Customer Attrition Analysis in FSI

Анализ вероятности ухода клиентов в финансовом секторе

Ha платформе Microsoft SQL Server 2008 R2 Enterprise Edition

Максим Гончаров Специалист по платформе приложений maxgon@microsoft.com

Данный материал опубликован на сайте http://businessdataanalytics.ru/

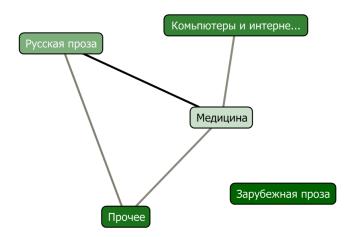
Data Mining в аналитическом CRM

- □ Операционный CRM автоматизировать процессы взаимоотношений, собрать данные, вовремя получать операционную информацию.
- □ Аналитический CRM узнать своих клиентов.
 - Описательный анализ.
 - Какая структура клиентской базы? Какой профиль идеального клиента?
 - Какие есть взаимосвязи между характеристиками клиентов?
 - Какие события происходят одновременно? Какие услуги приобретаются вместе?
 - Какие события наступают последовательно? Откажется ли клиент от сотрудничества при данном шаблоне взаимодействия?
 - Предиктивный анализ.
 - Откликнется ли клиент на данную маркетинговую кампанию?
 - Какова ценность клиента с данными характеристиками?
 - Какой размер прибыли будет в следующем месяце?
 - Какие из потенциальных клиентов вероятно совершат приобретение услуги в следующем месяце?

Кластерный анализ

Сегментация – обнаружение структуры данных

- На какие группы можно разбить клиентов?
- Какая структура продаж?
- Какие характеристики у мошеннических транзакций?

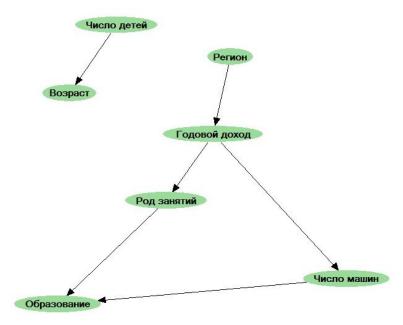


Разбиение постоянных клиентов книжного интернет-магазина по интересам.

Анализ цепочек влияния

Байесовские сети – обнаружение цепочек влияния факторов

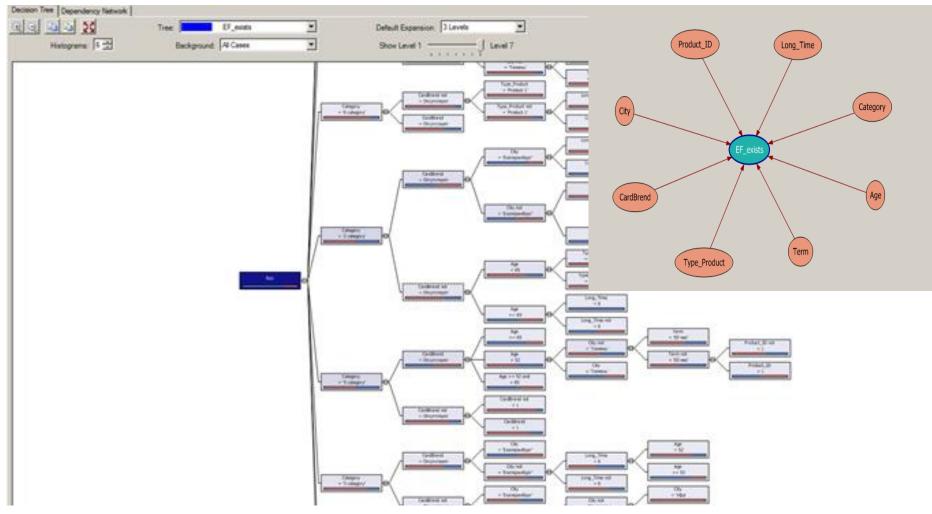
- Как связаны образование, доход и регион проживания?
- Как влияет семейное положение на покупку товара?



