

Федеральное государственное образовательное бюджетное
учреждение
высшего образования
**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ
РОССИЙСКОЙ ФЕДЕРАЦИИ»**
(Финансовый университет)

Краснодарский филиал Финуниверситета

Кафедра «Математика и информатика»

СОГЛАСОВАНО

ООО «Портал-Юг»
Генеральный директор



Е.В. Мостовой

«21» февраля 2024 г.

УТВЕРЖДАЮ

Краснодарский филиал
Финуниверситета

Директор



Э.В.Соболев

«21» февраля 2024 г.

Дюдин М.С.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
КЛАСТЕРНЫЙ АНАЛИЗ**

студентов, обучающихся по направлению подготовки

01.03.02 Прикладная математика и информатика

в соответствии с образовательными стандартами Финансового
университета

(программа подготовки бакалавров)

*Рекомендовано Ученым советом Краснодарского филиала Финуниверситета
(протокол № 12 от 20.02.2024)*

*Одобрено кафедрой «Математика и информатика»
(протокол № 13 от 27.02.2024)*

Краснодар 2024

УДК: 519.251:622

ББК: 22.172 Г 46

Д95

Рецензенты: Н.Ю. Нарыжная кандидат технических наук, доцент кафедры «Математика и информатика» Краснодарского филиала Финуниверситета. Н.Г. Пьянкова - доцент кафедры «Математика и информатика» Краснодарского филиала Финуниверситета.

Дюдин М.С. Рабочая программа дисциплины кластерный анализ для обучающихся по направлению 01.03.02 Прикладная математика и информатика, профиль «Анализ данных и принятие решений в экономике и финансах». – Краснодар: Краснодарский филиал Финуниверситета, кафедра «Математика и информатика», 2024г.

Дисциплина Кластерный анализ относится к модулю профиля по направлению подготовки 01.03.02-Прикладная математика и информатика.

В рабочей программе дисциплины определены ее цель, требования к результатам освоения дисциплины, содержание программы, тематика аудиторных занятий, формы самостоятельной работы, оценочные средства для текущего контроля и промежуточной аттестации, учебно-методическое и информационное обеспечение.

Рабочая программа дисциплины «Кластерный анализ»

*Формат 60*90/16. Гарнитура Times New Roman*

Усл. п.л. 2,0. Изд. № _от.

Тираж 100 экз.

Заказ № .

Отпечатано в Краснодарском филиале Финуниверситета

© Дюдин М.С.

© Краснодарский филиал Финуниверситета, 2024

Содержание

1.Наименование дисциплины	4
2.Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине	4
3.Место дисциплины в структуре образовательной программы	5
4.Объём дисциплины и виды учебной работы	5
5.Содержание дисциплины	7
5.1.Содержание разделов дисциплины	7
5.2.Учебно-тематический план	8
5.3.Содержание практических занятий	10
6.Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы	11
6.1.Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы.....	11
6.2.Перечень вопросов, заданий, тем для подготовки к текущему контролю.....	12
7.Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине	13
8.Перечень основной и дополнительной учебной литературы	23
9.Перечень ресурсов информационно-телекоммуникационной сети «Интернет»	24
10.Методические указания для обучающихся по освоению дисциплины ...	24
11.Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем.....	25
12.Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.....	25

1. Наименование дисциплины

Дисциплина «Кластерный анализ»

2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Дисциплина «Кластерный анализ» обеспечивает инструментарий формирования следующих компетенций: ПКП-5.

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (владения, умения и знания), соотнесенные с компетенциями/индикаторами достижения компетенции
ПКП-5	Способность применять методы и инструменты анализа данных и машинного обучения при подготовке аналитического обоснования финансово-экономических решений	1. Демонстрирует знание методов и инструментов анализа данных и машинного обучения. 2. Демонстрирует знание принципов подготовки аналитического обоснования финансово-экономических решений. 3. Владеет навыками использования методов и инструментов анализа данных и машинного обучения при подготовке аналитического обоснования	<p>Знать: математические методы обработки и классификации данных, способы их реализации на языках программирования;</p> <p>Уметь: проводить обработку и исследование данных в табличном процессоре; используя специализированные модули на Python</p> <p>Знать: основные требования и рекомендации к представлению данных, особенности их визуализации</p> <p>Уметь: проводить обработку данных, оформлять документы в распространенных форматах, проводить наглядную визуализацию данных</p> <p>Знать: методы анализа данных, их классификации, способы реализации алгоритмов машинного обучения на Python</p> <p>Уметь: исследовать данные и использованием табличных процессоров и программ на Python, применять машинное обучение для их классификации</p>

3. Место дисциплины в структуре образовательной программы

Дисциплина «Кластерный анализ» является обязательной дисциплиной модуля профиля направленности программы бакалавриата профиля «Анализ данных и принятие решений в экономике и финансах».

