



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



УТВЕРЖДАЮ:  
Директор ИДДО

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Шиндина Т.А.
	Идентификатор	Rd0ad64b2-5hindinaTA-e12224c9

(подпись)

Т.А. Шиндина  
(расшифровка подписи)

ДОПОЛНИТЕЛЬНАЯ ОБРАЗОВАТЕЛЬНАЯ ПРОГРАММА  
*повышения квалификации*

Наименование программы	Основы анализа текстовых данных
Форма обучения	очно-заочная
Выдаваемый документ	удостоверение о повышении квалификации
Новая квалификация	не присваивается
Центр ДО	ОДПО, Центр профессиональной переподготовки преподавателей "Управление в высшем образовании"

Зам. директора ИДДО  
(должность)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Усманова Н.В.
	Идентификатор	R3b653adc-UsmanovaNatV-90b3fa4

(подпись)

Н.В.  
Усманова  
(расшифровка подписи)

Начальник ОДПО  
(должность)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Крохин А.Г.
	Идентификатор	R6d4610d5-KrokhinAG-aa301f84

(подпись)

А.Г. Крохин  
(расшифровка подписи)

Руководитель ОДПО,  
ЦПП УВО  
(должность)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Максимова А.А.
	Идентификатор	R6a033f13-VorozhtsovaAA-daecd87

(подпись)

А.А.  
Максимова  
(расшифровка подписи)

Руководитель образовательной программы  
(должность)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Мохов А.С.
	Идентификатор	R55ae9104-MokhovAS-2434a28b

(подпись)

А.С. Мохов  
(расшифровка подписи)

Москва

## **1. ОБЩАЯ ХАРАКТЕРИСТИКА ПРОГРАММЫ**

**Цель:** формирование знаний и навыков в области анализа и обработки текстовых данных (Text mining).

**Программа составлена в соответствии:**

- с Федеральным государственным образовательным стандартом высшего образования по направлению подготовки 27.03.04 Управление в технических системах, утвержденным приказом Минобрнауки от 31.07.2020 г. № 87126.08.2020 г. № 59489.

**Форма реализации:** обучение с применением дистанционных образовательных технологий.

**Форма обучения:** очно-заочная.

**Режим занятий:**

Расписание занятий по дополнительной образовательной программе может устанавливаться в зависимости от набора в группы. Конкретные даты проведения занятий указываются в договоре на оказание образовательных услуг. Данные расписания хранятся в электронной системе учета хода реализации программы. При любом графике занятий учебная нагрузка устанавливается не более 40 часов в неделю, включая все виды аудиторной и внеаудиторной (самостоятельной) учебной работы слушателя.

**Требования к уровню подготовки слушателя, необходимые для освоения программы:** слушатели должны иметь высшее образование или обучаться в высшем учебном заведении и обладать знанием основ статистики и теории вероятностей..

**Выдаваемый документ:** при успешном прохождении программы и сдаче итоговой аттестации выдается удостоверение о повышении квалификации установленного образца.

**Срок действия итоговых документов**

Срок действия итоговых документов регламентируется на основе правил по работе с персоналом в сфере деятельности данной программы, устанавливается на основе содержания программы и составляет (в годах): 3.

## 2. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ

### 2.1. Компетенции

В результате освоения дополнительной образовательной программы слушатель должен обладать компетенциями (табл. 1).

Таблица 1

Компетентностно-ориентированные требования к результатам освоения программы

Компетенция	Требования к результатам
ОПК-6: Способен разрабатывать и использовать алгоритмы и программы, современные информационные технологии, методы и средства контроля, диагностики и управления, пригодные для практического применения в сфере своей профессиональной деятельности	Знать: - методы и алгоритмы интеллектуального анализа текстовых данных.
	Уметь: - решать задачи интеллектуального анализа текстовых данных.
	Владеть: - современными способами обработки текстовых данных.

В результате освоения программы слушатель должен быть способен реализовывать трудовые функции в соответствии с профессиональным стандартом (табл. 2).

Уровень квалификации \_\_\_\_\_.

