

МИНОБРНАУКИ РОССИИ

федеральное государственное бюджетное образовательное учреждение высшего
образования
«Санкт-Петербургский государственный экономический университет»



УТВЕРЖДАЮ
Проректор по образовательной
деятельности
В.Г. Шубаева
2023 г.

Методы анализа данных

Рабочая программа дисциплины

Направление подготовки/ Специальность 01.03.02 Прикладная математика и информатика
Направленность (профиль) программы/ Специализация Прикладная математика и информатика в экономике и управлении
Уровень высшего образования Бакалавриат
Форма обучения очная
Год набора 2023

Составитель(и):

Старший преподаватель, Ивахненко Дарья Александровна

Часов по учебному плану	180	Виды контроля в семестрах: Экзамен: семестр 7 Курсовая работа: семестр 7
в том числе:		
контактная работа	80	
самостоятельная работа	64	
практическая подготовка	0	
часов на контроль	36	

Распределение часов дисциплины:

Семестр:	7
Вид занятий	Часы
Лекционные занятия	38
Практические занятия	42
Лабораторные работы	0
Итого аудиторных часов	80
Самостоятельная работа	64
Часы на контроль	36
Итого академических часов	180
Общая трудоемкость в зачетных единицах	5

Санкт-Петербург
2023

СОДЕРЖАНИЕ

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ	3
2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ	3
3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ	3
4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ*	3
5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ.....	4
5.1 Рекомендуемая литература	4
5.2 Перечень лицензионного и свободно распространяемого программного обеспечения, в т.ч. отечественного производства	5
5.3 Перечень информационных справочных систем (ИСС) и современных профессиональных баз данных (СПБД).....	5
6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ	6
7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩЕГОСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ.....	7
8. ОСОБЕННОСТИ ОСВОЕНИЯ ДИСЦИПЛИНЫ ДЛЯ ИНВАЛИДОВ И ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ	8
ФОНД ОЦЕНОЧНЫХ СРЕДСТВ.....	10
1.1 Контрольные вопросы и задания к промежуточной аттестации	10
1.2 Темы письменных работ.....	10
1.3 Контрольные точки	11
1.4 Другие объекты оценивания	11
1.5 Самостоятельная работа обучающегося	11
1.6 Шкала оценивания результата	11

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель:	Познакомить обучающихся с современными методами анализа данных и алгоритмами машинного обучения для решения прикладных задач.
--------------	---

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина Б1.В Методы анализа данных относится к части, формируемой участниками образовательных отношений Блока 1.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Код и наименование компетенции выпускника	Код и наименование индикатора достижения компетенций	Планируемые результаты обучения по дисциплине
ПК-2 - Способен выполнить интеллектуальный анализ данных различной природы с использованием современных математических методов, инструментальных средств и языков программирования	ПК-2.1 - Владеет современными математическими методами анализа данных	<p>Знать: подходы к решению прикладных задач анализа данных, основные методы машинного обучения, а также инструментальные средства для их реализации</p> <p>Уметь: применять методы машинного обучения для решения прикладных задач анализа данных.</p> <p>Владеть: инструментальными средствами анализа данных и машинного обучения.</p>

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ*

Номер и наименование тем и/или разделов/тем	Содержание дисциплины	Объем дисциплины (академические часы)			
		Контактная работа			СРО
		ЗЛТ	ПЗ	ЛР	
Тема 1. Постановка задачи машинного обучения. Алгоритм решения прикладных задач анализа данных.	Постановка задачи обучения по прецедентам. Классы задач машинного обучения. Методы оценки качества предсказательных моделей. Алгоритм построения моделей машинного обучения. Понятия недообучения и переобучения.	2	2		4
Тема 2. Линейные модели для задач классификации и регрессии.	Модель линейной регрессии. Модель логистической регрессии. Модель многоклассовой логистической регрессии. Численные методы обучения линейных моделей. Регуляризация линейных моделей.	12	12		20
Тема 3. Метод	Понятие отступа. Разделяющая гиперплоскость.	8	8		12

