

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ
РОССИЙСКОЙ ФЕДЕРАЦИИ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«ДАГЕСТАНСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»
Экономический факультет

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ

Технологии Big Data

Кафедра «Информационных систем и технологий программирования»

Образовательная программа
38.03.01 Экономика

Направленность (профиль) программы
**Финансы и кредит, Бухгалтерский учет, анализ и аудит, Региональная
экономика**

Уровень высшего образования
бакалавриат

Форма обучения
очная

Статус дисциплины: входит в обязательную часть ОПОП

Махачкала, 2021

Рабочая программа дисциплины «Технологии Big Data» составлена в 2021 году в соответствии с требованиями ФГОС ВО – бакалавриат по направлению подготовки 38.03.01 Экономика от 12 августа 2020 г. №954.

Разработчик(и): кафедра информационных систем и технологий программирования, зав. каф., к.э.н., доцент Исмиханов З.Н.

Рабочая программа дисциплины одобрена:
на заседании кафедры ИСиТП от «29» июня 2021г., протокол № 11
Зав. кафедрой _____ Исмиханов З.Н.
(подпись)

на заседании Методической комиссии факультета ИиИТ
от «29» июня 2021г., протокол № 11.

Председатель _____ Бакмаев А.Ш.
(подпись)

на заседании Методической комиссии экономического факультета
от «5» июля 2021г., протокол № 10.

Председатель _____ Сулейманова Д.А.

Рабочая программа дисциплины согласована с учебно-методическим управлением «9» июля 2021 г.

Начальник УМУ _____ Гасангаджиева А.Г.
(подпись)
(подпись)

Аннотация рабочей программы дисциплины

Дисциплина «Технологии Big Data» входит в обязательную часть ОПОП бакалавриата по направлению подготовки 38.03.01 Экономика (уровень бакалавриата).

Дисциплина реализуется на экономическом факультете кафедрой информационных систем и технологий программирования.

Содержание дисциплины охватывает круг вопросов, связанных с рассмотрением основных методов анализа данных, математического моделирования и принятия решений применительно к решению задач в социально-экономической, финансовой и банковской сферах.

Дисциплина нацелена на формирование следующих компетенций выпускника: ОПК-2, ОПК-5.

Преподавание дисциплины предусматривает проведение следующих видов учебных занятий: лекции, практические занятия, лабораторные занятия, самостоятельная работа. Рабочая программа дисциплины предусматривает проведение следующих видов контроля успеваемости: текущий контроль в форме опроса, тестов, контрольных работ и промежуточный контроль в форме зачета.

Объем дисциплины 3 зачетные единицы, в том числе в академических часах по видам учебных занятий:

Очная форма обучения

Семестр	Учебные занятия							СРС	Форма промежуточной аттестации (зачет, дифференцированный зачет, экзамен)
	в том числе:								
	всего	Контактная работа обучающихся с преподавателем					консультации		
		всего	Лекции	Лабораторные занятия	Практические занятия	КСР			
6	108	48	16	16	16			60	Зачет

1. Цели освоения дисциплины

Целями изучения дисциплины являются изучение методов и подходов к анализу данных различного объема, включая предварительную обработку данных и статистический анализ, освоение различных моделей машинного обучения, предназначенных для решения задач кластеризации, классификации и регрессии и применение их для решения прикладных задач из различных сфер человеческой деятельности.

2. Место дисциплины в структуре ОПОП бакалавриата

Предшествующими курсами, на которых непосредственно базируется дисциплина «Технологии Big Data» являются «Математика для экономистов», «Введение в информационные технологии», «Инструментальные методы цифровой экономики», «Статистика», «Эконометрика», «Микроэкономика», «Макроэкономика». Результаты обучения по дисциплине могут быть использованы при выполнении ВКР.