Изучение дисциплины базируется на знаниях студентов, полученных в процессе подготовки бакалавров по направлению 01.03.02. «Прикладная математика и информатика» по следующим дисциплинам: «Основы информатики, дискретной математики», «Языки и методы программирования». Дисциплина «Прикладные модели и методы факторного, дискриминантного и кластерного анализа» является дополнительным теоретическим и практическим основанием для последующих математических и финансово-экономических дисциплин подготовки магистров, использующих соответствующие количественные методы.

4. Объём дисциплины и виды учебной работы

Очная форма обучения

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 6
Общая трудоёмкость дисциплины	3/108	108
Аудиторные занятия	34	34
Лекции	16	16
Практические занятия и семинарские занятия, в т.ч.	18	18
Самостоятельная работа (в том числе подготовка реферата)	74	74
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Зачет	Зачет

Очно – заочная форма обучения

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Семестр 7
Общая трудоёмкость дисциплины	3/108	108
Аудиторные занятия	24	24
Лекции	8	8
Практические занятия и семинарские занятия, в т.ч.	16	16
Самостоятельная работа (в том числе подготовка реферата)	84	84
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Зачет	Зачет

5. Содержание дисциплины

5.1.Содержание разделов дисциплины

Раздел 1. Корреляционный анализ многомерной генеральной совокупности

- 1.1. Основные понятия.
- 1.2. Точечные оценки параметров. Примеры вычисления выборочных характеристик.
- 1.3. Оценка значимости наличия связи.
- 1.4. Интервальные оценки коэффициентов корреляции.
- 1.5. Частные и множественные корреляции.
- 1.6. Методы проверки гипотезы о нормальном распределении генеральной совокупности.

Раздел 2. Регрессионный анализ

- 2.1. Основы линейного регрессионного анализа. Уравнение регрессии. Интервальные оценки коэффициентов регрессии.
- 2.2. Показатели качества регрессионной модели.
- 2.3. Построение прогнозов по регрессионной модели.

Раздел 3. Методы снижения размерности многомерных признаков

- 3.1. Метод главных компонент.
- 3.2. Линейная модель метода главных компонент.
- 3.2. Факторный анализ.
- 3.4. Многомерное шкалирование.

Раздел 4. Методы многомерной классификации

- 4.1. Дискриминантный анализ.
- 4.2. Проблема классификации.
- 4.3. Дискриминирующие функции.
- 4.4. Кластерный анализ.
- 4.5. Иерархические алгоритмы.

5.2. Учебно-тематический план

Очная форма обучения

№ п/ п	Наименование темы (раздела) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Все го	Аудиторная работа			Самост оатель ная работа	
			Об ще е	Лекц ии	Прак т/сем инар ские заня тия		
1.	Корреляционны й анализ многомерной генеральной совокупности	27	15	4	5	25	Самостоятельные работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям. Контрольная работа.
2.	Регрессионный анализ	27	10	4	5	25	
3.	Методы снижения размерности многомерных признаков	27	5	4	4	20	
4.	Методы многомерной классификации	27	4	4	4	4	
Всего по дисциплине		108	34	16	18	74	

Очно – заочная форма обучения

№ п/ п	Наименование темы (раздела) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Все го	Аудиторная работа			Самост оатель ная работа	
			Об ще е	Лекц ии	Прак т/сем инар ские заня тия		
1.	Корреляционны й анализ многомерной генеральной совокупности	27	10	2	5	20	Самостоятельные работы. Участие в решении задач на практических занятиях. Собеседования по домашним заданиям. Контрольная работа.
2.	Регрессионный анализ	27	10	2	5	20	
3.	Методы снижения размерности многомерных признаков	27	2	2	3	20	
4.	Методы многомерной классификации	27	2	2	3	24	
Всего по дисциплине		108	24	8	16	84	

