

**Рабочая программа дисциплины (модуля) «Машинное обучение», включающая  
оценочные и методические материалы**

**1. Требования к результатам обучения по дисциплине (модулю)**

**1.1. Перечень компетенций, формируемых дисциплиной (модулем) в процессе освоения образовательной программы**

Группа компетенций	Категория компетенций	Коды и содержание компетенций
Профессиональные	-	ПК-3. Способен на основе результатов анализа больших данных разрабатывать компьютерные модели, обеспечивающие информационно-управляющие системы умного цифрового производства

**1.2. Компетенции и индикаторы их достижения, формируемых дисциплиной (модулем) в процессе освоения образовательной программы**

Код компетенции	Код индикатора компетенции	Содержание индикатора компетенции
ПК-3	ПК-3.1	Составляет описание проводимых исследований, готовит данные для составления обзоров и отчетов
	ПК-3.2	Разрабатывает имитационные модели компонентов и процессов автоматизированных систем умного производства, а также модели составных частей процессов жизненного цикла объектов высокотехнологичных продуктов и производств
	ПК-3.3	Проектирует информационно-управляющие системы умного производства и разрабатывает их программные модули

**1.3. Результаты обучения по дисциплине (модулю)**

**Цель изучения дисциплины (модуля)** – изучение современных математических методов машинного обучения, предназначенных для анализа данных и построения предсказательных моделей.

В результате изучения дисциплины (модуля) обучающийся должен

**знать:**

- основы архитектуры, устройства и функционирования вычислительных систем;
- устройство и принципы функционирования информационных систем;
- методологии разработки программного обеспечения.

**уметь:**

- применять методологии управления проектами разработки программного обеспечения.

**владеть:**

- навыками оценки сроков, ресурсоемкости, себестоимости проекта по разработке системного ПО;
- инструментальными средствами разработки.

**2. Объем, структура и содержание дисциплины (модуля)**

**2.1. Объем дисциплины (модуля)**

Виды учебной работы	Формы обучения
	Очная
<b>Общая трудоемкость:</b> зачетные единицы/часы	3/108
<b>Контактная работа:</b>	48
Лекции	16
Лабораторные работы	32
Практические занятия, семинары	0
<b>Промежуточная аттестация:</b> зачет	0
<b>Самостоятельная работа (СР)</b>	60

**2.2. Темы (разделы) дисциплины (модуля) с указанием отведенного на них количества часов по формам образовательной деятельности**

### Очная форма обучения

Формы обучения								
№ п/п	Наименование тем (разделов)	Виды учебной работы (в часах)						СР
		Контактная работа						
		Занятия лекционного типа		Занятия семинарского типа				
		Л	Иные	ПЗ	С	ЛР	Иные	
1.	Основы машинного обучения	2	-	-	-	-	-	12
2.	Классификаторы	4	-	-	-	6	-	12
3.	Регрессионные модели	4	-	-	-	10	-	12
4.	Кластеризация	2	-	-	-	10	-	12
5.	Глубокое обучение	4	-	-	-	6	-	12

### 2.3. Содержание дисциплины (модуля), структурированное по темам (разделам) и видам работ

#### Содержание лекционного курса

№ п/п	Наименование тем (разделов)	Содержание лекционного курса
1.	Основы машинного обучения	Основные концепции машинного обучения
2.	Классификаторы	Алгоритмы построения классификаторов
3.	Регрессионные модели	Алгоритмы построения регрессоров
4.	Кластеризация	Выделение кластеров в данных
5.	Глубокое обучение	Нейронные сети для анализа данных

#### Содержание занятий семинарского типа

№ п/п	Наименование тем (разделов)	Тип	Содержание занятий семинарского типа
1.	Классификаторы	ЛР	Построение классификационных моделей и оценка их эффективности
2.	Регрессионные модели	ЛР	Построение регрессивных моделей и оценка их эффективности
3.	Кластеризация	ЛР	Решение задачи кластеризации данных и оценка его эффективности
4.	Глубокое обучение	ЛР	Решение классификационных и регрессионных задач нейросетевыми методами

