

Программа учебной дисциплины
«Анализ больших данных в экономико-социологических исследованиях»

Утверждена
Академическим советом ОП
Протокол № 4 от 26.06.2019 г.

Разработчик	Управителей Филипп Александрович, Старший преподаватель, Департамент больших данных и информационного поиска
Число кредитов	6
Контактная работа (час.)	56
Самостоятельная работа (час.)	172
Курс, Образовательная программа	2 (М) курс, Прикладные методы социального анализа рынков
Формат изучения дисциплины	Без использования онлайн курса

1. Цель, результаты освоения дисциплины и пререквизиты

Цели:

1. Дать студентам представление об основных задачах и инструментах, стоящих перед аналитиками в современных бизнес-компаниях.
2. Показать студентам возможные перспективы профессионального развития и заложить необходимые базовые навыки.

Планируемые результаты обучения (ПРО):

1. Студенты понимают цели и задачи, которые стоят перед аналитиками в бизнес-компаниях, преимущественно IT-сферы.
2. Студенты знают разные метрики активности и вовлечения пользователей, понимают их разницу, способны высчитать метрики или нарисовать соответствующие графики.
3. Студенты понимают, как высчитываются монетизационные показатели проекта, способны различать различные модели монетизации, могут оценить и проинтерпретировать lifetime value пользователя.
4. Студенты понимают, как проводить АБ-тесты для новых элементов продукта: способны написать дизайн исследования, рассчитать выборку и проинтерпретировать результаты.
5. Студенты понимают основные модели привлечения пользователей продукта через рекламные кампании, способны оценивать и предсказывать эффективность рекламных кампаний.

Пререквизиты:

1. Для успешной работы на курсе студенты должны знать какой-либо из языков программирования, используемых для анализа данных (предпочительно R, возможно Python или прочие языки).
2. Студенты должны иметь базовые знания по математической статистики и методам проверки эмпирических гипотез.

2. Содержание учебной дисциплины

Тема (раздел дисциплины)	Объем в часах	Планируемые результаты обучения (ПРО), подлежащие контролю	Формы контроля
	лк		
	см		
	онл/ср		
Введение в цели и задачи продуктовой аналитики	2	<ul style="list-style-type: none"> Студенты понимают цели и задачи, которые стоят перед аналитиками в бизнес-компаниях, преимущественно IT-сферы. 	КР1.
	0		
	10		
Метрики активности и вовлечения пользователей	2	<ul style="list-style-type: none"> Студенты знают разные метрики активности и вовлечения пользователей, понимают их разницу, способны высчитать метрики или нарисовать соответствующие графики. 	КР1.
	10		
	42		
Основы юнит-экономики и метрики монетизации пользователей	2	<ul style="list-style-type: none"> Студенты понимают, как высчитываются монетизационные показатели проекта, способны различать различные модели монетизации, могут оценить и проинтерпретировать lifetime value пользователя. 	ДЗ1, КР2.
	12		
	52		
Создание и тестирование фич (feature).	4	<ul style="list-style-type: none"> Студенты понимают, как проводить АБ-тесты для новых элементов продукта: способны написать дизайн исследования, рассчитать выборку и проинтерпретировать результаты. 	ДЗ2.
	18		
	52		
Основы маркетинговой аналитики	2	<ul style="list-style-type: none"> Студенты понимают основные модели привлечения пользователей продукта через рекламные кампании, способны оценивать и предсказывать эффективность рекламных кампаний. 	•
	4		
	16		
Часов по видам учебных занятий:	12		
	44		
	172		

Содержание разделов дисциплины:

