МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ,
МИНИСТЕРСТВО НАУКИ, ВЫСШЕГО ОБРАЗОВАНИЯ И ИННОВАЦИЙ
КЫРГЫЗСКОЙ РЕСПУБЛИКИ

МОО ВО Кыргызско-Российский Славянский университет
имени первого Президента Российской Федерации Б.Н. Ельцина

УТВЕРЖДАЮ
Пр. декана факультета ГГФ
_____ Н. М. Комарнов
22.09.2025

Аналитика больших данных

рабочая программа дисциплины (модуля)

Закреплена за кафедрой	**Информационных и вычислительных технологий**
Учебный план	g090404_25_1пи_рпис.plx Направление подготовки 09.04.04 - РФ, 710400 - КР Программная инженерия Магистерская программа "Разработка программно-информационных систем"
Квалификация	**магистр**
Форма обучения	**очная**
Общая трудоемкость	**4 ЗЕТ**

Часов по учебному плану	144	Виды контроля в семестрах:
в том числе:		зачет с оценкой 3
аудиторные занятия	38	
самостоятельная работа	105,9	

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	**3 (2.1)**		Итого	
Недель	18			
Вид занятий	УП	РП	УП	РП
Лекции	16	16	16	16
Практические	22	22	22	22
Контактная работа в период теоретического обучения	0,1	0,1	0,1	0,1
В том числе инт.	8	8	8	8
В том числе в форме практ.подготовки	22	22	22	22
Итого ауд.	38	38	38	38
Контактная работа	38,1	38,1	38,1	38,1
Сам. работа	105,9	105,9	105,9	105,9
Итого	144	144	144	144

Программу составил(и):
к.т.н., доцент, Манжикова С.Ц.;к.т.н., доцент, Верзунов С.Н. _____

Рецензент(ы):
д.т.н., проф. кафедры ИВТ, Лыченко Н.М. _____

Рабочая программа дисциплины
Аналитика больших данных

разработана в соответствии с ФГОС 3++:
Федеральный государственный образовательный стандарт высшего образования - магистратура по направлению подготовки 09.04.04 Программная инженерия (приказ Минобрнауки России от 19.09.2017 г. № 932)

составлена на основании учебного плана:
Направление подготовки 09.04.04 - РФ, 710400 - КР Программная инженерия
Магистерская программа "Разработка программно-информационных систем"
утвержденного учёным советом вуза от 30.06.2025 протокол № 13

Рабочая программа одобрена на заседании кафедры
Информационных и вычислительных технологий

Протокол от 03.09.2025 г. № 1
Срок действия программы: 2025-2026 уч.г.
Зав. кафедрой д.т.н. проф. Лыченко Н.М.

ЛИСТ РЕГИСТРАЦИИ ИЗМЕНЕНИЙ
к рабочей программе дисциплины (модуля)

Аналитика больших данных

наименование дисциплины

№ п/п	Основание внесения изменения	Раздел/ пункт	Содержание изменения
1.	№357-П от 24 июня 2026- О преобразовании Естественно-технического факультета и Факультета архитектуры дизайна и строительства в Политехнический институт КРСУ	Титульный лист. Пункт Разработано	Заменить слова «Разработано Кафедра Информационных и вычислительных технологий» словами «Разработано: Высшая школа Цифровых технологий»
2.			
3.			
4.			
5.			

1. ЦЕЛИ ОСВОЕНИЯ ДИСЦИПЛИНЫ	
1.1	Курс имеет своей целью: формирование у студентов профессиональной компетенции в области разработки и использования систем обработки и анализа больших массивов данных.
1.2	Знания,полученные в результате освоения дисциплины, помогут при сборе и анализе огромных объемов структурированной или неструктурированной информации, при разработке моделей данных и получении новых знаний. Все это необходимо выпускнику, освоившему программу, для решения различных задач практической и научно-исследовательской деятельности.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ООП	
Цикл (раздел) ООП:	Б1.В.ДВ.01
2.1	**Требования к предварительной подготовке обучающегося:**
2.1.1	Машинное обучение и анализ данных
2.1.2	Проектирование распределенных информационных систем
2.1.3	Системы обработки экспериментальных данных
2.2	**Дисциплины и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:**
2.2.1	Подготовка к защите и защита выпускной квалификационной работы
2.2.2	Производственная практика по получению профессиональных умений и опыта профессиональной деятельности
2.2.3	Преддипломная практика

