

Программа учебной дисциплины Анализ данных в бизнесе

Утверждена

Академическим советом ООП

Протокол № от «__» _____ 20__ г.

Автор	Лапшин В.А., Миоров А.В., Романенко А.А, Сагайдак Г.Р., Титова Н.Н., Филипенков Н.В., Цуканов М.В.
Число кредитов	6
Контактная работа (час.)	80
Самостоятельная работа (час.)	146
Курс	3
Формат изучения дисциплины	без использования онлайн курса

I. ЦЕЛЬ, РЕЗУЛЬТАТЫ ОСВОЕНИЯ ДИСЦИПЛИНЫ И ПРЕРЕКВИЗИТЫ

Целями освоения дисциплины «Анализ данных в бизнесе» является получение представления об особенностях задач анализа данных в бизнесе, знакомство с конкретными примерами бизнес-задач, использующих анализ данных, а также знакомство со специализированным программным обеспечением SAS для решения этих задач.

В результате освоения дисциплины студент должен:

Знать:

- Основные формулировки, особенности и характеристики прикладных задач анализа данных в бизнесе, возникающих в области клиентской аналитики, в розничных сетях продаж товаров и при анализе и оценке рисков.

- Математические методы и модели для решения задач анализа данных в бизнесе, возникающих в области клиентской аналитики, в розничных сетях продаж товаров и при анализе и оценке рисков, принципы проверки и представления результата решения этих задач.

- Семейство программных технологий SAS для обработки и анализа данных, включая программные продукты SAS для загрузки и предобработки данных, интерактивного исследования данных, построения и применения прогнозных и описательных моделей, применяемых в области клиентской аналитики, в розничных сетях продаж товаров и при анализе и оценке рисков.

Уметь:

- Формулировать, решать и оценивать результат решения задач анализа данных в бизнесе, возникающих в области клиентской аналитики, в розничных сетях продаж товаров и при анализе и оценке рисков.

- Использовать программные средства загрузки, обработки, визуализации и интерактивного исследования данных, а также строить и применять на практике описательные и прогнозные модели интеллектуального анализа данных и машинного обучения с использованием технологий компании SAS в области клиентской аналитики, в розничных сетях продаж товаров и при анализе и оценке рисков.

- Оформлять и представлять свои результаты в виде бизнес-презентации.

Настоящая дисциплина относится к циклу/блоку дисциплин «По выбору».

Для специализации «Машинное обучение и приложения» настоящая дисциплина является дисциплиной по выбору.

Изучение данной дисциплины базируется на следующих дисциплинах:

- Машинное обучение 1.
- Прикладной статистический анализ данных.
- Непрерывная оптимизация.
- Алгоритмы и структуры данных.
- Основы и методология программирования.
- Базы данных.

Для освоения учебной дисциплины студенты должны владеть следующими знаниями и компетенциями:

- Знать основы алгоритмов и структур данных и уметь программировать на любом императивном языке программирования.
- Знать основы языка SQL и иметь представление об организации и принципах работы современных СУБД.
- Знать основы математического анализа, линейной алгебры, аналитической геометрии, теории непрерывной оптимизации, теории вероятностей и математической статистики.
- Знать и уметь применять на практике основные методы и подходы прикладного статистического анализа данных и машинного обучения.

II. СОДЕРЖАНИЕ УЧЕБНОЙ ДИСЦИПЛИНЫ

Содержание темы может распределяться по лекционным и практическим занятиям.

1. Раздел 1 — клиентская аналитика.

Лекции: 6 часов, практические занятия: 6 часов.

- Введение в клиентскую аналитику
 - Лекция (2 часа). Цели и задачи клиентской аналитики. Понятие путешествия клиента. Клиентская аналитика в индустриях. Обзор этапов развития клиентской аналитики. Решения и инструменты клиентской аналитики. Сквозной бизнес-процесс клиентской аналитики.
 - Практическое занятие (2 часа). Работа с данными для подготовки к моделированию. Подготовка АВТ- витрин для кластеризации магазинов и построения модели МВА.
- Построение моделей и визуализация данных
 - Лекция (2 часа). Обзор подходов к анализу данных, визуализация данных, предсказательное моделирование, методология анализа, сегментация, построение экспериментов.
 - Практическое занятие (2 часа). Работа с Visual Statistics: анализ качества данных, кластеризация, деревья решений, регрессия.
- Финансовое обоснование проектов, основанных на анализе данных
 - Лекция (2 часа). Основные понятия финансового планирования. Структура финансового плана проекта. Влияние клиентской аналитики на драйверы финансового плана. Подходы к прогнозированию изменений драйверов за счет реализации проектов на основе клиентской аналитики.
 - Практическое занятие (2 часа).

