

Документ подписан простой электронной подписью
Информация о владельце:

ФИО: Исаев Игорь Магомедович

Должность: Проректор по учебной работе

Дата подписания: 21.09.2023 10:09:34

Уникальный идентификатор документа:

d7a26b9e8ca85e98ec3de2eb454b4659d061f249

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**Федеральное государственное автономное образовательное учреждение
высшего образования**

«Национальный исследовательский технологический университет «МИСИС»

Рабочая программа дисциплины (модуля)

Методы статистического анализа данных

Закреплена за подразделением

Кафедра АСУ

Направление подготовки

09.03.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА

Профиль

Квалификация

Бакалавр

Форма обучения

очная

Общая трудоемкость

4 ЗЕТ

Часов по учебному плану

144

Формы контроля в семестрах:

в том числе:

зачет с оценкой 5

аудиторные занятия

68

самостоятельная работа

76

Распределение часов дисциплины по семестрам

Семестр (<Курс>.<Семестр на курсе>)	5 (3.1)		Итого	
	Неделя 18			
Вид занятий	УП	РП	УП	РП
Лекции	34	17	34	17
Лабораторные	17	34	17	34
Практические	17		17	
Итого ауд.	68	51	68	51
Контактная работа	68	51	68	51
Сам. работа	76	93	76	93
Итого	144	144	144	144

Программу составил(и):

Рабочая программа

Методы статистического анализа данных

Разработана в соответствии с ОС ВО:

Самостоятельно устанавливаемый образовательный стандарт высшего образования - бакалавриат Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский технологический университет «МИСИС» по направлению подготовки 09.03.01 ИНФОРМАТИКА И ВЫЧИСЛИТЕЛЬНАЯ ТЕХНИКА (приказ от 02.04.2021 г. № 119 о.в.)

Составлена на основании учебного плана:

09.03.01 Информатика и вычислительная техника, 09.03.01-БИВТ-23.plx , утвержденного Ученым советом НИТУ МИСИС в составе соответствующей ОПОП ВО 22.06.2023, протокол № 5-23

Утверждена в составе ОПОП ВО:

09.03.01 Информатика и вычислительная техника, , утвержденной Ученым советом НИТУ МИСИС 22.06.2023, протокол № 5-23

Рабочая программа одобрена на заседании

Кафедра АСУ

Протокол от 05.07.2022 г., №10

Руководитель подразделения Темкин Игорь Олегович, д.т.н., доцент

1. ЦЕЛИ ОСВОЕНИЯ

1.1	научиться применять методы анализа данных для количественной оценки социально-экономической ситуации на микро, макро и глобальном уровнях; проведения статистического анализа (анализ данных (Data mining), интеллектуальная обработка информации, машинное обучение, искусственный интеллект, статистические и кибернетические методы анализа данных).
-----	---

2. МЕСТО В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ

Блок ОП:		Б1.В.ДВ.02
2.1	Требования к предварительной подготовке обучающегося:	
2.1.1	Алгоритмы дискретной математики	
2.1.2	Математика	
2.1.3	Комбинаторика и теория графов	
2.1.4	Технологии программирования	
2.1.5	Физика	
2.1.6	Компьютерная и инженерная графика	
2.1.7	Объектно-ориентированное программирование	
2.1.8	Основы дискретной математики	
2.1.9	Вычислительные машины, сети и системы	
2.1.10	Программирование и алгоритмизация	
2.2	Дисциплины (модули) и практики, для которых освоение данной дисциплины (модуля) необходимо как предшествующее:	
2.2.1	Введение в IoT системы	
2.2.2	Веб-аналитика	
2.2.3	Интеллектуальный анализ данных	
2.2.4	Математические модели социально-экономических систем	
2.2.5	Методология разработки программного обеспечения	
2.2.6	Методы оптимизации	
2.2.7	Научно-исследовательская работа	
2.2.8	Научно-исследовательская работа	
2.2.9	Научно-исследовательская работа	
2.2.10	Научно-исследовательская работа	
2.2.11	Научно-исследовательская работа	
2.2.12	Основы разработки цифровых платформ управления	
2.2.13	Системы реального времени	
2.2.14	Современные инструментальные средства анализа данных	
2.2.15	Современные инструменты управления проектами	
2.2.16	Технологии решения задач машинного обучения	
2.2.17	Инструментальные средства обработки изображений	
2.2.18	Нейросетевые технологии в прикладных задачах управления	
2.2.19	Оптимизационное моделирование сложных систем	
2.2.20	Программирование встраиваемых систем	
2.2.21	Технологии цифрового дублирования	
2.2.22	Управление проектами	
2.2.23	Подготовка к процедуре защиты и защита выпускной квалификационной работы	
2.2.24	Поиск решений в пространстве состояний	

