

SAS Institute Inc

SAS Global Business Knowledge Series (BKS)

Дополнительные продвинутые курсы по
технологиям SAS

Для специалистов в области аналитики



10

Уважаемые клиенты и партнеры компании SAS!

Позвольте представить Вам новую на нашем рынке услугу SAS Global Business Knowledge Series. Курсы, входящие в эту программу, объединены под эгидой дополнительных «продвинутых» знаний в определенных специализациях. Они не заменяют собой регулярных курсов, но дополняют их знаниями по специфике применения в бизнес-сфере. Общий список курсов размещен на второй странице этой брошюры. Любой из предложенных здесь курсов можно провести в корпоративном формате, отправив заявку минимум за три недели до желательной даты проведения. Описание курсов размещено на глобальном сайте www.sas.com.

В данной брошюре вы найдете описание тех курсов, которые российский офис SAS собирается провести в формате «открытого» обучения, т.е. куда могут быть собраны представители из различных организаций.

Расписание по данным курсам вы сможете получить дополнительным письмом или увидеть на нашем сайте www.sas.com/russia.

Мы надеемся, что курсы из серии Business Knowledge будут актуальны и полезны для ваших организаций и приведут к росту компетенций ваших специалистов в области аналитики. Уверенны, что следование принципу SAS «The power to know», принесет ощутимую выгоду и приведет ваши организации к лидерству в своих отраслях экономики.

С надеждой на сотрудничество,

Руководитель Учебного центра SAS в России и СНГ

Ольга Стасевич

Общий список курсов из Business Knowledge Series

BKS курс	инструктор	Длительность курса
Data Miner		
Advanced Analytics for Customer Intelligence Using SAS	Bart Baesens, PhD or Christophe Mues, PhD	3 дня
Applying Survival Analysis for Business Time-to-Event Problems	Michael Berry or Gordon Linoff	2 дня
Customer Segmentation Using SAS Enterprise Miner	Goutam Chakraborty, Ph.D.	2 дня
Data Mining Techniques: Theory and Practice	Michael Berry or Gordon Linoff	3 дня
Exploratory Analysis for Large and Complex Problems	Jeff Zeanah	2 дня
Exploratory Data Mining with Application to Life and Social Sciences	Patricia Cerrito, Ph.D.	3 дня
Multiple Comparisons in SAS: Classical to Modern Methods	Peter Westfall, PhD	3 дня
Survival Data Mining: Predictive Hazard Modeling for Customer History Data	Bob Lucas, PhD	3 дня
The Art and Science of Insurance Fraud Detection	Terry Woodfield, PhD	½ дня Live Web
Web Analytics and Web Intelligence Using SAS	Bart Baesens, PhD or Christophe Mues, PhD	2 дня
Market Researcher		
Conjoint Analysis: Evaluating Consumer Preferences Using SAS Software	Mike Patetta	½ дня Live Web
Discrete Choice Modeling Using SAS Software	Mike Patetta	2 дня Live Web
Getting the Most Out of Testing in Direct/Internet Marketing	Goutam Chakraborty, PhD	2 дня Live Web
Graphical Techniques for Marketing Research	Mike Patetta	½ дня Live Web
Data Manager		
Data Cleaning Techniques	Ron Cody, M.D.	2 дня
Health Care Data and the SAS System (This course has a strong US centric focus and may not be appropriate outside of North America)	Craig Dickstein	2 дня
SAS Functions by Example	Ron Cody, M.D.	2 дня
Statistical Analyst		

SAS Global Business Knowledge Series (BKS)

