



**Федеральное государственное автономное образовательное учреждение
высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"**

Факультет компьютерных наук

**Рабочая программа дисциплины «Анализ данных в бизнесе»
в 2021/2022 учебном году на потоке образовательных программ:**

- ["Прикладная Математика и Информатика"](#) Факультета Компьютерных Наук НИУ ВШЭ
- ["Прикладной анализ данных"](#) Факультета Компьютерных Наук НИУ ВШЭ
- ["Международный институт экономики и финансов"](#) Международный институт экономики и финансов НИУ ВШЭ
- ["Бизнес-информатика"](#) Высшая школа бизнеса НИУ ВШЭ
- ["Экономика"](#) Факультет Экономических наук НИУ ВШЭ

Разработчик(и) программы:

Романенко А.А, aromanenko@hse.ru

Титова Н.Н., nntitova@hse.ru

Филипенков Н.В. nfilipenkov@hse.ru

Воробьева М.С. msvorobeva@hse.ru

Лапшин В.А., к.ф.-м.н., vlapshin@hse.ru

Мировов А.В. Alexander.Mironov@sas.com

Лобок Т.С. tlobok@hse.ru

Утверждена Академическим советом образовательной программы

«__»_____ 2021 г., № протокола _____

Академический руководитель образовательной программы

_____ [подпись]



_____, 2021

Настоящая программа не может быть использована другими подразделениями университета и другими вузами без разрешения подразделения-разработчика программы.

1 Область применения и нормативные ссылки

Настоящая программа учебной дисциплины устанавливает требования к образовательным результатам и результатам обучения студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих дисциплину «Анализ данных в бизнесе», учебных ассистентов и студентов направления подготовки/специальности «Машинное обучение и приложения», обучающихся по образовательной программе «Прикладная математика и информатика» и «Прикладной анализ данных».

Курс читается дополнительно для студентов других факультетов:

- ["Международный институт экономики и финансов"](#) Международный институт экономики и финансов НИУ ВШЭ
- ["Бизнес-информатика"](#) Высшая школа бизнеса НИУ ВШЭ
- ["Экономика"](#) Факультет Экономических наук НИУ ВШЭ

Программа учебной дисциплины разработана в соответствии с:

- Образовательным стандартом НИУ ВШЭ ([Ссылка на программу 2017 года, ее надо освежить](#))
<https://www.hse.ru/data/2017/09/04/1321436546/2.03.02%D0%9F%D1%80%D0%B8%D0%BA%D0%BB%D0%B0%D0%B4%D0%BD%D0%B0%D1%8F%20%D0%BC%D0%B0%D1%82%D0%B5%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0%20%D0%B8%20%D0%B8%D0%BD%D1%84%D0%BE%D1%80%D0%BC%D0%B0%D1%82%D0%B8%D0%BA%D0%B0.pdf>;
- Образовательной программой 01.03.02. «Прикладная математика и информатика».
- Объединенным учебным планом университета по образовательной программе «Прикладная математика и информатика», утвержденным в 2017г.

2 Цели освоения дисциплины

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

Актуальная информация по курсу расположена на странице на WIKI

[http://wiki.cs.hse.ru/%D0%90%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85_%D0%B2_%D0%B1%D0%B8%D0%B7%D0%BD%D0%B5%D1%81%D0%B5_\(%D0%9A%D0%B0%D1%84%D0%B5%D0%B4%D1%80%D0%B0_SAS\)#.D0.9B.D0.B5.D0.BA.D1.86.D0.B8.D0.B8](http://wiki.cs.hse.ru/%D0%90%D0%BD%D0%B0%D0%BB%D0%B8%D0%B7_%D0%B4%D0%B0%D0%BD%D0%BD%D1%8B%D1%85_%D0%B2_%D0%B1%D0%B8%D0%B7%D0%BD%D0%B5%D1%81%D0%B5_(%D0%9A%D0%B0%D1%84%D0%B5%D0%B4%D1%80%D0%B0_SAS)#.D0.9B.D0.B5.D0.BA.D1.86.D0.B8.D0.B8)

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

В результате освоения дисциплины студент должен:



Знать:

- Ключевые показатели эффективности и основные метрики операционной и финансовой деятельности, используемые в разных отраслях экономики, отраслевую и функциональную специфику реализации задач анализа данных в бизнесе.
- 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 и в других отраслях экономики.
- 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.

