



Для чего предназначен пакет SAS Enterprise Miner?

Пакет SAS Enterprise Miner позволяет оптимизировать процесс интеллектуального анализа данных в целом, начиная от организации доступа к данным и заканчивая оценкой готовой модели, и поддерживает выполнение всех необходимых процедур в рамках единого интегрированного решения с гибкими возможностями коллективной работы.

В чем значимость пакета SAS Enterprise Miner?

Этот пакет представляет собой наиболее мощное и полнофункциональное из всех имеющихся на рынке решение для интеллектуального анализа данных. Он предоставляет не имеющие аналогов средства создания и развертывания моделей, а также широкие возможности интеграции. Пакет поставляется как распределенное клиент-серверное приложение, что особенно удобно для осуществления анализа данных в масштабах крупных организаций.

Для кого предназначен пакет SAS Enterprise Miner?

Пакет SAS Enterprise Miner предназначен для специалистов по анализу данных, маркетинговых аналитиков, маркетологов, специалистов по анализу рисков, специалистов по выявлению мошеннических действий, а также инженеров и ученых, ответственных за принятие ключевых решений в бизнесе или исследовательской деятельности.

SAS® Enterprise Miner 5.1

Извлечение подлинных знаний - реальные результаты с меньшими затратами времени и усилий

Превращение неуклонно растущих объемов необработанных данных в полезную информацию остается серьезной проблемой во многих организациях, поскольку знания, необходимые для повышения эффективности и выявления новых возможностей, часто оказываются скрыты в огромных массивах данных. Какие покупатели приобретут те или иные продукты и когда они это сделают? Как установить цены, чтобы обеспечить прибыльность? Какие клиенты покидают нашу компанию и что нужно сделать, чтобы удержать их? Какое влияние оказывают планы технического обслуживания и эксплуатации на время безотказной работы компонентов?

Для ответа на эти и другие подобные вопросы требуются мощные аналитические решения, позволяющие извлекать знания из огромных массивов данных. Методы углубленного интеллектуального анализа данных - отбора, исследования и моделирования огромных массивов данных - могут позволить обнаружить неизвестные ранее закономерности и усовершенствовать процесс принятия решений в масштабах всего предприятия. При эффективном использовании таких методов задача может оказаться колоссальной.

Однако процесс интеллектуального анализа данных может оказаться громоздким и неэффективным. Большинство представленных на рынке решений ограничены с точки зрения интеграции, масштабирования или типов решаемых проблем. В результате эксперты, занимающиеся количественным анализом, вынуждены тратить драгоценное время на поиск, подготовку и обработку разнородных данных вместо моделирования и применения своего опыта для решения задач бизнеса. По мере усложнения задач бизнеса эта проблема становится все более серьезной.

Современным организациям требуется мощное многофункциональное решение, обеспечивающее совместную работу над проектами интеллектуального анализа данных в масштабах всего предприятия и допускающее настройку с учетом разнообразных по-

требностей предприятия. Пакет SAS Enterprise Miner обеспечивает эффективную обработку огромных объемов данных, освобождая время для создания точных прогностических и описательных моделей, а также предоставляет простые способы публикации результатов анализа для различных аудиторий, что позволяет встраивать эти модели в бизнес-процессы предприятия.

Основные преимущества

- Широкий набор инструментальных средств, поддерживающих процессы интеллектуального анализа данных.** Каковы бы ни были конкретные потребности вашего предприятия, SAS предлагает гибкое интегрированное решение, обеспечивающее поддержку всех шагов, необходимых для решения той или иной проблемы бизнеса. Процесс перехода от необработанных исходных данных к точным аналитическим моделям, отвечающим на потребности бизнеса, становится гладким и эффективным. Пакет предоставляет интегрированную среду для коллективной работы, обеспечивающую эффективное сотрудничество подразделений статистического моделирования, менеджеров и ИТ-специалистов.
- Понятный графический интерфейс пользователя (GUI) помогает бизнес-аналитикам и специалистам по статистике создавать в более сжатые сроки большее количество моделей.** Среда для формирования диаграмм процессов обработки данных пакета SAS Enterprise Miner устраняет необходимость ручного кодирования и значительно сокращает время разработки моделей бизнес-аналитиками и специалистами по статистике. Кроме того, диаграммы выступают в качестве самоописательных шаблонов, которые можно легко изменять или применять для решения новых про-

блем, не повторяя анализ с самого начала. Существует возможность обмена диаграммами между аналитиками в масштабах предприятия.

