (например, Ван Рисбергена) исходят из предположения об одноклассовости, что ограничивает их применимость.

B. Подход Системы "Эйдос": Система "Эйдос" обрабатывает многоклассовые объекты, представляя физический объект как суперпозицию логических объектов, каждый из которых относится к одному классу.

C. Особенности Системы: Система "Эйдос" доступна бесплатно, что является важным преимуществом по сравнению с дорогими коммерческими аналогами.

VI. Заключение и Итоговые Замечания

A. Итог о Важности Проверки: Всегда необходимо проверять достоверность модели перед ее использованием для решения практических задач. Использование непроверенной модели – это профанация и авантюризм.

B. Риски и Профессиональная Ответственность: Применение непроверенных моделей несет большие риски. Необходимо осознавать ограничения используемых методов и данных.

Детальная расшифровка текста:

I. Введение и Постановка Проблемы

A. Начальная Настройка и Приветствия

Здравствуйте, ребята. На микрофон.

Здравствуйте.

Доброе утро, Евгений Вениаминович. Всем доброе утро.

Доброе утро, Геннадий Александрович.

Отметьте, пожалуйста.

Вот сейчас сделаю экран.

Вот. Ну начнём с обзора. Потому что постановление лабораторной работы, на которую мы будем ссылку рассматривать.

B. Ограничения Матриц Абсолютных Частот

Я не буду это комментировать, потому что вы это всё знаете.

Кирилл, скоро у тебя будет там Mercedes, которая. Вот это хорошая машина. Гранта.

Ну вот. Пересмотрим модель.

Мы остановились на чём? О том, что я показал модель абсолютных частот. Это ещё называется матрица сопряжённости.

Матрица абсолютных частот, матрица сопряжённости или корреляционная матрица. Как она построена? Значит, по колонкам у нас обобщающие категории, группы, которым принадлежат объекты выборки. А по строкам значения свойств или значения факторов. Я их называю признаки, сокращённо. А колоночки называю классами. Просто чтобы не говорить, что это градация классификационной шкалы и там значение свойства. Просто значение свойства - это признак. А градация классификационной шкалы - это класс. Градация описательная шкалы - это признак.

Вот мы видим, что здесь у нас число встреч какого-то признака, каждого признака у объектов определённой категории. Я вам говорил, что на основе такой матрицы непосредственно вот её делать выводы не основательно. Почему? Нет возможности корректно делать выводы, потому что у разных категорий разное число объектов. Вот, допустим, здесь мы видим три раза встречается признак размер подрубку и у средств связи, и у элементов компьютера.

И казалось бы, этот признак одинаково характерен для этих категорий объектов. Но это не так, потому что средств связи три, а элементов компьютера семь.

II. Относительные Частоты и Начальный Анализ

A. Использование Относительных Частот для Сравнения

Поэтому мы должны перейти к относительным величинам. Можно в процентах измерять. Это сделано.

Если мы посмотрим на размер подрубку, этот же самый признак, то у нас окажется, что 100% средств связи этим признаком обладают. И 42%,8 элементов компьютера. То есть здесь уже ясно, что для средств связи он гораздо более характерен, чем для элементов компьютера.

B. Условные и Безусловные Частоты

И можем ещё посмотреть в среднем, насколько часто он встречается по всей выборке. То есть эти вот 100% и 42,8 - это условные относительные частоты. Условные, в смысле, они для объектов определённых категорий. При условии, что объект относится к определённой категории. А безусловная относительная частота - это по всей выборке, по всем категориям.

И мы можем сравнить относительные условные относительные частоты с безусловными и друг с другом. И сделать на основе этого вполне обоснованные выводы, что этот признак более характерен для средств связи, менее характерен для элементов компьютера, но всё-таки и для элементов компьютера он более характерен, чем в среднем для всех объектов выборки.

Ну то есть всё-таки это характерный признак и для элементов компьютера, просто менее характерный, чем для средств связи. Вот. То есть мы можем вот так сравнивать и делать выводы на основе этого.

C. Необходимость Автоматизированного Анализа Больших Матриц

И было бы всё ничего, если бы матрицы эти могли быть, они могут быть очень большой размерности. Эти вот признаки, их может быть сотни, тысяч. Ну реально так, конечно, редко бывает. Реально бывает там 500, 700 там, 1800. Ну были случаи у меня, когда их было там и 11.000, например. А вот классов может быть в системе Эйдос вообще до 2.500. Вот. Ну обычно там их 20, 30, 40, ну 100, например, вот такое количество какое-то. То есть какие-то разумные количества. Ну как-то раз было и 500 там, 380, вот что-то такое, где-то 500.

Так вот, если будут матрицы 500, например, колоночек, здесь 14 колоночек, а представьте, что их 500, а строк, скажем, 11.000. Ну тогда понятно совершенно, что вот такое вот сравнение с помощью зрачков, как говорится, когда вот мы сидим и смотрим глазами просто на эту матрицу, и с помощью своего естественного интеллекта пытаемся всё это осмыслить, то это, конечно, делать бесперспективно, потому что очень большая размерность данных. Мы просто это не сможем прочитать. Если мы это распечатаем, будем читать, нам потребуется несколько лет, чтобы это прочитать, наверное. Ну даже, допустим, мы это и прочитаем. Ну и что? Мы же не сможем сделать выводы на основе этого. А если мы, мы должны делать выводы на основе уже сравнений.

III. Методы Автоматизированного Сравнения

Поэтому пришла такая мысль, она пришла мне в семьдесят девятом году, автоматизировать это сравнение. Как я это сделал? Я это сделал разными способами.

A. Хи-квадрат (Пирсон)

Один способ, вообще известный, это хи-квадрат. Это разница между фактической частотой наблюдения итого признака в житом классе и теоретической частотой. Эта мера разработана Пирсоном, разработчиком вообще параметрической статистики всей. Прекрасная, прекрасно себя она показывает. Чаще всего эта модель на основе хи-квадрат, она самая достоверная и наиболее так чётко всё в ней просматривается, все закономерности.

B. Мера Информации (Фартиевич) и ROI

Но если мы хотим сравнить два числа, какое из них больше, а какое меньше, то кроме вычитания, способа, который применил сам Пирсон, ещё есть возможность делить их друг на друга, да? То есть если мы сравниваем 7 и 5, какое число больше? Вычитаем из семи пять, получается больше нуля, значит семь больше. А если мы путём деления их сравниваем, тогда 7 делим на 5, получается величина больше единицы, значит семь больше, чем пять. Так вот, если мы потом хотим просуммировать все эти вот результаты сравнений, то понятно, что если никакой связи между наличием этого признака и принадлежностью объекта к житому классу нет, величина должна быть равна нулю, получается. Хи-квадрат мера Пирсона, она этому условию соответствует. То есть если фактическая частота и теоретическая равны, то разница их равна нулю. То есть нет никакой связи между наличием этого итого признака и принадлежностью объекта к житому классу. Если больше нуля, тогда есть связь, если меньше нуля, тоже есть связь, но отрицательная. То есть этот признак препятствует принадлежности объекта к этому классу.