Разбор кейса: подготовка финансового плана организации, внедряющей целевой маркетинг

2. Раздел 2 — задачи анализа данных в розничных сетях продаж товаров

Лекции: 6 часов, практические занятия: 6 часов.

- Введение в задачи анализа данных в ритейле.
 - Лекция (2 часа). Структура Ритейл-компаний, как элемента цепочки поставок; ключевые KPI компании, логистические KPI ритейл-компаний; задача прогнозирования спроса и использование прогноза спроса в различных бизнес-процессах, кластеризация магазинов, оптимизация запасов, ценообразование.
 - Практическое занятие (2 часа). Задача восстановления спроса, прогнозирование спроса на новые товары.
- Задачи оптимизации ассортимента и запасов товаров в ритейл-сети.
 - Лекция (2 часа). Постановка задачи оптимизации ассортимента, постановка в виде задачи ML. Основные ограничения при определении ассортимента ТТ. Постановки задачи оптимизации запасов в сети. Основные стратегии при оптимизации запасов в сети (ss, bs). Мультиэшелонный подход, эффект хлыста.
 - Практическое занятие (2 часа). Решение задачи кластеризации магазинов для управления ассортиментом сети. Реализация алгоритма оптимального распределения товаров в сети с целью уменьшения оборачиваемости.
- Оптимизация цен, прогнозирование объёмов промо
 - Лекция (2 часа). Задача оптимизации цен. Модели эластичности спроса. Примеры каузальных переменных для модели эластичности спроса. Бизнес-ограничения при оптимизации цен. Задача прогнозирования промо в сети, постановка в виде задачи ML
 - Практическое занятие (2 часа). Решение задачи расчёта эластичности спроса на реальных данных. Пример решения задачи прогнозирования объёмов промо в ритейл сети, постановка в виде задачи ML.

3. Раздел 3 — основы оценки рисков.

Лекции: 6 часов, практические занятия: 6 часов.

Введение: роль оценки рисков в управлении рисками.

- Понятие о кредитном риске
 - Лекция (2 часа). Виды кредитного риска, виды задач в управлении кредитным риском. Виды данных для кредитного анализа. Предварительная обработка данных: фильтрация и заполнение пропусков; виды пропусков (MCAR, MAR, MNAR); «ошибка выжившего» и другие ошибки, связанные с игнорированием особенностей данных. Выбор переменных для анализа. Требования к модели кредитного анализа: точность, устойчивость, прозрачность. Модели в кредитном анализе: обобщённые линейные модели, деревья решений, модели интенсивности дефолта. Портфельный подход. Учёт макроэкономических и других внешних факторов.
 - Практическое занятие (2 часа). Загрузка данных в среду SAS. Типичные преобразования данных для кредитного анализа в SAS: фильтрация, заполнение пропусков, нормализация (например, преобразование WOE), визуализация данных. Интерактивный выбор переменных в SAS (fine classing, coarse classing). Анализ отказанных заявок (Reject Inference). Оценка основных моделей для кредитного анализа.
- Понятие о рыночном риске
 - Лекция (2 часа). Меры рыночного риска. Value-at-Risk и основные методы его оценки: дельта-нормальный, исторический и модельный. Основы портфельной арифметики. Базовые модели для оценки рыночного риска: факторные модели, GARCH, «тяжёлые хвосты».
 - Практическое занятие (2 часа). Загрузка рыночных данных в среду SAS. Типичные преобразования рыночных данных в SAS: вычисление доходностей, их

средних значений и волатильностей, заполнение пропусков, визуализация рядов рыночных данных. Оценка Value-at-Risk дельта-нормальным, историческим и модельным методами. Оценка параметров стохастического процесса по рыночным данным

- Понятие о валидации моделей оценки риска.
 - Лекция (1 час). Валидация моделей оценки рыночного риска. Бэкстестинг Value-at-Risk. Концепция скользящего окна. Проверка гипотез об уровне покрытия и о независимости пробитий VaR.
 - Практическое занятие (1 час). Статистические тесты в SAS. Программа для расчёта последовательности пробитий.
 - Литература: [Christoffersen, Ch. 13],
 - Лекция (1 час). Валидация моделей кредитного риска. Показатели качества модели. Статистические тесты стабильности выборки (PSI анализ, анализ gini во времени). Мониторинг модели во времени.
 - Практическое занятие (1 час). Статистические тесты в SAS. Программа в SAS

4. Тема 4. Техника презентации.

Лекции: 2 часа.

Практические занятия: 2 часа.

- Лекция (2 часа): общие принципы построения эффективных презентаций.
- Практическое занятие (2 часа): групповое обсуждение конкретных презентаций.