- 1. Введение в цели и задачи продуктовой аналитики**
Лекция: Направления деятельности аналитиков: веб-аналитика, маркетинговая аналитика, продуктовая аналитика, анализ данных и машинное обучение. Основные задачи продуктовых аналитиков. Процессы и роль аналитика в продукте. Жизненный цикл продукта. Роль аналитиков на каждой этапе.
- 2. Метрики активности и вовлечения пользователей**
Лекция: Оценка активности пользователей - DAU, MAU. Методы оценки удержания (retention) и оттока (churn), проблема операционализации оттока. Rolling retention, недостатки метода. Retention rate. Основные ошибки при оценке удержания. Практикумы: Подключение к базам данных. PostgreSQL, основы языка SQL. Визуализация и интерпретация DAU/MAU, значений удержания. Предсказание оттока пользователей: конструирование и отбор переменных, преобразование шкал. Качество прогностической модели.
- 3. Основы юнит-экономики и метрики монетизации пользователей**
Лекция: Что такое юнит-экономика. Конверсия в целевое действие, скорость конверсии. ARPU, ARPPU, ARPDAU, структура прибыли. ROI. Основные ошибки при вычислении монетизационных метрик. Lifetime value (LTV), модели Pareto/NBD, BG/NBD, BG/BB. RFM-анализ. Понятие бизнес-модели. Практикумы: Вычисление значений конверсии. Расчет LTV. Продуктовые выводы и рекомендации на основе монетизационных метрик. Предсказание классов платящих и неплатящих пользователей. Деревья решений, случайный лес. PMML-спецификация.
- 4. Создание и тестирование фич (feature).**
Лекции: Понятие фичи в разработке IT-продуктов. Понятие Unique selling proposition (USP). Задачи аналитиков при разработке фич. А/Б-тесты. Парадигмы А/Б-тестов (частотные тесты, sequential-подход, многорукие бандиты). Частые ошибки в тестировании. Размер выборки. Практикумы: Анализ ключевых фич популярных it-продуктов (Яндекс.Музыка, Авито, Twitter, Aviasales, или другие продукты на выбор студентов). Обсуждение возможных новых фич. Постановка гипотезы и построение дизайна А/Б-теста новой фичи. Определение размера выборки. Анализ и интерпретация результатов теста.
- 5. Основы маркетинговой аналитики**
Исследования рынка. Модели привлечения пользователей в IT-продукты. Рекламные площадки (Facebook Ads, Google Adwords, прочие.). Проблема атрибуции. Оценка эффективности рекламных кампаний, бенчмарки. Практикумы: Кривые LTV. Оценка окупаемости кампании на 90 день, прогнозирование с помощью линейных моделей.

3. Оценивание

- **КР1**, Не блокирующее, Контрольная работа
Порядок проведения: не более 1 академического часа очно, во время семинарских занятий. Контрольная работа состоит из нескольких вариантов, включающих в себя задачи на сбор данных из удаленных баз данных и отрисовку базовых графиков метрик активности и вовлечения пользователей. Работа сдается в виде скрипта на R (Python или другой ЯП - по договоренности со студентом)
- **ДЗ1**, Не блокирующее, Домашнее задание
Задача на сегментацию пользователей с помощью деревьев решений. Интерпретация результатов и формирование бизнес-рекомендаций. Работа сдается в виде R-markdown-файла (или же любой другой формат, предполагающий использование языков программирования, по договоренности со студентом).
- **КР2**, Не блокирующее, Контрольная работа
Порядок проведения: не более 1 академического часа очно, во время семинарских занятий. Контрольная работа состоит из нескольких вариантов, включающих в себя задачи на сбор данных из удаленных баз данных, расчет и интерпретация основных метрик платежного поведения пользователей. Работа сдается в виде скрипта на R (Python или другой ЯП - по договоренности со студентом)
- **ДЗ2**, Не блокирующее, Домашнее задание
Задача на написание дизайна АБ-теста, анализ и интерпретацию результатов уже проведенных тестов (или модельных данных). Работа сдается в виде R-markdown-файла (или же любой другой формат, предполагающий использование языков программирования, по договоренности со студентом).

Формула округления:

округление вверх, в пользу студента

Шкала оценки: Десятибалльная

Вид формулы оценивания: Линейная

Формула оценивания:

Окончательная оценка = Округление($1/4 * КР1 + 1/4 * ДЗ1 + 1/4 * КР2 + 1/4 * ДЗ2$)

4. Примеры оценочных средств

1. Импортируйте данные по заходам пользователей в приложение и постройте интерактивный графики DAU и MAU.
2. Импортируйте данные по платежам и оцените скорость конверсии в платящих пользователей, а так же метрики ARPU/ARPPU.