Особенностью дисциплины является то, что в процессе изучения дисциплины обучающимися осуществляется умение использовать возможности нейросетевых технологий и машинного обучения, позволяющие реализовывать решение любой задачи из различных отраслей экономики и областей деятельности человеческого общества. Но наиболее эффективно с помощью искусственных нейронных сетей решаются задачи высокой вычислительной сложности, а также трудноформализуемые и неформализуемые, нередко неразрешимые средствами традиционной математики.

3. Компетенции обучающегося, формируемые в результате освоения дисциплины (перечень планируемых результатов обучения и процедура освоения).

Код и наименование компетенции из ОПОП	Код и наименование индикатора достижения компетенций (в соответствии с ОПОП)	Планируемые результаты обучения	Процедура освоения
ОПК-2 Способен осуществлять сбор, обработку и статистический анализ данных, необходимых для решения поставленных экономических задач	ОПК-2.1. Осуществляет сбор, обработку и анализ статистической информации, необходимой для решения поставленных экономических задач	Знает: теоретические основы методов анализа больших данных и машинного обучения; типы задач, решаемые с использованием методов анализа больших данных и машинного обучения; современные технологии анализа больших данных и машинного обучения Умеет: ставить задачи и	

		<p>разрабатывать программу исследования с использованием анализа больших данных и машинного обучения; выбирать соответствующие методы решения экспериментальных и теоретических задач; применять методы и подходы анализа больших данных и машинного обучения для решения экспериментальных и теоретических задач; интерпретировать, представлять и применять результаты, полученные с использованием анализа больших данных и машинного обучения.</p> <p>Владеет:</p> <p>основными методами, подходами и инструментарием анализа больших данных и машинного обучения; навыком критически анализировать современные проблемы инноватики с использованием анализа больших данных и машинного обучения. Должен демонстрировать способность и готовность: с использованием методов и подходов анализа больших данных и машинного обучения критически анализировать современные проблемы инноватики, ставить задачи и разрабатывать программу исследования, выбирать</p>	
--	--	--	--

		соответствующие методы решения экспериментальных и теоретических задач, интерпретировать, представлять и применять полученные результаты	
ОПК-5. Способен использовать современные информационные технологии и программные средства при решении профессиональных задач	ОПК-5.1. Выбирает соответствующие содержанию профессиональных задач современные информационные технологии и программное обеспечение	Знает: алгоритмы и программы, пригодные для практического применения; основы информатики, математики, линейной алгебры и программирования для анализа данных, конструирования и разработки интеллектуальных алгоритмов; Умеет: разрабатывать алгоритмы и программы, пригодные для практического применения; применять основы информатики, математики, линейной алгебры и программирования для анализа данных, конструирования и разработки интеллектуальных алгоритмов Владеет: навыками разработки алгоритмов и программ, пригодных для практического применения, применять основы информатики, математики, линейной алгебры и программирования для анализа данных, конструирования и разработки интеллектуальных алгоритмов	Устный опрос, письменный опрос; выполнение проекта.

4. Объем, структура и содержание дисциплины.

4.1. Объем дисциплины составляет 3 зачетные единицы, 108 академических часов.

4.2. Структура дисциплины.