3. РЕЗУЛЬТАТЫ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ, СООТНЕСЕННЫЕ С ФОРМИРУЕМЫМИ КОМПЕТЕНЦИЯМИ

ПК-6: Готовность применять системный подход и математические методы в формализации и решении задач управления
Знать:
ПК-6-31 системный подход и математические методы в формализации и решении задач управления (данные и знания, описательные статистики, корреляционный и регрессионный анализ, кластер-анализ, распознавание образов, деревья решений, нейронная сеть, метод опорных векторов).

ОПК-1: Способен применять естественнонаучные и общинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности
Знать:
ОПК-1-31 естественнонаучные и общинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности. статистические методы классификации и группировки, анализа взаимосвязей и динамики явлений и организации статистических работ.
ПК-6: Готовность применять системный подход и математические методы в формализации и решении задач управления
Уметь:
ПК-6-У1 применять системный подход и математические методы в формализации и решении задач управления (данные и знания, описательные статистики, корреляционный и регрессионный анализ, кластер-анализ, распознавание образов, деревья решений, нейронная сеть, метод опорных векторов).
ОПК-1: Способен применять естественнонаучные и общинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности
Уметь:
ОПК-1-У1 применять естественнонаучные и общинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности, статистические методы классификации и группировки, анализа взаимосвязей и динамики явлений и организации статистических работ.
ПК-6: Готовность применять системный подход и математические методы в формализации и решении задач управления
Владеть:
ПК-6-В1 методами системного подхода и математическими методами в формализации и решении задач управления, теорией вероятностей и математической статистикой, многомерными статистическими методами, эконометрикой, анализом временных рядов.
ОПК-1: Способен применять естественнонаучные и общинженерные знания, методы математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности
Владеть:
ОПК-1-В1 методами математического анализа и моделирования, теоретического и экспериментального исследования в профессиональной деятельности, статистическими методами классификации и группировки, анализа взаимосвязей и динамики явлений и организации статистических работ.

4. СТРУКТУРА И СОДЕРЖАНИЕ

Код занятия	Наименование разделов и тем /вид занятия/	Семестр / Курс	Часов	Формируемые индикаторы компетенций	Литература и эл. ресурсы	Примечание	КМ	Выполняемые работы
	Раздел 1. Введение в дисциплину. Понятие Data mining как мультидисциплинарной области.							
1.1	Виды данных. Задачи анализа данных: классификация, группировка, прогнозирование, нахождение ассоциаций и зависимостей, визуализация. /Лек/	5	2	ОПК-1-31 ОПК-1-У1 ПК -6-31			КМ1,К М2,КМ 3	
1.2	Основные разделы, на которых базируется анализ данных: статистика, базы данных и знаний, распознавание образов, искусственный интеллект, машинное обучение. /Лек/	5	2	ПК-6-У1			КМ2	

