



## **Выявление мошенничества на рынке финансовых услуг**

Методы выявления мошенничества  
с использованием программного  
обеспечения SAS® Enterprise Miner™.

## Содержание

1	Мошенничество в финансовой сфере.....	3
1.1	Примеры мошенничества в отдельных секторах .....	3
1.2	Сложности выявления мошенничества .....	4
2	Традиционные методы обнаружения мошенничества .....	5
2.1	Простые программные подходы .....	5
2.2	Обучение и поддержка .....	5
2.3	Дополнительные требования по подготовке отчетов .....	6
3	Обнаружение мошенничества при помощи средств добычи данных .....	7
3.1	Обнаружение необычных данных .....	7
3.2	Выявление необычных взаимозависимостей в данных .....	8
3.3	Выявление изменений поведенческих характеристик .....	9
3.4	Соответствие характеристик мошенничества методикам добычи данных .....	9
3.5	Прогнозирование мошенничества на основании общих характеристик.....	11
4	Пример из практики1: Выявление мошенничества с закладными .....	11
4.1	Описание входных данных .....	11
4.2	Пример данных .....	12
4.3	Выявление мошенничества .....	13
4.3.1	Исследовательский анализ .....	13
4.3.2	Прогностическое моделирование .....	14
4.3.3	Анализ ассоциаций и последовательностей.....	17
4.3.4	Анализ ссылок .....	18
4.4	Выводы .....	19
5	Пример из практики 2: Выявление мошенничества с кредитными картами .....	19
5.1	Описание входных данных .....	20
5.2	Анализ .....	21
5.3	Заключение .....	23

# 1. Мошенничество в финансовой сфере

От мошенничества и подделок страдают все области финансового сектора. Они происходят в банковской, страховой, инвестиционной сферах, в брокерской деятельности, бывают мошенничества с ценными бумагами и при операциях на товарных биржах. Общий ущерб от этого вида незаконной деятельности огромен. Так, например, считается, что от 10 до 20% всех требований на выплату страховок фабрикуются мошенническим путем. В результате страховая отрасль теряет ежегодно примерно 20 млрд. долл. только в Соединенных Штатах (New York Central Mutual 1999). Можно предположить, что подобные выплаты выливаются примерно в такую же сумму и в Европе. Потери, вызванные махинациями с кредитными карточками, оцениваются в 1 млрд. долл. за год (Cassidy 1997), а «отмывание» денег «стоит» государствам 500 млрд. долл. ежегодно (Steele 1999). Вот несколько наглядных примеров, характеризующих масштаб проблемы.

В Нью-Йорке сотрудники четырех медицинских учреждений обманном путем получали деньги со страховых компаний. В результате расследования, проведенного прокуратурой района Бруклин, были арестованы четыре человека, вовлеченные в широкомасштабную систему систематического обмана страховых компаний. Предъявленное мошенникам обвинение состояло из 140 пунктов. Ответчикам вменялось в вину уничтожение оригинальных историй болезни и их замена фальшивыми, а также выставление страховым компаниям счетов за обследования, которые никогда не проводились. Общая сумма выплат по таким подложным заявкам составила 100 тыс. долл. Страховые компании, ставшие жертвами махинации, могли бы понести убытки на сумму в 5 млн. долл. (Farrell 1999).

Государственный банк штата Огайо стал жертвой сложной махинации, которая стоила этому учреждению 15 млн. долл. за три года. Руководил аферой бывший член руководства банка, который остался его клиентом и произвел незаконный перевод миллионов долларов из банка в строительную компанию под видом кредитов. Анализ пакета кредитов дал основание подозревать возможность мошенничества. Банк предъявил мошенникам обвинение в проведении фиктивных займов, которые шли на приобретение мошенниками недвижимости. Кроме того, они предоставляли заведомо ложную информацию о клиентах строительных компаний в совет директоров банка (Provost 1999).

## 1.1. Примеры мошенничества в отдельных секторах

Мошенничество в финансовой отрасли не ограничивается случаями подлога медицинских страховок или отмывания денег. Можно перечислить некоторые наиболее распространенные виды мошенничества в отдельных секторах:

- |                                   |  |
|-----------------------------------|--|
| <b>Банковская сфера:</b>          | <ul style="list-style-type: none"> <li>• Кража кредитных карт или информации о кредитных картах и осуществление покупок по этим картам;</li> <li>• Отмывание денег, полученных в результате нелегальной деятельности, путем прямого размещения наличных средств на счетах местных и иностранных банков или путем инвестирования их в реальный или фиктивный бизнес;</li> </ul> |
| <b>Страхование собственности:</b> | <ul style="list-style-type: none"> <li>• Инсценирование грабежа</li> <li>• Подача фиктивных заявок на возмещение убытков от якобы имевшего места воровства или потери имущества</li> </ul>   |
| <b>Страхование автомобилей:</b>   | <ul style="list-style-type: none"> <li>• Подача фиктивных заявок на возмещение убытков по поводу угона транспортного средства</li> <li>• Установка использованных запасных частей при ремонте и подача заявок на возмещение убытков по стоимости новых запчастей</li> </ul>  |

- Намеренное создание аварийных ситуаций, предполагающих массовую подачу заявок на возмещение ущерба пассажирам и порчу собственности
  - подача заявок на возмещение ущерба в случае аварии лицами, чьи автомобили не участвовали в данном инциденте
- Здравоохранение:**
- Проведение обследований или получение медицинских услуг без необходимости
  - Прописывание дорогостоящих всесторонних обследований
  - Выставление счетов за услуги, которые не были оказаны
  - Выставление счетов на лечение, проведенное старшим медицинским персоналом, но выполненное реально вспомогательным персоналом
  - Выставление счетов по высоким индивидуальным расценкам, путем разбиения на отдельные операции стандартных комплексных обследований
  - Выставление счетов за подобные, но более сложные и дорогие процедуры, при фактическом проведении более простых и дешевых
- Ипотека и закладные:**
- Фальсификация оценки стоимости недвижимости
  - Брокерская деятельность:
  - Продажа или приобретение ценных бумаг с учетом информации, полученной незаконным путем (инсайдерская деятельность);
  - Попытки «подставлять» клиентов путем предоставления им заведомо ложной информации, такой как гарантированные проценты возврата инвестиций в ценные бумаги, или фальсификация наименований предлагаемых к приобретению инструментов ("первичные банковские сертификаты"), что должно послужить для клиента гарантией надежности этих инструментов для вложения средств;
  - Искусственное повышение цен такими методами, как купля-продажа ценных бумаг между брокерами.

## 1.2. Сложности выявления мошенничества

В разделе "Рекомендуемая литература" приведены ссылки на дополнительные источники информации по таким вопросам, как мошенничество, добыча данных и хранилища данных.

Наиболее трудны для обнаружения случаи мошенничества в страховании, поскольку они относятся к так называемому "невывяляемому" типу. Иными словами, в случае мошенничества со страховками отсутствуют явные свидетельства подделки, такие как заявление о краже кредитной карты, которые бы позволили, в конечном итоге, определить, что требование на выплату страховой суммы основывается на сфальсифицированных документах. В этом случае для обнаружения мошенничества нужно выявлять подозрительные закономерности при обращении в страховые компании или связи между лицами, предоставляющими ложную информацию о себе. Кроме того, многие компании сейчас обнаруживают, что у них имеется очень много данных о своих клиентах, но при этом очень мало полезной информации. Традиционные методы обнаружения мошенничества несут в себе слишком много проблем и имеют серьезные ограничения.

В этой статье рассматриваются традиционные методы обнаружения мошенничества на рынке финансовых услуг и ограничения этих методов. Также показано, как

технология добыча данных (data mining) может помочь преодолеть эти ограничения за счет использования более сложного и статистически достоверного анализа, позволяющего выявить мошенничество. В заключение рассматриваются два примера успешного обнаружения мошенничества - махинации с закладными и с кредитными картами. Эти примеры иллюстрируют возможности применения технологии добычи данных при обнаружении мошенничества с помощью программного обеспечения SAS® Enterprise Miner™.

## **2. Традиционные методы обнаружения мошенничества**

Общепринятые методы обнаружения и предотвращения мошенничества основаны на проведении индивидуальных расследований с возможным применением компьютерных технологий, а также на обучении и поддержке клиентов.

### **2.1. Простые программные подходы**

Компьютерные технологии могут облегчить обнаружение мошенничества, используя такие простые программные методы, как подготовка отчетов об исключительных ситуациях. В таких отчетах события, удовлетворяющие тем или иным заранее определенным критериям, получают специальную пометку. Так, например, в отчете об исключительных ситуациях при страховании здоровья могут быть помечены все операции по удалению миндалин, стоимость которых превышает определенный, заранее установленный уровень. Такие системы используются со вполне очевидной целью - избежать крупных расходов, обращая внимание на случаи, по которым могут быть взысканы наиболее крупные суммы. Несовершенство этого метода состоит в том, что мошенники могут узнать используемые пороговые значения и не превышать их в своих сфальсифицированных документах. При этом мошенничество так и не будет раскрыто. Более сложный мониторинг может предполагать использование дополнительных пороговых значений, таких как ставки страхования, а также другие показатели, специально разработанные для выявления мошеннической деятельности.

