



Министерство науки
и высшего образования РФ
ФГБОУ ВО «НИУ «МЭИ»
Институт дистанционного
и дополнительного образования



**ОЦЕНОЧНЫЕ И МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ДЛЯ КОНТРОЛЯ
ОСВОЕНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ
ДОПОЛНИТЕЛЬНОЙ ОБРАЗОВАТЕЛЬНОЙ ПРОГРАММЫ
повышения квалификации
«Основы анализа текстовых данных»,**

Текущий контроль

Текущий контроль проводится в соответствии с характеристиками контрольных заданий и представлен в табл. 1.

Таблица 1

Характеристика заданий текущего контроля

Наименование дисциплины (модуля)	Форма контроля/наименование контрольной точки	Пример задания	Критерии оценки
Основы анализа текстовых данных			
Постановка задачи машинного обучения	Тестирование	1. Каким типом признака является "город проживания"? а. количественный б. бинарный с. номинальный д. порядковый 2. Какие данные называют "скрытыми"? а. новые б. нетривиальные с. полезные для практики 3. Что такое "машинное обучение"? а. обучение на примерах (прецедентах) б. обучение по заданным правилам	<i>Оценка:</i> зачтено <i>Описание характеристики выполнения знания:</i> Оценка "зачтено" выставляется если задание выполнено правильно или с незначительными недочетами. <i>Оценка:</i> не зачтено <i>Описание характеристики выполнения знания:</i> Оценка "не зачтено" выставляется если задание не выполнено в отведенный срок или результат не соответствует заданию
Особенности задачи	Тестирование	1. При использовании модели "Мешок слов" текст	<i>Оценка:</i> зачтено <i>Описание</i>

обработки текстовых данных		<i>Машинное обучение и анализ данных</i> будет представляться в виде: а. В виде одного токена "Машинное обучение и анализ данных" б. В виде пяти токенов "Машинное", "обучение", "и", "анализ", "данных" с. В виде трех токенов "Машинное обучение", "и", "анализ данных" 2. В чем особенность матрицы "документ-термин"? а. матрица является разреженной б. Высокая размерность с. матрица является диагональной д. матрица состоит из весов, показывающих взаимосвязь терминов в документах 3. Модель "Мешок слов" является а. частично структурированной моделью б. неструктурированной моделью с. структурированной моделью	<i>характеристики выполнения знания:</i> Оценка "зачтено" выставляется если задание выполнено правильно или с незначительными недочетами. <i>Оценка:</i> не зачтено <i>Описание</i> <i>характеристики выполнения знания:</i> Оценка "не зачтено" выставляется если задание не выполнено в отведенный срок или результат не соответствует заданию
Методы классификации данных	Тестирование	1. Чему будет равен центроид для четырех объектов: (3,3; 7,2) (1,0; 4,0) (3,7; 3,8) (2,0; 5,0) Ответ записать по правилам математического округления до 1 знака после запятой в таком же формате, как и приведенные выше объекты 2. Настраиваемые параметры метода k-БС а. Количество признаков для поиска б. способ взвешивания	<i>Оценка:</i> зачтено <i>Описание</i> <i>характеристики выполнения знания:</i> Оценка "зачтено" выставляется если задание выполнено правильно или с незначительными недочетами. <i>Оценка:</i> не зачтено <i>Описание</i> <i>характеристики выполнения знания:</i> Оценка "не зачтено" выставляется если задание не выполнено в отведенный срок или результат не

		с. Количество ближайших соседей d. Метрика близости е. способ расчета центроидов 3. Вероятность появления объекта X в классе Q оценивается как a. Вероятность появления класса Q в выборке b. произведение вероятностей появления признаков x в классе Q с. сумма вероятностей появления признаков x в классе Q	соответствует заданию												
Кластеризация данных, выявление дубликатов текстовых документов	Тестирование	1. Чем характерны ленточные кластеры? a. межкластерное расстояние больше внутрикластерного b. межкластерное расстояние меньше внутрикластерного с. Зависимость образована не по сходству, а по типу регулярности 2. После кластеризации на разное количество кластеров измерена средняя внутрикластерная дисперсия, результаты приведены в таблице. <table><tr><td>2</td><td>3</td><td>4</td><td>5</td><td>6</td><td>7</td></tr><tr><td>5,4</td><td>5,1</td><td>3,9</td><td>3,7</td><td>3,6</td><td>3,6</td></tr></table> Какое оптимальное количество кластеров? 3. Внутрикластерную дисперсию следует a. уменьшать b. увеличивать	2	3	4	5	6	7	5,4	5,1	3,9	3,7	3,6	3,6	Оценка: зачтено Описание характеристики выполнения знания: Оценка "зачтено" выставляется если задание выполнено правильно или с незначительными недочетами. Оценка: не зачтено Описание характеристики выполнения знания: Оценка "не зачтено" выставляется если задание не выполнено в отведенный срок или результат не соответствует заданию
2	3	4	5	6	7										
5,4	5,1	3,9	3,7	3,6	3,6										

Промежуточная аттестация

Промежуточная аттестация по программе проводится в форме зачета, экзамена или отчета о стажировке в соответствии с учебным планом. Характеристика заданий представлена в табл. 2.

