МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

Федеральное государственное бюджетное образовательное учреждение высшего образования

«ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ
УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ»
(ТУСУР)

УТВЕРЖДАЮ
Проректор по учебной работе
_____ П. В. Сенченко
«___» _____ 20__ г.

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Анализ больших данных

Уровень образования: **высшее образование - магистратура**
Направление подготовки / специальность: **38.04.05 Бизнес-информатика**
Направленность (профиль) / специализация: **Предпринимательство и организация бизнеса в сфере информационных технологий**
Форма обучения: **очная**
Факультет: **ФСУ, Факультет систем управления**
Кафедра: **АОИ, Кафедра автоматизации обработки информации**
Курс: **1**
Семестр: **2**
Учебный план набора 2020 года

Распределение рабочего времени

№	Виды учебной деятельности	2 семестр	Всего	Единицы
1	Лекции	18	18	часов
2	Лабораторные работы	36	36	часов
3	Всего аудиторных занятий	54	54	часов
4	Самостоятельная работа	54	54	часов
5	Всего (без экзамена)	108	108	часов
6	Общая трудоемкость	108	108	часов
		3.0	3.0	З.Е.

Зачёт с оценкой: 2 семестр

Томск

Согласована на портале № 52309

Рассмотрена и одобрена на заседании кафедры
протокол № 332 от «10» 12 2019 г.

ЛИСТ СОГЛАСОВАНИЯ

Рабочая программа дисциплины составлена с учетом требований федерального государственного образовательного стандарта высшего образования (ФГОС ВО) по направлению подготовки (специальности) 38.04.05 Бизнес-информатика, утвержденного 08.04.2015 года, рассмотрена и одобрена на заседании кафедры АОИ «___» _____ 20__ года, протокол №_____.

Разработчики:

к.т.н., доцент каф. АОИ	_____	Н. В. Пермякова
к.т.н., доцент каф. АОИ	_____	Е. Ю. Агеев

Заведующий обеспечивающей каф. АОИ	_____	А. А. Сидоров

Рабочая программа дисциплины согласована с факультетом и выпускающей кафедрой:

Декан ФСУ	_____	Н. Ю. Салмина
Заведующий выпускающей каф. АОИ	_____	А. А. Сидоров

Эксперты:

Заведующий кафедрой автоматизации обработки информации (АОИ)	_____	А. А. Сидоров
Доцент кафедры автоматизации обработки информации (АОИ)	_____	Н. Ю. Салмина

Согласована на портале № 52309

1. Цели и задачи дисциплины

1.1. Цели дисциплины

Формирование представления о сборе, обработке и анализе данных средствами языка Python, в интерактивной среде Jupyter Notebook.

Знакомство с методами и средствами работы с Большими данными, нереляционными базами данных, экосистемой Hadoop.

1.2. Задачи дисциплины

– Развитие навыков подготовки, очистки и обработки данных, представления результатов
– анализа данных в различной форме, с учетом уровня подготовленности целевой аудитории и решаемых задач.
– Формирование четкого представления современных методов инжиниринга Больших
– данных, средств и способов построения высокоэффективных горизонтально масштабируемых хранилищ сверхбольших объемов данных.
– Знакомство с программными продуктами управления и обработки Больших данных.

2. Место дисциплины в структуре ОПОП

Дисциплина «Анализ больших данных» (Б1.В.ОД.4) относится к блоку 1 (вариативная часть).

Последующими дисциплинами являются: Защита выпускной квалификационной работы, включая подготовку к защите и процедуру защиты, Практика по получению первичных профессиональных умений и навыков (рассред.), Практика по получению профессиональных умений и опыта профессиональной деятельности (технологическая практика), Преддипломная практика.

3. Требования к результатам освоения дисциплины

Процесс изучения дисциплины направлен на формирование следующих компетенций:

– ПК-1 способностью готовить аналитические материалы для оценки мероприятий и выработки стратегических решений в области ИКТ ;

В результате изучения дисциплины обучающийся должен:

– **знать** математические методы подготовки и обработки данных. Методы и средства системного анализа, оптимизации, управления, принятия решений и обработки информации.
– **уметь** разрабатывать алгоритмы решения задач; работать с программными средствами общего назначения; разрабатывать программное обеспечение средствами объектно-ориентированного языка Python и его интерактивной реализации Jupyter Notebook и использовать его на практике
– **владеть** современными методами исследования и информационно-коммуникационными технологиями. Новыми методами исследования и обработки данных и их применению в самостоятельной научно-исследовательской деятельности в области профессиональной деятельности.