Уровни формирования компетенций:

РБ — ресурсная база, в основном теоретические и предметные основы (знания, умения);

СД — способы деятельности, составляющие практическое ядро данной компетенции;

МЦ — мотивационно-ценностная составляющая, отражает степень осознания ценности компетенции человеком и готовность ее использовать

В результате освоения дисциплины студент осваивает компетенции:

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
Способен описывать проблемы и ситуации профессиональной деятельности, используя язык и	ПК-1	СД	Формулирует основные задачи работы с данными в бизнесе, знает соответствующие математические методы и подходы для их решения	Лекции, практические занятия, домашние задания, проект	Практические задания, проект, экзамен



аппарат математики					
Способен провести сбор, обработку и анализ данных с использованием существующих методов машинного обучения	ПК-7	СД	Применяет соответствующие методы обработки и анализа данных, способен оценивать качество получаемых моделей и применять их для решения задач прогнозирования и выявления знаний	Лекции, практические занятия, домашние задания, проект	Практические задания, проект, экзамен
Способен проявлять творческий подход, инициативу и настойчивость в достижении целей (как профессиональной деятельности, так и личных)	ПК-18	МЦ	Способен самостоятельно формализовать задачу, критически оценивает и представляет результат своей работы	Лекции, практические занятия, домашние задания, проект	Практические задания, проект, экзамен

4 Место дисциплины в структуре образовательной программы

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.



Методические указания по освоению дисциплины.

Для лучшего усвоения дисциплины рекомендуется:

- пройти онлайн-курсы:
 - SAS Programming I: Essentials
<https://support.sas.com/edu/schedules.html?ctry=us&id=277>
 - SAS Statistics 1: Introduction to ANOVA, Regression, and Logistic Regression
<https://support.sas.com/edu/schedules.html?ctry=us&id=1979>
- читать:
 - Г.И.Ивченко, Ю.И.Медведев. Математическая статистика
 - О дисперсионном анализе в SAS/STAT:
http://support.sas.com/documentation/cdl/en/statug/66103/HTML/default/viewer.htm#statug_introanova_toc.htm

5 Тематический план учебной дисциплины для 3 и 4 курса

Для студентов 3 курса курс длится 3 и 4 модуль и включает практический проект.

Курс	3
Модуль	3 и 4 модуль
Число кредитов	6
Часы по видам работ (лекции, семинары, практические занятия) (академ. час.)	64
Самостоятельная работа (академ. час.)	164
Общее число академ. часов	228
Формат изучения дисциплины	Full time

Для студентов 4 курса курс длится 3 модуль и завершается экзаменом.

Курс	4
Модуль	Только 3 модуль
Число кредитов	5
Часы по видам работ (лекции, семинары, практические занятия) (академ. час.)	44
Самостоятельная работа (академ. час.)	146
Общее число академ. часов	190
Формат изучения дисциплины	Full time



Более подробно описание часов по темам и модулям

Модуль	№	Название раздела	Всего часов	Аудиторные часы (1 лекция/семинар 2 академ. Часа по 40 минут)				Самостоятельная работа
				Лекции	Семинары	Практические занятия	Другие виды работы ¹	
3 модуль	1	Клиентская аналитика	42	4	4			34
	2	Текстовая аналитика	42	4	4			34
	3	Задачи анализа данных в розничных сетях продаж товаров	46	6	6			34
	4	Основы оценки рисков	46	6	6			34
	5	ModelOps	14	2	2			10
Итого за 3 модуль			190	22	22			146
4 модуль	6	Учебный проект	38			20		18
Итого за 4 модуль			38			20		18
Итого за 3 + 4 модуль			228	22	22	20		164

