УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«УНИВЕРСИТЕТ УПРАВЛЕНИЯ «ТИСБИ»

Кафедра бизнес - аналитика

«Утверждаю»
Зав. кафедрой
А.М. Найда
Протокол заседания
кафедры № 7
от « 26 » февраля 2026 г.

Рабочая программа дисциплины

Наименование дисциплины	Основы анализа больших данных
Направление подготовки	38.04.02 Менеджмент
Профиль подготовки	Менеджмент устойчивого развития
Год набора	2025, 2026

Составитель:
канд. хим. наук, доц. Е.И. Уткина
асс. Кочешкова Э.Д.

Казань

Содержание

1. Цели и задачи учебной дисциплины 3
2. Место дисциплины в структуре ОПОП 4
3. Требования к результатам освоения дисциплины 5
4. Структура и содержание дисциплины 6
4.1. Модульно-тематический план и пояснительная записка с указанием этапов формирования компетенции 6
4.2. Содержание дисциплины по темам (разделам) 8
4.3. Планы практических и семинарских занятий 9
5. Учебно-методическое обеспечение самостоятельной работы студентов 13
6. Учебно-методическое и информационное обеспечение дисциплины 15
7. Материально-техническое обеспечение дисциплины 16
8. Оценка компетенций по изучаемой дисциплине 16
Приложение 1. Методические указания для обучающихся по освоению дисциплины
Приложение 2. Фонд оценочных средств для проведения текущей и промежуточной аттестации по дисциплине

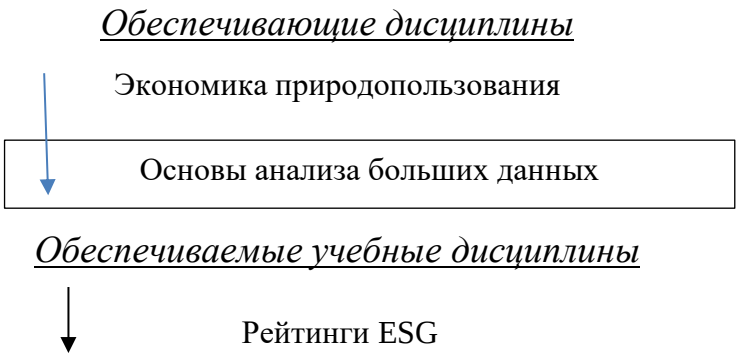
1. Цели и задачи учебной дисциплины

Целью освоения дисциплины «Основы анализа больших данных» является формирование у студентов целостного представления о технологиях, методах и инструментах для работы с большими объемами информации, а также получение практических навыков по поиску, структурированию, обработке и интеллектуальному анализу данных для извлечения полезных закономерностей и принятия обоснованных решений.

Задачами дисциплины являются: изучение теоретических основ работы с большими данными, освоение современных математических, статистических и программных средств для их обработки и анализа, а также развитие практических навыков поиска информации, выявления закономерностей, построения моделей и принятия обоснованных решений на основе полученных результатов.

2. Место дисциплины в структуре ОПОП

Данная дисциплина относится к обязательной части Блока 1 учебного плана. До начала изучения дисциплины «Основы анализа больших данных» развитием у студента должны быть сформированы компоненты компетенций (ЗУВы) полученных в результате изучения дисциплин: Экологический менеджмент, Экономика природопользования. Дисциплина находится во взаимосвязи с дисциплинами согласно схеме:



3. Требования к результатам освоения дисциплины

Дисциплина участвует в формировании следующих компетенций в соответствии с ФГОС ВО по направлению подготовки 38.04.02 Менеджмент, профиль «Менеджмент устойчивого развития»:

ОПК-2. способен применять современные техники и методики сбора данных, продвинутые методы их обработки и анализа, в том числе использовать интеллектуальные информационно-аналитические системы, при решении управленческих и исследовательских задач

После освоения дисциплины студент должен получить следующие образовательные результаты, соотнесённые с индикаторами достижения компетенций.