- **Упрощение выявления надежной бизнес-информации.**

Пакет SAS Enterprise Miner оснащен рядом встроенных функций контроля, работающих в рамках единой оболочки и обеспечивающих сравнение результатов различных методов моделирования - как с точки зрения статистики, так и с точки зрения бизнеса. Полученные модели можно публиковать для совместного использования в рамках предприятия при помощи уникального репозитория моделей, представляющего собой первую на рынке систему управления моделями.

- **Развертывание моделей в масштабах предприятия с непревзойденной простотой.**

Скоринг - это применение модели к новым данным и конечный результат деятельности по интеллектуальному анализу данных. Пакет SAS Enterprise Miner автоматизирует умительный процесс подбора коэффициентов и предоставляет готовый программный код для скоринга на всех стадиях создания модели. Этот программный код может использоваться в различных средах (в пакетном режиме или в реальном времени) - в системе SAS, в Web или непосредственно в реляционных базах данных. Это позволяет быстрее перейти к практическому использованию результатов анализа данных.

Обзор программного продукта

Пакет SAS Enterprise Miner 5.1 поставляется в виде современной распределенной клиент-серверной системы для углубленного интеллектуального анализа данных в крупных организациях. Пакет позволяет оптимизировать процессы анализа данных, поддерживая все необходимые шаги в рамках единого решения, а также обеспечивает гибкое сотрудничество больших рабочих групп в рамках единого проекта. Система обеспечивает расширенную интеграцию с системами управления данными и развертывания моделей, а гибкие возможности выбора конфигурации пакета в зависимости от требований бизнеса устраняют необходимость приобретения специализированных решений.

Продуманный и логичный графический интерфейс для успешного анализа данных

В пакете SAS Enterprise Miner реализован удобный подход, основанный на создании диаграмм процессов обработки данных и позволяющий устранить необходимость ручного кодирования и ускорить разработку моделей. Благодаря проверенной методике анализа данных (SEMMA), направляющей исследователя к правильному результату, опытные специалисты в области статистики и начинающие бизнес-аналитики получают возможность создавать больше моделей лучшего качества. Подход SEMMA подразумевает, что все процессы выполняются в рамках гибкой оболочки, поддерживающей выполнение всех необходимых работ по обработке и анализу данных, разбитых на пять основных шагов - создание выборки, исследование, модификация, моделирование и оценка. Уникальный подход SEMMA, разработанный SAS, сочетает структурированность процесса и логическую организацию инструментальных средств, поддерживающих выполнение каждого из шагов. Благодаря легко изменяемому, сохраняемому и используемому совместно диаграммам процессов обработки данных подход SEMMA упрощает применение методов статистического исследования и визуализации, позволяет выбирать и преобразовывать наиболее значимые переменные, создавать модели с этими переменными, чтобы предсказать результаты, подтвердить точность модели и подготовить модель к развертыванию.

Высокопроизводительный инструментальный интеллект для углубленного интеллектуального анализа данных

Новая версия пакета SAS Enterprise Miner 5.1 была заново спроектирована с использованием новаторской архитектуры Java-клиент / SAS-сервер, позволяющей отделить вычислительный сервер, выполняющий обработку данных, от пользовательского интерфейса. Это обеспечивает беспрецедентную гибкость в выборе конфигурации эффективного решения в диапазоне от однопользовательской системы до крупнейших решений корпоративного масштаба. Обработку данных можно выполнять на мощных серверах, а конечные пользователи могут перемещаться из офиса домой или в отделенные филиалы, не теряя связи с аналитическими проектами и сервисами. Некоторые сер-

верные задачи, интенсивно использующие ресурсы процессора, например, сортировка и агрегация данных, отбор переменных и регрессионный анализ, сделанные многопоточными, что позволяет распределить их выполнение между несколькими процессорами.

Процессы в Enterprise Miner могут работать параллельно и в асинхронном режиме. Масштабные или повторяющиеся процессы обучения модели или скоринга могут быть выполнены в виде пакетного задания, назначенного на наименее загруженные часы работы аналитического сервера.

