

Министерство науки и высшего образования РФ
Федеральное государственное бюджетное образовательное учреждение
высшего образования
«Национальный исследовательский университет «МЭИ»

Направление подготовки/специальность: 01.04.02 Прикладная математика и информатика

Наименование образовательной программы: Математическое и программное обеспечение вычислительных машин и компьютерных сетей

Уровень образования: высшее образование - магистратура

Форма обучения: Очная

Рабочая программа дисциплины
МЕТОДЫ И СРЕДСТВА АНАЛИЗА ДАННЫХ


Блок:	Блок 1 «Дисциплины (модули)»
Часть образовательной программы:	Часть, формируемая участниками образовательных отношений
№ дисциплины по учебному плану:	Б1.Ч.04
Трудоемкость в зачетных единицах:	1 семестр - 4;
Часов (всего) по учебному плану:	144 часа
Лекции	1 семестр - 16 часов;
Практические занятия	1 семестр - 32 часа;
Лабораторные работы	не предусмотрено учебным планом
Консультации	1 семестр - 2 часа;
Самостоятельная работа	1 семестр - 93,5 часа;
в том числе на КП/КР	не предусмотрено учебным планом
Иная контактная работа	проводится в рамках часов аудиторных занятий
включая:	
Лабораторная работа	
Промежуточная аттестация:	
Экзамен	1 семестр - 0,5 часа;

Москва 2023

ПРОГРАММУ СОСТАВИЛ:

Преподаватель

(должность)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Михайлов И.С.
	Идентификатор	Ra29e5243-MikhailovIS-1df6126d

(подпись)


И.С. Михайлов

(расшифровка подписи)

СОГЛАСОВАНО:

Руководитель
образовательной
программы

(должность, ученая степень, ученое
звание)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Маран М.М.
	Идентификатор	R7be141f2-MaranMM-804b01e2


(подпись)

М.М. Маран

(расшифровка подписи)

Заведующий выпускающей
кафедры

(должность, ученая степень, ученое
звание)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Варшавский П.Р.
	Идентификатор	R9a563c96-VarshavskyPR-efb4bbd

(подпись)

П.Р.

Варшавский

(расшифровка подписи)

1. ЦЕЛИ И ЗАДАЧИ ОСВОЕНИЯ ДИСЦИПЛИНЫ

Цель освоения дисциплины: Изучение моделей и методов интеллектуального анализа данных и машинного обучения в задачах поиска информации, обработки и анализа данных, а также приобретение навыков исследователя данных и разработчика математических моделей, методов и алгоритмов анализа данных, изучение современных подходов к разработке программного обеспечения, соответствующих языковых и инструментальных средств.

Задачи дисциплины

- Освоение основных моделей и методов машинного обучения и анализа данных.;
- Приобретение навыков по реализации и применению указанных моделей и методов.;
- Приобретение навыков по применению программных средств, в которых реализованы указанные модели и методы.;
- Формирование устойчивых навыков анализа реальных данных с помощью изученных методов..

Формируемые у обучающегося **компетенции** и запланированные **результаты обучения** по дисциплине, соотнесенные с **индикаторами достижения компетенций**:

Код и наименование компетенции	Код и наименование индикатора достижения компетенции	Запланированные результаты обучения
ПК-2 Способен применять методы проектирования для обеспечения реализации результатов анализа	ИД-1 _{ПК-2} Формализует описания бизнес-процессов	знать: - методы системного анализа; - основные подходы к реализации методов искусственного интеллекта для анализа данных; - алгоритмы анализа больших объёмов информации. уметь: - реализовать методы системного анализа; - реализовать алгоритмы анализа больших объёмов данных; - современные платформы и языки программирования высокого уровня.
ПК-2 Способен применять методы проектирования для обеспечения реализации результатов анализа	ИД-2 _{ПК-2} Демонстрирует умение следить за изменениями требований	знать: - этапы жизненного цикла информационных систем; - подходы к тестированию информационных систем. уметь: - осуществлять поддержку разработки информационных систем; - выполнять доработку модулей информационных систем.