Декомпозиция компетенций

Индикатор	**Результаты обучения по дисциплине**
Компетенция ОПК-2	
ОПК-2.1 Применяет современные техники и методики сбора данных, продвинутые методы их обработки и анализа при решении управленческих и исследовательских задач	ОПК-2.1 З.6 Знает основные понятия и терминологию в области технологий обработки и анализа больших данных ОПК-2.1 У.9 Умеет разрабатывать и анализировать концептуальные и теоретические модели прикладных задач анализа больших данных ОПК-2.1 В.9 Владеет навыками создания и решения моделей, необходимых в сфере управления; программными инструментами для работы с большими наборами данных; методикой прогнозирования.

Этапы формирования выбранных компетенций (или их частей – ЗУВов) можно проследить по пояснительной записке и модульно-тематическому плану дисциплины.

4. Структура и содержание дисциплины

4.1. Модульно-тематический план и пояснительная записка с указанием этапов формирования компетенций

Общая трудоемкость дисциплины составляет 2 зачетные единицы (72 академических часа).

Модульная разбивка учебной дисциплины					
Направление подготовки: «Менеджмент» профиль «Менеджмент устойчивого развития»					
Наименование модулей	Количество ауд. часов		Самостоят. работа.	Всего часов.	Индикаторы компетенции/ЗУВы
	Лекции очная/ заочная	Практ. очная/ заочная	очная/ заочная		
Модуль 1 «Введение в предмет»					
Тема 1: Введение в анализ данных и язык программирования Python	2/2	2/-	6/10	10/12	ОПК-2.1 З.6
Тема 2: Описательная статистика и визуализация данных*	4/2	4/-	6/12	14/12	
Модуль 2 «Задачи больших даннх					
Тема 3: Задача регрессии	3/-	3/2	6/12	12/14	ОПК-2.1 З.6 У.9 В.9
Тема 4: Задача кластеризации	6/-	6/4	6/12	18/16	
Подготовка к зачету			18/18	18/18	
ИТОГО	15/4	15/4	42/64	72/72	-

* Данная тема изучается с элементами интерактивных методов обучения, которые отражены в Пояснительной записке данного курса

Пояснительная записка

с этапами формирования компетенции

к модульному курсу «Основы анализа больших данных»

Данный курс состоит из 2модулей.

Модуль 1 «Введение в предмет» включает в себя 2 учебные темы.

В результате сдачи модуля студент должен:

- Знать: основные понятия и терминологию в области технологий обработки и анализа больших данных.

По результатам освоения модуля проводится устный и тестовый опрос, практическая работа, проверяется выполнение домашнего задания из раздела самостоятельная работа.

Модуль 2 «Задачи больших данных» включает 2 темы**.**

В результате сдачи модуля у студента должен

- Знать: основные понятия и терминологию в области технологий обработки и анализа больших данных

- Уметь: разрабатывать и анализировать концептуальные и теоретические модели прикладных задач анализа больших данных

- Владеть: навыками создания и решения моделей, необходимых в сфере управления; программными инструментами для работы с большими наборами данных; методикой прогнозирования..

По результатам освоения модуля проводится устный и тестовый опрос, практическая работа, проверяется выполнение домашнего задания из раздела самостоятельная работа

4.2. Содержание дисциплины по темам

Тема 1. Введение в анализ данных и язык программирования Python.

Введение в предметную область. Примеры использования методов машинного обучения для решения прикладных задач. Повторение основ программирования на языке Python.

Тема 2. Описательная статистика и визуализация данных

Знакомство с различными методами предобработки данных, описательными статистиками и основными способами визуализации данных. Работа с пропущенными значениями и выбросами. Визуализация данных с помощью Python.

Тема 3. Задача регрессии.

Метод наименьших квадратов. Измерение ошибки (метрики качества) в задачах регрессии (MSE, $RMSE$, MAE, R2). Многомерная регрессия, проблема мультиколлинеарности. Регрессия, линейная по параметрам. kNN-регрессия. Отбор и кодирование признаков. Визуальное сравнение общего и предсказанного моделью трафика. Проверка качества.

Тема 4. Задача кластеризации.