6 Формы контроля знаний студентов 3 и 4 курса

Тип контроля	Форма контроля	1 год		Параметры **
		3	4	
Текущий	3 Домашних задания	3		Отчёт о выполнении задания с приложением кодов написанных программ.
	Командный проект		*	Презентация результатов проекта
Итоговый	Экзамен	* Для 4 курса в конце 3 модуля	* Для 3 курса в конце 4 модуля	Письменный экзамен, 80 мин, вопросы в виде теста с вариантами ответов.

7 Критерии оценки знаний, навыков

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.

¹ Указать другие виды аудиторной работы студентов, если они применяются при изучении данной дисциплины.



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.

Содержание дисциплины

	Наименование разделов и дисциплин	Всего академ. часов в	В том числе, ак. часов		Форма контроля
			Лекции (1 пара=2 ак. часа)	Практические занятия и семинары (1 пара=2 ак. часа)	
Раздел 1. Клиентская аналитика		8	4	4	-
1	Тема 1.1. Введение в клиентскую и онлайн аналитику	4	2	2	
2	Тема 1.2. Построение прогнозных моделей и визуализация данных	4	2	2	Домашняя работа №1
Раздел 2. Текстовая аналитика		8	4	4	
1	Тема 2.1. Введение в задачи анализа текстовых данных	4	2	2	
2	Тема 2.2. Инструменты и методы текстовой аналитики.	4	2	2	
Раздел 3. Задачи анализа данных в розничных сетях продаж товаров		12	6	6	
1.	Тема 3.1. Введение в задачи анализа данных в ритейле. Прогнозирование спроса	4	2	2	
2	Тема 3.2. Описательная аналитика в Ритейл: кластеризация магазинов, сегментация товаров, восстановление спроса	4	2	2	Домашняя работа №2
3.	Тема 3.3. Задачи оптимизации запасов товаров в ритейл-сети, оптимизация цен, оптимизация ассортимента	4	2	2	



Раздел 4. Основы оценки рисков		12	6	6	
1.	Тема 4.1. Введение в кредитные риски	4	2	2	
2.	Тема 4.2. Введение в рыночные риски	4	2	2	Домашняя работа №3
3.	Тема 4.3. Валидация моделей	4	2	2	
ModelOps - Операционализация моделей машинного обучения		4	2	2	
Итоговый контроль за 3 модуль			Экзамен для студентов 4 курса		
Учебный проект для студентов 3 курса			10 практических занятий		
Итоговый контроль за 3 и 4 модуль			Экзамен для студентов 3 курса		

Раздел представляется в удобной форме (список, таблица). Изложение строится по разделам и темам. Содержание темы может распределяться по лекционным и практическим занятиям.

1 час- 40 минут. (академический час)

3 модуль

1. Раздел 1 — Клиентская аналитика. Всего 8 академ. часов (Лекции: 4 часа, семинары: 4 часа.)

- Введение в клиентскую и онлайн аналитику
 - о 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. Что такое онлайн-аналитика? Какие цели и задачи она решает? Сбор и обработка данных. Инструменты онлайн-аналитики.
 - о Семинар (2 часа). Работа с данными для подготовки к моделированию. Инструменты и методы анализа слабоструктурированных данных в оптимизации маркетинговых коммуникаций. Построение моделей и визуализация данных. Работа с SAS Visual Statistics: анализ качества данных, кластеризация, деревья решений, регрессия.
- Построение прогнозных моделей и визуализация данных
 - о Лекция (2 часа). Обзор существующих методов по обработке и анализа данных (Сегментация, анализ покупательских корзин, модель отклика, рекомендательная система) Бизнес-интерпретация результатов построения моделей прогнозного моделирования.
 - о Семинар (2 часа). Поведенческое профилирование на основе построенных моделей.