опорных векторов.	Определение функции потерь в задаче классификации. Оценка сверху для пороговой функции потерь. Формулировка задачи линейного программирования для нахождения оптимальной разделяющей полосы. Решение задачи классификации методом множителей Лагранжа. Типы двойственных переменных в методе опорных векторов. Ядра и спрямляющие пространства. Решение задачи регрессии методом опорных векторов.				
Тема 4. Метрические методы для задач классификации и регрессии.	Понятия компактности и функции расстояния в метрическом классификаторе. Обобщенный метрический классификатор. Виды функций расстояния в прикладных задачах. Метрические алгоритмы в задачах регрессии. Метод ближайших соседей для решения задач классификации и регрессии. Метод окна Парзена. Метод окна Парзена переменной ширины. Проклятие размерности.	4	4		8
Тема 5. Логические методы для задач классификации и регрессии. Композиции алгоритмов.	Бинарное решающее дерево. Понятие критерия информативности: энтропийный критерий, критерий Джини, дисперсионный критерий. Алгоритм построения решающего дерева ID3. Недостатки жадных алгоритмов построения решающих деревьев. Алгоритм C4.5. Обработка пропусков в деревьях решений. Методы обработки категориальных переменных. Пороговые правила для количественных переменных. Алгоритм CART. Теорема Кондорсе. Метод бэггинга Бреймана. Метод случайных подпространств. Случайный лес. Алгоритм AdaBoost. Алгоритм градиентного бустинга.	12	16		20
Контроль:					36
Всего по дисциплине:		38	42	0	64

*ЗЛТ – занятия лекционного типа, ПЗ – все виды занятий семинарского типа, кроме лабораторных работ, ЛР – лабораторные работы, СРО – самостоятельная работа обучающегося

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1 Рекомендуемая литература

Библиографическое описание издания (автор, заглавие, вид, место и год издания, кол. стр.)	Электронные ресурсы
Бринк, Х. Машинное обучение / Х. Бринк, Д. Ричардс, М. Феверолф. Санкт-Петербург : Питер, 2017. - 336 с.	http://ibooks.ru/reading.php?short=1&productid=355472
Гуриков С.Р. Основы алгоритмизации и программирования на Python : учебное пособие. — Электрон. дан. – М.: ФОРУМ:	http://znanium.com/go.php?id=970143

ИНФРА-М, 2019. – 343 с.	
Плас, Дж. Вандер. Python для сложных задач: наука о данных и машинное обучение / Дж. Вандер Плас. Санкт-Петербург : Питер, 2018. - 576 с.	https://ibooks.ru/bookshelf/376830/reading
Ниворожкина Л.И. Статистические методы анализа данных : учебник / Л.И. Ниворожкина, С.В. Арженовский, А.А. Рудяга [и др.] ; под общ. ред. Л.И. Ниворожкиной . — Электрон. дан. - М.; Издательский Центр РИОР: ООО "Научно-издательский центр ИНФРА-М", 2016. - 333 с.	http://znanium.com/go.php?id=556760
Вьюгин, В. В. Математические основы машинного обучения и прогнозирования: Учебное пособие / Вьюгин В.В. - Москва :МЦНМО, 2014. - 304 с.	https://znanium.com/catalog/product/958689

5.2 Перечень лицензионного и свободно распространяемого программного обеспечения, в т.ч. отечественного производства

- 7-Zip
- Python
- Anaconda Individual Edition
- LibreOffice
- Adobe Reader
- ОС Альт образование 10

5.3 Перечень информационных справочных систем (ИСС) и современных профессиональных баз данных (СПБД)

№	Наименование СПБД/ ИСС
1.	Электронная библиотека Grebennikon.ru – www.grebennikon.ru
2.	Научная электронная библиотека eLIBRARY – www.elibrary.ru
3.	Научная электронная библиотека КиберЛеника – www.cyberleninka.ru
4.	База данных ПОЛПРЕД Справочники – www.polpred.com
5.	База данных OECD Books, Papers & Statistics на платформе OECD iLibrary www.oecd-ilibrary.org
6.	Справочная правовая система КонсультантПлюс (инсталлированный ресурс СПБГЭУ или www.consultant.ru)
7.	Справочная правовая система «ГАРАНТ» (инсталлированный ресурс СПБГЭУ или www.garant.ru)
8.	Информационно-справочная система «Кодекс» (инсталлированный ресурс

	СПбГЭУ или www.kodeks.ru)
9.	Электронная библиотечная система BOOK.ru - www.book.ru
10.	Электронная библиотечная система ЭБС ЮРАЙТ – www.urait.ru
11.	Электронно-библиотечная система ЗНАНИУМ (ZNANIUM) – www.znanium.com
12.	Электронная библиотека СПбГЭУ – opac.unecon.ru

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Для реализации данной дисциплины имеются специальные помещения для проведения занятий лекционного типа, занятий семинарского типа, курсового проектирования (выполнения курсовых работ) групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, а также помещения для самостоятельной работы.