А если мы делаем это сравнение путём деления, тогда вот это n и g-то разделится на теоретическую частоту, а оно равно сумме по строке частот признаков этой матрицы абсолютных частот, сумма по колонкам, по колонке, делённая на общее число признаков по всей матрице сопряжённости. Тогда получается вот это вот выражение. И для того, чтобы его нормировать к нулю, чтобы потом использовать эффективные интегральные критерии, то есть потом просто сложить эти частные критерии и получить некое одно число, характерное для класса житого, нужно взять логарифм от этого отношения, а можно просто вычесть единицу. Если мы берём логарифм, тут называется мера Фартиевича. Если вычитаем единицу, то вы, наверное, в курсе, слышали, это называется коэффициент возврата инвестиций, ROI, Return on Investment.

C. Сравнение Методов (Subtraction vs. Division)

Значит, получается, что в науке, в разных областях, были выработаны уже способы разумного сравнения этих вот чисел, разные способы. И если мы посмотрим на вот это отношение, которое получилось под логарифмом или в коэффициенте ROI, то мы увидим, что это фактически э-э, условная относительная частота, делённая на безусловную. Потому что n и g-то, разделённое на n g-то - это P и g-то. То есть безусловная частота наблюдения этого признака в объектах житой категории, житого класса. А n и-то, разделённое на n - это безусловная относительная частота наблюдения этого признака по всем объектам всей выборки. То есть получается, мы что видим? Э-э, что когда мы сравниваем эти фактическую и теоретическую частоты не путём вычитания, а путём деления, то это у нас получается логарифм условной относительной частоты, делённой на безусловную относительную частоту встречи этого итого признака в житой категории. Ну это и есть классическое выражение Фартиевича для количества информации, которое мы получаем, узнав, что объект имеет итый признак, о том, что он относится к житому классу.

Таким образом, вы видите, как интересно получается, что хи-квадрат тесно связан с количеством информации по Фартиевичу и тесно связан с коэффициентом возврата инвестиций, который применяется в экономике. Посмотрим на сами эти модели. Вот эта модель посчитана на основе варианта Фартиевича, количества информации. Мы здесь видим, что вот этот размер подрубку, он даёт нам этот признак 0,779 бит информации о том, что это средство связи и 0,531 бит о том, что это элемент компьютера. Как и должно быть. Мы знаем, что для средств связи он более характерен, чем для элементов компьютера. И для элементов компьютера тоже количество информации положительное, потому что вероятность встретить этот признак в этом классе объектов тоже выше, чем в среднем.

Вот. И мы видим, что есть числа и отрицательные. Здесь это что означает? Что у каких-то объектов определённой категории есть некоторые признаки, которые у них встречаются реже, чем в среднем. И наблюдение такого признака несёт информацию о непринадлежности к данной категории. То же самое мы видим в хи-квадрат. Модель хи. Здесь тоже положительное значение говорит о связи, наличия связи между наличием этого признака и принадлежностью к классу. А величина отрицательная тоже о наличии связи, но уже отрицательной. То есть наличие этого признака несёт информацию или говорит о том, что объект не принадлежит данному классу. Мы видим, что мера хи-квадрат у средств связи, вот этот размер подрубку 2,1, а у элементов компьютера 0,9. То есть практически в два раза больше связь между наличием этого признака и принадлежностью к классу у средств связи, чем у элементов компьютера. И если эта величина близка к нулю, то это значит, что фактически никакой связи и нет между наличием этого признака и принадлежностью объекта к классу.

IV. Ключевая Роль Проверки Достоверности Модели (Достоверность)

A. Определение Вероятности и Практические Ограничения

Ну я вам приводил такой пример, когда это там, наверное, когда рассказывал, э-э, про то, как осуществляется распознавание. Вот если есть у нас кто-то за дверью, неизвестно кто, надо установить кто. И мы можем задавать вопросы ему. Мы можем спросить там, допустим: "У тебя волосы длинные?" Он говорит: "Да". Это признак, который намного чаще встречается у девушек, чем в среднем по выборке. То есть он несёт информацию о том, что это девушка. Потом мы спрашиваем: "А брюки?" Он говорит: "Да, брюки". А брюки, наличие этого признака несёт маленькое количество информации, небольшое о том, что это ребята, и небольшое количество информации о том, что это не девушки. Почему небольшое? Потому что почти все девушки тоже в брюках. И разница между э-э, относительной частотой встречи этого признака в группе ребят и в группе девушки, она мало отличается от средней. Поэтому получается количество информации очень незначительное в этом признаке. А если бы я ещё спрошу: "А мобильный телефон есть?" Он скажет: "Да, есть". Это никакой информации не несёт о том, парень это или девушка. Почему? Потому что этот признак есть у всех э-э, с относительной частотой 100% встречается. И во всех группах, и в целом по всей выборке.

То есть получается три вида признаков: несущих много информации о принадлежности к какому-то классу или непринадлежности к другому. Признаки, ну это детерминистского типа признаки. Ну не могут быть не совсем детерминистские, а близкого к этому. А вообще детерминистские признаки - это те, которые однозначно говорят о принадлежности объекта к классу. То есть встречаются только в объектах этого класса, а в объектах другого класса не встречаются. Ну и признаки, которые не совсем так, но почти, почти так, близко к этому. И признаки статистические. Есть близкие к детерминистским, и близкие э-э, к бесполезным. Типа признак брюки. Вообще бесполезные. И что-то среднее между ценными признаками и не очень ценными.

Иногда странные эффекты наблюдаются. Какая-то рамочка появляется. Что это значит, эта рамочка была? Не знаю. Вот. Так что модель Info 1, она автоматизирует и другие модели системы когнитивного моделирования, автоматизирует сравнение условных и безусловных относительных частот.

