



АНАЛИЗ ДАННЫХ

УНИКАЛЬНЫЙ КУРС ЭКОНОМИЧЕСКОГО ФАКУЛЬТЕТА МГУ

СИДОРЕНКО В.Н., ТИЩЕНКО С.А.
МГУ ИМ. М.В. ЛОМОНОСОВА
МОСКВА – 2019

Авторы



Сидоренко В.Н.

кандидат экономических наук с 1998 года
кандидат физико-математических наук с 2000 года
кандидат юридических наук с 2003 года
медаль РАН для молодых ученых
автор 28 книг и 43 учебных курсов



Тищенко С.А

консультант Министерства устойчивого развития (MEDDE) -
Франция
приглашенный профессор университета Ямагучи - Япония
обладатель премии Карла Менгера Американского
математического общества
победитель 41-го конкурса научных работ молодых ученых
МГУ

Анализ данных

Уровень высшего образования: **бакалавриат**

Направление подготовки: **менеджмент**

Статус дисциплины: **обязательная**

Семестр: 4

Объем дисциплины составляет **5 зачетных единиц**, из которых 30 часов составляют лекции, 46 - семинары

Кейсы:

- Алгоритмы MapReduce, **Hadoop** и работа с Большими Данными
- Применение методов **Data mining** в бизнесе

Кейсы:

- **Машинное обучение и нейронные сети**
- **Проверка гипотез, экспериментальный дизайн и тестирование экономических решений в бизнесе**

Деловые игры:

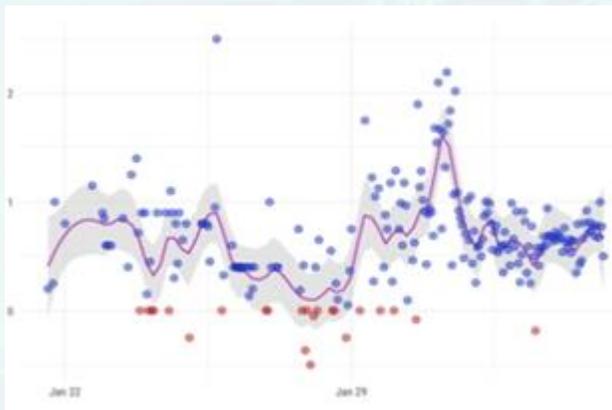
- Внедрение технологии **Блокчейн** в корпорации
- **Цифровая экономика** и трансформация рынка труда
- Корпоративные данные и информационная безопасность

Теоретические основы

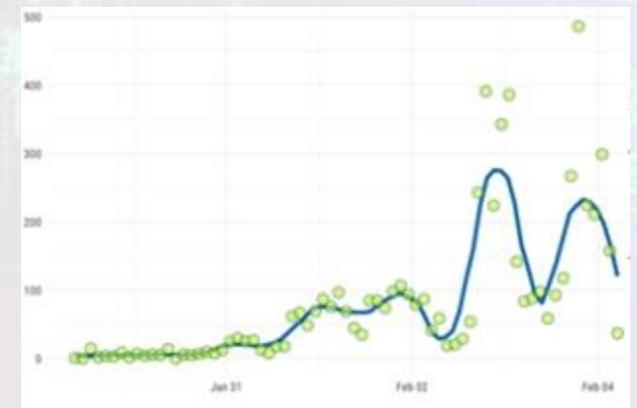
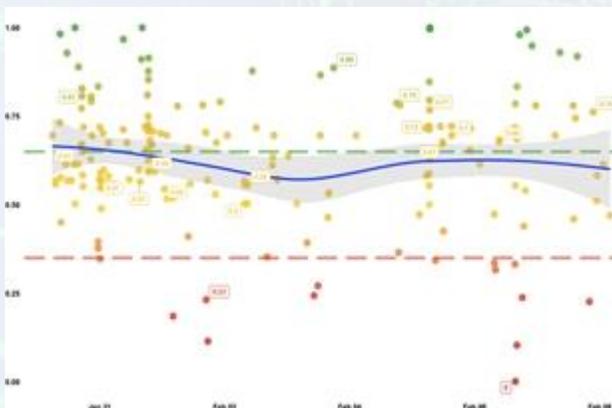
- Особенности работы с данными в конкурентной среде
- Факторный анализ и фундаментальная теорема Тэрстоуна
- Методы кластерного анализа
- Экспериментальный дизайн и тестирование экономических решений

The background of the slide is a light blue gradient with a faint, semi-transparent overlay of various financial data visualizations. These include line graphs with data points, bar charts, and numerical values in different colors (green, red, black) and fonts, suggesting a data analysis or financial market theme.

Методология решения задач анализа данных в среде R



Создание визуальных и графических образов в процессе комплексного подхода анализа данных



The background of the slide is a light blue gradient overlaid with a complex network of faint, semi-transparent white lines and data points, resembling a financial or economic data visualization. There are several line graphs with data points and various numerical values scattered across the background, such as '53.398', '59.908', '57.546', and '57.546'.

**Межкафедральная взаимосвязь с
другими курсами бакалавриата
Экономического факультета МГУ**

Тема 1. Введение в анализ данных и особенности работы с данными в конкурентной среде (9 часов)

Введение в анализ данных. Жизненный цикл данных. Основные модели структур данных. Базы данных. Конкурентные данные. Особенности работы с данными в конкурентной среде. Поиск, создание, организация и управление данными коммерческой организации. Работа с информацией внутри компании с целью повышения инновационной активности. Элементы алгоритмики для менеджеров. Алгоритмы, как технология. Технические и программные средства реализации информационных процессов, модели решения функциональных и алгоритмических задач, Асимптотические обозначения. Примеры алгоритмов поиска и сортировки данных.