Интегрированный комплекс разнообразных методов моделирования

Пакет SAS Enterprise Miner обеспечивает небывалую глубину анализа, предоставляя не имеющий аналогов по разнообразию набор инструментов и алгоритмов прогностического и описательного моделирования, включающий деревья решений, нейронные сети, самоорганизующиеся нейронные сети, методы рассуждения, основанные на механизмах поиска в памяти (memory-based reasoning), линейную и логическую регрессию, кластеризацию, ассоциации, временные ряды и многое другое. Благодаря интеграции множества моделей и алгоритмов в одном приложении пакет Enterprise Miner позволяет производить последовательное сравнение моделей, созданных на основе различных методов, оставаясь в рамках единого графического интерфейса. Встроенные средства оценки формируют единую среду для сравнения различных методов моделирования, как с точки зрения статистики, так и с точки зрения бизнеса, позволяя выявить наиболее подходящие методы для имеющихся данных. Результатом является качественный анализ данных, выполненный с учетом специфических проблем конкретного бизнеса.

Мощный набор инструментов для подготовки, агрегации и исследования данных

Подготовка данных к анализу часто является самым продолжительным процессом всего аналитического исследования - но не в случае SAS Enterprise Miner. В состав пакета входят различные инструменты, которые помогут вам осуществить подготовку данных, например, сделать выборку или разбивку данных, осуществить вставку недостаю-

щих значений, провести кластеризацию, объединить источники данных, устранить лишние переменные, выполнить обработку на языке SAS посредством специализированного узла SAS code, осуществить преобразование переменных и фильтрацию недостоверных данных. Кроме того, пакет оснащен функциями описательной статистики, а также расширенными средствами визуализации, позволяющими исследовать огромные объемы данных, представленных в виде многомерных графиков, и производить графическое сравнение результатов моделирования.

Платформенно-независимый пользовательский интерфейс пакета SAS Enterprise Miner 5.1 создан на базе Java и предоставляет пользователям широкий набор средств статистической графики с гибкими возможностями настройки и управления. Для создания специальных графиков предусмотрен удобный Java-мастер. Все графики и лежащие в их основе таблицы динамически связаны между собой и поддерживают интерактивные режимы работы.

Интегрированные средства сравнения моделей и пакеты результатов

Пакет SAS Enterprise Miner предоставляет ряд встроенных оценочных функций, позволяющих сравнить результаты различных методов моделирования, как в терминах бизнеса, так и с использованием статистической диагностики. Это дает уникальную возможность измерить эффективность модели в терминах ее прибыльности, позволяя специалистам по количественному анализу опубликовать значимые результаты и обсудить их с бизнес-пользователями. Кроме того, аналитики получают возможность наблюдать за обновляемыми моделями и отслеживать улучшение их точности с течением времени. Созданные диаграммы можно сохранять и импортировать в виде XML-файлов, что облегчает процесс их передачи другим аналитикам. SAS Enterprise Miner дает возможность создавать сжатые пакеты с результатами моделирования, в которых хранится вся информация о процессе обработки данных, включая предварительную обработку данных, логику моделирования, результаты моделирования и оценочный код. Эти пакеты результатов могут быть зарегистрированы на сервере метаданных (SAS Metadata Server), откуда их потом могут извлекать для изучения специалисты по анализу данных и представители бизнеса. Специальный модуль с Web-интерфей-

сом для просмотра репозитория моделей - единственный в отрасли Web-модуль, обеспечивающий эффективное управление и распространение портфелей моделей в рамках предприятия.

Скоринг по модели и беспрецедентная простота развертывания модели

Итогом работ по интеллектуальному анализу данных является развертывание созданной модели - это заключительная стадия, на которой реализуется экономическая отдача от проведенных исследований. Процесс применения модели к новым данным, известный как скоринг, может оказаться утомительным и часто требует ручного написания или преобразования программного кода, что может значительно задержать развертывание модели и увеличить вероятность внесения ошибок. Очень важно, чтобы код для скоринга по модели был полным и кратким, а также поддерживал различные программные среды для развертывания модели. Пакет SAS Enterprise Miner позволяет автоматизировать этот процесс и поддерживает создание исчерпывающих скоринг-кодов на языках SAS, C, Java и PMML. Пакет создает код не только для аналитических моделей, но и для предварительной обработки данных.

Когда оценочный код создан, вы можете как провести скоринг наборов данных непосредственно в Enterprise Miner, так и экспортировать скоринг-код и выполнить скоринг на другой машине, а также оторгнуть формулу для скоринга для применения в пакетном режиме или в режиме реального времени в Web или непосредственно в реляционных базах данных. Это способствует ускорению развертывания моделей и освобождает время аналитиков для создания новых высококачественных моделей.