Я несколько раз применил здесь термин информация, то есть информация, вероятность. Вместо термина относительная частота применил вероятность. И, соответственно, можно говорить об условной вероятности, безусловной вероятности. Я могу вам сказать, как они связаны. Значит, вероятность - это есть предел, к которому стремится относительная частота при неограниченном увеличении объёма выборки. Мы когда говорим, то иногда не слишком большую ответственность за то, что говорим, даже особо не вдумываемся в то, что говорим. Насколько вот эти слова важны и что они значат. И вот когда я вот это сейчас вот сказал определение, это определение из теории вероятности просто. То сразу возникает вопрос: а что понимается под неограниченным увеличением объёма выборки? Значит, я могу вам сказать, что реально такого неограниченного увеличения объёма выборки быть не может. Всегда она ограничена. Больше того, она довольно жёстко бывает ограничена. То есть часто встречаются небольшие выборки. Поэтому о вероятностях говорить, говорить не стоит. И во всяком случае нужно понимать, что когда мы используем термин вероятность, то мы допускаем некую погрешность в наших высказываниях, уклоняемся от истины, скажем так. Почему? Потому что относительная частота отличается от вероятности. Между ними есть некоторая разница всегда. Вот. И это вот и есть та погрешность нашего высказывания в случае, если мы применяем термин вероятность. То есть есть некоторая неточность. Но эта неточность, она стремится к нулю при увеличении объёма выборки. Но насколько быстро? Я вам скажу так, что где-то когда 480 наблюдений, около 400, ну больше 400, то уже около 5% погрешность между относительной частотой и вероятностью. Когда 1250 наблюдений, это из статистики известно, при равномерном случайном распределении величины случайной, то вот такая такая погрешность наблюдается, так она уменьшается. При 1250 получается погрешность около 2,5% различия между относительной частотой и вероятностью. И потом она стремится к вероятности, но всё медленнее и медленнее. Например, чтобы получить 1%, э-э, погрешности, нужно уже 10.000 наблюдений. То есть мы заметили, что сначала она быстро стремится к вероятности, потом всё медленнее и медленнее. И никогда её не достигает.

B. Цель Моделей (Решение Задач)

А теперь представьте себе, что если у нас э-э, информация стоит денег. Ну, допустим, мы опрашиваем респондентов и строим какие-то модели на основе этих опросов и анализируем эти модели. Опрос респондента, скажем, стоит рубль, к примеру. Ну это на самом деле рубль не стоит, но пусть пусть единица будет. Там 100 руб. или 1.000 руб. стоит. Вот. Ну давайте, допустим, будем считать, что он рубль стоит. Ну тогда получается, что для того, чтобы получить информацию с погрешностью 5%, нужно 400 руб. потратить. Чтобы 2%, 2,5, то есть 1.000, 1.200 руб., скажем, да? А если мы хотим 1%, ну тогда 10.000 нужно потратить, понимаете? Это уже как бы это самое, начинаешь задумываться, оно надо вот эти 1,5%, там ещё получить. 2,5% - 1.200 стоит, а 1% стоит э-э, 10.000. Понимаете? То есть мы добиваемся того, что получаем ещё 1% точности, то есть на 1% погрешность уменьшается, а платим за это 9.000. И вопрос возникает такой: а стоит этот процент точности 9.000? Если там мы занимаемся, допустим, инструментальными какими-то, где там у нас 15 знаков после запятой, там, допустим, считается достоверным результатом. Ну тогда может быть, может быть, там миллиарды долларов или может даже и больше, в большом адронном коллайдере там. А если мы занимаемся социологией, политологией, то этот 1%, может быть, вообще не очень важен. Хотя вот мы сейчас смотрим на выборы США, видим, что там, в общем-то, сейчас-то там явное преимущество. А вот выборы были Буш-Гор, там доли процента составили разницу общее число голосов. И какой-то ящик, который нашли где-то под столом, сыграл роль, и Буш стал президентом. Ну так говорят, по крайней мере, шутка такая основная. Ну, в общем, короче говоря, бывает, что это и стоит того, чтобы посчитать вот эту смену. Но редко так бывает. И обычно люди считают деньги, говорят: "Ну ладно, нам, в принципе, нас устроит вот 2,5%. Нам больше не надо". Вот. И они не тратят денег на опросы этих респондентов. Ну примерно вот смысл такой вот всего этого. То есть я хочу сказать, что точность стоит денег, понимаете? Чтобы получить более высокую точность, нужно больше объём данных. А сбор и вот система, обработка, всё это стоит денег. Поэтому возникает вопрос, насколько это необходимо, насколько это оправдано эти затраты.

Вот. Ну и когда мы модели получили, возникает вопрос: а можно ли пользоваться этими моделями? Значит, я сейчас вам задам такой вопрос, ну он такой риторический, сократовский называется вопрос. То есть я его спрашиваю, задаю этот вопрос вам не потому, что я не знаю ответа и хотел бы от вас получить ответ. А потому что я хочу вас навести на определённую мысль, к которой вы должны прийти сами, а не просто я вам должен это сказать. Самому вам это сказать не лучше, чтобы вы сами к этому пришли. Вот я хочу вас спросить, ребята, тем более лабораторная работа, вот, так что вы должны быть в диалоге со мной на этой лабораторной работе. Будет значит, на лекции ещё ладно, можно там послушать, хотя тоже было бы неплохо, если бы задавали вопросы в конце лекции. Вот. Но вопрос такой: э-э, зачем вообще нужна модель? Зачем мы создаём эти модели? Вводим данные, там какие-то расчёты проводим. Зачем это нужно? Кто знает? Ну под рукой ответа нет на этот вопрос.

Ну, наверное, чтобы упростить всё, проверить какую-то прежде чем осуществлять.

Ну, это ты ответил ты правильно, только на другой вопрос, не на тот, который я задал. Это ты ответил на вопрос, зачем проверять достоверность модели, который сейчас у нас следующий будет рассматриваться. А вообще, зачем она нужна?

Для анализа, прогноза, чтобы представить одну ситуацию.

Ну да, это вот конкретно ответили, а можно как-то более обобщённо ответить? Для решения различных задач. В частности, действительно, для прогнозирования, сравнения ситуаций, это всё это верно. Ну это просто частные случаи. А эти задачи, которые можно решать, их можно классифицировать. Я их классифицирую на три группы. Вот. Группировку сейчас я вам на экране покажу. Вот. Три группы задач. Вот это моё, по крайней мере, деление, моя классификация. Ну, по-моему, и другие нормальные люди тоже примерно так же классифицируют. Она логична. То есть есть задачи идентификации, распознавания, прогнозирования - это один блок задач. Второй блок - это задачи принятия решений или управления. И третье - это решение задач исследования предметной области путём исследования модели. Вот для того, чтобы эти задачи решать, каждая из них имеет очень много разных вариантов, в том числе и такие, которые с трудом позволяют разграничить эти задачи. Вот, допустим, принятие решений, оно тоже может потом прогнозирование осуществлять, понимаете, которое лежит в основе идентификации. Или в основе идентификации. То есть это всё переплетается. Потом при принятии решений, в простейшем варианте, может быть, мы и можем чисто эту задачу решить просто саму по себе. И то она может рассматриваться как задача обратная задача прогнозирования. Потому что при прогнозировании мы по факторам определяем, что произойдёт с объектом моделирования. А при принятии решений мы наоборот, по целевому состоянию объекта моделирования определяем, какие нужны факторы, чтобы объект перешёл в это состояние. То есть это обратная задача прогнозирования. Всё равно она связана с задачей прогнозирования. Вот. Но в развитой форме принятие решений предполагает ещё и использование некоторых результатов исследования предметной области, типа кластерного анализа. Понятно, да? То есть это мы рассмотрим попозже. А сейчас я могу сказать, что ответ был, в принципе, правильный, то модели нужны для того, чтобы решать задачи.

