Министерство науки и высшего образования Российской Федерации

ФГБОУ ВО «БАЙКАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ
Проректор по учебной работе
д.юр.н., доц. Васильева Н.В.

_____

26.06.2023г.

Рабочая программа дисциплины
Б1.У.1. Data Science

Направление подготовки: 09.04.03 Прикладная информатика
Направленность (профиль): Цифровые технологии в экономике
Квалификация выпускника: магистр
Форма обучения: очная

Курс	2
Семестр	21
Лекции (час)	14
Практические (сем, лаб.) занятия (час)	28
Самостоятельная работа, включая подготовку к экзаменам и зачетам (час)	138
Курсовая работа (час)	
Всего часов	180
Зачет (семестр)	21
Экзамен (семестр)	

Иркутск 2023

Программа составлена в соответствии с ФГОС ВО по направлению 09.04.03 Прикладная информатика.

Автор А.В. Родионов

Рабочая программа обсуждена и утверждена на заседании кафедры математических методов и цифровых технологий

Заведующий кафедрой А.В. Родионов

Дата актуализации рабочей программы: 28.06.2024

1. Цели изучения дисциплины

Цели дисциплины - формирование представление о типах задач, возникающих в области Data Science и методах их решения, которые помогут студентам выявлять, формализовать и успешно решать практические задачи анализа данных, возникающие в процессе их профессиональной деятельности.

2. Перечень планируемых результатов обучения по дисциплине (модулю), соотнесенных с планируемыми результатами освоения образовательной программы

Компетенции обучающегося, формируемые в результате освоения дисциплины

Код компетенции по ФГОС ВО	Компетенция
ПК-5	Способен управлять аналитическими ресурсами и компетенциями

Структура компетенции

Компетенция	Формируемые ЗУНы
ПК-5 Способен управлять аналитическими ресурсами и компетенциями	З. Знать теоретические основы управления аналитическими ресурсами и компетенциями У. Уметь управлять аналитическими ресурсами и компетенциями Н. Владеть навыками управления аналитическими ресурсами и компетенциями

3. Место дисциплины (модуля) в структуре образовательной программы

Принадлежность дисциплины - БЛОК 1 ДИСЦИПЛИНЫ (МОДУЛИ): Часть, формируемая участниками образовательных отношений.

Предшествующие дисциплины (освоение которых необходимо для успешного освоения данной): "Компьютерный анализ и интерпретация данных", "Системы поддержки принятия решений", "Методы принятия решений"

Дисциплины, использующие знания, умения, навыки, полученные при изучении данной: "Облачные и блокчейн-технологии в бизнесе"

4. Объем дисциплины (модуля) в зачетных единицах с указанием количества академических часов, выделенных на контактную работу обучающихся с преподавателем (по видам учебных занятий) и на самостоятельную работу обучающихся

Общая трудоемкость дисциплины составляет 5 зач. ед.,180 часов.

Вид учебной работы	Количество часов
Контактная(аудиторная) работа	
Лекции	14
Практические (сем, лаб.) занятия	28
Самостоятельная работа, включая подготовку к экзаменам и зачетам	138
Всего часов	180

5. Содержание дисциплины (модуля), структурированное по темам (разделам) с указанием отведенного на них количества академических часов и видов учебных занятий

5.1. Содержание разделов дисциплины

№ п/п	Раздел и тема дисциплины	Семе-стр	Лек-ции	Семинар Лаборат. Практич.	Само-стоят. раб.	В интера-ктивной форме	Формы текущего контроля успеваемости
1	Data Science и его задачи	21	4	8	36		Задачи и инструменты DataScience.. Язык запросов SQL
2	Хранилища данных	21	5	10	50		Документно-ориентированная модель хранения данных. Реляционная и графовая модели хранения данных
3	Введение в искусственный интеллект и основные методы машинного обучения для работы с данными	21	5	10	52		Использование аналитических платформ для проведения анализа данных. Итоговое тестирование по курсу. Методы работы с таблицами в Python. Агрегация и визуализация данных. Проведение первичного анализа данных
	ИТОГО		14	28	138		