Высочайшая гибкость благодаря открытости и расширяемости

Пакет Enterprise Miner предоставляет настраиваемую и расширяемую среду интеллектуального анализа данных, позволяющую добавлять инструментальные средства и интегрировать персонализированный код на языке SAS. Стандартную инструментальную библиотеку, входящую в состав пакета SAS Enterprise Miner 5.1, легко расширить при помощи средств настройки, использующих язык SAS и XML-логику. Кроме того, есть возможность использования экспериментального интерфейса Java API, поз-

воляющего встраивать процессы пакета Enterprise Miner в различные пользовательские приложения. Эта возможность может оказаться особенно плодотворной для компаний, стремящихся создать собственное аналитическое приложение, сочетающее в себе, например, возможности создания OLAP-отчетов и выполнения интеллектуального анализа данных в рамках единого интерфейса.

Встроенная стратегия обнаружения данных

Интеллектуальный анализ данных становится особенно эффективным, если он является составной частью интегрированной стратегии предоставления информации. Пакет Enterprise Miner органично интегрируется с другими предложениями SAS, например, пакетом SAS ETL Studio, средствами аналитической обработки OLAP, прогностическим и другими аналитическими модулями, а также с приложением SAS Text Miner. Такое сочетание различных приложений формирует комплексное решение, способное полностью удовлетворить потребности в анализе данных и добыче знаний. В результате вы получаете возможность избежать проблем, связанных с извлечением, очисткой и преобразованием данных. Вы получаете доступ именно к тем данным, которые вам нужны, и в тот момент, когда они нужны, и можете вернуться на несколько шагов назад и реструктурировать данные с учетом специфики изучаемой проблемы.

Современная, распределенная система интеллектуального анализа данных, ориентированная на крупные предприятия

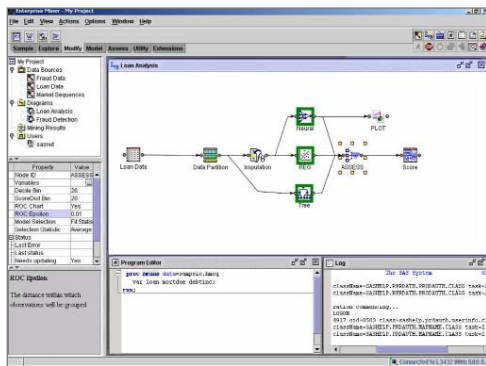
Пакет SAS Enterprise Miner может быть развернут с использованием Web-портала для тонких клиентов, что обеспечивает удобный доступ к пакету для множества пользователей при минимальных затратах на обслуживание клиентских программ.

При другом варианте установки вся система может быть установлена на автономном ПК. Кроме того, в отличие от многих приложений интеллектуального анализа данных, которые работают только на одной-двух платформах, пакет SAS Enterprise Miner поддерживает серверные системы Windows, а также различные UNIX-платформы, что делает пакет предпочтительным для организаций, ведущих крупномасштабные проекты в области интеллектуального анализа данных.

Основные характеристики пакета SAS Enterprise Miner 5.1

Разнообразные интерфейсы

- Простой графический интерфейс, обеспечивающий создание диаграмм процессов обработки данных:
 - Быстрое создание большого числа качественных моделей.
 - Возможность доступа через Web-интерфейс.
 - Доступ к среде программирования SAS.
 - Возможность обмена диаграммами в формате XML.
 - Возможность повторного использования диаграмм в качестве шаблонов для других проектов и пользователей.
- Пакетная обработка:
 - Включает в себя все те же возможности, что и графический интерфейс.
 - Основана на языке SAS macro.
- Экспериментальный интерфейс Java API.
- Репозиторий моделей с Web-интерфейсом:
 - Управление большими портфелями моделей.
 - Поиск моделей по заданному алгоритму, целевой переменной и т.п.
 - Публикация результатов в виде ступенчатых диаграмм, деревьев и скоринг-кодов, удобных для специалистов в области бизнеса и анализа данных.



Простой и понятный графический интерфейс пакета SAS Enterprise Miner позволяет создавать большее число моделей в сжатые сроки.

Масштабируемая обработка

- Серверная обработка - обучение модели в асинхронном режиме. Аккуратная остановка обработки (по заданным критериям).
- Параллельная обработка - одновременный запуск нескольких диаграмм или инструментов.
- Многопоточные прогностические алгоритмы.
- Все хранение и обработка данных - на серверах.