#### Содержание самостоятельной работы

№ п/п	Наименование тем (разделов)	Содержание самостоятельной работы
1.	Основы машинного обучения	Работа с лекционным материалом, поиск и обзор литературы и электронных источников информации по индивидуально заданной проблеме курса; Работа в электронном курсе (изучение теоретического материала, выполнение тестов и контрольных работ, разработка программ); Выполнение домашних заданий; Разработка программ для решения нестандартных задач; Анализ научных публикаций по заранее определенной преподавателем теме.
2.	Классификаторы	Работа с лекционным материалом, поиск и обзор литературы и электронных источников информации по индивидуально заданной проблеме курса; Работа в электронном курсе (изучение теоретического материала, выполнение тестов и контрольных работ, разработка программ); Выполнение домашних заданий; Подготовка к лабораторным работам; Разработка программ для решения нестандартных задач; Анализ научных публикаций по заранее определенной преподавателем теме.

3.	Регрессионные модели	Работа с лекционным материалом, поиск и обзор литературы и электронных источников информации по индивидуально заданной проблеме курса; Работа в электронном курсе (изучение теоретического материала, выполнение тестов и контрольных работ, разработка программ); Выполнение домашних заданий; Подготовка к лабораторным работам; Разработка программ для решения нестандартных задач; Анализ научных публикаций по заранее определенной преподавателем теме.
4.	Кластеризация	Работа с лекционным материалом, поиск и обзор литературы и электронных источников информации по индивидуально заданной проблеме курса; Работа в электронном курсе (изучение теоретического материала, выполнение тестов и контрольных работ, разработка программ); Выполнение домашних заданий; Подготовка к лабораторным работам; Разработка программ для решения нестандартных задач; Анализ научных публикаций по заранее определенной преподавателем теме.
5.	Глубокое обучение	Работа с лекционным материалом, поиск и обзор литературы и электронных источников информации по индивидуально заданной проблеме курса; Работа в электронном курсе (изучение теоретического материала, выполнение тестов и контрольных работ, разработка программ); Выполнение домашних заданий; Подготовка к лабораторным работам; Разработка программ для решения нестандартных задач; Анализ научных публикаций по заранее определенной преподавателем теме.

### **3. Оценочные материалы для проведения текущего контроля успеваемости и промежуточной аттестации обучающихся по дисциплине (модулю)**

По дисциплине (модулю) предусмотрены следующие виды контроля качества освоения:

- текущий контроль успеваемости;
- промежуточная аттестация обучающихся по дисциплине (модулю).

#### **3.1.1. Оценочные материалы для проведения текущей аттестации по дисциплине (модулю)**

№ п/п	Контролируемые темы (разделы)	Наименование оценочного средства
1.	Основы машинного обучения	Устный опрос, контрольная работа
2.	Классификаторы	Устный опрос, контрольная работа
3.	Регрессионные модели	Устный опрос, контрольная работа
4.	Кластеризация	Устный опрос, контрольная работа
5.	Глубокое обучение	Устный опрос, контрольная работа

#### **3.1.1. Типовые контрольные задания или иные материалы, необходимые для оценки знаний, умений, навыков и (или) опыта деятельности в процессе текущего контроля успеваемости**

##### **Устный опрос**

- Классификация алгоритмов машинного обучения.
- Регрессия, линейная регрессия, линейные модели регрессии.
- Целевая функция линейной регрессии.
- Классификация. Логистическая регрессия.
- Целевая функция логистической регрессии.
- Регуляризация.

- Нейронные сети. Структура нейронной сети.
- Алгоритм обратного распространения ошибки.
- Деревья решений. Обучение дерева решений.
- Алгоритм Random Forest.
- Алгоритм AdaBoost.
- Каскад классификаторов. Обучение каскада.
- Кластеризация. Алгоритм k-means