5.2. Лекционные занятия, их содержание

№ п/п	Наименование разделов и тем	Содержание
1	Что такое DataScience. Типы задач.	Что такое DataScience. Типы задач.
2	Инструменты аналитика	Используемые для анализа информационные системы/языки программирования/доступа к данным
3	Реляционные хранилища данных	Реляционные хранилища данных
4	Нереляционные модели хранения данных	Графовая модель храниения данных, Документно-ориентированная модель хранения данных

№ п/п	Наименование разделов и тем	Содержание
5	NoSQL решения. MapReduce / Hadoop	NoSQL решения. MapReduce / Hadoop
6	Технологии анализа данных	Технологии анализа данных разных типов - текстовых данных, аудио и видео.
7	Big Data. Основные понятия. Хранение и обработка больших данных.	Big Data. Основные понятия. Хранение и обработка больших данных.

5.3. Семинарские, практические, лабораторные занятия, их содержание

№ раздела и темы	Содержание и формы проведения
1	Что такое DataScience. Типы задач.. Лабораторные занятия с использованием ПК
1	Инструменты аналитика. Лабораторные занятия с использованием ПК
2	Реляционные хранилища данных. Лабораторные занятия с использованием ПК
2	Нереляционные модели хранения данных. Лабораторные занятия с использованием ПК
2	NoSQL решения. MapReduce / Hadoop. Лабораторные занятия с использованием ПК
3	Технологии анализа данных. Лабораторные занятия с использованием ПК
3	Big Data. Основные понятия. Хранение и обработка больших данных.. Лабораторные занятия с использованием ПК

6. Фонд оценочных средств для проведения промежуточной аттестации по дисциплине (полный текст приведен в приложении к рабочей программе)

6.1. Текущий контроль

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: (З.1…З.n, У.1…У.n, Н.1…Н.n)	Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы (Наименование оценочного средства)	Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания (по 100-балльной шкале)
1	1. Data Science и его задачи	ПК-5	З.Знать теоретические основы управления аналитическими ресурсами и компетенциями У.Уметь управлять аналитическими ресурсами и компетенциями Н.Владеть навыками управления	Задачи и инструменты DataScience.	Полностью выполненная работа -15 баллов, частично - доля правильно выполненных заданий*15 баллов, если доля меньше 0,5 - 0 баллов (15)

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: (З.1…З.n, У.1…У.n, Н.1…Н.n)	Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы (Наименование оценочного средства)	Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания (по 100-балльной шкале)
			аналитическими ресурсами и компетенциями		
2		ПК-5	З.Знать теоретические основы управления аналитическими ресурсами и компетенциями У.Уметь управлять аналитическими ресурсами и компетенциями Н.Владеть навыками управления аналитическими ресурсами и компетенциями	Язык запросов SQL	Полностью выполненная лабораторная работа -10 баллов, частично - доля правильно выполненных заданий*10 баллов, если доля меньше 0,5 - 0 баллов (10)
3	2. Хранилища данных	ПК-5	З.Знать теоретические основы управления аналитическими ресурсами и компетенциями У.Уметь управлять аналитическими ресурсами и компетенциями Н.Владеть навыками управления аналитическими ресурсами и компетенциями	Документно-ориентированная модель хранения данных	Полностью выполненная лабораторная работа -15 баллов, частично - доля правильно выполненных заданий*15 баллов, если доля меньше 0,5 - 0 баллов (15)
4		ПК-5	З.Знать теоретические основы управления аналитическими ресурсами и компетенциями У.Уметь управлять аналитическими ресурсами и компетенциями Н.Владеть навыками управления аналитическими ресурсами и компетенциями	Реляционная и графовая модели хранения данных	Полностью выполненная лабораторная работа -10 баллов, частично - доля правильно выполненных заданий*10 баллов, если доля меньше 0,5 - 0 баллов (10)
5	3. Введение в искусственный интеллект и основные методы машинного обучения для работы с	ПК-5	З.Знать теоретические основы управления аналитическими ресурсами и компетенциями У.Уметь управлять аналитическими	Использование аналитических платформ для проведения анализа данных	Полностью выполненная лабораторная работа -15 баллов, частично - доля правильно выполненных