Доступ к данным

- Доступ более чем к 50 различным файловым структурам.
- Интеграция с пакетом SAS ETL Studio посредством SAS Metadata Server:
 - SAS ETL Studio можно использовать для определения исходных, обучающих таблиц для пакета Enterprise Miner.
 - SAS ETL Studio можно использовать для извлечения и развертывания скоринг-кода пакета Enterprise Miner.

Выборки

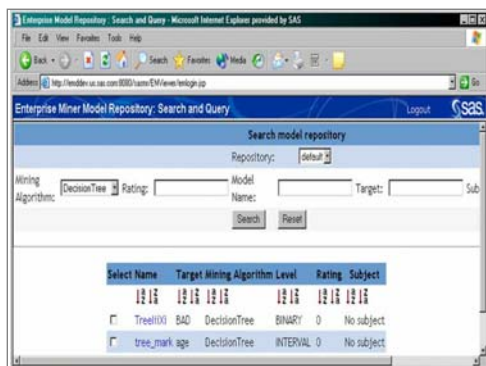
- Простая случайная.
- Стратифицированная.
- Взвешенная.
- Кластерная.
- Систематическая.
- Первые N наблюдений.
- Выборка редких событий.

Разбивка данных

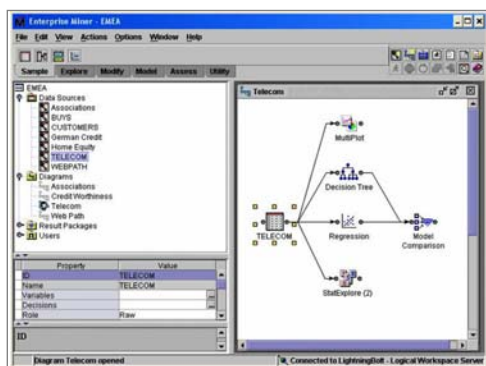
- Создание обучающих, проверочных и тестовых наборов данных.
- Обеспечение качественного обобщения моделей на основании контрольных данных.
- Стандартная стратификация по целевому классу.
- Сбалансированная разбивка по любой классовой переменной.

Преобразования

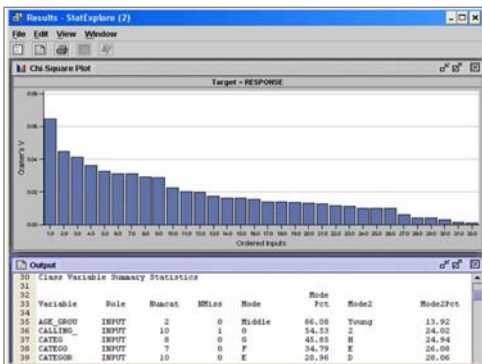
- Простые: логарифмическое, квадратный корень, обратное, квадратичное, экспоненциальное, стандартизованное.
- Накопительные: bucketed (с разбивкой по областям), квантильное, оптимизированная разбивка по взаимосвязи с целевыми значениями.
- Оптимизированные: максимизация нормализации, максимизация корреляции с целевыми значениями, выравнивание распределения по целевым уровням.



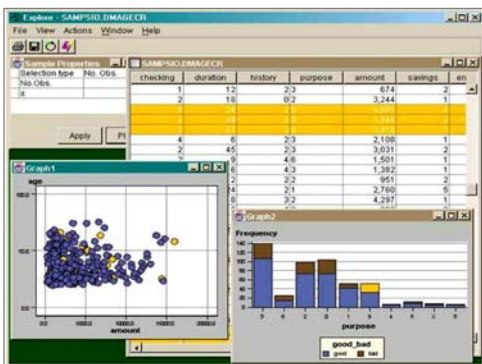
Модуль Enterprise Miner Repository обеспечивает управление моделями.



Удобное задание источника данных для многократного использования в проекте.



Enterprise Miner позволяет быстро получать описательные статистики.



Enterprise Miner предоставляет гибкие возможности визуализации данных.

Фильтрация недостоверных данных

- Применение различных распределительных порогов, позволяющих исключить значения из экстремальных интервалов.
- Объединение классовых значений, встречающихся менее p раз.

Замена данных

- С использованием мер центрированности.
- На основе распределения.
- Заполнение дерева суррогатными значениями.
- Методом усреднения расстояний.
- С использованием устойчивых M -оценок.
- С использованием стандартных констант.

