



**ФЕДЕРАЛЬНОЕ АГЕНТСТВО ВОЗДУШНОГО ТРАНСПОРТА  
(РОСАВИАЦИЯ)**

**ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ БЮДЖЕТНОЕ  
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ  
«САНКТ-ПЕТЕРБУРГСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ  
ГРАЖДАНСКОЙ АВИАЦИИ»**

**УТВЕРЖДАЮ**  
Ректор / Ю.Ю. Михальчевский  
« 23 » \_\_\_\_\_ 2022 года

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ**

**Анализ данных и машинное обучение**

Направление подготовки  
**38.03.01 Экономика**

Направленность программы (профиль)  
**Экономика предприятия и организации воздушного транспорта**

Квалификация выпускника  
**бакалавр**

Форма обучения  
**очная**

Санкт-Петербург  
2022

## **1 Цели освоения дисциплины**

Цель освоения дисциплины «Анализ данных и машинное обучение»: формирование компетенций для успешной профессиональной деятельности выпускника в области экономических и математических знаний, формирование комплексного и научного представления о методах выявления и количественного описания взаимосвязей между различными экономическими показателями, а также закономерностей их изменения, приобретение практических навыков применения аппарата математической статистики, методов анализа данных и алгоритмов машинного обучения, в сочетании с современными информационными технологиями, для обработки массивов эмпирических данных при построении моделей экономических процессов.

Задачами освоения дисциплины являются:

- изучение методов сбора, анализа и обработки данных, необходимых для решения экономических задач;
- изучение методов анализа результатов эконометрического моделирования;
- формирование умения анализировать и интерпретировать данные отечественной и зарубежной статистики о социально-экономических процессах и явлениях, выявлять тенденции изменения социально-экономических показателей;
- формирование навыка работы с инструментальными средствами, предназначенными для обработки экономических данных в соответствии с поставленной задачей.

Дисциплина «Анализ данных и машинное обучение» обеспечивает подготовку обучающегося к решению задач профессиональной деятельности расчетно-экономического, аналитического типа.

## **2 Место дисциплины в структуре ОПОП ВО**

Дисциплина «Анализ данных и машинное обучение» представляет собой дисциплину, относящуюся к Блоку 1 цикла дисциплин ОПОП ВО по направлению подготовки 38.03.01 «Экономика», профиль «Экономика предприятия и организации воздушного транспорта».

Дисциплина «Анализ данных и машинное обучение» базируется на результатах обучения, полученных при изучении дисциплин: «Высшая математика», «Экономическая статистика».

Дисциплина «Анализ данных и машинное обучение» является обеспечивающей для дисциплин: «Планирование и прогнозирование на воздушном транспорте», «Маркетинговые исследования и стратегии на воздушном транспорте», «Финансовый анализ», «Инновации и цифровая трансформация бизнеса».

Дисциплина изучается в 3 и 4 семестрах.

### 3 Компетенции обучающегося, формируемые в результате освоения дисциплины

Процесс освоения дисциплины «Анализ данных и машинное обучение» направлен на формирование следующих компетенций: ОПК-2; ОПК-3; ОПК-5; ОПК-6.

Код компетенции/ индикатора	Результат обучения: наименование компетенции, индикатора компетенции
ОПК-2	Способен осуществлять сбор, обработку и статистический анализ данных, необходимых для решения поставленных экономических задач (ОПК-2)
ИД <sup>3</sup> <sub>ОПК-2</sub>	Применяет статистические, количественные и качественные методы при решении конкретных экономических задач.
ОПК-3	Способен анализировать и содержательно объяснять природу экономических процессов на микро- и макроуровне (ОПК-3)
ИД <sup>2</sup> <sub>ОПК-3</sub>	Анализирует экономические процессы на микро- и макроуровне
ОПК-5	Способен использовать современные информационные технологии и программные средства при решении профессиональных задач (ОПК-5)
ИД <sup>1</sup> <sub>ОПК-5</sub>	Разбирается в многообразии современных информационных технологий.
ИД <sup>2</sup> <sub>ОПК-5</sub>	Использует современные программные средства при решении профессиональных задач
ОПК-6	Способен понимать принципы работы современных информационных технологий и использовать их для решения задач профессиональной деятельности (ОПК-6)
ИД <sup>1</sup> <sub>ОПК-6</sub>	Понимает сущность и принципы работы современных информационных технологий.
ИД <sup>2</sup> <sub>ОПК-6</sub>	Использует современные информационные технологии для решения задач профессиональной деятельности

Планируемые результаты изучения дисциплины:

Знать:

- основные понятия эконометрического подхода, основные методы оценивания неизвестных параметров эконометрических моделей, методы проверки статистических гипотез о параметрах построенных моделей, основные методы диагностики эконометрических моделей;
- методы анализа данных (datamining);
- основные алгоритмы машинного обучения (machine learning);

– особенности анализа экономических данных и методы интерпретации результатов математического моделирования;

– роль больших данных, их источники и методы их исследования при принятии решений для обеспечения устойчивого и безопасного развития экономики; роль человеческого капитала в развитии безопасной цифровой экономики в целом (на макроуровне), и в цифровизации различных хозяйственных процессов (на микроуровне).

Уметь:

– применять стандартные методы построения эконометрических моделей, обрабатывать статистическую информацию и получать статистически обоснованные выводы, делать содержательные выводы из результатов эконометрического моделирования;

– применять основные алгоритмы анализа данных и машинного обучения при решении экономических задач;

– анализировать взаимосвязь развития цифровых технологий и информационных потребностей экономики и общества; выявлять тенденции развития секторов экономики, связанных с созданием, хранением, транзитом и использованием больших данных.

Владеть:

– методами обработки и анализа информации в соответствии с поставленными задачами;

– основными методами обработки статистических данных;

– навыками применения прикладных пакетов программ для ПЭВМ;

– навыками использования источников экономических данных для получения и использования достоверной и актуальной социально-экономической информации в рамках решения профессиональных задач.

#### **4 Объем дисциплины и виды учебной работы**

Общая трудоемкость дисциплины составляет 8 зачетных единиц, 288 академических часов.

Наименование	Всего часов	Семестры	
		3	4
Общая трудоемкость дисциплины	288	144	144
Контактная работа:	127	70,5	56,5
лекции	46	28	18
практические занятия	70	42	28
семинары	–	–	–
лабораторные работы	8	–	8
курсовой проект (работа)	–	–	–
Самостоятельная работа студента	119	65	54

Наименование	Всего часов	Семестры	
		3	4
Промежуточная аттестация:	45	9	36
контактная работа	3	0,5	2,5
самостоятельная работа по подготовке к экзамену, зачету с оценкой	42	8,5	33,5

## 5 Содержание дисциплины

### 5.1 Соотнесения тем (разделов) дисциплины и формируемых компетенций

Темы (разделы) дисциплины	Количество часов	Компетенции				Образовательные технологии	Оценочные средства
		ОПК-2	ОПК-3	ОПК-5	ОПК-6		
3 семестр							
Тема 1. Движущие силы, этапы, инструменты и методы цифровой трансформации общества	15	+	+		+	ВК, Л, ПЗ, СРС	УО, Д
Тема 2. Основные сведения из теории вероятностей, математической статистики, линейной алгебры, теории сигналов и систем	20	+	+	+	+	Л,ПЗ, СРС	УО, Д, РС, ПАР
Тема 3. Инструментальные средства эконометрического исследования	20	+	+	+	+	Л,ПЗ, СРС	УО, Д, РС, ПАР
Тема 4. Предобработка данных	20	+	+	+	+	Л,ПЗ, СРС	УО, Д, РС, ПАР
Тема 5. Регрессионный анализ	20	+	+	+	+	Л,ПЗ, СРС	УО, Д, РС, ПАР
Тема 6. Системы эконометрических уравнений	20	+	+	+	+	Л,ПЗ, СРС	УО, Д, РС, ПАР

Темы (разделы) дисциплины	Количество часов	Компетенции				Образовательные технологии	Оценочные средства
		ОПК-2	ОПК-3	ОПК-5	ОПК-6		
Тема 7. Временные ряды	20	+	+	+	+	Л,ПЗ, СРС	УО, Д, РС, ПАР
Промежуточная аттестация	9						
Итого за 3 семестр	144						
<b>4 семестр</b>							
Тема 8. Динамические эконометрические модели	22	+	+	+	+	Л,ПЗ,ЛР, СРС	УО, РС, ПАР, Т, ЗЛР
Тема 9. Линейные алгоритмы машинного обучения	22	+	+	+	+	Л,ПЗ,ЛР, СРС	УО, РС, ПАР, Т, ЗЛР
Тема 10. Нелинейные алгоритмы машинного обучения	22	+	+	+	+	Л,ПЗ,ЛР, СРС	УО, РС, ПАР, ЗЛР
Тема 11. Глубокое обучение	24	+	+	+	+	Л,ПЗ,ЛР, СРС	УО, РС, ПАР, ЗЛР
Тема 12. Методы прогнозирования	18	+	+	+	+	Л, ПЗ, СРС	УО, РС, ПАР
Промежуточная аттестация	36						
Итого за 4 семестр	144						
Итого по дисциплине	288						

Сокращения: Л – лекция, ПЗ – практическое занятие, ВК – входной контроль, СРС – самостоятельная работа студента, Д – доклад, УО – устный опрос, РС – решение ситуационных задач, ПАР – письменная аудиторная работа, ЛР – лабораторная работа; ЗЛР – защита лабораторной работы, Т – тест.