4. Объем дисциплины и виды учебной работы

Общая трудоемкость дисциплины составляет 3.0 зачетных единицы и представлена в таблице 4.1.

Таблица 4.1 – Трудоемкость дисциплины

Виды учебной деятельности	Всего часов	Семестры
		2 семестр
Аудиторные занятия (всего)	54	54
Лекции	18	18
Лабораторные работы	36	36
Самостоятельная работа (всего)	54	54
Подготовка к контрольным работам	2	2
Оформление отчетов по лабораторным работам	36	36

Согласована на портале № 52309

Проработка лекционного материала	8	8
Выполнение контрольных работ	8	8
Всего (без экзамена)	108	108
Общая трудоемкость, ч	108	108
Зачетные Единицы	3.0	3.0

5. Содержание дисциплины

5.1. Разделы дисциплины и виды занятий

Разделы дисциплины и виды занятий приведены в таблице 5.1.

Таблица 5.1 – Разделы дисциплины и виды занятий

Названия разделов дисциплины	Лек., ч	Лаб. раб., ч	Сам. раб., ч	Всего часов (без экзамена)	Формируемые компетенции
2 семестр					
1 Понятие "Большие данные" история появления и характерные свойства данных, называемых Большими.	2	4	7	13	ПК-1
2 Подходы к анализу данных. Стандартные методики анализа на примере CRISP-DM. Роль данных и результатов их анализа при принятии решений.	2	4	8	14	ПК-1
3 Типы данных, статистические методы обработки, виды статистик. Традиционный подход анализа данных и новых подход анализа Больших данных.	2	12	17	31	ПК-1
4 Методы машинного обучения при построении прогнозов на основе анализа данных. Регрессия и классификация, способы оценки результата.	4	12	14	30	ПК-1
5 Подходы к формированию доказательных гипотез на основе анализа данных. Способы и средства визуализации данных.	2	0	1	3	ПК-1
6 Инжиниринг больших данных. Экосистема Hadoop, конвейер обработки Больших данных, NoSQL базы данных. Lambda архитектура.	6	4	7	17	ПК-1
Итого за семестр	18	36	54	108	
Итого	18	36	54	108	

5.2. Содержание разделов дисциплины (по лекциям)

Содержание разделов дисциплин (по лекциям) приведено в таблице 5.2.

Таблица 5.2 – Содержание разделов дисциплин (по лекциям)

Названия разделов	Содержание разделов дисциплины (по лекциям)	Трудоемкость, ч	Формируемые компетенции
2 семестр			
1 Понятие "Большие	Возрастающая роль данных в нашей жиз-	2	ПК-1

Согласована на портале № 52309

данные" история появления и характерные свойства данных, называемых Большими.	ни. "Подключенный" мир, Интернет-коммуникации и бизнес, социальные сети,Интернет вещей. Экспоненциальный ростобъема создаваемых и обрабатываемыхданных, изменяющий условия жизни иработы. Когда данные становятся большими, характерные свойства Больших данных, почему потребовался этот термин.Открытые данные и персональные данные. Структурированные и неструктурированные данные. Данные в покое и данные в движении. Базовые технологии хранения данных		
	Итого	2	
2 Подходы к анализу данных. Стандартные методики анализа на примере CRISP-DM. Роль данных и результатов их анализа при принятии решений.	Стандартные методики анализа данных.Методика Cross Industry Standard Process for Data Mining (CRISP-DM). Роль языка Python в сфере анализа данных. Типы анализа данных: описательный, прогностический и рекомендательный анализ. Фактор времени в анализе данных. Принципиальные различия в подходах анализа обычных данных и Больших данных. Жизненный цикл анализа данных.	2	ПК-1
	Итого	2	
3 Типы данных, статистические методы обработки, виды статистик. Традиционный подход анализа данных и новых подход анализа Больших данных.	Исследовательский анализ данных. Данные от датчиков. Наблюдения, реальныезначения и измеренная величина. Типыпеременных. Статистические методы.Описательная и индуктивная статистика.Популяция и репрезентативная выборка.Статистические методы при обработкеБольших данных. Стандартные распределения, дисперсия и центральность. Применение статистических методов при описательном анализе данных. Особенностиприменения библиотеки Pandas. Корреляционный анализ данных.	2	ПК-1
	Итого	2	
4 Методы машинного обучения при построении прогнозов на основе анализа данных. Регрессия и классификация, способы оценки результата.	Что такое машинное обучение.Прогнозный анализ данных с помощью машинного обучения. Суть процесса машинного обучения и типичные приложения. Регрессионный анализ, линейная регрессия, области применения регрессионного анализа. Задачи классификации,алгоритмы решения классификационных задач, алгоритм решающего дерева, области применения.	4	ПК-1
	Итого	4	