Описательная статистика

- Одномерные статистические таблицы и графики:
 - Интервальные переменные - n , среднее, медиана, минимум, максимум, стандартное отклонение, масштабированное отклонение и процент отсутствия.
 - Классовые переменные - число категорий, счетчики, модальные, процентные модальные, процент отсутствия.
 - Графики распределения.
 - Статистическая разбивка для каждого уровня целевых классов.
- Двумерные статистические таблицы и графики:
 - Упорядоченные графики корреляции Пирсона и Спирмана.
 - Упорядоченный график хи-квадрат с возможностью группировки непрерывных исходных данных по n группам.
 - График коэффициентов вариации.
- Отбор переменных по logworth-критерию.
- Другие интерактивные графики:
 - «Тепловые» карты, отражающие корреляцию или ассоциацию типа хи-квадрат первоначальных значений с целевыми признаками по сегментам.
 - Графики стоимости переменных, ранжирующие первоначальные значения на основании их стоимости по целевому признаку.
 - Распределения классовых переменных по целевым признакам и/или сегментным переменным.
- Графики масштабированного среднего отклонения.

Графика/визуализация

- Графики, создаваемые в пакетном и интерактивном режимах: графики разброса, гистограммы, многомерные графики, круговые диаграммы, диаграммы с областями, пузырьковые диаграммы.
- Удобный Java-мастер для построения графиков:
 - Заголовки и сноски.
 - Возможность применения к данным предложения WHERE.
 - Возможность выбора из нескольких цветовых схем.
 - Простота масштабирования осей.
 - Использование данных, полученных в результате анализа в пакете Enterprise Miner, для создания специализированных графиков.
- Динамическая загрузка данных в клиентское приложение при помощи нескольких методик выборки.
- Графики и таблицы интерактивно связаны между собой и поддерживают выполнение таких операций, как очистка и связывание.
- Удобное копирование данных и графиков в другие приложения, а также возможность их сохранения в виде файлов GIF или TIF.

Кластеризация

- По выбору пользователя или автоматический выбор k лучших кластеров.
- Различные стратегии кодирования классовых переменных в процессе анализа.
- Управление недостающими данными.
- Графики профилей переменных сегментов, отражающие распределение исходных данных и других факторов в рамках каждого кластера.
- Профиль дерева решений, использующий исходные данные для составления прогноза о принадлежности кластеру.
- Оценочный код PMML.

Анализ рыночной корзины

- Выявление ассоциаций и причинно-следственных связей:
 - Сетевой график правил, упорядоченный по степени достоверности.
 - Статистические графики подъема, достоверности, прогноза достоверности и поддержки правил.
 - Статистическая гистограмма частотных показателей в заданных границах поддержки и достоверности.
 - График зависимости разброса достоверности от прогнозируемой достоверности.
 - Таблица описания правил.
 - Сетевой график правил.
- Органичная интеграция правил с другими исходными данными обеспечивает расширенное прогностическое моделирование.
- Удобный вывод правил обеспечивает кластеризацию клиентов по их покупательным и поведенческим характеристикам.
- Оценочный код PMML.

Анализ Web-активности

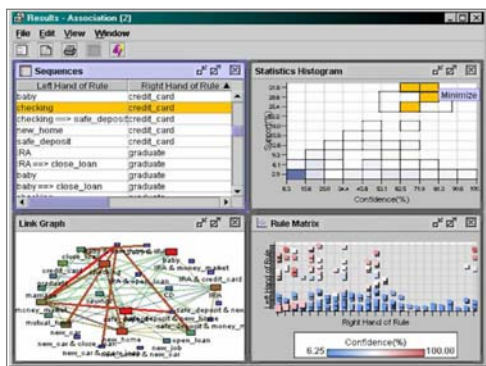
- Масштабируемое и эффективное выявление наиболее популярных интернет-маршрутов на основе анализа данных об интернет-активности пользователей.
- Выявление наиболее частых последовательностей в последовательных данных любого типа.