## 5.2 Темы (разделы) дисциплины и виды занятий

Наименование темы (раздела) дисциплины	Л	ПЗ	С	ЛР	СРС	КР	Всего часов
<b>3 семестр</b>							

Наименование темы (раздела) дисциплины	Л	ПЗ	С	ЛР	СРС	КР	Всего часов
Тема 1. Движущие силы, этапы, инструменты и методы цифровой трансформации общества	4	6			5		15
Тема 2. Основные сведения из теории вероятностей, математической статистики, линейной алгебры, теории сигналов и систем	4	6			10		20
Тема 3. Инструментальные средства эконометрического исследования	4	6			10		20
Тема 4. Предобработка данных	4	6			10		20
Тема 5. Регрессионный анализ	4	6			10		20
Тема 6. Системы эконометрических уравнений	4	6			10		20
Тема 7. Временные ряды	4	6			10		20
Итого за 3 семестр	28	42			65		135
4 семестр							
Тема 8. Динамические эконометрические модели	4	6		2	10		22
Тема 9. Линейные алгоритмы машинного обучения	4	6		2	10		22
Тема 10. Нелинейные алгоритмы машинного обучения	4	6		2	10		22
Тема 11. Глубокое обучение	4	6		2	12		24
Тема 12. Методы прогнозирования	2	4			12		18
Итого за 4 семестр	18	28		8	54		108
Итого за 3 и 4 семестры	46	70	–	8	119	–	243
Промежуточная аттестация							45
Итого по дисциплине							288

Сокращения: Л – лекции, ПЗ – практические занятия, С – семинары, ЛР – лабораторные работы, СРС – самостоятельная работа студента, КР – курсовая работа.

### 5.3 Содержание дисциплины

#### 3 семестр

**Тема 1. Движущие силы, этапы, инструменты и методы цифровой трансформации общества**

Цифровые двойники. Методы моделирования. Основные методы статистического анализа данных. Основные понятия эконометрики. Виды эконометрических моделей. Большие данные (Big Data, BD). Мировые тенденции развития технологий BD. Цели и методы интеллектуального анализа данных (Data Mining, DM). Обзор методов машинного обучения (Machine Learning, ML). Введение в предметную область. Примеры использования методов эконометрики, DM и ML для решения прикладных задач.

## **Тема 2. Основные сведения из теории вероятностей, математической статистики, линейной алгебры, теории сигналов и систем**

Случайные величины и их числовые характеристики. Описательная статистика. Виды распределений. Закон больших чисел и предельные теоремы. Нормальный закон распределения. Точечные и интервальные оценки параметров. Статистические оценки параметров распределения. Выборочные характеристики. Методы статистического вывода. Проверка статистических гипотез. Дисперсионный и корреляционный анализ.

Операции над векторами и матрицами.

Онлайн алгоритмы. Конкурентный анализ.

Основы теории сигналов и систем. Рекурсивные фильтры.

## **Тема 3. Инструментальные средства эконометрического исследования**

Обзор программных средств, используемых для количественного анализа при решении экономических задач. Электронные таблицы. Системы компьютерной математики общего назначения: MATLAB, Scilab, Octave. Системы статистического анализа Statistica, SAS, SPSS, Stata, MINITAB, STADIA, Eviews, Gretl, Prognoz Platform. Языки программирования Python и R. Специализированные библиотеки для DM и ML.

## **Тема 4. Предобработка данных**

Источники экономических данных. Представление статистической информации. Обобщающие статистические показатели. Методы предобработки данных. Извлечение признаков (Feature Extraction). Преобразования признаков (Feature transformations): кодирование нечисловых данных, нормировка и калибровка, заполнение пропусков. Выбор признаков (Feature selection): статистические подходы, визуализация, отбор с использованием моделей.

## **Тема 5. Регрессионный анализ**

Линейная регрессия. Функциональная, статистическая и корреляционная зависимость. Метод наименьших квадратов. Матричное представление. Функции ошибок регрессионных моделей.

Нелинейная регрессия. Типы нелинейности в регрессионной зависимости: нелинейность по экзогенным переменным, нелинейность по параметрам. Сведение нелинейного по переменным уравнения к линейному с помощью преобразований. Смещенность оценок параметров, полученных МНК. Коэффициент



детерминации для нелинейных моделей. Метод последовательных приближений нахождения оценок параметров. Регрессия с фиктивными переменными. Logit- и probit-модели для бинарных эндогенных переменных.

Спецификация переменных в уравнениях множественной регрессии. Мультиколлинеарность факторов. Оценка значимости совместного предельного вклада группы переменных с помощью F-теста.

Гетероскедастичность и автокоррелированность остатков. Взвешенный и обобщённый методы наименьших квадратов. Автокорреляция.

### **Тема 6. Системы эконометрических уравнений**

Классификация систем эконометрических уравнений. Структурная и приведённая формы модели. Проблема идентификации. Идентифицируемые, неидентифицируемые, сверхидентифицируемые модели. Методы оценивания параметров структурной модели: косвенный МНК, двухшаговый МНК, трехшаговый МНК. Метод максимального правдоподобия. Понятие о методе главных компонент, как средстве борьбы с мультиколлинеарностью данных. Примеры применения систем эконометрических уравнений.

### **Тема 7. Временные ряды**

Аддитивная и мультипликативная модели временного ряда. Циклическая, трендовая и случайная компоненты ряда. Задачи эконометрического исследования временных рядов. Автокорреляционная функция ряда и выявление структуры ряда. Аналитическое выравнивание методом скользящей средней. Способы сглаживания: простое и взвешенное среднее, экспоненциальное сглаживание. Моделирование сезонных и циклических колебаний. Моделирование тенденции временного ряда.

Оценка взаимосвязи двух временных рядов. Методы исключения ложной корреляции. Коинтеграция временных рядов.

## **4 семестр**

### **Тема 8. Динамические эконометрические модели**

Явные модели Бокса-Дженкинса. Модели ARIMA и SARIMA. Компоненты авторегрессии и скользящего среднего. Итеративная стратегия разработки модели: проверка стационарности ряда, выбор исходной модели, оценка параметров, анализ остатков. Модель авторегрессии с распределённым лагом первого порядка (ADL модель). Модели с распределённым лагом (DL модели): конечномерные и бесконечномерные. Неявные модели. Сведение модели адаптивных ожиданий к модели авторегрессии.

### **Тема 9. Линейные алгоритмы машинного обучения**

Логистическая регрессия. Линейный дискриминантный анализ. Линейный дискриминант Фишера. Снижение размерности пространства признаков. Задача классификации.

### **Тема 10. Нелинейные алгоритмы машинного обучения**

Алгоритм К-ближайших соседей. Наивный байесовский классификатор. Машины опорных векторов. Деревья классификации и регрессии. Переобучение решающих деревьев. Случайный лес. Построение модели дерева решений для задачи прогнозирования. Построение модели случайного леса на примере задачи кредитного скоринга. Кодирование признаков и заполнение пропущенных данных. Анализ качества построенной модели. Бэггинг. Бустинг. Градиентный бустинг в задаче регрессии. Градиентный бустинг в задаче классификации. Градиентный бустинг над деревьями.

### **Тема 11. Глубокое обучение**

Основные сведения об искусственных нейронных сетях (ИНС). Экономические задачи распознавания образов. Анализ экономических показателей с помощью ИНС. Решение задач классификации с помощью ИНС.

