

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



УТВЕРЖДАЮ
Проректор
по учебной и методической работе

 В.Г.Шубаева

« 15 » 06 20 19 г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

БОЛЬШИЕ ДАННЫЕ

| | |
|---------------------------------------|--|
| Направление подготовки | 09.03.02 Информационные системы и технологии |
| Направленность (профиль) программы | Цифровизация экономической деятельности |
| Уровень высшего образования | бакалавриат |
| Форма обучения | очная |

Составитель:



к.т.н., доцент Верзун Н.А.

Санкт-Петербург
2019

СОДЕРЖАНИЕ

| | | |
|------|--|----|
| 1. | ЦЕЛЬ ДИСЦИПЛИНЫ..... | 3 |
| 2. | МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ | 3 |
| 3. | ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ..... | 3 |
| 4. | ОБЪЕМ И СТРУКТУРА ДИСЦИПЛИНЫ..... | 4 |
| 5. | СОДЕРЖАНИЕ РАЗДЕЛОВ И ТЕМ ДИСЦИПЛИНЫ | 4 |
| 6. | ЗАНЯТИЯ СЕМИНАРСКОГО ТИПА | 6 |
| 7. | МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ..... | 6 |
| 7.1. | Методические указания для обучающегося по освоению дисциплины | 6 |
| 7.2. | Организация самостоятельной работы..... | 7 |
| 8. | ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ | 8 |
| 9. | РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ..... | 8 |
| 9.1. | Учебно-методическое и информационное обеспечение дисциплины | 8 |
| 9.2. | Материально-техническое обеспечение учебного процесса..... | 9 |
| 10. | ОСОБЕННОСТИ ОСВОЕНИЯ ДИСЦИПЛИНЫ ДЛЯ ИНВАЛИДОВ И ЛИЦ С ОГРАНИЧЕННЫМИ ВОЗМОЖНОСТЯМИ ЗДОРОВЬЯ | 11 |
| 11. | ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ..... | 11 |

1. ЦЕЛЬ ДИСЦИПЛИНЫ

Цель дисциплины «Большие данные» ознакомление с основами и принципами работы с большими данными. Получение знаний о сущности и признаках больших данных, о существующих и перспективных математических 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.Ф «Большие данные» в структуре образовательной программы отнесена к части, формируемой участниками образовательных отношений Блока 1.

3. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ

Планируемые результаты обучения по дисциплине, соотнесенные с установленными в образовательной программе индикаторами достижения компетенций, представлены в таблице 3.1.

Таблица 3.1 – Планируемые результаты обучения по дисциплине, соотнесенные с установленными в образовательной программе индикаторами достижения компетенций

| Код и наименование компетенции выпускника | Код и наименование индикаторов достижения компетенций | Планируемые результаты обучения по дисциплине |
|---|--|---|
| 1 | 2 | 3 |
| ПК-2. Способен к анализу, синтезу и моделированию процессов функционирования информационных систем и систем в проектной области | ПК-2.1. Обладает аналитическими навыками работы с большими данными | Знать: понятие, сущность и признаки больших данных. Особенности и принципы работы с данными большого объема. Роль и место больших данных в решении аналитических и исследовательских задач в различных предметных областях. Существующие и перспективные математические методы обработки и анализа данных большого объема. Уметь: применять математические методы обработки и анализа данных большого объема для решения бизнес-задач различных предметных областей. Выбирать и применять инструментальные средства для анализа, исследований и моделирования процессов и явлений в профессиональной деятельности. |

4. ОБЪЕМ И СТРУКТУРА ДИСЦИПЛИНЫ

Трудоемкость дисциплины составляет 3 зачетных единицы, 108 часов.

Форма промежуточной аттестации: зачет – 6 семестр.

Распределение фонда времени по темам дисциплины представлено в таблице 4.1.