Таблица 2

Практико-ориентированные требования к результатам освоения программы

Трудовые функции	Требования к результатам
------------------	--------------------------

### 2.2. Характеристика нового вида профессиональной деятельности, новой квалификации

Не предусмотрено

## 3. СТРУКТУРА И СОДЕРЖАНИЕ ПРОГРАММЫ (РАБОЧИЕ ПРОГРАММЫ ДИСЦИПЛИН (МОДУЛЕЙ))

### 3.1. Трудоемкость программы

Трудоемкость программы включая все виды аудиторной и внеаудиторной (самостоятельной) работы составляет:

- 1 зачетных единиц;

36 ак. ч.

Структура программы с указанием наименования дисциплин (модулей) и их трудоемкости представлена в табл. 3.

Учебный план дополнительной образовательной программы представлен в приложение А., являющийся неотъемлемой частью программы.

Таблица 3

## Структура программы и формы аттестации

№	Наименование дисциплин (модулей)	всего	Контактная работа, ак. ч					Самостоятельная работа, ак. ч	Стажировка, ак. ч	Форма аттестации			
			всего	аудиторные занятия	электронное обучение	обучение с ДОГ	контроль			текущий контроль (тест, опрос и пр.)	промежуточная аттестация (зачет, экзамен, защита отчета о стажировке)	итоговая аттестация (итоговый зачет, итоговый экзамен, доклад по результатам стажировки, итоговый аттестационный экзамен, итоговая аттестационная работа)	
1	2	3	4	5	6	7	8	9	11	12	13	14	
1	Основы анализа данных	8	4		4			4			Нет		
1.1.	Постановка задачи машинного обучения	8	4		4			4		Тестирование			
2	Анализ текстовых данных (Text Mining)	8	4		4	0		4			Нет		
2.1.	Особенности задачи обработки текстовых данных	8	4		4	0		4		Тестирование			
3	Задача классификации текстовых документов	8	4		4	0		4			Нет		
3.1.	Методы классификации данных	8	4		4	0		4		Тестирование			
4	Другие задачи, решаемые в рамках Text Mining	10	4		4			6			Нет		
4.1.	Кластеризация данных, выявление дубликатов текстовых документов	10	4		4			6		Тестирование			
5	Итоговая аттестация	20	0				03	17				Итоговый зачет	
	<b>ИТОГО:</b>	<b>360</b>	<b>163</b>	<b>0</b>	<b>16</b>	<b>0</b>	<b>03</b>	<b>197</b>	<b>0</b>				

**3.2. Содержание программы (рабочие программы дисциплин (модулей))**

Содержание дисциплин (модулей) представлено в табл. 4.

Таблица 4

## Содержание дисциплин (модулей)

№	Наименование дисциплин (модулей)	Содержание дисциплин (модулей)
1.	Основы анализа данных	
1.1.	Постановка задачи машинного обучения	Типы задач, решаемых методами машинного обучения (МО). Формальная постановка задачи машинного обучения. Задачи анализа тестовой информации, решаемые методами МО. Критерии качества: аккуратность, полнота, точность, F-мера, площадь под кривой ошибок, матрица неточностей. Способы формирования выборок. Обучающая, экзаменационная и тестовая выборки. Обучение моделей. Явление переобучения модели. Несбалансированные классы. Методы борьбы с несбалансированностью: oversampling, undersampling. Специальные стратегии сэмплинга в условиях несбалансированных классов
2.	Анализ текстовых данных (Text Mining)	
2.1.	Особенности задачи обработки текстовых данных	Text Mining и особенности задач, связанных с анализом текстов. Онтологии и тезаурусы. Статистический подход к анализу текстовой информации. Проблема снижения размерности в задачах Text Mining. Предварительная обработка данных: стемминг, лемматизация. Слова, не несущие информации. Выявление информативных признаков. Взвешивание как способ выявления информативных терминов. Статистический подход к выявлению информативных терминов. Теоретико-информационный подход к выявлению информативных терминов. Модель представления текстовых данных в математическом виде. Модель «Мешок слов» (Bag of words). Частично- и полностью структурированные модели.
3.	Задача классификации текстовых документов	
3.1.	Методы классификации данных	Систематизация и обзор методов классификации. Метод ближайших соседей. Метод деревьев решений. Метод опорных векторов. Метод логистической регрессии. Наивный байесовский метод. Профильные методы классификации. Ансамблевые методы классификации. Оценка разнородности методов классификации. Бустинг, бэггинг. Метод случайного леса деревьев решений.
4.	Другие задачи, решаемые в рамках Text Mining	
4.1.	Кластеризация данных, выявление дубликатов текстовых документов	Постановка задачи кластеризации, ее особенности. Иерархическая кластеризация. EM-алгоритм кластеризации. Семейство алгоритмов k-means, другие алгоритмы кластеризации. Виды дубликатов: полные