Column ▼	Value ▼	Favors Yes	▼ Favors No	₩.
Cars	2			
Cars	o			
Marital Status	Single			
Marital Status	Married			
Cars	1			
Region	Pacific			
Region	North America			

Цепочки влияния характеристик клиента

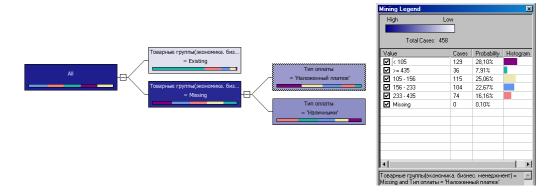
Факторы влияния



Что влияет на приобретение банковского продукта

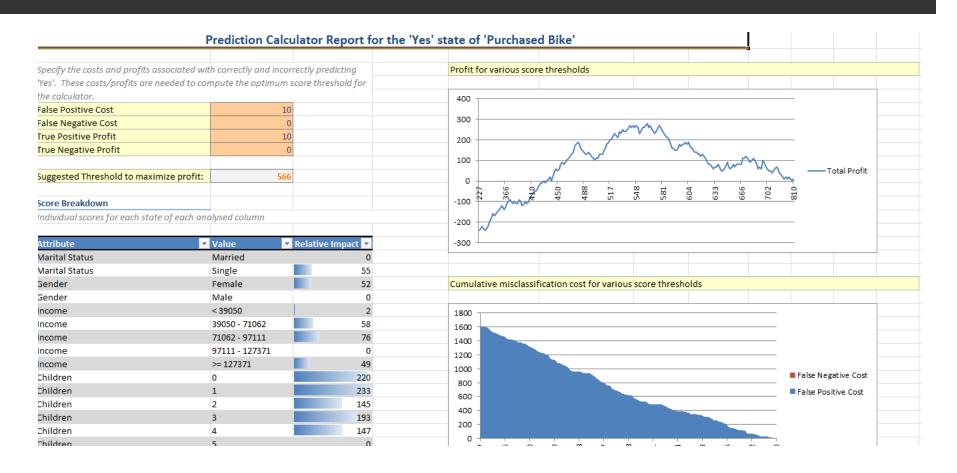
Классификация

- Откликнется ли клиент с данными характеристиками на наше предложение?
- Что характерно для клиента, собирающегося отказаться от услуг нашей компании?
- Скоринг, приоритезация лидов.



Дерево решения: что влияет на размер заказа продажи?

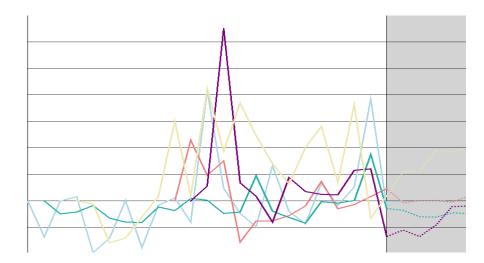
Кредитный скоринг



Скоринговая модель

Прогнозирование

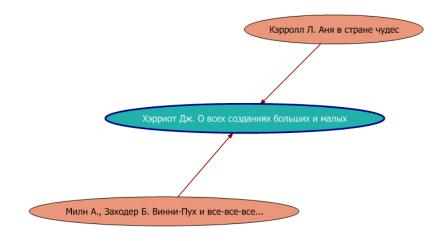
- Прогноз продаж по товарным категориям и группам клиентов.
- Прогноз обращений в колл-центр.



Прогноз продаж банковских услуг по категориям

Кросс-продажи

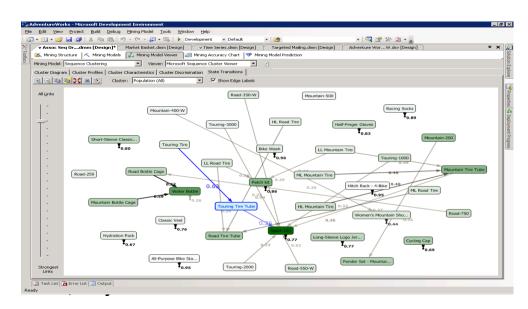
- Какие товары и услуги продаются совместно?
- Cross-sales, up-sales.
- Какие события происходят вместе?