5. Командный проект.

Студенты самостоятельно записываются на один из предложенных проектов (см. примеры в разделе 10) и в течение всего 4-го модуля выполняют их. Очные практические занятия проходят в группах студентов, выбравших один и тот же проект. На практических занятиях в компьютерном классе студенты начинают основные этапы работы по проекту под руководством специалиста-практика. При необходимости в занятие включается теоретическая информация.

- Практические занятия (36 часов).
- Очная защита проектов (4 часа).
- Литература выдаётся командам индивидуально в соответствии с выбранным ими заданием.

III. ОЦЕНИВАНИЕ

Оценки за все домашние задания выставляются по 2-балльной шкале, где «2» — задание решено полностью, «1» — задание решено не полностью или с недочётами, «0» — задание не решено или решено неверно.

Если домашнее задание разбито на несколько частей, то каждая часть оценивается по 2-балльной шкале как описано выше, а затем оценки усредняются с равными весами без округления.

Перевод оценки за домашние задания из 2-балльной шкалы в 10-балльную проводится путём умножения оценки на 5 без округления.

Оценка за экзамен выставляется по 10-балльной шкале.

Оценка за командный проект выставляется по 10-балльной шкале.

Пусть оценки за 3 домашних задания по 10-балльной шкале — O_1, O_2, O_3 , а оценка за экзамен в конце первого модуля по 10-балльной шкале — $O_{\text{экз}}$.

Итоговая оценка $O_{\text{итог}}$ рассчитывается по формуле

$$O_{\text{итог}} = \min(0.3O_1 + 0.3O_2 + 0.3O_3 + 0.2O_{\text{экз}}, 10)$$

Округление происходит только в самом конце — в итоговой оценке. Округление арифметическое.

IV. ПРИМЕРЫ ОЦЕНОЧНЫХ СРЕДСТВ

4.1 Пример домашнего задания по клиентской аналитике

Вам даны данные магазина

- Транзакционные данные с привязкой клиентам за год
- Справочник по клиентам

Домашнее задание будет состоять из двух частей:

1. Оценка качества данных. Подготовка витрины данных для проведения моделирования;
2. Построение регрессионной модели для проверки корректности полученных данных.

Компания хочет сделать коммуникацию с наиболее вероятными покупателями, чтобы меньше потратить средств на рассылку ненужных писем. Покупателям будет предложена скидка /подарок/купон.

Задача:

Необходимо определить клиентов, которые с наибольшей вероятностью сделают заказ в категории, если им отправить предложение.

Целевой флаг – была ли покупка в категории (для каждого варианта она своя)

В рамках задания необходимо провести корректное импортирование и провести проверку целостности полученных данных.

Делаем проверку на:

- Пропущенные и нулевых значения в полях;
- Наличие некорректных знаков;
- Полноту данных по клиентам.

Корректируем данные – исправляем ошибки, восстанавливаем пропущенные значения.

Результатом первого блока будет:

- ✓ Выводы по данным. Все ли параметры были достоверны? Что Вы сделали, чтобы исправить ошибки?

Пример – Скорректировали названия категорий. Исключили магазины, которые закрылись;

- ✓ Сводные отчеты по основным параметрам данных* + визуализации (в формате Excel/SAS), для интерпретации:
 - Основные показатели по транзакционным данным LFL** по категориям/городам/магазинам/общий итог;
 - Основные показатели по клиентским данным по полу, возрасту, основным показателям продаж в рамках каждого клиента (сколько сделал заказов, выручки, кол-во товаров, маржи, средний чек;
- ✓ Витрина данных для построение регрессионной модели.

Результатом второго блока будет:

- ✓ Предоставление финального списка переменных, вошедших в финальную модель, их коэффициенты (β -коэффициенты) и их вероятности (p-value). Коэффициенты должны быть статистически значимыми и логически корректными;
 - Проекты в инструментах SAS (Guide, Miner) или код на языке программирования (Python/ R) с инструкцией для запуска на компьютере.

*Основными показателями являются - выручка, уникальное кол-во чеков, сумма маржи, сумма начисленной скидки, кол-во товаров.

**LFL (Like-For-Like) — это показатель эффективности торговли, позволяющий оценить динамику развития торговой сети или отдельного магазина, целиком исключив органические факторы: появление новых точек или закрытие уже существующих.

4.2 Пример домашнего задания по ритейлу

Дана выгрузка реальных данных о рекомендуемых объёмах пополнения каждого магазина в сети для конкретной линейки товаров (линейка содержит несколько SKU) и о величине прогнозируемого спроса по каждому SKU в магазинах. Бизнес задача: требуется найти оптимальные объёмы пополнения каждого SKU при заданных бизнес-ограничениях.