Согласована на портале № 52309

5 Подходы к формированию доказательных гипотез на основе анализа данных. Способы и средства визуализации данных.	Каким образом данные и результаты иханализа могут применяться при формулировании гипотез, обосновании решений ипредложений. Дедуктивная и индуктивнаялогика, формулирование логических выводов. Опасность ошибок и заблуждений.Эффект визуального представления результатов анализа данных. Средства визуализации, библиотека Matplotlib, модульPyplot, возможности и свойства. Онлайнсервис Plotly, создание и экспорт графического представления, работа в офлайн режиме. Общепринятые способы отображения данных. Знакомство с библиотекой Folium для отображения картографических данных.	2	ПК-1
	Итого	2	
6 Инжиниринг больших данных. Экосистема Hadoop, конвейер обработки Больших данных, NoSQL базы данных. Lambda архитектура.	Структура современного дата-центра.Технологии виртуализации и облачныевычисления. Уровни абстракции, гипервизоры, контейнеры. Аналитика в облаке(Cloud analytics) и аналитика на периферии (Edge analytics). Виртуальный датацентр, IaaS, PaaS и SaaS сервисы. Виртуализованное хранилище данных, виртуализованные сетевые подключения. Средстваобеспечения масштабируемости, высокойдоступности и отказоустойчивости хранилищ Больших данных. Программное обеспечение Hadoop и файловая системаHDFS, обработка данных в Hadoop - MapReduce. Экосистема Hadoop. Конвейер обработки больших данных: загрузка данных, хранение и обработка. Проблемы загрузки Больших данных, программноеобеспечение Kafka. Проблемы храненияБольших данных, NoSQL база данныхCassandra. Сравнение Hadoop и Cassandra.Проблемы обработки Больших данных, программное обеспечение Spark. Обработка данных в оперативной памяти с Resilient Distributed Data sets (RDDs).Сравнение Spark и Map Reduce. Архитектура системы для хранения и обработки Больших данных - Lambda архитектура.	6	ПК-1
	Итого	6	
Итого за семестр		18	

5.3. Разделы дисциплины и междисциплинарные связи с обеспечивающими (предыдущими) и обеспечиваемыми (последующими) дисциплинами

Разделы дисциплины и междисциплинарные связи с обеспечивающими (предыдущими) и обеспечиваемыми (последующими) дисциплинами представлены в таблице 5.3.

Согласована на портале № 52309

Таблица 5.3 – Разделы дисциплины и междисциплинарные связи

Наименование дисциплин	№ разделов данной дисциплины, для которых необходимо изучение обеспечивающих и обеспечиваемых дисциплин					
	1	2	3	4	5	6
Последующие дисциплины						
1 Защита выпускной квалификационной работы, включая подготовку к защите и процедуру защиты	+	+	+	+	+	+
2 Практика по получению первичных профессиональных умений и навыков (рассред.)	+	+	+	+	+	+
3 Практика по получению профессиональных умений и опыта профессиональной деятельности (технологическая практика)	+	+	+	+	+	+
4 Преддипломная практика	+	+	+	+	+	+

5.4. Соответствие компетенций, формируемых при изучении дисциплины, и видов занятий

Соответствие компетенций, формируемых при изучении дисциплины, и видов занятий представлено в таблице 5.4.

Таблица 5.4 – Соответствие компетенций, формируемых при изучении дисциплины, и видов занятий

Компетенции	Виды занятий			Формы контроля
	Лек.	Лаб. раб.	Сам. раб.	
ПК-1	+	+	+	Контрольная работа, Отчет по лабораторной работе, Опрос на занятиях, Тест, Зачёт с оценкой

6. Интерактивные методы и формы организации обучения

Не предусмотрено РУП.

7. Лабораторные работы

Наименование лабораторных работ приведено в таблице 7.1.

