

Федеральное государственное бюджетное образовательное учреждение высшего
профессионального образования
Московский государственный университет имени М.В. Ломоносова
Высшая школа управления и инноваций

УТВЕРЖДАЮ

(и.о.декана)

_____/В.В.Печковская/

«9» июня 2022 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

PYTHON: ВВЕДЕНИЕ В АНАЛИЗ ДАННЫХ

Уровень высшего образования:

Магистратура

Направление подготовки (специальность):

27.04.05 «Инноватика» (3++)

Форма обучения:

очная

Рабочая программа рассмотрена и одобрена
На заседании Совета факультета
(протокол № 3, 9 июня 2022 г.)

Москва 2022

Рабочая программа дисциплины (модуля) разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки / специальности 27.04.05 «Инноватика» (программа магистратуры), утвержденным приказом МГУ от 22 июля 2011 года № 729 (в редакции приказов МГУ от 22 ноября 2011 года № 1066, от 21 декабря 2011 года № 1228, от 30 декабря 2011 года № 1289, от 22 мая 2015 года № 490, от 30 июня 2016 года № 746, от 30 декабря 2020 года №1376).

Год (годы) приема на обучение: 2022.

I. Цели и задачи учебной дисциплины

Целью изучения дисциплины «Python: введение в анализ данных» являются формирование у студентов знаний в области анализа больших массивов данных, понимании типов данных, умении построения прогнозных, регрессионных моделей, а также моделей классификации и кластеризации.

Задачами дисциплины являются:

- изучение типов и структур данных языка программирования Python;
- первичный анализ данных и описательная статистика;
- “очистка” и подготовка данных для последующего построения модели;
- формирование навыков и умений, необходимых для построения моделей на основе алгоритмов машинного обучения

В результате изучения данного курса обучающиеся получают знания о типах и структурах данных языка программирования Python, о принципах использования различных типов в данных в зависимости от решаемой задачи, о методах анализа больших данных и алгоритмах машинного обучения, приобретут навыки и умения построения моделей прогнозирования, классификации, кластеризации, выявления аномальных значений

II. Место дисциплины в структуре ОПОП ВО

Дисциплина «Python: анализ данных» является дисциплиной по выбору профессионального блока вариативной части программы магистратуры.

Изучение дисциплины базируется на знаниях и умениях, полученных обучающимися в процессе изучения математических и экономических дисциплин: «Математическая статистика», «Математический анализ», «Линейная алгебра», «Количественные методы в экономике»

Для успешного освоения дисциплины обучающийся должен:

Знать:

- фундаментальные положения математической статистики, математического анализа и линейной алгебры;

Уметь:

- использовать междисциплинарные системные связи наук;
- анализировать и оценивать ;
- применять математический инструментарий для анализа данных

Владеть:

- навыками математического анализа;
- навыками выбора наиболее актуальных направлений научных исследований, ставить задачи исследования и определять способы решения поставленных задач;
- самостоятельно приобретать и использовать в практической деятельности новые знания и умения в различных сферах деятельности.

III. Требования к результатам освоения дисциплины

В результате освоения дисциплины должны быть сформированы следующие компетенции:

Планируемые результаты обучения

Код и наименование компетенции	Код и наименование индикатора	Планируемые результаты
Универсальные компетенции		
<p>УК-1. Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий, формулировать научно обоснованные гипотезы, применять методологию научного познания в профессиональной деятельности</p>	<p>УК-1.1. Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними</p>	<p>Знать:</p> <ul style="list-style-type: none"> – основные методы критического анализа; – методологию системного подхода; <p>Уметь:</p> <ul style="list-style-type: none"> – выявлять проблемные ситуации, используя методы анализа, синтеза и абстрактного мышления; – осуществлять поиск решений проблемных ситуаций на основе действий, эксперимента и опыта; – производить анализ явлений, обрабатывать полученные результаты, делать обоснованные выводы; – определять в рамках выбранного алгоритма вопросы (задачи), подлежащие дальнейшей разработке и предлагать способы их решения; <p>Владеть:</p> <ul style="list-style-type: none"> – технологиями выхода из проблемных ситуаций, навыками выработки стратегии действий; – навыками критического анализа; – навыками применения системного подхода к анализу проблемных ситуаций. <ul style="list-style-type: none"> – навыками интерпретации полученных данных в ходе анализа проблемной ситуации и формирования обоснованных выводов.