### **2.2. Обучение и поддержка**

Обучение и поддержка покупателей и клиентов - это еще один важный компонент традиционного подхода к выявлению мошенничества. Например, можно существенно повысить надежность страхования в области здравоохранения, если направлять получателю денег подробный отчет от медицинского учреждения, предоставляющего медицинские услуги. Это оказывается достаточно эффективным в случаях, когда мошенники используют чужие номера счетов или украденные карточки медицинского страхования. Недостатки этого метода связаны с тем, что отчеты о медицинских услугах могут содержать непреднамеренные ошибки, кроме того, в сфальсифицированных заявках могут указываться услуги, не внесенные в отчет, к тому же и сам получатель денег может не понимать или не помнить о том, какие конкретно медицинские услуги ему оказывались, или целенаправленно скрывать эту информацию.

Для успешного обнаружения мошенничества нужны оба вышеприведенных подхода; тогда их сильные стороны окажутся еще более эффективными, а недостатки будут нивелироваться. Несмотря на то, что отчет об исключительных ситуациях не несет в себе знания о том, что действительно произошло, он выдает последовательную и непредвзятую информацию. С другой стороны, именно последовательность и непредвзятость анализа прежде всего подвергается сомнению, если он проводится людьми. Правда, в этом случае можно полагаться на пациентов, которые нередко вспоминают, что происходило в действительности, и предоставляют необходимую дополнительную информацию. Недостаток обоих подходов состоит в том, что существует множество случаев, которые с трудом поддаются автоматизации, и большая часть аналитической работы в действительности выполняется людьми, а это занимает нередко месяцы, а то и годы. Улучшенные средства компьютерного мониторинга ускоряют процесс расследования подозрительных заявок, причем не просто указывают на показатели, не соответствующие норме, но идентифицируют случаи мошенничества, а также предоставляют надежный прогноз на будущее.

### **2.3. Дополнительные требования по подготовке отчетов**

Традиционные подходы к обнаружению и предотвращению мошенничества можно усовершенствовать, если ввести подготовку отчетов в число стандартных бизнес-процедур.

Строгое соблюдение требований по подготовке отчетности для внутреннего персонала, а также для внешней аудитории, например, правительственных агентств, способно во многих случаях остановить мошенников, а также упростить выявление мошенничества. Хорошим примером в данном случае могут послужить требования к отчетности, оформленные недавно в США в законодательном порядке.

- Для предотвращения «отмывания» денег закон о банковской тайне предписывает соответствующим банкам, а также другим финансовым организациям подавать в Министерство финансов Отчет о валютных транзакциях (Currency Transaction Report, CTR), если объем бизнес-транзакции в произвольный банковский день превысил сумму в 10 тыс. долл. В этом отчете содержится подробная информация о личности клиента и форме оплаты, а также о финансовых учреждениях, вовлеченных в данную транзакцию. Кроме того, банкам вменяется в обязанность подавать отчет о переводе за рубеж валюты или других кредитно-денежных средств (Report of International Transportation of Currency or Monetary Instruments, CMIR), содержащий развернутую аналогичную информацию обо всех транзакциях с иностранными счетами (General Accounting Office 1994).
- Для частичного возврата ежегодных потерь в сумме 23 млрд. долл. от выплат по сфальсифицированным или ошибочным заявкам по медицинскому страхованию, Управление финансирования здравоохранения (Health Care Financing Administration, HCFA) и Американская медицинская ассоциация (American Medical Association, AMA) издали правила, согласно которым практикующие врачи обязаны предоставлять подробные истории болезни пациентов, сведения о методах обследования и процедурах, о принятых решениях, проведенных консультациях, согласовании методов лечения, диагностике. Вся предоставляемая информация должна сопровождаться указанием дат и сроков проведения тех или иных операций, и их продолжительности (Elliott 1998).
- Для пресечения воровства в страховании, Бюро страховых услуг (Insurance Services Office, Inc., ISO) объединяет данные по страхованию транспортных средств, находящиеся в распоряжении Национального бюро страховых расследований (National Insurance Crime Bureau, NICB), с базой данных заявок на выплату страховых, которую поддерживает Группа страхового обслуживания США (American Insurance Services Group, AISG) с целью формирования единой базы данных, содержащей сведения обо всех телесных повреждениях, порче собственности, компенсации служащим и заявкам, связанным с транспортными средствами (ISO 1999).

Требования к подаче отчетов обычно не могут предотвратить мошенничество; «отмывание» денег по-прежнему остается столь же распространенным видом преступной деятельности, каким оно было до выхода нового закона. Однако это может послужить для преступников предупреждением, уменьшить число случаев мошенничества, а также дать полезную информацию для правоохранительных органов об уже совершившихся преступлениях. Помимо непосредственного выявления преступников, данные отчетов могут дать основание подозревать мошенничество. В случае с «отмыванием» денег такими индикаторами могут стать сведения о профиле компаний, вовлеченных в подозрительную транзакцию, их местоположение, частота и время совершения операций с валютой. Дополнение традиционных методов определения мошенничества требованием обязательной отчетности повышает надежность определения случаев мошенничества, так как данные оказываются уже собранными и подготовленными для расследования. Однако расследования все еще остаются в недостаточной степени автоматизированными и требуют проведения анализа записей непосредственно человеком.

### **3. Обнаружение мошенничества при помощи средств добычи данных**

Методы, предполагающие применение средств добычи данных, разительно отличаются от традиционных подходов к обнаружению мошенничества тем, что выходят далеко за рамки простых отчетов об исключительных ситуациях. Эти средства выявляют подозрительные случаи на основе шаблонов данных, позволяющих сделать предположение о мошенничестве. Шаблоны данных, указывающие на возможность мошенничества, могут обладать одной или несколькими следующими характеристиками:

- Необычные величины данных, каким-либо образом отличающиеся от нормы.
- Необычные взаимосвязи между величинами данных или записей.
- Изменения в поведении сторон, участвующих в транзакции.

Эти шаблоны данных и их выявление при помощи методов добычи данных более подробно описаны в последующих разделах.

### 3.1. Обнаружение необычных данных

Необычно высокое число заявок на выплату компенсаций в случае аварии, сравнительно высокие цены при распродажах, или необычные сочетания диагностических или терапевтических вмешательств представляют собой необычные данные. Есть три наиболее часто встречающихся индикаторов необычных данных.

- Необычное значение, встретившееся только однажды.
- Значение, оказывающееся необычным, при сопоставлении его со значениями в группе сравнения.
- Необычные сочетания значений, которые сами по себе являются приемлемыми.

Данные, содержащие необычное значение, которое встречается только однажды, проще всего выявлять путем анализа выбросов. При этом методе выбросы определяются как необычные, но приемлемые значения. Применение количественных статистических инструментов, таких как определение среднего значения или стандартного отклонения, а также представление данных в виде различных графиков и диаграмм, может оказаться эффективным для выявления необычных значений для непрерывно изменяющихся переменных. Аналогичные измерения, проведенные для частоты появлений, могут стать хорошим индикатором для категориальных переменных.

Выявление значений, которые оказываются необычными при сопоставлении с референтной группой, представляет собой более трудную задачу. Так, например, когда объектом анализа являются цены на недвижимость, референтная группа имеет принципиальное значение для выявления мошенничества. Цена определенного объекта может и не быть слишком большой, если речь идет, к примеру, о домах вообще, однако, она может оказаться чрезмерной, если рассматривать именно подобные владения одной площади и типа, в одном регионе, в одной определенной экономической ситуации. Убедиться в необычности того или иного значения можно, применяя методы исчерпывающего анализа данных к информации, собранной торговыми агентами на данной небольшой территории или к ценам на аналогичную недвижимость в том же рыночном секторе. Методы выявления необычных значений в данной ситуации могут включать в себя кластерный анализ, позволяющий установить референтные значения для каждого сегмента рынка, с последующим анализом выбросов.

Обнаружение необычных сочетаний значений может оказаться трудной задачей, которая в дополнение к выполнению операций добычи данных потребует серьезного программирования. Как правило, алгоритмы добычи данных значительно лучше определяют наиболее распространенные сочетания, ассоциации или последовательности. Так, например, если применить их для решения задачи выявления потенциально фальшивых медицинских счетов, в большинстве которых указывается необычное сочетание медицинских назначений, то будет достаточно сложно автоматически определить сочетания, наиболее характерные для фальсификации. Этот метод предполагает наличие исходных медицинских знаний, то есть знания того, какие сочетания назначений можно считать нормой. Без этих знаний работа алгоритмов ассоциации и формирования последовательностей будет неэффективной.