- 1) суммарный объём пополнения по всем SKU не должен превышать объём запасов каждого SKU, находящиеся на центральном складе;
- 2) следует минимизировать дату, когда все запасы закончатся во всех магазинах ритейл сети.

Задание для студентов:

1. Требуется сформулировать задачу оптимизации при заданных бизнес-ограничениях.
2. Найти аналитическое решение задачи
3. Реализовать код на языке SAS, решающий поставленную задачу оптимизации.
4. Рассчитать оптимальные объёмы отправки для всех SKU в реальных данных и сравнить результаты с выданным решением.

По итогам оформить краткий отчёт о ходе решения задачи; к отчёту приложить коды написанных программ.

Примеры решения задач будут разобраны на семинаре.

4.3 Пример домашнего задания по оценке рыночных рисков

Приемлемое решение задания будет объяснено на практических занятиях.

Дана выгрузка реальных данных о ходе торгов на бирже в течение исторического периода.

1. Загрузить данные и проверить их качество: выбросы, пропуски и т.д. В случае необходимости, провести первичную обработку данных.
2. Проверить гипотезу о нормальности данных и оценить Value-at-Risk дельта-нормальным методом.
3. Выбрать динамическую модель для данных, оценить её параметры — и оценить Value-at-Risk модельным методом.
4. По итогам оформить краткий отчёт, включающий все использованные модели, и методы; к отчёту приложить коды написанных программ.

4.4 Пример заданий проектов

4.4.1 Пример проекта по ритейлу

Построение прогноза по новым товарам

Что это за проект?

Ключевой вопрос при появлении нового товара в ритейл сети – а в каком магазине он будет более быстро продаваться? Для того чтобы ответить, куда везти товар, а куда не стоит, требуется спрогнозировать спрос на этот товар в каждом магазине, НЕ зная истории продаж товара. В рамках проекта предлагается решить задачу прогнозирования спроса по новым товарам с помощью методов Machine Learning.

Требования к разработке готовой системы

Система прогнозирования спроса по новым товарам должна уметь

1. получать на вход данные о новых товарах в заранее оговоренном формате;

2. разбивать имеющиеся данные на train\validate\test подвыборки;
3. строить финальный алгоритм прогнозирования спроса (решающее правило) на базе таких алгоритмов как LinearRegression, RandomForest, GradientBoosting и др; при этом обязательным элементом на пути построения решения будет:
 - a. автоматическая фильтрация данных;
 - b. автоматическое извлечение признаков;
 - c. автоматическая генерация новых признаков;
4. выдавать оценку по точности полученных алгоритмов на test выборке, а также другие статистики и отчёты о настройке финального алгоритма.

Все алгоритмы и структуры данных, с которыми работают данные алгоритмы, должны быть реализованы на платформе SAS.

Почему данное решение необходимо?

До сих пор проблема предсказания спроса на новые товары является белым пятном для большинства ритейл компаний. В данном проекте предлагается разработать подход с помощью методов машинного обучения, претендующий на универсальность: т.е. будет разработан фреймворк для построения прогнозов по новым товарам независимый от природы данных.

Наличие данного фреймворка позволит процесс построения качественного прогноза по новым товарам.

Чему научатся студенты? Что самое интересное в проекте?

Студенты смогут освоить и закрепить язык SAS и решения, написанные на SAS.

Для решения задач проекта студента погрузятся в специфику проблемы прогнозирования новых товаров в ритейл компаниях, осознают экономическое обоснование данной задачи. Обязательными элементами проекта является анализ реальных данных, реализация алгоритмов по отбору признаков, построения workflow для обучения алгоритма прогнозирования.

Наиболее занимательная (творческая) часть – это изучение реальных данных и построение прогноза для них. Для более живого понимания важности данного прогноза для ритейл компаний, студентам будет демонстрироваться отчет о том, сколько денег сэкономит их система благодаря более точному прогнозу.

Организация работы (Как студенты будут работать в команде?)

Команда из 2-3 человек. По ходу реализации системы работы будут распараллелены, чтобы каждый получил опыт и в анализе данных, и в разработке и в тестировании системы.

Компоненты (Из каких частей состоит проект?)

- 1) Изучение данных, изучение задачи прогнозирования новых товаров.
- 2) Анализ данных, написание фреймворка фильтрации данных.
- 3) Написание фреймворка для выявления наиболее значимых факторов.
- 4) Фреймворк по построению финального алгоритма, обучение алгоритмов, оценка точности построенного алгоритма.
- 5) Тестирование системы на реальных данных

Какие будут использоваться технологии?