Таблица 7.1 – Наименование лабораторных работ

Названия разделов	Наименование лабораторных работ	Трудоемкость, ч	Формируемые компетенции
2 семестр			
1 Понятие "Большие данные" история появления и характерные свойства данных, называемых Большими.	Анализ криминальной обстановки в Сан-Франциско	4	ПК-1
	Итого	4	
2 Подходы к анализу данных. Стандартные методики анализа на примере CRISP-DM. Роль данных и результатов их анализа при принятии	Измерение и анализ данных Интернет соединения	4	ПК-1
	Итого	4	

Согласована на портале № 52309

решений.			
3 Типы данных, статистические методы обработки, виды статистик. Традиционный подход анализа данных и новых подход анализа Больших данных.	Измерение и анализ Интернет-соединения с хранилищем данных в виде SQL базы данных	4	ПК-1
	Описательная статистика в Python	4	
	Корреляционный анализ в Python	4	
	Итого	12	
4 Методы машинного обучения при построении прогнозов на основе анализа данных. Регрессия и классификация, способы оценки результата.	Простая линейная регрессия в Python	4	ПК-1
	Классификация с помощью дерева решений	4	
	Обработка ошибок при использовании метода линейной регрессии	4	
	Итого	12	
6 Инжиниринг больших данных. Экосистема Hadoop, конвейер обработки Больших данных, NoSQL базы данных. Lambda архитектура.	Измерение и анализ Интернет-соединениявизуализация результатов	4	ПК-1
	Итого	4	
Итого за семестр		36	

8. Практические занятия (семинары)

Не предусмотрено РУП.

9. Самостоятельная работа

Виды самостоятельной работы, трудоемкость и формируемые компетенции представлены в таблице 9.1.

Таблица 9.1 – Виды самостоятельной работы, трудоемкость и формируемые компетенции

Названия разделов	Виды самостоятельной работы	Трудоемкость, ч	Формируемые компетенции	Формы контроля
2 семестр				
1 Понятие "Большие данные" история появления и характерные свойства данных, называемых Большими.	Проработка лекционного материала	1	ПК-1	Зачёт с оценкой, Контрольная работа, Отчет по лабораторной работе, Тест
	Оформление отчетов по лабораторным работам	4		
	Подготовка к контрольным работам	2		
	Итого	7		
2 Подходы к анализу данных. Стандартные методики анализа на примере CRISP-	Выполнение контрольных работ	2	ПК-1	Зачёт с оценкой, Контрольная работа, Отчет по лабораторной работе, Тест
	Проработка лекционного материала	2		
	Оформление отчетов по	4		

Согласована на портале № 52309

DM. Роль данных и результатов их анализа при принятии решений.	лабораторным работам			
	Итого	8		
3 Типы данных, статистические методы обработки, виды статистик. Традиционный подход анализа данных и новых подход анализа Больших данных.	Выполнение контрольных работ	4	ПК-1	Зачёт с оценкой, Контрольная работа, Отчет по лабораторной работе, Тест
	Проработка лекционного материала	1		
	Оформление отчетов по лабораторным работам	12		
	Итого	17		
4 Методы машинного обучения при построении прогнозов на основе анализа данных. Регрессия и классификация, способы оценки результата.	Выполнение контрольных работ	1	ПК-1	Зачёт с оценкой, Контрольная работа, Отчет по лабораторной работе, Тест
	Проработка лекционного материала	1		
	Оформление отчетов по лабораторным работам	12		
	Итого	14		
5 Подходы к формированию доказательных гипотез на основе анализа данных. Способы и средства визуализации данных.	Проработка лекционного материала	1	ПК-1	Зачёт с оценкой, Отчет по лабораторной работе, Тест
	Итого	1		
6 Инжиниринг больших данных. Экосистема Hadoop, конвейер обработки Больших данных, NoSQL базы данных. Lambda архитектура.	Выполнение контрольных работ	1	ПК-1	Зачёт с оценкой, Контрольная работа, Отчет по лабораторной работе, Тест
	Проработка лекционного материала	2		
	Оформление отчетов по лабораторным работам	4		
	Итого	7		
Итого за семестр		54		
Итого		54		

10. Курсовой проект / курсовая работа

Не предусмотрено РУП.