### 3.2. Выявление необычных взаимосвязей в данных

На наличие необычных зависимостей в данных могут указывать две или более несвязанных записи, имеющие одинаковые значения некоторых переменных. Это может быть, например, непропорционально большое число жертв дорожно-транспортного происшествия, лечащихся у одного врача. С другой стороны, необычными иногда можно считать соотношения, которые связывают анализируемые записи, например, названия компаний, с разными именами, но одинаковыми адресами, или сделки с различными земельными участками, но с одними и теми же покупателями, продавцами и агентами.

Средства добычи данных позволяют автоматизировать и более эффективно использовать дополнительные данные, полученные из подробных отчетов, для выявления и прогнозирования мошенничества путем применения сложного и статистически достоверного анализа.

Первый тип взаимосвязей наиболее прост для выявления. Здесь требуется, разумеется, чтобы общность анализируемых полей была действительно необычной и существенной. Распространенные атрибуты, такие как пол или национальность в этом случае использовать не имеет смысла. Хорошим примером необычного взаимоотношения в данных может служить ситуация, когда две или более компаний вовлечены, например, в операции перевода фондов и при этом имеют разные названия, но один и тот же почтовый адрес. В данном случае весьма логично заподозрить «отмывание» денег.

Затем для каждого счета можно при помощи своего рода набора правил дать заключение с какой вероятностью он может оказаться поддельным. Этот процесс называется scoring.

Учитывая, что данные транзакции включают в себя сотни переменных, и что число транзакций может быть весьма велико, эти необычные сочетания вполне могут остаться незамеченными, если не проводить специального анализа. В принципе, можно провести идентификацию только путем определения частоты появления значений определенных переменных. На практике, однако, осуществить это довольно трудно, поскольку объем данных очень велик и переменные могут иметь множество значений, например, номеров счета и адресов. Автоматические инструментальные средства оценки частоты появления значений по многим переменным и сочетаниям переменных являются обязательным элементом решений добычи данных. Наличие необычных взаимоотношений в данных не обязательно указывает на мошенничество, но служит сигналом к проведению дальнейшего расследования.

Другой подход к выявлению необычных взаимоотношений в данных состоит в обнаружении так называемых «почти дублирующихся записей», то есть таких записей, переменные в которых – при наличии возможных незначительных различий – содержат в основном идентичную информацию. В этом случае эффективно применять кластерный анализ с большим числом кластеров анализируемых переменных. Эти кластеры рассматриваются на наличие подозрительных групп и исследуются более пристально.

Если в связи записей просматривается некая логика, например, как в инсценированных дорожно-транспортных происшествиях, методы выявления мошенничества требуют четкой идентификации взаимосвязи между записями. В простейшем случае эта взаимосвязь может выражаться в конкретной записи, например, как в случае, когда жертвы ДТП обращались за помощью к одному и тому же врачу, который является неотъемлемым элементом сценария. Эта связь может быть вскрыта при помощи анализа выбросов, нацеленного на переменную TREATED BY для всех случаев инцидентов. При этом необходимо отбросить все записи, в которых врачи, перечисленные в TREATED BY, лечили только одного или нескольких пострадавших. Доктора, оставшиеся в списке, подлежат более углубленному анализу. Это итеративный процесс, который может дать и ряд ошибочных предположений, особенно если применять его к слишком большому объему данных. В то же время, инсценированные ДТП, как правило, происходят на одной территории, что сокращает размер базы данных и сложность задач, и, следовательно, облегчает решение.

Более сложен для анализа случай, когда связь между записями проявляется в виде многочисленных промежуточных записей. Это отражает ситуацию, когда группа людей последовательно перепродает дома друг другу по все более высокой цене. Эта операция может служить подготовительным этапом для получения фиктивной закладной или страховки недвижимости. Выявление такого набора записей требует навигации от продавца к покупателю собственности, и, если покупатель вскоре вновь продает ее, от этого покупателя к покупателю новому и так далее, пока не будет установлена цикличность транзакций. Поскольку в большинстве случаев законопослушный покупатель собственности планирует впоследствии продать то, что у него имеется сейчас, или уже продал свою собственность, для выявления настоящего мошенничества требуется интенсивный процесс поиска. Этот поиск можно конечно и запрограммировать, например, при помощи языка Structured Query Language (SQL) для формирования запроса к базе данных; алгоритмы, эффективно обрабатывающие такие запросы, в добыче данных называются «анализом ссылок».

Некое лицо подает необычно большое число требований на возмещение ущерба в случае дорожно-транспортного происшествия. Цена недвижимости слишком высока по сравнению с ценами на аналогичные владения в данной местности. Врач делает назначения на необычные сочетания диагностических и терапевтических процедур

### 3.3. Выявление изменений поведенческих характеристик

Данные, не соответствующие устоявшейся норме, представляют собой изменения поведенческих характеристик, которые могут свидетельствовать о мошенничестве. Среди примеров таких изменений поведения: интенсивные покупки, совершаемые в короткий промежуток времени, крупные закупки по кредитным картам, многочисленные звонки в необычные для данного клиента регионы или повторяющиеся счета за нестандартные медицинские процедуры. В случае если изменение поведения проявляется в виде конкретного события, например, крупной покупки по кредитной карте, подозрительное действие легко проверяется путем анализа выбросов. В случае, если от-

клонение поведения проявляется в виде целой последовательности событий, как в случае со звонками в необычные регионы, для того, чтобы определить, имеет ли место мошенничество, требуется установить частоту повторяемости звонков.

### 3.4. Соответствие характеристик мошенничества методикам добычи данных

Таблица 1 позволяет установить соответствие характеристик данных, анализируемых с целью выявления мошенничества, с соответствующими методиками добычи данных.

Характеристика	Цель	Примерный сценарий мошеннической операции	Метод добычи данных
Необычные данные	Выявить отдельные необычные, но в целом приемлемые значения Выявить значения, необычные по сравнению с референтной группой Выявить необычные сочетания значений, по отдельности не выходящих за рамки нормы	Некое лицо подает необычно большое число требований на возмещение ущерба в случае дорожно-транспортного происшествия Цена недвижимости слишком высока по сравнению с ценами на аналогичные владения в данной местности Врач делает назначения на необычные сочетания диагностических и терапевтических процедур	Анализ выбросов; частотный анализ Кластерный анализ; анализ выбросов Различные алгоритмы
Необычные взаимоотношения	Выявить связи между не связанными между собой записями Выявить практически идентичные записи Выявить прямые связи между взаимосвязанными записями Выявить связи между записями, выражающиеся посредством многочисленных пересекающихся записей	Две или более компании с различными названиями, участвующие в передаче фондов, имеют один и тот же почтовый адрес. В многочисленных сделках с недвижимостью участвуют одни и те же лица в качестве продавцов, покупателей и агентов Жертвы инсценированных дорожно-транспортных происшествий обращаются за лечением к одному и тому же врачу Транзакции, в которых участвуют многочисленные компании, с переводом денег через различные финансовые институты	Частотный анализ Кластерный анализ Анализ выбросов Анализ ссылок
Изменения поведенческих характеристик	Выявить единичный случай необычного поведения на базе исторической информации Выявить многочисленные случаи необычного поведения	Со счета кредитной карты частного лица снимаются крупные суммы в оплату покупок Врач постоянно выставляет счета за нестандартные процедуры	Анализ выбросов Частотный анализ

Таблица 1. Соответствие методик добычи данных выявлению определенных типов мошенничества

### 3.5. Прогнозирование мошенничества на основании общих характеристик

После идентификации конкретных случаев мошенничества полученные характеристики, свойственные целому ряду таких случаев, можно использовать для выявления и предотвращения потенциально мошеннических транзакций. В числе таких транзакций могут быть как уже совершившиеся, так и те, что еще только могут осуществиться.

В обоих случаях для идентификации специфических характеристик данных, которые дают основание подозревать мошенничество, применяется прогностическая добыча данных. Для применения этой методики требуется достаточно объемный набор транзакций, которые можно обобщать для выполнения прогнозов путем применения регрессионного анализа, деревьев решений или нейронных сетей. В зависимости от используемого типа анализа, прогнозы могут быть в большей степени как численными (регрессия, нейронные сети), так и текстовыми, подобными письменному естественному языку, и включать в себя информацию о наиболее общих бизнес-методах.