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ ВО

Дисциплина относится к основной профессиональной образовательной программе Математическое и программное обеспечение вычислительных машин и компьютерных сетей (далее – ОПОП), направления подготовки 01.04.02 Прикладная математика и информатика, уровень образования: высшее образование - магистратура.

Базируется на уровне среднего общего образования.

Результаты обучения, полученные при освоении дисциплины, необходимы при выполнении выпускной квалификационной работы.

3. СТРУКТУРА И СОДЕРЖАНИЕ ДИСЦИПЛИНЫ

3.1 Структура дисциплины

Общая трудоемкость дисциплины составляет 4 зачетных единицы, 144 часа.

№ п/п	Разделы/темы дисциплины/формы промежуточной аттестации	Всего часов на раздел	Семестр	Распределение трудоемкости раздела (в часах) по видам учебной работы										Содержание самостоятельной работы/ методические указания	
				Контактная работа							СР				
				Лек	Лаб	Пр	Консультация		ИКР		ПА	Работа в семестре	Подготовка к аттестации /контроль		
КПР	ГК	ИККП	ТК												
1	2	3	4	5	6	7	8	9	10	11	12	13	14	15	
1	Интеллектуальный анализ данных (Data Mining). Этапы исследования данных с помощью методов Data Mining. Типы закономерностей. Методы Data Mining.	18	1	2	-	6	-	-	-	-	-	10	-	<p><u>Самостоятельное изучение теоретического материала:</u> Изучение дополнительного материала по разделу "Интеллектуальный анализ данных (Data Mining). Этапы исследования данных с помощью методов Data Mining. Типы закономерностей. Методы Data Mining."</p> <p><u>Изучение материалов литературных источников:</u></p> <p>[1], п.1 [3], п.1 [4], п.2</p>	
1.1	Интеллектуальный анализ данных (Data Mining). Этапы исследования данных с помощью методов Data Mining. Типы закономерностей. Методы Data Mining.	18		2	-	6	-	-	-	-	-	10	-		
2	Задачи Data Mining. Задача кластеризации. Смесь нормальных распределений. EM-алгоритм. K-means и его модификации.	18		2	-	6	-	-	-	-	-	10	-		<p><u>Подготовка к лабораторной работе:</u> Для выполнения заданий по лабораторной работе "Реализация алгоритма K-means для решения задачи кластеризации" необходимо предварительно изучить тему и задачи выполнения лабораторной работы, а так же изучить вопросы вариантов обработки результатов по изученному в разделе "Задачи Data Mining. Задача кластеризации. Смесь нормальных распределений. EM-алгоритм. K-means и его модификации."</p>
2.1	Задачи Data Mining. Задача кластеризации. Смесь нормальных распределений. EM-алгоритм. K-means и его модификации.	18		2	-	6	-	-	-	-	-	10	-		

													<u>Изучение материалов литературных источников:</u> [4], п.3	
3	Байесовская классификация. Иерархическая классификация. Расстояние между кластерами. Алгоритм DBSCAN.	18	3	-	5	-	-	-	-	-	-	10	-	<u>Подготовка к лабораторной работе:</u> Для выполнения заданий по лабораторной работе "Реализация алгоритма DBSCAN для решения задачи кластеризации" необходимо предварительно изучить тему и задачи выполнения лабораторной работы, а так же изучить вопросы вариантов обработки результатов по изученному в разделе "Байесовская классификация. Иерархическая классификация. Расстояние между кластерами. Алгоритм DBSCAN." материалу.
3.1	Байесовская классификация. Иерархическая классификация. Расстояние между кластерами. Алгоритм DBSCAN.	18	3	-	5	-	-	-	-	-	-	10	-	<u>Изучение материалов литературных источников:</u> [2], п.3
4	Задачи классификации и регрессии. Деревья решений. Алгоритм CART, другие алгоритмы.	18	3	-	5	-	-	-	-	-	-	10	-	<u>Подготовка к лабораторной работе:</u> Для выполнения заданий по лабораторной работе "Реализация алгоритм CART для построения дерева решений" необходимо предварительно изучить тему и задачи выполнения лабораторной работы, а так же изучить вопросы вариантов обработки результатов по изученному в разделе "Задачи классификации и регрессии. Деревья решений. Алгоритм CART, другие алгоритмы." материалу.
4.1	Задачи классификации и регрессии. Деревья решений. Алгоритм CART, другие алгоритмы.	18	3	-	5	-	-	-	-	-	-	10	-	<u>Изучение материалов литературных источников:</u> [1], п.4 [3], п.3
5	Классификация текстов.	18	3	-	5	-	-	-	-	-	-	10	-	<u>Подготовка к лабораторной работе:</u> Для выполнения заданий по лабораторной работе "Классификация текстов. Байесовский классификатор" необходимо предварительно изучить тему и задачи выполнения лабораторной работы, а так же изучить
5.1	Классификация текстов.	18	3	-	5	-	-	-	-	-	-	10	-	