11. Рейтинговая система для оценки успеваемости обучающихся

11.1. Балльные оценки для элементов контроля

Таблица 11.1 – Балльные оценки для элементов контроля

Элементы учебной	Максимальный	Максимальный	Максимальный	Всего за

Согласована на портале № 52309

деятельности	балл на 1-ую КТ с начала семестра	балл за период между 1КТ и 2КТ	балл за период между 2КТ и на конец семестра	семестр
2 семестр				
Зачёт с оценкой			20	20
Контрольная работа	5	5	5	15
Отчет по лабораторной работе	10	20	20	50
Тест	5	5	5	15
Итого максимум за период	20	30	50	100
Нарастающим итогом	20	50	100	100

11.2. Пересчет баллов в оценки за контрольные точки

Пересчет баллов в оценки за контрольные точки представлен в таблице 11.2.

Таблица 11.2 – Пересчет баллов в оценки за контрольные точки

Баллы на дату контрольной точки	Оценка
≥ 90% от максимальной суммы баллов на дату КТ	5
От 70% до 89% от максимальной суммы баллов на дату КТ	4
От 60% до 69% от максимальной суммы баллов на дату КТ	3
< 60% от максимальной суммы баллов на дату КТ	2

11.3. Пересчет суммы баллов в традиционную и международную оценку

Пересчет суммы баллов в традиционную и международную оценку представлен в таблице 11.3.

Таблица 11.3 – Пересчет суммы баллов в традиционную и международную оценку

Оценка (ГОС)	Итоговая сумма баллов, учитывает успешно сданный экзамен	Оценка (ECTS)
5 (отлично) (зачтено)	90 - 100	A (отлично)
4 (хорошо) (зачтено)	85 - 89	B (очень хорошо)
	75 - 84	C (хорошо)
	70 - 74	D (удовлетворительно)
3 (удовлетворительно) (зачтено)	65 - 69	
	60 - 64	E (посредственно)
2 (неудовлетворительно) (не зачтено)	Ниже 60 баллов	F (неудовлетворительно)

12. Учебно-методическое и информационное обеспечение дисциплины

12.1. Основная литература

1. Юре, Л. Анализ больших наборов данных / Л. Юре, Р. Ананд, Д. У. Джеффри ; перевод с английского А. А. Слинкин. — Москва [Электронный ресурс]: ДМК Пресс, 2016. — 498 с. — ISBN 978-5-97060-190-7 — Режим доступа: https://reader.lanbook.com/book/93571#1 (дата обращения: 27.09.2021).

12.2. Дополнительная литература

1. Рындина, С. В. Цифровая трансформация бизнеса [Электронный ресурс] [Электронный ресурс]: использование аналитики на основе больших данных : учебное пособие / С. В. Рындина.

Согласована на портале № 52309

— Пенза : ПГУ, 2019. — 182 с. — ISBN 978-5-907262-04-1. — Режим доступа: https://reader.lanbook.com/book/162301#1 (дата обращения: 27.09.2021).

12.3. Учебно-методические пособия

12.3.1. Обязательные учебно-методические пособия

1. Анализ больших данных, гр. 400-М, 420-М1, Весенний семестр 2020/2021 у.г. (Журнал) https [Электронный ресурс]: //sdo.tusur.ru/mod/resource/view.php?id=163365 — Режим доступа: https://sdo.tusur.ru/mod/resource/view.php?id=163365 (дата обращения: 27.09.2021).

12.3.2. Учебно-методические пособия для лиц с ограниченными возможностями здоровья и инвалидов

Учебно-методические материалы для самостоятельной и аудиторной работы обучающихся из числа лиц с ограниченными возможностями здоровья и инвалидов предоставляются в формах, адаптированных к ограничениям их здоровья и восприятия информации.

Для лиц с нарушениями зрения:

- в форме электронного документа;
- в печатной форме увеличенным шрифтом.

Для лиц с нарушениями слуха:

- в форме электронного документа;
- в печатной форме.

Для лиц с нарушениями опорно-двигательного аппарата:

- в форме электронного документа;
- в печатной форме.

12.4. Профессиональные базы данных и информационные справочные системы

1. Дополнительно к профессиональным базам данных рекомендуется использовать информационные, справочные и нормативные базы данных https://lib.tusur.ru/ru/resursy/bazy-dannyh

13. Материально-техническое обеспечение дисциплины и требуемое программное обеспечение

13.1. Общие требования к материально-техническому и программному обеспечению дисциплины

13.1.1. Материально-техническое и программное обеспечение для лекционных занятий

Для проведения занятий лекционного типа, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации используется учебная аудитория с количеством посадочных мест не менее 22-24, оборудованная доской и стандартной учебной мебелью. Имеются демонстрационное оборудование и учебно-наглядные пособия, обеспечивающие тематические иллюстрации по лекционным разделам дисциплины.