Проект задуман как учебный — для помощи в освоении ПО SAS, а именно следующих продуктов:

- 1) SAS BASE и SAS Enterprise Guide – для анализа данных и разработки алгоритмов.
- 2) SAS Enterprise Miner и SAS Forecast Server.

По согласованию с руководителем проекта, студенты могут использовать любое другое ПО.

Какие начальные требования?

Навыки анализа данных. Навыки программирования (C, C++, C#, Python, R и др.) и знание SQL-like языков.

Темы вводных занятий

1. Постановка задачи прогнозирования спроса, план работа по проекту, погружение в ритейл.
2. План проекта.
3. Типовые задачи при построении решение методами машинного обучения.

Критерии оценки

8-10 – разработана полноценная система для построения прогноза, создано её описание, система позволяет значительно улучшить базовый результат (тривиальный прогноз*).

6-7 – разработана система для построения прогноза, система не ухудшает базовый результат (тривиальный прогноз).

4-5 – проведение анализа данных, выявление зависимостей и закономерностей. Разработка архитектуры и верхнеуровневое описание алгоритмов.

* Тривиальный прогноз строится алгоритмом типа decision stump (на базе одного признака).

4.4.2 Примеры проектов по клиентской аналитике

1. «Оптимизатор коммуникаций»

Задача:

Подготовка рекомендации по оптимальному набору коммуникаций (предложение + канал) для достижения максимального экономического эффекта.

Результат:

Программное решение, получающее в качестве входных данных (1) обучающие выборки с историей контактов, (2) выборки для скоринга и (3) параметры каналов: стоимость и емкость. На выходе: рекомендуемый список для коммуникации формата «клиент – предложение – канал» и прогноз дополнительной выручки и стоимости коммуникации.

2. «Обогащение информации о заемщике»

Задача:

Сбор дополнительной информации об организациях, подающих заявку на кредит, из открытых источников: в СМИ (о крупных сделках, поглощениях, конфликтах), в реестре судебных решений РФ и проч.

Результат:

Скрипт, собирающий публикации, относящиеся к интересующим организациям, выделение ключевых слов, подготовка справки с «тональностью» новостей

3. «Анализ рыночной корзины (Market Basket Analysis)»

Задача:

Построение модели совместных покупок на основе транзакционных данных по клиентам. Исходя из полученного результата необходимо выявить правила общих покупок по продуктовым группам.

Результат:

На основе разработанной модели ассоциативных правил, реализованной на платформе SAS, сформировали правила и вес связей совместных покупок в чеках. Определи наиболее эффективные связи между категориями – выяснили какие категории товаров приводят к покупке ключевых категорий – приносящих наибольшую прибыль.

4. «Сбор данных из открытых источников для конкурентного анализа рынка»

Задача:

Сбор данных по локациям конкурентов по схожей специфике бизнеса (адреса магазинов с фиксированием изменений - открытия/закрытия, отзывы) и населения в городах из открытых источников. Создание обновляемого отчета с фиксированием изменений в данных.

Результат:

Скрипт, собирающий адреса и интересующие данные из открытых источников по конкурентам в файл с последующей фиксацией изменений, которые произошли с ранней версией отчета. Подготовка отчета по произошедшим изменениям.

5. «Создание модели по прогнозированию оборота/трафика новых объектов»

Задача:

Построение модели для прогнозирования оборота /трафика будущих новых объектов или изменений конфигураций действующих.

Результат:

Разработана модель по прогнозированию гипотетического оборота/трафика новых объектов на платформе SAS. Сформированы факторы и предпосылки, при помощи которых было рассчитано прогнозное значение, с указанием всех рисков или же положительных факторов, влияющих на результат прогнозирования.

4.4.3 Примеры проектов по риск-менеджменту

Анализ портфельных рисков банка

Что это за проект?

Торговый портфель каждого банка состоит из разнообразных финансовых инструментов – акций, фьючерсов, опционов, свопов. Каждый из этих сложных инструментов, называемых производными, был создан, чтобы помочь взять риски под контроль, но их оценка и анализ – отдельная большая задача. В рамках этого проекта мы расширим проект из домашнего задания – познакомимся с основными видами ПФИ, выстроим фреймворк для их оценки, расчёта рыночного и процентного риска, сценарного анализа.