### **Тема 12. Методы прогнозирования**

Исходные понятия прогнозирования, его сущность, предмет и объект. Типология прогнозов. Основные принципы и функции прогнозирования. Классификация методов прогнозирования. Кривые роста. Временной ряд и тренд. Выбор формы кривой. Регрессионный анализ в прогнозировании. Одиночные и связанные временные ряды. Многофакторные модели прогнозирования. Прогнозирование на основе эконометрической модели. Метод экспертных оценок. Прогнозирование макроэкономических показателей.

Нестохастические методы прогнозирования. Универсальное прогнозирование. Калибруемость прогнозов. Прогнозирование с произвольным ядром. Алгоритм взвешенного большинства. Алгоритм взвешивания экспертных решений. Рандомизированные прогнозы. Бустинг. Агрегирующий алгоритм.

Элементы теории игр. Теоретико-игровая интерпретация теории вероятностей. Игры на универсальные предсказания.

## **5.4 Практические занятия**

Номер темы дисциплины	Тематика практических занятий	Трудоемкость (часы)
3 семестр		
1	ПЗ 1. Цифровая трансформация общества.	2
1	ПЗ 2. Обзор эконометрических задач	
1	ПЗ 3. Методы анализа данных и машинного обучения	2
2	ПЗ 4. Основные сведения из теории вероятностей и математической статистики	2
2	ПЗ 5. Основные сведения из линейной алгебры	2

Номер темы дисциплины	Тематика практических занятий	Трудоемкость (часы)
2	ПЗ 6. Онлайн алгоритмы. Основы теории сигналов и систем	2
3	ПЗ 7. Математические пакеты общего назначения. Электронные таблицы	2
3	ПЗ 8. Системы статистического анализа	2
3	ПЗ 9. Языки программирования Python и R. Специализированные библиотеки для анализа данных и машинного обучения	2
4	ПЗ 10. Источники экономических данных	2
4	ПЗ 11. Предобработка статистических данных	2
4	ПЗ 12. Визуализация данных	2
5	ПЗ 13. Линейная парная регрессия	2
5	ПЗ 14. Нелинейная регрессия	2
5	ПЗ 15. Множественная регрессия	2
6	ПЗ 16. Системы эконометрических уравнений	2
6	ПЗ 17. Метод максимального правдоподобия	2
6	ПЗ 18. Метод главных компонент	2
7	ПЗ 19. Исследование временных рядов	2
7	ПЗ 20. Моделирование тенденций и сезонных колебаний	2
7	ПЗ 21. Коинтеграция временных рядов	2
	Итого за 3 семестр	42
4 семестр		
8	ПЗ 22. Модели ARIMA и SARIMA	2
8	ПЗ 23. Линейные нестационарные модели временных рядов	2
8	ПЗ 24. Прогнозирование временных рядов	2
9	ПЗ 25. Логистическая регрессия	2
9	ПЗ 26. Дискриминантный анализ	2
9	ПЗ 27. Решение задач классификации	2
10	ПЗ 28. Алгоритм К-ближайших соседей	2
10	ПЗ 29. Машины опорных векторов	
10	ПЗ 30. Решающие деревья	
11	ПЗ 31. Экономические задачи распознавания образов	
11	ПЗ 32. Анализ экономических показателей с помощью ИНС	2
11	ПЗ 33. Решение задач классификации с помощью ИНС	2
12	ПЗ 34. Многофакторные модели прогнозирования	2

Номер темы дисциплины	Тематика практических занятий	Трудоемкость (часы)
12	ПЗ 35. Агрегирующий алгоритм прогнозирования	2
	Итого за 4 семестр	28
Итого по дисциплине		70

### 5.5 Лабораторный практикум

Номер темы дисциплины	Тематика лабораторных занятий	Трудоемкость (часы)
8	ЛР 1. Модели ARIMA и SARIMA	2
9	ЛР 2. Линейные алгоритмы машинного обучения	2
10	ЛР 3. Нелинейные алгоритмы машинного обучения	2
11	ЛР 4. Глубокое обучение	2
Итого по дисциплине		8

### 5.6 Самостоятельная работа

Номер темы дисциплины	Виды самостоятельной работы	Трудоемкость (часы)
3 семестр		
1	Повторение материалов лекции. Изучение теоретического материала [1,6]. Подготовка к устному опросу. Подготовка к письменной аудиторной работе. Подготовка доклада по выбранной теме.	5
2	Повторение материалов лекции. Изучение теоретического материала [1-3]. Подготовка к устному опросу. Подготовка к письменной аудиторной работе. Подготовка доклада по выбранной теме.	10
3	Повторение материалов лекции. Изучение теоретического материала [1-3, 5]. Подготовка к устному опросу. Подготовка к письменной аудиторной работе. Подготовка доклада по выбранной теме.	10
4	Повторение материалов лекции. Изучение теоретического материала [1-3].	10

Номер темы дисциплины	Виды самостоятельной работы	Трудоемкость (часы)
	Подготовка к устному опросу. Подготовка к письменной аудиторной работе.	
5	Повторение материалов лекции. Изучение теоретического материала [1-3]. Подготовка к устному опросу. Подготовка к письменной аудиторной работе.	10
6	Повторение материалов лекции. Изучение теоретического материала [1-3]. Подготовка к устному опросу. Подготовка к письменной аудиторной работе.	10
7	Повторение материалов лекции. Изучение теоретического материала [1-3, 6]. Подготовка к устному опросу. Подготовка к письменной аудиторной работе.	10
	Итого за 3 семестр	65
4 семестр		
8	Повторение материалов лекции. Изучение теоретического материала [4]. Подготовка к устному опросу. Подготовка к письменной аудиторной работе. Подготовка к лабораторной работе.	10
9	Повторение материалов лекции. Изучение теоретического материала [4]. Подготовка к устному опросу. Подготовка к письменной аудиторной работе. Подготовка к лабораторной работе.	10
10	Повторение материалов лекции. Изучение теоретического материала [4]. Подготовка к устному опросу. Подготовка к письменной аудиторной работе. Подготовка к лабораторной работе.	10
11	Повторение материалов лекции. Изучение теоретического материала [4]. Подготовка к устному опросу. Подготовка к письменной аудиторной работе. Подготовка к лабораторной работе.	12
12	Повторение материалов лекции. Изучение теоретического материала [4]. Подготовка к устному опросу. Подготовка к письменной аудиторной работе.	12
	Итого за 4 семестр	54
Итого по дисциплине		119

## 5.7 Курсовые работы

Курсовые работы учебным планом не предусмотрены.

## 6 Учебно-методическое и информационное обеспечение дисциплины

### а) основная литература:

1. Кремер, Н. Ш. **Эконометрика**: учебник и практикум для вузов / Н. Ш. Кремер, Б. А. Путко ; под редакцией Н. Ш. Кремера. — 4-е изд., испр. и доп. — Москва : Издательство Юрайт, 2022. — 308 с. — (Высшее образование). — ISBN 978-5-534-08710-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/488678> (дата обращения: 27.01.2022).

2. **Эконометрика**: учебник для вузов / И. И. Елисеева [и др.] ; под редакцией И. И. Елисеевой. — М.: Издательство Юрайт, 2022. — 449 с. — (Высшее образование). — ISBN 978-5-534-00313-0. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/488603> (дата обращения: 27.01.2022).

3. Галочкин, В. Т. **Эконометрика**: учебник и практикум для вузов / В. Т. Галочкин. — М.: Издательство Юрайт, 2022. — 293 с. — (Высшее образование). — ISBN 978-5-534-14974-6. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/490094> (дата обращения: 27.01.2022).

4. Алексеев, Д. С. **Технологии интеллектуального анализа данных**: учебник для вузов / Д. С. Алексеев, О. В. Щекочихин. — СПб: Лань, 2022. — 176 с. — ISBN 978-5-8114-8299-3. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/187559> (дата обращения: 27.01.2022). — Режим доступа: для авториз. пользователей.

5. Кузьмин, В. И. **Методы анализа данных**: учебное пособие / В. И. Кузьмин, А. Ф. Гадзаов. — 2-е изд., перераб. и доп. — М.: РТУ МИРЭА, 2020. — 155 с. — Текст : электронный // Лань: электронно-библиотечная система. — URL: <https://e.lanbook.com/book/171433> (дата обращения: 27.01.2022). — Режим доступа: для авториз. пользователей.