Постановка задачи кластеризации. Обзор основных методов ее решения (kNN, k- Means, иерархическая кластеризация и др.). Проблема разбиения объектов на группы по степени близости объектов в группах. Функции расстояния в пространстве объектов. Проблема выбора адекватного числа кластеров. Графическое представление алгоритмов кластеризации. Разбор примеров построения моделей кластеризации. Отбор и кодирование признаков. Метрики качества модели кластеризации. Анализ качества построенной модели

4.3. Планы семинарских и практических занятий

Методические указания

Подготовка по материалам семинарского занятия позволяет закрепить полученные путем проработки лекций знания и приобрести навыки решения практических задач. К каждому занятию предложены литературные источники, где можно найти разбираемые задания и практические работы. При проработке материала семинарского занятия необходимо обратить внимание на раздел данной программы «Самостоятельная работа», в котором также предложено выполнить задания и проверить усвоение материала лекции при помощи контрольных вопросов. *Все семинарские занятия проводятся в интерактивной форме.*

Тема 1: Введение в анализ данных и язык программирования Python.

Основные вопросы

1. загрузите данные (датасет) из файла `diamonds.csv` и создайте датафрейм с именем `df`
2. какой размер датасета?
3. какие переменные в датасете количественные, а какие категориальные? есть ли в датасете пропущенные значения и если есть, то для каких переменных?
4. найдите наибольшую и наименьшую цену бриллианта
5. найдите наибольший и наименьший вес бриллианта в каратах
6. найдите медианную цену бриллианта
7. найдите средний вес бриллианта в каратах (ответ округлите до двух десятичных знаков)

8 найдите описательные статистики для категориальных переменных из датасета

Тема 2: . Описательная статистика и визуализация данных

Основные вопросы

1. Загрузите данные (датасет). Вычислите и визуализируйте корреляционную матрицу для указанных переменных.
2.

Тема 3 . Задача регрессии

Основные вопросы

1. Для набора данных wage2 рассмотрим линейную регрессию wage на age, educ, exper. Подгоните модель и на обучающей выборке вычислите метрики подгонки: R2, MSE, MAE, MAPE, RMSE.

Тема 4. : Задача кластеризации

Основные вопросы

1. Загрузите данные (датасет) из файла `countries.csv`. Проведите z-нормировку данных. Проведите кластеризацию методом k-средних с тремя

кластерами. Вычислите метрики кластеризации Silhouette Coefficient, Calinski-Harabasz, Davies-Bouldin. Визуализируйте кластеры.

5. Учебно-методическое обеспечение самостоятельной работы студентов

Самостоятельная работа студентов является составной частью их учебной работы и имеет целью закрепление и углубление полученных знаний, умений и навыков, поиск и приобретение новых знаний.

Самостоятельная работа студентов включает в себя освоение теоретического материала на основе лекций, основной и дополнительной литературы; подготовку к семинарским занятиям в индивидуальном и групповом режиме. Советы по самостоятельной работе с точки зрения использования литературы, времени, глубины проработки темы и др., а также контроль за деятельностью студента осуществляется во время семинарских занятий.

Предлагаемый план самостоятельной подготовки:

-прочитать лекционный материал, и при необходимости, предлагаемые источники литературы, которые представлены в разделе 6 рабочей программы;

-ответить на контрольные вопросы, проверив свои знания;

-перейти к проработке материала к семинарскому занятию: ответить на вопросы, выполнить задания;

-проработать материал к семинару, предложенный в разделе самостоятельная работа: выполнить задания, закрепляющие усвоение темы.

Тема 1. Введение в анализ данных и язык программирования Python.

Введение в предметную область. Примеры использования методов машинного обучения для решения прикладных задач. Повторение основ программирования на языке Python.

Тема 2. Описательная статистика и визуализация данных

Знакомство с различными методами предобработки данных, описательными статистиками и основными способами визуализации данных. Работа с пропущенными значениями и выбросами. Визуализация данных с помощью Python.

Тема 3. Задача регрессии.

Метод наименьших квадратов. Измерение ошибки (метрики качества) в задачах регрессии (MSE, $RMSE$, MAE, R2). Многомерная регрессия, проблема мультиколлинеарности. Регрессия, линейная по параметрам. kNN-регрессия. Отбор и кодирование признаков. Визуальное сравнение общего и предсказанного моделью трафика. Проверка качества.

Тема 4. Задача кластеризации.