13.1.2. Материально-техническое и программное обеспечение для лабораторных работ

Компьютерный класс № 2

учебная аудитория для проведения занятий практического типа, учебная аудитория для проведения занятий лабораторного типа, помещение для курсового проектирования (выполнения курсовых работ), помещение для проведения групповых и индивидуальных консультаций

634034, Томская область, г. Томск, Вершинина улица, д. 74, 430 ауд.

Описание имеющегося оборудования:

- Персональный компьютер Intel Core i5-9600K 6x3.7ГГц, ОЗУ – 16 Гб, жесткий диск SSD – 250 Гб – 25 шт.;
- Персональный компьютер Intel Core i5-9500 6x3.0ГГц, ОЗУ – 16 Гб, жесткий диск SSD – 500 Гб – 1 шт.;
- Проектор Epson EB-982W;
- Экран для проектора;
- Магнитно-маркерная доска;
- Комплект специализированной учебной мебели;
- Рабочее место преподавателя.

Согласована на портале № 52309

Программное обеспечение:
- Django, Модифицированная лицензия BSD
- GIMP 2.9.8, GNU GPLv3
- Git 2.11.03, GNU GPLv2
- Microsoft Office 2010 Standard
- PDF-XChange Editor, свободно распространяемое ПО
- Python, свободнораспространяемое ПО совместимое с GNU GPL

Учебная аудитория
учебная аудитория для проведения занятий лекционного типа, учебная аудитория для проведения занятий практического типа
634034, Томская область, г. Томск, Вершинина улица, д. 74, 418 ауд.
Описание имеющегося оборудования:
- Персональный компьютер Intel Core i5-9500 6x3.0ГГц, ОЗУ – 16 Гб, жесткий диск SSD – 500 Гб;
- Проектор Epson EB-982W;
- Экран для проектора;
- Магнитно-маркерная доска;
- Комплект специализированной учебной мебели;
- Рабочее место преподавателя.
Программное обеспечение:
- Epson iProjection
- Google Chrome, Open Source
- LibreOffice
- Microsoft PowerPoint Viewer
- PDF-XChange Editor, свободно распространяемое ПО
- Windows 10 Professional, AZURE DEV TOOLS FOR TEACHING

13.1.3. Материально-техническое и программное обеспечение для самостоятельной работы

Для самостоятельной работы используются учебные аудитории (компьютерные классы), расположенные по адресам:
- 634050, Томская область, г. Томск, Ленина проспект, д. 40, 233 ауд.;
- 634045, Томская область, г. Томск, ул. Красноармейская, д. 146, 201 ауд.;
- 634034, Томская область, г. Томск, Вершинина улица, д. 47, 126 ауд.;
- 634034, Томская область, г. Томск, Вершинина улица, д. 74, 207 ауд.

Состав оборудования:
- учебная мебель;
- компьютеры класса не ниже ПЭВМ INTEL Celeron D336 2.8ГГц. - 5 шт.;
- компьютеры подключены к сети «Интернет» и обеспечивают доступ в электронную информационно-образовательную среду университета.

Перечень программного обеспечения:
- Microsoft Windows;
- OpenOffice;
- Kaspersky Endpoint Security 10 для Windows;
- 7-Zip;
- Google Chrome.

13.2. Материально-техническое обеспечение дисциплины для лиц с ограниченными возможностями здоровья и инвалидов

Освоение дисциплины лицами с ограниченными возможностями здоровья и инвалидами осуществляется с использованием средств обучения общего и специального назначения.

Согласована на портале № 52309

При занятиях с обучающимися **с нарушениями слуха** предусмотрено использование звукоусиливающей аппаратуры, мультимедийных средств и других технических средств приема/передачи учебной информации в доступных формах, мобильной системы преподавания для обучающихся с инвалидностью, портативной индукционной системы. Учебная аудитория, в которой занимаются обучающиеся с нарушением слуха, оборудована компьютерной техникой, аудиотехникой, видеотехникой, электронной доской, мультимедийной системой.

При занятиях с обучающимися **с нарушениями зрениями** предусмотрено использование в лекционных и учебных аудиториях возможности просмотра удаленных объектов (например, текста на доске или слайда на экране) при помощи видеоувеличителей для комфортного просмотра.

