

Министерство науки и высшего образования Российской Федерации
ФГБОУ ВО «БАЙКАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»

УТВЕРЖДАЮ

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



30.06.2022г.

Рабочая программа дисциплины

Б1.У.9. Анализ больших данных

Направление подготовки: 09.03.03 Прикладная информатика
Направленность (профиль): Информационные системы и технологии в
управлении

Квалификация выпускника: бакалавр

Форма обучения: очная, заочная

	Очная ФО	Заочная ФО
Курс	3	3
Семестр	32	32
Лекции (час)	18	6
Практические (сем, лаб.) занятия (час)	36	10
Самостоятельная работа, включая подготовку к экзаменам и зачетам (час)	54	92
Курсовая работа (час)		
Всего часов	108	108
Зачет (семестр)	32	32
Экзамен (семестр)		

Иркутск 2022

Программа составлена в соответствии с ФГОС ВО по направлению 09.03.03
Прикладная информатика.

Автор В.В. Братищенко

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

Заведующий кафедрой А.В. Родионов

Дата актуализации рабочей программы: 30.06.2023

Дата актуализации рабочей программы: 28.06.2024

1. Цели изучения дисциплины

Целью изучения дисциплины является получение компетенций в области применения технологии больших данных.

Задачами изучения дисциплины являются

- знакомство с особенностями размещения и обработки больших данных,
- овладение методами применения структур хранения распределенных данных,
- изучение технологий обработки распределенных данных,
- изучение типовых моделей исследования больших данных.

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-4	Способен разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС

Структура компетенции

Компетенция	Формируемые ЗУНы
ПК-4 Способен разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	З. Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У. Уметь разрабатывать архитектуру, дизайна, информационное обеспечение и прототип ИС Н. Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС

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): "Основы алгоритмизации", "Программирование", "Теория вероятностей и математическая статистика", "Организация ЭВМ и систем", "Базы данных", "Операционные системы"

Дисциплины, использующие знания, умения, навыки, полученные при изучении данной: "Введение в нейронные сети и машинное обучение"

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 3 зач. ед., 108 часов.

Вид учебной работы	Количество часов (очная ФО)	Количество часов (заочная ФО)
Контактная(аудиторная) работа		
Лекции	18	6
Практические (сем, лаб.) занятия	36	10

Самостоятельная работа, включая подготовку к экзаменам и зачетам	54	92
Всего часов	108	108

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

Заочная форма обучения

№ п/п	Раздел и тема дисциплины	Семестр	Лекции	Семинар Лаборат. Практич.	Самостоят. раб.	В интерактивной форме	Формы текущего контроля успеваемости
1	Общие сведения о больших данных	32	1	1	8		Задание 2. Загрузка статистических данных с интернет-сайтов. Тест 1. Технология больших данных
2	Структуры распределенных данных	32	1	1	8		Задание 1. Сбор данных из разных источников, объединение в одну таблицу и обработка в PowerBI. Тест 2. Многомерный анализ данных
3	Технологии хранения и использования больших данных	32	1	2	16		Задание 3. Облачные технологии хранения данных. Задание 4. Конструирование хранилища данных. Тест 3. Исследование зависимостей
4	Распределенная обработка данных	32	1	2	24		Тест 4. Исследование зависимостей
5	Задачи и модели исследования зависимостей данных	32	0				Задание 5. Решение задачи классификации. Задание 6. Решение задачи кластеризации. Задание 7. Решение задачи регрессии. Задание 8. Сбор,

№ п/п	Раздел и тема дисциплины	Семестр	Лекции	Семинар Лаборат. Практич.	Самостоят. раб.	В интерактивной форме	Формы текущего контроля успеваемости
							предобработка данных. Решение задачи классификации текстов. Задание 9. Решение задачи ассоциации. Задание 10. Прогнозирование рядов. Тест 5. Обработка текстов
5.1	Введение в искусственный интеллект и основные методы машинного обучения для работы с табличными данными	32	1	0	8		
5.2	Обучение с подкреплением	32	0,5	2	12		
5.3	Системы глубокого обучения	32	0,5	2	16		
	ИТОГО		6	10	92		