Тема 2. Алгоритмы MapReduce, Hadoop и работа с Большими Данными (15 часов)

Рассматриваются современные методы и модели обработки больших данных в крупномасштабных системах, в том числе и экономических системах. Приведены алгоритмы обработки данных, на основе которых можно анализировать и прогнозировать конъюнктуру рынка. Методы работы с Big Data и Data mining в управлении и экономике. Алгоритмы MapReduce, Hadoop. В рамках семинарских занятий студентами выполняется кейс иллюстрирующий эффективность алгоритма Hadoop при решении задач на массивах больших данных.

Тема 3. Машинное обучение и нейронные сети (17 часов)

Теория и практика машинного обучения. Вопросы, связанные с анализом данных: модели, алгоритмы, методы и их реализация. Анализ временных рядов. Способы применения вычислительной мощности современной компьютерной техники при решении управленческих задач. Нейронные сети. Machine learning. Анализ обучающейся сети. Использование вычислительной мощности современной компьютерной техники при решении управленческих задач. Методы анализа больших данных с применением машинного обучения. В рамках семинарских занятий студентами выполняется кейс, иллюстрирующий актуальность применения машинного обучения в процессе оптимизации управленческих процессов компании.

Тема 4. Применение методов Data mining в бизнесе (8 часов)

Вопросы, связанные с набирающей популярность областью интеллектуального анализа данных - Data Mining. Изучаются основные технологические тренды, сопровождающие Data Mining, вопросы терминологии. Рассматриваются основные методы и инструменты Data Mining, связанные с высокопроизводительной интеллектуальной аналитической обработкой данных, направленной на то, чтобы оперативно извлекать из значительных массивов, накопленных и поступающих данных ценные экспертные знания, поддерживая эффективную управленческую деятельность. В рамках семинарских занятий студентами выполняется кейс, иллюстрирующий актуальность применения Data Mining в процессе выработки конкурентной стратегии компании.

Тема 5. Методология решения задач анализа данных в среде R и/или Python (26 часов)

Специализированные статистические программы. Применение, преимущества и недостатки R и/или Python. Как скачать и установить R и/или Python. Как начать работать с R и/или Python. Графический интерфейс для R и/или Python. Как загружать данные. R и/или Python как калькулятор. Типы данных. Основы программирования в R и/или Python. Анализ связей. Анализ структуры.

Тема 6. Создание визуальных и графических образов в процессе комплексного подхода анализа данных (22 часа)

Возможности языков программирования R при графической и сетевой визуализации данных. Создание визуальных и графических образов в процессе комплексного подхода в управлении бизнесом. Визуализация данных эффективным образом, создание визуального или графического образа в процессе комплексного подхода, который объединяет анализ данных и визуализацию данных с использованием возможностей языков R. Методики принятия управленческих решений на основании анализа больших объемов данных.

Тема 7. Факторный анализ (16 часов)

Задачи и методы многомерного статистического анализа в экономике. Теоретические основы многомерного статистического анализа. Особенности обработки многомерных статистических данных. Сущность методов факторного анализа. Фундаментальная теорема факторного анализа Тэрстоуна. Общий алгоритм и теоретические проблемы факторного анализа. Метод главных компонент. Метод главных факторов. Факторный анализ соответствия. Кластерный анализ. Общая характеристика методов кластерного анализа. Меры сходства. Иерархический кластерный анализ. Дискриминантный анализ.

Тема 8. Экспериментальный дизайн и тестирование экономических решений в бизнесе (15 часов)

Сценарный анализ при принятии решений в бизнесе. Экспериментальный дизайн и тестирование экономических решений в бизнесе. Выполнение при управлении компанией бизнес-экспериментов, экспериментальный дизайн и тестирование. Методы проверки стратегических гипотез, эффективности новой упаковки продукта или маркетингового подхода и др., позволяющие апробировать что-либо в некоторой части организации с последующим обобщением на всю организацию. В рамках семинарских занятий студентами выполняется кейс, иллюстрирующий актуальность применения экспериментального дизайна в процессе разработки организационной стратегии компании.

Тема 9. Анализ числовой информации (28 часов)

В рамках семинарских занятий студенты решают задачи на анализ числовой информации. Тем самым проходят подготовку к, распространенным при приеме на работу, числовым тестам на анализ числовой информации.

Тема 10. Функции и методы управления подразделениями, группами сотрудников при переходе предприятия к цифровой экономике и внедрению современных методов анализа конкурентных данных (20 часов)

В рамках учебного курса студенты принимают участие в трех деловых играх: "Внедрение технологии блокчейн в корпорации", "Цифровая экономика и трансформация рынка труда" и "Корпоративные данные и информационная безопасность". В процессе деловых ролевых игр студенты знакомятся с особенностями процессов внедрения методов цифровой экономики в бизнесе.

Общая нагрузка

30 пар лекций и 368 пар семинаров

Материально-техническое обеспечение дисциплины

- Текстовый редактор Emacs (или любой аналогичный текстовый редактор) - GNU GPL - <http://www.gnu.org/software/emacs/>
- Deductor Academic – <https://basegroup.ru/deductor/description>
- Компилятор R – GNU - <http://www.r-project.org>, интегрированная среда разработки RStudio – GNU AGP - <http://www.rstudio.com/ide/>
- Среда разработки Anaconda для Python/R – <https://www.anaconda.com>
- Anylogic PLE – <https://www.anylogic.ru>

The background features a complex overlay of financial data, including line graphs, bar charts, and various numerical values in different colors (red, green, blue, yellow). The overall aesthetic is that of a busy trading floor or a data visualization dashboard.

Спасибо!

v_sidorenko@mail.ru

tichtch@mail.ru