5.3.Содержание практических занятий

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на практических занятиях, рекомендуемые источники из разделов 8,9	Формы проведения занятий
Корреляционный анализ многомерной генеральной совокупности	Решение задач в интерактивной форме. Рекомендуемые источники п.8, 9	Практикум с интерактивом
Регрессионный анализ	Решение задач в интерактивной форме. Рекомендуемые источники п.8, 9	Практикум с интерактивом
Методы снижения размерности многомерных признаков	Решение задач в интерактивной форме, проверка самостоятельной работы и разбор ошибок. Рекомендуемые источники п.8, 9	Практикум с интерактивом
Методы многомерной классификации	Решение задач в интерактивной форме, проверка самостоятельной работы и разбор ошибок, аудиторная контрольная работа. Рекомендуемые источники п.8, 9	Практикум с интерактивом

По структуре практические занятия состоят из:

- 1) выборочной проверки корректности выполнения домашнего задания;
- 2) разбора типичных ошибок, возникших в самостоятельной работе;
- 3) рассмотрения теоретических оснований для практики текущей темы;
- 4) разбора практических методов и решения соответствующих задач;
- 5) корректировки заданий для самостоятельной работы студентов.
- 6) решения задач в интерактивной форме.

6.Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

6.1.Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы

При изучении дисциплины «Кластерный анализ» обязательными являются следующие формы самостоятельной работы:

- разбор теоретического материала по пособиям, конспектам лекций или видеолекциям;
- самостоятельное изучение указанных теоретических вопросов;
- решение задач по темам практических занятий;
- выполнение домашней работы;
- подготовка к зачету.

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Корреляционный анализ многомерной генеральной совокупности	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.	Разбор теории, изучение теоретических вопросов, решение задач по темам практических, выполнение ДЗ и подготовка к зачету
Регрессионный анализ	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.	Разбор теории, изучение теоретических вопросов, решение задач по темам практических, выполнение ДЗ и подготовка к зачету
Методы снижения размерности многомерных признаков	Работа с учебной литературой. Решение типовых задач. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.	Разбор теории, изучение теоретических вопросов, решение задач по темам практических, выполнение ДЗ и подготовка к зачету

6.2.Перечень вопросов, заданий, тем для подготовки к текущему контролю

Примерный вариант контрольной работы

1. Дайте определение аксиоме расстояния?
2. Перечислите, что применяется в кластерном анализе?
3. Перечислите качественные признаки кластерного анализа?
4. Перечислите количественные признаки кластерного анализа?
5. Какое свойство кластера показывает плотность?
6. В каком методе Результирующие кластеры имеют тенденцию быть представленными «цепочками»
7. В каком методе используется оптимизация дисперсии внутрикластерных расстояний?
8. Что показывает График средних значений признаков?
9. Дайте определение Манхэттенскому расстоянию?
10. Что получается в результате ослабления критерий уникальности объектов?
11. Дайте определение монокластеру?
12. Примеры практического использования методов многомерного статистического анализа.
13. Для непараметрической модели метода наименьших квадратов в случае линейной функции одной переменной разработайте алгоритмы
14. а) расчета доверительных границ для коэффициентов модели;
15. б) проверки гипотез относительно этих коэффициентов.
16. Докажите, что сумма исходных значений зависимой переменной должны быть равна сумме восстановленных значений.
17. Критерии качества регрессионной модели.
18. Использование непараметрических оценок плотности для
19. Восстановления зависимости.

Критерии балльной оценки различных форм текущего контроля успеваемости

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях кафедры «Математика и информатика».

7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в п.1. «Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Технологии обработки данных».

Фонд оценочных средств включает контрольные материалы для проведения текущего контроля и промежуточной аттестации.

7.1 Описание показателей и критериев оценивания компетенций

Планируемые результаты освоения компетенции (индикатора достижения компетенции)	Уровень освоения				Оценочное средство
	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»	
ПКП-5 Способность применять методы и инструменты анализа данных и машинного обучения при подготовке аналитического обоснования финансово-экономических решений.					
1. Демонстрирует знание методов и инструментов анализа данных и машинного обучения.					
Знать: —положения теории вероятностей и мат. статистики, свойства непрерывных и дискретных случайных величин, типичные	Фрагментарное представление	Неполное представление	Сформированные, но содержащие отдельные пробелы представления	Сформированные систематические представления	Вопросы для оценки знаний и умений, задания в виде расчетных задач, тестовые задания