Таблица 4.1 – Распределение фонда времени по темам дисциплины

| Номер и наименование тем и разделов | Объем дисциплины (ак. часы) | | | |
|--|--------------------------------|-----------|----------|-----------|
| | Контактная работа | | | СРО |
| | ЗЛТ | ПЗ | ЛР | |
| <i>I</i> | 2 | 3 | 4 | 5 |
| <i>РАЗДЕЛ 1. Большие данные как базовая технология цифровой экономики</i> | 8 | 12 | - | 14 |
| Тема 1.1. Введение. Основные понятия | 4 | 4 | - | 6 |
| Тема 1.2. Основные принципы работы с Big Data | 4 | 8 | - | 8 |
| <i>РАЗДЕЛ 2. Задачи управления большими данными</i> | 14 | 30 | - | 30 |
| Тема 2.1. Хранение и управление Big Data | 2 | 4 | - | 4 |
| Тема 2.2. Неструктурированная информация | 2 | 6 | - | 6 |
| Тема 2.3. Методы и техники анализа Big Data | 6 | 10 | - | 10 |
| Тема 2.4. Технологии и инструментальные средства | 4 | 10 | - | 10 |
| Всего по дисциплине: | 22 | 42 | - | 44 |

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

5. СОДЕРЖАНИЕ РАЗДЕЛОВ И ТЕМ ДИСЦИПЛИНЫ

РАЗДЕЛ 1. Большие данные как базовая технология цифровой экономики

Тема 1.1. 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, неструктурированность). 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.

Тема 1.2. 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.

РАЗДЕЛ 2. Задачи управления большими данными

Тема 2.1. Хранение и управление Big Data

Платформа Apache Hadoop. Компоненты Hadoop. Принципы построения Hadoop систем. Распределенная файловая система HDFS.

NoSQL базы данных: особенности, характеристики, преимущества использования. Типы хранилищ данных. Платформа Oracle для больших данных.

Тема 2.2. 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, нечеткий 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.

Тема 2.3. 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.

Тема 2.4. Технологии/инструментальные средства

Система MapReduce (модель распределенных 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. ЗАНЯТИЯ СЕМИНАРСКОГО ТИПА

Таблица 6.1 – Практические занятия/Семинарские занятия

| № темы | Тема занятия | Вид занятия / Оценочное средство |
|--------|--|----------------------------------|
| 1 | 2 | 3 |
| 1.1 | Примеры решений Big Data в различных предметных областях | СЗ: доклад/дискуссия |
| 1.2 | Проблемы и перспективы использования технологий Big Data в экономике | СЗ: доклад/дискуссия |
| 2.1 | Альтернативы реляционным СУБД | СЗ: доклады |
| 2.2 | Анализ неструктурированной текстовой информации | ПЗ: Решение практических задач |
| 2.3 | Ознакомление с возможностями анализа данных в Deductor | ПЗ: Решение практических задач |
| 2.4 | Возможности R для статистической обработки и визуализации данных. | ПЗ: Решение практических задач |
| | Аналитические платформы | СЗ: доклад |

* ПЗ – практические занятия, СЗ – семинарские занятия, ЛР – лабораторные работы

7. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ

7.1. Методические указания для обучающегося по освоению дисциплины

Для формирования четкого представления об объеме и характере знаний и умений, которыми надо будет овладеть по дисциплине в самом начале учебного курса, обучающийся должен ознакомиться с учебно-методической документацией:

- рабочей программой дисциплины: с целями и задачами дисциплины, ее связями с другими дисциплинами образовательной программы, перечнем знаний и умений, которыми в процессе освоения дисциплины должен владеть обучающийся;
- порядком проведения текущего контроля успеваемости и промежуточной аттестации;
- графиком консультаций преподавателей кафедры.

Систематическое выполнение учебной работы на занятиях лекционных и семинарских типов, а также выполнение самостоятельной работы позволит успешно освоить дисциплину.