2. Раздел 2 — Текстовая аналитика. Всего 8 академ. часов (Лекции: 4 часа, семинары: 4 часа.)

- Введение в задачи анализа текстовых данных.
 - о Лекция (2 часа). Обзор задач бизнеса, которые решаются с помощью классификации текстов или извлечения фактов из неструктурированных данных (отчетов, пользовательских отзывов, новостей и т.д.). Рассмотрение индустрий, где такие задачи встречаются.



- Семинар (2 часа). Тематическое моделирование. Практикум по тематическому моделированию на SAS Viya.
- Инструменты и методы текстовой аналитики.
 - Лекция (2 часа). Обзор методов обработки текста и извлечения знаний из неструктурированных данных. Обзор программных продуктов: свободного ПО, библиотек Python и SAS Viya – которые эти методы реализуют.
 - Семинар (2 часа). Автоматизация работы с юридическими документами. Практикум по извлечению информации из протоколов общих собраний акционеров на SAS Viya.

3. Раздел 3 — Задачи анализа данных в розничных сетях продаж товаров. Всего 12 академ. часов (Лекции: 6 часов, семинары: 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.
 - Семинар (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.
 - Семинар (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
 - Семинар (2 часа). Решение задачи расчёта эластичности спроса на реальных данных. Пример решения задачи прогнозирования объёмов промо в ритейл сети, постановка в виде задачи ML.

4. Раздел 4 — основы оценки рисков. Всего 12 академ. часов (Лекции: 6 часов, семинары: 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.
 - Семинар (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».
 - Семинар (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.
 - Семинар (1 час). Статистические тесты в SAS. Программа для расчёта последовательности пробитий.
 - Лекция (1 час). Валидация моделей кредитного риска. Показатели качества модели. Статистические тесты стабильности выборки (PSI анализ, анализ gini во времени). Мониторинг модели во времени.
 - Семинар (1 час). Статистические тесты в SAS. Программа в SAS
- 5. Раздел 5 — Оптимизация маркетинговых кампаний для повышения отклика и увеличения прибыли.**
- Задачи анализа данных в розничной торговле.
 - Лекция (2 часа). Описания концепции и разработки комплекса предиктивных моделей. Анализ бизнес ограничений, демонстрирующий зависимость прибыли от изменения доступности ограниченного ресурса (sensitivity-анализ). Процессы управления кампаниями с использованием оптимизации.
 - Семинар (2 часа). Разработки комплекса предиктивных моделей и анализ бизнес ограничений.

4 модуль

Раздел 6 — Командный проект.

Студенты самостоятельно записываются на один из предложенных проектов (см. примеры в разделе 10) и по ссылке

<https://docs.google.com/document/d/1NN9Pva4OJwZBpIR6n2Oqf4ILTVfjMrsv11sQM19behM/edit#heading=h.e45y7bybvwox>

и в течение всего 4-го модуля работают над проектом. Очные и онлайн практические занятия проходят в группах студентов (3 человека на 1 проект) с куратором раз в неделю. На практических занятиях в компьютерном классе или дистанционно студенты начинают основные этапы работы по проекту под руководством специалиста-практика. При необходимости в занятие включается теоретическая информация.

- Практические занятия (20 академ. Часов).
- Самостоятельная работа над проектом (18 академ. Часов).
- Подготовка презентационного материала с куратором (опционально)
- Очная защита проектов (30 минут выступление + вопросы).



- Литература выдаётся командам индивидуально в соответствии с выбранным ими заданием.

8 Образовательные технологии

Программная среда для работы: студенты будут работать с платформой SAS Viya. Все необходимое ПО размещается в облаке, предоставляемом SAS, для доступа достаточно браузера.