### б) дополнительная литература:

6. Кремер, Н. Ш. **Математика для экономистов: от арифметики до эконометрики**. Учебно-справочное пособие : учебник для вузов / Н. Ш. Кремер, Б. А. Путко, И. М. Тришин ; под общей редакцией Н. Ш. Кремера. — 5-е изд., испр. и доп. — М.: Издательство Юрайт, 2022. — 760 с. — (Высшее образование). — ISBN 978-5-534-14218-1. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/488582> (дата обращения: 27.01.2022).

7. **Статистика: учебник для вузов** / под редакцией И. И. Елисеевой. — 3-е изд., перераб. и доп. — М.: Издательство Юрайт, 2022. — 361 с. — (Высшее образование). — ISBN 978-5-534-04082-1. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/488653> (дата обращения: 27.01.2022).

8. **Статистика. Практикум:** учебное пособие для академического бакалавриата / И. И. Елисеева [и др.] ; под редакцией И. И. Елисеевой. — Москва : Издательство Юрайт, 2019. — 514 с. — (Бакалавр. Академический курс). — ISBN 978-5-9916-3688-9. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/425262> (дата обращения: 27.01.2022).

9. Демидова, О. А. **Эконометрика:** учебник и практикум для вузов / О. А. Демидова, Д. И. Малахов. — М.: Издательство Юрайт, 2022. — 334 с. — (Высшее образование). — ISBN 978-5-534-00625-4. — Текст: электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/489325> (дата обращения: 27.01.2022).

10. Антохонова, И. В. **Методы прогнозирования социально-экономических процессов:** учебное пособие для вузов / И. В. Антохонова. — 2-е изд., испр. и доп. — М.: Издательство Юрайт, 2022. — 213 с. — (Высшее образование). — ISBN 978-5-534-04096-8. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://urait.ru/bcode/492661> (дата обращения: 27.01.2022).

11. Губенко, А.В. **Экономика воздушного транспорта [Текст]:** Учебник для вузов. Допущ. УМО /А. В. Губенко, М. Ю. Смуров, Д. С. Черкашин. — СПб.: Питер, 2009. — 288 с. Количество экземпляров 342.

в) перечень ресурсов информационно-телекоммуникационной сети «Интернет»:

12. **Coursera: Специализация «Машинное обучение и анализ данных»** [Электронный ресурс]. – Режим доступа: <https://ru.coursera.org/specializations/machine-learning-data-analysis> , свободный (дата обращения: 27.01.2022).

13. **Федеральная служба государственной статистики.** Официальный сайт Федеральной службы государственной статистики. – Режим доступа: <http://www.gks.ru/>, свободный (дата обращения 27.01.2021).

14. **Министерство финансов РФ** [Электронный ресурс]официальный сайт Министерства финансов РФ. - Режим доступа: <http://www.minfin.ru/>, свободный (дата обращения 20.01.2022)

15. **Правительство РФ** [Электронный ресурс] официальный сайт Правительства РФ. - Режим доступа: <http://www.government.ru/>, свободный (дата обращения 20.01.2022).

16. Воронцов К.В. **Машинное обучение.** Школа Анализа данных. Национальный открытый университет Интуит, 2015. – Режим доступа: <https://www.intuit.ru/studies/courses/13844/1241/info>, свободный (дата обращения 27.01.2021).

17. **Библиотека СПбГУ ГА** [Электронный ресурс]. – Режим доступа: <http://spbguga.ru/objects/e-library/>, свободный (дата обращения 20.01.2022).

г) программное обеспечение (лицензионное), базы данных, информационно-справочные и поисковые системы:

18. **Консультант Плюс**[Электронный ресурс]: официальный сайт компании Консультант Плюс. – Режим доступа: <http://www.consultant.ru/>, свободный (дата обращения 20.01.2022).

19. **Гарант** [Электронный ресурс] официальный сайт компании Гарант. - Режим доступа:<http://www.aero.garant.ru/>, свободный (дата обращения 20.01.2022)

## 7 Материально-техническое обеспечение дисциплины

Для обеспечения образовательного процесса материально-техническими ресурсами используется аудитория №534, оборудованная МОК (мультимедийный обучающий комплекс) - компьютер, проектор, интерактивная доска, а также аудитории №536, 538, 541, 543.

Материалы INTERNET, мультимедийные курсы, оформленные с помощью Microsoft Power Point, используются при проведении лекционных и практических занятий.

Наименование дисциплины	Наименование специальных помещений и помещений для самостоятельной работы	Оснащенность специальных помещений и помещений для самостоятельной работы	Перечень лицензионного программного обеспечения
Анализ данных и машинное обучение	Аудитория 534	Комплект учебной мебели: парты и стулья (местимостъ: 26 посадочных мест) МОК (мультимедийный обучающий комплекс) - компьютер, проектор, интерактивная доска	Microsoft Windows 7 Professional Microsoft Windows Office Professional Plus 2007 Acrobat Professional 9 Windows International Kaspersky Anti-Virus Suite для WKS и FS Konsi- SWOT ANALYSIS Konsi - FOREXSAL
Анализ данных и машинное обучение	Аудитория 536	Комплект учебной мебели Вместимостъ: 26 посадочных мест	
Анализ данных и машинное обучение	Аудитория 538	Комплект учебной мебели Вместимостъ: 24 по-	



		садочных места	
Анализ данных и машинное обучение	Аудитория 541	Комплект учебной мебели Вместимость: 28 посадочных мест	
Анализ данных и машинное обучение	Аудитория 543	Комплект учебной мебели Вместимость: 44 посадочных места	

## 8 Образовательные и информационные технологии

В рамках изучения дисциплины предполагается использовать следующие образовательные технологии.

Входной контроль проводится преподавателем в начале изучения дисциплины с целью коррекции процесса усвоения обучающимися дидактических единиц при изучении базовых дисциплин.

Лекция составляет основу теоретического обучения в рамках дисциплины и направлена на систематизированное изложение накопленных и актуальных научных знаний. Лекция предназначена для раскрытия состояния и перспектив развития экономических знаний в современных условиях. На лекции концентрируется внимание обучающихся на наиболее сложных и узловых вопросах, стимулируется их активная познавательная деятельность.

Ведущим методом в лекции выступает устное изложение учебного материала, который сопровождается одновременной демонстрацией слайдов, при необходимости привлекаются открытые Интернет-ресурсы, а также демонстрационные и наглядно-иллюстрационные материалы и практические примеры.

Цель практических занятий – закрепить теоретические знания, полученные обучающимися на лекциях и в результате самостоятельного изучения соответствующих тем, а также приобрести начальные практические навыки. Рассматриваемые в рамках практического занятия задачи, ситуации, примеры и проблемы имеют профессиональную направленность и содержат элементы, необходимые для формирования компетенций в рамках подготовки обучающихся. Практические занятия предусматривают участие обучающихся в выполнении отдельных элементов работ, связанных с будущей профессиональной деятельностью.

Практические занятия по дисциплине являются составляющими практической подготовки обучающихся, так как предусматривают их участие в выполнении отдельных элементов работ, связанных с будущей профессиональной деятельностью.

Самостоятельная работа обучающихся является составной частью учебной работы. Ее основной целью является формирование навыка самостоятельного приобретения знаний по некоторым вопросам теоретического курса,

закрепление и углубление полученных знаний, самостоятельная работа со справочниками, периодическими изданиями и научно-популярной литературой.

Самостоятельная работа включает выполнение учебных заданий, в том числе и индивидуальных.

## **9 Фонд оценочных средств для проведения текущего контроля успеваемости и промежуточной аттестации по итогам освоения дисциплины**

Уровень и качество знаний обучающихся оцениваются по результатам текущего контроля успеваемости и промежуточной аттестации по итогам освоения дисциплины.

Оценочные средства включают: решение ситуационных задач, письменную аудиторную работу, задания, выдаваемые на самостоятельную работу по темам дисциплины (подготовка докладов), устный опрос пройденного материала.

Устный опрос проводится на практических занятиях с целью контроля усвоения теоретического материала, излагаемого на лекции.

Обсуждение докладов обучающихся проходит в рамках практических занятий по темам дисциплины. Преподаватель, как правило, выступает в роли консультанта при заслушивании докладов, осуществляет контроль полученных обучающимися результатов. Преподаватель следит, чтобы выступление не сводилось к простому воспроизведению текста. Выступление должно строиться свободно, убедительно и аргументировано. При этом обучающийся может обращаться к своим записям, приводить выдержки из периодической печати, сайтов интернета и т. д.