Очная форма обучения

№ п/п	Раздел и тема дисциплины	Семестр	Лекции	Семинар Лаборат. Практич.	Самостоят. раб.	В интерактивной форме	Формы текущего контроля успеваемости
1	Общие сведения о больших данных	32	1	2	4		Тест 1. Технология больших данных. Тест 2. Общие сведения
2	Многомерный анализ (OLAP)	32	2	4	4		Тест 3. Многомерный анализ данных. Тест 4. Построение дашбордов
3	Использование Power Pivot	32	2	4	6		Задание 1. Объединение и анализ данных в Power Pivot. Тест 5. OLAP в Power Pivot MS Excel
4	Применение Power BI	32	2	6	6		Задание 2. Объединение и анализ данных в Power BI. Задание 3. Выбор данных в Интернете и анализ данных в

№ п/п	Раздел и тема дисциплины	Семестр	Лекции	Семинар Лаборат. Практич.	Само- стоят. раб.	В интера- ктивной форме	Формы текущего контроля успеваемости
							Power BI. Тест 6. OLAP-анализ с использованием Power BI
5	Облачные технологии анализа на примере Yandex DataLens	32	2	4	6		Задание 4. Облачные технологии многомерного анализа на примере Yandex Datalens. Тест 7. OLAP-анализ с использованием Yandex Datalens
6	ABC-XYZ анализ	32	1	4	6		Задание 5. Yandex Datalens ABC-XYZ анализ. Тест 8. ABC-XYZ анализ
7	Распределенное хранение и обработка данных	32	1	4	4		Задание 6. Распределенная обработка в Питоне
8	Задачи и модели исследования зависимостей данных	32		0			Задание 7. Исследование и трансформация данных. Задание 8. Решение задачи классификации. Задание 9. Решение задачи регрессии. Задание 10. Решение задачи кластеризации. Задание 11. Прогнозирование рядов. Задание 12. Сбор, предобработка данных. Решение задачи классификации текстов. Задание 13. Решение задачи ассоциации. Тест 9. Трансформация данных. Тест 10. Задачи и модели

№ п/п	Раздел и тема дисциплины	Семестр	Лекции	Семинар Лаборат. Практич.	Самостоят. раб.	В интерактивной форме	Формы текущего контроля успеваемости
							Data Mining
8.1	Введение в искусственный интеллект и основные методы машинного обучения для работы с табличными данными	32	1	0	4		
8.2	Трансформация данных	32	2	2	4		
8.3	Задачи регрессии и классификации данных	32	2	4	6		
8.4	Задача кластеризации данных	32	2	2	4		
	ИТОГО		18	36	54		

5.2. Лекционные занятия, их содержание

№ п/п	Наименование разделов и тем	Содержание
1	Общие сведения о больших данных	Области применения больших данных. Структурирование больших данных. Технологии анализа больших данных. Этапы анализа: выбор целей и задач анализа, сбор данных, подготовка данных, предварительное исследование данных, выбор задач и моделей исследования зависимостей данных, развертывание и внедрение
2	OLAP-технология анализа	Структуры данных OLAP, показатели и измерения, операции OLAP-анализа
3	Дизайн дашбордов	Общие сведения. Назначение дашборда. Показатели и измерения. Визуальные компоненты дашборда. Объединение визуальных компонент в дашборд.
4	Применение Power Pivot для анализа данных	Создание модели данных. Вычисление полей и показателей. Построение сводных таблиц и диаграмм
5	Анализ данных с применением PowerBI	Определение источников данных, загрузка и связывание таблиц, вычисление полей и показателей, использование визуальных компонентов
6	Облачные технологии анализа	Применение Yandex DataLens для определения источников, наборов данных, вычислений, визуальных элементов для построения аналитических панелей - дашбордов
7	ABC-XYZ анализ	Выполнение ABC-XYZ анализа в PowerBI и Datalens
8	Распределенная обработка данных	MapReduce – модель распределенной обработки данных, предложенная компанией Google для обработки больших объемов данных на компьютерных кластерах Hadoop Distributed File System (HDFS) – распределённая файловая система, позволяющая хранить информацию практически неограниченного объёма. Hadoop YARN – фреймворк для управления ресурсами кластера и менеджмента задач, в том числе включает фреймворк MapReduce
9	Исследование	Основные задачи систем искусственного интеллекта.