Книги, покупаемые совместно

Цепочки последовательностей

- Приобретение каких продуктов следует ожидать в следующем месяце от человека, открывшего в этом месяце счет и получившего дебетовую карту?
- С какой вероятностью от нас уйдет клиент при данной последовательности действий?



Временные шаблоны покупок велосипедных аксессуаров

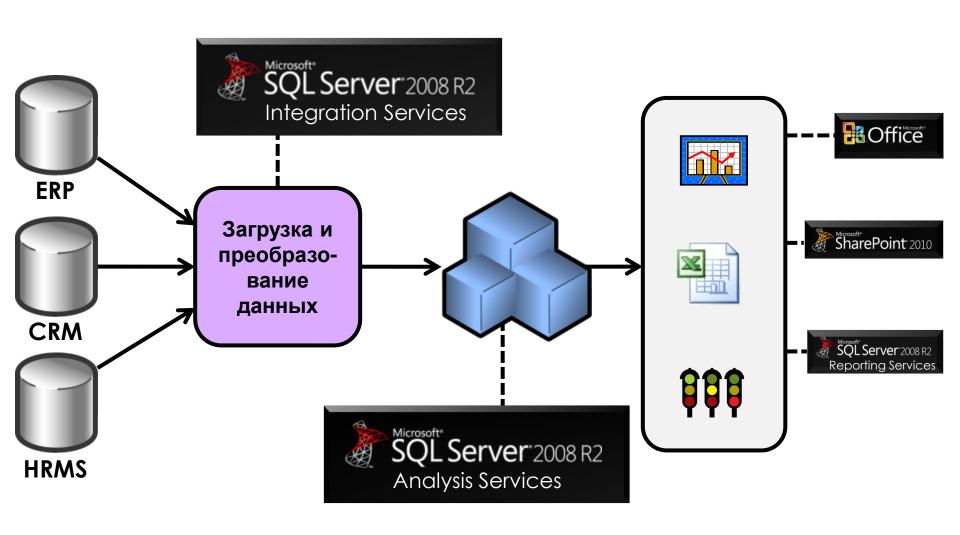
Бизнес-задачи

Ответы, которые позволяет получить Data Mining интересны не сами по себе, а как инструмент решения бизнес-задач.

Н	аг	ın	И	М	e	n	•
	uı	· M		IAI	J	Μ	

- Какие продукты надо предлагать данному клиенту?
- Какова вероятность того, что данный сектор потенциальных клиентов отреагирует на рекламную кампанию?
- Каков размер поступления оплат в будущем месяце?
- Как велик риск выдачи кредита данному клиенту банка?
- Какие группы услуг надо предлагать совместно?
- □ Какая вероятность ухода клиента? (Customer Attrition/Churn)

Общая схема BI-решения



Состав решений Microsoft BI





DATA INFRASTRUCTURE & BUSINESS INTELLIGENCE PLATFORM

Microsoft® SQL Server® 2008 R2

Инфраструктура данных и ВІ-платформа

- Analysis Services
- Reporting Services
- Master Data Services
- Integration Services
- Data Mining
- Data Warehousing
- StreamInsight

Состав решений Microsoft BI







Инфраструктура эффективности бизнеса

- Панели индикаторов и карты показателей
- Excel Services
- Web-формы и процессы
- Совместная работа
- Поиск
- Управление контентом
- Интеграция с приложениями
- PowerPivot для SharePoint

Состав решений Microsoft BI







Предоставление через знакомый интерфейс

- Самостоятельный доступ к информации
- Исследование и анализ данных
- Интеллектуальный анализ данных
- Визуализация данных
- Контекстная визуализация

Customer Attrition – цели

- Сохранение клиентов дешевле приобретения новых.
- Новые клиенты рискованные, склонны к смене поставщика, выгодные старые клиенты – лояльные, проверенные.
- Выборочно сохранять выгодных клиентов актуально в условиях насыщенных рынков, когда каждый новый клиент – переманенный от конкурента.
- Повышение прогнозируемости и стабильности поступлений.