По мере появления работающих прогнозов, можно провести оценку информации, уже накопленной в базах данных, на предмет выявления мошеннических транзакций, которые остались невыявленными. Кроме того, функция прогноза может использоваться в режиме реального времени для обнаружения мошеннической транзакции в момент ее проведения. Во многих случаях это помогает предугадать мошенничество и предпринять необходимые меры для его пресечения. Качество и надежность инструментов прогнозирования, используемых при выведении коэффициентов, вероятнее всего, со временем будут повышаться. Таким образом, есть основания рассчитывать, что оперативность выявления мошенничества будет повышаться синхронно с ростом объемов данных. Достоинство этого метода состоит в том, что его надежность можно оценить и проверить статистическими методами. Если надежность высока, анализ будет указывать в подавляющем большинстве на действительные случаи мошенничества, а не выдавать набор подозрительных данных, с одинаковой вероятностью как указывающих, так и не указывающих на мошенничество.

## 4. Пример из практики 1: Выявление мошенничества с закладными

Искажение стоимости недвижимости может привести к мошенничеству со страховками на недвижимость или с закладными. Наиболее известна следующая схема – дома действительно покупаются группой людей, а затем перепродаются между членами этой группы по завышенным ценам. В итоге формируется «стоимость», которая указывается в страховых документах на возмещение ущерба или используется при оформлении закладных. Или затем собственность продается ничего не подозревающим покупателям со стороны по искусственно взвинченным ценам. Пример анализа, позволившего выявить этот тип мошенничества, описывается в данном разделе.

### 4.1. Описание входных данных

Вся база данных недвижимости состоит из записей, характеризующих цены на собственность, как сформированные мошенническими методами, так и реальные рыночные цены (последние будут использованы для прогнозирования возможных будущих мошеннических действий). Прогнозный анализ проводится на основании базы данных, содержащей 10 тыс. записей. Исследовательский анализ (анализ ассоциативных связей и ссылок) базируется на подмножестве из 400 записей. В таблице 2 перечислены переменные, содержащиеся в базе данных.

Имя переменной	Описание
PROPERTY	Код недвижимости проданной в каждой транзакции
PRICE	Продажная цена собственности в долларах
DATE	Дата проведения (мм/дд/гг)
FRAUD	Признак того, что данная транзакция является мошенничеством путем циклического обмена собственности
SELLER	Фамилия продавца собственности
SELLER__F	Имя продавца собственности
BUYER	Фамилия покупателя собственности
BUYER_F	Имя покупателя собственности

Таблица 2. Описание входных данных для базы недвижимости. Переменная PROPERTY может быть определенным образом сопоставлена или выведена из собственности.

## 4.2. Пример данных

Из примерно 10 тыс. записей всей базы данных около 5% транзакций представляют мошеннические операции по схеме многократной перепродажи с повышением продажной цены между членами преступной группы. Если бы данные поступали непосредственно из транзакционной системы учета изменений в правах собственности, тогда 5%-ный уровень мошенничества был бы неестественно высоким. Очень часто, однако, тенденция выявления мошенничества наблюдается уже в процессе предварительного отбора данных. Так, например, данные могут концентрироваться вокруг лиц, кредитные карты которых были прежде зафиксированы в заявках или в документах на получение займов по закладным от других кредиторов. Следовательно, не обязательно вслепую отбирать данные из транзакционной системы. В таблице 3 приведены первые 12 записей из подмножества в 400 записей.

Таблица 3. Транзакции с объектами недвижимости

Цель применения аналитических средств добычи данных состоит в повышении про-

PROPERTY	PRICE	DATE	FRAUD	SELLER	SELLER_F	BUYER	BUYER_F
10	111281.34	07/18/92	No	Adams	Mike	Marelli	Cosme
3	156626.68	05/19/92	No	Aird	Pauline	Marignier	Gwen
8	94769.50	06/25/92	No	Allen	Judith	Marshall	Philip
6	105200.12	06/22/92	No	Andrade	Indra	Marsland	Suzanne
12	156045.30	07/28/92	No	Apps	Indra	Mason	Carolyn
9	130916.56	07/02/92	No	Arnold	Michael	Mather	Kevin
2	145170.77	05/05/92	No	Balcombe	Alison	Maurer	Calvin
4	134858.07	05/23/92	No	Banerji	Angela	Mayes	Nicholas
7	114146.35	06/22/92	No	Basker	Christine	Mayr	Lauren
5	113885.36	06/01/92	No	Basnett	Karen	Menendez	Martin
1	94141.77	04/28/92	No	Baston	Nicholas	Miles	Janice
11	84238.96	07/24/92	No	Baxandall	Jason	Miller	Kathleen

цента записей, характеризующих мошенничество, в последующих наборах данных путем корректного нахождения дополнительных случаев мошенничества, и отбрасывания таких записей, которые не свидетельствуют о незаконных махинациях. Схемы мошеннических операций включают в себя циклический обмен собственности, как показано в таблице 4.

Таблица 4. Пример мошеннических транзакций с циклической перепродажей собственности

PROPERTY	PRICE	DATE	FRAUD	SELLER	SELLER_F	BUYER	BUYER_F
213	115489.98	06/04/92	No	Wright	Kay	Aird	Pauline
214	115489.98	07/28/92	No	Wroot	Aileen	Adams	Mike
1001	100332.87	05/05/92	es	Burgess	Nick	Casey	
1001	120399.45	08/03/92	es	Casey	Joan	Crocombe	Agnes
1001	144479.33	10/20/92	es	Crocombe	Agnes	Burgess	Nick
1002	17360017	04/10/92	es	Hills	Shayne	Hole	Maureen
1002	208320.21	06/22/92	es	Hole	Maureen	Horn	Charlene
1002	249984.25	07/02/92		Horn	Charlene	Koll	Heidi
1002	299981.10	08/19/92	es	Koll	Heidi	Hills	Shayne
1005	138813.83	06/11/92	es	O'Sullivan	Roland	Roddick	Sandra

В таблице также есть транзакции, подверженные одному или более обменам в течение определенных периодов времени, но в них мошенничество не выявляется, поскольку отсутствует циклическая составляющая обмена. Некоторые из циклических мошеннических обменов включают в себя также подгруппы покупателей или продавцов, которые покупают или продают не один объект собственности. Следовательно, некоторые из этих фиктивных продавцов или покупателей «работают» с несколькими объектами собственности.

### 4.3. Выявление мошенничества

В данном случае, для выявления мошенничества можно применить несколько подходов, основанных на добыче данных. Очевидно, исследовательский анализ может быть сосредоточен на объектах собственности, цена которых чрезмерно возросла и(или) которые стали объектом слишком большого числа перепродаж за необычно короткое время.

#### 4.3.1. Исследовательский анализ

Установить, какие объекты собственности перепродавались слишком часто, достаточно легко при помощи узла Filter Outliers (Анализ выбросов) программного обеспечения SAS® Enterprise Miner™ (рис. 1). Отбрасывая все объекты собственности, фигурирующие в менее чем трех транзакциях, получаем набор данных, который теперь содержит 20% мошеннических транзакций, что означает существенное повышение по сравнению с первоначальными 5%. Однако выявить записи о мошеннических транзакциях все еще весьма затруднительно, когда речь идет о базе данных, содержащей многие миллионы записей. Добавление к условию многократных продаж ограничения по времени сделает отбор еще более точным и сузит диапазон поиска. Это, в свою очередь, приведет к росту процентного соотношения мошеннических транзакций.

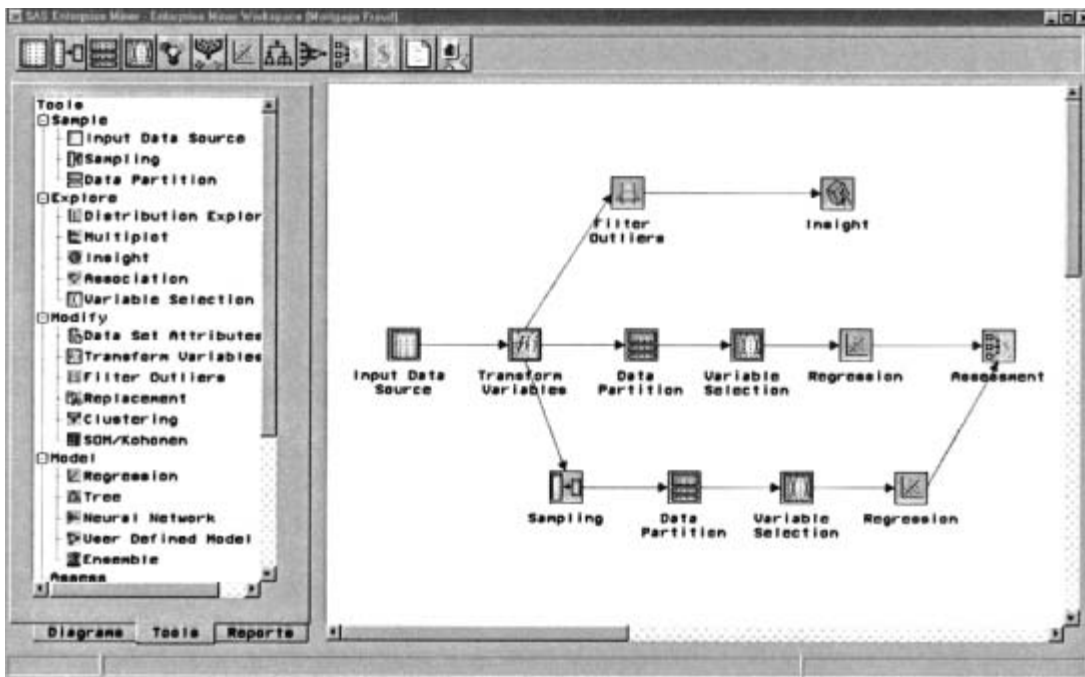


Рисунок 1. Исследовательский анализ и моделирование с функцией Over-Sampling и без нее