№ п/п	Наименование разделов и тем	Содержание
	зависимостей. Обзор	Классификация, кластеризация, регрессия. Типы машинного обучения: с учителем, без учителя, с частичным привлечением учителя, обучение с подкреплением. Классификация на примере алгоритма k-ближайших соседей (kNN) Метрики оценки классификации: полнота, точность, F1, ROC, AUC. Обучающая и тестовая выборка. Работа с категориальными признаками. Перцептрон, логистическая регрессия, полносвязные нейронные сети. Алгоритмы, основанные на применении решающих деревьев. Ансамбли решающих деревьев: случайный лес, градиентный бустинг. Линейные и нелинейные модели регрессии. Кластеризация. k-means, DBSCAN, агломеративная кластеризация. Метрики оценки кластеризации.
10	Преобразование и очистка данных	Предварительная обработка данных является важным этапом аналитики. Собранные в одну таблицу данные из разных источников могут отличаться по форматам и классификаторам, содержать ошибки и пропуски, измеряться в разных единицах и противоречить друг другу. Они могут быть излишне детальными или наоборот содержать агрегированные значения. Предварительную обработку рассматривают как отдельный этап анализа – ETL – Extract, Transform, Load — «извлечение, преобразование, загрузка».
11	Задачи классификации, регрессии и кластеризации	Решение задачи классификации – отнесения объекта (например, заемщика) по его атрибутам (данные заемщика) к некоторому классу (надежный или ненадежный). Модель классификации настраивается по данным наблюдений атрибутов и классов объектов (кредитная история). Регрессионные модели для определения средних значений выходных переменных по значениям входных переменных. Задача кластеризации – объединение объектов в кластеры по степени близости характеристик.

5.3. Семинарские, практические, лабораторные занятия, их содержание

№ раздела и темы	Содержание и формы проведения
3	Разработка модели данных, вычислений и аналитической панели в Power Pivot. Загрузка и соединение таблиц, определение вычисляемых полей и показателей, задание иерархий, построение аналитической панели, включающей сводную таблицу, диаграммы, срезы
4	Разработка модели данных, вычислений и аналитической панели в PowerBI. Загрузка и соединение таблиц, определение вычисляемых полей и показателей, задание иерархий, построение аналитической панели, включающей сводную таблицу, диаграммы, срезы
5	Разработка модели данных, вычислений и аналитической панели в Datalens. Загрузка и соединение таблиц, определение вычисляемых полей и показателей, задание иерархий, построение аналитической панели, включающей сводную таблицу, диаграммы, срезы
6	ABC-XYZ анализ. Создание визуальных элементов и показателей для ABC-

№ раздела и темы	Содержание и формы проведения
	XYZ анализа в среде Datalens
7	Распределенная обработка данных. Применение технологии MapReduce для решения задач обработки распределенных данных. Инструментальные среды проектирования распределенной обработкой
8	Применение задач и моделей исследования зависимостей. Исходные данные исследования зависимостей. Предварительная обработка исходных данных. Задачи и модели классификации. Использование и сравнение алгоритмов классификации: kNN, решающие деревья и их ансамбли, логистическая регрессия. Задачи и модели кластеризации. Задачи и модели ассоциаций. Задачи и модели анализа текстов