Дополнительно могут проводиться практические занятия в системе SAS OnDemand for Academics (доступна онлайн с любого компьютера).

Домашние задания рекомендуется выполнять в этой же системе, однако по согласованию с проверяющим, студенты могут использовать и другие программные продукты / языки программирования.

8.1 Методические указания студентам

Рекомендуется использовать этот курс как повод познакомиться с технологиями SAS и овладеть ими хотя бы на минимальном уровне. Да, все задания курса можно выполнить и с использованием тех технологий, которыми вы уже владеете. Курс предоставляет вам учебные задачи, специально подобранные для знакомства с технологиями SAS.

9 Оценочные средства для текущего контроля и аттестации студента

9.1 Примеры домашних заданий

9.1.1 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.*

Актуальная версия домашнего задания №1 на 2020 год

<https://drive.google.com/file/d/16WFqo1xfPvnlGLwyorWFGdiL2ZM8xp8D/view?usp=sharing>

9.1.2 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.

Актуальная версия домашнего задания №2 на 2020 год

<https://drive.google.com/file/d/1R6BLH40CLk7uewRn02JIGj64Ov2peJgK/view?usp=sharing>

9.1.3 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.

Актуальная версия домашнего задания №3 на 2020 год Домашнее задание №3

Необходимо построить скоринговую модель, оценивающую вероятность дефолта клиента на стадии заведения кредитной заявки.

Для этого необходимо:

0. Скачать данные по ссылке

<https://drive.google.com/drive/u/0/folders/16CMypnLu7Fv7IgsYOZimQK-7MaFZEWEZ>

Каждый студент выбирает 2 выборки "accept" и "reject", которые начинаются с варианта студента.

Номера вариантов для ДЗ №3 необходимо взять такие же как и для ДЗ №1

Выполненное задание необходимо отправить в следующем виде:

1) Файл/скрипты с построенными моделями (обязательно должны быть комментарии, без комментариев задание считается нерешенным)

2) Excel файл с ответами на следующие вопросы:

1. Какая доля 1 в выборке "accept"?
2. Необходимо рассчитать для всех интервальных переменных следующее:
 - Доля пропущенных значений
 - Медиана
 - Среднее
 - Среднеквадратическое отклонение
 - Есть ли аномальные значения, выбросы?
 - Information Value
3. Необходимо рассчитать для всех категориальных переменных следующее:
 - Мода
 - Доля пропущенных значений
 - Information Value
 - Есть ли выбросы, аномальные значений
4. Построить логистическую регрессию только на одобренных заявках с преобразованными переменными WoE. Какое значение GINI? F1 мера?
5. Провести анализ Reject Inference. Какая доля отказанных заявок?
6. Построить логистическую регрессию на всех заявках с преобразованными переменными WoE. Какое значение GINI, F1? Изменилась ли модель?
7. Какую модель вы рекомендуете для внедрения в продуктивную среду? Дать развернутое пояснение



9.1.4 Пример домашнего задания по оценке кредитных рисков

Дана выборка данных о кредитных заявках.

1. Провести анализ переменных. Какие переменные имеют пропущенные значения? Как вы будете обрабатывать эти пропущенные значения?
2. Необходимо ли проводить преобразование переменных?
3. Есть ли выбросы в данных?
4. Построить несколько моделей оценки кредитного риска.
5. Выбрать из построенных моделей наилучшую и обосновать выбор.
6. Провести анализ отказанных заявок.
7. Провести валидацию модели: анализ входящего потока (PSI анализ) и статистические тесты.
8. Выдать рекомендации по мониторингу модели

9.2 Пример заданий проектов

Актуальные темы на 2020 год с описанием представлены тут:

<https://docs.google.com/document/d/1NN9Pva4OJwZBpIR6n2Oqf4ILTvfjMrsv11sQM19behM/edit#heading=h.e45y7bybvwox>

9.2.1 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 теоретического и практического материала изученного в 3 модуле, а также ПО SAS, а именно следующих продуктов:

- 1) SAS Visual Analytics (on SAS Viya)
- 2) SAS VDMML (on SAS Viya)



- 3) SAS BASE, SAS/STAT – Statistical analysis и SAS Enterprise Guide – для анализа данных и разработки алгоритмов.
- 4) SAS Enterprise Miner и SAS Forecast Server.