Решение ситуационных задач представляет собой практическое применение теоретических знаний к конкретной хозяйственной ситуации (совокупности хозяйственных операций, осуществляемых в рамках организации).

Письменная аудиторная работа выполняется обучающимися на практических занятиях по индивидуальным вариантам на основании задания, выдаваемого преподавателем по соответствующей теме дисциплины и представляет собой оценку практического применения полученных теоретических знаний.

Контроль выполнения задания, выполняемого на практических занятиях, преследует собой цель своевременного выявления плохо усвоенного материала дисциплины для последующей корректировки.

Промежуточная аттестация по итогам освоения дисциплины проводится в виде зачета с оценкой в 3 семестре и экзамена в 4 семестре. К моменту сдачи экзамена должны быть успешно пройдены предыдущие формы контроля.

Экзамен позволяет оценить уровень освоения компетенций за весь период изучения дисциплины. Билет включает два теоретических вопроса и задачу.

### **9.1. Балльно-рейтинговая система оценки текущего контроля успеваемости и знаний и промежуточной аттестации студентов**

Не применяется.

## **9.2 Методические рекомендации по проведению процедуры оценивания знаний, умений и навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций**

Устный опрос оценивается следующим образом:

«зачтено»: обучающийся дает ответ на поставленный вопрос по существу и правильно отвечает на уточняющие вопросы;

«не зачтено»: обучающийся отказывается отвечать на поставленный вопрос, либо отвечает на него неверно и при формулировании дополнительных (вспомогательных) вопросов.

Тест оценивается: до 60% - неудовлетворительно; 61..75% - удовлетворительно, 76..89% - хорошо; 90..100% - отлично.

Решение ситуационных задач оценивается:

«зачтено»: обучающийся самостоятельно правильно решает задачу, дает обоснованную оценку по итогу решения;

«не зачтено»: обучающийся отказывается от выполнения задачи или не способен ее решить самостоятельно, а также с помощью преподавателя.

Доклад:

«зачтено»: грамотное и непротиворечивое изложение сути вопроса при использовании современных источников. Обучающийся способен сделать обоснованные выводы, а также уверенно отвечать на заданные в ходе обсуждения вопросы;

«не зачтено»: неудовлетворительное качество изложения материала и неспособность обучающегося сделать обоснованные выводы или рекомендации.

Письменная аудиторная работа:

«зачтено»: работа зачитывается в том случае, если задание выполнено полностью, в соответствии с поставленными требованиями и сделаны необходимые выводы;

«не зачтено»: работа не зачитывается в том случае, если обучающийся не выполнил задания, или результат выполнения задания не соответствует поставленным требованиям, а в заданиях и (или) ответах имеются существенные ошибки.

## **9.3 Темы курсовых работ (проектов) по дисциплине**

В учебном плане курсовых работ не предусмотрено.

## **9.4 Контрольные вопросы для проведения входного контроля остаточных знаний по обеспечивающим дисциплинам**

Примерные вопросы входного контроля:

1. Генеральная совокупность и выборка.
2. Вычисление выборочного среднего и дисперсии.
3. Формула для вычисления моментов непрерывной случайной величины.

4. Перечислить основные вероятностные распределения.
5. Привести примеры использования вероятностных методов в экономике.
6. Вычисление доверительных интервалов.

### 9.5 Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания

Компетенции	Показатели оценивания (индикаторы достижения) компетенций	Критерии оценивания
I этап		
ОПК-2	ИД <sup>3</sup> <sub>ОПК-2</sub>	<p>Знать:</p> <ul style="list-style-type: none"> <li>– основные понятия эконометрического подхода, основные методы оценивания неизвестных параметров эконометрических моделей, методы проверки статистических гипотез о параметрах построенных моделей, основные методы диагностики эконометрических моделей;</li> <li>– роль больших данных, их источники и методы их исследования при принятии решений для обеспечения устойчивого и безопасного развития экономики; роль человеческого капитала в развитии безопасной цифровой экономики в целом (на макроуровне), и в цифровизации различных хозяйственных процессов (на микроуровне).</li> </ul> <p>Уметь:</p> <ul style="list-style-type: none"> <li>– применять стандартные методы построения эконометрических моделей, обрабатывать статистическую информацию и получать статистически обоснованные выводы, делать содержательные выводы из результатов эконометрического моделирования;</li> </ul> <p>Владеть:</p> <ul style="list-style-type: none"> <li>– методами обработки и анализа информации в соответствии с поставленными задачами;</li> <li>– основными методами обработки статистических данных</li> </ul>
ОПК-3	ИД <sup>2</sup> <sub>ОПК-3</sub>	<p>Знать:</p> <ul style="list-style-type: none"> <li>– основные понятия эконометрического подхода, основные методы оценивания неизвестных параметров эконометрических моделей, методы проверки статистических гипотез о параметрах построенных моделей, основные методы диагностики эконометрических моделей;</li> </ul> <p>Уметь:</p> <ul style="list-style-type: none"> <li>– применять стандартные методы построения эконометрических моделей, обрабатывать статистическую информацию и получать статистически обоснованные вы-</li> </ul>

Компетенции	Показатели оценивания (индикаторы достижения) компетенций	Критерии оценивания
		<p>воды, делать содержательные выводы из результатов эконометрического моделирования;</p> <p>Владеть:</p> <ul style="list-style-type: none"> <li>– методами обработки и анализа информации в соответствии с поставленными задачами</li> </ul>
ОПК-5	<p>ИД<sup>1</sup><sub>ОПК-5</sub></p> <p>ИД<sup>2</sup><sub>ОПК-5</sub></p>	<p>Знать:</p> <ul style="list-style-type: none"> <li>– основные понятия эконометрического подхода, основные методы оценивания неизвестных параметров эконометрических моделей, методы проверки статистических гипотез о параметрах построенных моделей, основные методы диагностики эконометрических моделей;</li> </ul> <p>Уметь:</p> <ul style="list-style-type: none"> <li>– применять стандартные методы построения эконометрических моделей, обрабатывать статистическую информацию и получать статистически обоснованные выводы, делать содержательные выводы из результатов эконометрического моделирования;</li> <li>– применять основные алгоритмы анализа данных и машинного обучения при решении экономических задач;</li> </ul> <p>Владеть:</p> <ul style="list-style-type: none"> <li>– навыками применения прикладных пакетов программ для ПЭВМ;</li> </ul>
ОПК-6	<p>ИД<sup>1</sup><sub>ОПК-6</sub></p> <p>ИД<sup>2</sup><sub>ОПК-6</sub></p>	<p>Знать:</p> <ul style="list-style-type: none"> <li>– роль больших данных, их источники и методы их исследования при принятии решений для обеспечения устойчивого и безопасного развития экономики; роль человеческого капитала в развитии безопасной цифровой экономики в целом (на макроуровне), и в цифровизации различных хозяйственных процессов (на микроуровне).</li> </ul> <p>Владеть:</p> <ul style="list-style-type: none"> <li>– основными методами обработки статистических данных</li> </ul>
II этап		
ОПК-2	ИД <sup>3</sup> <sub>ОПК-2</sub>	<p>Знать:</p> <ul style="list-style-type: none"> <li>– особенности анализа экономических данных и мето-</li> </ul>