Уменьшение размерности

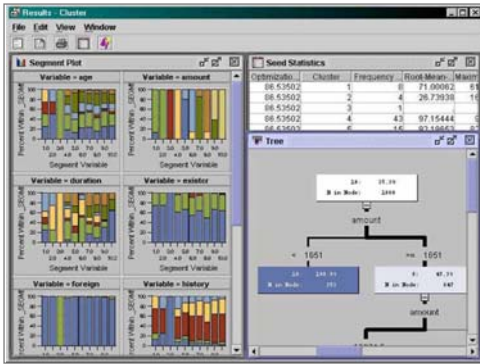
- Выбор переменных:
 - Удаление переменных, не связанных с целевыми признаками, на основе критериев отбора хи-квадрат или R2.
 - Удаление переменных из иерархий.
 - Удаление переменных со многими недостающими значениями.
 - Сокращение числа классовых переменных с большим количеством уровней.
 - Группировка непрерывных исходных данных для выявления нелинейных взаимосвязей.
 - Выявление взаимодействий.
- Главные компоненты:
 - Вычисление собственных значений и собственных векторов на основании матриц корреляции и ковариации.
 - Графики: масштабированное отклонение, логарифмические собственные значения, кумулятивные пропорциональные собственные значения.
 - Исследование выбранных основных компонентов при помощи методов предиктивного моделирования.
- Исследование временных рядов:
 - Сокращение объемов транзакционных данных на основе формирования временных рядов с использованием разнообразных методов аккумуляции и преобразования.
 - Методы анализа включают сезонный анализ, анализ тенденций, анализ временных областей, сезонную декомпозицию.
 - Исследование сокращенных временных рядов при помощи методов кластерного и предиктивного моделирования.
- Управление временными метриками при помощи описательных данных.

Утилита SAS Code Node

- Обеспечивает запись кода SAS для упрощения сложных процедур подготовки и преобразования данных.
- Позволяет использовать процедуры других продуктов SAS.
- Поддерживает импорт внешних моделей.
- Позволяет создавать собственные модели и узлы Enterprise Miner.
- Содержит макропеременные, упрощающие ссылку на источники данных, переменные и т.п.
- Расширяемая логика формирования оценочного кода.



Возможности по анализу профилей рыночных корзин.



Проведение сегментации данных

Исчерпывающие средства моделирования

- Выбор моделей на базе обучающей, проверочной или тестовой выборки данных с использованием различных критериев, таких как: прибыли или убытки, AIC, SBC, среднеквадратичная ошибка, частота ошибок классификации, ROC, Джини, KS (Колмогорова-Смирнова).
- Поддерживает двоичные, номинальные, порядковые и интервальные исходные данные и целевые признаки.
- Удобный доступ к оценочному коду и всем источникам данных.
- Отображение нескольких результатов в одном окне позволяет лучше оценить эффективность модели.

Регрессии

- Линейная и логистическая.
- Пошаговая, с прямой и обратной выборкой.
- Построитель условий для уравнений: полиномиальных, основных взаимодействий, поддержка иерархии эффектов.
- Перекрестная проверка.
- Правила для иерархии эффектов.
- Методы оптимизации: сопряженные градиенты, метод двойных ломаных, метод Ньютона-Рафсона с линейным или гребневым поиском, квазиньютоновский метод, метод доверительных областей.
- Оценочный код PMML.

Деревья решений

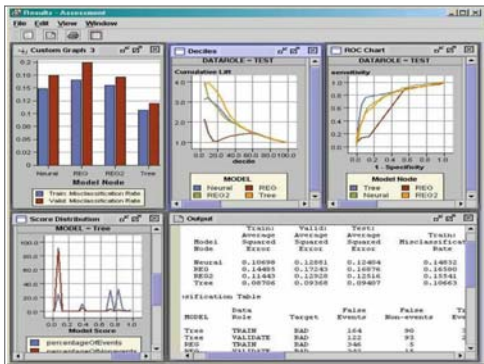
- Общая методология:
 - CHAID (автоматическое выявление взаимодействия по методу хи-квадрат).
 - Деревья классификации и регрессии.
 - C 4.5.
 - Отбор деревьев на основе целевых значений прибыльности или роста с соответствующим отсечением ветвей.
- Критерии расщепления: вероятностный критерий хи-квадрат, вероятностный F-критерий, критерий Джини, критерий энтропии, уменьшение дисперсии.
- Автоматический вывод идентификаторов листьев дерева в качестве входных значений для последующего моделирования.
- Отображение правил на английском языке.
- Вычисление значимости переменных для предварительного отбора.
- Уникальное представление консолидированной диаграммы дерева.
- Интерактивная работа с деревом на настольном ПК:
 - Интерактивное расширение и обрезание деревьев.
 - Задание специальных точек разбиения, включая двоичные или многовариантные разбиения.
- Более 13 динамически связанных таблиц и графиков, позволяющих произвести более качественную оценку дерева.
- Возможность распечатать диаграмму дерева на одном или нескольких листах.
- В основе - новая быстрая процедура ARBORETUM.