По согласованию с руководителем проекта, студенты могут использовать любое другое ПО или выполнить задание на Python.

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).

9.2.2 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.

9.2.3 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:

- 5) SAS BASE, SAS Enterprise Guide, SAS Studio – для работы с данными и написания программ;
- 6) SAS ETS для построения эконометрических моделей;
- 7) 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 задание не вошло много интересных моментов.

9.3 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 « a » 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) $F_{\text{факт}} > F_{\text{табл}}$
- 19) Для какой компании характерна высокая сезонность выручки?
- a) Авиакомпания, специализирующаяся на пассажирских перевозках
 - b) Metallurgical combination
 - c) Тепличное хозяйство, выращивающее огурцы



- 20) В какой отрасли в оценке финансового положения компании большую роль играют забалансовые активы?
- Авиация
 - Телекоммуникации
 - Розничная торговля
- 21) С какими проблемами при обработке данных, скорее всего, столкнется аналитик, исследующий базу с персональной информацией клиентов банка и сведениями об их профилях в социальных сетях?
- Плохое качество данных
 - Преднамеренное вмешательство в данные
 - Очевидность выявляемых закономерностей
- 22) К какому типу задач описательной аналитики относится расшифровка информации с NFC датчиков самолета?
- Выявление новизны
 - Профилирование
 - Анализ последовательностей
- 23) Какие задачи прогнозной аналитики с наибольшей вероятностью будет решать Data Scientist в FMCG компании?
- Выявление мошенничества
 - Мониторинг процессов
 - Прогнозирование выручки

9.4 Вопросы для самоконтроля

Клиентская аналитика

- Какие задачи решает клиентская аналитика?
- В каких индустриях клиентская аналитика имеет значительный потенциал в настоящее время?
- Какие базовые прогнозные модели позволяют оценить корректность используемых данных?
- Какие бывают форматы данных? Какой формат лучше всего переводить переменные (числа, дата, время, валюта и тд.)?
- Какие основные принципы оценки качества данных?
- Как проверяется качество построенной прогнозной регрессионной модели?
- Каким методом определяются коэффициенты регрессии? О чем говорят нам значение этих коэффициентов?
- Что такое средняя ошибка аппроксимации?
- Что такое дисконтирование денежного потока?
- Что такое анализ чувствительности проекта?
- Что такое драйверы бизнес-кейса?
- Какая структура типичного бизнес-плана?
- 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?

10 Порядок формирования оценок по дисциплине

Пусть оценки за 3 домашних задания по 10-балльной шкале — O_1, O_2, O_3 , а оценка за экзамен в конце второго модуля по 10-балльной шкале — $O_{\text{экз}}$, оценка за проект во втором модуле $O_{\text{пр}}$ выставляется по 10-балльной шкале по итогам защиты проекта.

Итоговая оценка для студентов 3 курса, которые прошли 3 и 4 модуль $O_{\text{итог}}$ определяется по формуле:

$$O_{\text{итог}} \text{ для ПМИ, ПАД, БЭК} = 0.1 * O_1 + 0.1 * O_2 + 0.1 * O_3 + 0.2 * O_{\text{экз}} + 0.5 * O_{\text{пр}}$$

Округление происходит в итоговой оценке в сторону ближайшего целого.