Компетенции	Показатели оценивания (индикаторы достижения) компетенций	Критерии оценивания
		<p>ды интерпретации результатов математического моделирования;</p> <p>Уметь:</p> <ul style="list-style-type: none"> <li>– применять основные алгоритмы анализа данных и машинного обучения при решении экономических задач;</li> <li>– анализировать взаимосвязь развития цифровых технологий и информационных потребностей экономики и общества; выявлять тенденции развития секторов экономики, связанных с созданием, хранением, транзитом и использованием больших данных.</li> </ul> <p>Владеть:</p> <ul style="list-style-type: none"> <li>– навыками использования источников экономических данных для получения и использования достоверной и актуальной социально-экономической информации в рамках решения профессиональных задач.</li> </ul>
ОПК-3	ИД <sup>2</sup> <sub>ОПК-3</sub>	<p>Владеть:</p> <ul style="list-style-type: none"> <li>– методами обработки и анализа информации в соответствии с поставленными задачами;</li> <li>– основными методами обработки статистических данных</li> </ul>
ОПК-5	ИД <sup>1</sup> <sub>ОПК-5</sub> ИД <sup>2</sup> <sub>ОПК-5</sub>	<p>Знать:</p> <ul style="list-style-type: none"> <li>– методы анализа данных;</li> <li>– основные алгоритмы машинного обучения;</li> </ul> <p>Уметь:</p> <ul style="list-style-type: none"> <li>– применять основные алгоритмы анализа данных и машинного обучения при решении экономических задач;</li> </ul> <p>Владеть:</p> <ul style="list-style-type: none"> <li>– методами обработки и анализа информации в соответствии с поставленными задачами;</li> <li>– основными методами обработки статистических данных;</li> <li>– навыками использования источников экономических данных для получения и использования достоверной и актуальной социально-экономической информации в рамках решения профессиональных задач.</li> </ul>
ОПК-6	ИД <sup>1</sup> <sub>ОПК-6</sub>	<p>Уметь:</p> <ul style="list-style-type: none"> <li>– применять стандартные методы построения эконо-</li> </ul>

Компетенции	Показатели оценивания (индикаторы достижения) компетенций	Критерии оценивания
	ИД <sub>ОПК-6</sub> <sup>2</sup>	метрических моделей, обрабатывать статистическую информацию и получать статистически обоснованные выводы, делать содержательные выводы из результатов эконометрического моделирования; Владеть: – методами обработки и анализа информации в соответствии с поставленными задачами

### Шкала оценивания при проведении промежуточной аттестации

*«Отлично»* выставляется обучающемуся, показавшему всесторонние, систематизированные, глубокие знания по рассматриваемой компетенции и умение уверенно применять их на практике при решении задач, свободное и правильное обоснование принятых решений. Отвечая на вопрос, может быстро и безошибочно проиллюстрировать ответ собственными примерами. Обучающийся самостоятельно правильно решает задачу, дает обоснованную оценку итогам решения.

*«Хорошо»* выставляется обучающемуся, если он твердо знает материал, грамотно и по существу излагает его, умеет применять полученные знания на практике, но допускает в ответе или в решении задачи некоторые неточности, хорошо владеет всем содержанием, видит взаимосвязи, но не всегда делает это самостоятельно без помощи преподавателя. Обучающийся решает задачу верно, но при помощи преподавателя.

*«Удовлетворительно»* выставляется обучающемуся, показавшему фрагментарный, разрозненный характер знаний, недостаточно правильные формулировки базовых понятий, нарушения логической последовательности в изложении программного материала, но при этом он владеет основными разделами учебной программы в рамках заданной компетенции, необходимыми для дальнейшего обучения и может применять полученные знания по образцу в стандартной ситуации. Отвечает только на конкретный вопрос, соединяет знания из разных разделов курса только при наводящих вопросах преподавателя. Ситуационная задача решена не полностью, или содержатся незначительные ошибки в расчетах.

*«Неудовлетворительно»* выставляется обучающемуся, который не знает большей части основного содержания учебной программы дисциплины в рамках компетенций, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач. Не раскрыты глубина и полнота при ответах. Задача не решена даже при помощи преподавателя.

## **9.6 Типовые контрольные задания для проведения текущего контроля и промежуточной аттестации по итогам обучения по дисциплине**

### **9.6.1 Примерные контрольные задания для проведения текущего контроля успеваемости**

#### **Примерный перечень вопросов теста (4 семестр)**

1. Что такое линейная регрессия?
  - а) это функциональная зависимость, которая позволяет прогнозировать изменения непрерывных числовых параметров;
  - б) модель зависимости непрерывной переменной  $y$  от объясняющих ее факторов, в которой функция зависимости является линейной;
  - в) модель зависимости дискретной переменной  $y$  от объясняющих ее факторов, в которой функция зависимости является линейной .
  
2. Основная характеристика задачи бинарной классификации:
  - а) классификация осуществляется по одному признаку;
  - б) зависимая переменная может принимать только два значения;
  - в) классификация осуществляется по двум признакам.
  
3. Классификация относится к стратегии:
  - а) обучения без учителя;
  - б) обучения с учителем;
  - в) оба ответа неверны.
  
4. Явление переобучения характеризуется
  - а) чрезмерно точным соответствием модели конкретному набору обучающих примеров, при котором модель теряет способность к обобщению ;
  - б) возникновением, в случае слишком долгого обучения, недостаточного числа обучающих примеров или слишком сложной структуры модели ;
  - в) возникновением, в случае слишком долгого обучения, слишком сложной структуры модели.
  
5. Задача классификации сводится к
  - а) нахождению частых зависимостей между объектами или событиями;



- б) определению класса объекта по его характеристикам;
- в) определению по известным характеристиками объекта значение некоторого его параметра;
- г) поиску независимых групп и их характеристик в всем множестве анализируемых данных.

6. Множество примеров, используемое для конструирования модели, называется

- а) обучающим множеством;
- б) тестовым множеством;
- в) валидационным множеством.

### **Примерный перечень вопросов устного опроса**

3 семестр

1. Предмет исследования эконометрики.
2. Привести классификацию эконометрических задач.
3. Какие методы и зачем используются для предобработки статистических данных?
4. Объективные причины использования и разновидности несплошного наблюдения.
5. Классификация эконометрических переменных.
6. Что называют панельными данными?
7. Описать модель парной регрессии.
8. Каким образом оцениваются значения коэффициентов регрессионной модели?
9. Сформулировать теорему Гаусса-Маркова.
10. Как проводится проверка гипотезы о значимости коэффициентов модели парной регрессии?

4 семестр

1. Перечислить линейные методы машинного обучения и дать им краткую характеристику.
2. Перечислить нелинейные методы машинного обучения и дать им краткую характеристику.

3. Привести пример постановки задачи распознавания образов при решении экономических задач.
4. Привести пример постановки регрессионной задачи при решении экономических задач.
5. Перечислить основные типы ИНС и дать им краткую характеристику.
6. Кратко описать агрегирующий алгоритм прогнозирования.
7. Описать алгоритм идентификации модели временного ряда.
8. Как логистическая регрессия может быть использована при решении экономических задач?

## Примеры практических заданий

### Задание 1.

Ответьте на вопросы о данных по авиарейсам за 4 месяца.

1) Считайте выборку из файла при помощи функции `pd.read_csv` и ответьте на следующие вопросы:

- Имеются ли в данных пропущенные значения?
- Сколько всего пропущенных элементов в таблице "объект-признак"?
- Сколько объектов имеют хотя бы один пропуск?
- Сколько признаков имеют хотя бы одно пропущенное значение?

2) Преобразуйте каждый признак `FeatureName` из указанных в паре новых признаков `FeatureName_Hour`, `FeatureName_Minute`, разделив каждое из значений на часы и минуты. Не забудьте при этом исключить исходный признак из выборки. В случае, если значение признака отсутствует, значения двух новых признаков, его заменяющих, также должны отсутствовать.

3) Некоторые из признаков, отличных от целевой переменной, могут оказывать чересчур значимое влияние на прогноз, поскольку по своему смыслу содержат большую долю информации о значении целевой переменной. Изучите описание набора данных и исключите признаки, сильно коррелирующие с ответами. Ваш выбор признаков для исключения из выборки обоснуйте.

4) Приведите данные к виду, пригодному для обучения линейных моделей. Для этого вещественные признаки надо отмасштабировать, а категориальные — привести к числовому виду. Также надо устранить пропуски в данных. Реа-

лизуйте функцию, которая принимает на вход данные с признаками и выполняет следующие шаги:

- Замена пропущенных значений на нули для вещественных признаков и на строки 'nan' для категориальных.
- Масштабирование вещественных признаков с помощью StandardScaler.
- One-hot-кодирование категориальных признаков с помощью DictVectorizer или функции `pd.get_dummies`. Метод должен возвращать преобразованный DataFrame, который должна состоять из масштабированных вещественных признаков и закодированных категориальных (исходные признаки должны быть исключены из выборки).

5) Разбейте выборку и вектор целевой переменной на обучение и контроль в отношении 70/30.

6) Обучите линейную регрессию на 1000 объектах из обучающей выборки и выведите значения MSE и  $R^2$  и на этой подвыборке и контрольной выборке. Проинтерпретируйте полученный результат — насколько качественные прогнозы строит полученная модель? Какие проблемы наблюдаются в модели?

