

МИНОБРНАУКИ РОССИИ

федеральное государственное бюджетное образовательное учреждение высшего образования

«Санкт-Петербургский государственный экономический университет»



УТВЕРЖДАЮ

Проректор по образовательной деятельности

В.Г. Шубаева

20 23 г.

Большие данные

Рабочая программа дисциплины

Направление подготовки/ Специальность 09.03.02 Информационные системы и технологии

Направленность (профиль) программы/ Специализация Цифровизация экономической деятельности

Уровень высшего образования Бакалавриат

Форма обучения очная

Год набора 2023

Составитель(и):

к.т.н, Верзун Наталья Аркадьевна

Часов по учебному плану	216	Виды контроля в семестрах: Экзамен: семестр 7
в том числе:		
контактная работа	80	
самостоятельная работа	100	
практическая подготовка	0	
часов на контроль	36	

Распределение часов дисциплины:

Семестр:	7
Вид занятий	Часы
Лекционные занятия	38
Практические занятия	42
Лабораторные работы	
Итого аудиторных часов	80
Самостоятельная работа	100
Часы на контроль	36
Итого академических часов	216
Общая трудоемкость в зачетных единицах	6

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.....	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.....	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.....	11
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:	Ознакомление с основами и принципами работы с большими данными. Получение знаний о сущности и признаках больших данных, о существующих и перспективных методах хранения, обработки, анализа разнородных данных большого объема; формирование умений выбирать и применять математические методы и инструментальные средства для обработки и анализа данных большого объема при решении бизнес-задач в различных сферах деятельности.
--------------	---

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Дисциплина Б1.В Большие данные 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-8 - Способен работать с большими данными при цифровизации бизнес-процессов на предприятии	ПК-8.2 - Умеет применять методы машинного обучения при работе с большими данными	<p>Знать: особенности и принципы работы с данными большого объема; роль и место больших данных в решении аналитических и исследовательских задач в различных предметных областях при цифровизации бизнес-процессов на предприятии; существующие и перспективные математические методы обработки, анализа и визуализации данных большого объема.</p> <p>Уметь: применять математические методы обработки и анализа данных большого объема для решения бизнес-задач различных предметных областей; использовать методы машинного обучения при работе с большими данными; выбирать и применять инструментальные средства для анализа, исследований и моделирования процессов и явлений в профессиональной деятельности..</p> <p>Владеть: методами и инструментальными средствами обработки, анализа и визуализации данных и навыками их использования в профессиональной деятельности..</p>

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ*

Номер и наименование тем и/или	Содержание дисциплины	Объем дисциплины (академические часы)
--------------------------------	-----------------------	---------------------------------------

разделов/тем		Контактная работа			СРО
		ЗЛТ	ПЗ	ЛР	
Раздел I. Большие данные как базовая технология цифровой экономики.					
Тема 1. Введение. Основные понятия.	Обзор технологий третьей платформы информатизации. Понятие Big Data. Роль и значение технологий Big Data в становлении цифровой экономики. Характеристики, особенности больших данных (объем, скорость обработки, разнородность, не структурированность). Задачи, возникающие при работе с большими данными. Источники больших данных. Применение Big Data в различных сферах: государственное управление, ритейл, ЖКХ, маркетинг, финансы, социальные сети и пр. Примеры.	2	2		12
Тема 2. Основные принципы работы с Big Data.	ИТ-инфраструктура для задач класса Big Data. Организация и управление Big Data. Аналитическая обработка Big Data и выявление закономерностей. Средства поддержки принятия решения. Проблемы технологий Big Data.	6	6		14
Раздел II. Задачи управления большими данными.					
Тема 3. Хранение и управление Big Data.	Платформа Apache Hadoop. Компоненты Hadoop. Принципы построения Hadoop систем. Распределённая файловая система HDFS. NoSQL базы данных: особенности, характеристики, преимущества использования. Типы хранилищ данных. Платформа Oracle для больших данных.	6	6		12
Тема 4. Неструктурированная информация.	Автоматическая обработка текстов. Лингвистическая обработка текстов. Основные задачи, подходы к их решению. Компьютерная лингвистика. Технологии анализа и поиска текстовой информации RCO (Russian Context Optimizer). Архитектура RCO. Возможности: поиск, анализ, синтез, нечёткий поиск, семантическая сеть, рубрицирование. ABBYY Compreno – семантический разбор текста и визуализация. Алгоритм извлечения информации в ABBYY Compreno. RDF (Resource Description Framework) – граф. Sentiment analysis – анализ тональности – автоматическое извлечение субъективных мнений из текста или речи.	8	8		22
Тема 5. Методы и техники анализа Big Data.	Методы и техники анализа: дескриптивные и предиктивные методы, предобработка данных; методы класса Data Mining; ассоциативные правила; секвенциальный анализ (анализ последовательностей), кластерный анализ; корреляционный и регрессионный анализ; краудсорсинг, смешение и интеграция данных; машинное обучение (с учителем и без учителя); искусственные нейронные сети; сетевой анализ; оптимизация; генетические алгоритмы; распознавание образов; прогнозная аналитика; имитационное моделирование; пространственный анализ; статистический анализ; визуализация аналитических данных.	10	14		28