Требования к разработке готовой системы

Система анализа рисков должна уметь

- 1) Получать на вход информацию о структуре портфеля банка и временных рядах с историей котировок;
- 2) Оценивать стоимость следующих видов фин.инструментов:
 - a. Акции
 - b. Облигации

- c. Фьючерсы
 - d. Опционы
 - e. Свопы
 - f. ... (студенты могут добавлять свои финансовые инструменты, например, бинарные опционы)
- 3) Проводить симуляции Монте-Карло, описывающие динамику рынка – историческим, параметрическим методом или на основании статистических моделей:
 - a. GBM
 - b. Vasicek
 - c. GARCH
 - 4) Рассчитывать различные метрики риска, такие, как Value-at-Risk, Expected Shortfall, чувствительность к изменению процентных ставок, курсов валют и котировок;
 - 5) Проводить оценку точности написанных методов и выбранных подходов; Все алгоритмы и структуры данных, с которыми работают данные алгоритмы, должны быть реализованы на платформе SAS.

Организация работы (Как студенты будут работать в команде?)

Команда из 2-3 человека. По ходу реализации системы работы будут распараллелены, чтобы каждый получил практический опыт по основным направлениям – оценка деривативов, моделирование рыночных данных и анализ результатов расчёта.

Компоненты (Из каких частей состоит проект?)

- 1) Знакомство с основными видами финансовых инструментов.
- 2) Реализация методов оценки стоимости финансовых инструментов.
- 3) Анализ и подготовка исторических рыночных данных.
- 4) Построение эконометрических моделей динамики рыночных данных.
- 5) Проведение симуляций стоимости портфеля с использованием построенных моделей и написанных методов оценки, анализ результатов, проверка правдоподобности результатов.

Какие будут использоваться технологии?

Проект задуман как учебный — для помощи в освоении ПО SAS, а именно следующих продуктов:

- 1) SAS BASE, SAS Enterprise Guide, SAS Studio – для работы с данными и написания программ;
- 2) SAS ETS для построения эконометрических моделей;
- 3) SAS Risk Dimensions – для проведения расчётов

Темы вводных занятий

1. Производные финансовые инструменты;
2. Структура проекта по анализу риска;
3. План проекта.
4. Разбор отдельных компонент проекта.

Критерии оценки

9-10 – система анализа рисков включает весь диапазон указанных ФИ, а полученная оценка Value-at-Risk по акциям является более точной, чем та, что была получена в рамках домашнего задания.

7-8 – система анализа рисков включает несколько различных ФИ, а полученная оценка Value-at-Risk по акциям сопоставима по точности с той, что была получена в рамках домашнего задания.

4-6 – расчёт реализован для одного ФИ, полученная оценка ухудшает точность по сравнению с базовым подходом.

1. Оценка кредитного риска

Задача:

Построение модели для оценки рыночного риска.

По сути, проект является углублением соответствующего домашнего задания. В задании не вошло много интересных моментов.

4.5 Примеры заданий экзамена

- 1) Укажите соотношение между мерами Value-at-Risk (VaR) и Expected Shortfall (ES) для одного и того же непрерывного распределения убытков на одном и том же уровне значимости.
 - a) $VaR > ES$.
 - b) $VaR < ES$.
 - c) Соотношение зависит от вида конкретного распределения.
 - d) Даже в рамках одного и того же вида распределения соотношение может быть любым.
- 2) Пусть выборка состоит из всего двух кредитов: высококачественного ($PD=3\%$) и низкокачественного ($PD=10\%$). И пусть имеется идеальная модель, которая корректно отличает высококачественный кредит от низкокачественного. Известно, что по итогам этих кредитов аналитик успешно вычислил коэффициент Джини для идеальной модели. Какова вероятность, что он получился отрицательным?
 - a) 0%
 - b) 3%
 - c) 21%
 - d) 30%
- 3) Известно, что Value-at-Risk для некоторого портфеля на горизонте в 2 недели на уровне 99% был оценён как 2.5% от стоимости портфеля. Какая из приведённых ниже оценок VaR для этого же портфеля на этом же уровне на горизонте в 1 день вероятна?
 - a) 0.25%
 - b) 0.8%
 - c) 8%
 - d) 25%
- 4) Что такое дисконтирование денежного потока?
 - a) Скидка заказчику за ранний платеж
 - b) Плата за обналичивание средств, поступивших на счет компании
 - c) Приведение будущих денежных потоков к текущему периоду
- 5) Ценность денег со временем:
 - a) Становится больше
 - b) Становится меньше
 - c) Не изменяется
- 6) В каких проектах необходимо учитывать изменение стоимости денег во времени?
 - a) В любых проектах, требующих затрат
 - b) В длительных проектах (больше 1 года)
 - c) В коротких проектах (менее 1 года)
- 7) Планируя маркетинговую активность, необходимо принимать решение на основании:
 - a) Анализа инкрементального денежного потока