### **Контрольная работа**

#### **Примерные задания**

1. Выберите верные утверждения про площадь под ROC-кривой (AUC-ROC)
    - Площадь под ROC-кривой для идеального алгоритма зависит от доли положительных объектов в выборке
    - ROC-кривая строится в осях "доля верных отрицательных классификаций" и "доля верных положительных классификаций"
    - ROC-кривая строится в осях "доля ошибочных положительных классификаций" и "доля верных положительных классификаций"
    - Площадь под ROC-кривой для идеального алгоритма равна единице
  2. Как определяется ответ решающего дерева в задаче классификации для заданного объекта?
    - Ответ определяется на основе значения в любой из вершин дерева
    - Ответ определяется на основе суммы значений из всех вершин, по которым прошел данный объект
    - По значению в листе, который попал в данный объект
  3. Выберите утверждения, которые относятся к Bag-of-words матрица признаков плотная
    - матрица признаков разреженная
    - игнорирует порядок слов в предложении
    - учитывает порядок слов в предложении
    - учитывает близость слов по значению
  4. Что такое дисперсия у модели?
  5. Как можно провалидировать\проверить вашу модель?
  6. Что произойдет если линейной регрессии на вход подать три колонки  $x$ ,  $y$ ,  $z$ , где  $z=x+y$ ?
  7. Какой из нижеприведенных методов позволяет привести слово к его нормальной форме?
    - Стемминг
    - Сегментация
    - Лемматизация
    - Токенизация
  8. У вас есть матрица мешка слов из твитов, в которой каждый твит рассматривается как один документ. Какие из следующих утверждений верны?
    - Удаление стоп-слов из данных повлияет на размерность данных.
    - Нормализация слов в данных снизит размерность данных
    - Преобразование всех слов в нижний регистр не повлияет на размерность данных
  9. Выберите утверждения, которые относятся к Bag-of-words матрица признаков плотная
  10. Опишите 3 основных этапа обучения ULMFit.
- Контрольная работа №2*

1. Какими из перечисленных ниже свойств обладают нейронные сети?
  - Могут восстанавливать очень сложные разделяющие поверхности
  - Могут дообучаться на новых данных
  - Дают хорошо интерпретируемые модели
  - До определенного момента не переобучаются при увеличении числа параметров
  - Практически не зависят от выбора параметров оптимизационного алгоритма
2. За что отвечают функции активации в нейронных сетях?
3. Что такое learning rate?
4. Что произойдет если learning rate будет слишком большой или слишком маленький?
5. Какую проблему частично позволяет решить использование CRF в задаче NER
6. Кратко опишите архитектуру на основе LSTM, которую используют для решения задачи NER
7. Опишите 2 основных варианта символьного представления токена
8. Какой вид attention чаще всего используется в современных моделях?
9. Как по корпусу получается BPE и какие преимущества дает?
10. Чем отличается scaled dot product attention трансформера от обычного dot product attention?

### **3.1.2. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков и (или) опыта деятельности в ходе текущего контроля успеваемости**

#### **Устный ответ**

Оценка знаний предполагает дифференцированный подход к обучающемуся, учет его индивидуальных способностей, степень усвоения и систематизации основных понятий и категорий по дисциплине. Кроме того, оценивается не только глубина знаний поставленных вопросов, но и умение использовать в ответе практический материал. Оценивается культура речи, владение навыками ораторского искусства.

*Критерии оценивания:* последовательность, полнота, логичность изложения, анализ различных точек зрения, самостоятельное обобщение материала, использование профессиональных терминов, культура речи, навыки ораторского искусства. Изложение материала без фактических ошибок.

Оценка «*отлично*» ставится в случае, когда материал излагается исчерпывающе, последовательно, грамотно и логически стройно, при этом раскрываются не только основные понятия, но и анализируются точки зрения различных авторов. Обучающийся не затрудняется с ответом, соблюдает культуру речи.

Оценка «*хорошо*» ставится, если обучающийся твердо знает материал, грамотно и по существу излагает его, знает практическую базу, но при ответе на вопрос допускает несущественные погрешности.

Оценка «*удовлетворительно*» ставится, если обучающийся освоил только основной материал, но не знает отдельных деталей, допускает неточности, недостаточно правильные формулировки, нарушает последовательность в изложении материала, затрудняется с ответами, показывает отсутствие должной связи между анализом, аргументацией и выводами.

Оценка «*неудовлетворительно*» ставится, если обучающийся не отвечает на поставленные вопросы.

#### **Контрольная работа**

Оценивается не только глубина знаний поставленных вопросов, но и умение изложить письменно.

*Критерии оценивания:* последовательность, полнота, логичность изложения, анализ различных точек зрения, самостоятельное обобщение материала. Изложение материала без фактических ошибок.

Оценка «отлично» ставится в случае, когда соблюдены все критерии.