3. КОМПЕТЕНЦИИ ОБУЧАЮЩЕГОСЯ, ФОРМИРУЕМЫЕ В РЕЗУЛЬТАТЕ ОСВОЕНИЯ ДИСЦИПЛИНЫ (МОДУЛЯ)	
ПК-1: Владение навыками программной реализации систем с параллельной обработкой данных и высокопроизводительных систем	
Знать:	
Уровень 1	- методы проектирования систем с параллельной обработкой данных, высокопроизводительных систем и их компонент - основные элементы архитектуры высокопроизводительных вычислительных систем
Уровень 2	- основные технологические этапы разработки параллельных программ; - методы декомпозиции последовательных алгоритмов; - средства, позволяющие избежать случаев низкой эффективности распараллеливания.
Уметь:	
Уровень 1	- создавать алгоритмы параллельных вычислений - организовать взаимодействие процессов (потоков) и синхронизацию доступа к разделяемым ресурсам с помощью современных методов
Уровень 2	- оценивать сложность используемых алгоритмов и выбирать оптимальные алгоритмы для современных программ; - анализировать последовательные программы для выявления возможности их распараллеливания; - оценивать эффективность работы распараллеленных программ.
Владеть:	
Уровень 1	- средствами параллельного программирования; - способами проверки правильности работы программных комплексов с параллельной обработкой данных
Уровень 2	- приемами распараллеливания алгоритмов и программ; - средствами разработки приложений, обеспечивающими проведение параллельного вычислительного эксперимента.

В результате освоения дисциплины обучающийся должен

3.1	**Знать:**
3.1.1	основные характеристики больших данных, знать основные технологии, применяемые для хранения и поиска больших данных.
3.2	**Уметь:**
3.2.1	применять методы анализа больших данных, уметь реализовывать приложения для аналитики больших данных.
3.3	**Владеть:**
3.3.1	навыками проведения задач анализа больших данных с использованием аналитических платформ.

4. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Компетен-ции	Литература	Инте ракт.	Пр. подг.	Примечание
	Раздел 1. Введение в большие данные							

1.1	Определение термина большие данные и базовая модель. Применение больших данных. Роль больших данных в национальной экономике /Лек/	3	2		Л1.1Л2.1 Л2.2Л3.1 Э2			
1.2	Загрузка данных из различных хранилищ. Форматы хранения данных JSON, CSV, XML /Пр/	3	4		Л1.1Л2.2 Л2.1Л3.1 Э2		4	
1.3	Использование статистических пакетов для интеллектуального анализа данных /Ср/	3	12		Л1.1Л2.1 Л2.2Л3.1 Э2			
	Раздел 2. Жизненный цикл проекта по аналитике больших данных							
2.1	Основные этапы жизненного цикла проекта /Лек/	3	4		Л1.1Л2.1 Л2.2Л3.1 Э2 Э3			
2.2	Визуализация данных /Пр/	3	4		Л1.1Л2.1 Л2.2Л3.1 Э1		4	
2.3	Методология KDD. Задачи предобработки данных. Технология ETL. /Ср/	3	26		Л1.1Л2.1 Л2.2Л3.1 Э1			
	Раздел 3. Основные техники работы с большими данными							
3.1	Сбор и консолидация данных, визуализация данных, язык R для аналитики, библиотека CRAN, работа с СУБД /Лек/	3	4		Л1.1Л2.1 Л2.2Л3.1 Э1			
3.2	Hadoop, HDFS, Map/Reduce, YARN, Storm, Apache Spark /Лек/	3	2		Л1.1Л2.1 Л2.2Л3.1 Э1	2		"лекция-провокация"
3.3	Алгоритмы обработки больших данных /Лек/	3	2		Л1.1Л2.1 Л2.2Л3.1 Э1	2		"лекция-провокация"
3.4	Алгоритм кластеризации к-средних /Пр/	3	2		Л1.1Л2.1 Л2.2Л3.1 Э1		2	
3.5	Алгоритмы поиска ассоциативных правил /Пр/	3	6		Л1.1Л2.1 Л2.2Л3.1 Э1	2	6	работа в парах
3.6	Основы работы с платформой Apache Hadoop /Пр/	3	2		Л1.1Л2.1 Л2.2Л3.1 Э1	2	2	работа в группах
3.7	Профили кластеров. Взаимосвязь кластерного и регрессионного анализа. Использование пакета Deductor для решения задач кластерного анализа /Ср/	3	10		Л1.1Л2.1 Л2.2Л3.1 Э4			
3.8	Алгоритмы построения ассоциативных правил. Рекомендации по генерации правил. Алгоритм Apriori. /Ср/	3	12		Л1.1Л2.1 Л2.2Л3.1 Э4			
3.9	Примеры алгоритмов построения деревьев решений. Использование статистических пакетов Deductor, Statistica, Excel для построения деревьев решений /Ср/	3	10		Л1.1Л2.1 Л2.2Л3.1 Э4			
	Раздел 4. Приложение больших данных							