Постановка задачи кластеризации. Обзор основных методов ее решения (kNN, k- Means, иерархическая кластеризация и др.). Проблема разбиения объектов на группы по степени близости объектов в группах. Функции

расстояния в пространстве объектов. Проблема выбора адекватного числа кластеров. Графическое представление алгоритмов кластеризации. Разбор примеров построения моделей кластеризации. Отбор и кодирование признаков. Метрики качества модели кластеризации. Анализ качества построенной модели

6. Учебно-методическое и информационное обеспечение дисциплины

Основная литература:

1. Анализ данных: учебник для вузов / под редакцией В. С. Мхитаряна. — Москва: Издательство Юрайт, 2025. — 448 с. — (Высшее образование). — ISBN 978-5-534- 19964-2. — Текст: электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/560311

2. Миркин, Б. Г. Базовые методы анализа данных: учебник и практикум для вузов / Б. Г. Миркин. — 3-е изд., перераб. и доп. — Москва: Издательство Юрайт, 2025. — 297 с. — (Высшее образование). — ISBN 978-5-534-19709-9. — Текст: электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/560414

Дополнительная литература:

1. Чернышев, С. А. Основы программирования на Python : учебник для вузов / С. А. Чернышев. — 2-е изд., перераб. и доп. — Москва : Издательство Юрайт, 2025. — 349 с. — (Высшее образование). — ISBN 978-5-534-17139-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: https://urait.ru/bcode/567821

Интернет-ресурсы, современные профессиональные базы данных, информационно-справочные системы:

www.iprbookshop.ru – Электронно-библиотечная система IPR books

https://www.garant.ru/ - Информационно-правовой портал «Гарант»

Научная электронная библиотека КиберЛеника – www.cyberleninka.ru.

Научная электронная библиотека eLIBRARRY – www.elibrary.ru

Электронная библиотечная система ЭБС ЮРАЙТ – www.urait.ru

Банк России https://cbr.ru/ https://www.cbr.ru/statistics/macro_itm/sec_st/issue_sector/ https://www.cbr.ru/develop/ur/ https://www.cbr.ru/develop/ur/na/

Федеральная служба государственной статистики (Росстат) https://rosstat.gov.ru/

Государственная корпорация развития ВЭБ.РФ https://xn--90ab5f.xn--p1ai/ustojchivoe- razvitie/zeljonoe-finansirovanie/vypuski- finansovykh-instrumentov/

Экспертно-аналитическая платформа Infragreen https://infragreen.ru

RAEX Sustainability: вебинары https://raex-rr.com/events/?q=Вебинар&page_future=1& page_past=1 и телеграм-канал @raex_sustainability

7. Материально-техническое обеспечение дисциплины

В процессе изучения данной дисциплины используется учебная аудитория, кабинет для самостоятельной работы студентов, читальный зал, видеопроекционное оборудование, компьютер, оснащенный типовым пакетом системного и офисного ПО, в соответствии с Реестром материально-технического обеспечения аудиторного фонда Университета управления «ТИСБИ».

8. Оценка компетенций по изучаемой дисциплине

Для оценки компетентности рекомендуется использовать рейтинговую оценку знаний, умений и навыков студента по окончанию изучения каждого Модуля в соответствии с Положением о модульно-рейтинговой системе организации образовательного процесс. Итоговая оценка (в баллах) складывается из баллов, набранных по каждому Модулю (семестровая оценка) и баллов, набранных, непосредственно на экзамене (зачете).

Расчет набранных баллов по дисциплине осуществляется в следующей последовательности:

$C = \frac{M_1 + M_2 + ...M_n}{n} \times 0{,}6$, где М – количество баллов по модулю; n – количество модулей

$З = K \times 0{,}4$, где К - количество баллов на экзамене (зачете);

$И = C + З + П$, где П – поощрительные баллы (от 1 до 5).

Уровень освоения компетенций	Количество баллов
компетенции не сформированы	до 59 баллов
компетенции сформированы	от 60 до 100 баллов

Уровень сформированности компетенции, ее основные признаки и инструменты оценки приведены в табл. 8.1.

Таблица 8.1.