При занятиях с обучающимися **с нарушениями опорно-двигательного аппарата** используются альтернативные устройства ввода информации и другие технические средства приема/передачи учебной информации в доступных формах, мобильной системы обучения для людей с инвалидностью.

14. Оценочные материалы и методические рекомендации по организации изучения дисциплины

14.1. Содержание оценочных материалов и методические рекомендации

Для оценки степени сформированности и уровня освоения закрепленных за дисциплиной компетенций используются оценочные материалы в составе:

14.1.1. Тестовые задания

Загрузите данные из предоставленного файла CSV в датафрейм Pandas

Выполните описательный статистический анализ данных, хранящихся в датафрейме

Постройте график, гистограмму, круговую диаграмму данных, хранящихся в датафрейме

Примените машинное обучение для классификации предоставленного набора данных

Примените машинное обучение для выявления линии тренда в наборе данных

14.1.2. Темы контрольных работ

Понятие Большие данные, история появления и характерные свойства данных, называемых Большими.

Стандартные методики анализа данных. Роль данных и результатов их анализа при принятии решений.

Статистические методы обработки данных, виды статистик. Особенности анализа Больших данных.

Методы машинного обучения при построении прогнозов на основе анализа данных.

Формирование доказательных гипотез на основе анализа данных.

Способы и средства визуализации данных с помощью Python

14.1.3. Темы опросов на занятиях

Возрастающая роль данных в нашей жизни. "Подключенный" мир, Интернет-коммуникации и бизнес, социальные сети,

Интернет вещей. Экспоненциальный рост

объема создаваемых и обрабатываемых

данных, изменяющий условия жизни и

работы. Когда данные становятся большими, характерные свойства Больших данных, почему потребовался этот термин.

Открытые данные и персональные данные. Структурированные и неструктурированные данные. Данные в покое и данные в движении. Базовые технологии хранения данных

Стандартные методики анализа данных.

Методика Cross Industry Standard Process for Data Mining (CRISP-DM). Роль языка Python в сфере анализа данных. Типы анализа данных: описательный, прогностический и рекомендательный анализ. Фактор времени в анализе данных. Принципиальные различия в подходах анализа обычных данных и Больших данных. Жизненный цикл анализа данных.

Что такое машинное обучение.

Прогнозный анализ данных с помощью машинного обучения. Суть процесса машинного обучения и типичные приложения. Регрессионный анализ, линейная регрессия, области примене-

Согласована на портале № 52309

ния регрессионного анализа. Задачи классификации,

алгоритмы решения классификационных задач, алгоритм решающего дерева, области применения.

Каким образом данные и результаты их

анализа могут применяться при формулировании гипотез, обосновании решений и

предложений. Дедуктивная и индуктивная

логика, формулирование логических выводов. Опасность ошибок и заблуждений.

Эффект визуального представления результатов анализа данных. Средства визуализации, библиотека Matplotlib, модуль

Pyplot, возможности и свойства. Онлайн

сервис Plotly, создание и экспорт графического представления, работа в офлайн режиме. Общепринятые способы отображения данных. Знакомство с библиотекой Folium для отображения картографических данных.

Структура современного дата-центра.

Технологии виртуализации и облачные

вычисления. Уровни абстракции, гипервизоры, контейнеры. Аналитика в облаке

(Cloud analytics) и аналитика на периферии (Edge analytics). Виртуальный датацентр, IaaS, PaaS и SaaS сервисы. Виртуализованное хранилище данных, виртуализованные сетевые подключения. Средства

обеспечения масштабируемости, высокой

доступности и отказоустойчивости хранилищ Больших данных. Программное обеспечение Hadoop и файловая система

HDFS, обработка данных в Hadoop - Map

Reduce. Экосистема Hadoop. Конвейер обработки больших данных: загрузка данных, хранение и обработка. Проблемы загрузки Больших данных, программное

обеспечение Kafka. Проблемы хранения

Больших данных, NoSQL база данных

Cassandra. Сравнение Hadoop и Cassandra.

Проблемы обработки Больших данных, программное обеспечение Spark. Обработка данных в оперативной памяти с Resilient Distributed Data sets (RDDs).

Сравнение Spark и Map Reduce. Архитектура системы для хранения и обработки Больших данных - Lambda архитектура.

14.1.4. Вопросы для зачёта с оценкой

Каково назначение функции Лямбда в Python?

Каким образом выполняется переименование столбцов датафрейма?

Как выглядит код удаления столбца Datetime?