Customer Attrition - анализируемые факторы

- Демографические данные
 - **♦** По∧
 - ❖ Возраст
 - ❖ Доход
 - **...**
- □ Поведение
 - Приобретаемые услуги
 - Прибыльность, динамика прибыльности,
 - Число обращений в колл-центр, тип проблем.
 - RFM, ABC, XYZ,...
- Макроэкономические показатели
 - Состояние, экономики, отрасли
 - **...**

Customer Attrition - анализируемые факторы

Predictors			1	Non-
	Quantitative	Qualitative	Time-varying	
Customer Behaviour Predictor Category				
Product-specific ownership			l	
Current account		x	x	
Savings account		x	x	
Term deposits and pension savings		x	x	
Mortgage		×	x	
Personal loan and tax credit		x	x	
Mutual funds		×	x	
Shares and bonds		×	x	
Total product ownership	x		x	
Card-specific ownership			l	
Off-line Debit Card		×	x	
Credit Card		×	x	
On-line Debit Card		×	x	
Online Debit /Eurocheque Card		×	x	
Total card ownership	x		×	
Interpurchase time	x		×	
Monetary value	x		×	
Homebanking		×	×	
Phonebanking		×	×	
Customer Demographics Predictor Category				
Age	x		x	
Gender		×	l	×
Education level	×		l	x
Social status code			l	
Elite		×	l	x
High status		×	l	x
Medium status		x	l	x
Social potential code	×			×
Macro Environment Predictor Category				
Merger		×	l x	
Prosperity index	x		l â	

Customer Attrition – поведенческие характеристики

- □ Обладание конкретным продуктом (да/нет)
- □ Общее число банковских продуктов (новый продукт уменьшает вероятность ухода до <1%)</p>
- Среднее время между покупками (увеличение времени ведет к повышению вероятности ухода)
- Использование новых каналов взаимодействия с банком: интернет-банкинг, мобильный банк и т.д. (частое взаимодействие с банком уменьшает вероятность ухода)
- Использование новых услуг: категоризация расходов, advisory services, personal budgeting и т.д.

Customer Attrition – демографические характеристики

- Возраст (старшие уходят реже)По∧ (мужчины уходят чаще)
- Образование (более образованные реже уходят)
- □ Род занятий
- Доход
- Тип занимаемой должности
- География
- □ Число детей
- □ Недвижимость
- Автомобили
- Уровень социального статуса (функция от дохода, образования, сферы деятельности, должности и т.д.) более статусные реже уходят
- Уровень социального потенциала

Customer Attrition – макроэкономические характеристики

- Макроэкономические индексы качества жизни (чаще уходят при хорошей экономической ситуации)
- Макроэкономические индексы ощущаемой стабильности
- Индексы удовлетворенности обслуживанием (анкетирование, маркетинговые исследования)
- Изменения законодательства и регулирования
- Организационные изменения (слияния, IPO) изменения ведут к повышению вероятности ухода

Категории моделей

- □ Описательные модели
 - Кластеризация. Какие существуют группы клиентов?
 - ❖ Байесовские сети. Какие взаимосвязи между атрибутами существуют?
 - ❖ Ассоциативный анализ. Какие характеристики встречаются совместно?
 - ❖ Временные ассоциации. Какие события предшествуют уходу?
- 🔲 Предиктивные модели
 - Какова вероятность ухода группы клиентов в течение периода?
 - Деревья решений
 - Нейронные сети
 - Логистическая регрессия
 - **>** ...

Тип предиктивных моделей

Анализ ухода клиентов – бинарная классификация (клиент уйдет?).

- Статическая классификация. Анализ производится без учета изменений во времени. Одна классификационная модель строится на основании усредненных за прошлый анализируемый период характеристик ушедших в течение следующего периода клиентов и оставшихся.
- Динамическая классификация. Клиенты анализируется в течение длительного времени при помощи скользящего окна. Строятся несколько классификационных моделей для каждого значения времени дожития (для клиентов, не ушедших в течение 1 года, двух лет и т.д.)