Оценка уровня сформированности компетенции **ОПК-2. способен применять современные техники и методики сбора данных, продвинутые методы их обработки и анализа, в том числе использовать интеллектуальные информационно-аналитические системы, при решении управленческих и исследовательских задач**

№	**Уровни сформированности компетенции**	**Основные признаки уровня**	**Инструменты оценки сформированности уровня**
1	**Пороговый уровень (как минимально допустимый)** (обязательный для всех студентов-выпускников вуза по завершении освоения ОПОП ВО) (от 60 до 70 баллов)	Знает основные понятия и терминологию в области технологий обработки и анализа больших данных	Тестирование зачет
2	**Базовый уровень (относительно порогового уровня) (**От 71 до 85 баллов**)**	Знает основные понятия и терминологию в области технологий обработки и анализа больших данных Умеет разрабатывать и анализировать концептуальные и теоретические модели прикладных задач анализа больших данных	Выступление на семинаре Тестирование зачет
3	**Повышенный уровень (относительно порогового уровня)** (От 86 до 100 баллов)	Знает основные понятия и терминологию в области технологий обработки и анализа больших данных Умеет разрабатывать и анализировать концептуальные и теоретические модели прикладных задач анализа больших данных Владеет навыками создания и решения моделей, необходимых в сфере управления; программными инструментами для работы с большими наборами данных; методикой прогнозирования	Выступление на семинаре с презентацией Тестирование зачет

Приложение 1

Методические рекомендации для обучающихся по освоению дисциплины.

Студентам на первом занятии необходимо ознакомиться с Рабочей программой дисциплины, где прописаны цели, задачи и трудоемкость дисциплины. Перед началом изучения дисциплины необходимо повторить учебный материал обеспечивающих учебных дисциплин предшествующих курсов.

Затем необходимо ознакомиться с порядком изучения дисциплины, т.е. модульно-тематическим планом и пояснительной запиской с указанием этапов формирования заявленных компетенций.

И, наконец, ознакомиться с порядком оценивания результатов обучения, для чего необходимо изучить следующие документы: Положение о модульно-рейтинговой системе оценивания и Принципы оценки уровня знаний, умений и навыков (характеристика ответа).

Студент должен внимательно изучить перечень основной (дополнительной) литературы и взять необходимые учебники в библиотеке.

При сдаче модулей упор делается на выявление основных факторов, их анализ и определения путей повышения экономической эффективности, полученных в результате анализа.

При подготовке к семинарскому занятию необходимо уточнить план проведения занятий, подготовить необходимую документацию. Практические занятия проводятся после лекционного изучения темы. Решение задач и выполнение заданий, приведенных в программе учебной дисциплины обязательно.

При изучении данного курса преподавателем используются интерактивные методы обучения, что помогает эффективнее сформировать заявленные компетенции. Если занятия проводятся в малых группах, то каждая группа обеспечивается необходимой документацией. Занятие проводится в постоянном сравнении расчетов и выступлении участников команд.

В результате каждая из команд выносит на всеобщее обсуждение свои результаты и может быть оценена как со стороны преподавателя, так и со стороны студентов другой команды.

Приложение 2

УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«УНИВЕРСИТЕТ УПРАВЛЕНИЯ «ТИСБИ»

Кафедра проектного менеджмента и управления качеством

Фонд оценочных средств для проведения текущей и промежуточной аттестации по дисциплине «Основы анализа больших данных»

Направление подготовки 38.04.02 Менеджмент

Профиль подготовки Менеджмент устойчивого развития

Год набора 2025, 2026

Содержание

1. Паспорт фонда оценочных средств
2. Наполнение фонда оценочных средств по формам контроля
2.1. Фонд оценочных средств и шкала оценивания для текущего контроля
2.1.1 Выступление на семинаре
2.1.2 Тестирование
2.2. Фонд оценочных средств и шкала оценивания для промежуточного контроля.
2.2.1 Фонд оценочных средств для проверки знаний и умений (вопросы к зачету)
2.2.2 Фонд оценочных средств для проверки сформированности навыков (задачи к зачету).