Для студентов 4 курса, которые прошли только 3 модуль

$$O_{\text{итог}} \text{ для МИЭФ, БИ} = 0.225 * O_1 + 0.225 * O_2 + 0.225 * O_3 + 0.235 * O_{\text{экз}}$$



Оценки за все домашние задания выставляются по 10-балльной шкале для каждой домашней работы могут немного скорректированы, где

- «8-10» — задание решено полностью, разработана полноценная система для построения прогноза, создано её описание, система позволяет значительно улучшить базовый результат (тривиальный прогноз*).
- «6-7» — задание решено не полностью или с недочётами, разработана система для построения прогноза, система не ухудшает базовый результат (тривиальный прогноз).
- «4-5» — задание решено с существенными недочётами, проведение анализа данных, выявление зависимостей и закономерностей. Разработка архитектуры и верхнеуровневое описание алгоритмов.
- «0-3» — задание не решено или решено неверно.
- * Тривиальный прогноз строится алгоритмом типа decision stump (на базе одного признака).

Оценка за командный проект и экзамен выставляется по 10-балльной шкале.

11 Учебно-методическое и информационное обеспечение дисциплины

11.1 Базовый учебник

Отсутствует

11.2 Основная литература

По Разделу 1:

1. Sweetwood A. (2016) The Analytical Marketer: How to Transform Your Marketing Organization. Harvard Business Review Press.
2. Anderson C. (2015) Creating a Data-Driven Organization: Practical Advice from the Trenches. O'Reilly Media, Inc.
3. SAS Documentation (2015). SAS® Visual Analytics 7.2, 7.3, and 7.4: Getting Started with Analytical Models
(<https://support.sas.com/documentation/cdl/en/vaamgs/71329/PDF/default/vaamgs.pdf>)
4. SAS(R) Visual Analytics 7.3: User's Guide
(<https://support.sas.com/documentation/cdl/en/vaug/68648/HTML/default/viewer.htm#n0lfkxxrf3zhugn14g96lr8spatc.htm>)
5. 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>)
6. 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>)
7. LeSueur J. (2007) Marketing Automation. Practical Steps to More Effective Direct Marketing. Wiley and SAS Business Series.
8. McKinsey&Company (2000) How to write a business plan.
(http://10k.inc.hse.ru/files/McKINSEY_GUIDE_to_business_plan.pdf)
9. 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.
3. Sherbrooke C. (2004) Optimal Inventory Modeling of Systems. Multi-Echelon Techniques. Second Edition. Springer US.

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. ЛеФевевер Л. (). Искусство объяснять. Манн, Иванов и Фербер.

По Разделу 5:

1. Croll A., Yoskovitz B. Lean Analytics: Use Data to Build a Better Startup Faster. New York: O'Reilly Media Inc, 2013.
2. Koller T., Goedhart M., Wessels D. Valuation: Measuring and Managing the Value of Companies. New York, John Wiley & Sons, 2015.
3. McKinsey & Co. Big Data, Analytics, and the Future of Marketing & Sales. 2015.
4. Porter, M. Competitive Strategy: Techniques for Analyzing Industries and Competitors. New York: Free Press, 1980
5. Shearn M. The Investment Checklist: The Art of In-Depth Research. New York: John Wiley & Sons, 2011.
6. Соловьева О. В. МСФО и ГААП: учет и отчетность. — ИД ФБК-ПРЕСС Москва, 2003

11.3 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/>)

11.4 Программные средства

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).

12 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.
- Программная среда для работы: студенты будут в основном работать с платформой SAS Viya. Все необходимое ПО размещается в облаке, предоставляемом SAS, для доступа достаточно браузера.
- Учебные аудитории для лекционных занятий по дисциплине обеспечивают использование и демонстрацию тематических иллюстраций, соответствующих программе дисциплины в составе:
- ПЭВМ с доступом в Интернет (операционная система, офисные программы, антивирусные программы);
- мультимедийный проектор с дистанционным управлением.
- Учебные аудитории для лабораторных и самостоятельных занятий по дисциплине оснащены компьютерами, с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.
- Можем проводить занятия дистанционно.