1.3	Классификация методов анализа данных. Этапы анализа данных: выявление закономерностей, прогнозирование, анализ исключений. Сферы применения анализа данных. /Лек/	5	2	ОПК-1-У1 ОПК-1-В1			КМ4	
1.4	Отличия от SQL и OLAP-технологий. Этапы анализа данных: выявление закономерностей, прогнозирование, анализ исключений. Сферы применения анализа данных: финансы и банковское дело, маркетинг, медицина, генетика, биоинформатика, интернет. /Ср/	5	30	ПК-6-31 ПК-6-У1			КМ5	
	Раздел 2. Статистические методы анализа данных.							
2.1	Проверка гипотез о вероятностной природе данных (стационарности, нормальности, независимости, однородности, оценивание параметров функции распределения). /Лек/	5	2	ОПК-1-У1 ОПК-1-В1			КМ1,К М2	
2.2	Выявление связей и закономерностей в данных (регрессионный анализ, корреляционный анализ). /Лек/	5	3	ОПК-1-У1 ОПК-1-В1			КМ5	
2.3	Основные методы многомерного статистического анализа (дискриминантный анализ, кластерный анализ, анализ главных компонент, факторный анализ) /Лаб/	5	6	ПК-6-31 ПК-6-У1			КМ4	
2.4	Динамические модели и прогноз на основе временных рядов. /Лаб/	5	6	ОПК-1-В1 ПК-6-31			КМ5,К М4	
2.5	Инструменты Data mining. Анализ данных в программных системах Excel, Statistica. /Ср/	5	30	ОПК-1-У1 ОПК-1-В1			КМ3	
	Раздел 3. Кибернетические методы анализа данных.							
3.1	Понятие о машинном обучении, методах искусственного интеллекта. /Лек/	5	2	ОПК-1-В1			КМ1,К М2	
3.2	Нейронные сети, их архитектура (однослойные, многослойные, с обратной связью). /Лек/	5	4	ОПК-1-В1			КМ5,К М4	
3.3	Эволюционные и генетические алгоритмы. /Лаб/	5	6	ПК-6-31 ПК-6-У1			КМ2	
	Раздел 4. Деревья решений.							

4.1	Методы построения деревьев решений. Критерии качества дерева решений (критерий Джини, энтропийный и регуляризирующий критерии). /Лаб/	5	6	ОПК-1-В1 ПК-6-31 ПК-6-У1			КМ5	
4.2	Основные операции с деревьями решений: ветвление, разрастание, редуцирование. Процедура кросс-проверки качества дерева. /Лаб/	5	4	ПК-6-31 ПК-6-У1			КМ4,КМ5	
4.3	Принятие решений на основе набора деревьев. Бустинг-метод. /Лаб/	5	6	ОПК-1-У1 ОПК-1-В1			КМ4,КМ5	
4.4	Методы анализа данных, основанные на использовании метрики: метод опорных векторов, метод ближайших соседей. /Ср/	5	33	ОПК-1-У1 ОПК-1-В1			КМ5	

5. ФОНД ОЦЕНОЧНЫХ МАТЕРИАЛОВ

5.1. Контрольные мероприятия (контрольная работа, тест, коллоквиум, экзамен и т.п), вопросы для самостоятельной подготовки

Код КМ	Контрольное мероприятие	Проверяемые индикаторы компетенций	Вопросы для подготовки
КМ1		ПК-6-31;ПК-6-У1;ПК-6-В1	<ol style="list-style-type: none"> 1. Основные сферы применения Data mining. 2. Основные вопросы теории распознавания образов. 3. Основные вопросы искусственного интеллекта. 4. Статистические методы анализа данных. 5. Основные методы выявления вероятностной природы данных. 6. Выявление связей и закономерностей в данных. 7. Основные методы анализа временных рядов и их применение при обработке данных. 8. Нейронные сети и их применение при анализе данных. 9. Метод МГУА и его применение для анализа данных. 10. Метод опорных векторов и его применение для анализа данных.
КМ2		ОПК-1-В1;ОПК-1-У1;ПК-6-В1	<ol style="list-style-type: none"> 1. Методы построения деревьев решений и их применение для анализа данных. 2. Методы кластер-анализа и их применение для обработки данных. 3. Бустинг-метод в анализе данных. 4. Применение корреляционно-регрессионного анализа. 5. Применение факторного анализа . 6. Применение методов анализа временных рядов. 7. Применение методов построения деревьев решений. 8. Анализ данных с использованием баз и хранилищ данных, SQL и OLAP технологии. 9. Инструменты анализа данных в системе Statistica. 10. Инструменты анализа данных в системе Excel.