- b) Анализа полного денежного потока
 - c) Анализа отклика на кампанию
- 8) Анализ чувствительности проекта это:
- a) Экономическое моделирование результатов проекта в зависимости от различных значений ключевых параметров кампании
 - b) Запуск пилотного проекта и подтверждение ключевых параметров кампании
 - c) Опрос экспертов на предмет их понимания вероятных значений ключевых параметров кампании
- 9) Драйверы бизнес-кейса это:
- a) Те параметры анализируемого проекта, которые влияют на финансовые показатели организации
 - b) Программное обеспечение, которое необходимо установить, чтобы реализовать проект
 - c) Менеджеры организации, ответственные за внедрение проекта
- 10) Если наблюдаемое значение критерия больше критического значения, то гипотеза ...
- a) H_1 отвергается
 - b) H_0 отвергается
 - c) H_0 принимается
- 11) Коэффициент регрессии a показывает ...
- a) как меняется переменная y при увеличении переменной x на 1%
 - b) прогнозируемое значение зависимой переменной при $x = 0$
 - c) прогнозируемое значение зависимой переменной при $x > 0$
- 12) Коэффициент регрессии b показывает ...
- a) на сколько единиц в среднем изменяется переменная y при увеличении независимой переменной x на единицу
 - b) прогнозируемое значение зависимой переменной при $x = 0$
- 13) прогнозируемое значение зависимой переменной при $x < 0$ Допустимый предел значений средней ошибки аппроксимации ...%
- a) не более 8-10
 - b) более 8-10
 - c) не более 10-20
- 14) Коэффициенты регрессии (a , b) в выборочном уравнении регрессии определяются методом (ами) ...
- a) взвешенных наименьших квадратов
 - b) наименьших квадратов
 - c) Гаусса
- 15) Если значение выборочного коэффициента корреляции близко к единице, это говорит
- a) о наличии тесной связи между факторами
 - b) об отсутствии связи между факторами
 - c) о слабой связи между факторами.
- 16) Найти величину параметра «а» некоторой линейной регрессии, если $y = 20$, $b = 0,1$, $x = 10$.
- a) $a = 19$
 - b) $a = 21$
 - c) $a = 120$
- 17) Для оценки качества построенной регрессии служит эконометрическая характеристика
- a) Дисперсия
 - b) коэффициент детерминации
 - c) выборочный коэффициент корреляции.

- 18) Какова будет связь между фактическим и табличным значениями F-критерия, если нулевая гипотеза об отсутствии связи признаков отклоняется, и делается вывод о существенности этой связи?
- a) $F_{\text{факт}} = F_{\text{табл}}$
 - b) $F_{\text{факт}} < F_{\text{табл}}$
 - c) $F_{\text{факт}} > F_{\text{табл}}$

4.6 Вопросы для самоконтроля

Клиентская аналитика

1. Какие задачи решает клиентская аналитика?
2. В каких индустриях клиентская аналитика имеет значительный потенциал в настоящее время?
3. Какие базовые прогнозные модели позволяют оценить корректность используемых данных?
4. Какие бывают форматы данных? Какой формат лучше всего переводить переменные (числа, дата, время, валюта и тд.)?
5. Какие основные принципы оценки качества данных?
6. Как проверяется качество построенной прогнозной регрессионной модели?
7. Каким методом определяются коэффициенты регрессии? О чем говорят нам значение этих коэффициентов?
8. Что такое средняя ошибка аппроксимации?
9. Что такое дисконтирование денежного потока?
10. Что такое анализ чувствительности проекта?
11. Что такое драйверы бизнес-кейса?
12. Какая структура типичного бизнес-плана?
13. Назовите несколько показателей, на основании которых принимается решение, реализовывать проект или нет?

Ритейл

14. Чем отличается восстановленный спрос от реальных продаж в магазинах ритейл-сети?
15. Укажите как минимум три различных бизнес-процесса, где может использоваться прогноз спроса.
16. Укажите два ключевых KPI Распределительной логистики
17. Объясните бизнес-смысл величин ORDER_UPTO_LEVEL и REORDER_LEVEL.
18. В чем отличия между стратегиями оптимизации запасов в сети Ss и Bs?
19. В чем заключается эффект «хлыста» в задаче оптимизации запасов в мультиэшелонной сети.
20. Что такое эластичность спроса по цене?
21. Приведите минимум 5 разных каузальных переменных, которые могут быть использованы для построения модели эластичности. Для каждого из них приведите минимум один пример задачи, когда отсутствие их в модели приведёт к некорректному определению эластичности.
22. Что такое гедоническая модель цены (hedonic price)?

Риск-менеджмент

23. Каковы типичные проблемы с рыночными данными и как их решать?