4.2.1. Структура дисциплины в очной форме

№ п/п	Разделы и темы дисциплины по модулям	Семестр	Виды учебной работы, включая самостоятельную работу студентов (в часах)					Формы текущего контроля успеваемости и промежуточной аттестации
			Лекции	Практические занятия	Лабораторные занятия	...	Самостоятельная работа в т.ч. экзамен	
Модуль 1. Введение в анализ данных. Очистка, интеграция и преобразование данных.								
1	Тема 1. Введение в интеллектуальный анализ данных: основные понятия, области применения современных технологий обработки и интеллектуального анализа больших данных.	6	2	2	2		10	опрос, тестирование контрольная работа
2	Тема 2. Библиотека pandas. Структуры данных в pandas, работа со структурами данных. Операции над данными. Комбинирование данных из разных источников. Обработка пропущенных значений.	6	2	2	2		14	опрос, тестирование контрольная работа
Итого по модулю 1:		36	4	4	4		24	
Модуль 2. Программные модули и пакеты для работы с многомерными массивами данных. Визуализация данных.								
	Тема 3. Библиотеки NumPy, SciPy: основные функции. Визуализация данных с matplotlib и pandas.	6	2	2	2		10	опрос, тестирование контрольная работа
	Тема 4. Статистика: описание одиночного набора данных, показатели центра распределения, показатели вариации, корреляция и причинная зависимость. Проверка статистических гипотез.	6	4	4	4		8	опрос, тестирование контрольная работа
Итого по модулю 2:		36	6	6	6		18	
Модуль 3. Методы машинного обучения.								
	Тема 5. Задачи машинного обучения. Применение машинного обучения в Data Science. Алгоритмы машинного обучения. Библиотека Scikit-Learn.	6	2	2	4		8	опрос, тестирование контрольная работа
	Тема 6. Алгоритмы машинного обучения.	6	4	4	4		10	опрос, тестирование

	Способы и типы машинного обучения. Библиотека Scikit-Learn. Оценка и улучшение качества моделей машинного обучения: перекрестная проверка, поиск по сетке, метрики качества моделей и их вычисление.							контрольная работа
	Итого по модулю 3:	36	6	6	6		18	
	Итого	108	16	16	16		60	

4.3. Содержание дисциплины, структурированное по темам (разделам).

4.3.1. Содержание лекционных занятий по дисциплине.

Модуль 1. Введение в анализ данных. Очистка, интеграция и преобразование данных.

Тема 1. Введение в интеллектуальный анализ данных: основные понятия, области применения современных технологий обработки и интеллектуального анализа больших данных.

Этапы анализа данных. Структурированные и неструктурированные данные. Сбор и подготовка данных. Большие данные. Организация сбора и хранения больших наборов данных. Nadoop.

Тема 2. Библиотека pandas. Структуры данных в pandas, работа со структурами данных. Операции над данными. Комбинирование данных из разных источников. Обработка пропущенных значений.

Восстановление пропущенных значений в массивах данных. Этапы анализа данных. Структурированные и неструктурированные данные. Сбор и подготовка данных.

Модуль 2. Программные модули и пакеты для работы с многомерными массивами данных. Визуализация данных.

Тема 3. Библиотеки NumPy, SciPy: основные функции. Визуализация данных с matplotlib и pandas.

Работа с библиотеками NumPy и SciPy.

Тема 4. Статистика: описание одиночного набора данных, показатели центра распределения, показатели вариации, корреляция и причинная зависимость. Проверка статистических гипотез.

Модуль 3. Методы машинного обучения.

Тема 5. Задачи машинного обучения. Применение машинного обучения в Data Science. Алгоритмы машинного обучения. Библиотека Scikit-Learn.

Тема 6. Алгоритмы машинного обучения. Способы и типы машинного обучения. Библиотека Scikit-Learn. Оценка и улучшение качества моделей машинного обучения: перекрестная проверка, поиск по сетке, метрики качества моделей и их вычисление.

4.3.2. Содержание практических занятий по дисциплине.

Модуль 1. Введение в анализ данных. Очистка, интеграция и преобразование данных.

Тема 1. Введение в интеллектуальный анализ данных: основные понятия, области применения современных технологий обработки и интеллектуального анализа больших данных.

Этапы анализа данных. Структурированные и неструктурированные данные. Сбор и подготовка данных. Большие данные. Организация сбора и хранения больших наборов данных. Hadoop.

Тема 2. Библиотека pandas. Структуры данных в pandas, работа со структурами данных. Операции над данными. Комбинирование данных из разных источников. Обработка пропущенных значений.

Восстановление пропущенных значений в массивах данных. Этапы анализа данных. Структурированные и неструктурированные данные. Сбор и подготовка данных.

Модуль 2. Программные модули и пакеты для работы с многомерными массивами данных. Визуализация данных.