Для выявления объектов собственности с сильно завышенными ценами полезно составить сжатый набор данных, в котором каждый объект собственности представлен только одной записью, демонстрирующей начальную и конечную цены на него в течение определенного промежутка времени или определенного количества транзакций. Используя узел Transform Variables (рис. 1), можно сконструировать дополнительные переменные, как, например, такие, которые позволяют отслеживать рост цены и число продаж за определенный временной интервал. В полном наборе данных записи о мошеннических транзакциях встречаются с вероятностью примерно 5%. При помощи узла Filter Outliers можно исключить записи о собственности, участвовавшей только в одной транзакции, как говорилось ранее. Далее, пропуская только записи о тех объектах собственности, цены на которые значительно выросли или которые часто встречались в различных транзакциях за определенный промежуток времени, можно сформировать концентрированный набор данных, содержащий примерно 25% записей о мошеннических транзакциях. Когда речь идет о больших наборах данных, это означает, что еще слишком много случаев подлежат ручному анализу; однако, небольшие наборы можно создать и проанализировать вручную при помощи этого средства. Выявленные случаи мошенничества могут быть затем обобщены в более крупном наборе данных при помощи прогностических моделей.

#### 4.3.2. Прогностическое моделирование

При прогностическом моделировании предпринимаются попытки выполнить обобщение на базе шаблонов, выявленных в тренировочных наборах данных (6). Для этого требуются исторические или предварительно классифицированные данные, где уже проведена идентификация случаев мошенничества.

Консолидированные данные (одна запись на объект недвижимости) «улучшены» дополнительными переменными и разбиты на две части. Одна часть работает с полным набором данных, другая с выборкой. Так как случаи мошенничества редки, выборка в данном сценарии является средством концентрации случаев мошенничества, так чтобы стало возможным обобщение их характеристик. Этот случай изображен на рис. 1, где стратифицированная выборка, использующая 20% данных, приводит к увеличению случаев мошенничества в наборе с 5% до 10%.

Затем данные подразделяются на тренировочный и оценочный наборы данных (7), используя узел Data Partition (рис. 1) и подаются на узел Variable Selection node. Этот узел оценивает относительную важность различных переменных для цели, выделяет дискретные элементы для интервальных переменных и исключает переменные, которые либо не влияют на цель, либо имеют много значений, либо коррелируют с

(6). Тренировочные наборы данных используются для «тренировки» модели, т. е. Для подбора и оценки параметров модели. На рис. 1 изображена стандартная диаграмма для прогностической модели (Ripley 1996, стр. 354).

(7). Оценочный набор данных используется для точной настройки и/или выбора лучшей модели. После того, как лучшая модель выбрана и проверена, она может использоваться для оценки всей базы данных (Ripley 1996, стр. 354).

другими переменными. Результатом этого анализа являются переменные, описанные в табл. 5, которые могут быть использованы для построения модели, а также оценка их значимости для определения целевого отклонения.

Имя переменной	Описание	Роль	Причина отклонения
PRICE	Продажная цена собственности	отклонено	предпочтительнее использовать переменную A16PRICE
DATE	Дата выполнения транзакции по продаже	отклонено	предпочтительнее использовать переменную A16DATE
FREQUENCY	Число продаж данной собственности	отклонено	предпочтительнее использовать групповую переменную G_FREQ
INCREASE	Величина роста продажной цены собственности от одной продажи к другой	входная переменная	
SPAN	Период времени между продажами	входная переменная собственности	
REL_FREQ	Относительная частота то есть частота, на период времени	входная переменная	
G_FREQ	Оптимизированный вариант переменной	входная переменная	времени между продажами
FREQUENCY	автоматически создаваемый Enterprise Miner		
A16SPAN*	16 диапазонов интервалов, выведенных из временных интервалов между продажами для всех транзакций по продажам	входная переменная	
A16PRICE*	16 диапазонов цен, выведенных на основании продажных цен по всем ценам транзакций	входная переменная	
A16REL_F*	16 диапазонов относительной частоты, выведенных из входных частот и интервалов всех транзакций по продажам	входная переменная	
A16DATE*	16 диапазонов дат, выведенных на основании даты всех транзакций по продажам	входная переменная	

Табл. 5. оценка переменных для моделирования

\*Автоматически создаваемые EM переменные, являющиеся вариантами оригинальных переменных со значениями, разбитыми на 16 интервалов.

Затем строится модель с использованием узла Regression (Рис. 1). В качестве альтернативы можно было бы использовать деревья решений или нейронные сети. Полученные в результате анализа релевантные переменные изображены на Рис. 2.

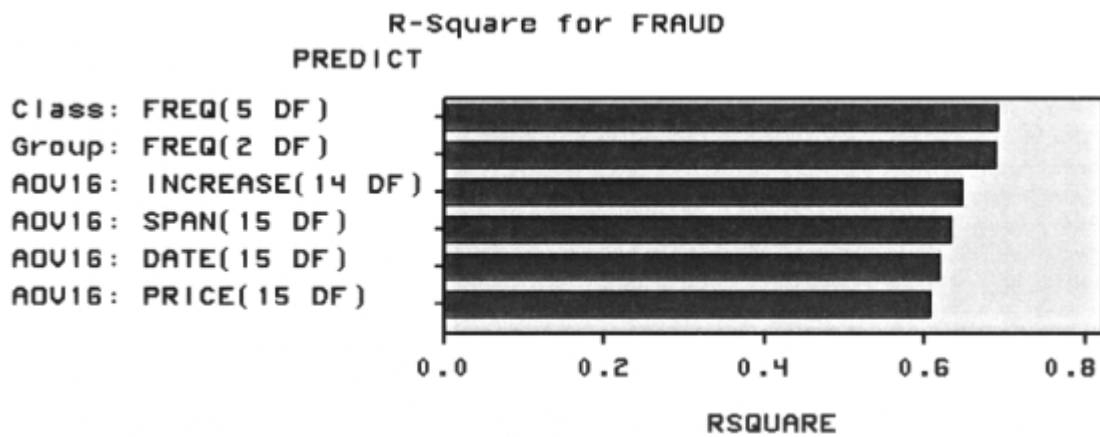


Рисунок 2. Определение целевой релевантности входных переменных

Этот анализ демонстрирует также, как исследовательскую добычу данных можно использовать для инициации прогнозной модели, которая затем приведет к обнаружению ранее невыявленных случаев мошенничества и соответствующих объектов недвижимости. Выявление этих объектов собственности даст аналитикам возможность извлечь записи о вовлеченных в эти операции продавцах и покупателях из исходного файла транзакций. Важно отдавать себе отчет в том, что этот начальный транзакционный набор данных не дает никаких полезных инструментов прогнозирования. При таком типе мошенничества любая отдельно взятая транзакция выглядит абсолютно законно.

На рисунке 3 приведено сравнение двух прогнозов в узле Assessment. В этом сравнении анализ исходных данных слегка превосходит по эффективности анализ предварительно отобранных данных. Если рассмотреть первую десятку указанных в прогнозе случаев мошенничества, вероятность того, что анализ исходных данных (рис. 3, красная линия) выдаст данные о мошеннических транзакциях, в 10 превышает вероятность выдачи таких данных анализом случайно отобранных записей (рис. 3, синяя линия). Действительно, первые десять записей, полученные по этой модели содержат все случаи мошенничества.

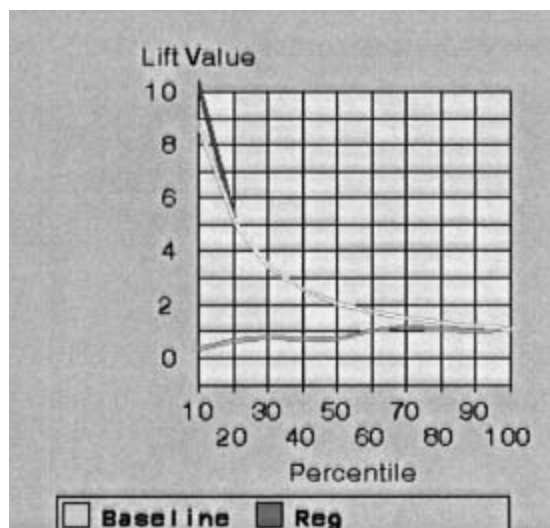


Рисунок 3. Приведено сравнение двух прогнозов в узле Assessment.