КМ3		ПК-6-31;ПК-6-В1;ОПК-1-31	<ol style="list-style-type: none"> 1. Статистическое оценивание, понятие среднего, дисперсии, среднеквадратического отклонения, доверительного интервала. 2. Понятие о выборочном методе. Основные виды выборки. 3. Оценивание выборочных средних, суммарных значений, долей. 4. Характеристики случайных величин 5. Основные вероятностные распределения, используемые в статистическом анализе: нормальное распределение, распределение Хи-квадрат, Стьюдента и Фишера. 6. Проверка статистических гипотез, уровень значимости, Р-значение. 7. Понятия статистического показателя (характеристики, признака), наблюдения, таблицы данных. Основные виды показателей (количественные, порядковые и нечисловые). 8. Этапы построения моделей – анализ данных, подбор кривой, формализация описания кривой, оценивание параметров, проверка качества модели. 9. Оценивание параметров парной регрессионной модели методом наименьших квадратов. 10. Матричное представление множественной регрессионной модели. Оценивание параметров множественной
КМ4		ПК-6-У1;ОПК-1-В1;ОПК-1-У1	<ol style="list-style-type: none"> 1. Интервальные оценки параметров регрессионной модели. 2. Коэффициент детерминации и его свойства. 3. Скорректированный коэффициент детерминации и его свойства. 4. Нелинейные модели регрессии и их линеаризация. Коэффициент эластичности. 5. Проверка значимости регрессии с помощью t-критерия Стьюдента, F-критерия Фишера. 6. Регрессионный анализ при наличии нечисловых переменных. Модель с переменной структурой. Метод введения фиктивных переменных. 7. Корреляционный анализ статистических показателей. 8. Выборочный коэффициент линейной корреляции и его использование при изучении взаимосвязи показателей 9. Частный и множественный коэффициенты корреляции. 10. Корреляционный анализ в случае нескольких переменных. Корреляционная матрица. 11. Задача дискриминантного анализа. 12. Метод дендритов в кластер-анализе.
КМ5		ПК-6-31;ОПК-1-В1;ОПК-1-31	<ol style="list-style-type: none"> 1. Способы оценивания расстояния между группами (кластерами), близости внутри групп. 2. Метод К-средних в кластер-анализе. 3. Иерархические методы группировки. Понятие дендрограммы 4. Линейный дискриминант Фишера. 5. Задача дисперсионного анализа. Понятия факторов, уровней, таблицы данных для дисперсионного анализа. Однофакторный дисперсионный анализ. 6. Основные возможности пакета Excel для регрессионного, корреляционного, дисперсионного анализа. 7. Нахождение параметров тренда, исключение тренда. 8. Понятия тренда, трендовой модели временного ряда. Проверка ряда на наличие тренда. 9. Показатели временного ряда: темпы, коэффициенты роста, прироста. 10. Сезонная модель временного ряда. Оценивание сезонных эффектов. 11. Понятие лага (порядка модели), общая линейная авторегрессионная модель временного ряда.
<p>5.2. Перечень работ, выполняемых по дисциплине (модулю, практике, НИР) - эссе, рефераты, практические и расчетно-графические работы, курсовые работы, проекты и др.</p>			

5.3. Оценочные материалы, используемые для экзамена (описание билетов, тестов и т.п.)**5.4. Методика оценки освоения дисциплины (модуля, практики. НИР)****6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ****6.1. Рекомендуемая литература****6.3 Перечень программного обеспечения**

П.1	LMS Canvas
П.2	MS Teams
П.3	MATLAB
П.4	MATCAD
П.5	Statistica Neural Networks

6.4. Перечень информационных справочных систем и профессиональных баз данных**7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ**

Ауд.	Назначение	Оснащение
Л-826	Лаборатория	доска и маркеры, персональные компьютеры ОС Windows с администраторскими правами доступа, с проводными сетевыми платами, с СОМ-портами количеством не менее 6, сетевое коммуникационное оборудование CISCO: 6 коммутаторов и 6 маршрутизаторов, обжатые кабели витая пара прямые и кроссовые количеством не менее 12 каждый, консольные кабели количеством не менее 6
Л-829	Учебная аудитория	комплект учебной мебели на 15 рабочих мест, оборудованных компьютерами, рабочее место преподавателя с персональным компьютером, без доступа к ИТС «Интернет»
Л-830	Лекционная аудитория	комплект учебной мебели на 60 рабочих мест, рабочее место преподавателя с персональным компьютером, без доступа к ИТС «Интернет»
Л-831	Учебная аудитория	комплект учебной мебели на 15 рабочих мест, оборудованных компьютерами, рабочее место преподавателя с персональным компьютером, без доступа к ИТС «Интернет»