Advanced Analytics Using the MIXED Procedure	F. Michael Speed, PhD	2 дня
Conjoint Analysis: Evaluating Consumer Preferences Using SAS Software	Mike Patetta	½ дня Live Web
Discrete Choice Modeling Using SAS Software	Mike Patetta	2 дня Live Web
Fitting Tobit and Other Limited Dependent Variable Models	Mike Patetta	½ дня Live Web
Getting the Most Out of Testing in Direct/Internet Marketing	Goutam Chakraborty, PhD	2 дня Live Web
Imputation Techniques in SAS	Patricia Berglund	½ дня Live Web
Introduction to Statistical Concepts: Means, Standard Deviations, Confidence Intervals and Hypotheses Testing	F. Michael Speed, PhD	½ дня Live Web
Introduction to Structural Equation Modeling	Werner Wothke, PhD	½ дня Live Web
Modeling Trend, Cycles, and Seasonality in Time Series Data Using PROC UCM	Rajesh Selukar, PhD	½ дня Live Web
Modern Multiple Comparisons and Multiple Tests Using SAS	Peter Westfall, PhD	½ д дня Live Web
Multilevel Modeling of Hierarchical and Longitudinal Data Using SAS	Catherine Truxillo, PhD or Chris Daman	3 дня
Robust Regression Techniques in SAS/STAT	Mike Patetta, PhD	½ дня Live Web
Operations Researcher		
Custom Design for Experiments (Using JMP)	Mark Bailey	½ дня Live Web
Forecaster		
Introduction to Applied Econometrics	Oral Capps, PhD	3 дня
Stationarity Testing and Other Time Series Topics	Dave Dickey, PhD	½ дня Live Web
Risk Management		
Credit Risk Modeling Using SAS	Bart Baesens, PhD or Christophe Mues, PhD	4 дня
Credit Scorecard Development and Implementation	Naeem Siddiqi	2 дня
Data Integration and ETL		
Warehouse Architecture and Design Principles	Steve Morton	3 дня
Customer Intelligence		
Advanced Analytics for Customer Intelligence Using SAS	Bart Baesens, PhD or Christophe Mues, PhD	3 дня
Planning and Interpreting Marketing Campaigns	Tony Woods	2 дня
Propensity Scoring Models Using the SAS System	Tony Woods	2 дня
Segmentation of a Customer Database	Tony Woods	2 дня
Project Manager		
Maximizing Technology and Resource Investment with a BI Competency Center	Aiman Zeid	1 дня

Курсы, представленные в расписании Российского учебного центра SAS

	название	Длительность курса	Стоимость (предварительная)	Дата в расписании (предварительная)
Data Miner				
1	Customer Segmentation Using SAS Enterprise Miner	2 дня	1000 евро	2-3 сентября
2	Data Mining Techniques: Theory and Practice.	3 дня	1700 евро	6-8 сентября
Risk Management				
3	Credit Risk Modeling Using SAS	4 дня	2000 евро	24-27 августа
4	Credit Scorecard Development and Implementation	2 дня	1000 евро	26-27 октября
Customer Intelligence				
5	Advanced Analytics for Customer Intelligence Using SAS	3 дня	1500 евро	1-3 ноября

Содержание

Общий список курсов из Business Knowledge Series	3
Курсы, представленные в расписании Российского учебного центра SAS.....	5
Содержание	6
1. Сегментация клиентов при использовании SAS Data Miner (Customer Segmentation Using SAS Enterprise Miner)	8
Goals:	8
Предварительные требования:.....	8
Аудитория:.....	8
Содержание курса:	8
Материалы курса:	9
2. Методы извлечения данных: теория и практика (Data Mining Techniques: Theory and Practice).10	
Цели курса:.....	10
Аудитория:.....	10
Предварительные требования:.....	10
Описание курса:	10
Материалы курса:	13
3. Моделирование кредитных рисков с использованием технологий SAS.....	14
Цели курса:.....	14
Аудитория:.....	14
Предварительные требования:.....	14
Описание курса:	14
Материалы курса:	17
4. Разработка и применение Credit Scorecard (Credit Scorecard Development and Implementation)....	18
Цели курса:.....	18
Аудитория:.....	18
Предварительная подготовка:.....	18

SAS Global Business Knowledge Series (BKS)

Описание программы курса:	18
Материалы курса:	19
5. Продвинутая аналитика для Customer Intelligence при использовании SAS (Advanced Analytics for Customer Intelligence Using SAS)	20
Цели курса:.....	20
Аудитория:.....	20
Предварительные требования:.....	20
Описание курса:	21
Материалы курса:	23

1. Сегментация клиентов при использовании SAS Data Miner (Customer Segmentation Using SAS Enterprise Miner)

Длительность курса - 2 дня.

Курс проводится профессором по маркетингу Государственного Университета штата Оклахома (Oklahoma State University, США), [Goutam Chakraborty, Ph.D.](#),

Этот практический курс посвящен проблемам сегментного анализа рынка в рамках business data mining, по окончании которого вы получите не только теоретические знания, но и практический опыт. Разделы курса включают теорию сегментации, а также сведения о четырех основных аналитических инструментах сегментации: иерархической кластеризации, K-means кластеризации, методе RFM cell и методе SOM/Kohonen.