### 4.3.3. Анализ ассоциаций и последовательностей

Набор данных содержит также группы и подгруппы людей, постоянно участвующих в циклических операциях по перепродаже собственности. Анализ ассоциаций позволяет выявить эти группы. Для этого необходимо использовать исходный набор данных, преобразуя его таким образом, чтобы для каждого объекта был «виден» его продавец и, в отдельной записи, покупатель каждого объекта. В таблице 6 приведены выборки из этого набора данных.

PROPERTY	DATE	FRAUD	SELLER
1001	05/30/87	Yes	Burgess
1001	05/30/87	Yes	Casey
1001	09/04/87	Yes	Casey
1001	09/04/87	Yes	Crocombe
1001	11/15/87	Yes	Crocombe
1001	11/15/87	Yes	Burgess
1006	08/12/87	Yes	Hills
1006	08/12/87	Yes	Hole
1006	08/17/87	Yes	Hole
1006	08/17/87	Yes	Horn
1006	10/21/87	Yes	Horn
1006	10/21/87	Yes	Koll
1006	01/12/88	Yes	Koll
1006	01/12/88	Yes	Hills
1008	07/02/87	Yes	Moroz
1008	07/02/87	Yes	Morton
1008	09/02/87	Yes	Morton
1008	09/02/87	Yes	Myles
1008	10/03/87	Yes	Myles
1008	10/03/87	Yes	O'Connor
1008	11/17/87	Yes	O'Connor
1008	11/17/87	Yes	Moroz

Таблица 6. Структура данных для анализа ассоциаций

В таблице 7 показано, как анализ ассоциаций идентифицирует группы лиц, часто действующих совместно (переменная SET\_SIZE). Аналогичные результаты получаются при анализе других шаблонов, например, последовательностей требований на возмещение ущерба, повторяющихся через определенные временные интервалы.

SETSIZE	COUNT	PERSON1	PERSON2	PERSON3	PERSON4
00		O'Connor	Myles	Moroz	
00		O'Gonnor	Morion	Moroz	
00		Myles	Morion	Moroz	
00		Mitchell	Meyer	Margol	
00		Mitcrell	Meyer	Leidel	
00		Mitchell	Margol	Leidel	
00		Meyer	Margol	Leidel	
00		Koll	Horn	Hole	
00					
00		Koll	Hole	Hills	
00		Horn	Hole	Hills	
00		Crocombe	Casey	Burgess	
00		O'Connor	Myles	Morion	Moroz
00		Mitcrell	Meyer	Margo)	Leidel
00		Koll	Horn	Hole	Hills

Таблица 7. Результаты анализа ассоциаций

#### 4.3.4. Анализ ссылок

Анализ ссылок – это наиболее прямой путь к выявлению мошенничества путем многократных циклических перепродаж. SAS® Enterprise Miner™ дает пользователям возможность создавать, анализировать и манипулировать представленными графически ссылками. На рис. 4 изображены все характеристики таблицы 3, сведенные в единый граф.

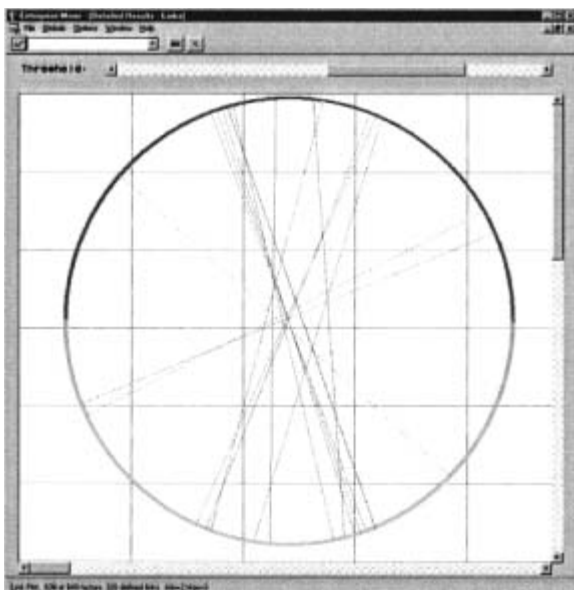


Рисунок 4. Граф ссылок всех характеристик

Отбрасывая ссылки с частотой меньше 2, можно выделить операции циклической перепродажи, как показано на рис. 5.

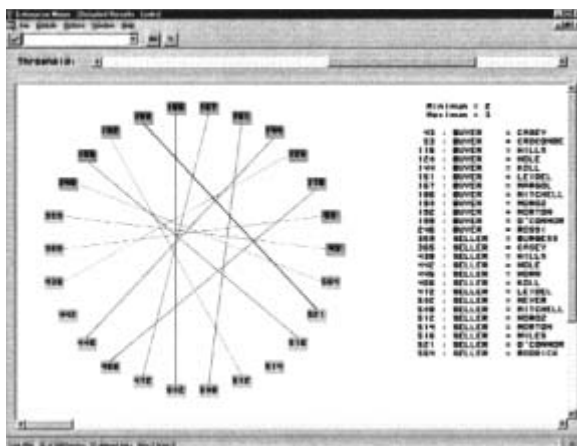


Рисунок 5. Граф ссылок с выделенными подмножествами, соответствующими циклической перепродаже

#### 4.4. Выводы

Добыча данных может оказаться весьма эффективным инструментом для анализа сложных случаев, таких как невыявляемые мошенничества. Для выявления мошенничества часто бывает необходимо использовать несколько типов анализа, выбирая и объединяя перечисленные методы в зависимости от характеристик имеющихся в распоряжении данных и конкретных приложений.

В представленном примере выявления мошенничества использовался прогностический анализ и несколько типов исследовательского анализа. Прогностический анализ вырабатывает код SAS, который может быть применен к новым данным для предотвращения возможного мошенничества. Поскольку случаи мошенничества, как правило, редки, непосредственно применение прогностической модели может не дать адекватных результатов,

пока не будет найдено достаточное количество соответствующих случаев. Для их обнаружения к данным применяются различные исследовательские методики. В данном примере, наиболее эффективной технологией оказался анализ ассоциаций и ссылок. Подобная взаимозависимость исследовательских методик анализа для обнаружения новых случаев мошенничества (основанных, например, на новых схемах мошеннических операций, применяемых злоумышленниками) и прогностических моделей для обобщения этих случаев способствует повышению эффективности обнаружения мошенничества в принципе. Конечной целью является выработка системы коэффициентов, которые не только указывают на состоявшееся мошенничество, но предупреждают о потенциальном преступлении до того, как оно совершилось.

## 5. Пример из практики 2: Выявление мошенничества с кредитными картами

Мошенничество с кредитными картами представляет собой ни что иное, как кражу кредитных карт или информации о кредитных картах и последующее совершение крупных или, наоборот, небольших покупок с помощью этих карт или полученной незаконным путем информации. В случае кражи карт вероятность скорого выявления преступного деяния весьма высока, поскольку наблюдается отчетливая тенденция к немедленному изменению характера покупок, которые продолжают практически постоянно вплоть до того момента, пока деньги на счету не будут исчерпаны. Если же похищена только информация о кредитной карте, характер изменения покупательской активности не столь очевиден, и может занять несколько циклов выставления счетов.

### 5.1. Описание входных данных

Вся база данных кредитных карт состоит из 4736 записей о транзакциях. Данные поступают примерно из 20% записей и идентифицируются целевой переменной BAD\_YN как случаи мошенничества. Все остальные переменные в базе характеризуют индивидуальные карты или транзакции. В таблице 8 перечислены переменные в базе данных кредитных карт.