Survival Analysis

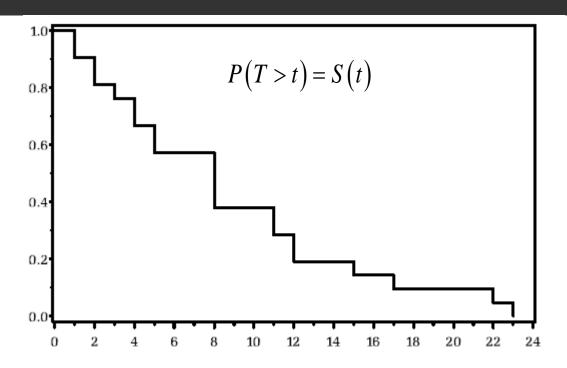
Анализ дожития (survival). Динамика вероятности того,
 что клиент останется в течение следующего периода.

Все клиенты смещаются по временной оси X в точку 0. По оси Y откладывается часть клиентов, оставшихся к этому времени.

 Анализ опасности (hazard). Динамика вероятности ухода в следующем периоде при условии, что клиент остался до текущего.

Все клиенты смещаются по временной оси X в точку 0. По оси Y откладывается часть оставшихся к этому времени клиентов, ушедших в следующем периоде.

Survival Function



T – случайная величина, описывающая время ухода клиента, S(t) – вероятность дожития до времени t.

Функция дожития – вероятность того, что клиент останется после момента времени †, монотонно убывающая функция.

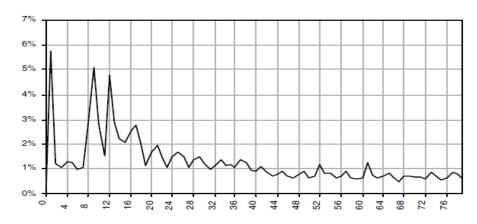
Hazard Function

Вероятность ухода клиента в течении времени $(t,t+\Delta t]$ при условии дожития до времени t.

$$P(t < T \le t + \Delta t | T > t) = \frac{P(t < T \le t + \Delta t, T > t)}{P(T > t)} = \frac{P(t < T \le t + \Delta t)}{P(T > t)} = \frac{F(t + \Delta t) - F(t)}{S(t)}$$

Функция опасности (hazard function) – производная, плотность вероятности ухода клиента в момент времени † при условии дожития до этого времени.

$$h(t) = \lim_{\Delta t \to 0} \frac{P(t < T \le t + \Delta t \mid T > t)}{\Delta t} \frac{1}{S(t)} \lim_{\Delta t \to 0} \frac{F(t + \Delta t) - F(t)}{\Delta t} = \frac{1}{S(t)} F'(t) = \frac{f(t)}{S(t)}$$



Для большинства европейских банков существуют 2 критических периода ухода: в течение 1 года и через 20 лет.

Expected future lifetime

Среднее время, оставшееся до ухода

Ожидаемое время, оставшееся до ухода клиента при условии его дожития до времени t_0 :

$$T(t_{0}) = E(T - t_{0} | T > t_{0}) = \frac{1}{P(T > t_{0})} \int (T(\omega) - t_{0}) 1_{(t_{0}, \infty]} \circ T(\omega) dP(\omega) =$$

$$= \frac{1}{S(t_{0})} \int \left(\int_{t_{0}}^{T(\omega)} dt \right) 1_{(t_{0}, \infty]} \circ T(\omega) dP(\omega) = \frac{1}{S(t_{0})} \int \left(\int_{(t_{0}, T(\omega))}^{T(\omega)} \circ t dt \right) 1_{(t_{0}, \infty)} \circ T(\omega) dP(\omega) =$$

$$= \frac{1}{S(t_{0})} \int \int (1_{(t, \infty)} \circ T(\omega)) (1_{(t_{0}, \infty)} \circ T(\omega)) (1_{(t_{0}, \infty)} \circ t) dt dP(\omega) =$$