Планируемые результаты освоения компетенции (индикатора достижения компетенции)	Уровень освоения				Оценочное средство
	«неудовлетв орительно»	«удовлетво рительно»	«хорошо»	«отлично»	
законы распределения					
Уметь: –подбирать подходящий закон распределения для моделировани я исследуемого показателя, определять его числовые характеристик и используя алгоритмы на языке Python	Фрагмента рное умение	Несистемати ческое применение умений	В целом успешное, но содержащее отдельные пробелы умение	Сформирова нное умение	Вопросы для оценки знаний и умений, задания в виде расчетных задач, тестовые задания
2. Демонстрирует знание принципов подготовки аналитического обоснования финансово-экономических решений.					
Знать: –ограничения существующи х теоретико- вероятностны х моделей, методы мат. моделировани я из смежных областей	Фрагмента рное представле ние	Неполные представлен ия	Сформирова нные, но содержащие отдельные пробелы представлен ия	Сформирова нные систематиче ские представлен ия	Вопросы для оценки знаний и умений, задания в виде расчетных задач, тестовые задания
Уметь: –проверять значимость построенной модели с использовани ем модулей Statistics, Pandas Python, выбирать оптимальный способ решения	Фрагмента рное умение	Несистемати ческое применение умений	В целом успешное, но содержащее отдельные пробелы умение	Сформирова нное умение	Вопросы для оценки знаний и умений, задания в виде расчетных задач, тестовые задания

Планируемые результаты освоения компетенции (индикатора достижения компетенции)	Уровень освоения				Оценочное средство
	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»	
поставленной задачи					
3. Владеет навыками использования методов и инструментов анализа данных и машинного обучения при подготовке аналитического обоснования					
Знать: —стандартные алгоритмы проверки статистических гипотез, составления вероятностных моделей особенности их реализации на Python	Фрагментарное представление	Неполные представления	Сформированные, но содержащие отдельные пробелы представления	Сформированные систематические представления	Вопросы для оценки знаний и умений, задания в виде расчетных задач, тестовые задания
Уметь: -с помощью языка Python создавать мат. модели реальных экономических и IT процессов.	Фрагментарное умение	Несистематическое применение умений	В целом успешное, но содержащее отдельные пробелы умение	Сформированное умение	Вопросы для оценки знаний и умений, задания в виде расчетных задач, тестовые задания

7.2 Задания или иные материалы, необходимые для оценки знаний и умений, характеризующих формирование компетенций в процессе освоения ОП ВО

7.2.1 Вопросы для оценки знаний и умений, характеризующих формирование компетенций

Шифр компетенции	Вопросы	Правильный ответ
ПКП-5	1. Какая точечная оценка является смещенной?	дисперсия
	2. Какой стат. критерий используется для проверки значимости отличия к-та корреляции от нуля?	критерий Стьюдента
	3. Какой критерий используется для	критерий Пирсона или

проверки соответствия закона распределения нормальному закону?	критерий Колмогорова
4. Какой столбец в таблице MS Excel показывает границы интервальной оценки к-тов регрессии?	"Верхние 95%", «нижние 95%»
5. Какой показатель оценивает средний квадрат отклонения от среднего?	дисперсия
6. Чем вызвана изменчивость точечных оценок параметров?	вычислением по выборке из генеральной совокупности
7. Какие признаки следует выбирать для составления многофакторной регрессионной модели?	Показывающие корреляцию, значимо отличную от нуля
8. Какой закон распределения описывается кривой Гаусса?	нормальный закон
9. От чего зависит достоверность оценки среднего значения показателя?	от объема выборки и дисперсии
10. Как меняется стандартная ошибка среднего при увеличении объема выборки?	уменьшается
11. Как называют значения, которые сильно выбиваются из набора данных, не вписываются в модель по какой-либо причине?	выбросы
12. Какие форматы файлов с данными может загружать программа, написанная на Python?	любые, если алгоритм может их обработать
13. Как определить, на сколько интервалов надо разбить диапазон значений признака для проверки критерия Пирсона?	по формуле Стерджесса
14. Какой критерий используется для проверки закона распределения?	критерий Пирсона или критерий Колмогорова
15. Как называется показатель, оценивающий взаимосвязь двух величин?	коэффициент корреляции
16. Какая статистика применяется для сравнения двух дисперсий?	Фишера
17. Если надо проверить, что среднее больше определенного значения, то какой критерий надо применить?	критерий Стьюдента (t-статистика)
18. Какой метод применяется для определения коэффициентов линейной регрессии?	метод наименьших квадратов
19. Какие значения чаще всего принимают величины, подчиняющиеся нормальному закону распределения?	значения возле среднего
20. В чем смысл правила 3 сигма?	отклонение от среднего более чем на 3 стандартных отклонения маловероятно