Имя переменной	Объяснение	Роль	Тип
BAD_YN	Случаи мошенничества	Целевая переменная	binary
COUNT	Частота ежедневного использования карты	Входная переменная	interval
LOG_AMT	Логарифм величины транзакции, поделенный на логарифм времени между транзакциями	Входная переменная	interval
LOG_CUM	Логарифм общего объема транзакций за день, поделенный на логарифм среднего времени между транзакциями в течение дня	Входная переменная	interval
ST_C_RAT	ST_C_AMT, поделенное на интервал времени между использованиями	Входная переменная	interval
ST_C_AMT	Стандартизованный объем ежедневного использования	Входная переменная	interval
ST_AMT	Стандартизованный объем транзакции	Входная переменная	interval
ST_INT	Стандартизованный интервал времени между использованиями	Входная переменная	interval
CLTP2	Тип покупки по объему	Входная переменная	ordinal
T_CLUST	Тип покупки по вероятности мошенничества	Входная переменная	ordinal

Таблица 8. Переменные в базе данных кредитных карт

Большинство этих переменных являются измененными/стандартизованными на базе начальных переменных, например:

- J LOG\_AMT - 16 сегментов значений переменной, полученной делением логарифма от величины объема средств по транзакции на логарифм времени, прошедшего с момента совершения предыдущей транзакции к данному счету
- J LOG\_CUM – переменная, полученная делением логарифма общего объема средств по транзакциям за день на логарифм среднего интервала между транзакциями за день
- J ST\_C\_RAT – стандартизованное отношение суммарного объема транзакций за день к кумулятивному интервалу между транзакциями за день
- J ST\_C\_AMT – стандартизованный объем по транзакциям за день

Подобные трансформации могут быть осуществлены с использованием узла Transform Variable node. Две последние переменные (CLTP2, T\_CLUST) представляют собой группировку характера покупок либо по величине (объему) покупки, либо по степени мошенничества. В таблице 9 показаны в качестве примера первые 12 записей из базы данных кредитных карт.

NUM	BAD_YN	COUNT	LOG_AMT	LOG_CUM	ST_C_RAT	ST_C_AMT	ST_AMT	ST_INT	CLTP2	T_CLUST
1	No	2	2.0866377441	2.1740741315	0.0002168675	0.0081	0.02	0.1245	4	0
2	Yes	2	2.5619471721	2.6721216099	0.0005284553	0.0065	0.016	0.041	3	9
3	No	2	18.194602975	19.194602975	0.1	0.03	0.06	0.001	3	6
4	No	3	2.2950686732	2.3699208302	0.001294964	0.081	0.204	0.2075	4	6
5	No	2	2.925860955	3.1323474464	0.0012819444	0.00923	0.0166	0.024	3	6
6	No	2	3.1611819439	3.3374733783	0.0032679739	0.025	0.05	0.0255	2	6
7	No	3	3.3741713532	2.998801687	0.0028673835	0.04	0.06	0.021	2	6
8	No	4	.	3.061835387	0.0039007092	0.055	0.06	0.0005	2	6
9	No	2	11.416716353	12.014542851	0.06	0.027	0.056	0.0015	2	6
10	No	2	.	.	0.1573333333	0.0236	0.0404	0.0005	4	6
11	No	2	.	.	.	0.074	0.148	0	4	8
12	No	3	.	.	0.66	0.099	0.1	0.0005	4	8

Таблица 9. Записи транзакций по кредитным картам

## 5.2. Анализ

На рисунке 6 приведена типичная диаграмма процессов, которая в данном случае используется для выявления мошенничества с кредитными картами. Узел Data Replacement «отвечает» за недостающие значения; узел Data Partition генерирует наборы данных для обучения и контроля; узел Variable Selection сегментирует ряд переменных, определяет значимые переменные и отбрасывает ненужные переменные. Используются два моделирующих узла, Decision Tree (дерево решений) и Neural Network (нейронная сеть), а результаты их работы сравниваются в узле Assessment.

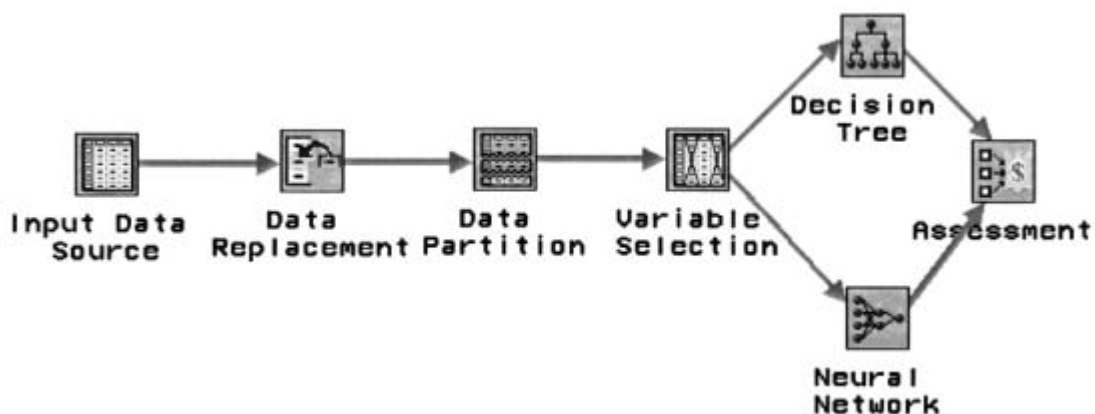


Рисунок 6. Диаграмма процессов для выявления мошенничества с кредитными картами

Сравнение результатов, выполненное в узле Assessment, может быть визуализировано в виде различных графиков и таблиц. Один из наиболее информативных и наиболее широко используемых – это диаграмма взвешенности (lift chart), приведенная на рис.7.

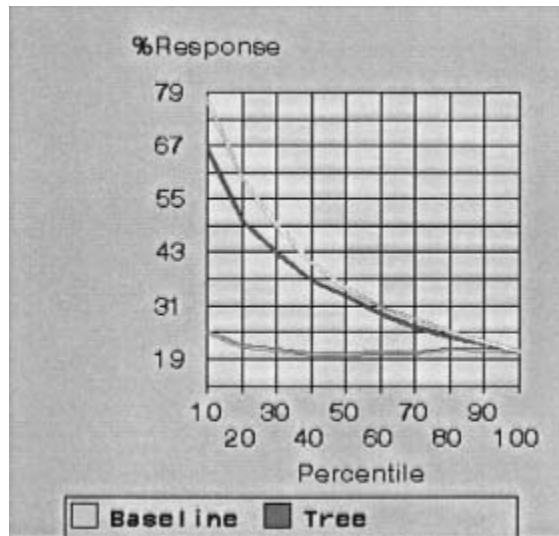


Рисунок 7. Диаграмма взвешенности для сравнения результатов

В диаграмме взвешенности данные отсортированы по степени вероятности, с которой они могут свидетельствовать о мошенничестве исходя из прогнозов двух моделей - дерева решений (рис. 7, красная линия) и нейронной сети (рис. 7, желтая линия). На диаграмме видно, что среди 10% данных, которые, согласно прогнозу нейронной сети наиболее вероятно свидетельствуют о мошенничестве, в действительности являются таковыми на 78%. Соответственно, для дерева решений этот показатель составляет 66%. Для 20% данных эти цифры равны, соответственно, 61% и 50%. В любом случае эти показатели значительно превышают вероятность того, что случайно выбранная запись будет соответствовать мошенничеству (что показано на рис.7 синей линией). Диаграмма также показывает, что использование нейронной сети в целом дает лучшие результаты, чем дерева решений.

Преимущество дерева решений, однако, состоит в том, что его результаты легче интерпретировать, нежели результаты нейронной сети. В этом легко убедиться, посмотрев на результаты, полученные с применением дерева решений, приведенные на рис. 8 и 9. Показаны три верхних уровня, разделенных для удобства чтения на левую и правую половину, показанные на отдельных рисунках.