Цели:

- Понять и научиться применять оценочные и поведенческие инструменты и методы сегментации в работе с клиентскими данными
- Научиться использовать как дескриптивную, так и прогнозную сегментацию. Научиться описывать и обосновывать сегменты
- Научиться оценивать стабильность сегментов в динамике по времени
- Научиться задавать вероятность сегментации к новым данным

Предварительные требования:

Знания по технологиям SAS будут желательны, но не обязательны. Не требуются навыки работы с SAS Enterprise Miner или SAS Enterprise Guide

Аудитория:

Курс предназначен для всех тех, кто хочет научиться выявлять значимые сегменты на основе имеющихся клиентских данных, фокусируясь на решении практических бизнес задач, а также бизнес-аналитикам, менеджерам, специалистам по маркетингу, программистам и другим, желающим почерпнуть полезные знания и навыки данного курса.

Содержание курса:

Введение

Определение сегментации клиентов
Клиентская сегментация в рамках бизнес задач

Основы сегментации

- базисы сегментации
- признаки сегментации
- методы сегментации

Иерархическая кластеризация для сегментации

- введение в кластерный анализ
- сходства, различия, метрики различий
- типы кластеризации

Приложения для иерархической кластеризации

- поиск сегментов по данным опроса клиентов сектора B2B
- изменение сегментов
- определение профиля сегментов
- применение метода Ward's для поиска сегментов по данным опроса клиентов сектора B2B
- валидация сегментов с признаками и другими важными управленческими переменными

k-Means Clustering

- mechanics of k-means clustering
- приложения для k-means clustering
- скоринг новых данных
- оценка стабильности кластерных решений в динамике по времени
- вероятность кластерной принадлежности

Априорная сегментация с использованием RFM

- применение RFM cell-based сегментации
- кластеры сходные по продуктам (самостоятельное изучение)
- применение сходства по продуктам в сегментации

Выбор переменных, SOM/Kohonen, Wrap-up, and Take-aways

- выбор переменных для сегментации
- SOM/Kohonen for segmentation
- wrap-up and take-aways

Материалы курса:

Слушатели получают печатные материалы курса

2. Методы извлечения данных: теория и практика (Data Mining Techniques: Theory and Practice)

Длительность курса - 3 дня.

Данный курс проводится основателями Data Miners Inc. и соавторами книги «*Data Mining Techniques and Mastering Data Mining*» [Michael J. A. Berry](#) или [Gordon S. Linoff](#)

Этот курс познакомит вас с методологией data mining, расширенным вариантом методологии SAS SEMMA, на основе которого создан SAS Enterprise Miner. На этом курсе вы также узнаете о широкой линейке алгоритмов data mining и получите не только теоретические знания, но и практические навыки и умения. В рамках данного курса вы пройдете все стадии проекта data mining, начиная с определения проблемы и выборки данных, и заканчивая исследованием данных, трансформацией данных, определением подвыборок, portioning, моделированием и оценкой.

Цели курса:

- научиться использовать технологии data mining
- строить и использовать дерево решений и нейронные сети для целей моделирования и скоринга
- использовать методы анализа выживания и создавать кривые выживания

Аудитория:

Бизнес-аналитики, руководители бизнес-аналитиков, статистики.

Предварительные требования:

Не требуются знания и навыки работы с инструментами (statistical or data mining).

Описание курса:

Введение в Data Mining

что такое data mining?
прямой and не прямой data mining
модели
создание профилей и прогнозирование

Методология Data Mining

- зачем нужна методология?
- как специалисты по data mining неосторожно узнают то, что не является правдой
- перевод бизнес проблем в проблемы data mining
- важность стабильности моделей
- поиск правильных входных переменных

- создание подвыборок с целью определения сбалансированных модельных наборов
- разбиение с целью создания обучающих, валидационных и тестовых модельных наборов
- подготовка данных
- оценка моделей

Исследование данных

- разработка прогнозов касательно данных
- структура данных
- типы данных
- значения данных
- исследование дистрибуций
- сводная статистика
- гистограмма
- использование SAS Enterprise Miner для изучения данных

Регрессионная модель

- тестирование гипотезы на значимость переменных
- статистический уровень значимости
- доверительный предел
- вариативность и стандартные отклонения
- стандартизированные значения
- корреляции
- линейная регрессия
- логистическая регрессия
- использование SAS Enterprise Miner для построения регрессионных моделей