7.2.2 Практико-ориентированные задания

Шифр компетенции	Практико-ориентированные задания	Правильный ответ
ПKN-2	1. Напишите программу, которая считывает данные в 4-м столбце таблицы в файле “GAZP.csv” в список	<pre>f=open('GAZP.csv','r') x=[] for i in range(10): z=f.readline().split(',')[3] try: z=float(z) x.append(z) except: pass</pre>
	2. Напишите программу, загружающую таблицу из файла как data frame	<pre>import pandas as pd table1= pd.read_csv('GAZP.csv') df=pd.DataFrame(table1)</pre>
	3. Напишите программу, оценивающую центр распределения данных, содержащихся в списке y.	<pre>import statistics mean_y=statistics.mean(y) median_y=statistics.median(y) try: mode_y=statistics.mode(y) except: mode_y='не существует'</pre>
	4. Напишите программу, составляющую словарь, где ключи – названия показателей (среднее, ст отклонение), а значения – их величины.	<pre>import statistics mean_y=statistics.mean(y) stdev_y=statistics.stdev(y) dict_y=dict_y={'среднее':mean_y,'ст отклонение':stdev_y}</pre>
	5. Как можно импортировать функцию mean() из библиотеки statistics, чтобы не указывать название библиотеки при её использовании?	<pre>from statistics import mean print(mean([1,2,3]))</pre>
	6. Вычислите коэффициент вариации в Python	<pre>import statistics mean_y=statistics.mean(y) stdev_y=statistics.stdev(y) if mean_y != 0: c_var=stdev_y/mean_y else: c_var='не существует, т.к. среднее равно 0'</pre>
	7. Посчитайте размах данных по списку y	<pre>min_y=min(y) max_y=max(y) range_y=max_y-min_y</pre>
	8. Постройте график по 3 точкам, используя библиотеку matplotlib	<pre>import matplotlib.pyplot as plt x = [1,2,3] y = [2,4,1] plt.plot(x, y) plt.xlabel('x - axis') plt.ylabel('y - axis') plt.title('graph') plt.show()</pre>
	9. Напишите программу для построения графика по	<pre>import matplotlib.pyplot as plt import pandas as pd df = pd.read_csv('fdata.csv', sep=',', parse_dates=True,</pre>

	загруженной из файла таблице	<pre>index_col=0) df.plot() plt.show()</pre>
	10. Посчитайте коэффициент корреляции Пирсона между двумя показателями	<pre>import numpy as np x = np.array([0, 1, 3]) y = np.array([2, 4, 5]) print(np.corrcoef(x, y))</pre>
	11. Посчитайте дневные доходности акций «Газпром», их среднее и стандартное отклонение	<pre>from statistics import mean, stdev f=open('GAZP.csv','r') x=[] for i in range(10): z=f.readline().split(',')[3] try: z=float(z) x.append(z) except: pass for i in range(len(x)-1): d[i]=(x[i]-x[i-1])/x[i-1] print(mean(d),stdev(d))</pre>
	12. Найдите среди всех заработных плат (Wage) самую высокую и запишите это значение в переменную rich. Посчитайте, сколько зарабатывают сотрудники за год (Wage) вместе, и сохраните значение в переменной total. Выведите результаты кода через запятую: rich, low, total.	<pre>import pandas as pd table1 = pd.read_csv('personel.csv') rich = table1.Wage.max() low = table1.Value.min() total = table1.Wage.sum() print(rich, low, total)</pre>
	13. Напишите программу, которая загрузит из файла таблицу с данными по сотрудникам и посчитает средний возраст, округленный до целых (столбец Age)	<pre>import pandas as pd table1 = pd.read_csv('personel.csv') age = round(table1.Age.mean())</pre>
	14. Напишите программу на Python для построения двух или более линий с легендами, различной ширины и цвета.	<pre>import matplotlib.pyplot as plt # line 1 x1 = [10,20,30] y1 = [20,40,10] # line 2 x2 = [10,20,30] y2 = [40,10,30] plt.xlabel('x - axis') plt.ylabel('y - axis') plt.title('Two or more lines') plt.plot(x1,y1, color='blue', linewidth = 3, label = 'line1 - width-3') plt.plot(x2,y2, color='red', linewidth = 5, label = 'line2 - width-5') plt.legend() plt.show()</pre>
	15. Напишите программу на Python для отображения	<pre>import datetime as DT from matplotlib import pyplot as plt from matplotlib.dates import date2num</pre>