Тема 3. Библиотеки NumPy, SciPy: основные функции. Визуализация данных с matplotlib и pandas.

Работа с библиотеками NumPy и SciPy.

Тема 4. Статистика: описание одиночного набора данных, показатели центра распределения, показатели вариации, корреляция и причинная зависимость. Проверка статистических гипотез.

Модуль 3. Методы машинного обучения.

Тема 5. Задачи машинного обучения. Применение машинного обучения в Data Science. Алгоритмы машинного обучения. Библиотека Scikit-Learn.

Тема 6. Алгоритмы машинного обучения. Способы и типы машинного обучения. Библиотека Scikit-Learn. Оценка и улучшение качества моделей машинного обучения: перекрестная проверка, поиск по сетке, метрики качества моделей и их вычисление.

4.3.3. Содержание лабораторно занятий по дисциплине

Модуль 1. Название модуля

Лабораторная работа №1 «Основы языка Python»

Цель: ознакомиться с основами языка Python, получить умения для выполнения дальнейших лабораторных работ. Задачи:

- изучить типизацию данных;

- научиться пользоваться циклами «for» и «while»;
- рассмотреть «ветвление» в Python;
- отработать задачи с использованием конструкции «try-except»;
- разобрать функции и пространства имён.

Лабораторная работа №2 «Классификация данных»

Цель: научиться работать с данными при помощи визуальных инструментов и разобрать азы классификации при помощи построения простейшего классификатора со статичными параметрами.

Задачи:

- научиться анализировать данные;
- сформировать понятие математических срезов;
- получить умения в работе с визуальными инструментами;
- построить классификатор на основе данных полученных при анализе;
- научиться калибровать нейросеть для получения более точных ответов.

Лабораторная работа №3 «Классификация методом «К-ближайших соседей"»».

Цель: изучить метод простейший метод классификации данных «К-ближайших соседей» и научиться производить оценку данных с помощью визуальных инструментов Python.

Задачи:

- детально разобрать метод машинного обучения «К-ближайших соседей»;
- научиться работать с информацией;
- сформировать понятие математических срезов;
- получить умения в работе с визуальными инструментами;
- научиться калибровать нейросеть для получения более точных ответов.

Лабораторная работа № 4 «Основы работы с Pandas»

Цель: научиться пользоваться библиотекой Pandas и её встроенными объектами для визуализации данных в датасетах.

Задачи:

- получить умения по использованию библиотеки Pandas;
- сформировать понятия о DataFrame и Series;
- научиться строить графики с помощью scatter matrix (матрица рассеивания) и matplotlib.

Лабораторная работа № 5 «Деревья решений»

Цель: познакомить обучающихся с методом машинного обучения, построенном на деревьях решений, а также научить строить сами деревья.

Задачи:

- рассмотреть понятие дерева решений; • рассмотреть варианты применения данной классификации;
- обучить модель на основе классов;
- отобразить дополнительный класс на модели и посмотреть результат;
- рассмотреть плюсы и минусы данной модели.

5. Образовательные технологии

Информационная функция лекции предполагает передачу необходимой информации по теме, которая должна стать основой для дальнейшей самостоятельной работы студента.

Мотивационная функция должна заключаться в стимулировании интереса студентов к науке. На лекции необходимо заинтересовывать, озадачить студентов с целью выработки у них желания дальнейшего изучения той или иной экономической проблемы.

Воспитательная функция ориентирована на формирование у молодого поколения чувства ответственности, закладку нравственных, эстетических норм поведения в обществе и коллективе, формирование патриотических взглядов, мотивов социального поведения и действий, финансово-экономического мировоззрения.

Обучающая функция реализуется посредством формирования у студентов навыков работы с первоисточниками и научной и учебной литературой.

6. Учебно-методическое обеспечение самостоятельной работы студентов.