Деревья решений

- деревья решений как инструмент классификации и исследования данных
- деревья решений для моделирования и скоринга
- деревья решения для выборки переменных
- альтернативное представление деревьев решений
- алгоритмы для построения дерева решений
- критерии разбиения
- распознавание нестабильности и чрезмерной подгонки в моделях дерева решений
- нахождение взаимосвязей между переменными

- использование SAS Enterprise Miner для построения дерева решений

Нейронные сети

- происхождение нейронных сетей
- нейронные сети в сравнении с регрессией
- алгоритмы для формирования нейронных сетей
- требования к подготовке данных для построения нейронных сетей
- подбор соответствующих вводимых данных для нейронных сетей
- создание моделей нейронных сетей с использованием SAS Enterprise Miner

Вывод путем сопоставления

- сходства и различия
- метрики, предназначенные для различных видов данных
- роль training set для вывода путем сопоставлений (MBR)
- объединение мажоритарной выборки нескольких соседних записей
- other K-nearest neighbor techniques
- collaborative filtering
- использование узла SAS Enterprise Miner MBR

Кластеризация

- подробнее о сходствах и различиях
- алгоритм k -means
- аналитическая кластеризация
- агломеративная кластеризация
- подготовка данных к кластеризации
- интерпретация кластеров
- поиск кластеров с помощью SAS Enterprise Miner

Анализ выживаемости

- происхождение анализа выживаемости
- чем отличаются бизнес данные от клинических данных
- риски и графики рисков
- кривые удержания и выживания
- расчет survival после удержания
- вычисление рисков эмпирическим путем
- parametric hazard models параметризованные модели рисков
- оценка (censoring)

- конкурирующие риски
- прогнозирование, основанное на survival
- использование кода SAS в SAS Enterprise Miner для создания кривых выживания

Правила ассоциаций

- анализ «рыночной корзины»
- правила ассоциаций
- анализ последовательных моделей
- использование SAS Enterprise Miner для поиска ассоциаций в розничных данных

Анализ ссылок

- история теории графов
- сфера применения
- использование link analysis для получения производных переменных
- алгоритм graph-coloring
- алгоритм Kleinberg's

Генетические алгоритмы (Genetic Algorithms)

- техники и проблемы оптимизации (SAS/OR software)
- прочие алгоритмы
- проблемы линейного программирования
- генетические алгоритмы (genetic algorithms)

Материалы курса:

Слушатели получают печатные материалы курса, а также копию книги «*Data Mining Techniques: For Marketing, Sales, and Customer Relationship Management, Second Edition*».

3. Моделирование кредитных рисков с использованием технологий SAS

Длительность 4 дня

Данный курс проводится докторами философских наук, помощниками профессора Школы Управления в Университете Саутгемптона (Великобритания), [Bart Baesens, Ph.D.](#) либо [Christophe Mues, Ph.D.](#)

На этом курсе слушатели научатся разрабатывать модели кредитных рисков в рамках требований Basel II. Курс изобилует не только теоретическими и техническими тонкостями, но также и практическими деталями внедрения, что показано на примерах реальных case studies и упражнениях.

Цели курса:

- разработать модели вероятности дефолта (PD), потери в случае дефолта (LGD) и величину, подверженную риску в случае дефолта (EAD)
- провести валидацию, бэк-тестирование и бенчмаркет моделей кредитных рисков
- выполнить стресс-тестирование моделей кредитных рисков
- разработать модели кредитных рисков для портфелей с низкой вероятностью дефолта
- научиться использовать продвинутые технологии для улучшенного моделирования кредитных рисков

Аудитория:

Курс предназначен для любых специалистов, вовлеченных в построение моделей кредитных рисков или ответственных за отслеживание работы моделей кредитных рисков.

Предварительные требования:

Опыт работы с кредитными рисками и начальный уровень знаний методов статистической классификации. Работа с программными продуктами SAS, SAS Enterprise Miner приветствуется, но является необязательной.

Описание курса:

Обзор Basel I и Basel II

- прикладной скоринг, поведенческий скоринг и скоринг дохода
- прогнозные модели банкротства
- кредитные рейтинги
- правила Basel I и Basel II
- стандартный подход vs. подход IRB к кредитным рискам
- PD vs. LGD vs. EAD
- Ожидаемые потери vs. непредвиденные потери

Создание выборок и обработка данных

- определение выборки
- типы переменных
- отсутствующие значения
- нормализация популяции
- разведочный анализ данных
- категоризация
- weight of evidence coding and information value
- сегментация
- reject inference (hard cut-off augmentation, parceling, etc.)

