

Сибирский институт управления – филиал РАНХиГС

Факультет экономики и финансов

Кафедра информатики и математики

**АННОТАЦИЯ РАБОЧЕЙ ПРОГРАММЫ
ДИСЦИПЛИНЫ**

АНАЛИЗ И ВИЗУАЛИЗАЦИЯ ДАННЫХ

Б1.В.05

по направлению подготовки 38.03.01 Экономика

направленность (профиль): «Государственные и муниципальные финансы»

квалификация: Бакалавр

формы обучения: очная, очно-заочная, заочная

Год набора — 2021

Автор–составитель:

старший преподаватель кафедры «Бухгалтерский учет, экономический анализ и аудит»
РАНХиГС Баркова Е.А.

кандидат физико-математических наук, доцент, заведующий кафедрой информатики и
математики СИУ – филиала РАНХиГС Е.А. Рапоцевич

Новосибирск, 2020

Цель освоения дисциплины:

Сформировать компетенцию, необходимую при осуществлении профессиональной деятельности – способен применять информационные технологии для решения прикладных финансово-экономических задач.

План курса:

Тема 1. Основы анализа данных и машинного обучения

Тема осваивается с применением дистанционных образовательных технологий (размещается на платформе СДО РАНХиГС)

Раздел 1. Введение в теорию анализа данных.

Основы обработки данных. Основы алгоритмизации. Существующие наборы данных, визуализация модели классификации. Объекты и признаки. Типы шкал. Показатели вариации. Линейные и нелинейные модели регрессии.

Раздел 2. Практика анализа данных и машинного обучения

Понижение размерности. Модель алгоритмов машинного обучения. Классификаторы: виды и принципы построения. Методы оценки точности решений. Визуализация.

Тема 2. Подготовка данных к анализу

Очистка и предобработка данных.

Оценка качества данных. Фильтрация данных. Обработка дубликатов и противоречий. Выявление аномальных значений. Восстановление пропущенных значений. Сокращение размерности. Сэмплинг.

Трансформация данных.

Группировка данных. Слияние данных. Квантование. Нормализация и кодирование.

Тема 3. Визуализация данных

Введение в визуализацию. Визуализаторы общего назначения. OLAP-анализ. Визуализаторы для оценки качества моделей. Визуализаторы, применяемые для интерпретации результатов анализа.

Визуализация данных средствами MS Power BI.

Семейство MS Power BI. Конструктор кубов и аналитических представлений. Совместная работа с представлениями – серверные сервисы Power BI. Решение задач объединения данных из разных источников, вычисляемые поля и Меры, элементы языка DAX, построение различных визуальных представлений по данным из учетных систем и CRM-систем.

Тема 4. Классификация и регрессия (Статистические методы)

Введение в классификацию и регрессию. Множественная линейная регрессия. Регрессия с категориальными входными переменными. Методы отбора переменных в регрессионные модели. Ограничения применимости регрессионных моделей. Нелинейные регрессионные модели. Основы логистической регрессии. Множественная логистическая регрессия. Оценка эффективности и сравнение моделей. ROC-анализ. Оценка ошибки модели. Издержки ошибочной классификации. LIFT- и PROFIT-кривые.

Тема 5. Классификация и регрессия (Машинное обучение)

Введение в деревья решений. Алгоритмы построения деревьев решений. Упрощение деревьев решений. Введение в нейронные сети. Принципы построения нейронных сетей. Процесс обучения нейронной сети. Обучение в условиях несбалансированности классов.

Тема 6. Кластеризация

Введение в кластеризацию. Меры близости, используемые в алгоритмах кластеризации. Методы кластерного анализа. Иерархическая кластеризация. Алгоритм кластеризации k-means. Сети Кохонена. Карты Кохонена.

Тема 7. Анализ и прогнозирование временных рядов

Введение в прогнозирование. Временной ряд и его компоненты. Модели прогнозирования. Исследование наборов данных и выбор метода прогнозирования. Ошибки прогноза. Оценка адекватности выбранного метода прогнозирования. Методы сглаживания и скользящие средние.

Декомпозиция временного ряда. Регрессионный анализ временных рядов.

Формы текущего контроля и промежуточной аттестации

Тема (раздел)	Методы текущего контроля успеваемости
Основы анализа данных и машинного обучения	Онлайн тестирование, Баллы О-Л курса
Подготовка данных к анализу	Опрос Практико-ориентированное задание (компьютерный практикум)
Визуализация данных	Практико-ориентированное задание (компьютерный практикум) Практическая контрольная работа
Классификация и регрессия (Статистические методы)	Опрос Практико-ориентированное задание (компьютерный практикум)
Классификация и регрессия (Машинное обучение)	Тестирование Практико-ориентированное задание (компьютерный практикум) Практическая контрольная работа
Кластеризация	Опрос Практико-ориентированное задание (компьютерный практикум)
Анализ и прогнозирование временных рядов	Опрос Практико-ориентированное задание (компьютерный практикум)

Промежуточная аттестация проводится в форме зачета с оценкой.

Основная литература:

1. Анализ данных: учебник для вузов / В. С. Мхитарян [и др.] ; под редакцией В. С. Мхитаряна. — Москва: Издательство Юрайт, 2020. — 490 с. — (Высшее образование). — ISBN 978-5-534-00616-2. — Текст : электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/450166>
2. Билл, Фрэнкс Революция в аналитике: как в эпоху Big Data улучшить ваш бизнес с помощью операционной аналитики / Фрэнкс Билл; перевод И. Евстигнеева; под редакцией В. Мылов. — Москва: Альпина Паблишер, 2017. — 320 с. — ISBN 978-5-9614-5302-7. — Текст: электронный // Электронно-библиотечная система IPR BOOKS : [сайт]. — URL: <http://www.iprbookshop.ru/58563.html>