Можно ли объединить таблицы в базе данных, как это сделать?

Как при выводе значения переменной в текстовой строке оставить только 2 цифры после десятичной точки, если исходное значение содержит больше цифр?

Как в Python работает конструкция with … as

Какое значение в Python имеет оператор in как с его помощью организовать цикл for?

Столбец Datetime содержит информацию о дате и времени. Поясните, какое действие выполнит следующая команда: df['Date'] = df['Datetime'].apply(lambda dt_str:

pd.to_datetime(dt_str).date())

Какая функция или метод Python позволяет выполнить корреляционный анализ?

Что такое ковариация и как с помощью NumPy найти ковариацию двух переменных?

Как выполняются изменения файла в базе данных Cassandra?

а) При внесении изменений они дописываются в конец файла

б) Создается новый файл с изменениями, а старый удаляется

в) Сохраняются обе версии файла до и после внесения изменений

г) Изменения вносятся в существующий файл и он сохранятся

Что является первым компонентом конвейера больших данных:

а) сохранение данных

б) обработка данных

Согласована на портале № 52309

в) прием данных
г) передача данных
Какое решение повышает доступность приложений для работы с большими данными, сохраняя часто запрашиваемые данные в памяти для быстрого доступа?
а) sharding
б) memcaching
в) load balancing
г) distributed databases

14.1.5. Темы лабораторных работ

Анализ криминальной обстановки в СанФранциско
Измерение и анализ данных Интернет соединения
Измерение и анализ Интернет-соединения с хранилищем данных в виде SQL базы данных
Описательная статистика в Python
Корреляционный анализ в Python
Простая линейная регрессия в Python
Классификация с помощью дерева решений
Обработка ошибок при использовании метода линейной регрессии
Измерение и анализ Интернет-соединения
визуализация результатов

14.2. Требования к оценочным материалам для лиц с ограниченными возможностями здоровья и инвалидов

Для лиц с ограниченными возможностями здоровья и инвалидов предусмотрены дополнительные оценочные материалы, перечень которых указан в таблице 14.

Таблица 14 – Дополнительные материалы оценивания для лиц с ограниченными возможностями здоровья и инвалидов

Категории обучающихся	Виды дополнительных оценочных материалов	Формы контроля и оценки результатов обучения
С нарушениями слуха	Тесты, письменные самостоятельные работы, вопросы к зачету, контрольные работы	Преимущественно письменная проверка
С нарушениями зрения	Собеседование по вопросам к зачету, опрос по терминам	Преимущественно устная проверка (индивидуально)
С нарушениями опорно-двигательного аппарата	Решение дистанционных тестов, контрольные работы, письменные самостоятельные работы, вопросы к зачету	Преимущественно дистанционными методами
С ограничениями по общемедицинским показаниям	Тесты, письменные самостоятельные работы, вопросы к зачету, контрольные работы, устные ответы	Преимущественно проверка методами исходя из состояния обучающегося на момент проверки

14.3. Методические рекомендации по оценочным материалам для лиц с ограниченными возможностями здоровья и инвалидов

Для лиц с ограниченными возможностями здоровья и инвалидов предусматривается доступная форма предоставления заданий оценочных средств, а именно:

- в печатной форме;
- в печатной форме с увеличенным шрифтом;
- в форме электронного документа;
- методом чтения ассистентом задания вслух;
- предоставление задания с использованием сурдоперевода.

Лицам с ограниченными возможностями здоровья и инвалидам увеличивается время на подготовку ответов на контрольные вопросы. Для таких обучающихся предусматривается доступная форма предоставления ответов на задания, а именно:

Согласована на портале № 52309

– письменно на бумаге;
– набор ответов на компьютере;
– набор ответов с использованием услуг ассистента;
– представление ответов устно.

Процедура оценивания результатов обучения лиц с ограниченными возможностями здоровья и инвалидов по дисциплине предусматривает предоставление информации в формах, адаптированных к ограничениям их здоровья и восприятия информации:

Для лиц с нарушениями зрения:
– в форме электронного документа;
– в печатной форме увеличенным шрифтом.

Для лиц с нарушениями слуха:
– в форме электронного документа;
– в печатной форме.

Для лиц с нарушениями опорно-двигательного аппарата:
– в форме электронного документа;
– в печатной форме.

При необходимости для лиц с ограниченными возможностями здоровья и инвалидов процедура оценивания результатов обучения может проводиться в несколько этапов.

Согласована на портале № 52309