Разработка PD моделей для Basel II

- основные концепции классификации
- методы классификации: логистическая регрессия, дерево принятия решений,
- линейное программирование, k-nearest neighbor, коммулятивная логистическая регрессия,
- выборка входных данных путем фильтрации, пошаговой регрессии, p-values (strategy curve, marginal good-bad rates)
- установка (strategy curve, marginal good-bad rates)
- оценка эффективности скоринговых карт
- разбивка данных: однократная выборка, holdout sample, перекрестная проверка
- метрики эффективности, такие, как ROC curve, CAP curve, and KS-statistic
- разделение данных: single sample, holdout sample, cross-validation performance metrics, such as ROC curve, CAP curve, and KS-statistic defining ratings
- выравнивание скоринговых карт и их внедрение

Разработка LGD и моделей EAD для Basel II

- моделирование потерь в случае дефолта (LGD)
- определение LGD путем рыночного подхода и work-out approach
- взвешенный по времени vs. взвешенный дефолтом vs. exposure weighted LGD
- выбор коэффициента скидки и периода урегулирования проблем
- работа с неполной задолженностью (incomplete workouts)
- LGD, обусловленные экономическим спадом
- моделирование LGD с использованием сегментации
- моделирование LGD с использованием регрессии

- формирование Бета-распределения (Beta distribution) для LGD
- моделирование LGD с помощью двухступенчатой модели
- моделирование величины подверженной риску в случае дефолта (EAD): оценка коэффициента перерасчета кредита (credit conversion factors) (CCF)
- cohort/fixed time horizon/momentum approach for CCF
- причины риска (risk drivers) для CCF
- CAP curves for LGD and CCF
- связь между PD, LGD, and EAD
- расчет ожидаемых потерь (expected loss (EL))

Проверка на соответствие (validating), контрольное тестирование (backtesting) и стресс-тестирование

- проверка на соответствие PD, LGD, и EAD моделей
- количественная vs. качественной проверки на соответствие
- контрольное тестирование (back testing) для PD, LGD, and EAD
- контрольное тестирование на стабильность модели (показатель стабильности системы)
- backtesting model discrimination
- backtesting model calibration using the binomial, Vasicek, and chi-squared tests
- traffic light indicator approach
- планирование контрольных тестирований (backtesting action plans) through-the-cycle (TTC) versus point-in-time (PIT) validation
- бенчмаркинг внутренний vs. внешний бенчмаркинг
- Kendall's tau and Kruskal's gamma for benchmarking
- использование тестирования
- качество данных
- документация
- корпоративное управление и управленческий контроль
- стресс-тестирование моделей на PD, LGD, и EAD
- статическое vs. динамическое стресс-тестирование
- сопоставительный анализ изменений
- управление скоринговыми картами
- low default portfolios (LDPs): внедрение и контрольное тестирование
- вероятностные методы (likelihood approaches) LDPs
- rating mapping approaches to LDPs

Новые методики разработки PD, LGD, и EAD моделей для Basel II

- обзор традиционных методик разработки скоринговых карт

- нейронные связи: нейронная модель the neuron model, многоуровневый перцептрон (multilayer perceptrons) (MLPs), подготовка MLP
- метод опорных векторов (support vector machines): модель классификации SVM (SVM classification model) создание скоринговых моделей с помощью SVMs (short)
- кейсы: использование логистической регрессии и метод опорных векторов (support vector machines) для разработки системы кредитного рейтинга страны (country rating system)

Анализ долговечности для скоринга прибыли (Profit Scoring)

- анализ долговечности для разработки для моделей развития «времени жизни» клиента
- чувствительные проблемы
- кривые выживания, кривые versus hazard
- анализ Kaplan Meier
- параметрический анализ долговечности (parametric survival analysis)
- пропорциональная регрессия рисков (proportional hazards regression)
- использование анализа долговечности для моделирования кредитных рисков

Материалы курса:

Слушатели получают печатные материалы курса

4. Разработка и применение Credit Scorecard (Credit Scorecard Development and Implementation)

Длительность курса - 2 дня.