	<p>УК-1.2. Разрабатывает и обосновывает стратегию действий по решению проблемной ситуации на основе системного и междисциплинарных подходов.</p>	<p>Знать основные положения разработки стратегии действий по решению проблемной ситуации;</p> <p>Уметь:</p> <ul style="list-style-type: none"> – разрабатывать и обосновывать стратегию действий по решению проблемной ситуации; – использовать системный и междисциплинарные подходы к решению проблемной ситуации; <p>Владеть навыками разработки стратегии действий по решению проблемной ситуации на основе системного и междисциплинарных подходов.</p>
	<p>УК-1.3. Использует логико-методологический инструментарий для решения проблемной</p>	<p>Знать основные положения использования логико-методологического</p>

	ситуаций.	инструментария; Уметь использовать логико-методологический инструментарий для решения проблемной ситуаций; Владеть навыками применения логико-методологического инструментария для решения проблемной ситуаций.
	УК-1.4. Формулирует научно обоснованные гипотезы, применяет методологию научного познания в профессиональной деятельности.	Знать: – основные положения формулирования научно обоснованных гипотез; – методы научного познания; Уметь: – формулировать научно обоснованные гипотезы; – применять методологию научного познания в профессиональной деятельности; Владеть: – навыками формулирования научно обоснованных гипотез в решении задач профессиональной деятельности; – навыками применения методов научного познания в решении профессиональных задач.
Общепрофессиональные компетенции		
ОПК-1. Способен анализировать и выявлять естественно-научную сущность проблем управления в технических системах на основе положений, законов и	ОПК-1.1. Демонстрирует знание законов, естественно-научных и математических методов для использования в профессиональной деятельности в области	Знать: фундаментальные законы природы и основные физические математические принципы; Уметь: применять физические законы и математические методы

методов в области математики, естественных и технических наук	управления в технических системах.	<p>для решения задач теоретического и прикладного характера в области управления в технических системах;</p> <p>Владеть: навыками использования</p> <p>знаний математики, естественных и технических наук при решении практических задач в области управления в технических системах;</p>
	<p>ОПК 1.2. Проводит анализ и выявляет естественно-научную сущность проблемы управления в технической системе.</p>	<p>Знать:</p> <ul style="list-style-type: none"> – естественнонаучные методы познания; – методологию научных исследований; <p>Уметь:</p> <ul style="list-style-type: none"> – проводить анализ проблемы и выявлять её естественнонаучную сущность; – применять законы математики, естественных и технических наук для анализа проблемы управления в технической системе; <p>Владеть навыками определения естественнонаучной сущности проблемы управления в технической системе.</p>

<p>ОПК-2. Способен формулировать задачи управления в технических системах и обосновывать методы их решения</p>	<p>ОПК-2.1. Формулирует задачи управления в технических системах на основе знаний, профильных разделов математических и естественнонаучных дисциплин</p>	<p>Знать: фундаментальные разделы, профильные разделы математических и естественнонаучных дисциплин;</p> <p>Уметь формулировать задачи профессиональной деятельности на основе знаний, профильных разделов математических и естественнонаучных дисциплин;</p> <p>Владеть методами формулирования задач профессиональной деятельности на основе знаний в области математики, естественных и технических наук.</p>
<p>ОПК-3. Способен самостоятельно решать задачи управления в технических системах на базе последних достижений науки и техники</p>	<p>ОПК-3.1. Применяет результаты и тенденции последних достижений науки и техники для решения задач в области управления в технических системах</p>	<p>Знать: особенности развития последних достижений науки и техники в области управления в технических системах;</p> <p>Уметь:</p> <ul style="list-style-type: none"> – выявлять тенденции технологического развития в наукоемких сферах деятельности; – решать задачи управления в технических системах с использованием современных технологий; <p>Владеть: навыками применения современных технологий для решения задачи управления в технических системах.</p>