Тема 6. Технологии и инструментальные средства.	Система MapReduce (модель распределённых вычислений): алгоритмы, графы. Statistica Big Data Analytics инструмент статистического и графического анализа, прогнозирования, data mining и др. Возможности. Примеры применения. Аналитическая платформа Deductor. Состав. Назначение. Возможности. Примеры применения. Язык программирования R для статистической обработки данных и работы с графикой. Возможности R для статистической обработки данных. Методы обработки информационных блоков.	6	6		12
Контроль:					36
Всего по дисциплине:		38	42	0	100

*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
Парфенов Ю.П. Постреляционные хранилища данных: учебное пособие для вузов / Ю. П. Парфенов; под науч. ред. Н. В. Папуловской. — М.: Издательство Юрайт, 2019 — 121 с.	https://www.urait.ru/bcode/438577
Гасанов Э.Э. Интеллектуальные системы. Теория хранения и поиска информации : учебник для бакалавриата и магистратуры / Э. Э. Гасанов, В. Б. Кудрявцев. — 2-е изд., испр. и доп. — Электрон. дан. — Москва : Издательство Юрайт, 2019. — 271 с.	https://www.urait.ru/bcode/437023
Миркин, Б. Г. Введение в анализ данных: учебник и практикум / Б. Г. Миркин. — Электрон. дан. — М.: Издательство Юрайт, 2019. — 174 с.	https://www.urait.ru/bcode/432851
Верзун Н.А. Введение в инфокоммуникационные технологии и сети Future Networks : учебное пособие / Н.А.Верзун, М.О.Колбанев, А.В.Омельян. — Санкт-Петербург : Изд-во СПбГЭУ, 2016. — 51 с.	http://opac.unecon.ru/elibrary... BE%D0%BD%D0%BD%D1%8B%D0%B5.pdf
Верзун Н.А. Сетевая архитектура цифровой экономики : [монография] / Н.А.Верзун, М.О.Колбанев, А.В.Омельян. — Санкт-Петербург : Изд-во СПбГЭУ, 2018. — 156 с.	http://opac.unecon.ru/elibrary... 80%D0%BE%D0%B2%D0%BE%D0%B9.pdf

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

- 7-Zip
- R
- ОС Альт образование 10
- LibreOffice

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.unicon.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. 2010 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 60 посадочных мест, рабочее место преподавателя, стол м/м - 1 шт., доска меловая - 2 шт., кафедра - 1 шт., стул - 2 шт., Компьютер Intel i3-2100 2.4 Ghz /4Gb/500Gb/Acer V193 19" - 1 шт., Мультимедийный проектор Optoma x 400 - 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. 2020 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 25 посадочных мест, рабочее место преподавателя, доска маркерная на колесиках 2 шт., стол 1шт., стул 6шт., жалюзи 2шт., вешалка стойка 2шт. Компьютер Intel I5-7400/16Gb/1Tb/ видеокарта NVIDIA GeForce GT 710/Монитор DELL S2218H - 25 шт., Шкаф телекоммуникационный настенный ЦМО ШРН-Э-6.650 - 1 шт., Коммутатор ProCurve Switch 2626 - 1 шт., Мультимедийный проектор Optoma x 400 - 1 шт., Экран подпружинен. ручной MW Cinerollo 200x200см (S/N) - 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»