Данный курс проводится [Naeem Siddiqi](#), менеджером по продуктам кредитного скоринга, Глобальная практика по управлению рисками в SAS, автором книги «*Credit Risk Scorecards*»

На этом бизнес-ориентированном курсе слушатели получают необходимые знания для планирования, разработки, внедрения и поддержки скоринговых карт внутри организации. Тематика управления кредитными рисками раскрывается в наиболее полной и качественной мере; вниманию слушателей также предлагаются стратегии по внедрению скоринговых карт.

Цели курса:

- создать проектный и бизнес планы по разработке скоринговых карт
- разработать и протестировать на корректность аналитические скоринговые карты по кредитным рискам в пошаговом режиме
- сгенерировать скоринговые карты и отчеты по производительности (generate scorecard and portfolio performance reports).

Аудитория:

Менеджеры по кредитным рискам/кредитному скорингу, специалисты по работе с Data Miner; все специалисты, вовлеченные в процесс проверки/валидации и аудита моделей; разработчики стратегии по рискам; руководители/администраторы по кредитным рискам (credit risk executives).

Предварительная подготовка:

Для посещения курса опыт работы с SAS и навыки программирования не требуются. Слушатели должны быть знакомы с логистической регрессией. Знания SAS Enterprise Miner желательны, но не являются обязательными.

Приветствуется опыт работы в бизнес сфере, в следующих направлениях:

банки/финансы
розничные компании
страхование
телекоммуникации

Описание программы курса:

Введение в управление кредитными рисками

- определение рисков по типам
- экономическая выгода от использования системы управления кредитными рисками
- обсуждение приложений по управлению кредитными рисками

Роли разработки скоринговых карт

- определение ролей разработки скоринговых карт и обозначение их важности

Разработка скоринговых карт. Этап 1: предварительная подготовка и планирование

- постановка задач проекта
- составление проектного плана
- составление бизнес-плана

Разработка скоринговых карт. Этап 2: Обзор данных (Data Review) и параметры проекта

- исследование данных на предмет наличия и качества
- сбор данных с целью определения параметров проекта

Разработка скоринговых карт. Этап 3: Разработка и создание базы данных

- создание и установка данных
- соответствие ценовым вероятностям

Разработка скоринговых карт. Этап 4: Разработка моделей

- работа с пропущенными и отклоняющимися значениями (handling missing values and outliers)
- осуществление анализа начальных характеристик
- создание предварительных скоринговых карт с помощью логистической регрессии
- осуществление reject inference
- масштабирование скоринговой карты
- выборка лучшей скоринговой карты
- валидация скоринговой карты

Разработка скоринговых карт. Этап 5: Создание отчетов по управлению скоринговыми картами

- создание отчетов для обеспечения мониторинга и процесса принятия решений

Разработка скоринговых карт. Этап 6: Внедрение скоринговых карт

- валидация предпроектной фазы
- скоринговые стратегии
- установки блокировок
- разработка стратегий к действию, используя скоринговые карты, понимание правил работы алгоритмов и оверрайтов
- пост-проектная фаза
- создание скоринговых карт и отчетов о результатах мониторинга reports

Материалы курса:

Слушатели получают печатные материалы курса, а также копию книги автора

Naеem Siddiqi «*Credit Risk Scorecards: Developing and Implementing Intelligent Credit Scoring*»

5. Продвинутая аналитика для Customer Intelligence при использовании SAS (Advanced Analytics for Customer Intelligence Using SAS)

Длительность курса - 3 дня.

Данный курс проводится докторами философских наук, помощниками профессоров Школы Управления в Университете Саутгемптона (Великобритания), [Bart Baesens, Ph.D.](#) либо [Christophe Mues, Ph.D.](#)

Этот продвинутый курс поможет понять, как адаптировать ультрасовременные техники data mining под комплексные приложения по клиентской аналитике. Курс изобилует не только теоретическими и техническими тонкостями, но также и практическими деталями внедрения, что показано на примерах реальных case studies и упражнениях.

Цели курса:

- научиться применять мощные и самые передовые техники data mining
- обеспечить практическое применение данных техник для того, чтобы оптимизировать стратегические бизнес решения
- получить представление о том, как совершенно новые технологии data mining могут изменить ваши ключевые бизнес-процессы
- развернуть, отследить и провести оптимальное контрольное тестирование системы data mining.