C. Dangers of Unvalidated Models (Dissertation Anecdote)

Ну и теперь вот вопрос, на который вы ответили, хотя я его и не задавал, потому что подумал, а вы уже там, наверное, телепатия у вас, конечно, там развита. То есть вы забегаете вперёд, по сути дела, предугадываете ход мысли мои и изложения. Вот. И отвечаете даже на вопросы, которые я ещё не успел задать, только подумал о них. И к ним подвожу, собственно говоря, изложение. Значит, все эти задачи, ребята, чтобы их решать, для этого нужно предварительно убедиться в том, что модель что? Что модель достоверна. То есть, что она адекватно отражает, правильно отражает моделируемую предметную область. Потому что если это не так, то мы при идентификации будем ошибаться. А идентификация - это слово, термин научный, который имеет синонимы: распознавание, классификация, идентификация, диагностика. Представьте себе, в медицине неверно болезнь продиагностировали пациенту. И будут лечить его от другой болезни. Ничего хорошего в результате не произойдёт, соответственно, правильно? Вот. То есть это очень чревато такая ошибка. Нежелательными последствиями. А если мы возьмём недостоверную модель и будем прогнозировать с помощью неё, а потом будем э-э, какие-то как-то корректировать свою деятельность с учётом этого прогноза. А модель недостоверная, прогноз такой, который сделан на основе недостоверной модели, скорее всего, не осуществится. Если осуществится, ну просто случайно так может получиться, что он осуществится. Но скорее всего, что нет. И тогда получится, что мы будем э-э, исходить в своей деятельности из неверных представлений о будущем, понимаете? Вот. Соответственно, наши затраты будут убытком практически, скажем так, пустую или в лучшем случае, а то ещё и какой-то вред нанесут. Ну, по крайней мере, такой вред, что они не принесут пользы, той или иной прибыли, той или иной пользы.

А если мы будем принимать решения на основе неверной модели, тогда что у нас будет получаться? Вот мы хотим оказать воздействие на фирму, чтобы она перешла от состояния предбанкротного в нормальное. Берём эту недостоверную модель, вырабатываем решения соответствующие, рекомендации. И эта фирма нам доверяется и выполняет эти рекомендации. Как вы думаете, у неё получается перейти в нормальное состояние, если модель недостоверна, если она неправильно отражает зависимости между причинами и следствиями в этой предметной области? Да на предприятие вообще может и обанкротиться в результате, понимаете? То есть весьма рискованно э-э, такие давать рекомендации, которые вот на основе непроверенной модели. Или, скажем так, недостоверной модели. А теперь будем, хотя это разные вещи, непроверенная модель и недостоверная модель. Сейчас я вам расскажу различие.

D. Assessing Model Validity

Ну и третья группа задач - исследование моделируемой предметной области. Вот мы когда исследуем модель э-э, и делаем вывод о том, что вот какие-то получили выводы какие-то на основе модели, то эти выводы имеют отношение к объекту исследования, объекту моделирования. Но только в том случае, ребята, они имеют отношение к объекту исследования, познания, к объекту познания. Только в том случае, если модель его отражает. Модель правильно отражает объект познания. А если она его не отражает, у нас нет никаких оснований переносить результаты исследования модели на объект э-э, моделирования. Понимаете? То есть мы не можем считать, что если мы такие закономерности или свойства этого объекта выявили на основе модели, то это действительно вот такими свойствами этот объект исследования обладает. Потому что вообще-то модель может его и не отражать, а точнее, неверно отражать. Значит, соответственно, и выводы о его свойствах у нас получатся неверными на основе этой модели. Понимаете суть? Мы будем так писать в книжках, что у него такие свойства там, а у него нет таких свойств. Потому что модель недостоверна.