1. ПАСПОРТ ФОНДА ОЦЕНОЧНЫХ СРЕДСТВ

Формы контроля / **Формируемые компетенции и индикаторы**	**ОПК-2**
	ОПК-2.1
Формы текущего контроля	
Выступление на семинаре	З6
Тестирование	З6, У9
Формы промежуточного контроля	
Зачет	В9

З- знания, У- умения, В- владение

2. Наполнение фонда оценочных средств по формам контроля

2.1. Фонд оценочных средств и шкала оценивания для текущего контроля

2.1.1. Выступление на семинаре

Выступление на семинаре является формой контроля для оценки уровня освоения компетенций, применяемой на семинарских занятиях. Выступление на семинаре может проводиться с использованием форм устного опроса, обсуждения докладов, эссе, выполненных индивидуальных заданий и проблемных вопросов. Выступление на семинаре, таким образом, является обязательной для всех студентов формой текущего контроля знаний. Примерные вопросы к семинару отражены в разделе «Самостоятельная работа». Дополнительно задаются вопросы, направленные на выявление уровня понимания студентом сути проблемной ситуации, поиска аналогов и решений. Выступление, сопровождаемое презентацией, оценивается наиболее высоко.

Примеры вопросов к семинарским занятиям:

1. Понятие Больших данных.

2) Особенности сбора, хранения, обработки и анализа Больших данных

3) Требования к распределенным информационным системам

4) Средства построения распределенных информационных систем

5) Технология Map-Reduce

6) Система Apache Hadoop

7) Базы данных NoSQL. Особенности, классификация

8) Возможности NoSQL-баз данных по обеспечению целостности, доступности скорости обработки информации. CAP-теорема.

9) Способы репликации и кластеризации баз данных

10) Документо-ориентированные базы данных

11) Возможности СУБД MongoDB

12) Работа с документо-ориентированными БД на языке JSON.

Критерии оценивания выступления на семинаре

Результат	Балл
Демонстрирует полное понимание поставленного вопроса, логично и последовательно отвечает на вопрос. Дает развернутый ответ с практическими примерами, использует презентацию	100-90
Дает полный и логически правильный ответ на вопрос, но сформулировать примеры по рассматриваемому вопросу не может	80-89
Демонстрирует частичное понимание сути вопроса, с ответами на вопросы по теме затрудняется	70-79
Способен сформулировать основные подходы к управлению качеством	60-69
Демонстрирует непонимание вопроса, отвечает с наличием грубых ошибок в ответе, либо не отвечает на вопросы	Менее 60

2.1.2. Тестирование

Тестирование в письменной форме проводиться для закрепления знаний по изученной теме. Примеры тестовых вопросов:

1. Что такое бизнес-процесс?

A) Любая деятельность в корпоративных масштабах

B) Коммерческая деятельность с целью получения прибыли

C) Совокупность бизнес-функций

D) Последовательность действий по преобразованию входов в выходы, удовлетворяющие потребителя

2. Описать структуру системы бизнес-процессов, показать состав процессов одного уровня абстракции и взаимосвязи между ними можно с помощью диаграммы в нотации

A) EPC

B) IDEF0

C) BPMN

D) DFD

3. Архитектура предприятия — это

A) Искусство проектировать и строить бизнес-центры и производственные здания

B) Концептуальная структура организация системы

C) Единая система, которая описывает существующие организационные структуры, цели и показатели их достижения, линейку создаваемых продуктов/услуг, которые приносят доход, а также инфраструктуру (программное и аппаратное обеспечение, оборудование), используемые в работе

D) Стиль управления

4. Требование «Пользовательский GUI должен предоставлять возможность языковой локализации: выбор языка (русский/английский) для надписей на элементах» — это

A) Требование стейкхолдера (stakeholder requirement)

B) Нефункциональное требование (non-functional requirement)

C) Бизнес-требование (business requirement)

D) Функциональное требование (functional requirement)

5. Владелец бизнес-процесса — это

A) ответственный исполнитель

B) лицо, которое отвечает за результат процесса, заинтересовано в нем, обладает ресурсами и полномочиями для его выполнения

C) функциональный менеджер

D) спонсор проекта

Критерии оценивания тестирования

Результат	Балл
Даны правильные ответы на все вопросы	100-90
Даны правильные ответы на 80% вопросов теста	80-89
Даны правильные ответы на 60% вопросов	70-79
Даны правильные ответы менее чем на 60% вопросов	60-69
Даны правильные ответы на 20% вопросов теста	Менее 60

2.2. Фонд оценочных средств и шкала оценивания для промежуточного контроля

2.2.1. Фонд оценочных средств для проверки знаний/умений

Вопросы и задачи к зачету

1. Назовите основные области применения анализа данных в финансах.

