МИНОБРНАУКИ РОССИИ

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

УТВЕРЖДАЮ
Проректор по учебной и методической работе
_____ В.Г. Шубаева
«23» июня 2021 г.

Математические основы анализа данных

Рабочая программа дисциплины

Направление подготовки/ *Специальность*	*01.04.02 Прикладная математика и информатика*
Направленность (профиль) программы/ *Специализация*	*Математическое и компьютерное моделирование в экономике и управлении*
Уровень высшего образования	*Магистратура*
Форма обучения	*очная*
Год набора	*2022*

Составитель*(и)*:

Старший преподаватель, Ивахненко Дарья Александровна

Часов по учебному плану	216	**Виды контроля в семестрах:**
в том числе:		
контактная работа	42	Экзамен: семестр 2
самостоятельная работа	138	
практическая подготовка	0	
часов на контроль	36	

Распределение часов дисциплины:

Семестр:	2
Вид занятий	Часы
Лекционные занятия	18
Практические занятия	24
Лабораторные работы	
Итого аудиторных часов	**42**
Самостоятельная работа	138
Часы на контроль	36
Итого академических часов	**216**
Общая трудоемкость в зачетных единицах	**6**

Санкт-Петербург
2022

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* 4

5. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ 5

5.1 Рекомендуемая литература 5

5.2 Перечень лицензионного и свободно распространяемого программного обеспечения, в т.ч. отечественного производства 6

5.3 Перечень информационных справочных систем (ИСС) и современных профессиональных баз данных (СПБД) 6

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ 6

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩЕГОСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ 8

8. ОСОБЕННОСТИ ОСВОЕНИЯ ДИСЦИПЛИНЫ ДЛЯ ИНВАЛИДОВ И ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ 9

ФОНД ОЦЕНОЧНЫХ СРЕДСТВ 11

1.1 Контрольные вопросы и задания к промежуточной аттестации 11

1.2 Темы письменных работ 11

1.3 Контрольные точки 11

1.4 Другие объекты оценивания 12

1.5 Самостоятельная работа обучающегося 12

1.6 Шкала оценивания результата 12

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель:	Представить обучающимся методы и средства анализа данных, развить навыки применения методов анализа данных для решения прикладных задач.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина Б1.О Математические основы анализа данных относится к обязательной части Блока 1.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Код и наименование компетенции выпускника	**Код и наименование индикатора достижения компетенций**	**Планируемые результаты обучения по дисциплине**
ОПК-2 - Способен совершенствовать и реализовывать новые математические методы решения прикладных задач	ОПК-2.2 - Совершенствует и реализует современные математические методы анализа данных, разрабатывает новые алгоритмы	Знать: теоретические основы методов анализа данных и области их применения Уметь: применять математический аппарат для разработки и построения моделей машинного обучения. Владеть: средствами для разработки и реализации моделей машинного обучения.
ОПК-4 - Способен комбинировать и адаптировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности	ОПК-4.3 - Использует современные системы управления базами данных для формирования информационной составляющей при решении прикладных задач	Знать: основные принципы построения предсказательных и прогнозных моделей Уметь: проводить предобработку данных для построения предсказательных и прогнозных моделей. Владеть: инструментами анализа данных и навыками построения и оценки качества математических моделей анализа данных.
ПК-1 - Способен собрать информацию, организовать ее хранение и обработку на основе современных информационно-коммуникационных	ПК-1.2 - Готовит аналитические отчеты и/или научные публикации, используя	Знать: методы предварительного и визуального анализа данных различной природы Уметь: формулировать гипотезы и строить выводы относительно проведенного анализа данных. Владеть: навыками анализа данных различной природы

технологий, подготовить аналитический отчет и/или научную публикацию	современные средства визуализации и стандарты оформления	с помощью компьютерных средств.
УК-1 - Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий	УК-1.2 - Находит и критически анализирует информацию, необходимую для решения поставленной задачи, сопоставляет разные источники информации с целью выявления их противоречий и поиска достоверных суждений	Знать: современные подходы к анализу данных и актуальные области исследований Уметь: применять полученные знания для совершенствования существующих подходов к анализу данных. Владеть: математическим и компьютерным инструментарием для решения поставленных задач.