Тренировка модели и изучение её поведения.

Нейронные сети

- Узел нейронной сети:
 - Гибкие архитектуры сетей с развитыми функциями комбинирования и активации.
 - 10 методов обучения сети.
 - Предварительная оптимизация.
 - Автоматическая стандартизация входных параметров.
 - Поддержка направленных связей.
- Узел самоорганизующейся нейронной сети:
 - Автоматизированное создание многоуровневых перцептронов для поиска оптимальной конфигурации.
 - Выбор функций типа и активации из четырех различных типов архитектур.
- Оценочный код PMML.
- Узел нейронной сети анализа данных (DM Neural node):
 - Создание модели с уменьшением размерности и выбором функций.
 - Быстрое обучение сети.
 - Линейное и нелинейное оценивание.



Сравнительный анализ нескольких моделей.

Технические требования пакета SAS® Enterprise Miner

Поддерживаемые клиентские платформы

Microsoft Windows (32-разрядная)
 Windows NT 4 Workstation,
 Windows 2000 Professional,
 Windows XP Professional
 AIX (64-разрядная) релиз 5.1
 HP-UX (64-разрядная) релиз 11i (11.11)
 Solaris 8 или 9 (64-разрядная)

Поддерживаемые серверные платформы

Microsoft Windows (32-разрядная, 64-разрядная)
 Windows NT 4 Server 4.0, Windows 2000,
 Windows Server 2003
 AIX (64-разрядная) релиз 5.1
 HP-UX (64-разрядная), релиз 11i (11.11)
 Linux для Intel (32-разрядная)
 Red Hat Linux 8.0, Red Hat Advanced Server 2.1,
 SuSE Linux Enterprise Server 8
 Solaris 8 или 9 (64-разрядная)
 Tru64 UNIX (64-разрядная) Version 5.1A или 5.1B

Минимальная тактовая частота процессора
 1 ГГц

Требования к оперативной памяти
 512 МБ для клиента
 512 МБ для сервера

Требования к объему дисковой памяти
 Не менее 40 МБ для клиента
 3 Гб для сервера (в среднем для Win XP - см. замечания по инсталляции SAS).

Требования к ПО
 Основной пакет SAS, SAS/STAT
 Web-клиент - Java 1.4.1, в противном случае - не нужен (в состав SAS входит частная версия JRE 1.4.1).



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

SAS Institute
 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>

Индуктивный вывод правил

- Метод рекурсивного прогностического моделирования.
- Особенно хорошо подходит для моделирования редких событий.

Двухуровневое моделирование

- Последовательное и параллельное моделирование для классовых и интервальных целевых признаков.
- Выбор модели в виде дерева решений, регрессии или нейронной сети на каждом уровне.
- Управление применением прогноза для классов к прогнозу интервалов.
- Точная оценка экономической выгоды клиентов.

Методы вывода путем сопоставления

- Метод отбора ближайших к соседям для категоризации или прогноза наблюдений.
- Запатентованные методы создания дерева и поиска с уменьшенной размерностью.

Множества моделей

- Объединение прогнозов моделей для создания потенциально более сильного решения.
- Среди методов: усреднение, мажоритарная выборка, выбор максимального значения.

Сравнение моделей

- Сравнение нескольких моделей в рамках единой инструментальной оболочки для всех источников данных.
- Автоматический выбор лучшей модели на основе заданного пользователем критерия.
- Расширенная статистика соответствия и диагностики.
- Ступенчатые диаграммы.
- Кривые ROC.
- Диаграммы прибылей и убытков с возможностью выбора решения.
- Матрица неточностей (классификации).
- График распределения вероятностных оценок классовых целевых признаков.
- Ранжирование и распределение оценок интервальных целевых признаков.

Количественная оценка

- Интерактивная количественная оценка узла в рамках графического интерфейса
- Автоматическая генерация оценочного кода на языках SAS, C, Java и PMML.
- Моделирование сбора, кластеризации, преобразования и вычисления недостающих значений для оценочных кодов на языках SAS, C и Java.
- Развертывание моделей в нескольких средах.

Инструментальные средства

- Узел удаления переменных.
- Узел слияния данных.
- Узел метаданных, позволяющий изменять столбцы метаданных, например, роль, уровень измерений и порядок.