Помещения оснащены оборудованием и техническими средствами обучения.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду университета.

Наименование учебных аудиторий, перечень	Адрес (местоположение) учебных аудиторий
Ауд. 2023 Компьютерный класс (для проведения практических занятий, курсового проектирования (выполнения курсовых работ) с применением вычислительной техники). Оборудован мультимедийным комплексом. Специализированная мебель и оборудование: Учебная мебель на 48 посадочных мест, рабочее место преподавателя (компьютерный стол - 1 шт.), доска маркерная на колесиках - 1 шт., доска маркерная 3-х секционная - 1 шт., кафедра - 1 шт., стол - 1 шт., стул изо - 7 шт., стул -1 шт., жалюзи -3 шт., Компьютер i5-8400/8GB/500GB_SSD/Viewsonic VA2410-mh -34 шт., Коммутатор Cisco Catalyst 2960-48PST-L (в т.ч. Сервисный контракт SmartNet CON-SNT-2964STL) - 1 шт., Точка беспроводного доступа Wi-Fi Тип1 UBIQUITI UAP-AC-PRO - 1 шт., Проектор NEC M350 X - 1 шт., Коммутатор локальной вычислительной сети (48 портов) Cisco WS-C2960+48PST-L - 1 шт., Коммутатор ProCurve Switch 2626 - 1 шт., Компьютер Intel pentium x2 g3250 /500gb/монитор philips 21.5" - 1 шт., IP видекамера Ubiquiti - 1 шт., Беспроводная точка доступа/UNI FI AP PRO/Ubiquiti - 1 шт. Наборы демонстрационного оборудования и учебно-наглядных пособий: мультимедийные приложения к лекционным курсам и практическим занятиям, интерактивные учебно-наглядные пособия.	191023, г. Санкт-Петербург, ул. Канал Грибоедова, 30/32, литер «А», «Б», «Р»
Ауд. 2022 Лаборатория "Лабораторный комплекс" Специализированная мебель	191023, г. Санкт-

и оборудование: Учебная мебель на 19 посадочных мест (19 компьютерных стола, 19 черных кресла) рабочее место преподавателя (компьютерный стол 1шт., кресло 1шт.), доска меловая односекционная 1шт., доска маркерная на колесиках 1 шт., стол 1шт., стул 1шт., жалюзи 1шт., вешалка стойка 1шт. Компьютер Intel i5 4460/1Тб/8Гб/монитор Samsung 23" - 1 шт., Компьютер Intel i5 4460/1Тб/8Гб/ монитор Samsung 23" - 18 шт. Наборы демонстрационного оборудования и учебно-наглядных пособий: мультимедийные приложения к лекционным курсам и практическим занятиям, интерактивные учебно-наглядные пособия.	Петербург, ул. Канал Грибоедова, 30/32, литер «А», «Б», «Р»
---	---

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩЕГОСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ

Приступая к изучению дисциплины, обучающемуся необходимо ознакомиться со следующими документами:

- учебно-методической документацией;
- локальными нормативными актами, регламентирующими основные вопросы организации и осуществления образовательной деятельности, в том числе регламентирующие порядок проведения текущего контроля успеваемости и промежуточной аттестации обучающихся;
- графиком консультаций сотрудников профессорско-преподавательского состава.

Уровень и глубина освоения дисциплины определяются активной и систематической работой обучающихся на лекционных занятиях, занятиях семинарского типа, выполнением самостоятельной работы, в том числе в части выделения наиболее значимых и актуальных проблем для дальнейшего изучения. Особым условием качественного освоения дисциплины является эффективная организация труда, позволяющая распределить учебную нагрузку равномерно в соответствии с графиком учебного процесса.