3. С помощью алгоритмов дерева решений создайте набор правил выдачи скидок на покупку разным сегментам пользователей.
4. Расчитайте, сколько потребуется привлечь пользователей для оценки эффективности фиши на удержание пользователей, из предположения, что внедряемая фиша повысит удержание на седьмой день с 7,5% до 9%.

5. Ресурсы

5.1. Рекомендуемая основная литература

п/п	Наименование
1	K. Rosamont Churn Analysis – Part 1: Model Selection
2	Г. Джеймс, Д. Уиттон, Т. Хастис, Р. Тибуширани Введение в статистическое обучение с примерами на языке R, перевод с английского С.Э. Мостицкого. — Москва : ДМК Пресс, 2017. — 456 с. — ISBN 978-5-97060-495-3.
3	D. Birant Data Mining Using RFM Analysis
4	S. Gupta, D. Hanssens, B. Hardie, W. Kahn, V. Kumar, N. Lin, N. Ravishanker, S. Modelin Modeling Customer Lifetime Value
5	П.Нестеров Байесовские многорукие бандиты против А/В тестов

5.2. Рекомендуемая дополнительная литература

п/п	Наименование
1	Е. Гильманов Процессы и место аналитика в продукте
2	А. Долотов Почти уволен. Как я построил отдел аналитики Яндекса
3	А. Кравченко Открытый курс машинного обучения. Тема 6. Построение и отбор признаков
4	Ultimate Guide to Unit Economics
5	А.Минин Юнит экономика. ч.1: история появления
6	D. Chen, S. Laing Sain, K Guo Data mining for the online retail industry: A case study of RFM model-based customer segmentation using data mining
7	E. Wadsworth Buy'Til You Die-A Walkthrough.
8	Ю.Кашицкий Открытый курс машинного обучения. Тема 3. Классификация, деревья решений и метод ближайших соседей

5.3. Программное обеспечение

п/п	Наименование	Условия доступа/скачивания
1	Microsoft Windows 7 Professional RUS Microsoft Windows 8.1 Professional RUS Microsoft Windows 10	<i>Из внутренней сети университета (договор)</i>
2	Microsoft Office Professional Plus 2010	<i>Из внутренней сети университета (договор)</i>
3	R programming language	https://cran.r-project.org/

4	RStudio IDE	https://www.rstudio.com/products/rstudio/download/
---	-------------	---

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

п/п	Наименование	Условия доступа/скачивания
	<i>Профессиональные базы данных, информационно-справочные системы</i>	
1	Электронно-библиотечная система Юрайт	URL: https://biblio-online.ru/
	<i>Интернет-ресурсы (электронные образовательные ресурсы)</i>	
1	Открытое образование	URL: https://openedu.ru/

5.5. Материально-техническое обеспечение дисциплины

Учебные аудитории для лекционных 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 семинарских и самостоятельных занятий по дисциплине оснащены ПЭВМ, с возможностью подключения к сети Интернет и доступом к электронной информационно-образовательной среде НИУ ВШЭ.

Компьютерные классы оборудованы ПЭВМ с доступом в Интернет, операционными системами и программным обеспечением, необходимыми для освоения дисциплины. При необходимости допускается замена оборудования его виртуальными аналогами.

6. Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося) а для инвалидов также в соответствии с индивидуальной программой реабилитации инвалида, могут предлагаться следующие варианты восприятия учебной информации с учетом их индивидуальных психофизических особенностей, в том числе с применением электронного обучения и дистанционных технологий:

6.1.1. для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; в форме аудиофайла (перевод учебных материалов в аудиоформат); в печатной форме на языке Брайля; индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания и консультации.

6.1.2. для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; видеоматериалы с субтитрами; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания и консультации.

6.1.3. для лиц с нарушениями опорно-двигательного аппарата: в печатной

форме; в форме электронного документа; в форме аудиофайла; индивидуальные задания и консультации.