	<p>ОПК-3.2. Использует фундаментальные знания для решения базовых задач управления в технических системах</p>	<p>Знать: общие методы решения базовых задач управления в технических системах;</p> <p>Уметь:</p> <ul style="list-style-type: none"> – применять знания естественных наук для построения математических моделей объектов и процессов; – применять методы и способы решения базовых задач в технических системах; <p>Владеть навыками решения базовых задач управления в технических системах.</p>
Профессиональные компетенции		
<p>ПК-2. Способен выявлять и оценивать тенденции технологического развития в наукоемких сферах деятельности, осуществлять технологическое прогнозирование</p>	<p>ПК-2.1. Выявляет и оценивает тенденции технологического развития в наукоемких сферах деятельности</p>	<p>Знать:</p> <ul style="list-style-type: none"> – методы построения концептуальных, математических и имитационных моделей; – передовой отечественный и зарубежный опыт в области развития науки и техники; – методы прогнозирования, технико-экономических исследований научно-технических решений и нормативного проектирования инновационных видов продукции и процессов; <p>Уметь:</p> <ul style="list-style-type: none"> – анализировать научную, научно-техническую информацию; – выявлять и оценивать тенденции технологического развития в наукоемких сферах

		<p>на основе анализа, обобщения и систематизации передового опыта в сфере инноватики по материалам ведущих научных журналов и изданий, с использованием электронных библиотек и интернет-ресурсов;</p> <ul style="list-style-type: none"> – оценивать возможные результаты внедрения передовых технологических решений; <p>Владеть навыками подготовки заключений и отзывов на инновационные предложения повышения эффективности в наукоемких сферах деятельности.</p>
	<p>ПК-2.2. Осуществляет технологическое прогнозирование</p>	<p>Знать основные положения и методы технологического прогнозирования;</p> <p>Уметь:</p> <ul style="list-style-type: none"> – использовать источники информации для анализа данных, необходимых для составления прогноза; – применять методы анализа данных и построения математических моделей; – применять программные средства планирования, мониторинга, контроля исполнения, формирования прогнозных данных; – выполнять технико-экономический анализ проектных, конструкторских и технологических решений для выбора оптимального варианта реализации инноваций; – прогнозировать тенденции

		<p>развития науки и техники в профессиональной сфере;</p> <p>Владеть навыками формирования прогноза технологического развития.</p>
<p>ПК-9. Способен планировать и осуществлять мероприятия по адаптации организации к изменяющимся условиям рынка с учётом тенденций развития науки и техники, руководить процессом организационных изменений при внедрении новой техники и технологий</p>	<p>ПК-9.1. Планирует и осуществляет мероприятия по адаптации организации к изменяющимся условиям рынка с учётом тенденций развития науки и техники</p>	<p>Знать:</p> <ul style="list-style-type: none"> – принципы и основные положения теории решения нестандартных задач, законы эволюции сложных систем, принципы функционального моделирования технических систем и типовые методы их совершенствования; – классификация и основные методы моделирования бизнес-процессов; <p>Уметь:</p> <ul style="list-style-type: none"> – оценивать инновационный потенциал организации; – планировать мероприятия по внедрению и сокращению сроков освоения новой техники и технологии, рациональному использованию ресурсов, повышению эффективности деятельности организации, улучшению качества продукции, совершенствованию организации труда; <p>Владеть навыками разработки плана совершенствования организации производства, труда и управления на основе внедрения новейших технических</p>