2. Какие структуры данных в Python используются для хранения табличной информации?

3. Объясните, как Python помогает в обработке данных. Назовите 2–3 библиотеки и их назначение.

4. Объясните, как машинное обучение может помочь в прогнозировании рыночных трендов.

5. Какие методы предобработки данных применяются для работы с пропущенными значениями?

6. Что такое выбросы? Как их можно обнаружить выбросы в данных графически и численно?

7. Для каких целей в финансах используют гистограммы и box-plot?

8. Как гистограмма помогает анализировать распределение данных? Приведите пример из финансовой сферы.

9. В чем разница между MSE и R2? Как интерпретировать R2 = 0.85?

10. Сравните метрики MSE и MAE. В каких случаях предпочтительна каждая?

11. Как с помощью метрики R2 оценить качество моделей прогнозирования?

12. Как с помощью метрики MSE оценить качество моделей прогнозирования?

13. Как с помощью метрики MAE оценить качество моделей прогнозирования?

14. Как с помощью метрики MAPE оценить качество моделей прогнозирования?

15. Как с помощью метрики RMSE оценить качество моделей прогнозирования?

16. Что такое мультиколлинеарность и как она влияет на модель регрессии?

17. Объясните принцип работы kNN-регрессии. Какие у нее преимущества и недостатки?

18. Функции расстояния.

19. Чем отличается k-means от иерархической кластеризации?

20. Как выбрать оптимальное число кластеров?

21. Какие метрики качества используются для оценки кластеризации?

22. Как с помощью метрики Silhouette Coefficient оценить качество моделей кластеризации?

23. Как с помощью метрики Calinski-Harabasz оценить качество моделей кластеризации?

24. Как с помощью метрики Davies-Bouldin оценить качество моделей кластеризации?

2.2.2 Фонд оценочных средств для проверки сформированных навыков (задачи к зачету)

Задача 1

Для набора данных wage2 рассмотрим задачу прогнозирования для переменных зависимая/target log(wage) объясняющая/предикторы/features hours, age, educ подгоните на исходном датасете модель k-NN с параметрами: k = 5, weights=distance.

Рассмотрим индивидуумов с характеристиками:

index hours age educ exper

0 41 32 13 11

1 46 28 11 12

2 42 36 12 17

Вычислите прогноз для wage для индивидуума с index = 2 .

Задача 2

Для набора данных wage2 рассмотрим линейную регрессию wage на age, educ, exper. Подгоните модель и на обучающей выборке вычислите метрики подгонки: R2, MSE, MAE, MAPE, RMSE.

Задача 3

Для набора данных wage2 рассмотрим задачу прогнозирования для переменных зависимая/target log(wage) объясняющая/предикторы/features hours, age, educ

постройте на исходном датасете модель k-NN с параметрами: k = 5, weights=distance (не разбивайте датасет на обучающую и тестовую часть). На обучающей выборке вычислите метрики подгонки: R2, MSE, MAE, MAPE, RMSE.

Критерии оценки уровня усвоения знаний, умений и навыков по результатам зачета

Характеристика ответа	**Европейская оценка**	**Рубежные баллы**	**Оценка**	**Уровень сформированности компетенций**
Дан полный, развернутый ответ на поставленный теоретический вопрос, показана совокупность осознанных знаний об объекте, проявляющаяся в свободном ориентировании понятиями, умении выделить существенные и несущественные его признаки, причинно-следственные связи. Знание об объекте демонстрируется на фоне понимания его в системе данной науки и междисциплинарных связей. Ответ формулируется в терминах науки, изложен	**A**	**100-96**	**Зачтено**	**Повышенный уровень сформированности компетенций**