При подготовке к учебным занятиям обучающимся предоставляется возможность посещения консультаций сотрудников профессорско-преподавательского состава СПбГЭУ согласно расписанию, установленному в графике консультаций.

Аудиторная и внеаудиторная работа обучающихся должна быть направлена на формирование:

- фундаментальных основ мировоззрения обучающихся и естественнонаучного познания;
- базисных знаний, соответствующих направлению подготовки и заявленной профессиональной области, формирующих целевую и профессиональную основу для подготовки кадров;

- профессиональных компетенций ориентированных на удовлетворение потребностей рынка труда;
- индивидуальной траектории посредством освоения уникального набора профессиональных компетенций дополняющих компетентностную модель обучающегося, за счет ориентации на конкретные профессиональные специализированные области знаний, определяемые представителями рынка труда;
- метанавыков обучающихся, таких как: командная работа и лидерство, анализ данных, цифровые навыки, разработка и реализация проектов, межкультурное взаимодействие.

8. ОСОБЕННОСТИ ОСВОЕНИЯ ДИСЦИПЛИНЫ ДЛЯ ИНВАЛИДОВ И ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ

Обучение обучающихся с ограниченными возможностями здоровья при необходимости осуществляется на основе адаптированной рабочей программы с использованием специальных методов обучения и дидактических материалов, составленных с учетом особенностей психофизического развития, индивидуальных возможностей и состояния здоровья таких обучающихся (обучающегося).

В целях освоения учебной программы дисциплины инвалидами и лицами с ограниченными возможностями здоровья Университет обеспечивает:

- для инвалидов и лиц с ограниченными возможностями здоровья по зрению: размещение в доступных для обучающихся, являющихся слепыми или слабовидящими, местах и в адаптированной форме справочной информации о расписании учебных занятий; присутствие ассистента, оказывающего обучающемуся необходимую помощь; выпуск альтернативных форматов методических материалов (крупный шрифт или аудиофайлы);

- для инвалидов и лиц с ограниченными возможностями здоровья по слуху: надлежащими звуковыми средствами воспроизведение информации;

- для инвалидов и лиц с ограниченными возможностями здоровья, имеющих нарушения опорно-двигательного аппарата: возможность беспрепятственного доступа обучающихся в учебные помещения, туалетные комнаты и другие помещения кафедры, а также пребывание в указанных помещениях.

Обучающиеся из числа инвалидов и лиц с ОВЗ обеспечены печатными и (или) электронными образовательными ресурсами в формах, адаптированных к ограничениям их здоровья. Образование обучающихся с

ограниченными возможностями здоровья может быть организовано как совместно с другими обучающимися, так и в отдельных группах или в отдельных организациях.

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

1.1 Контрольные вопросы и задания к промежуточной аттестации

- 1 Постановка задачи обучения по прецедентам. Классы задач машинного обучения. Методы оценки качества предсказательных моделей. Алгоритм построения моделей машинного обучения.
- 2 Теоретическое обоснование модели линейной регрессии.
- 4 Модель логистической регрессии для задачи бинарной классификации.
- 6 Методы борьбы с переобучением линейных моделей.
- 7 Метод опорных векторов для задачи бинарной классификации.
- 8 Метод опорных векторов для задачи многоклассовой классификации.
- 10 Понятия компактности и функции расстояния в метрическом классификаторе. Обобщенный метрический классификатор.
- 11 Метод ближайших соседей для решения задач классификации и регрессии.
- 13 Бинарное решающее дерево. Понятие критерия информативности: энтропийный критерий, критерий Джини, дисперсионный критерий. Алгоритм построения решающего дерева ID3.
- 14 Недостатки жадных алгоритмов построения решающих деревьев. Алгоритм C4.5.
- 15 Обработка пропусков в деревьях решений. Методы обработки категориальных переменных. Пороговые правила для количественных переменных.
- 16 Алгоритм построения решающего дерева CART.
- 17 Теорема Кондорсе. Метод бэггинга Бреймана. Метод случайных подпространств. Случайный лес.
- 18 Алгоритм AdaBoost. Выбор функции потерь.
- 19 Алгоритм градиентного бустинга.
- 20 Выбор функции потерь в алгоритме градиентного бустинга. Методы преобразования входных данных.