Теперь, в чём различие между э-э, моделью, достоверность которой не изучалась и недостоверной моделью? Значит, модель, достоверность которой не изучалась, не исследовалась, она может быть как достоверной, так и недостоверной. И в различной степени. То есть о ней вообще ничего не известно. В частности, она может быть недостоверной. Поэтому, если, допустим, идёт какая-то защита, я присутствовал на очень серьёзных защитах, очень много раз. И могу вам даже привести реальный пример. Ну, без фамилии там и без этой детализации, ну вот так суть. Э-э, одна девушка защищалась, соискатель учёной степени. И вот она так рассказывала, что она там то, то сделала. Ну, я скажу так, между нами возникало некое сомнение, что она это сделала, потому что там модель математическая была из книжки приведена э-э, из раздела массового обслуживания, теория массового обслуживания. Один к одному просто там формулы были переписаны. А в этой книжке, там внизу странички, а мы эту книжку знаем, вот с Юрием Ивановичем Серебрянским, который недавно скончался, вот 12 октября. Он эту книжку писал, понимаете? Я по ней преподавал. Вот. Моделирование лекции ему читал, когда он меня просил. Вот я помню, что внизу этой страницы, там есть фраза такая, что обращаем ваше внимание, ну тех, кто читает эту книжку, что эта модель, она не обобщённый характер имеет. То есть она применима не для всех систем, а только для автоматизированных телефонных станций, где она была разработана. То есть эта модель массового обслуживания для телефонных станций. И написано, что она телефонные станции описывает хорошо. Но, допустим, для супермаркетов она вообще непригодна. Или для, допустим, ну, ядерных процессоров, многопроцессорных систем, например. Она непригодна для них. Там это прямо написано явно, понимаете, чёрным по белому в книжке этой. Вот. А она как раз моделировала э-э, супермаркет, систему массового обслуживания. А там внизу написано, что она неприменима для этого, понимаете? Вот эта девушка, значит, эту модель показала, там все это всё покивали, покивали, типа всё нормально. Валерий Иванович говорит: "Слушай, что она несёт? Ты посмотри". Я говорю: "Да вообще, кошмар какой-то". Это же вообще, вот на этом месте можно защиту прекращать, понимаете? Вот, потому что там, ну, неверная модель применяется. Дальше там всё на основе модели все выводы делаются. Вот. Ну мы сидим все там тихонечко так хихикаем. Делаем такой вид, что у меня уже рот не закрывается от удивления. Вот. И тут, значит, вдруг она потом говорит: "Вот мы разработали такую модель". А потом слушайте внимательно. "И выработали вот такие рекомендации. И какая-то фирма там в Адыгее, она из Адыгеи была, вот, эта фирма применила эти рекомендации". Мы всё, ё-моё, сидим вообще вот так вот уже э-э, за голову хватаемся, думаем. Так она, во-первых, модель применила, которая некорректна для этих целей применять. А потом, не проверив её достоверность, она и выработала рекомендации, эта фирма несчастная, эти рекомендации применила. Представляете себе? И что с этой фирмой стало? Можете себе вообразить? Ну, единственное спасает, знаете что? Я вам сейчас скажу, что спасло эту фирму. Знаете? Угадаете, что её спасло? Давайте, спрашивайте, что спасло эту фирму? А я вам скажу, что спасло. Ну что, я вроде Ирине Игоревне когда не общаюсь. Ну её спасло то, что на самом деле никто там ничего не вырабатывал, никто ничего не применял. Никто ничему не сообщал там. Вот только это её спасло. Вот если бы они реально действительно выработали рекомендации, но они было это сложно сделать, уже эта модель не была ещё автоматизирована, как то, что будем делать. Чтобы численно посчитать. Вот если бы они выработали и применили бы эти рекомендации, то этой фирме хана пришла бы, понимаете? И тут, значит, Валерий Иванович говорит: "Вы знаете, у меня вопросы есть". Ну это когда уже обсуждение. Вот как вы вообще осмелились применять модель для выработки рекомендаций, э-э, когда вы ещё не знаете, достоверна эта модель или нет? Вы представляете себе? Это очень рискованно, это вообще авантюра, понимаете, то, что мы с вами говорим. Вот когда революцию совершили октябрьскую в семнадцатом году, кто-нибудь проверял на маленьких странах, там, на кроликах, на мышах, проверял, работает эта модель или нет, Маркса, Энгельса, Ленина, или не работает? Взяли бы, проверили на какой-нибудь такой вот миллиметровой стране, которую на карте не видно. А потом бы уже взяли бы там, масштабировали, э-э, для страны, которая шестая часть света. Понимаете? Хотя бы так. Вот. Хотя то, что в маленькой стране получается, в большой может не получиться. Потому что там, э-э, в Монте-Карло, например, получается, а там во всей России сделать невозможно. Или во всём мире. В Люксембурге там высокий уровень жизни. Можно ли сделать в России высокий уровень жизни по тем рецептам, по которым живёт Люксембург? Нельзя. Почему? Очень большая разница, понимаете? Раз Люксембург - это как юбилейный район примерно Краснодара. Вот это тоже пошучу немножко, понимаете? Так вот, э-э, соответственно, отсюда вывод следует, что, значит, ну, Валерий Иванович говорит: "Вы же вообще-то, э-э, очень большой риск э-э, для этой фирмы создали, потому что ваши рекомендации вообще-то могли бы не не некорректными, неправильными быть". Потому что, ну, про то, что модель некорректна, он не стал говорить, потому что это вообще рушит всю диссертацию, просто перечеркнуть можно и выкинуть. Вот. А он говорит: "Хотя бы проверили достоверность модели, а потом уже вырабатывали решения на её основе. Если она достоверна, естественно". А так разве можно? Это он говорит: "Разве так можно делать?" Она говорит: "А мы, - говорит, - слушайте внимательно ответ этой соискательницы, который меня тоже шокировал очень". Ну и она уже когда-то, ну, уже всё нормально. Вот. Она говорит: "А вы знаете, вот эта фирма применила наши рекомендации, всё получилось хорошо". Что? Вы слушаете, что я рассказываю вообще, нет? Ребята. Всё получилось хорошо. Тогда Валерий Иванович уже говорит: "Ну значит, вам повезло, говорит. Вам просто очень повезло, понимаете? Благодарите Бога, что он вас спас". Хотя вы действовали очень э-э, опрометчиво, неразумно, рискованно. Но вот вам просто повезло, понимаете, у вас получилось хорошо. А могло получиться не очень хорошо, и вообще нехорошо, понимаете? Вот и всё. Весь вывод из этой ситуации. То есть это уже называется профанирование, профанация, понимаете, не наука, а профанация науки. И крайне авантюрные и рискованные действия. И всё из-за чего? А маленькая такая деталь. Вот этот ромбик. Ромбик, видите, вот этот ромбик? Выбор наиболее достоверной модели. Здесь надо нарисовать ещё выход вот сюда вот. И нарисовать: достоверной модели нет. Я это нарисую, сделаю. Достоверной модели нет вообще. И мы тогда вообще не можем эти задачи решать. Ну слово наиболее убрать. Ну или, наверное, надо её оставить. Вот. Ну, короче говоря, э-э, очень важный момент - исследовать достоверность модели. Вот к первому вопросу мы сейчас и переходим. То есть если достоверность модели не исследована и неизвестна, то её применение - это вообще профанация и авантюризм, и очень большой риск.

V. Обработка Многоклассовых Объектов и Специфика Системы

A. Многоклассовость Объектов vs. Одноклассовые Допущения

А если, значит, она высокая, достаточно высокая, тогда можно ей пользоваться более-менее спокойно, уверенно, если э-э, наша обучающая выборка, на основе которой она создана, репрезентативна по отношению к той генеральной совокупности, где мы собираемся применять модель. То есть вот мы создали модель, а это на основе исходных данных. А эти исходные данные, они хорошо отражают закономерности в той предметной области, где мы эту модель будем использовать. И относятся сейчас к этой фирме. И тогда всё совершенно корректно, можно использовать, потому что там действительно те же самые закономерности, которые мы выявили на основе обучающей выборки. А если мы взяли, применили эту модель, взяли, сделали там на данных этих Адыгеи эту модель, а потом взяли в Вашингтонщине там или или Аризонщине, или где-нибудь там применили. Вот, или в Калифорнийщине. Ну, в общем, вы поняли, да? Вот. И вообще там может эта модель ничего не отражает практически того, что там происходит. То есть она не соответствует условиям, допустим, Западной Америки, которые очень сильно отличаются от условий, так сказать, Адыгеи. И получается, что в этом случае э-э, применение этой модели там вообще неправомерно. Является просто чистой авантюрой. А когда наоборот, ребят, наоборот, разработали в США какую-то модель, мы берём её, вот там же в Нью-Джерси разработали, есть центр. Рядышком. А мы берём эту модель и применяем вот здесь у нас в Краснодаре. Это что, лучше что ли? Вот когда они наши модели применяют, мы смеёмся, это полный бред, понимаете, применять наши модели там. Мы это понимаем, осознаём. При этом мы берём их модели и применяем у себя. И не осознаём, что это вообще тоже такой же точно бред. Это нельзя делать. Нужно свои разрабатывать, отражающие нашу реальность.