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 Обзор технологий третьей платформы информатизации.
- 2 Понятие Big Data. Роль и значение технологий Big Data в становлении цифровой экономики.
- 3 Характеристики, особенности больших данных (объем, скорость обработки, разнородность, не структурированность).
- 4 Задачи, возникающие при работе с большими данными.
- 5 Источники больших данных.
- 6 Применение Big Data в различных сферах: государственное управление, ритейл, ЖКХ, маркетинг, финансы, социальные сети и пр. Примеры.
- 7 ИТ-инфраструктура для задач класса Big Data.
- 8 Организация и управление Big Data.
- 9 Аналитическая обработка Big Data и выявление закономерностей.
- 10 Средства поддержки принятия решения Big Data.
- 11 Проблемы технологий Big Data.
- 12 Платформа Apache Hadoop. Компоненты Hadoop. Принципы построения Hadoop систем.
- 13 Распределённая файловая система HDFS.
- 14 NoSQL базы данных: особенности, характеристики, преимущества использования.
- 15 Типы хранилищ данных. Платформа Oracle для больших данных.
- 16 Автоматическая обработка текстов.
- 17 Лингвистическая обработка текстов. Основные задачи, подходы к их решению.
- 18 Компьютерная лингвистика.
- 19 Технологии анализа и поиска текстовой информации RCO (Russian Context Optimizer).
- 20 Архитектура RCO. Возможности: поиск, анализ, синтез, нечёткий поиск, семантическая сеть, рубрицирование.
- 21 ABBYY Compreno – семантический разбор текста и визуализация.
- 22 Алгоритм извлечения информации в ABBYY Compreno.
- 23 RDF (Resource Description Framework) – граф.
- 24 Sentiment analysis – анализ тональности – автоматическое извлечение субъективных мнений из текста или речи.
- 25 Методы и техники анализа: дескриптивные и предиктивные методы,
- 26 Методы и техники анализа: предобработка данных;
- 27 Методы и техники анализа: методы класса Data Mining;
- 28 Методы и техники анализа: ассоциативные правила;
- 29 Методы и техники анализа: секвенциальный анализ (анализ последовательностей),
- 30 Методы и техники анализа: кластерный анализ;
- 31 Методы и техники анализа: корреляционный и регрессионный анализ;
- 32 Методы и техники анализа: смешение и интеграция данных;
- 33 Машинное обучение (с учителем и без учителя);
- 34 Методы и техники анализа: искусственные нейронные сети;
- 35 Методы и техники анализа: сетевой анализ;
- 36 Методы и техники анализа: оптимизация;
- 37 Методы и техники анализа: генетические алгоритмы;
- 38 Методы и техники анализа: распознавание образов;

- 39 Методы и техники анализа: имитационное моделирование;
- 40 Методы и техники анализа: пространственный анализ;
- 41 Методы и техники анализа: статистический анализ;
- 42 Методы и техники анализа: визуализация аналитических данных.

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	Контрольное тестирование	с помощью технических средств и информационных систем	1-6
2	Информационно-аналитическая работа	письменно	5,6
3	Текущий контроль	с помощью технических средств и информационных систем	1-6

1.4 Другие объекты оценивания

Рабочей программой дисциплины не предусмотрено.

1.5 Самостоятельная работа обучающегося

Наименования самостоятельной работы	Номера тем
Выполнение расчетных, аналитических, расчетно-графических и др. заданий	4,5
Подготовка к лекционным и практическим занятиям	2-6
Подготовка к экзамену	1-6

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.