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: (З.1…З.n, У.1…У.n, Н.1…Н.n)	Контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности, характеризующих этапы формирования компетенций в процессе освоения образовательной программы (Наименование оценочного средства)	Описание показателей и критериев оценивания компетенций на различных этапах их формирования, описание шкал оценивания (по 100-балльной шкале)
	данными		ресурсами и компетенциями Н.Владеть навыками управления аналитическими ресурсами и компетенциями		заданий*15 баллов, если доля меньше 0,5 - 0 баллов (15)
6		ПК-5	З.Знать теоретические основы управления аналитическими ресурсами и компетенциями У.Уметь управлять аналитическими ресурсами и компетенциями Н.Владеть навыками управления аналитическими ресурсами и компетенциями	Итоговое тестирование по курсу	Полностью выполненная работа -20 баллов, частично - доля правильно выполненных заданий*20 баллов, если доля меньше 0,5 - 0 баллов (20)
7		ПК-5	З.Знать теоретические основы управления аналитическими ресурсами и компетенциями У.Уметь управлять аналитическими ресурсами и компетенциями Н.Владеть навыками управления аналитическими ресурсами и компетенциями	Методы работы с таблицами в Python. Агрегация и визуализация данных. Проведение первичного анализа данных	Полностью выполненная лабораторная работа -15 баллов, частично - доля правильно выполненных заданий*15 баллов, если доля меньше 0,5 - 0 баллов (15)
				Итого	**100**

6.2. Промежуточный контроль (зачет, экзамен)

Рабочим учебным планом предусмотрен Зачет в семестре 21.

ВОПРОСЫ ДЛЯ ПРОВЕРКИ ЗНАНИЙ:

1-й вопрос билета (40 баллов), вид вопроса: Тест/проверка знаний. Критерий: Сумма баллов за правильные ответы. Правильный ответ на каждый вопрос - 4 балла, неправильный ответ - 0 баллов за вопрос..

Компетенция: ПК-5 Способен управлять аналитическими ресурсами и компетенциями

Знание: Знать теоретические основы управления аналитическими ресурсами и компетенциями

1. NoSQL, основные понятия, отличия от реляционных СУБД
2. Python. Сильные и слабые стороны по сравнению с другими технологиями.
3. Анализ текста. Латентное размещение Дирихле.
4. Ассоциативные правила. Определение. Достоверность и поддержка. Отличия построения ассоциативного правила от решающего правила задачи классификации. Привести примеры использования ассоциативных правил.
5. Визуализация данных в Python. Matplotlib и Seaborn. Обзор различных видов графиков
6. Визуализация. Дать определение визуализации. Показать важность визуализации в аналитике больших данных. Привести примеры использования визуализации.
7. Генераторы и Lambda-функции в языке Python
8. Жизненный цикл» проекта по аналитике больших данных.
9. Классификация. Признаковое описание объекта и таблица объект-свойства. Постановка задачи. Отличия задачи классификации от задачи регрессии. Определение модели и алгоритма. Процесс обучения.
10. Кластеризация. Метрики. Матрица парных расстояний. Постановка задачи кластеризации. Отличие от задачи классификации. Привести примеры использования алгоритмов кластеризации.
11. Машинное обучение в SciKit-Learn
12. Модуль для работы с числовыми данными NumPy
13. Научные проблемы больших данных.
14. Обработка больших данных с помощью библиотек Pandas, CSV и OpenPyXL
15. Определение класса. Инкапсуляция. Полиморфизм. Наследование. Пример задания класса в Python.
16. Парадигма Map Reduce. Описать принцип работы. Перечислить слабые и сильные стороны. Обозначить области применимости. Привести примеры использования.
17. Понятие о больших данных
18. Проблема переобучения. Регуляризация. Cross validation. Привести примеры использования алгоритмов классификации.
19. Работа с файлами. Чтение и запись в файл в Python.
20. Типовая архитектура проекта в области больших данных.
21. Типы хранилищ данных NoSQL. Хранилище «ключ — значение»
22. Функции в Python. Особенности, определения. Рекурсия.