C. Специфика Системы (Доступность)

Значит, я могу сказать, что многие это, конечно, понимают, но помалкивают. А некоторые даже и не помалкивают. Вот в Перми есть доцент Алексеев, Александр Олегович. Он э-э, исследовал несколько моделей американские, петербургские модели исследовал, э-э, описывающие состояние предприятий строительной отрасли в Перми. В Перми, в Перми. Вот. Оказалось, американская модель вообще э-э, на уровне случайного угадывания практически, недостоверно она работает. Э-э, петербургская модель чуть-чуть получше, но тоже очень плохо. А модель, которая сделана в системе Эйдос, работает великолепно, на уровне 90 там 2, 92-93% достоверности. Значит, как вы думаете, почему, ребята? Причина в чём заключается?

Наверное, в том, что использованы были американские данные изначальные, из-за этого?

Американские данные, они совершенно не отражают условия в Перми.

Причина не в том, что были взяты данные, которые именно соответствуют ситуации.

Ну конечно, они не являются репрезентативными. Они не являются репрезентативными. То есть, если взять генеральную совокупность, включающую нас, нашу Пермь, российскую, то вот эти данные американские не были нерепрезентативными. А петербургские? Петербургские лучше? Как вы думаете? Петербург и Пермь, они очень похожи, да? Как вы считаете? Да Петербург больше похож на э-э, Нью-Джерси, чем на Пермь. Ну понимаете? Он больше похож на западные города, европейские, чем на российские. Просто действительно объективно похож по своим показателям. Значит, петербургские немножко получше, но не намного, понимаете, не принципиально.

B. Подход Системы "Эйдос" (Логические Объекты)

И вот я сейчас хочу самое главное сказать. Э-э, модель, разработанная в системе Эйдос в Перми, она лучше не потому, что система Эйдос лучше, чем те системы, которые в Петербурге и в Америке используются. Там тоже люди соображают очень неплохо. И тоже они используют хорошие методы, корректные и сильные. И тоже у них есть программные системы. То есть свет клином не сошёлся на системе Эйдос в этом смысле. То есть есть и системы получше, чем система Эйдос. Я знаю это, уверен в этом абсолютно. Ну, может быть, не во всех отношениях получше, но в некоторых. У системы Эйдос есть одно очень важное достоинство. Знаете какое? Может быть, немножко смешно это звучит. Ну догадайтесь, какое у неё достоинство? Почему она имеет некоторое преимущество перед другими системами? Никогда не догадаетесь, я так думаю, что вы не можете догадаться. Разве что вот Ирина может догадаться. Или Сумрак, или Галина. Или Дима. Дима может, Виктория догадаться. Ну Артур точно догадался, просто Артур просто обязан догадаться.

В ней нет ничего лишнего?

Нет, она просто бесплатная. Ну да. Значит, просто можно взять её и воспользоваться, понимаете? Применить её для решения своих задач. А эти вот американские системы, самые дешёвые системы аналогичного функционального назначения, 150.000 долларов стоят. Вот возьми её и воспользуйся ей в Перми. Понимаете? Это нереально. Даже для Пермского национального университета. Это просто, ну, идиотизм. Откуда я эту цифру взял? Я показывал систему Эйдос э-э, в управлении президента, э-э, управлении делами президента Российской Федерации. Замначальнику управления делами, профессору Райкову. Ну он говорит: "Такое, что хотите сказать, что вы сделали систему э-э, лучше американских?" Говорит: "Да нет, я не хочу". Ну вот. "А за сколько вы хотите её продать?" Я говорю: "Да я вообще-то не по этому пришёл". Вот. Вот. Ну вот он тогда и говорит: "Самые дешёвые американские системы, аналогичные функциональной системе Эйдос, стоят 150.000 долларов. Самые дешёвые, простенькие". А получше системы, они стоят там и 300, и 400, и 800. Это Wall Street, и, допустим, система, она 800.000 долларов стоит. Просто система программная. Понимаете? Вот и всё. То есть э-э, доцент Алексеев, он подумал, думал он очень недолго, я вам скажу. Или заплатить 150.000 долларов за самую простенькую американскую систему интеллектуальную, решающую такие задачи. Или скачать с моего сайта систему Эйдос и попробовать. Он думал, знаете сколько? Одну сотую секунды. Артур, слышишь, да? Может, даже одну тысячную секунды он думал. Он вообще не думал. Он взял, скачал её и попробовал. И получилось достоверность гораздо выше, чем у этих систем там Wall Street, там и всё прочее. Понимаете? Почему? Потому что если бы взяли систему Wall Street и применили в Перми, то, конечно, бы получилось лучше, чем в системе Эйдос, я уверен. Но её там не применили, понимаете? Вот такая вот маленькая деталь. Её там не применили и не применят никогда. Вот и всё. Вот такое у неё достоинство есть. Ну, а второе достоинство, оно не настолько серьёзное, но тоже важное. Она работает реально. То есть она вполне разумно сделана, убедительно, аргументированно, модели там всё это вполне обосновано. И самое главное, что это всё работает действительно, понимаете, работоспособно. В тех областях, где есть эксперты, они говорят: "Да, так это ж мы и так знаем". Вот. А в тех областях, где их нет, мы становимся экспертами, проведя исследования и узнав какие-то зависимости. Мы оказываемся теми, кто лучше всего разбирается в этой предметной области. И разобрались мы в этой предметной области новой, которой ещё нет экспертов, благодаря системе Эйдос. Она выявляет зависимости и очень хорошо их э-э, в различных формах представляет. В том числе и в аналитической форме. Я вам это не показывал, но она и для регрессионного анализа готовит данные. Вот. То есть всё это там есть. Почему? Да возникала такая необходимость, и мы это, я это сделал. Просто это уже не уровень лабораторных работ, это уровень диссертации. Вот и всё.