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 обучения по прецедентам. Средства анализа данных	Постановка задачи машинного обучения. Классы задач. Примеры задач. Типы данных и методы их обработки. Компьютерные средства анализа данных.	4			12
Тема 2. Первичный и визуальный анализ данных	Пакеты анализа данных в языке программирования Python. Операции над массивами данных. Методы извлечения данных. Индексация данных. Группировка данных. Статистические методы анализа данных. Основные способы визуализации вещественных признаков. Методы визуализации категориальных признаков. Визуализация основных статистик признаков. Методы заполнения пропусков в данных. Методы обнаружения и обработки выбросов.	4	8		36
Тема 3. Линейные предсказательные и	Теоретическое обоснование модели линейной регрессии. Модель логистической регрессии.	4	8		36

прогнозные модели	Модель многоклассовой логистической регрессии. Обобщенные линейные модели. Методы обучения линейных моделей. Метод опорных векторов. Особенности применения линейных моделей в задачах прогнозирования.				
Тема 4. Метрические методы машинного обучения	Обобщенный метрический классификатор. Виды функций расстояния в прикладных задачах анализа данных. Частные случаи метрических алгоритмов в задачах классификации регрессии: метод ближайших соседей, метод окна Парзена. Проклятие размерности.	2	2		18
Тема 5. Логические алгоритмы машинного обучения	Бинарное решающее дерево. Выбор критерия информативности в задачах классификации и регрессии. Алгоритмы построения решающего дерева. Определение набора предикатов для решающего дерева. Обработка пропусков в решающем дереве. Методы кодирования категориальных переменных. Теорема Кондорсе. Метод бэггинга Бреймана. Метод случайных подпространств. Случайный лес. Алгоритмы AdaBoost и LogitBoost. Алгоритм градиентного бустинга. Выбор функции потерь в алгоритме градиентного бустинга.	4	6		36
Контроль:					**36**
Всего по дисциплине:		**18**	**24**	**0**	**138**

*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
Плас, Дж. Вандер. Python для сложных задач: наука о данных и машинное обучение / Дж. Вандер Плас. Санкт-Петербург : Питер, 2018. - 576 с.	http://ibooks.ru/reading.php?short=1&productid=356721
Вьюгин, В. В. Математические основы машинного обучения и прогнозирования: Учебное пособие / Вьюгин В.В. - Москва	https://znanium.com/catalog/product/958689

:МЦНМО, 2014. - 304 с.	
Прохоренок, Н.А.Python 3 и PyQt 5. Разработка приложений / Н.А. Прохоренок, В.А. Дронов. Санкт-Петербург : БХВ-Петербург, 2016. - 832 с.	http://ibooks.ru/reading.php?short=1&productid=353570

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

- 7-Zip
- Microsoft Office Professional
- Microsoft Windows Professional
- Anaconda
- Python

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 eLIBRARRY – 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. 309 Учебная аудитория (для проведения занятий лекционного типа и занятий семинарского типа, курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации), оборудована мультимедийным комплексом.Специализированная мебель и оборудование: Учебная мебель на 96 посадочных мест; доска меловая - 1 шт.; тумба - 1 шт.; Компьютер Intel Core 2 Duo E7300 2.6/2Gb/120Gb/19Samsung 943N - 1 шт., Мультимедийный проектор Optoma EX-632 - 1 шт., Экран 183x240 в доп.комплект. - 1 шт., Акустическая система JBL CONTROL 25 WH - 2 шт., Микшер-усилитель TA-1120 - 1 шт. Наборы демонстрационного оборудования и учебно-наглядных пособий: мультимедийные приложения к лекционным курсам и практическим занятиям, интерактивные учебно-наглядные пособия.	196084, г. Санкт-Петербург, Московский пр., д. 103, лит. А, пом. 1Н, 2Н
Ауд. 210 Учебная аудитория (для проведения занятий лекционного типа и занятий семинарского типа, курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации), оборудована мультимедийным комплексом.Специализированная мебель и оборудование: Учебная мебель на 27 посадочных мест; доска меловая - 1 шт.; тумба - 1 шт.; Компьютер HP GQ652AW#ACB dc7800 USDT E 6550 1.0G.DVD-ROM/ 2Gb/80Gb - 1 шт., Проектор NEC M350X в компл. - 1 шт., Акустическая система JBL CONTROL 25 WH (пара колонок) - 1 шт., Микшер-усилитель TA-1120 - 1 шт., Экран с электроприводом Screen Media Champion 203x153cm. MW 4:3. 4-уг. корпус - 1 шт. Наборы демонстрационного оборудования и учебно-наглядных пособий: мультимедийные приложения к лекционным курсам и практическим занятиям, интерактивные учебно-наглядные пособия.	196084, г. Санкт-Петербург, Московский пр., д. 103, лит. А, пом. 1Н, 2Н
Ауд. 302 Учебная аудитория (для проведения занятий лекционного типа и занятий семинарского типа, курсового проектирования (выполнения курсовых работ), групповых и индивидуальных консультаций, текущего	196084, г. Санкт-Петербург, Московский пр., д.