Самостоятельная работа студентов включает в себя:

- изучение материалов лекционных занятий, рекомендованной литературы и источников;
- подготовка к лабораторной работе;
- подготовка домашних заданий и выполнение самостоятельной работы;
- подготовка к промежуточной аттестации.

Вид самостоятельной работы	Примерная трудоёмкость, а.ч.	Формируемые компетенции
	очная	
Текущая СРС		
работа с лекционным материалом, с учебной литературой	4	ОПК-2, ОПК-5
опережающая самостоятельная работа (изучение нового материала до его изложения на занятиях)	4	ОПК-2, ОПК-5
самостоятельное изучение разделов дисциплины	8	ОПК-2, ОПК-5
выполнение домашних заданий, домашних контрольных работ	6	ОПК-2, ОПК-5
подготовка к лабораторным работам, к практическим и семинарским занятиям	10	ОПК-2, ОПК-5
подготовка к контрольным работам, коллоквиумам, зачётам	10	ОПК-2, ОПК-5
Творческая проблемно-ориентированная СРС		
выполнение расчётно-графических работ	4	ОПК-2, ОПК-5
поиск, изучение и презентация информации по заданной проблеме, анализ научных публикаций по заданной теме	4	ОПК-2, ОПК-5

исследовательская работа, участие в конференциях, семинарах, олимпиадах	6	ОПК-2, ОПК-5
анализ данных по заданной теме, выполнение расчётов, составление схем и моделей на основе собранных данных	6	ОПК-2, ОПК-5
Итого СРС:	60	

7. Фонд оценочных средств для проведения текущего контроля успеваемости, промежуточной аттестации по итогам освоения дисциплины.

7.1. Типовые контрольные задания

Описание задания «Машинное зрение и распознавание лиц в реальном времени»

Цель: сформировать умения по использованию библиотеки OpenCV самостоятельному поиску информации, связанной с решением поставленной задачи.

Задачи:

- закрепить умения работы с машинным зрением;
- сформировать умение поиска информации;
- сформировать умения по разработке полноценного приложения;
- научиться отрабатывать ошибки при разработке приложения.

7.2. Методические материалы, определяющие процедуру оценивания знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций.

Общий результат выводится как интегральная оценка, складывающаяся из текущего контроля - 50% и промежуточного контроля - 50%.

Текущий контроль по дисциплине включает:

- посещение занятий - 10 баллов,
- участие на практических занятиях - 30 баллов,
- выполнение лабораторной работы – 10 баллов.

Промежуточный контроль по дисциплине включает:

- устный опрос - 10 баллов,
- письменная контрольная работа - 40 баллов,

8. Учебно-методическое обеспечение дисциплины.

а) основная литература

1. Барский А.Б. Введение в нейронные сети [Электронный ресурс]: учебное пособие/ Барский А.Б.— Электрон. текстовые данные. — Москва, Саратов: Интернет Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. — 357 с.— [Электронный ресурс]. [http:// www.iprbookshop.ru/89426.html](http://www.iprbookshop.ru/89426.html)
2. Яхьяева Г.Э. Нечеткие множества и нейронные сети [Электронный ресурс]: учебное пособие/ Яхьяева Г.Э.— Электрон. текстовые данные.— Москва, Саратов: Интернет-Университет Информационных Технологий (ИНТУИТ), Вузовское образование, 2017.— 320 с.— [Электронный ресурс].- <http://www.iprbookshop.ru/67390.html>.

б) дополнительная литература

1. Самуйлов С.В. Алгоритмы и структуры обработки данных [Электронный ресурс]: учебное пособие/ Самуйлов С.В.— Электрон. текстовые данные.— Саратов: Вузовское образование, 2016.— 132 с.— Режим доступа: <http://www.iprbookshop.ru/47275.html>.

9. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины.