№	Наименование дисциплин (модулей)	Содержание дисциплин (модулей)
		дубликаты, явные дубликаты, нечеткие дубликаты. Коэффициент ассоциативности Жаккара. Семейство методов шинглов. Методы выявления дубликатов Winnowing, SpotSigs, I-Match, коэффициент Джаро-Винклера

Аннотации рабочих программ дисциплин (модулей) представлены в приложении Б.

#### 4. ПРАКТИЧЕСКАЯ ПОДГОТОВКА

Информация о практической подготовке в структуре дополнительной образовательной программы представлена в приложение В.

В рамках учебного плана дополнительной образовательной программы используются традиционные образовательные технологии, а также интерактивные технологии, представленные в табл. 5.

Таблица 5

Характеристика образовательной технологии

Наименование	Краткая характеристика
Тестирование	Тестирования по нескольким темам курса

#### 5. ОЦЕНОЧНЫЕ И МЕТОДИЧЕСКИЕ МАТЕРИАЛЫ ДЛЯ КОНТРОЛЯ ОСВОЕНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ

##### 5.1. Текущий контроль

Текущий контроль проводится в соответствии с характеристиками контрольных заданий и представлен в Таблице 1 приложения Г.

##### 5.2. Промежуточная аттестация

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

##### 5.3. Итоговая аттестация

Итоговая аттестация по программе проводится в форме *итогового зачета*. Характеристика заданий представлена Таблице 3 приложения Г.

##### 5.4. Независимый контроль качества обучения

Порядок независимой оценки качества дополнительной образовательной программы представлен в приложении Г.

## **6. ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ И РЕСУРСНОЕ ОБЕСПЕЧЕНИЕ**

### **6.1. 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. Бородкин, Нац. исслед. ун-т "МЭИ" (НИУ"МЭИ") . – Москва : Изд-во МЭИ, 2020 . – 52 с. - ISBN 978-5-7046-2284-0 .

<http://elibr.mpei.ru/elibr/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.,Щекочихин О. В.- "Технологии интеллектуального анализа данных", Издательство: "Лань", Санкт-Петербург, 2022 - (176 с.)  
<https://e.lanbook.com/book/187559>;

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*

### **6.2. Кадровое обеспечение**

Для реализации дополнительной образовательной программы привлекаются преподаватели из числа штатных научно-педагогических работников ФГБОУ ВО «НИУ «МЭИ» и лица, представители работодателей или объединений работодателей. Информация о кадровом обеспечении дополнительной образовательной программы представлена в приложении Д.

Сведения о руководителе дополнительной образовательной программы представлены в приложении Е.

### 6.3. Финансовое обеспечение

План расходов и расчет обоснования стоимости по дополнительной образовательной программе представлены в приложении Ж.

Финансирование программы осуществляется за счет личных средств слушателей или заказчиков, по направлению которых проводится обучение. В качестве заказчика могут выступать работодатели, университеты (в том числе МЭИ), государственные структуры и прочие участники образовательного рынка.

### 6.4. Материально-техническое обеспечение

Материально-технические условия реализации дополнительной образовательной программы представлены в Приложении З.

Календарный график учебного процесса разрабатывается с учетом требований к качеству освоения и по запросам обучающихся (Приложение И). Расписание занятий разрабатывается на каждую реализуемую программу.

## ЛИСТ ИЗМЕНЕНИЙ (АКТУАЛИЗАЦИИ)

№ п/п	Содержание изменения (актуализации)	Дата утверждения изменений
-------	-------------------------------------	----------------------------

Руководитель  
образовательной  
программы  

---

(должность)

	Подписано электронной подписью ФГБОУ ВО «НИУ «МЭИ»	
	Сведения о владельце ЦЭП МЭИ	
	Владелец	Мохов А.С.
Идентификатор	R55ae9104-MokhovAS-2434a28b	

---

А.С. Мохов  
(расшифровка  
подписи)