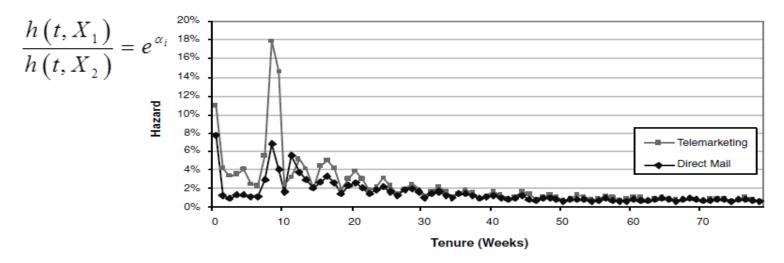
$$= \frac{1}{S(t_{0})} \int \int (1_{(t, \infty)} \circ T(\omega)) (1_{(t_{0}, \infty)} \circ t) dt dP(\omega) = \frac{1}{S(t_{0})} \int_{t_{0}}^{\infty} \left(\int_{t_{0}}^{T(\omega)} \int_{t_{0}}^{\infty} T(\omega) dP(\omega) dT(\omega) dT(\omega)$$

Регрессия Кокса

Зависимость функции опасности от характеристик клиента

$$\log h(t,X) = \alpha_0(t) + \sum_{i=1}^n \alpha_i x_i \Rightarrow h(t,X) = h_0(t) \prod_{i=1}^n e^{\alpha_i x_i}$$

Функция опасности описывается как некая базовая функция опасности h_0 (†), зависящая от времени, промодулированная характеристиками клиента $x_1, ..., x_n$. Если два клиента отличаются только характеристикой і, т.е. у одного она есть (1), а у другого нет (0), то отношение функций опасности будет равно константе.



Учет изменяющихся факторов

Наряду с постоянными характеристиками клиента (пол, раса), существуют характеристики, меняющиеся во времени. Их можно учитывать следующими методами:

- Вводить в модель характеристики, соответствующие прошлым периодам. Например, «количество банковских продуктов в текущий момент», «количество банковских продуктов в прошлом месяце» и т.д.
- Ввести в модель Кокса характеристики Х(t) как функции от времени и учитывать их влияние при помощи значений за прошлый период t-1:

$$h(t,X(t)) = h_0(t) \prod_{i=1}^{n} e^{\alpha_i x_i(t-1)}$$

Оценка параметров в регрессии Кокса

Параметры модели оцениваются методом максимального правдоподобия: функция правдоподобия данных – произведение плотности вероятности даты фактического ухода для клиентов, ушедших в течение рассматриваемого периода, и вероятности того, что время ухода оставшихся клиентов больше анализируемого периода:

$$\begin{split} L\left(X\mid\alpha\right) &= \left(\prod_{\text{i kinert ymer b moment }t_i} \underbrace{f\left(t_i\mid x_i,\alpha\right)}_{f = hS}\right) \left(\prod_{\text{i kinert ymer}} S\left(t_i\mid x_i,\alpha\right)\right) = \\ &= \prod_{i=1}^n \left[h\left(t_i\mid x_i,\alpha\right)S\left(t_i\mid x_i,\alpha\right)\right]^{\frac{1}{i} \text{(i kinertymer)}} S\left(t_i\mid x_i,\alpha\right)^{1-1} \text{(i kinertymer)} = \\ &= \prod_{i=1}^n \left[h\left(t_i\mid x_i,\alpha\right)\right]^{\frac{1}{i} \text{(i kinertymer)}} S\left(t_i\mid x_i,\alpha\right) \\ \text{УЧИТЫВОЯ } h &= -\left(\log S\right)' \Rightarrow S = \exp\left(-\int h\right) \text{ , polyyorm} \\ L(X|\alpha) &= \prod_{i=1}^n \left[h\left(t_i\mid x_i,\alpha\right)\right]^{\frac{1}{i} \text{(i kinertymer)}} \exp\left(-\int h\left(s\mid x_i,\alpha\right)ds\right) \end{split}$$

Параметры lpha получаем максимизацией этого выражения

Литература

- Michael J.A. Berry, Gordon S. Linoff, «Data Mining Techniques For Marketing, Sales, and Customer Relationship Management»
- Dirk Van den Poel, Bart Larivière, «Customer Attrition Analysis For Financial Services Using Proportional Hazard Models»,
- ☐ John P. Klein, «Survival Analysis Techniques for Censored and Truncated Data»
- John Fox, «Introduction to Survival Analysis»

Спасибо!

Максим Гончаров
Microsoft
Специалист по платформе приложений
maxgon@microsoft.com
+7 (903) 116-86-03