VI. Заключение и Итоговые Замечания

A. Итог о Важности Проверки

Поэтому мы переходим к вопросу о том, как измеряется достоверность. И хочу ваше внимание вот на что обратить. Как только где-нибудь, когда-нибудь в будущем, в будущем, ребята, вы услышите слово модель, вот от кого угодно, от любого, вот где-нибудь, на какой-нибудь защите, там, дипломной, диссертации, где-нибудь услышат, где даже доклад какой-нибудь на конференции, услышат слово модель. Кто-то там что-то сидит, и там что-то в телефончике там что-то наговаривает. Вот услышит там модель, сказали, выступающий сказал слово модель. Вы сразу должны оторваться и подумать, вспомнить то, что я сейчас вам буду говорить. Э-э, интересно, вот он сейчас скажет про достоверность модели или нет? Вот этот докладчик. Или в статье вы прочитаете. Если нет, ребята, если нет, то вы можете спокойно абсолютно закрывать на этом месте эту статью, доклад. Вот. Ничего там дальше интересного не будет. Почему? Потому что человек писал недалёкий, понимаете? Не понимающий элементарных вещей. Причём человек, может быть, даже опасный, потому что если он что-нибудь порекомендует там где-нибудь на основе своей модели, достоверность которой он не проверил, то неизвестно, чем это кончится, понимаете? Может, его потом с собаками будут искать этого рекомендатора. Я серьёзно говорю, совершенно. Что есть люди такие скажут: "А вот он нам дал такую рекомендацию, и мы потерпели какой-то убыток. Ну давайте-ка найдём его и накажем его за это, понимаете?" Вот в чём дело. Это же не смешно, когда там предприятие может потерпеть убытки на многомиллионные, скажем так. А можно было дом построить, а можно было вот рекомендацию эту применить. Вот выбираете, или рекомендацию, или дом. Применили рекомендацию, всё, дом вы не построите. Потом нашли этого человека, который эту рекомендацию вам дал, и сказали, что вы про него думаете, понимаете? Так прямо конкретно. Это же серьёзно всё, это ж не игрушки. Поэтому вы должны, как только услышите слово модель, сразу вспоминать про слово достоверность. Сразу же. И должны поинтересоваться: а проверяли её достоверность или нет? Если нет, то о чём вообще можно говорить тогда? О каких, о каких решениях, каких задач? Да никаких. Вообще нельзя пользоваться этой системой, этой моделью. Это крайне опасно, и это авантюра и профанация науки, если кто-то построил модель и, не проверив её достоверность, начал применять. Как у нас в семнадцатом году. Применили, ну, получилось что-то такое, что сейчас спорят все с переменным успехом. Надо было, конечно, на кроликах попробовать или на миллиметровых странах проверить сначала. Которые не видно на карте. Было бы тогда более корректно. Вот эти миллиметровые страны, они потом тебе не услышали, они обиделись. А почему на нас? Вы тогда на сантиметровых странах проверяете, это более будет похоже вот на вашу страну. То есть лучше будет отражать условия вашей страны. Ну и ладно. Ну давай на километровых странах, которые полглобуса занимают. С обоих сторон видны страны, глобуса. И так оно и сделали. Ну так оно и получилось, что очень много уехало, очень много погибло, потом ещё очень много сами сами поубивали. Потом война. Ну, в общем, очень тяжёлая история. И я думаю, что, конечно, это был очень рискованный эксперимент, такой социальный, крупномасштабный. Но нужно было иметь очень большую смелость, чтобы пойти на такой эксперимент. Без э-э, проверенной модели, ребята. Она была хорошо обоснована, логично, убедительно Марксом. Ну и Энгельсом, и Лениным. Вроде бы всё очень хорошо обосновали, но потом оказалось, что там совсем всё понятно. Ленин вообще высказался таким образом, что вот построили не совсем то, а точнее, совсем не то, что собирались. То есть он признал, что получилось не то совсем. Публично. Потом предложил НЭП. НЭП - это была, конечно, гениальная совершенно идея. Вот то, что сейчас в Китае - это вот и есть НЭП, то, что Ленин предложил. Но у нас он длился недолго, до смерти Ленина. А потом пришёл Иосиф Виссарионович и всё стало как раньше. Вот. Ну и что там получилось. А если бы НЭП продолжался, то у нас было бы ещё круче, чем в Китае, и чуть не распался никогда Советский Союз. И сейчас бы мы были бы первой страной мира, наверное. Я так думаю. Потому что в Китае это же самое происходило через 50 лет, там, через 60. После НЭПа. То есть у нас был бы огромный задел во времени. И надо не надо забывать, что мы строили в Китае заводы, там, и технологии им давали, всё, они ещё и сами их брали иногда. И так оно со всего мира наоборот к нам. Ну, в общем, была бы другая история абсолютно, понимаете?