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.п, У.1...У.п, Н.1...Н.п))	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. Общие сведения о больших данных	ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	Тест 1. Технология больших данных	Доля правильных ответов (3)
2		ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	Тест 2. Общие сведения	Доля правильных ответов (3)
3	2. Многомерный анализ (OLAP)	ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	Тест 3. Многомерный анализ данных	Доля правильных ответов (3)
4		ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	Тест 4. Построение дашбордов	Доля правильных ответов (3)
5	3. Использование Power Pivot	ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	Задание 1. Объединение и анализ данных в Power Pivot	Корректность запросов. Соответствие анализа постановке задачи (7)

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: (З.1...З.п, У.1...У.п, Н.1...Н.п)	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)
			У. Уметь разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС Н. Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС		
6		ПК-4	З. Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	Тест 5. OLAP в Power Pivot MS Excel	Доля правильных ответов (3)
7	4. Применение Power BI	ПК-4	З. Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У. Уметь разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС Н. Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС	Задание 2. Объединение и анализ данных в Power BI	Полнота представленных данных (7)
8		ПК-4	З. Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У. Уметь разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС Н. Владеть навыками разработки архитектуры, дизайна, информационного	Задание 3. Выбор данных в Интернете и анализ данных в Power BI	Корректность создания БД (7)

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: (З.1...З.п, У.1...У.п, Н.1...Н.п))	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)
			обеспечения и прототипа ИС		
9		ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	Тест 6. OLAP-анализ с использованием Power BI	Доля правильных ответов (3)
10	5. Облачные технологии анализа на примере Yandex DataLens	ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У.Уметь разрабатывать архитектуру, дизайна, информационное обеспечение и прототип ИС Н.Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС	Задание 4. Облачные технологии многомерного анализа на примере Yandex Datalens	Полнота описания процесса показателями и призаказами (7)
11		ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	Тест 7. OLAP-анализ с использованием Yandex Datalens	Доля правильных ответов (3)
12	6. ABC-XYZ анализ	ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У.Уметь разрабатывать архитектуру, дизайна, информационное обеспечение и прототип ИС Н.Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС	Задание 5. Yandex Datalens ABC-XYZ анализ	Корректность (7)
13		ПК-4	З.Знать, как разрабатывать	Тест 8. ABC-XYZ анализ	Доля правильных ответов (3)

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: (З.1...З.п, У.1...У.п, Н.1...Н.п)	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)
			архитектуру, дизайн, информационное обеспечение и прототип ИС		
14	7. Распределенное хранение и обработка данных	ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У.Уметь разрабатывать архитектуру, дизайна, информационное обеспечение и прототип ИС Н.Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС	Задание 6. Распределенная обработка в Питоне	Корректность программы (7)
15	8. Задачи и модели исследования зависимостей данных	ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У.Уметь разрабатывать архитектуру, дизайна, информационное обеспечение и прототип ИС Н.Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС	Задание 7. Исследование и трансформация данных	Полнота преобразований (7)
16		ПК-4	З.Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У.Уметь разрабатывать архитектуру, дизайна, информационное обеспечение и прототип ИС	Задание 8. Решение задачи классификации	Корректность применения моделей (7)

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: (З.1...З.п, У.1...У.п, Н.1...Н.п)	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)
			Н. Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС		
17		ПК-4	З. Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У. Уметь разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС Н. Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС	Задание 9. Решение задачи регрессии	Корректность моделей и интерпретации результатов (7)
18		ПК-4	З. Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У. Уметь разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС Н. Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС	Задание 10. Решение задачи кластеризации	Корректность применения моделей (7)
19		ПК-4	З. Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У. Уметь разрабатывать архитектуру, дизайн, информационное	Задание 11. Прогнозирование рядов	Обоснованность применения моделей (0)

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: З.1...З.п, У.1...У.п, Н.1...Н.п)	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)
			обеспечение и прототип ИС Н. Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС		
20		ПК-4	З. Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У. Уметь разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС Н. Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС	Задание 12. Сбор, предобработка данных. Решение задачи классификации текстов	Корректность применения моделей (0)
21		ПК-4	З. Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС У. Уметь разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС Н. Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС	Задание 13. Решение задачи ассоциации	Корректность применения методов (0)
22		ПК-4	З. Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС	Тест 9. Трансформация данных	Доля правильных ответов (3)
23		ПК-4	З. Знать, как разрабатывать	Тест 10. Задачи и модели Data Mining	Доля правильных ответов (3)