Таблица 2

Характеристика заданий промежуточной аттестации

Наименование дисциплины (модуля)	Пример задания	Критерии оценки
Основы анализа текстовых данных	Не предусмотрено	Не предусмотрено

Итоговая аттестация

Итоговая аттестация по программе проводится в форме *итогового зачета*. Характеристика заданий представлена в табл. 3.

Таблица 3

Характеристика заданий итоговой аттестации

Вид контроля	Краткая характеристика задания	Критерии оценки
Итоговая аттестация	1. 5 ближайших соседей к объекту X принадлежат к следующим классам: А, В, А, А, С К какому классу будет отнесен объект X? а. А б. В с. С 2. Решается задача задачи фильтрации спама. Даны точность и полнота классификации 3 классификаторов (расчеты относительно класса 'спам'). Какой из них стоит применять в данной задаче? а. Precision =0.85, Recall =0.85 б. Precision = 0.95, Recall = 0.8 с. Precision=0.8, Recall=0.95 3. Дан профиль класса Q: Класс - 0,8 классификация - 0,75 метод - 0,65 текст - 0,55 алгоритм - 0,5 задача - 0,45 Определить вес документа в классе (считаем что слова приведены к нормальной форме: "класса", "классов" -> "класс"):	<i>Оценка:</i> зачтено <i>Описание характеристики выполнения знания:</i> Оценки «зачтено» заслуживает слушатель, обнаруживший всестороннее, систематическое и глубокое знание учебного и нормативного материала, умеющий свободно выполнять задания, предусмотренные программой. <i>Оценка:</i> не зачтено <i>Описание характеристики выполнения знания:</i> Оценка «не зачтено» выставляется слушателю, обнаружившему пробелы в знаниях основного учебного материала, допускающему принципиальные ошибки в выполнении предусмотренных программой заданий.

	Одной из задач машинного обучения является задача классификации, которую часто приходится решать в условиях несбалансированности классов.	
--	---	--

Независимая оценка качества обучения

Независимая оценка качества обучения предполагает внутренний аудит программ ДПО и анкетирование слушателей и/или работодателей по вопросам удовлетворенности процессом и результатами обучения.

Учебно-методическое и информационное обеспечение

а) литература НТБ МЭИ:

1. Маннинг, К. Д. Введение в информационный поиск : пер. с англ. / К. Д. Маннинг, П. Рагхаван, Х. Шютце . – М. : Вильямс, 2011 . – 528 с. - ISBN 978-5-8459-1623-5 .;

2. Методы и модели анализа данных: OLAP и Data Mining : учебное пособие по специальности 071900 "Информационные системы и технологии" направления 654700 "Информационные системы" / А. А. Барсегян, и др. – СПб. : БХВ-Петербург, 2004 . – 336 с. + CD-ROM . - ISBN 5-941575-22-X .;

3. Мохов, А. С. Анализ и обработка текстовых данных : практикум по курсу "Интеллектуальные информационные системы" по направлению подготовки магистров 27.04.04 "Управление в технических системах" / А. С. Мохов, В. О. Толчеев, А. А. Бородин, Нац. исслед. ун-т "МЭИ" (НИУ"МЭИ") . – Москва : Изд-во МЭИ, 2020 . – 52 с. - ISBN 978-5-7046-2284-0 .

<http://elbib.mpei.ru/elbib/view.php?id=11207>;

4. Толчеев, В. О. Современные методы обработки и анализа текстовой информации : учебное пособие по курсу "Интеллектуальные информационные системы" по специальности "Управление и информатика в технических системах" / В. О. Толчеев, Моск. энерг. ин-т (МЭИ ТУ) . – М. : Изд-во МЭИ, 2006 . – 76 с. - ISBN 5-7046-1285-7 ..

б) литература ЭБС и БД:

1. Алексеев Д. С.,Щекожихин О. В.- "Технологии интеллектуального анализа данных", (2-е изд., стер.), Издательство: "Лань", Санкт-Петербург, 2024 - (176 с.)
<https://e.lanbook.com/book/362915>;

2. 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>.

в) используемые ЭБС:

Не предусмотрено

Руководитель
ОДПО, ЦДО
Сократ

Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
Сведения о владельце ЦЭП МЭИ	
Владелец	Максимова А.А.
Идентификатор	R6a033f13-VorozhtsovaAA-daecd82

А.А.
Максимова

Начальник ОДПО

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Селиверстов Н.Д.
	Идентификатор	Rf19596d9-SeliverstovND-39ee0b7

Н.Д.
Селиверстов