B. Риски и Профессиональная Ответственность

Вот. Значит, теперь давайте смотреть, как измеряется достоверность модели. Значит, для этого мы переходим в режим 3-4. Вот. И здесь вот у нас есть код. Вот в этом режиме, который сейчас я вам этот код в чат пошлю. Вот это не сам код, режим. А всякие достоверности модели. Вот эти вот формулы. Да, ребят, те, кто пользуется системой Эйдос, должны знать, что, значит, я изменил доступ к своему сайту системы, э-э, к сайту обновлений, к FTP-серверу, э-э, в целях безопасности. Вот. И поэтому обновить систему можно, используя режим 6.2. А в нём есть пункт шесть. Просто скачиваете оттуда обновление, этот патч, разворачиваете в папочке системы, когда система не запущена. И у вас система обновляется до текущей версии, которая 24 июня. То есть позавчерашняя. Понятно, да? Вот. И ещё сейчас одну важную информацию дам. Интересно. Хорошо. Смотрим на следующие моменты. Значит, вы поняли, что с этой вот мультиклассовостью нужно разбираться. Следующее. Смотрим на результаты идентификации. Система Эйдос представляет собой систему, которая определяет не просто принадлежность номер объекта к классу, а ещё определяет степень уверенности своей в этом решении. Вот, допустим, она видит и пишет, что мышка один, объект конкретный, относится к классу мышка. И она уверена в этом на 90 там 2,68%. Ещё она считает, что этот объект относится к классу элемент компьютера, и в этом она уверена на 72%. И фактически так и есть. Эти птички стоят там, где они фактически действительно, это так и есть. Ещё она думает, что это, может быть, относится к классу телефон. Но в этом она уверена только на 48%. А ещё она думает, что это может быть даже и монитор, но только на 28, на 29% она в этом уверена. Но она уверена в том, что это не вешалка, не сумка, не аксессуар, не мебель, не спорт-инвентарь. И вот так вот мы можем посмотреть на разные объекты. Вот мне нравится клавиатура два, смотрите. Клавиатура два, она относится к классу клавиатура на 95%, к классу элемент компьютера на 56%. А к классу аксессуары, сумки относятся на 4%. Это неверно, она к ним на самом деле фактически не относится. Птички здесь нет. Это ложно-положительное решение. Значит, это ошибка системы, ложная идентификация. Но ошибка небольшая. Почему? Потому что она сама система думает, что всего лишь на 4% этот объект относится к этому классу. Что значит относится на 4%, на 20%, на 90%? Что это вообще значит? А это значит, что этот объект похож именно настолько, похож на вот этот класс. Я думаю, что можно считать, что объект клавиатура 2 является взвешенной суперпозицией всех объектов, на которые он похож, с теми весами, которыми означают степень уверенности системы в этом решении. То есть если мы сложим аксессуар умножить на 4%, сумку умножить на 4%, средства связи умножить на 22% и так далее, всё это сложим, у нас получится клавиатура 2 в результате. Как разложение в ряды? Ребята, я послал вам ссылочку на статью новую свою, у которой будет иметь продолжение, или у неё даже есть продолжение. Сейчас это продолжение я вам покажу. Я её хочет принять к публикации. Вот. Вот такая статья. Скажем так, математизированная. То есть это нельзя сказать, что это там чисто математика, но она имеет самое непосредственное отношение к математике. Ну примерно как вот какое отношение математика имеет к физике, к экономике, к другим областям науки. Ну, прямое отношение имеет. Это язык, который применяется во многих областях науки. Причём, э-э, очень характерно, что когда наука достигает определённого уровня развития, тогда начинает больше пользоваться математикой, чем до этого. То есть можно сказать так, что можно в какой-то степени оценивать уровень развития науки использованием, степенью использования математики. Если эта степень высокая, тогда наука, значит, очень э-э, неплохо развита. Если слабая, использование математики, наука слабо развита. Значит, я вот доктор наук по специальности 080013, математические инструментальные методы экономики. У меня с Валерием Ивановичем мы сидим и смотрим на диссертацию по специальности 08 2005, управление народным хозяйством. Когда-то оно было народным хозяйством. Ну можно сказать, что оно, конечно, оно народное, но так иносказательно. Так вот, ребята, э-э, раньше там буквально на степени было всё. Так вот, э-э, смотрим, работа такая слабоватенькая. Слабоватенькая. Валерий Иванович говорит: "Да, эту работу правильно, что по пятой специальности её защищают. На тринадцатую она не тянет, она слишком слабая для для тринадцатой". А другую работу мы смотрим, тоже защита, а там тоже она по 05, а там хорошо так всё обосновано, математика развита, причём корректно применяется. Э-э, всё, значит, э-э, убедительно очень выглядит и понятно, что это правильно. Ну вот эта работа настолько хорошая, это вот это слова Валерия Ивановича, э-э, настолько хорошая эта работа, хотя она и по 05, но в принципе, она даже потянула бы даже на тринадцатую специальность, даже вот на 08 00 13 тоже, в принципе, даже она бы и по этой специальности можно было её защитить. Поняли, да, ребята, о чём речь? А чем отличается тринадцатая от 05? Знаете чем? Можно ли взять 05 и защитить по тринадцатой? Нет. Там нужна математика, чтобы была, понимаете, корректно применённая, развитая, мощная математика. А можно ли тринадцатую по 05? Можно, потому что это то же самое, когда вам дают, допустим, мастера спорта, говорите: "Да не надо мне мастера спорта, мне давайте первый разряд дадите, мне и хватит". Мне это устроит. То есть тринадцатая специальность - это как бы более сильная специальность, чем 05, понимаете? Более серьёзная. У нас на ней три доктора наук по специальности 08 00 13. Три. А по 05 сколько? Ну десятки. Поняли, да, ребята, о чём речь? По двенадцатой специальности присваиваются в технических науках, физмат науках или экономических науках. Мне как бы совет экономических наук. Вот. Ну вот и все дела. Вот так вот. Так вот, э-э, ребята, значит, я вот в этой статье, которую сейчас ссылочку, на которую я вам дал, я предложил считать, что распознавание - это разложение в ряд. А вопрос возникает, разложение в ряд чего? И в ряд по чему, по каким функциям? Я предложил считать, что функция, описывающая объект, вектор объекта, разлагается в ряд по векторам классов, а в качестве весовых коэффициентов разложения в ряд используются вот эти вот степени уверенности. Значит, это всё подробно описано в этих статьях. В ResearchGate описано с численным примером, а в журнале только теория описана, а численный пример в отдельной статье будет, которая сейчас вот размещаемом номере будет. Это очень глубокое, очень интересное, э-э, такое интересное представление о том, что такое распознавание. И вот мы здесь видим, смотрите, ребята, что, видите, э-э, система определила, в соответствии с интегральным критерием, который я потом вам расскажу попозже, когда будем прямые задачи рассматривать. Сейчас мы рассматриваем достоверность, а потом, потом будем подробно рассматривать задачи. Там я расскажу про эти интегральные критерии, про эти и так далее, и так далее. Так вот, здесь мы видим что, что система, она ошиблась, отнесла сумку к аксессуарам. Но она ошиблась немного, совсем чуть-чуть, всего 4% уверенность самой системы в том, что это так. То есть она относит, но очень сильно сомневается в этом. Но всё-таки она относит к этому, к этой категории, а не к категории стул. Стул вообще не относит, видите, с минусом она, да? То есть уровень сходства. А здесь она относит, но с очень низким уровнем сходства. Что делает Ван Рисберген? Он бабах и к сумматору True Positive, False Positive, то есть ложно-положительное решение, суммирует, что бы вы думали? Он единицу суммирует к сумматору. Он подсчитывает сами решения, ложные, истинные, положительные, отрицательные. Сколько таких решений? Я говорю: "Ну ты, Ваня, э-э, Рисберген, прямо, ну крутой ты какой-то, совершенно прямо человек, очень строгий, понимаете?" Вот, несправедливый, я считаю. Зачем ты так наказываешь систему и модель, суммируя единицу к сумматору False Positive? Когда обоснованно, оправданно суммировать 4 сотых, ребята, не единицу, а 4 сотых. Потому что 4% - это это 4 сотых. Она же сама говорю, не уверена в своём решении. Она уверена только в 4 сотых. Ну давайте и суммировать к сумматору False Positive 4 сотых, вот эту самую уверенность. И тогда мы увидим, что система, да, она ошибается, но она ошибается чаще на э-э, решениях с очень низким уровнем сходства. И мы это увидим сейчас, ребята. Вот смотрите. Берём мы строим график частотных, частотное распределение решений, истинных, числа, числа истинных и ложных решений, положительных и отрицательных. Истинные решения красным обозначены, ложные решения синим. Справа от оси Y - положительные решения, слева от оси Y - отрицательные решения. Вот замечательный чат в системе Teams. Смотрим. Решения отрицательные, о непринадлежности объекта к классу. Они, а здесь ложных и нету, здесь все истинные. Все решения истинные. А решения о принадлежности есть и истинные, и ложные. Ну здесь вот сами эти решения, а здесь сглаженная кривая, чтобы было видно смысл этих вот э-э, самой исходной кривой. Когда данных побольше, э-э, тестовой выборки, тогда здесь таких изломов нет, она тогда очень похожа на нормальное распределение, два нормальных распределения. А здесь очень маленькая выборка, всего 20 примеров, поэтому вот так выглядит. Значит, здесь смотрите, ребята, значит, ложные решения есть? Да, есть, но они есть при маленьких уровнях сходства. Так, ребят, на этом мы занятие заканчиваем. До следующего занятия. До свидания. До свидания.