		и телекоммуникационных средств.
	<p>ПК-9.2. Руководит процессом организационных изменений при внедрении новой техники и технологий</p>	<p>Знать:</p> <ul style="list-style-type: none"> – современные принципы и технологии менеджмента; – особенности осуществления организационных изменений; – программы (инструменты) корпоративных инноваций; <p>Уметь:</p> <ul style="list-style-type: none"> – оценивать готовность организации к осуществлению организационных изменений и корпоративных инноваций; – анализировать барьеры осуществления организационных изменений и разрабатывать меры по их устранению; – анализировать инновационный потенциал организации и выбирать соответствующие программы (инструменты) корпоративных инноваций; – принимать управленческие решения по осуществлению организационных изменений при внедрении новой техники и технологий; <p>Владеть</p> <ul style="list-style-type: none"> – навыками принятия управленческих решений по осуществлению организационных изменений при внедрении новой техники и технологий; – навыками руководства и организации процесса планирования

		организационных изменений.
ПК-17. Способен применять современные информационные технологии и технические средства для подготовки, публичного представления и защиты проекта (программы) в виде презентации.	ПК-17.1. Применяет современные информационные технологии и технические средства для подготовки презентации проекта (программы)	<p>Знать:</p> <ul style="list-style-type: none"> – технологии подготовки и проведения презентаций; – методы создания рекламных текстов; – основы работы с программными и техническими средствами по подготовке презентаций; <p>Уметь:</p> <ul style="list-style-type: none"> – составлять информационных материалы; – подготавливать презентации с использованием технических средств; <p>Владеть навыками подготовки презентации проекта (программы).</p>
	ПК-17.2. Публично представляет и защищает презентацию проекта (программы)	<p>Знать:</p> <ul style="list-style-type: none"> – правила аргументации и обоснования проекта (программы); – приёмы публичного выступления; <p>Уметь:</p> <ul style="list-style-type: none"> – убеждать собеседника; – проводить публичные презентации с использованием современных информационных технологий и технических средств; – проводить переговоры; – организовывать встречи, совещания, презентация в рамках реализации проекта (программы);

		Владеть навыками публичной защиты основных положений проекта (программы).
СПК-3. Способен применять методы анализа данных для решения профессиональных задач посредством применения современных инструментальных и программных средств	СПК-3.1. Применяет методы анализа данных для решения профессиональных задач посредством применения современных инструментальных и программных средств	<p>Знать:</p> <ul style="list-style-type: none"> – фундаментальные разделы математики; – методы системного анализа; <p>Уметь:</p> <ul style="list-style-type: none"> – использовать законы естественно-научных дисциплин в профессиональной деятельности и применять математический аппарат, методы оптимизации, теории вероятностей, математической статистики, системного анализа для принятия решений в области решения профессиональных задач; – применять методы анализа данных; – применять инструментальные и программные средства для анализа данных; <p>Владеть навыками анализа данных для решения профессиональных задач с применением современных технологий.</p>

Форма обучения: очная.

IV. Формы контроля

Контроль за освоением дисциплины осуществляется в каждом дисциплинарном разделе отдельно.

Рубежный контроль: тестирование и контрольная работа по отдельным разделам дисциплины.

Итоговая аттестация в 3 семестре – зачет в устные вопросы и тест.

Результаты текущего контроля и итоговой аттестации формируют рейтинговую оценку работы обучающегося. Распределение баллов по отдельным видам работ в процессе освоения дисциплины «Python: введение в анализ данных» осуществляется в соответствии с Приложением 1.

V. Объём дисциплины и виды учебной работы

Объём курса – 72 часа, 2 зачетные единицы, в том числе 24 часа – аудиторная нагрузка, из которых 6 часов – лекции, 18 часов – семинары, 48 часов – самостоятельная работа студентов. Изучается на 3 курсе (3 семестр), итоговая форма отчетности – зачет.