	<p>сетки и нарисуйте линейные графики значения закрытия Alphabet Inc. в период с 3 октября 2016 года по 7 октября 2016 года. Настроил линии сетки с рендерингом с большей сеткой (основной сеткой) и меньшей сетка (второстепенная сетка). Включите сетку, но отключите галочки.</p>	<pre> data = [(DT.datetime.strptime('2016-10-03', "%Y-%m-%d"), 772.559998), (DT.datetime.strptime('2016-10-04', "%Y-%m-%d"), 776.429993), (DT.datetime.strptime('2016-10-05', "%Y-%m-%d"), 776.469971), (DT.datetime.strptime('2016-10-06', "%Y-%m-%d"), 776.859985), (DT.datetime.strptime('2016-10-07', "%Y-%m-%d"), 775.080017)] x = [date2num(date) for (date, value) in data] y = [value for (date, value) in data] fig = plt.figure() graph = fig.add_subplot(111) # Plot the data as a red line with round markers graph.plot(x,y,'r-o') # Set the xtick locations graph.set_xticks(x) # Set the xtick labels graph.set_xticklabels([date.strftime("%Y-%m-%d") for (date, value) in data]) # naming the x axis plt.xlabel('Date') # naming the y axis plt.ylabel('Closing Value') # giving a title plt.title('Closing stock value of Alphabet Inc.') # Turn on the minor TICKS, which are required for the minor GRID plt.minorticks_on() # Customize the major grid plt.grid(which='major', linestyle='-', linewidth=0.5, color='red') # Customize the minor grid plt.grid(which='minor', linestyle=':', linewidth=0.5, color='black') # Turn off the display of all ticks. plt.tick_params(which='both', # Options for both major and minor ticks top='off', # turn off top ticks left='off', # turn off left ticks right='off', # turn off right ticks bottom='off') # turn off bottom ticks plt.show() </pre>
	<p>16. Напишите программу на Python для отображения гистограммы популярности языков программирования.</p>	<pre> import matplotlib.pyplot as plt x = ['Java', 'Python', 'PHP', 'JavaScript', 'C#', 'C++'] popularity = [22.2, 17.6, 8.8, 8, 7.7, 6.7] x_pos = [i for i, _ in enumerate(x)] plt.bar(x_pos, popularity, color='blue') plt.xlabel("Languages") plt.ylabel("Popularity") plt.title("Popularity of Programming Language\n" + "Worldwide, Oct 2017 compared to a year ago") plt.xticks(x_pos, x) # Turn on the grid plt.minorticks_on() plt.grid(which='major', linestyle='-', linewidth=0.5, color='red') # Customize the minor grid plt.grid(which='minor', linestyle=':', linewidth=0.5, color='black') plt.show() </pre>
	<p>17. Напишите программу с использованием модуля Pandas для удаления последних n строк таблицы, записанной в структуру данных DataFrame.</p>	<pre> import pandas as pd d = {'col1': [1, 2, 3, 4, 7, 11], 'col2': [4, 5, 6, 9, 5, 0], 'col3': [7, 5, 8, 12, 1, 11]} df = pd.DataFrame(data=d) print("Original DataFrame") print(df) print("\nAfter removing last 3 rows of the said DataFrame:") df1 = df.iloc[:3] print(df1) </pre>

	18. Решите задачу на построение линейной регрессии по двум спискам числовых значений	<pre>import numpy as np from sklearn.linear_model import LinearRegression x = np.array([5, 15, 25, 35, 45, 55]).reshape((-1, 1)) y = np.array([5, 20, 14, 32, 22, 38]) model = LinearRegression() model.fit(x, y) model = LinearRegression().fit(x, y) y_pred = model.predict(x) print('predicted response:', y_pred, sep='\n') y_pred = model.intercept_ + model.coef_ * x print('predicted response:', y_pred, sep='\n')</pre>
--	--	--