1.2 Темы письменных работ

- 1 Разработка и реализация алгоритма рекомендаций для краудсорсинговой платформы.
- 2 Реализация и применение алгоритмов восстановления пробелов в тексте.
- 3 Разработка и реализация алгоритма восстановления пробелов в тексте на основе методов машинного обучения.
- 4 Применение средств анализа данных для обнаружения браконьерства.
- 5 Разработка и реализация рекомендательной системы для онлайн-игры.
- 6 Применение методов анализа данных в командных видах спорта.
- 7 Анализ и прогнозирование данных спроса на услуги такси.
- 8 Разработка и реализация алгоритма распознавания аудиофрагментов.
- 9 Применение методов машинного обучения для распознавания языка жестов.
- 10 Реализация языковых моделей для автодополнения слов и обработки опечаток.
- 11 Решение задачи распознавания животных по снимкам с фотоловушек.
- 12 Разработка и реализация алгоритма обнаружения плагиата.

- 13 Применение методов машинного обучения для прогнозирования вероятностей попадания вагонов в ремонт.
- 14 Разработка и реализация алгоритма восстановления изображения из фрагментов.
- 15 Применение методов машинного обучения для анализа состояния водителя.
- 16 Реализация моделей распознавания эмоций человека по аудиофрагменту.
- 17 Разработка и реализация алгоритма агрегации оценок исполнителей на краудсорсинговой платформе.
- 18 Разработка и реализация алгоритма агрегации расшифровок аудиозаписей на краудсорсинговой платформе.
- 19 Применение методов машинного обучения для определения количества калорий по фотографии блюда.
- 20 Применение методов машинного обучения для прогнозирования потребления газа.

1.3 Контрольные точки

Номер контрольной точки	Тип контрольной точки	Способ проведения	Номера тем
1	Контрольная работа	с помощью технических средств и информационных систем	1-3
2	Контрольная работа	с помощью технических средств и информационных систем	4-5
3	Текущий контроль	с помощью технических средств и информационных систем	1-5

1.4 Другие объекты оценивания

Рабочей программой дисциплины не предусмотрено.

1.5 Самостоятельная работа обучающегося

Наименования самостоятельной работы	Номера тем
Выполнение домашних заданий	1-5
Курсовое проектирование	1-5
Выполнение расчетных, аналитических, расчетно-графических и др. заданий	1-5
Подготовка к лекционным и практическим занятиям	1-5
Подготовка к экзамену	1-5

1.6 Шкала оценивания результата

Шкалы оценивания и процедуры оценивания результатов обучения по дисциплине регламентируются Положением о текущем контроле

успеваемости и промежуточной аттестации обучающихся по программам высшего образования и Положением о балльно-рейтинговой системе.

Для оценки сформированности результатов обучения по дисциплине используется **балльно-рейтинговая система успеваемости обучающихся**:

Формой итогового контроля по дисциплине является экзамен (или дифференцированный зачет), итоговая оценка формируется в соответствии со шкалой, приведенной ниже в таблице:

Баллы	Оценка
≤ 54	неудовлетворительно
55-69	удовлетворительно
70-84	хорошо
≥ 85	отлично

Шкала оценивания результата

2 (балл до 54)	Демонстрирует непонимание проблемы. Многие требования, предъявляемые к заданию не выполнены. Демонстрируется первичное восприятие материала. Работа незакончена и /или это плагиат.
3 (балл 55-69)	Демонстрирует частичное понимание проблемы. Большинство требований, предъявляемых, к заданию выполнены. Владение элементами заданного материала. В основном выполненный материал понятен и носит целостный характер.
4 (балл 70-84)	Демонстрирует значительное понимание проблемы обозначенной дисциплиной. Все требования, предъявляемые к заданию выполнены. Содержание выполненных заданий раскрыто и рассмотрено с разных точек зрения.
5 (балл 85-100)	Демонстрирует полное понимание проблемы. Все требования, предъявляемые к заданию выполнены. Продemonстрировано уверенное владение материалом дисциплины. Выполненные задания носят целостных характер, выполнены в полном объеме, структурированы, представлены различные точки зрения, продемонстрирован творческий подход.