Вид учебной работы	Всего часов
Контактные занятия (всего)	24
В том числе:	-
Лекции	6
Практические занятия (ПЗ)	-
Семинары (С)	18
Лабораторные работы (ЛР)	-
Самостоятельная работа (всего)	48
В том числе:	-
Домашние задания	12
Проектная работа	12
Подготовка к тестированию	10
Подготовка к опросу	5
Подготовка к контрольной работе	5
Вид промежуточной аттестации Зачет	4
Общая трудоемкость (часы)	72
Зачетные единицы	2

VI. Структура и содержание дисциплины

п/п	Раздел	Содержание (темы)
1	Библиотеки языка программирования Python	Основные назначения и функции библиотек. Виды библиотек для анализа данных: Pandas, Numpy, Statsmodels, Sklearn, Seabourne Виды библиотек для визуализации данных

2	Типы и структуры данных в Python	Разновидности типов данных: Integer, float, bool, string, object Разновидности структур данных: Dataframe, series, массив, кортежи, списки и др.
3	Загрузка данных в программу и предварительный анализ	Загрузка данных разного формата (xlsx, csv, html, py и др) Выявления количества строк и колонок Выявление пропущенных значений Выявление типов данных в матрице
4	Функции Python для анализа данных	<ul style="list-style-type: none"> • Функции для получения описательной статистики (нахождение max, min, среднее, медианное значение, квантили) • Функции для визуализации плотности распределения данных (Нормальное Гауссово распределение) • Функции для создания бинарных переменных (dummies var) • Функции алгоритмов машинного обучения для построения моделей (Метод наименьших квадрат, метод опорных векторов, случайных лес, логистическая регрессия, временные ряды)
5	Построение регрессионных моделей	Цель построения линейных регрессий по методу МНК Выдвижение гипотез и постановка задачи (на основе рабочих данных) Построение регрессионной модели в Python Оценка значимости полученных коэфф и модели в целом (t-статистика, F-статистика) Оценка качества модели (R ²) Проверка предпосылок Гаусса-Маркова Интерпретация полученных результатов
6	Построение моделей классификации	Алгоритм Random Forest Логистическая регрессия Метод опорных векторов

Разделы дисциплин и виды занятий (ак. часы)

п/п	Наименование раздела	Лекция	Практи-	Лабора-			Форма
-----	----------------------	--------	---------	---------	--	--	-------

	дисциплины		ческие занятия	торные занятия	Семинар	СРС	текущего контроля
1	Библиотеки языка программирования Python	1	-	-		6	Опрос
2	Типы и структуры данных в Python	1	-	-	4	8	Тест Домашнее задание
3	Предварительный анализ	1	-	-	2	8	Опрос Домашнее задание
4	Функции Python для анализа данных	1	-	-	4	8	Тест
5	Построение регрессионных моделей	1	-	-	4	8	КР
6	Построение моделей классификации	1	-	-	4	6	Проект
	Промежуточная аттестация (зачет)					4	
	Итого	6	-	-	18	48	

Разделы дисциплины и междисциплинарные связи

№ п/п	Наименование обеспечиваемых (последующих) дисциплин	№ № разделов данной дисциплины, необходимых для изучения обеспечиваемых (последующих) дисциплин					
		1	2	3	4	5	6
1.					+		
2.		+			+	+	+
3.		+			+	+	

VII. Образовательные технологии

В процессе освоения дисциплины «Управление бизнес-процессами» используются следующие образовательные технологии:

1. Стандартные методы обучения:

- лекции;
- семинары;
- письменные или устные домашние задания;
- консультации преподавателей;

- самостоятельная работа студентов, в которую входит освоение теоретического материала, подготовка к семинарам, выполнение указанных выше письменных работ.

2. Методы обучения с применением интерактивных форм образовательных технологий:

- интерактивные лекции;
- анализ деловых ситуаций на основе кейс-метода и имитационных моделей;
- круглые столы;
- обсуждение подготовленных студентами рефератов;
- групповые дискуссии и проекты;
- обсуждение результатов работы студенческих исследовательских групп.