7) Обучите линейные регрессии с L1- и L2-регуляризатором, подберите лучшее значение параметра регуляризации при помощи кросс-валидации. Выведите значения и на обучающей и контрольной выборках. Удалось ли решить указанные вами ранее проблемы?

## **Задание 2.**

- 1) Разбейте предоставленный Вам преподавателем набор данных на обучающую и тестовую части в соотношении 8:2.
- 2) Проведите предобработку данных: заполнение пропусков, кодирование, масштабирование
- 3). Обучите, а затем провалидируйте на тестовых данных модель логистической регрессии
- 4) Вычислите значения метрик: recall, precision, F1-мера, AUC-ROC. Постройте ROC-кривую.

## **Темы докладов**

1. Использование ортогональных полиномов в представлении и анализе временных рядов.
2. Динамические модели эконометрики на основе уравнений состояния.

3. Рекурсивный метод наименьших квадратов и его применение в экономике.
4. Методы технического анализа курсов валют, акций и других активов.
5. Методы прогнозирования изменения макроэкономических показателей.

### Типовые ситуационные задачи

#### Задача 1.

Имеется информация за 10 лет относительно среднего дохода  $X$  и среднего потребления  $Y$  (млн.руб.):

Годы	90	91	92	93	94	95	96	97	98	99
X	10,5	11,6	12,3	13,7	14,5	16,1	17,3	18,7	20,1	21,8
Y	8,115	10,03	8,409	12,07	12,44	11,35	12,76	13,92	17,28	17,49

1. Оцените коэффициенты линейной регрессии  $Y = \beta_0 + \beta_1 X + \varepsilon$  по методу наименьших квадратов.
2. Проверьте статистическую значимость оценок  $b_0, b_1$  теоретических коэффициентов  $\beta_0, \beta_1$  при уровнях значимости  $\alpha = 0,05$ .
3. Рассчитайте 95%-е доверительные интервалы для теоретических коэффициентов регрессии.
4. Спрогнозируйте потребление при доходе  $X = 19,0$  и рассчитайте 95% доверительный интервал для условного математического ожидания  $M(Y|X = 19,0)$ .
5. Рассчитайте границы интервала, в котором будет сосредоточено не менее 95% возможных объемов потребления при доходе  $X = 19,0$ .
6. Оцените на сколько изменится потребление, если доход вырастет на 3 млн.руб.
7. Рассчитайте коэффициент детерминации  $R^2$ .
8. Рассчитайте  $F$ - статистику для коэффициента детерминации и оцените его статистическую значимость.

#### Задача 2.

Имеется информация за 15 лет относительно среднего дохода  $X$  и среднего потребления  $Y$  (млн.руб.):

Годы	X	Y	Годы	X	Y	Годы	X	Y
1986	10,5	8,8	1991	16,1	11,9	1996	23,1	20,5
1987	11,6	12,0	1992	17,3	13,5	1997	24,3	19,5
1988	12,3	13,0	1993	18,7	15,0	1998	25,5	19,1
1989	13,7	12,6	1994	20,1	18,2	1999	27,8	19,3
1990	14,5	11,2	1995	21,8	21,2	2000	30,0	24,0

1. Оцените коэффициенты линейной регрессии  $Y = \beta_0 + \beta_1 X + \varepsilon$  по методу наименьших квадратов.
2. Вычислите значение  $DW$  статистики Дарбина-Уотсона и проанализируйте наличие автокорреляции остатков.

3. При наличии автокорреляции переоцените уравнение регрессии, используя для этого один цикл метода Кохрана-Оркатта.

### Задача 3.

Известны данные для 30 домохозяйств (в условных единицах) по доходам ( $X$ ) и расходам ( $Y$ ):

X	26	28	31	32	34	35	37	40	41	43
Y	11,2	9,74	12,4	15	12,2	12,1	16,4	14,7	16,4	20,2
X	45	48	49	52	53	54	57	60	61	62
Y	14,9	19,2	23	24,4	21,2	17,8	22,8	28,2	21,6	20,5
X	63	66	67	68	69	70	75	77	79	80
Y	29,6	31	24,8	22,4	22,8	34,9	31,5	30,8	23,3	41,1

1. Оцените коэффициенты линейной регрессии  $Y = \beta_0 + \beta_1 X + \varepsilon$  по методу наименьших квадратов.
2. Примените тест Голдфелда-Квандта для изучения гипотезы об отсутствии гетероскедастичности остатков.
3. В случае гетероскедастичности остатков примените взвешенный метод наименьших квадратов, предполагая, что дисперсии отклонений  $\sigma_i^2$  пропорциональны  $x_i^2$ .
4. Определите, существенно ли повлияла гетероскедастичность на качество оценок в уравнении, построенном по обычному методу наименьших квадратов.

### Задача 4. LDA.

Районы некоторой области характеризуются социально-экономическими показателями, обозначения, наименования и значения которых приведены в табл. Имеются  $p$  обучающих выборок из нормально распределенных генеральных совокупностей с равными ковариационными матрицами. Для каждого варианта приведен набор из пяти показателей для анализа, количество и состав обучающих выборок. Необходимо провести классификацию районов, не вошедших в обучающие выборки, на  $p$  классов и дать экономическую интерпретацию результатов классификации.

## Примерный вариант письменной аудиторной работы

Известны статистические данные наблюдений за некоторый период времени. Требуется:

1. Построить график динамики уровней ряда.
2. Рассчитать значения сезонных компонент методом скользящей средней.
3. Устранить сезонную компоненту из исходных уровней ряда, построить уравнение, моделирующее динамику трендовой компоненты.
4. Проверить остатки на автокорреляцию по критерию Дарбина – Уотсона.
5. Дать прогноз фактора  $y_i$ .

## 9.6.2 Контрольные вопросы промежуточной аттестации по итогам освоения дисциплины

### 3 семестр

1. Основные задачи эконометрики.
2. Линейная парная регрессия.
3. Корреляция и ее использование в эконометрическом анализе.
4. Оценка значимости линейной парной регрессии.
5. Метод наименьших квадратов.
6. Доверительные интервалы для оценок параметров регрессии.
7. Множественная линейная регрессия. Отбор факторов.
8. Оценка значимости множественной регрессии в целом и по отдельным параметрам.
9. Множественная регрессия с фиктивными параметрами.
10. Гетероскедастичность и ее виды.
11. Автокорреляция остатков. Критерий Дарбина-Уотсона.
12. Обобщенный метод наименьших квадратов.
13. Системы линейных одновременных уравнений. Классификация.
14. Структурная и приведенная форма модели.
15. Проблема идентификации. Счетное правило.
16. Косвенный метод наименьших квадратов.
17. Двухшаговый метод наименьших квадратов.
18. Временной ряд. Общие понятия.

### 4 семестр

19. Автокорреляционная функция временного ряда.
20. Выделение регулярных составляющих временного ряда.
21. Прогноз уровней временного ряда.
22. Оценка взаимосвязи двух временных рядов.
23. Модель с распределенным лагом. Лаги Алмон.
24. Модель авторегрессии. Метод инструментальных переменных.
25. Основные понятия машинного обучения. Примеры прикладных задач.
26. Линейные методы классификации и регрессии: функционалы качества, методы настройки, особенности применения.
27. Метрики качества алгоритмов регрессии и классификации.
28. Линейная регрессия. Простая многомерная регрессия. Регрессия с полиномиальными признаками. Методы регуляризации: Ridge, Lasso, ElasticNet.
29. Логистическая регрессия.
30. Деревья решений. Методы построения деревьев. Их регуляризация.
31. Композиции алгоритмов. Разложение ошибки на смещение и разброс.
32. Случайный лес, его особенности.
33. Градиентный бустинг, его особенности при использовании деревьев в качестве базовых алгоритмов.
34. Характеристики нейронных сетей прямого распространения, их сопоставительный анализ с другими нейронными сетями.

35. Принципы и особенности обучения радиально-базисных нейронных сетей.

### Типовые задачи для проведения промежуточной аттестации

#### 3 семестр

**Задача 1.** На основании приведенных данных получена аналитически заданная модель множественной регрессии.

- 1) Дать интерпретацию построенной модели и отдельных ее коэффициентов.
- 2) Найти коэффициенты эластичности, бета-коэффициенты, дельта-коэффициенты, коэффициент детерминации. Объяснить рассчитанные значения.
- 3) Оценить точность модели с помощью средней относительной ошибки аппроксимации.