ТИПОВЫЕ ЗАДАНИЯ ДЛЯ ПРОВЕРКИ УМЕНИЙ:

2-й вопрос билета (30 баллов), вид вопроса: Задание на умение. Критерий: Полностью выполненное задание – 30 баллов, допущены незначительные ошибки - (20-30) баллов, существенные ошибки, не искажающие сути - (10-20) баллов, грубые ошибки - (0-10) баллов.

Компетенция: ПК-5 Способен управлять аналитическими ресурсами и компетенциями

Умение: Уметь управлять аналитическими ресурсами и компетенциями

Задача № 1. Создать новый проект и продемонстрировать умения использовать различные инструменты, алгоритмы и методы управления аналитическими ресурсами в соответствии с вариантом

ТИПОВЫЕ ЗАДАНИЯ ДЛЯ ПРОВЕРКИ НАВЫКОВ:

3-й вопрос билета (30 баллов), вид вопроса: Задание на навыки. Критерий: Полностью выполненное задание – 30 баллов, допущены незначительные ошибки - (20-30) баллов, существенные ошибки, не искажающие сути - (10-20) баллов, грубые ошибки - (0-10) баллов.

Компетенция: ПК-5 Способен управлять аналитическими ресурсами и компетенциями
Навык: Владеть навыками управления аналитическими ресурсами и компетенциями
Задание № 1. Создать новый проект и продемонстрировать навыки использования различных инструментов, алгоритмов и методов управления аналитическими ресурсами в соответствии с вариантом

ОБРАЗЕЦ БИЛЕТА

Министерство науки и высшего образования Российской Федерации Федеральное государственное бюджетное образовательное учреждение высшего образования **«БАЙКАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»** **(ФГБОУ ВО «БГУ»)**	Направление - 09.04.03 Прикладная информатика Профиль - Цифровые технологии в экономике Кафедра математических методов и цифровых технологий Дисциплина - Data Science

БИЛЕТ № 1

1. Тест (40 баллов).
2. Создать новый проект и продемонстрировать умения использовать различные инструменты, алгоритмы и методы управления аналитическими ресурсами в соответствии с вариантом (30 баллов).
3. Создать новый проект и продемонстрировать навыки использования различных инструментов, алгоритмов и методов управления аналитическими ресурсами в соответствии с вариантом (30 баллов).

Составитель _____ А.В. Родионов

Заведующий кафедрой _____ А.В. Родионов

7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины (модуля)

а) основная литература:
1. Воронова, Л. И. Big Data. Методы и средства анализа [Электронный ресурс] : учебное пособие / Л. И. Воронова, В. И. Воронов. — Электрон. текстовые данные. — М. : Московский технический университет связи и информатики, 2016. — 33 с. — 2227-8397. — Режим доступа: http://www.iprbookshop.ru/61463.html
2. Уэс Маккинли Python и анализ данных [Электронный ресурс] / Маккинли Уэс. — Электрон. текстовые данные. — Саратов: Профобразование, 2017. — 482 с. — 978-5-4488-0046-7. — Режим доступа: http://www.iprbookshop.ru/64058.html

б) дополнительная литература:
1. Буйначев, С. К. Основы программирования на языке Python [Электронный ресурс] : учебное пособие / С. К. Буйначев, Н. Ю. Боклаг ; под ред. Ю. В. Песин. — Электрон.

