

АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ ДИСЦИПЛИНЫ Б1В.02 «Анализ больших данных»

Автор: Кирилл Шилин, доктор технических наук, профессор кафедры эконометрии и мат.методов ЭМИТ РАНХиГС

Код и наименование направления подготовки, профиля: 38.04.02 Менеджмент
«Прорывные финансовые и цифровые технологии»

Квалификация (степень) выпускника: Магистр

Форма обучения: очная

Цель освоения дисциплины:

Сформировать следующие компетенции:

ПК-4 – способность управлять организациями, подразделениями, группами (командами) сотрудников, проектными сетями; ПК ОС-13 - способность принимать управленческие решения на основе анализа структурированных и неструктурированных данных; ОК-1 - способность к абстрактному мышлению, анализу и синтезу.

План курса:

Тема 1. Основы анализа данных. Системы поддержки принятия решений.

Сравнительный анализ OLAP и OLTP систем. Сферы их применения. Отличие информационного хранилища от баз данных. Принципы построения информационных хранилищ. Классификация информационных хранилищ. Модели информационных хранилищ. Многомерная модель данных. Нормальная форма. Денормализация моделей данных. Правила Кодда. Зачем применяется денормализация моделей? Размерностные модели. Отличие таблицы фактов от размерностной таблицы. Характеристика стандартов DataMining.

Тема 2. Проработка и очистка данных.

Характеристика этапа ETL (Extracting Transforming and Loading). Задачи, решаемые DataMining. Предназначение и средства разведочный анализ данных. Характеристика диаграммы «ящик с усами». Операции, выполняемые при агрегировании данных. Примеры использования статистических пакетов для разведочного анализа. Как выполняется сравнительный анализ графических средств анализа? Характеристика биржевых диаграмм. Использование диаграммы рассеяния.

Тема 3. Классификационный анализ без обучения. Кластерный анализ.

Основные характеристики кластера, «центроид» кластера, классификация методов кластерного анализа, примеры их применения в практической жизни. Меры близости, методы определения близости между кластерами, применение метода ближнего соседа, дальнего соседа, их сравнение. Характеристика метрик кластерного анализа. Содержание

«дендограммы» и организацию ее применения. Использование статистических пакетов для решения задач кластерного анализа.

Тема 4. Анализ взаимосвязей между переменными, ассоциативные правила

Задачи использования ассоциативных правил, примеры задач использования ассоциативных правил, обобщенных правил. Понятие транзакции, примеры применения транзакций. Показатели, используемые для построения правила, алгоритмы построения ассоциативных правил, алгоритм *apriori*. Общая характеристика пакета *Deductor*. Использование пакета *Deductor* для решения задач интеллектуального анализа данных.

Тема 5. Классификационный анализ с обучением

Определение задачи классификации, методы решения задачи классификации, особенности решения задач классификации с обучением. Деревья классификации и их свойства, примеры алгоритмы построения деревьев. Алгоритм *CART*, примеры его использования. Что такое искусственный нейрон? Его состав, классификация решающих функций нейрона, их сравнительный анализ, искусственные нейронные сети. Что понимается под архитектурой нейронных сетей?

В результате освоения дисциплины у студентов должны быть сформированы:

Код компетенции	Наименование компетенции	Код этапа освоения компетенции	Наименование этапа освоения компетенции
ПК-4	способность управлять организациями, подразделениями, группами (командами) сотрудников, проектными сетями	ПК-4.1	способен управлять организациями, подразделениями, группами (командами) сотрудников, проектными сетями
ПК ОС-13	способность принимать управленческие решения на основе анализа структурированных и неструктурированных данных	ПК ОС-13.1	способен принимать управленческие решения на основе анализа структурированных и неструктурированных данных
ОК-1	способность к абстрактному мышлению, анализу и синтезу	ОК-1.1	обладает абстрактным мышлением, анализом и синтезом

Формы и методы текущего контроля и промежуточной аттестации:

В ходе реализации дисциплины используются следующие методы текущего контроля и успеваемости обучающихся:

Опрос, тест, диспут, итоговый зачет

Основная литература:

1. Гмурман В.Е. Теория вероятностей и математическая статистика. Учебное пособие для бакалавров. - М.: Юрайт: ИД Юрайт, 2012. – 479 с.
2. Гусева Е.Н. Теория вероятностей и математическая статистика: учебное пособие. - Издательство: Флинта, 2011 г. – 220 с.- <http://rucont.ru/efd/246481>
3. Волкова П.А., Шипунов А.Б. Статистическая обработка данных в учебно-исследовательских работах. – М.: Форум, 2012. – 96 с.
4. Кремер Н.Ш. Теория вероятностей и математическая статистика: учебник. -Издательство: ЮНИТИ-ДАНА, 2012. – 551 с.
5. Плохотникова К.Э. Основы эконометрики в пакете Statistica: учебное пособие для вузов. – М.: Вузовский учебник, 2011. – 297 с.
6. Теория вероятностей и математическая статистика: учебное пособие для студентов вузов / под ред. В.И. Ермакова. – М.: ИНФРА-М, 2012. – 287 с.
7. Шумак О.А., Гераськин А.В. Статистика: учебное пособие для студентов вузов. – М.: РИОР: ИНФРА-М, 2012. – 311 с.
8. Вуколов Э.А. Основы статистического анализа. Практикум по статистическим методам и исследованию операций с использованием пакетов STATISTICA и EXCEL: учебное пособие для студентов вузов. – М.: ФОРУМ: ИНФРА-М, 2013. – 464 с.
9. Волкова П.А., Шипунов А.Б. Статистическая обработка данных в учебно исследовательских работах. – М.: Форум, 2012. – 96 с.
10. Зарядов И. С. Введение в статистический пакет R: типы переменных, структуры данных, чтение и запись информации, графика. М.: Издательство Российского университета дружбы народов, 2010. - 207 с.
11. Зарядов И. С. Статистический пакет R: теория вероятностей и математическая статистика. М.: Издательство Российского университета дружбы народов, 2010. - 141 с.
12. Зорин А. В., Федоткин М. А. Введение в прикладной статистический анализ в пакете R. Учебно-методическое пособие. Нижний Новгород: Нижегородский гос. ун-т им. Н. И. Лобачевского, 2010. - 50 с.
13. Шипунов А.Б., Балдин Е.М., Волкова П.А., Коробейников А.И., Назарова С.А., Петров С.В., Суфиянов В.Г. Наглядная статистика. Используем R! - М.: ДМК Пресс, 2012. - 298 с