													вопросы вариантов обработки результатов по изученному в разделе "Классификация текстов." материалу. <u>Изучение материалов литературных источников:</u> [1], п.5
6	Системы аналитической обработки данных. Линейные модели для классификации и регрессии. Машина опорных векторов.	18	3	-	5	-	-	-	-	-	10	-	<u>Подготовка к лабораторной работе:</u> Для выполнения заданий по лабораторной работе "Машина опорных векторов" необходимо предварительно изучить тему и задачи выполнения лабораторной работы, а так же изучить вопросы вариантов обработки результатов по изученному в разделе "Системы аналитической обработки данных. Линейные модели для классификации и регрессии. Машина опорных векторов." материалу. <u>Изучение материалов литературных источников:</u> [3], п.4
6.1	Системы аналитической обработки данных. Линейные модели для классификации и регрессии. Машина опорных векторов.	18	3	-	5	-	-	-	-	-	10	-	
	Экзамен	36.0	-	-	-	-	2	-	-	0.5	-	33.5	
	Всего за семестр	144.0	16	-	32	-	2	-	-	0.5	60	33.5	
	Итого за семестр	144.0	16	-	32		2		-	0.5		93.5	

Примечание: Лек – лекции; Лаб – лабораторные работы; Пр – практические занятия; КПр – аудиторные консультации по курсовым проектам/работам; ИККП – индивидуальные консультации по курсовым проектам/работам; ГК- групповые консультации по разделам дисциплины; СР – самостоятельная работа студента; ИКР – иная контактная работа; ТК – текущий контроль; ПА – промежуточная аттестация

3.2 Краткое содержание разделов

1. Интеллектуальный анализ данных (Data Mining). Этапы исследования данных с помощью методов Data Mining. Типы закономерностей. Методы Data Mining.

1.1. Интеллектуальный анализ данных (Data Mining). Этапы исследования данных с помощью методов Data Mining. Типы закономерностей. Методы Data Mining.

Введение в «анализ данных». Работа с данными. Этапы решения задачи анализа данных и их взаимосвязи. Интеллектуальный анализ данных. Этапы исследования данных с помощью методов Data Mining. Типы закономерностей. Методы Data Mining. Data Mining как моделирование и как KDD. Принцип Бонферрони. Процесс разработки с использованием Data Mining..

2. Задачи Data Mining. Задача кластеризации. Смесь нормальных распределений. EM-алгоритм. K-means и его модификации.

2.1. Задачи Data Mining. Задача кластеризации. Смесь нормальных распределений. EM-алгоритм. K-means и его модификации.

Задачи Data Mining. Задача кластеризации. Смесь нормальных распределений. EM-алгоритм. K-means и его модификации. Постановка задачи. Многомерное нормальное распределение. Maximum Likelihood. Альтернативные функции расстояния. «Проклятие размерности». Альтернативные критерии качества..

3. Байесовская классификация. Иерархическая классификация. Расстояние между кластерами. Алгоритм DBSCAN.

3.1. Байесовская классификация. Иерархическая классификация. Расстояние между кластерами. Алгоритм DBSCAN.