В процессе освоения дисциплины обучающимся следует:

- слушать, конспектировать излагаемый преподавателем материал;
- ставить, обсуждать актуальные проблемы курса, быть активным на занятиях;

– задавать преподавателю уточняющие вопросы с целью уяснения теоретических положений;

– выполнять задания практических занятий полностью и в установленные сроки.

При затруднениях в восприятии материала следует обратиться к основным литературным источникам. Если разобраться в материале не удалось, то обратиться к лектору (по графику его консультаций) или к преподавателю на занятиях семинарского типа.

Обучающимся, пропустившим занятия (независимо от причин), не имеющим письменного решения задач или не подготовившимся к данному занятию, рекомендуется не позже, чем в 2 - недельный срок явиться на консультацию к преподавателю и отчитаться по теме.

7.2. Организация самостоятельной работы

Под самостоятельной работой обучающихся понимается планируемая работа обучающихся, направленная на формирование указанных компетенций, выполняемая во внеаудиторное время по заданию и при методическом руководстве преподавателя, без его непосредственного участия.

Методическое обеспечение самостоятельной работы при наличии обучающихся лиц с ограниченными возможностями представляется в формах, адаптированных к ограничениям их здоровья.

Методическое обеспечение самостоятельной работы, в т. ч. для обучающихся с использованием ДОТ и лиц с ограниченными возможностями здоровья представлено:

Электронный учебно-методический комплекс дисциплины Большие данные размещен в СДО «Moodle» на сайте СПбГЭУ de.unicon.ru/

Виды самостоятельной работы по дисциплине представлены в таблице 7.2.1.

Таблица 7.2.1 – Организация самостоятельной работы обучающегося

| № темы | Вид самостоятельной работы |
|---------------|--|
| <i>1</i> | <i>2</i> |
| 2.1; 2.4 | Написание реферата и подготовка к презентации по заданной теме |
| 1.2; 1.3 | Подготовка к занятиям семинарского типа |
| 2.1; 2.3; 2.4 | Выполнение домашних заданий по практическим занятиям |
| все | Подготовка к итоговому тестированию |

Каждый вид СРО, указанный в таблице 7.2.1 обеспечен методическими материалами.

8. ОБРАЗОВАТЕЛЬНЫЕ ТЕХНОЛОГИИ

В рамках реализации дисциплины «Большие данные» используются разнообразные образовательные технологии как традиционные, так и с применением активных и интерактивных методов обучения.

Активные и интерактивные методы обучения:

- лекция-дискуссия (тема №1.1);
- проблемная лекция (тема №2.1);
- индивидуальные занятия на ПК (тема №2.4);

Лекция-дискуссия проводится: по материалам лекций; по итогам практических занятий; по проблемам, предложенным самими студентами, или преподавателем, если студенты затрудняются; по событиям и фактам из практики изучаемой сферы деятельности.

Проблемная лекция – метод, при котором важнейшим показателем «проблемности» характера обучения является наличие познавательной проблемы. «Проблема» может быть сформулирована на основе материалов истории науки, социальной практики, в контексте предстоящей профессиональной деятельности. Лекция характеризуется проблемным изложением материала: преподаватель ставит вопрос или формулирует проблемную задачу и показывает варианты ответов или способов решения, а студенты наблюдают за поиском и определяют свое отношение к полученному материалу.

Индивидуальные занятия на ПК – метод предполагает выполнение индивидуальных заданий по изучению пакетов прикладных программ, языков программирования, информационно-коммуникационных технологий.

9. РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

9.1. Учебно-методическое и информационное обеспечение дисциплины

Таблица 9.1.1 – Учебно-методическое обеспечение дисциплины

| Библиографическое описание издания (автор, заглавие, вид, место и год издания, кол. стр.) | основная/ дополнительная литература | Книгообеспеченность | |
|--|---|--|---------------------------|
| | | Кол-во. экз. в библ. СПбГЭ У | Электронные ресурсы |
| 1. Парфенов, Ю. П. Постреляционные хранилища данных : учебное пособие для вузов / Ю. П. Парфенов. — Электрон. дан. — Москва : Издательство Юрайт, 2019 ; Екатеринбург : Изд-во Урал. ун-та. — 121 с. | основная | — | ЭБС Юрайт |
| 2. 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 : | основная | — | ЭБС Юрайт |

| Библиографическое описание издания (автор, заглавие, вид, место и год издания, кол. стр.) | основная/ дополнитель ная литература | Книгообеспеченность | |
|--|---|--|---|
| | | Кол-во. экз. в библ. СПбГЭ У | Электронные ресурсы |
| Издательство Юрайт, 2019. — 271 с. | | | |
| 4. Миркин, Б. Г. Введение в анализ данных: учебник и практикум / Б. Г. Миркин. — Электрон. дан. — М.: Издательство Юрайт, 2019. — 174 с. | дополнительная | — | ЭБС Юрайт |
| 5. 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. — Сведения доступны также по Интернету: opac.unesco.ru . | дополнительная | 25 | ЭБ ОРАС.UNESCO N.RU |
| 6. 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. — Сведения доступны также по Интернету: opac.unesco.ru . | дополнительная | 5 | ЭБ ОРАС.UNESCO N.RU |

Таблица 9.1.2 – Перечень современных профессиональных баз данных (СПБД)

| № | Наименование СПБД |
|---|---|
| 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 |

Таблица 9.1.3 – Перечень информационных справочных систем (ИСС)

| № | Наименование ИСС |
|---|--|
| 1 | Справочная правовая система КонсультантПлюс (инсталлированный ресурс СПбГЭУ или www.consultant.ru) |
| 2 | Справочная правовая система «ГАРАНТ» (инсталлированный ресурс СПбГЭУ или www.garant.ru) |
| 3 | Информационно-справочная система «Кодекс» (инсталлированный ресурс СПбГЭУ или www.kodeks.ru) |
| 4 | Электронная библиотечная система BOOK.ru - www.book.ru |
| 5 | Электронная библиотечная система ЭБС ЮРАЙТ – www.urait.ru |
| 6 | Электронно-библиотечная система ЗНАНИУМ (ZNANIUM) – www.znanium.com |
| 7 | Электронная библиотека СПбГЭУ– opac.unesco.ru |

9.2. Материально-техническое обеспечение учебного процесса

Для реализации данной дисциплины имеются специальные помещения для проведения занятий лекционного типа, занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации.

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.

Таблица 9.2.1 – Перечень лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства (ПО)

| № | Наименование ПО |
|---|--|
| 1 | Microsoft Windows Professional (КОНТРАКТ № 244/20 «26» июня 2020 г.) |
| 2 | Microsoft Office Professional (КОНТРАКТ № 244/20 «26» июня 2020 г.) |
| 3 | 7-Zip (freeware) |

Таблица 9.2.2 – Перечень учебных аудиторий для проведения учебных занятий, оснащенных оборудованием и техническими средствами обучения

| Наименование учебных аудиторий, перечень оборудования и технических средств обучения | Адрес (местоположение) учебных аудиторий |
|--|--|
| Ауд. 2065 Учебная аудитория (для проведения занятий лекционного типа и занятий семинарского типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации). Специализированная мебель и оборудование: Учебная мебель на 80 посадочных места (столов 40шт., стульев 80шт.), рабочее место преподавателя, доска маркерная 1шт, кафедра 1шт., стул 2шт. Персональный компьютер в сборе Ledovo тип 1 (Core I3 2100+монитор Acer V193) - 1 шт., Интерактивный проектор Epson EB-485Wi - 1шт. 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» |

10. 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.

11. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ ДЛЯ ПРОВЕДЕНИЯ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ОБУЧАЮЩИХСЯ ПО ДИСЦИПЛИНЕ

Фонд оценочных средств для проведения аттестации уровня сформированности компетенций обучающихся по дисциплине оформляется отдельным документом и является приложением к рабочей программе дисциплины (модуля).