1. Портал доступа к электронным образовательным ресурсам ДГУ [Электронный ресурс] Режим доступа URL: <http://dgu.ru> (дата обращения: 1.07.2021)

2. Электронно-библиотечная система IPRbooks [Электронный ресурс] Режим доступа URL: <http://www.iprbookshop.ru> (дата обращения: 1.07.2021)

3. Электронно-библиотечная система «Университетская библиотека онлайн» [Электронный ресурс] Режим доступа URL: <http://biblioclub.ru> (дата обращения: 1.09.2021)
43

4. Электронно-библиотечная система издательства «Инфра» [Электронный ресурс] Режим доступа URL: <http://znanium.com> (дата обращения: 1.09.2021)

5. IT-портал [Электронный ресурс] Режим доступа URL: <http://citforum.ru>

6. Портал Национального открытого университета «Интуит» [Электронный ресурс] Режим доступа URL: <http://www.intuit.ru> (дата обращения: 1.09.2021)

7. Техническая документация фирмы Microsoft [Электронный ресурс] Режим доступа URL: <http://technet.microsoft.com/ru-ru/sysinternals>(дата обращения: 1.07.2021)

8. Moodle [Электронный ресурс]: система виртуального обучения: [база данных] / Даг.гос. ун-т. - Махачкала, г. - Доступ из сети ДГУ или, после регистрации из сети ун-та, из любой точки, имеющей доступ в интернет. -Режим доступа URL: <http://moodle.dgu.ru/>(дата обращения: 1.09.2021).

9. Учебный курс по программированию «Учите Питон» [Электронный ресурс] - Режим доступа: <http://pythontutor.ru/> (дата обращения: 1.07.2021)

10. Платформа учебных курсов Stepik [Электронный ресурс] - Режим доступа URL: <https://stepik.org/catalog?q=Python> (дата обращения: 1.07.2021)

10. Методические указания для обучающихся по освоению дисциплины.

Для изучения теоретического курса студентам необходимо использовать лекционный материал, учебники и учебные пособия из списка основной и дополнительной литературы, интернет источники. По дисциплине в конце каждого модуля проводится контрольная работа. В контрольную работу включаются теоретические вопросы и задачи тех типов, которые были разобраны на предшествующих практических занятиях. Рабочей программой дисциплины предусмотрена самостоятельная работа студентов в объеме 60 часов. Самостоятельная работа проводится с целью углубления знаний по дисциплине и предусматривает:

- чтение студентами рекомендованной литературы и усвоение теоретического материала дисциплины;

- подготовку к практическим занятиям; - выполнение индивидуальных заданий;

- подготовку к контрольным работам, зачету и экзаменам.

С самого начала изучения дисциплины студент должен четко уяснить, что без систематической самостоятельной работы успех невозможен. Эта работа должна регулярно начинаться сразу после лекционных и практических занятий, для закрепления только что пройденного материала. После усвоения теоретического материала можно приступить к самостоятельному решению задач из учебников и пособий, входящих в

список основной литературы.

11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень программного обеспечения и информационных справочных систем.

Интернет-ресурсы, мультимедиа, электронная почта для коммуникации со студентами. Обучающие интерактивные интернет-порталы, интернет-ресурсы, мультимедиа, электронная почта для коммуникации со студентами, системы программирования, Microsoft Office.

Компьютерные классы с доступом к сети интернет и с набором лицензионного базового программного обеспечения для проведения лабораторных занятий; Python IDE (Microsoft Visual Studio, Microsoft Visual Studio Code, PyCharm), для выполнения лабораторных заданий. Лекционная мультимедийная аудитория для чтения лекций с использованием мультимедийных материалов.

12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине.

Компьютерные классы и лаборатории, оборудованные для проведения лекционных, практических и лабораторных занятий средствами оргтехники, персональными компьютерами, объединенными в сеть с выходом в Интернет; установленное лицензионное и свободное программное обеспечение. Для выполнения лабораторных работ используется компьютерное оборудование с установленными программными продуктами MSOffice, Интерпретатор Python, Notebook. Аудиторные занятия проводятся в компьютерных классах с доступом к сети Интернет.