текстовые данные. — Екатеринбург : Уральский федеральный университет, ЭБС АСВ, 2014. — 92 с. — 978-5-7996-1198-9. — Режим доступа: http://www.iprbookshop.ru/66183.html

8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины (модуля), включая профессиональные базы данных и информационно-справочные системы

Для освоения дисциплины обучающемуся необходимы следующие ресурсы информационно-телекоммуникационной сети «Интернет»:

– Сайт Байкальского государственного университета, адрес доступа: http://bgu.ru/, доступ круглосуточный неограниченный из любой точки Интернет

– Электронно-библиотечная система IPRbooks, адрес доступа: https://www.iprbookshop.ru. доступ неограниченный

9. Методические указания для обучающихся по освоению дисциплины (модуля)

Изучать дисциплину рекомендуется в соответствии с той последовательностью, которая обозначена в ее содержании. Для успешного освоения курса обучающиеся должны иметь первоначальные знания в области анализа, обработки и представления данных.

На лекциях преподаватель озвучивает тему, знакомит с перечнем литературы по теме, обосновывает место и роль этой темы в данной дисциплине, раскрывает ее практическое значение. В ходе лекций студенту необходимо вести конспект, фиксируя основные понятия и проблемные вопросы.

Практические (семинарские) занятия по своему содержанию связаны с тематикой лекционных занятий. Начинать подготовку к занятию целесообразно с конспекта лекций. Задание на практическое (семинарское) занятие сообщается обучающимся до его проведения. На семинаре преподаватель организует обсуждение этой темы, выступая в качестве организатора, консультанта и эксперта учебно-познавательной деятельности обучающегося.

Изучение дисциплины (модуля) включает самостоятельную работу обучающегося.

Основными видами самостоятельной работы студентов с участием преподавателей являются:

• текущие консультации;

• коллоквиум как форма контроля освоения теоретического содержания дисциплин: (в часы консультаций, предусмотренные учебным планом);

• прием и разбор домашних заданий (в часы практических занятий);

• прием и защита лабораторных работ (во время проведения занятий);

• выполнение курсовых работ в рамках дисциплин (руководство, консультирование и защита курсовых работ в часы, предусмотренные учебным планом) и др.

Основными видами самостоятельной работы студентов без участия преподавателей являются:

• формирование и усвоение содержания конспекта лекций на базе рекомендованной лектором учебной литературы, включая информационные образовательные ресурсы (электронные учебники, электронные библиотеки и др.);

• самостоятельное изучение отдельных тем или вопросов по учебникам или учебным пособиям;

• написание рефератов, докладов;

• подготовка к семинарам и лабораторным работам;

• выполнение домашних заданий в виде решения отдельных задач, проведения типовых расчетов, расчетно-компьютерных и индивидуальных работ по отдельным разделам содержания дисциплин и др.

10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине (модулю), включая перечень программного обеспечения

В учебном процессе используется следующее программное обеспечение:

– ActivePython x64,
– Google Chrome,
– Ministep,
– MongoDB,
– MS Office,
– MS SQL Server и программы администрирования,
– MS Visio Professional,
– Python,
– Sql datamining,
– Visual studio,
– Модули: matplotlib, Mlxtend, PrefixSpan, NumPy, pandas, scikit-learn, SciPy, TensorFlow, языка Python,
– Модули: pymongo, PyMySQL, PyQT, SQL Alchemy языка Python,

11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине (модулю):

В учебном процессе используется следующее оборудование:

– Учебные аудитории для проведения: занятий лекционного типа, занятий семинарского типа, практических занятий, выполнения курсовых работ, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, укомплектованные специализированной мебелью и техническими средствами обучения,
– Мультимедийный класс