4.1	Имплементация технологии Demand/Responce /Лек/	3	2		Л1.1Л2.1 Л2.2Л3.1 Э3			
4.2	Работа с платформой Demand/Response /Пр/	3	4		Л1.1Л2.1 Л2.2Л3.1 Э3		4	
4.3	Развертывание VPP – виртуальной электростанции /КрТО/	3	0,1		Л1.1Л2.1 Л2.2Л3.1 Э3			
4.4	/ЗачётСОц/	3	35,9		Л1.1Л2.1 Л2.2Л3.1			

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

5.1. Контрольные вопросы и задания

Вопросы для проверки уровня обученности ЗНАТЬ
Основные характеристики больших данных.
Роль больших данных в экономике.
Основные конструкции языка R.
HDFS - основы организации.
Архитектура Hadoop.
Исполнение Map/Reduce.
Исполнение программ в Hadoop.
Основы YARN.
Аналитика потоковых данных в платформе Storm.
Архитектура Apache Spark.
Алгоритмы классификации.
Алгоритмы кластеризации.
Нейронные сети как реализация алгоритмов машинного обучения.
Предиктивные алгоритмы.
Применение технологий больших данных для задач управления в реальном времени.
Понятие бизнес-аналитики. Классификация средств «бизнес-аналитики».
Этапы анализа данных. KDD.
Понятие бизнес-аналитики. Классификация средств «бизнес-аналитики».
Этапы анализа данных. KDD.
Понятие бизнес-аналитики. Классификация средств «бизнес-аналитики».
Этапы анализа данных. KDD.
Общая характеристика QlikView.
Общая характеристика DeductorAcademic.
Общая характеристика средств интеллектуального анализа SQLServer.

Вопросы для проверки уровня обученности УМЕТЬ
Консолидация данных.
Визуализация данных.
Системы поддержки принятия решений. Хранилища данных.
Размерностные модели. OLAP-куб. Таблица размерностей. Таблица фактов.
Сравнительный анализ OLAP и OLTP-систем.
Консолидация данных в R.
Визуализация в R.
Организация данных в Apache Spark.
Обработка данных в GraphX.
Data Mining. Сдерстваобработки Data Mining.
Элементы математической статистики. Описательная статистика. Операции агрегирования данных.
Графические средства анализа. Диаграмма рассеяния. Гистограмма.
Начальные этапы KDD. ETL. Средства очистки и трансформации данных.
Классфикация метод предобработки и очистки данных.
Методы борьбы с аномалиями. Ящечная диаграмма.
Общая характеристика задач кластерного анализа.
Метрики кластерного анализа.
Методы определения близости между кластерами.
Иерархическая кластеризация. Дендограмма.
Метод к-средних.
Определение регрессионной модели. Логистическая регрессионная модель.

Вопросы для проверки уровня обученности ВЛАДЕТЬ
Ассоциативные правила. Свойства антимонотонности.
Метрики построения ассоциативных правил.

Алгоритм построения ассоциативных правил a’priori.
Общая характеристика деревьев решений.
Оценка качества классификации. Задачи классификации. ROC-кривая. Таблица сопряженности.
Использование логистической модели для классификации.
Нейронные сети. Перцептрон. Радиальные базисные сети.
Использование карты Кохоннена для решения задач классификации.

5.2. Темы курсовых работ (проектов)

не предусмотрены.

5.3. Фонд оценочных средств

Тесты представлены в Приложении 2.
Практические работы представлены в Приложении 3.

5.4. Перечень видов оценочных средств

Тесты, практические работы. Виды шкал оценивания представлены в Приложении 1.

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

6.1. Рекомендуемая литература

6.1.1. Основная литература

	Авторы, составители	Заглавие	Издательство, год
Л1.1	Федин Ф.О.	Анализ данных. Часть 2. Инструменты Data Mining : учебное пособие	М.: Московский городской педагогический университет 2014

6.1.2. Дополнительная литература

	Авторы, составители	Заглавие	Издательство, год
Л2.1	Айзек М.П., Серогодский В.В., Финков М.В., Прокди Р.Г.	Вычисления, графики и анализ данных в Excel 2010: Самоучитель	Наука и Техника, Санкт-Петербург, 2013
Л2.2	Федин Ф.О.	Анализ данных. Часть 1. Подготовка данных к анализу : учебное пособие	М.: Московский городской педагогический университет 2012

6.1.3. Методические разработки

	Авторы, составители	Заглавие	Издательство, год
Л3.1	сост. Истомина А.П.	Анализ данных качественных исследований: практикум	Ставрополь: Северо-Кавказский федеральный университет 2016

6.2. Перечень ресурсов информационно-телекоммуникационной сети "Интернет"