контроля и промежуточной аттестации), оборудована мультимедийным комплексом.Специализированная мебель и оборудование: Учебная мебель на 48 посадочных мест; доска меловая 1 шт.; тумба; Компьютер Intel i3 2100 3.1/2Gb/500 Gb - 1шт., Проектор цифровой Acer X1240 - 1 шт., Акустическая система - 2 шт., Экран Projecta Compact 153[200 см с эл\привод. - 1 шт., Микшер-усилитель (JPA-1120A) 120 Вт/100 В - 1 шт. Наборы демонстрационного оборудования и учебно-наглядных пособий: мультимедийные приложения к лекционным курсам и практическим занятиям, интерактивные учебно-наглядные пособия.	103, лит. А, пом. 1Н, 2Н

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 машинного обучения. Классы задач. Примеры задач. Типы данных и методы их обработки.
2 Статистические методы анализа данных. Основные способы визуализации вещественных признаков. Методы обнаружения и обработки выбросов.
3 Теоретическое обоснование модели линейной регрессии.
4 Модель логистической регрессии.
5 Модель многоклассовой логистической регрессии.
6 Обобщенные линейные модели. Канонический вид плотности распределения. Функция связи.
7 Методы обучения линейных моделей.
8 Метод опорных векторов для задачи классификации.
9 Метод опорных векторов для задачи регрессии.
10 Особенности применения линейных моделей в задачах прогнозирования.
11 Обобщенный метрический классификатор. Виды функций расстояния в прикладных задачах анализа данных.
12 Обобщенный метрический классификатор. Виды функций расстояния в прикладных задачах анализа данных.
13 Бинарное решающее дерево. Выбор критерия информативности в задачах классификации и регрессии. Определение набора предикатов для решающего дерева.
14 Алгоритмы построения решающего дерева.
15 Обработка пропусков в решающем дереве. Методы кодирования категориальных переменных.
16 Метод бэггинга Бреймана. Метод случайных подпространств. Случайный лес.
17 Алгоритм AdaBoost.
18 Алгоритм LogitBoost.
19 Алгоритм градиентного бустинга.
20 Выбор функции потерь в алгоритме градиентного бустинга.

1.2 Темы письменных работ

Рабочей программой дисциплины не предусмотрено.

1.3 Контрольные точки

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

		и информационных систем	

1.4 Другие объекты оценивания

Рабочей программой дисциплины не предусмотрено.

1.5 Самостоятельная работа обучающегося

Наименования самостоятельной работы	Номера тем
Выполнение домашних заданий	2-5
Подготовка к экзамену	1-5
Подготовка к лекционным и практическим занятиям	1-5
Выполнение расчетных, аналитических, расчетно-графических и др. заданий	2-5
Решение профессиональных задач	2-5

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. Продемонстрировано уверенное владение материалом дисциплины. Выполненные задания носят целостных характер, выполнены в полном объеме, структурированы, представлены различные точки зрения, продемонстрирован творческий подход.