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ

При изучении дисциплины используется балльно-рейтинговая технология, которая позволяет реализовать непрерывную и комплексную систему оценивания учебных достижений студентов. Непрерывность означает, что текущие оценки не усредняются, а непрерывно складываются на всем протяжении при изучении дисциплины в семестре. Комплексность означает учет всех форм учебной и творческой работы студента в течение семестра. Балльно-рейтинговая технология, включает в себя два вида контроля: текущий контроль и промежуточная аттестация по дисциплине. Лекционные занятия проводятся в форме контактной работы со студентами и с применением дистанционных образовательных технологий. Практические занятия проводятся в форме контактной работы со студентами и с применением дистанционных образовательных технологий, в компьютерном классе либо в аудитории с мультимедийным оборудованием. Консультирование студентов в процессе изучения дисциплины организуется кафедрой и осуществляется преподавателем в форме контактной работы со студентами с применением дистанционных образовательных технологий. Консультирование может осуществляться как в режиме on-line, так и заочно в форме ответов на вопросы студентов, направляемых преподавателю посредством размещения их в разделе «Консультации» в структуре изучаемой дисциплины в электронной информационно-образовательной среде (ЭИОС) университета. Роль консультаций должна сводиться, в основном, к помощи в изучении дисциплины (модуля), выполнении контрольных работ. Текущий контроль (ТК)-основная часть балльно-рейтинговой технологии, основанная на поэтапном контроле усвоения студентом учебного материала, выполнении индивидуальных заданий. Форма контроля: тестовые оценки в ходе изучения дисциплины, оценки за выполнение индивидуальных заданий, контрольных работ. Основная цель ТК: своевременная оценка успеваемости студентов, побуждающая их работать равномерно, исключая малые загрузки или перегрузки в течение семестра. ТК осуществляется программными средствами ЭИОС в период самостоятельной работы студента по его готовности.

Оценивание учебной работы студента осуществляется в соответствии с критериями оценивания, определяемые балльно-рейтинговой системой (БРС) рабочей программы учебной дисциплины. По результатам ТК, при достаточной личной организованности и усердии, студенты имеют возможность получить оценку при промежуточной аттестации по итогам текущей успеваемости. Промежуточная аттестация (ПА)-это проверка оценочными средствами уровня учебных

достижений студентов по всей дисциплине за семестр. Тесты формируются соответствующими программными средствами случайным образом из банка тестовых заданий по учебной дисциплине. ПА осуществляется с применением дистанционных образовательных технологий. Цель ПА: проверка базовых знаний дисциплины и практических навыков, полученных при изучении модуля (дисциплины) и уровня сформированности компетенций.

Методические указания по организации процедуры тестирования

Итоговый контрольный тест доступен студенту только во время тестирования, согласно расписания занятий или в установленное деканатом время.

Студент информируется о результатах текущей успеваемости.

Студент получает информацию о текущей успеваемости, начислении бонусных баллов и допуске к процедуре итогового тестирования от преподавателя или в ЭИОС.

Производится идентификация личности студента.

Студентам, допущенным к промежуточной аттестации, открывается итоговый контрольный тест.

Тест закрывается студентом лично по завершении тестирования или автоматически по истечении времени тестирования.

Используемая литература

1. Дюк В. А., Самойленко А. П. Data Mining: учебный курс. Спб.: Питер, 2001.
2. Айвазян С.А. Мхитарян В.С. прикладная статистика и основы эконометрики: Учебник. М., ЮНИТИ, 1998.
3. Дубров А.М и др. Многомерные статистические методы для экономистов и менеджеров. М.: ФиС, 2000
4. Справочник по прикладной статистике / под ред. Ллойда, Лидермана. Т.2. – М.: Финансы и Статистика, 1990.
5. Тюрин Ю.Н., Макаров А.А. Статистический анализ данных на компьютере. Под ред. В.Э. Фигурнова. М., Финансы и статистика, 1998.
6. Киселев М., Соломатин Е. Средства добычи знаний в бизнесе и финансах // Открытые системы. 1997. № 4. С. 41-44.
7. А.Н. Горбань Методы нейроинформатики. КГТУ, Красноярск, 1998.
8. Бериков В.Б. Анализ статистических данных с использованием деревьев решений. Учебное пособие. – Новосибирск. Изд-во НГТУ, 2002.