№ п/п	Этапы формирования компетенций (Тема из рабочей программы дисциплины)	Перечень формируемых компетенций по ФГОС ВО	(ЗУНы: (З.1...З.п, У.1...У.п, Н.1...Н.п)	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)
			архитектуру, дизайн, информационное обеспечение и прототип ИС		
				Итого	100

6.2. Промежуточный контроль (зачет, экзамен)

Рабочим учебным планом предусмотрен Зачет в семестре 32.

ВОПРОСЫ ДЛЯ ПРОВЕРКИ ЗНАНИЙ:

1-й вопрос билета (30 баллов), вид вопроса: Тест/проверка знаний. Критерий: Доля правильных ответов.

Компетенция: ПК-4 Способен разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС

Знание: Знать, как разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС

1. Базы данных NoSQL. Особенности, классификация
2. Документо-ориентированные базы данных. Возможности СУБД MongoDB
3. Понятие Больших данных. Особенности сбора, хранения, обработки и анализа Больших данных
4. Работа с документо-ориентированными БД.
5. Распределенная обработка и модели для решения задач анализа текстов
6. Распределенная обработка и модели для решения задач кластеризации
7. Система Apache Hadoop
8. Средства построения распределенных информационных систем
9. Структурированные данные и неструктурированные данные. Данные на естественном языке. Машинные данные. Графовые, или сетевые, данные. Аудио, видео и графика. Поточковые данные
10. Технология Map-Reduce
11. Требования к распределенным информационным системам

ТИПОВЫЕ ЗАДАНИЯ ДЛЯ ПРОВЕРКИ УМЕНИЙ:

2-й вопрос билета (30 баллов), вид вопроса: Задание на умение. Критерий: Корректность и эффективность решения.

Компетенция: ПК-4 Способен разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС

Умение: Уметь разрабатывать архитектуру, дизайна, информационное обеспечение и прототип ИС

- Задача № 1. Для предложенного датасета выполнить исследование и трансформацию данных
- Задача № 2. Для предложенного датасета построить дашборд в среде Power Pivot BI
- Задача № 3. Для предложенного датасета построить дашборд в среде Power Pivot MS Excel
- Задача № 4. Для предложенного датасета построить дашборд в среде Yandex Datalens
- Задача № 5. Для предложенного датасета решить задачу классификации
- Задача № 6. Для предложенного датасета решить задачу кластеризации
- Задача № 7. Для предложенного датасета решить задачу регрессии
- Задача № 8. Используя функции map и reduce выполнить имитацию распределенной обработки данных о продажах товаров для вычисления указанных показателей для измерения

ТИПОВЫЕ ЗАДАНИЯ ДЛЯ ПРОВЕРКИ НАВЫКОВ:

3-й вопрос билета (40 баллов), вид вопроса: Задание на навыки. Критерий: Корректность и эффективность решения.

Компетенция: ПК-4 Способен разрабатывать архитектуру, дизайн, информационное обеспечение и прототип ИС

Навык: Владеть навыками разработки архитектуры, дизайна, информационного обеспечения и прототипа ИС

Задание № 1. Для предложенной предметной области разработать решение задачи классификации, регрессии или кластеризации

Задание № 2. Для предложенной предметной области разработать схему анализа показателей

ОБРАЗЕЦ БИЛЕТА

Министерство науки и высшего образования Российской Федерации	Направление - 09.03.03 Прикладная информатика
Федеральное государственное бюджетное образовательное учреждение высшего образования	Профиль - Информационные системы и технологии в управлении
«БАЙКАЛЬСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ»	Кафедра математических методов и цифровых технологий
(ФГБОУ ВО «БГУ»)	Дисциплина - Анализ больших данных

БИЛЕТ № 1

1. Тест (30 баллов).
2. Для предложенного датасета решить задачу классификации (30 баллов).
3. Для предложенной предметной области разработать схему анализа показателей (40 баллов).