VIII. Учебно-методическое, информационное и материально-техническое

IX. обеспечение дисциплины

Учебно-методическое и информационное обеспечение дисциплины

а) Основная литература:

1. А. Мюллер, С. Гвидо. «Введение в машинное обучение с помощью Python. Руководство для специалистов по работе с данными» - Published by O'Reilly Media 2021.
2. Дж. Вандер Плас. «Python для сложных задач. Наука о данных и машинное обучение»// — СПб.: Питер, 2020г. — 576 с.: ил. — (Серия «Бестселлеры O'Reilly»).
3. Доугерти К “ Введение в эконометрику”/-М: ИНФРА-М,1999.-XIV, 402 с.
4. Жерон О. «Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем»//Пер. с англ. - СПб.: ООО "Альфа-книга": 2018. - 688 с.: ил. - Парал. тит. англ.
5. Маккини У. Python и анализ данных / пер. с англ. А. А. Слинкина. – М.: ДМК Пресс, 2020. – 540 с.
7. Эл Свейгарт «Автоматизация рутинных задач при помощи Python: практическое руководство для начинающих»// Пер. с англ,-М.:ООО “И.Д.Вильямс”, 2017.-592с
8. Эрик Мэтиз «Изучаем Python» [Текст] / Джестон Д., Нелис Й. – Альпина Паблишер 2015. – 641 с.
9. Ян Эрик Содем. «Программирование компьютерного зрения на Python»[Текст]/ Ян Эрик Содем-O'Reilly, 2021-
10. Beazley D, B.Jones «Python cookbook»// Published by O'Reilly Media, 2021.-667 с.

Перечень ресурсов информационно-телекоммуникационной сети «Интернет» и информационных справочных систем

Перечень профессиональных баз данных и информационных справочных систем

1. ЭБС «Юрайт» [раздел «ВАША ПОДПИСКА: учебники и учебные пособия издательства «Юрайт»]: сайт. – URL: <https://www.biblio-online.ru/catalog/>
2. ЭБС издательства «Лань» [учебные, научные издания, первоисточники, художественные произведения различных издательств; журналы] : сайт. – URL: <http://e.lanbook.com>
3. <https://www.econ.msu.ru/elibrary> – электронная библиотека Экономического факультета МГУ
4. <https://www.nbmgu.ru> – Научная библиотека МГУ

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. <https://pandas.pydata.org/> – документация библиотеки pandas
2. <https://www.statsmodels.org/stable/> документация библиотеки pandas
3. <https://scikit-learn.org/stable/index.html> документация библиотеки sklearn
4. <https://www.kaggle.com/> репозиторий проектов по анализу данных
5. <https://github.com/> - репозиторий проектов на python и других языках программирования
6. <https://habr.com/ru/post/277275/> - Метод наименьших квадратов

Рекомендуемые обучающие, справочно-информационные, контролирующие и прочие компьютерные программы, используемые при изучении дисциплины

№ п/п	Название рекомендуемых по разделам и темам программы технических и компьютерных средств обучения	Номера тем
1.	Jupyter notebook (Anaconda)	1-6
2.	MS Excel	2-4

Методические указания для обучающихся по освоению дисциплины

В процессе изучения курса обучающиеся обязаны соблюдать дисциплину, вовремя приходить на занятия, делать домашние задания, осуществлять подготовку к семинарам и контрольным работам, проявлять активность на занятиях.

При этом важное значение имеет самостоятельная работа, которая направлена на формирование у учащегося умений и навыков правильного оформления конспекта и работы с ним, работы с литературой и электронными источниками информации, её анализа, синтеза и обобщения. Для проведения самостоятельной работы обучающимся предоставляется список учебно-методической литературы.