Байесовская классификация. Апостериорное распределение параметров. Иерархическая классификация. Расстояние между кластерами. Неевклидовы пространства. Алгоритм DBSCAN. Виды объектов. Модификации алгоритма DBSCAN. Критерий Silhouette. Подходы к моделированию. Генеративные модели. Дискриминативные модели. Функции решения..

4. Задачи классификации и регрессии. Деревья решений. Алгоритм CART, другие алгоритмы.

4.1. Задачи классификации и регрессии. Деревья решений. Алгоритм CART, другие алгоритмы.

Основы регрессионного анализа. Простая линейная взаимосвязь. Уравнение регрессии. Подгонка линии регрессии. Понятие корреляции и ковариации. Оценка результатов классификации: разделение выборки, скользящий контроль, bootstrap. Метрики качества для вероятностных моделей. Метрики качества для функций решения. Кривая Receiver Operating Characteristic. Деревья решений. Алгоритм CART. Чистота узла. Теория информации. Выбор наилучшего разделения. Использование нескольких признаков. Укорачивание дерева. Вычислительная сложность. Отсутствующие значения. Алгоритм ID3. Алгоритм C4.5. Алгоритм C5.0..

5. Классификация текстов.

5.1. Классификация текстов.

Классификация текстов. Обработка текстов и Naive Bayes. Задачи Text Mining. Этапы обработки текстов. Декодирование. Разбиение на токены. Стоп-слова. Нормализация.

Стемминг. Лемматизация. Hear's law. Модели представления документов: Boolean model, Bag of words. Закон Ципфа. TF-IDF. Представление документов. Байесовский классификатор. Multinomial Naive Bayes. Сглаживание. Байесовские сети..

6. Системы аналитической обработки данных. Линейные модели для классификации и регрессии. Машина опорных векторов.

6.1. Системы аналитической обработки данных. Линейные модели для классификации и регрессии. Машина опорных векторов.

Системы аналитической обработки данных. Линейные модели для классификации и регрессии. ML – функция правдоподобия. ML – решение. Регуляризация. Логистическая регрессия. Градиентный спуск. Метод Ньютона. Обобщённые линейные модели. Случай линейно разделимых и линейно неразделимых классов. Задача оптимизации. Переобучение модели. Машина опорных векторов. Максимальный зазор. Смягчение ограничений. Функции ядра. Теорема Мерсера. Стохастический градиентный спуск..

3.3. Темы практических занятий

1. Машина опорных векторов (SVM);
2. Классификация текстов. Байесовский классификатор.;
3. Реализация алгоритм CART для построения дерева решений;
4. Реализация алгоритма DBSCAN для решения задачи кластеризации;
5. Реализация алгоритма K-means для решения задачи кластеризации.

3.4. Темы лабораторных работ

не предусмотрено

3.5 Консультации

3.6 Тематика курсовых проектов/курсовых работ

Курсовой проект/ работа не предусмотрены

3.7. Соответствие разделов дисциплины и формируемых в них компетенций

Запланированные результаты обучения по дисциплине (в соответствии с разделом 1)	Коды индикаторов	Номер раздела дисциплины (в соответствии с п.3.1)						Оценочное средство (тип и наименование)
		1	2	3	4	5	6	
Знать:								
алгоритмы анализа больших объёмов информации	ИД-1пк-2	+						Лабораторная работа/Реализация алгоритма K-means для решения задачи кластеризации
основные подходы к реализации методов искусственного интеллекта для анализа данных	ИД-1пк-2	+						Лабораторная работа/Машина опорных векторов (SVM) Лабораторная работа/Реализация алгоритма DBSCAN для решения задачи кластеризации
методы системного анализа	ИД-1пк-2						+	Лабораторная работа/Машина опорных векторов (SVM)
подходы к тестированию информационных систем	ИД-2пк-2	+						Лабораторная работа/Классификация текстов. Байесовский классификатор.
этапы жизненного цикла информационных систем	ИД-2пк-2			+				Лабораторная работа/Реализация алгоритма DBSCAN для решения задачи кластеризации
Уметь:								
современные платформы и языки программирования высокого уровня	ИД-1пк-2		+					Лабораторная работа/Реализация алгоритма K-means для решения задачи кластеризации
реализовать алгоритмы анализа больших объёмов данных	ИД-1пк-2						+	Лабораторная работа/Классификация текстов. Байесовский классификатор.
реализовать методы системного анализа	ИД-1пк-2				+			Лабораторная работа/Реализация алгоритм CART для построения дерева решений
выполнять доработку модулей информационных систем	ИД-2пк-2			+				Лабораторная работа/Реализация алгоритма DBSCAN для решения задачи кластеризации
осуществлять поддержку разработки информационных систем	ИД-2пк-2	+						Лабораторная работа/Машина опорных векторов (SVM)