Аудитория:

Всем тем, кто занимается оценкой, мониторингом и поддержкой прогнозных моделей разных типов систем клиентской аналитики; всем тем, кто использует технологии data mining для работы с разными типами систем клиентской аналитики

Предварительные требования:

Вам будут необходимы следующие знания и опыт в следующих областях:

- предварительная обработка данных (такая, как отсутствующие значения, выбросы, категоризации, выборка и т.д.)
- разработка прогнозных моделей с помощью логистической регрессии
- разработка прогнозных моделей с помощью дерева решений
- разработка дескриптивных моделей с использованием базовых техник по сегментации
- расчет производительности прогнозных моделей (lift curves, ROC curves и др.)

Данный опыт можно получить из курсов [Data Mining Techniques: Theory and Practice](#) и [Decision Tree Modeling](#)

Описание курса:

Прогнозные модели для клиентской аналитики: модель процессов The KDD
Повторение информации по предварительной обработке данных и Data Mining
Продвинутый план выборки

- кросс-валидация (stratified, leave-one-out)
- bootstrapping

Нейронные сети

- многоуровневые перцептроны (MLPs)
- MLP types (RBF, recurrent, etc.)
- weight learning (backpropagation, сопряженные градиенты (conjugate gradient, etc.)
- чрезмерно близкая подгонка (early stopping, and weight regularization architecture selection (grid search, SNC, etc.)
- выбор входных значений (input selection Hinton graphs, likelihood statistics, brute force, etc.)
- самоорганизующиеся карты для (self organizing maps (SOMs) for unsupervised learning
- кейс: SOMs для анализа коррупции в стране

Метод опорных векторов (Support Vector Machines) (SVMs)

- линейное программирование (linear programming)
- the kernel trick and Mercer theorem
- классификация и регрессия SVMs
- multiclass SVMs (one versus one, one versus all coding)
- настройка гиперпараметров используя методы кросс-валидации
- кейс: benchmarking SVM classifiers

Раскрытие нейронных сетей и «черного ящика» SVM (Opening up the Neural Network and SVM Black Box)

- rule extraction methods (pedagogical versus decompositional approaches such as neurorule, neurolinear, trepan, etc.
- двухуровневые модели (two-stage models)

Повторение сведений о дереве решений (C4.5, CART, CHAID)

Дерева регрессии

splitting/stopping/assignment criteria

Ensemble Methods

- bagging
- boosting
- stacking
- random forests

Альтернативное правило представления форматов (Alternative Rule Representation Formats)

- типы правил (oblique, M-of-N, fuzzy, etc.)
- таблица данных для принятия решений (лексикографическое упорядочивание, contraction methods, etc.)
- диаграммы принятия решений
- кейс: таблица данных для принятия решений и диаграммы принятия решений для реализации клиентского скоринга

Классификация Байесовых сетей

- naive Bayes
- tree augmented naive Bayes (TAN)
- неограниченная классификация Байесовых сетей
- Байесовский вывод (Bayesian inference)
- кейс: использование Байесовых сетей для прогнозирования оттока клиентов

Анализ долговечности

- цензурирование
- анализ Kaplan-Meier
- параметрический анализ долговечности
- анализ пропорциональных рисков
- использование нейронных сетей для анализа долговечности (survival analysis)
- кейс: использование анализа долговечности с нейронными сетями для реализации клиентского скоринга (neural network survival analysis for customer scoring)

Обучение при использовании данных локальной сети (Learning Using Networked Data)

- случайные поля Markov (random fields)
- гомофильность (homophily) (по ассоциациям)
- локальные классификаторы
- реляционные классификаторы (relational classifiers) (реляционная соседняя величина, вероятностная реляционная соседняя величина, реляционная логистическая регрессия, и т.д.)
-
- групповой вывод (collective inference) (Gibbs sampling, iterative classification, etc.)

Мониторинг и контрольное тестирование аналитических моделей

- количественный vs. качественный мониторинг моделей
- контрольное тестирование моделей (стабильность модели, binomial/Hosmer-Lemeshow test, traffic light indicator approach, влияние макро-экономических воздействий)
- модель бенчмаркетинга (internal versus external benchmarking, benchmarking statistics)
- качественная валидация аналитических моделей (качество данных, разработка модели, документация, вовлеченность менеджмента)
- кейс: контрольное тестирование модели клиентского скоринга

Прочие техники прогнозного моделирования (Short)

- semi-supervised learning
- генетические алгоритмы
- расплывчатые техники (fuzzy techniques)
- ant colony optimization
- пример: Antminer+

Материалы курса:

Слушатели получают печатные материалы курса