24. Что такое Value-at-Risk? Какие другие меры риска вы знаете?
25. Каковы основные методы оценки VaR по рыночным данным?
26. Какие основные динамические модели используются для оценки VaR по рыночному риску?
27. Каковы основные методы оценки параметров динамических моделей по рыночным данным?
28. Как проверяется качество построенной VaR-модели?
29. Каковы типичные проблемы с анкетными данными и как их решать?
30. В каких случаях требуется преобразование переменных?
31. Каковы основные виды скоринговых моделей?
32. Как проверяется качество построенной скоринговой модели?
33. Для чего нужен анализ отказанных заявок?

V. РЕСУРСЫ

1. Основная литература

По Разделу 1:

1. Shive W. and Mouton D. (2012) Improving Retail Decisions with Customer Analytics: Leveraging Actionable Customer Insights across the Retail Enterprise to Build Sales and Profits. Paper 286-2012, SAS Institute, Inc., Cary, NC
(<http://support.sas.com/resources/papers/proceedings12/286-2012.pdf>)
2. Baer D. and Grover S. (2016) Enhanced Segmentation Using SAS® Visual Analytics and SAS® Visual Statistics. Paper 6222-2016, SAS Institute, Inc., Cary, NC.
(<https://support.sas.com/resources/papers/proceedings16/SAS6222-2016.pdf>)
3. К. Уолш (2000) Ключевые показатели менеджмента: как анализировать, сравнивать и контролировать данные, определяющие стоимость компании. М.: Дело.
4. McKinsey&Company (2000) How to write a business plan.
(http://10k.inc.hse.ru/files/McKINSEY_GUIDE_to_business_plan.pdf)
5. Kaplan Publishing (2018) CIMA P2 Study Text. Advanced Management Accounting.
(<http://kaplan-publishing.kaplan.co.uk/cima/p2/pages/study-text.aspx>)

По Разделу 2:

1. Шапиро Дж (2006). Моделирование цепи поставок. Питер. Серия «Теория менеджмента».
2. Tijms H.C., Groenevelt H. (1984). Simple approximations for the reorder point in periodic and continuous review (s, S) inventory systems with service level constraints. European Journal of Operational Research, Vol. 17, Issue 2, August 1984, Pages 175-190.

По Разделу 3:

1. Christoffersen P. (2012) Elements of Financial Risk Management. 2nd ed. Elsevier Academic Press.

По Разделу 4:

1. Мортон С. (2016) Лаборатория презентаций. Формула идеального выступления. Альпина Паблишер.

2. Дополнительная литература

Обучающий портал с дополнительной литературой для работы с SAS Enterprise Guide (<https://support.sas.com/en/software/enterprise-guide.html#freetutorials>)

Учебные пособия:

- <https://support.sas.com/documentation/cdl/en/leforinforref/63324/PDF/default/leforinforref.pdf>
- <http://support.sas.com/documentation/onlinedoc/guide/tut71/en/menu.htm>
- <http://support.sas.com/documentation/cdl/en/mcrolref/69726/HTML/default/viewer.htm>
- <https://support.sas.com/documentation/onlinedoc/ets/>
- <https://support.sas.com/rnd/app/ets/examples/garchex/index.html>

Основы программирования на SAS Base (<https://habr.com/company/sas/blog/>)

3. Программное обеспечение

Для успешного освоения дисциплины, студент использует следующие программные средства:

- SAS Enterprise Miner, Forecast Server, Enterprise Guide, SAS OR, SAS ETS (в рамках облачного академического сервиса SAS OnDemand for Academics).
- SAS Viya Visual Analytics and Visual Statistics (в виде удаленного доступа к учебным серверам компании SAS или образовательным сервисам Teradata University - TUN).
- SAS Risk Dimensions (в виде удаленного доступа к учебным серверам компании SAS).
- Python/R (по желанию студента).

4. Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)

№ п/п	Наименование	Условия доступа
<i>Профессиональные базы данных, информационно-справочные системы</i>		
1.	Консультант Плюс	<i>Из внутренней сети университета (договор)</i>
2.	Электронно-библиотечная система Юрайт	URL: https://biblio-online.ru/
<i>Интернет-ресурсы (электронные образовательные ресурсы)</i>		
1.	Открытое образование	URL: https://openedu.ru/

5. Материально-техническое обеспечение дисциплины

Используются персональный компьютер (ноутбук) и проектор для проведения лекций и практических занятий, техническое оснащение компьютерных классов для выполнения лабораторных работ должно позволять осуществлять удаленный доступ к дата центрам компании SAS, устанавливать программных клиентов аналитических облачных сервисов на операционной системе Windows 8 и выше (64 bit), поддерживающей Java Runtime Environment (JRE) версии 8 и выше.