литературным языком, логичен, доказателен, демонстрирует авторскую позицию студента. При ответе вопрос студент демонстрирует применение знаний к реальным профессиональным ситуациям, объясняет решение задачи на уровне анализа, синтеза и дает свою оценку решения проблемы. Причем студент не затрудняется с ответом при видоизменении задания и правильно обосновывает принятое решение, владеет разносторонними навыками и приемами выполнения практических задач.				
Дан полный, развернутый ответ на поставленный теоретический вопрос, показана совокупность осознанных знаний об объекте, доказательно раскрыты основные положения темы; в ответе прослеживается четкая структура, логическая последовательность, отражающая сущность раскрываемых понятий, теорий, явлений. Знание об объекте демонстрируется на фоне понимания его в системе данной науки и междисциплинарных связей. Умеет тесно увязывать теорию с практикой. Задача решена правильно и с обоснованием принятого решения. Ответ изложен литературным языком в терминах науки. Могут быть допущены недочеты в определении понятий, исправленные студентом самостоятельно в процессе ответа.	**A**	**95-91**	**Зачтено**	
Дан полный, развернутый ответ на поставленный вопрос, доказательно раскрыты основные положения темы; в ответе прослеживается четкая структура, логическая	**A**	**90-86**	**Зачтено**	

последовательность, отражающая сущность раскрываемых понятий, теорий, явлений. Задача решена верно, правильно обосновывает принятую методику решения задачи. Ответ изложен литературным языком в терминах науки. В ответе допущены недочеты, исправленные студентом с помощью преподавателя.				
Дан полный, развернутый ответ на поставленный вопрос, показано умение выделить существенные и несущественные признаки, причинно-следственные связи. Ответ четко структурирован, логичен, изложен литературным языком в терминах науки. Студент владеет разносторонними навыками и приемами выполнения практических задач. Могут быть допущены недочеты или незначительные ошибки, исправленные студентом с помощью преподавателя.	**B**	**85-81**	**Зачтено**	**Базовый уровень сформированности компетенций**
Дан полный, развернутый ответ на поставленный вопрос, показано умение выделить существенные и несущественные признаки, причинно-следственные связи. Ответ четко структурирован, логичен, изложен в терминах науки. Студент владеет разносторонними навыками и приемами выполнения практических задач. Ответы на дополнительные вопросы логичны, изложены в терминах науки, однако допущены незначительные ошибки или недочеты, исправленные студентом с помощью "наводящих" вопросов преподавателя.	**C**	**80-76**	**Зачтено**	
Студент демонстрирует достаточные теоретические и	**C**	**75-71**	**Зачтено**	

практические знания. Дан полный, но недостаточно последовательный ответ на поставленный вопрос, но при этом показано умение выделить существенные и несущественные признаки и причинно-следственные связи. Ответ логичен и изложен в терминах науки. Могут быть допущены 1-2 ошибки в определении основных понятий или решении практической задачи, которые студент затрудняется исправить самостоятельно.				
Дан недостаточно полный и развернутый ответ. Логика и последовательность изложения имеют нарушения. Допущены ошибки в раскрытии понятий, употреблении терминов. Студент не способен самостоятельно выделить существенные и несущественные признаки и причинно-следственные связи. Студент может конкретизировать обобщенные знания, доказав на примерах их основные положения только с помощью преподавателя. Речевое оформление требует поправок, коррекции. Студент испытывает затруднения при выполнении практической задачи и не может связать теорию с практикой.	**D**	**70-66**	**Зачтено**	**Пороговый уровень сформированности компетенций**
Дан неполный ответ, логика и последовательность изложения имеют существенные нарушения. Допущены грубые ошибки при определении сущности раскрываемых понятий, теорий, явлений, вследствие непонимания студентом их существенных и несущественных признаков и связей. В ответе отсутствуют выводы. Умение раскрыть конкретные проявления	**E**	**65-61**	**Зачтено**	

обобщенных знаний не показано. Испытывает затруднения при выполнении практических задач. Речевое оформление требует поправок, коррекции.				
Дан неполный ответ. Присутствует нелогичность изложения. Студент затрудняется с доказательностью. Масса существенных ошибок в определениях терминов, понятий, характеристике фактов, явлений. В ответе отсутствуют выводы. Речь неграмотна. При ответе на дополнительные вопросы студент начинает понимать связь между знаниями только после подсказки преподавателя	**E**	**60**	**Зачтено**	
Студент испытывает значительные трудности в ответе на вопросы. Присутствует масса существенных ошибок в определениях терминов, понятий, характеристике фактов, явлений теории управления. Речь неграмотна. На дополнительные вопросы студент не отвечает. Задача не решена	**F**	**Менее 60**	**Не зачтено**	**Компетенции не сформированы**