**Задача 2.** Фирму интересует связь между числом клиентов и расходами на рекламу. Была собрана следующая информация (табл.)

Требуется:

- 1) Построить модель зависимости числа клиентов от затрат на рекламу.
- 2) Пояснить смысл параметров уравнения регрессии.
- 3) Рассчитать прогнозное значение результата, если прогнозное значения фактора X составят 80% от его максимального значения.
- 4) Результаты моделирования и прогнозирования отобразить на графике.

**Задача 3.** В табл. приведены данные по зависимой переменной  $y$ . Требуется проверить наличие гетероскедастичности с помощью параметрического теста Гольдфельда-Квандта при уровне значимости 0,05.

**Задача 4.** Для выборки

$y$	2	-1	1	3	-2	0
$x_1$	1	-1	2	0	2	1
$x_2$	1	2	-1	2	1	2

напишите матрицы  $X$  и  $y$  для модели регрессии

$$y = \beta_0 + \beta_1 x_1 + \beta_2 x_2 + \text{error}.$$

Запишите в матричном виде систему нормальных уравнений.

**Задача 5.** Для мужчин возрастной группы 28-38 лет была оценена модель регрессии зависимости почасовой оплаты  $wage$  от возраста  $age$ , стажа на текущем месте работы  $tenure$ , семейного статуса  $married$ , места жительства  $urban$ , коли-

чества братьев и сестер sibs, и результатов IQ-теста. Зависимая переменная:  $\ln(\text{wage})$ ; объем выборки 935.

	const	age	age <sup>2</sup>	tenure	married	urban	sibs	iq
coeff.	3.3868	0.1166	-0.0015	0.0113	0.1901	0.1879	-0.0092	0.0082
Prob.	0.0408	0.42423	0.3153	0.0000	0.0000	0.0000	0.0987	0.0000

Задание:

- 1 Дайте интерпретацию коэффициентов модели регрессии.
- 2 Дайте интерпретацию коэффициента  $R^2$  и проверьте значимость регрессии «в целом».
- 3 Проверьте значимость влияния фактора возраст и фактора квадрат возраста.
- 4 F-статистика для проверки совместной значимости age и  $age^2$  равна 8.4986. Значимо ли совместное влияние факторов возраст и квадрат возраста?
- 5 Как можно объяснить результаты, полученные в п.3 и п.4?

#### 4 семестр

**Задача 1.** Приведены данные об энергопотреблении зданий. Необходимо выполнить прогноз рейтинга энергопотребления (Energy Star Score) здания и проанализировать, какие признаки оказывают на него сильнейшее влияние.

**Задача 2.** Необходимо восстановить параметры  $a, b, c$  функции  $f$ , имея набор известных пар  $(x_1, f(x_1)), \dots, (x_n, f(x_n))$  и зная, что значения функции определяются по следующей формуле:

$$f(x) = ((a + \varepsilon_a)\sin x + (b + \varepsilon_b)\ln x)^2 + (c + \varepsilon_c)x^2$$

**Задача 3.** Выполнить кластеризацию приведенных данных на числовых признаках, кластер добавить к данным в качестве признака и обучить на этом классификатор. Сравнить метрики с кластером-признаком и без него. Сделать вывод.

**Задача 4.** Дана макроэкономическая кейнсианская модель потребления вида

$$\begin{cases} y = a_1 + b_1(C + D) + \varepsilon_1, \\ C = a_2 + b_2 \cdot y + b_3 \cdot y_{-1} + \varepsilon_2, \end{cases}$$

Информация за девять лет о приростах всех показателей дана в табл.



Год	D	$y_{-1}$	$y$	C
1	-6,8	46,7	3,1	7,4
2	22,4	3,1	22,8	30,4
3	-17,3	22,8	7,8	1,3
4	12,0	7,8	21,4	8,7
5	5,9	21,4	17,8	25,8
6	44,7	17,8	37,2	8,6
7	23,1	37,2	35,7	30,0
8	51,2	35,7	46,6	31,4
9	32,3	46,6	56,0	39,1
Сумма	167,5	239,1	248,4	182,7

Задание:

1. Трансформировать систему к приведенному виду.
2. Провести идентификацию модели.
3. Рассчитать параметры уравнений структурной модели.

### **10 Методические рекомендации для обучающихся по освоению дисциплины**

Методика преподавания дисциплины «Анализ данных и машинное обучение» характеризуется совокупностью методов, приемов и средств обучения, обеспечивающих реализацию содержания и учебно-воспитательных целей дисциплины, которая может быть представлена как некоторая методическая система, включающая методы, приемы и средства обучения. Такой подход позволяет более качественно подойти к вопросу освоения дисциплины обучающимися.

Учебные занятия начинаются и заканчиваются по времени в соответствии с утвержденным режимом СПбГУ ГА в аудиториях согласно семестровым расписаниям теоретических занятий. На занятиях, предусмотренных расписанием, обязаны присутствовать все обучающиеся.

Лекции являются одним из важнейших видов учебных занятий и составляют основу теоретической подготовки обучающихся по дисциплинам. Лекция имеет целью дать систематизированные основы научных знаний по дисциплине, раскрыть состояние и перспективы прогресса конкретной области науки и экономики, сконцентрировать внимание на наиболее сложных и узловых вопросах. Эта цель определяет дидактическое назначение лекции, которое заключается в том, чтобы ознакомить обучающихся с основным содержанием, категориями, принципами и закономерностями изучаемой темы и предмета обучения в целом, его главными идеями и направлениями развития. Именно на лекции формируется научное мировоззрение обучающегося, закладываются теоретические основы фундаментальных знаний будущего управленца, стимулируется его активная познавательная деятельность, решается целый ряд вопросов воспитательного характера.

Каждая лекция должна представлять собой устное изложение лектором основных теоретических положений изучаемой дисциплины или отдельной темы как логически законченное целое и иметь конкретную целевую установку. Особое место в лекционном курсе по дисциплине занимают вводная и заключительная лекции.

Вводная лекция должна давать общую характеристику изучаемой дисциплины, подчеркивать новизну проблем, указывать ее роль и место в системе изучения других дисциплин, раскрывать учебные и воспитательные цели и кратко знакомить обучающихся с содержанием и структурой курса, а также с организацией учебной работы по нему. Заключительная лекция должна давать научно-практическое обобщение изученной дисциплины, показывать перспективы развития изучаемой области знаний, навыков и практических умений.

Практические занятия проводятся в целях выработки практических умений и приобретения навыков при решении управленческих задач. Основным содержанием этих занятий является практическая работа каждого обучающегося. Назначение практических занятий – закрепление, углубление и комплексное применение на практике теоретических знаний, выработка умений и навыков обучающихся в решении практических задач. Вместе с тем, на этих занятиях, осуществляется активное формирование и развитие навыков и качеств, необходимых для последующей профессиональной деятельности. Практические занятия проводятся по наиболее сложным вопросам дисциплины и имеют целью углубленно изучить ее содержание, привить обучающимся навыки самостоятельного поиска и анализа информации, умение делать обоснованные выводы, аргументировано излагать и отстаивать свое мнение. Каждое практическое занятие заканчивается, как правило, кратким подведением итогов, указаниями преподавателя о последующей самостоятельной работе.


Промежуточная аттестация по итогам освоения дисциплины проводится в виде зачета с оценкой в 3 семестре и экзамена в 4 семестре. К моменту сдачи экзамена должны быть успешно пройдены предыдущие формы контроля. Экзамен позволяет оценить уровень освоения компетенций за весь период изучения дисциплины.

Рабочая программа дисциплины составлена в соответствии с требованиями ФГОС ВО по направлению подготовки 38.03.01 «Экономика».

Программа рассмотрена и утверждена на заседании кафедры № 17 «Экономики» «17» мая 2022 года, протокол № 10 .

Разработчик:


К.Т.Н.

  
(ученая степень, ученое звание, фамилия и инициалы разработчика)

Земсков Ю.В.

Заведующий кафедрой № 8 «Прикладной математики и информатики»

Д.Т.Н., доцент


  
(ученая степень, ученое звание, фамилия и инициалы заведующего кафедрой)

Костин Г.А.

Программа согласована:

Руководитель ОПОП

д.э.н., профессор

  
(ученая степень, ученое звание, фамилия и инициалы руководителя ОПОП)

Бородулина С.А.

Программа рассмотрена и одобрена на заседании Учебно-методического совета Университета «дд» июня 2022 года, протокол № 9.