Материально-техническое обеспечение дисциплины

Для проведения образовательного процесса необходима аудитория, оборудованная компьютером и проектором, необходимыми для демонстрации презентаций. Обязательное программное обеспечение – MS Office.

X. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

Темы курсовых работ

Курсовая работа по дисциплине «Python: введение в анализ данных» не предусмотрена.

Темы рефератов

1. Алгоритмы машинного обучения: типовые задачи
2. Суть метода наименьших квадратов
3. Алгоритм Random Forest для задач классификации
4. Задачи классификации в компьютерном зрении
5. Задачи классификации в приложениях голосовых помощников
6. Метод опорных векторов
7. Алгоритмы машинного обучения для задач банковского скоринга кредитоспособности клиентов
8. Алгоритмы машинного обучения для прогнозирования продаж
9. Прогнозирование стоимости недвижимости: методы оценки
10. Библиотека OpenCV
11. Библиотека Pandas

Вопросы для текущего контроля студентов

- 1.
2. .

Пример теста для контроля знаний обучающихся

Выберите правильные ответы (правильных ответов может быть несколько):

1. Выберите тип данных, не используемый в Python:
 - 1) Integer
 - 2) Uinteger
 - 3) String
 - 4) Bool
2. Метод наименьших квадратов:
 - 1) Метод, с помощью которого находятся оптимальные коэффициенты регрессии
 - 2) Метод, с помощью которого находится та линия регрессии, которая располагается ближе всего к точкам(значениям переменной)
 - 3) Метод для разработки приложений компьютерного зрения
3.
 - 1) describe
 - 2) info()
 - 3) dtype
 - 4)
4. Алгоритмы Random Forest способен решать задачи :
 - 1) Регрессии
 - 2) Классификации
 - 3) Классификации и регрессии.
 - 4) Не один из предложенных вариантов

5. Предпосылки Гаусса-Маркова(укажите верные ответы):
- 1) Мат.ожидание остатков регрессии =0
 - 2) Отсутствие мультиколлинеарности
 - 3) Отсутствие Гетероскедастичности
 - 4) Нормальное распределение зависимой переменной
 - 5) Отсутствие автокорреляции
6. Что описывает показатель R²:
- 1) Насколько % дисперсия Y описывается включенными в модель X's
 - 2) Значимость модели в целом
 - 3) Значимость константы модели
 - 4) Стандартное отклонение переменной
7. По каким критериям можно определить значимость коэффициентов регрессии:
- 1) P-value
 - 2) t-студент
 - 3) F-статистика Фишера
 - 4) Хи-квадрат
8. В уравнении $Y=1200+24000X(1)+600X(2)$, где Y-стоимость страховки, X(1)=1(человек курит) и X1=0(человек не курит), X(2)- количество детей, найти стоимость страховки для курящих людей при постоянстве прочих факторов:
- 1) 25200
 - 2) 24000
 - 3) 1200
 - 4) 1800
9. Какое значение должен принимать показатель DW:
- 1) Около 2
 - 2) В интервале от 0 до 4
 - 3) От -1 до 1
10. Какой интервал значений может принимать коэффициент линейной корреляции :
- 1) От 0 до 1
 - 2) От -1 до 1
 - 3) От 0 до 4
11. Причина, по которой модель случайного леса работает так хорошо, заключается в том, что(выберите правильные варианты):
- 1) Большое число относительно некоррелированных деревьев, работающих совместно, будет превосходить любую из их отдельных составляющих.
 - 2) Некоррелированные модели могут прогнозировать точнее, чем любой индивидуальный прогноз.
 - 3) Причина такого чудесного эффекта: деревья защищают друг друга от своих индивидуальных ошибок, по крайней мере до тех пор, пока они не будут постоянно ошибаться в одном и том же направлении.
12. Случайный лес имеет множество применений, таких как :
- 1) механизмы рекомендаций,
 - 2) классификация изображений