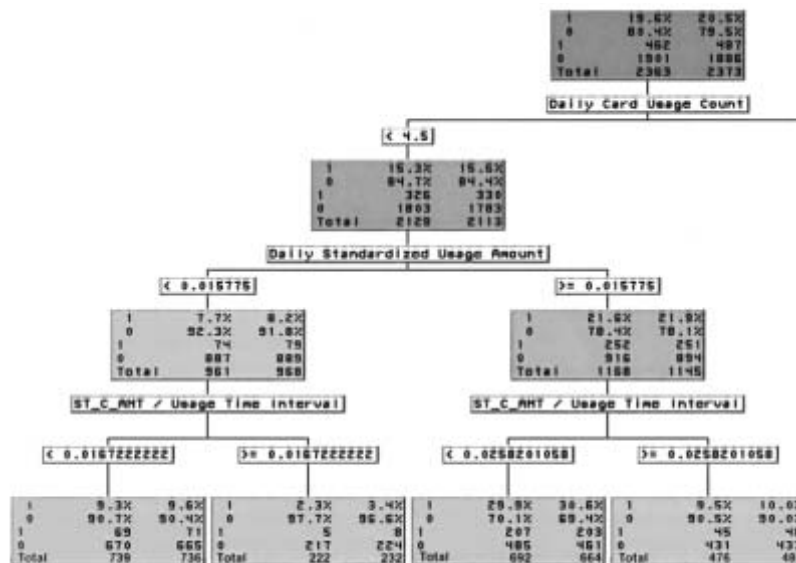


Рисунок 8. Первые три уровня левой половины дерева решений

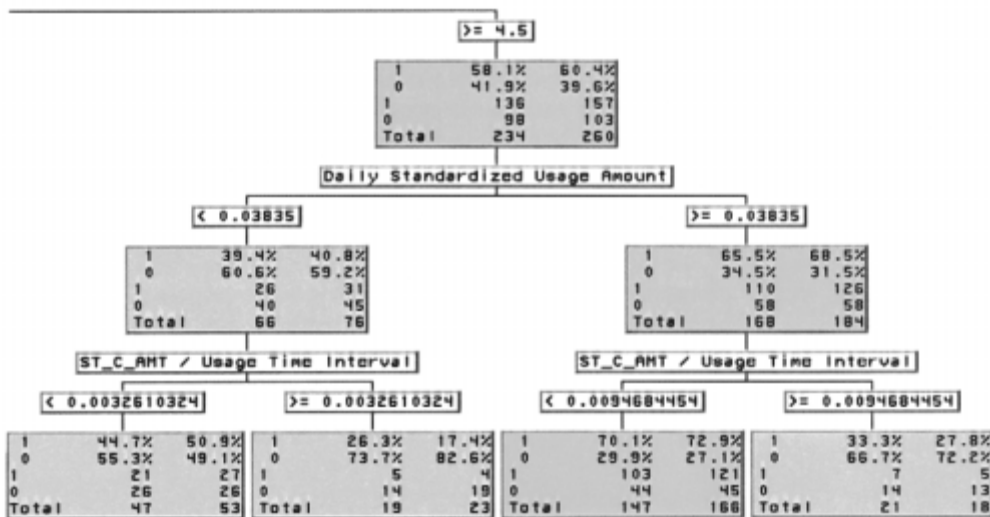


Рисунок 9. Первые три уровня правой половины дерева решений

Среди наиболее значимых переменных следующие:

- число использований карточки в день (переменная COUNT);
- стандартизованный объем ежедневного использования (переменная STD\_C\_AMT);
- соотношение объема по транзакциям к кумулятивному интервалу времени между транзакциями (ST\_C\_AMT/Usage Time Interval), указывающее на аномально большой объем покупок за короткое время.

Большинству случаев мошенничества соответствуют транзакции с высоким показателем использования и большими объемами трат в течение одних суток.

Как правило, большинство наиболее ответственных критериев, влияющих на целевые переменные, которые подлежат анализу (в данном случае это переменная BAD\_YN), сосредоточено ближе к корню дерева. В данном примере наиболее важным критерием, позволяющим отличить преступные деяния от законных, является значение Daily Card Usage Count. Затем анализируется Daily Standardized Usage Amount, за которым следует соотношение объема затрат ко времени, прошедшему со времени последнего использования карты.

### 5.3. Заключение

Используя прогностическую добычу данных, аналитик может построить модели для прогнозирования вероятности того, что та или иная транзакция окажется мошеннической. Если используются деревья решений, то модель легко визуализировать в виде последовательности правил. Таким образом, ее корректность легко может быть проверена специалистами бизнеса, после чего такая модель легко реализуется в виде программы обнаружения мошенничества, либо в качестве системы раннего предупреждения. Результаты работы нейронной сети труднее поддаются интерпретации, однако, этот метод может быть эффективнее дерева решений. Сравнение инструментов моделирования осуществляется в узле Assessment и позволяет аналитикам выбирать наиболее эффективную модель для решения конкретной задачи. Каждый инструмент моделирования вырабатывает коэффициенты (с соответствующим программным кодом), которые могут быть применены к новым данным, что повышает вероятность прогнозирования мошенничества в будущих транзакциях.

## 6. Ссылки

- Cassidy, Peter, 1997, "Credit Card Fraud a \$1 Billion Problem-How Much of It Is Yours?" NetscapeWorld, August, <http://www.netscapeworld.com/netscape-world/nw-08-1997/nw-08-cybersleuth.html> (accesses 14 Feb. 2000).
- Elliott, Jeffrey, 1998, "Software Developers Come Out Ahead," Healthcare Informatics Online, News and Trends, June, [http://www.healthcare-informatics.com/issues/1998/06\\_98/news.htm](http://www.healthcare-informatics.com/issues/1998/06_98/news.htm) (accessed 19 Apr. 2000).
- Farrell, Bill, (1999), "4 Grabbed in Medical Insurance Con," New York Daily News Online, July 1, [http://www.nydailynews.com/1999-07-01/News\\_and\\_Views/City\\_Beat7a-33555.asp](http://www.nydailynews.com/1999-07-01/News_and_Views/City_Beat7a-33555.asp) (accessed 17 Nov. 1999).
- General Accounting Office, 1999, "Money Laundering: U.S. Efforts to Fight It Are Threatened by Currency Smuggling," March 9, Chapter Report, GAO/GGD-94-73, <http://www.fas.org/irp/gao/ggd94073.htm> (accessed 3 Dec. 1999).
- Insurance Services Office, Inc., 1999, "ISO and NICB to Cooperate in Building an All-Claims Database to Fight Insurance Fraud," News from ISO, August 11, <http://www.iso.com/docs/pres047.htm> (accesses 3 Dec. 1999).
- New York Central Mutual, (1999), "Insurance Fraud, It's a Crime," Insurance Fraud, December 9, <http://www.nycm.com/Fraud/Fraud.htm> (accessed 14 Feb. 2000).
- Provost, Taran, (1999), "Ohio Bank Tells SEC: Exec, Client Conspired in 3-Year, \$15M Fraud," American Banker Online, October 27, <http://www.american-banker.com> (accessed 1 Dec. 1999).
- Steel, Billy, (1999), "Money Laundering - How Big is the Problem?" Billy's Money Laundering Information Website, April 26, [http://www.laundryman.u-net.com/page3\\_probl.html](http://www.laundryman.u-net.com/page3_probl.html) (accessed 14 Feb. 2000).

## 7. Сопровождающий документ

- SAS Institute Inc., (1998), SAS Institute White Paper, "Finding the Solution to Data Mining: A Map of the Features and Components of SAS® Enterprise Miner™ Software," Gary, NC: SAS Institute Inc.

## 8. Рекомендуемая литература

### Fraud

- FINCEN: The FinCEN Artificial Intelligence System: "Identifying Potential Money Laundering from Reports of Large Cash Transactions," Proc. 7th Annual Conf. IAAI, Menlo Park, CA: AAAI.
- Guidette, Christopher, and Benzing, Jeff, (February 1999), "NICB Turns Its Claims Databases Over To ISO In Move To Fight Insurance Fraud By Building One All-Claims Database," <http://www.iso.com/docs/pres060.htm> (accessed 23 Nov. 1999).
- Hoffman, Thomas, (October 1996), "Empire Strikes Back Against Legacy System," Computerworld, 30 (43), 12.

- Hoffman, Thomas, and Nash, Kim S, (July 1995), "Data Mining Unearths Customers," *Computerworld*, 29 (28), 1, 28.
- Investigative Data Mining, Ltd., "Further Reading on Fraud and Fraud Detection," <http://www.c-r.co.uk/IDM/framefurther.html> (accessed 29 Nov. 1999).
- Insurance Fraud Bureau of Massachusetts, "Insurance Fraud Register," <http://www.ifb.org/IFRR/ifrrjnd.htm> (accessed 29 Nov. 1999).
- Sparrow, Malcolm K (1996), *License to Steal: Why Fraud Plagues America's Health Care System*, New York: Westview Press.
- Teplitzky, Sandy, (Summer/Fall 1997), "Balanced Budget Act of 1997: More Health Care Fraud and Abuse Measures," <http://www.ober.com/pubs/health/current/health.htm> (accessed 23 Nov. 1999).
- Way, Paul, (August 1996), "Managing Knowledge: the CIO's Next Challenge," *Insurance & Technology*, 21 (8), 52.
- Way, Paul, (August 1999), "Decision Time for Decision Support," *Insurance & Technology*, 21 (8), 30-34.
- Williams, Graham J, and Huang, Zhexue, (1997), "Mining the Knowledge Mine: The Hot Spots Methodology for Mining Large Real World Databases," <http://www.cmis.csiro.au/Graham.Williams/papers/ai97.html> (accessed 23 Nov. 1999).
- Williams, Nia, (March 1994), "Data Mining with Neural Networks," *Insurance Systems Bulletin*, 9 (7), 3-4.



**Московское  
представительство  
109240, Москва,  
Николаямская ул., 13  
Тел.: +7 095 937 4151  
Факс: +7 095 937 4155  
<http://www.sas.com/russia>**

**SAS Institute Inc.  
Европейская штаб квартира  
Neuenheimer Landstr. 28-30  
P.O. Box 10 53 40  
D-69043 Heidelberg, Germany  
Тел.: +49 6221 4160  
Факс: +49 6221 474850**

**SAS Institute Inc.  
Мировая штаб-квартира  
SAS Campus Drive,  
Cary, NC 27513 USA  
Тел.: +1 919 677 8000  
Факс: +1 919 677 4444  
<http://www.sas.com>**