Составитель _____ В.В. Братищенко

Заведующий кафедрой _____ А.В. Родионов

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

а) основная литература:

1. Воронова Л.И., Воронов В.И. Big Data. Методы и средства анализа. учебное пособие. Электронный ресурс/ В.И. Воронов.- Москва: Московский технический университет связи и информатики, 2016.-33 с.
2. Чубукова И. А. Data Mining. учеб. пособие/ И. А. Чубукова.- М.: БИНОМ. Лаборатория знаний, 2006.-382 с.
3. [Билл Фрэнкс Революция в аналитике \[Электронный ресурс\] : как в эпоху Big Data улучшить ваш бизнес с помощью операционной аналитики / Фрэнкс Билл. — Электрон. текстовые данные. — М. : Альпина Паблишер, 2017. — 320 с. — 978-5-9614-5302-7. — Режим доступа: <http://www.iprbookshop.ru/58563.html>](http://www.iprbookshop.ru/58563.html)
4. [Воронов, В. И. Data Mining - технологии обработки больших данных : учебное пособие / В. И. Воронов, Л. И. Воронова, В. А. Усачев. — Москва : Московский технический университет связи и информатики, 2018. — 47 с. — ISBN 2227-8397. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : \[сайт\]. — URL: <http://www.iprbookshop.ru/81324.html>](http://www.iprbookshop.ru/81324.html)
5. [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>](http://www.iprbookshop.ru/61463.html)
6. [Чубукова И.А. Data Mining \[Электронный ресурс\] / И.А. Чубукова. — Электрон. текстовые данные. — М. : Интернет-Университет Информационных Технологий \(ИНТУИТ\), 2016. — 470 с. — 978-5-94774-819-2. — Режим доступа: <http://www.iprbookshop.ru/56315.html>](http://www.iprbookshop.ru/56315.html)

б) дополнительная литература:

1. Технологии анализа данных: Data Mining, Visual Mining, Text Mining, OLAP. учеб. пособие [для вузов]. рек. УМО вузов по унив. политехн. образованию. 2-е изд., перераб. и доп./ А. А. Барсегян [и др.].- СПб.: БХВ-Петербург, 2008.-375 с.
2. [Билл, Фрэнкс Революция в аналитике : как в эпоху Big Data улучшить ваш бизнес с помощью операционной аналитики / Фрэнкс Билл ; перевод И. Евстигнеева ; под редакцией В. Мылов. — Москва : Альпина Паблишер, 2017. — 320 с. — ISBN 978-5-9614-5302-7. — Текст : электронный // Электронно-библиотечная система IPR BOOKS : \[сайт\]. — URL: <http://www.iprbookshop.ru/58563.html>](http://www.iprbookshop.ru/58563.html)
3. [Кухаренко Б.Г. Интеллектуальные системы и технологии \[Электронный ресурс\] : учебное пособие / Б.Г. Кухаренко. — Электрон. текстовые данные. — М. : Московская государственная академия водного транспорта, 2015. — 116 с. — 2227-8397. — Режим доступа: <http://www.iprbookshop.ru/47933.html>](http://www.iprbookshop.ru/47933.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

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 информационных технологий, баз данных, интернет-технологий.

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:

- Python,
- Scilab,
- Модули: matplotlib, Mlxtend, PrefixSpan, NumPy, pandas, scikit-learn, SciPy, TensorFlow, языка Python,
- Модули: pymongo, PyMySQL, PyQt, SQL Alchemy языка Python,
- Надстройка интеллектуального анализа данных для MS Office,
- MongoDB,
- MS SQL Server и программы администрирования,
- ActivePython x64,
- Sql datamining,

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

В учебном процессе используется следующее оборудование:

- Помещения для самостоятельной работы, оснащенные компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечением доступа в электронную информационно-образовательную среду вуза,
- 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,
- Компьютерный класс,
- Наборы демонстрационного оборудования и учебно-наглядных пособий