7.2.3 Тесты

Шифр компетенции	Тестовые задания	Правильный ответ
ПКН-2	1. Что из ниже перечисленного, относится к обучающей выборке? 1) классификация данных 2) объекты с известными ответами 3) алгоритм решающий функцию 4) нет верного ответа	2
	2. Объекты состоят из признаков? 1) нет 2) да 3) в общем случае нельзя сказать 4) нет верного ответа	2
	3. Что не является данными в машинном обучении? 1) матрицы 2) объекты 3) признаки 4) алгоритм	4
	4. Задача классификации - это: 1) Множество объектов, разделенных на классы 2) Исследование влияние одного или нескольких признаков на объект 3) Определение порядка признака согласно рангу 4) нет верного ответа	1
	5. Задача регрессии - это: 1) множество объектов, разделенных на классы 2) исследование влияние одного или нескольких признаков на объект 3) определение признака согласно рангу 4) нет верного ответа	2
	6. Задача ранжирования - это: 1) множество объектов, разделенных на классы 2) исследование влияние одного или нескольких признаков на объект 3) определение признака согласно рангу 4) нет верного ответа	3
	7. Что служит индикатором ошибки для задач классификации?	1

	1) $\varphi(a, x) = [a(x) \neq y^{(*)}(x)]$ 2) $\varphi(a, x) = a(x) - y^{(*)}(x) $ 3) $\varphi(a, x) = (a(x) - y^{(*)}(x))^2$ 4) нет верного ответа	
	8. Как формула подходит для абсолютного значения ошибки для задач регрессии? 1) $\varphi(a, x) = a(x) - y^{(*)}(x) $ 2) $\varphi(a, x) = (a(x) - y^{(*)}(x))^2$ 3) $\varphi(a, x) = [a(x) \neq y^{(*)}(x)]$ 4) нет верного ответа	1
	9. Что является квадратичной ошибкой для задачи регрессии? 1) $\varphi(a, x) = [a(x) \neq y^{(*)}(x)]$ 2) $\varphi(a, x) = (a(x) - y^{(*)}(x))^2$ 3) $\varphi(a, x) = a(x) - y^{(*)}(x) $ 4) нет верного ответа	1
	10. Выберите неверные утверждения. 1) класс это множество объектов с определенным значением 2) в задачах регрессии допустимым ответом является действительное число или числовой вектор 3) в задачах ранжирования ответы получают сразу на множестве объектов 4) области минимального объема с достаточно гладкой границей являются основной составляющей задач ранжирования	4
	11. Какая, из ниже перечисленных задач, является задачей классификации на 4 класса? 1) $Y = \{0, 1\}^M$ 2) $Y = \{0, 1\}$ 3) $Y = \{-1; +1\}$ 4) $Y = \{1, 2, 3, 4\}$	4
	12. В задачах классификации признаки могут быть строковыми, вещественными, числовыми. 1) Да 2) Нет	1

	3) Перечислены не все 4) Нельзя дать универсальный ответ	
--	---	--

7.3. Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний

Наименование компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции	Примеры типовых контрольных заданий
Способность применять методы и инструменты анализа данных и машинного обучения при подготовке аналитического обоснования финансово-экономических решений (ПКП-5)	<p><u>Знать:</u> математические методы анализа и обработки данных, способы их реализации на языках программирования;</p> <p><u>Уметь:</u> проводить обработку и исследование данных в табличном процессоре; используя специализированные модули на Python</p> <p><u>Знать:</u> основные требования и рекомендации к представлению данных, особенности их визуализации</p> <p><u>Уметь:</u> проводить обработку данных, оформлять документы в распространенных форматах, проводить наглядную визуализацию данных</p> <p><u>Знать:</u> методы анализа данных, способы реализации алгоритмов машинного обучения на Python</p> <p><u>Уметь:</u> исследовать данные и использовать табличных процессоров и программ на Python, применять машинное обучение для их классификации</p>	<p>Построение линейной многофакторной регрессии в табличном процессоре; реализация аналогичного алгоритма на Python</p> <p>Статистическая обработка маркетинговой информации в табличном процессоре (описательные статистики, сравнение средних, корреляционного-регрессионный анализ)</p> <p>Задания на выполнение кластерного анализа по эмпирическим данным в табличном процессоре и на Python</p>

Примеры задания к зачету

1. Каким методом кластерного анализа является метод k-средних?
2. Методом k-средних необходимо разделить множество объектов на 3 группы. Сколько кластеров будет образовано по завершении кластеризации?
3. При использовании метода поиска сгущений необходимо указывать количество кластеров?
4. Матрица факторных нагрузок симметричная?
5. Сколько дискриминантных функций необходимо построить при наличии трех классов?

Пример зачетного билета

Билет №1

1. Оценка качества линейной регрессионной модели.
2. Имеются данные за несколько лет о торговом обороте Y западногерманского предприятия и его расходах на рекламу X . Данные представлены в табл. 4.

Таблица 4.

Расходы на рекламу и торговый оборот предприятия.