- 3) отбор признаков.
13. Среди самых популярных задач в машинном обучении выделяют (укажите верное):
- 1) классификация – отнесение объекта к одной из категорий на основании его признаков
 - 2) регрессия – прогнозирование количественного признака объекта на основании прочих его признаков
 - 3) кластеризация – разбиение множества объектов на группы на основании признаков этих объектов так, чтобы внутри групп объекты были похожи между собой, а вне одной группы – менее похожи
 - 4) детекция аномалий – поиск объектов, "сильно непохожих" на все остальные в выборке либо на какую-то группу объектов
14. Алгоритм состоит из четырех этапов:
- 1) Создайте случайные выборки из заданного набора данных.
 - 2) Для каждой выборки постройте дерево решений и получите результат предсказания, используя данное дерево.
 - 3) Проведите голосование за каждый полученный прогноз.
 - 4) Выберите предсказание с наибольшим количеством голосов в качестве окончательного результата.

Вопросы к зачету

1. Опишите сущность метода наименьших квадратов.
2. Выведите формулу коэфф а и b линейной регрессии.
3. Опишите сущность алгоритма “случайный лес”.
4. Опишите сущность логистической регрессии .
5. Что показывают коэфф регрессии, константа, показатели R², t-статистика, F-статистика, DW, Skew, p-value.
6. Чему равен tg угла наклона линии регрессии
7. Объясните основные метрики классификации Accuracy, Precision, Recall
8. Объясните типы и структуры данных , используемые в Python
9. Какие циклы в Python Вы знаете? Для чего нужны циклы. Приведите примеры.
10. С помощью каких функций в Python Вы можете создать бинарные переменные. Можно ли качественные переменные сделать бинарными. Если можно, то приведите примеры.
11. Для чего может использоваться логарифмическая или полу-логарифмическая регрессия. Какая будет интерпретация модели в этих случаях.

Примеры контрольной работы

В 1

1. Построить регрессионную модель в Jupyter notebook (Python). Таблица с данными предлагается. Проверить модель и коэфф на значимость. Проверить качество модели. Интерпретировать полученные результаты

В 2

Построить логистическую регрессию кредитного скоринга банковских клиентов.
Интерпретировать полученные результаты

В3

1. Построить прогнозную модель финансового актива на основе анализа временных рядов. Таблица с данными прилагается. Интерпретировать полученные результаты

Примеры домашнего задания

1. На основе имеющихся данных построить регрессионную модель. Интерпретировать полученные результаты
2. На основе имеющихся данных построить классификационную модель, используя алгоритм Random Forest
3. На основе имеющихся данных построить модель временных рядов для прогноза стоимости финансового актива

Приложение 1.

СИСТЕМА РЕЙТИНГОВОЙ ОЦЕНКИ И КОНТРОЛЯ ЗНАНИЙ СТУДЕНТОВ

№ п/п	СТРУКТУРА	Баллы по каждому модулю
1.	Оценка за активное участие в учебном процессе и посещение занятий: <div style="text-align: right; margin-right: 20px;"> Всех занятий Не менее 75% Не менее 50% Не менее 25% </div> Итого:	5 4 3 2 до 5
2.	устный опрос в форме собеседования (УО-1) письменный опрос в виде теста (ПР-1) письменная контрольная работа (ПР-2) устный опрос в форме коллоквиума (УО-2) письменная работа в форме реферата (ПР-4) Итого:	5 10 10 10 10 45
3.	Зачет	50
	ВСЕГО:	100

Пересчет на 5 балльную систему

2 (неудовлетворительно)	3 (удовлетворительно)	4 (хорошо)	5 (отлично)
< 50	50-64	65-84	85-100

Язык преподавания: русский.

Автор (авторы) программы: преподаватель Высшей школы управления и инноваций МГУ им. М.В. Ломоносова С.С. Ивлев

Преподаватель (преподаватели) программы: преподаватель Высшей школы управления и инноваций МГУ имени М.В. Ломоносова С.С. Ивлев