4. КОМПЕТЕНТНОСТНО-ОРИЕНТИРОВАННЫЕ ОЦЕНОЧНЫЕ СРЕДСТВА ДЛЯ КОНТРОЛЯ ОСВОЕНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ (ТЕКУЩИЙ КОНТРОЛЬ УСПЕВАЕМОСТИ, ПРОМЕЖУТОЧНАЯ АТТЕСТАЦИЯ ПО ДИСЦИПЛИНЕ)

4.1. Текущий контроль успеваемости

1 семестр

Форма реализации: Компьютерное задание

1. Классификация текстов. Байесовский классификатор. (Лабораторная работа)
2. Машина опорных векторов (SVM) (Лабораторная работа)
3. Реализация алгоритм CART для построения дерева решений (Лабораторная работа)
4. Реализация алгоритма DBSCAN для решения задачи кластеризации (Лабораторная работа)
5. Реализация алгоритма K-means для решения задачи кластеризации (Лабораторная работа)

Балльно-рейтинговая структура дисциплины является приложением А.

4.2 Промежуточная аттестация по дисциплине

Экзамен (Семестр №1)

Для получения итоговой оценки баллы промежуточной аттестации суммируются с балами текущей аттестации.

В диплом выставляется оценка за 1 семестр.

Примечание: Оценочные материалы по дисциплине приведены в фонде оценочных материалов ОПОП.

5. УЧЕБНО-МЕТОДИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

5.1 Печатные и электронные издания:

1. Чубукова И. А.- "Data Mining", (2-е изд.), Издательство: "ИНТУИТ", Москва, 2016 - (470 с.)
<https://e.lanbook.com/book/100582>;
2. Мартин О.- "Байесовский анализ на Python", Издательство: "ДМК Пресс", Москва, 2020 - (340 с.)
<https://e.lanbook.com/book/140585>;
3. Макшанов А. В., Журавлев А. Е.- "Технологии интеллектуального анализа данных", (2-е изд., стер.), Издательство: "Лань", Санкт-Петербург, 2019 - (212 с.)
<https://e.lanbook.com/book/120063>;
4. Златопольский Д. М.- "Основы программирования на языке Python", Издательство: "ДМК Пресс", Москва, 2017 - (284 с.)
<https://e.lanbook.com/book/97359>.

5.2 Лицензионное и свободно распространяемое программное обеспечение:

1. СДО "Прометей";
2. Windows / Операционная система семейства Linux;
3. Майнд Видеоконференции;
4. Visual Studio;
5. Python;

6. Keras.

5.3 Интернет-ресурсы, включая профессиональные базы данных и информационно-справочные системы:

1. ЭБС Лань - <https://e.lanbook.com/>
2. ЭБС "Университетская библиотека онлайн" - http://biblioclub.ru/index.php?page=main_ub_red
3. Научная электронная библиотека - <https://elibrary.ru/>
4. База данных ВИНТИ online - <http://www.viniti.ru/>
5. База данных журналов издательства Elsevier - <https://www.sciencedirect.com/>
6. Электронные ресурсы издательства Springer - <https://link.springer.com/>
7. База данных Web of Science - <http://webofscience.com/>
8. База данных Scopus - <http://www.scopus.com>
9. Национальная электронная библиотека - <https://rusneb.ru/>
10. ЭБС "Консультант студента" - <http://www.studentlibrary.ru/>
11. Журналы American Chemical Society - <https://www.acs.org/content/acs/en.html>
12. Журналы American Institute of Physics - <https://www.scitation.org/>
13. Журналы American Physical Society - <https://journals.aps.org/about>
14. База данных издательства Annual Reviews Science Collection - <https://www.annualreviews.org/>
15. База данных Association for Computing Machinery Digital Library - <https://dl.acm.org/about/content>
16. Журналы издательства Cambridge University Press - <https://www.cambridge.org/core>
17. База данных IEL издательства IEEE (Institute of Electrical and Electronics Engineers, Inc.) - <https://ieeexplore.ieee.org/Xplore/home.jsp?reload=true>
18. База данных Computers & Applied Sciences Complete (CASC) - <http://search.ebscohost.com>
19. База данных INSPEC на платформе компании EBSCO Publishing - <http://search.ebscohost.com>
20. Журналы Institute of Physics (IOP), Великобритания - <https://iopscience.iop.org/>
21. Журналы научного общества Optical Society of America (OSA) - <https://www.osapublishing.org/about.cfm>
22. Патентная база Orbit Intelligence компании Questel - <https://www.orbit.com/>
23. Журналы издательства Oxford University Press - <https://academic.oup.com/journals/>
24. База данных диссертаций ProQuest Dissertations and Theses Global - <https://search.proquest.com/pqdtglobal/index>
25. Журналы Журналы Royal Society of Chemistry - <https://pubs.rsc.org/>
26. Журналы издательства SAGE Publication (Sage) - <https://journals.sagepub.com/>
27. Журнал Science - <https://www.sciencemag.org/>
28. Журналы научного общества Society of Photo-Optical Instrumentation Engineers (SPIE) Digital Library - <https://www.spiedigitallibrary.org/>
29. Коллекция журналов Taylor & Francis Group - <https://www.tandfonline.com/>
30. Журналы по химии Thieme Chemistry Package компании Georg Thieme Verlag KG - <https://www.thieme-connect.com/products/all/home.html>
31. Журналы издательства Wiley - <https://onlinelibrary.wiley.com/>
32. Электронная библиотека МЭИ (ЭБ МЭИ) - <http://elib.mpei.ru/login.php>
33. Портал открытых данных Российской Федерации - <https://data.gov.ru>
34. База открытых данных Министерства труда и социальной защиты РФ - <https://rosmintrud.ru/opendata>

35. **База открытых данных профессиональных стандартов Министерства труда и социальной защиты РФ** - <http://profstandart.rosmintrud.ru/obshchiy-informatsionnyy-blok/natsionalnyy-reestr-professionalnykh-standartov/>
36. **База открытых данных Министерства экономического развития РФ** - <http://www.economy.gov.ru>
37. **База открытых данных Росфинмониторинга** - <http://www.fedsfm.ru/opendata>
38. **Электронная открытая база данных "Polpred.com Обзор СМИ"** - <https://www.polpred.com>
39. **Информационно-справочная система «Кодекс/Техэксперт»** - <Http://proinfosoft.ru>; <http://docs.cntd.ru/>
40. **Национальный портал онлайн обучения «Открытое образование»** - <https://openedu.ru>
41. **Официальный сайт Федерального агентства по техническому регулированию и метрологии** - <http://protect.gost.ru/>
42. **Открытая университетская информационная система «РОССИЯ»** - <https://uisrussia.msu.ru>
43. **Официальный сайт Министерства науки и высшего образования Российской Федерации** - <https://minobrnauki.gov.ru>
44. **Официальный сайт Федеральной службы по надзору в сфере образования и науки** - <https://obrnadzor>
45. **Федеральный портал "Российское образование"** - <http://www.edu.ru>

6. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ ДИСЦИПЛИНЫ

Тип помещения	Номер аудитории, наименование	Оснащение
Учебные аудитории для проведения лекционных занятий и текущего контроля	Ж-120, Машинный зал ИВЦ	сервер, кондиционер
	Г-406, Учебная аудитория	парта, стол преподавателя, стул, доска меловая, компьютерная сеть с выходом в Интернет, мультимедийный проектор, экран
Учебные аудитории для проведения практических занятий, КР и КП	М-706, Дисплейный класс каф. "ПМИИ"	стол преподавателя, стол компьютерный, стул, компьютерная сеть с выходом в Интернет, мультимедийный проектор, экран, компьютер персональный, кондиционер
Учебные аудитории для проведения промежуточной аттестации	М-706, Дисплейный класс каф. "ПМИИ"	стол преподавателя, стол компьютерный, стул, компьютерная сеть с выходом в Интернет, мультимедийный проектор, экран, компьютер персональный, кондиционер
Помещения для самостоятельной работы	НТБ-303, Компьютерный читальный зал	стол компьютерный, стул, стол письменный, вешалка для одежды, компьютерная сеть с выходом в Интернет, компьютер персональный, принтер, кондиционер
Помещения для консультирования	М-706а, Консультационный зал кафедры ПМИИ	парта со скамьей, стол, стул, доска меловая, мультимедийный проектор
	М-704, Преподавательская кафедры ПМИИ	стол, стул, шкаф, тумба, доска меловая, компьютерная сеть с выходом в Интернет, мультимедийный проектор, экран, компьютер персональный,

		холодильник, кондиционер
Помещения для хранения оборудования и учебного инвентаря	Е-403, Склад	стол для работы с документами, шкаф, шкаф для документов

БАЛЛЬНО-РЕЙТИНГОВАЯ СТРУКТУРА ДИСЦИПЛИНЫ

Методы и средства анализа данных

(название дисциплины)

1 семестр

Перечень контрольных мероприятий текущего контроля успеваемости по дисциплине:

- КМ-1 Реализация алгоритма K-means для решения задачи кластеризации (Лабораторная работа)
- КМ-2 Реализация алгоритма DBSCAN для решения задачи кластеризации (Лабораторная работа)
- КМ-3 Реализация алгоритм CART для построения дерева решений (Лабораторная работа)
- КМ-4 Классификация текстов. Байесовский классификатор. (Лабораторная работа)
- КМ-5 Машина опорных векторов (SVM) (Лабораторная работа)

Вид промежуточной аттестации – Экзамен.

Номер раздела	Раздел дисциплины	Индекс КМ:	КМ-1	КМ-2	КМ-3	КМ-4	КМ-5
		Неделя КМ:	4	8	12	14	16
1	Интеллектуальный анализ данных (Data Mining). Этапы исследования данных с помощью методов Data Mining. Типы закономерностей. Методы Data Mining.						
1.1	Интеллектуальный анализ данных (Data Mining). Этапы исследования данных с помощью методов Data Mining. Типы закономерностей. Методы Data Mining.		+	+		+	+
2	Задачи Data Mining. Задача кластеризации. Смесь нормальных распределений. EM-алгоритм. K-means и его модификации.						
2.1	Задачи Data Mining. Задача кластеризации. Смесь нормальных распределений. EM-алгоритм. K-means и его модификации.		+				
3	Байесовская классификация. Иерархическая классификация. Расстояние между кластерами. Алгоритм DBSCAN.						
3.1	Байесовская классификация. Иерархическая классификация. Расстояние между кластерами. Алгоритм DBSCAN.			+			
4	Задачи классификации и регрессии. Деревья решений. Алгоритм CART, другие алгоритмы.						
4.1	Задачи классификации и регрессии. Деревья решений. Алгоритм CART, другие алгоритмы.				+		
5	Классификация текстов.						
5.1	Классификация текстов.					+	
6	Системы аналитической обработки данных. Линейные модели для классификации и регрессии.						

	Машина опорных векторов.					
6.1	Системы аналитической обработки данных. Линейные модели для классификации и регрессии. Машина опорных векторов.					+
Вес КМ, %:		20	20	20	20	20