Годы, t	68	69	70	71	72	73	74	75
Расходы на рекламу $x(t)$, тыс. марок	4	4	5	6	8	8	10	11
Торговый оборот $y(t)$, млн.марок	4	5	6	6	8	10	12	13

Вычислите линейный коэффициент корреляции между случайными величинами X и Y . С помощью метода наименьших квадратов определите коэффициенты линейной регрессии $Y = aX + b$.

8.Перечень основной и дополнительной учебной литературы

а) Основная:

1. Козлов, А. Ю. Статистический анализ данных в MS Excel : учебное пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. — Москва : ИНФРА-М, 2023. — 320 с. — (Высшее образование: Бакалавриат). — DOI 10.12737/2842. - ISBN 978-5-16-004579-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1907518> (дата обращения: 01.07.2024).

2. Гобарева, Я. Л. Бизнес-аналитика средствами Excel : учебное пособие / Я.Л. Гобарева, О.Ю. Городецкая, А.В. Золотарюк. — 3-е изд., перераб. и доп. — Москва : Вузовский учебник : ИНФРА-М, 2023. — 350 с. + Доп. материалы [Электронный ресурс]. - ISBN 978-5-9558-0560-3. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/1893969> (дата обращения: 01.07.2024).

б) Дополнительная:

3. Нагаева, И. А. Основы алгоритмизации и программирования: практикум : учебное пособие / И. А. Нагаева, И. А. Кузнецов. - Москва : Берлин : Директ-Медиа, 2021. - 169 с. - ЭБС Университетская библиотека ONLINE. - URL: <https://biblioclub.ru/index.php?page=book&id=598404> (дата обращения: 17.01.2023). - Текст : электронный.

4. Ниворожкина, Л. И. Многомерные статистические методы в экономике : учебник / Л.И. Ниворожкина, С.В. Арженовский. — Москва : РИОР : ИНФРА-М, 2023. — 203 с. — (Высшее образование). — DOI: <https://doi.org/10.12737/21773>. - ISBN 978-5-369-01621-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1048326> (дата обращения: 01.07.2024).

9.Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

- 1.Сайт <http://www.statsoft.ru>.
- 2.Электронная библиотека (www.bibliotekar.ru).
- 3.Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/> (<http://library.fa.ru/files/elibfa.pdf>)
- 4.Электронно-библиотечная система Znanium <http://www.znanium.com>
5. Образовательный портал Финансового университета

10.Методические указания для обучающихся по освоению дисциплины

Самостоятельная работа студентов проходит аудиторно и внеаудиторно. Организации самостоятельной работы служит учебнотематический план изучения дисциплины. В этом плане указана тематика лекций, практических занятий, вопросы и задания для самостоятельного изучения.

При подготовке к лекции целесообразно предварительно познакомиться с ее содержанием по рекомендованным пособиям и выделить наиболее трудные вопросы. Во время лекций следует конспектировать содержание лекции. После занятий следует провести работу с конспектом:

отредактировать записи, оформить конспект. При оформлении целесообразно выделять специальным образом названия тем и формулировки вопросов, основные определения, формулировки теорем и примеры. Сделанные записи нужно сверить с учебниками и учебными пособиями и в случае расхождений проконсультироваться с преподавателем.

При подготовке к практическому занятию необходимо повторить или, если это требуется, изучить соответствующий теоретический материал. Во время занятия нужно точно записывать формулировки решаемых задач, вопросы, указания преподавателя к решению и разбираемые решения. После занятий необходимо просмотреть записанные решения и восстановить в решениях имеющиеся пробелы. В случае затруднений отметить соответствующие задания и обратиться за консультацией к преподавателю. Практические занятия проходят, как правило, в интерактивной форме и преподаватель учитывает активность студентов, направленную на решение предложенных задач, и в поиске ответов на вопросы.

Для успешного освоения данной дисциплины студентам следует иметь опыт работы в системе MS Excel и в пакете Statistica.

11.Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем

11.1. Комплект безлицензионного программного обеспечения:

- Astra Linux, Libre Office; Python 3; PyCharm

11.2.Современные профессиональные базы данных и информационные справочные системы:

Не требуются

11.3.Сертифицированные программные и аппаратные средства защиты информации

Не предусмотрены.

12.Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине

Лекционные и практические занятия проводятся в мультимедийных компьютерных классах.

Для освоения дисциплины возможно использование вычислительных средств – смартфон или планшет, в качестве дополнительных инструментов организации и осуществления образовательного процесса.