Э1	Пальмов С.В. Интеллектуальный анализ данных [Электронный ресурс] : учебное пособие / С.В. Пальмов. — Электрон. текстовые данные. — Самара: Поволжский государственный университет телекоммуникаций и информатики, 2017. — 127 c. — 2227-8397. — Режим доступа: http://www.iprbookshop.ru/75376.html	http://www.iprbookshop.ru/75376.html
Э2	Виктор Майер-Шенбергер Большие данные. Революция, которая изменит то, как мы живем, работаем и мыслим [Электронный ресурс] / Майер-Шенбергер Виктор, Кукьер Кеннет. — Электрон. текстовые данные. — М. : Манн, Иванов и Фербер, 2014. — 220 c. — 978-5-91657-936-9. — Режим доступа: http://www.iprbookshop.ru/39150.html	http://www.iprbookshop.ru/3
Э3	Адлер Ю.П. Статистическое управление процессами. «Большие данные» [Электронный ресурс] : учебное пособие / Ю.П. Адлер, Е.А. Черных. — Электрон. текстовые данные. — М. : Издательский Дом МИСиС, 2016. — 52 c. — 978-5-87623-969-3. — Режим доступа: http://www.iprbookshop.ru/64199.html	http://www.iprbookshop.ru/64199.html
Э4	Билл Фрэнкс Укрощение больших данных [Электронный ресурс] : как извлекать знания из массивов информации с помощью глубокой аналитики / Фрэнкс Билл. — Электрон. текстовые данные. — М. : Манн, Иванов и Фербер, 2014. — 340 c. — 978-5-00057-146-0. — Режим доступа: http://www.iprbookshop.ru/39433.html	http://www.iprbookshop.ru/39433.html

6.3. Перечень информационных и образовательных технологий

6.3.1 Компетентностно-ориентированные образовательные технологии

6.3.1.1	При проведении лекций используются интерактивные формы обучения (используются технологии типа «лекция-провокация», т.е. в процессе лекции делается преднамеренная ошибка с последующим опросом студентов на следующей лекции и организацией диалога «преподаватель-студент», «студент-студент» с целью выявления ошибки и установления истины). Практические занятия проводятся в интерактивной форме, в группах (используются технологии бригадного выполнения практической работы). В процессе их выполнения функциональные обязанности студентов разделены. Типичная бригада – 2 студента, один из которых изучает и готовит базу исходных данных, второй - возможные для использования модели, затем исходная база обрабатывается и оформляется отчет.
6.3.2 Перечень информационных справочных систем и программного обеспечения	
6.3.2.1	MS Windows 10;
6.3.2.2	MS Office 2013;
6.3.2.3	MathWorks Matlab 2012a;
6.3.2.4	MS Visual Studio;
6.3.2.5	Служба поддержики Microsoft https://support.microsoft.com/ru-ru;
6.3.2.6	Документация МАТЛАБ https://support.microsoft.com/ru-ru.

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ (МОДУЛЯ)

7.1	Дисциплина «Информатика» должна быть обеспечена необходимым аппаратным (персональные компьютеры) и программным обеспечением (операционная система ОС Windows и пакет прикладных программ MS Office, MatLab, MS Visual Studio). Для проведения лекционных занятий используется мультимедийное оборудование (проектор, экран, интерактивная доска).

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ ПО ОСВОЕНИЮ ДИСЦИПЛИНЫ (МОДУЛЯ)

Технологическая карта дисциплины представлена в Приложении 4.
Рекомендации по организации самостоятельной работы студента
1. Советы по планированию и организации времени, необходимого для изучения дисциплины. Рекомендуется следующим образом организовать время, необходимое для изучения дисциплины:
Изучение конспекта лекции в тот же день, после лекции – 10-15 минут;
Изучение конспекта лекции за день перед следующей лекцией – 10-15 минут;
Изучение теоретического материала по учебному пособию и конспекту – 1 час в неделю;
Подготовка к практическому занятию – 2-3 час.
Всего в неделю – 4 часа.
2. Описание последовательности действий студента
Для понимания материала и качественного его усвоения рекомендуется следующая последовательность действий.
После прослушивания лекции и окончания учебных занятий, при подготовке к занятиям следующего дня, нужно сначала просмотреть и обдумать текст лекции, прослушанной сегодня (10-15 минут).
При подготовке к лекции следующего дня, нужно просмотреть текст предыдущей лекции, подумать о том, какая может быть тема следующей лекции (10-15 минут).
В течение недели выбрать время (2-3часа) для работы с рекомендуемыми электронными учебными пособиями.
При подготовке к практическим занятиям следующего дня, необходимо сначала прочитать основные понятия и подходы по теме домашнего задания. При выполнении упражнения или задачи нужно сначала понять, что требуется в задаче, какой теоретический материал нужно использовать, наметить план решения задачи.
3. Рекомендации по использованию материалов учебно-методического комплекса. Все рекомендуемые учебные пособия размещены на серверах компьютерных классов в сетевой папке.