Оценка «хорошо» ставится, если обучающийся твердо знает материал, грамотно и по существу излагает его, знает практическую базу, но допускает несущественные погрешности.

Оценка «удовлетворительно» ставится, если обучающийся освоил только основной материал, но не знает отдельных деталей, допускает неточности, недостаточно правильные формулировки, нарушает последовательность в изложении материала, затрудняется с ответами, показывает отсутствие должной связи между анализом, аргументацией и выводами.

Оценка «неудовлетворительно» ставится, если обучающийся не отвечает на поставленные вопросы.

### 3.2. Оценочные материалы для проведения промежуточной аттестации

#### 3.2.1. Критерии оценки результатов обучения по дисциплине (модулю)

Шкала оценивания	Результаты обучения	Показатели оценивания результатов обучения
ОТЛИЧНО	Знает:	<ul style="list-style-type: none"> <li>- обучающийся глубоко и всесторонне усвоил материал, уверенно, логично, последовательно и грамотно его излагает, опираясь на знания основной и дополнительной литературы,</li> <li>- на основе системных научных знаний делает квалифицированные выводы и обобщения, свободно оперирует категориями и понятиями.</li> </ul>
	Умеет:	<ul style="list-style-type: none"> <li>- обучающийся умеет самостоятельно и правильно решать учебно-профессиональные задачи или задания, уверенно, логично, последовательно и аргументировано излагать свое решение, используя научные понятия, ссылаясь на нормативную базу.</li> </ul>
	Владет:	<ul style="list-style-type: none"> <li>- обучающийся владеет рациональными методами (с использованием рациональных методик) решения сложных профессиональных задач, представленных деловыми играми, кейсами и т.д.;</li> <li>При решении продемонстрировал навыки</li> <li>- выделения главного,</li> <li>- связкой теоретических положений с требованиями руководящих документов,</li> <li>- изложения мыслей в логической последовательности,</li> <li>- самостоятельного анализа факты, событий, явлений, процессов в их взаимосвязи и диалектическом развитии.</li> </ul>
ХОРОШО	Знает:	<ul style="list-style-type: none"> <li>- обучающийся твердо усвоил материал, достаточно грамотно его излагает, опираясь на знания основной и дополнительной литературы,</li> <li>- затрудняется в формулировании квалифицированных выводов и обобщений, оперирует категориями и понятиями, но не всегда правильно их верифицирует.</li> </ul>
	Умеет:	<ul style="list-style-type: none"> <li>- обучающийся умеет самостоятельно и в основном правильно решать учебно-профессиональные задачи или задания, уверенно, логично, последовательно и аргументировано излагать свое решение, не в полной мере используя научные понятия и ссылки на нормативную базу.</li> </ul>
	Владет:	<ul style="list-style-type: none"> <li>- обучающийся в целом владеет рациональными методами решения сложных профессиональных задач, представленных деловыми играми, кейсами и т.д.;</li> <li>При решении смог продемонстрировать достаточность, но не глубинность навыков,</li> <li>- выделения главного,</li> <li>- изложения мыслей в логической последовательности,</li> <li>- связки теоретических положений с требованиями руководящих документов,</li> <li>- самостоятельного анализа факты, событий, явлений, процессов в их взаимосвязи и диалектическом развитии.</li> </ul>
УДОВЛЕТВОРИТЕЛЬНО	Знает:	<ul style="list-style-type: none"> <li>- обучающийся ориентируется в материале, однако затрудняется в его изложении;</li> <li>- показывает недостаточность знаний основной и дополнительной литературы;</li> <li>- слабо аргументирует научные положения;</li> </ul>

		- практически не способен сформулировать выводы и обобщения; - частично владеет системой понятий.
	Умеет:	- обучающийся в основном умеет решить учебно-профессиональную задачу или задание, но допускает ошибки, слабо аргументирует свое решение, недостаточно использует научные понятия и руководящие документы.
	Владеет:	- обучающийся владеет некоторыми рациональными методами решения сложных профессиональных задач, представленных деловыми играми, кейсами и т.д.; При решении продемонстрировал недостаточность навыков - выделения главного, - изложения мыслей в логической последовательности, - связи теоретических положений с требованиями руководящих документов, - самостоятельного анализа факты, событий, явлений, процессов в их взаимосвязи и диалектическом развитии.
НЕУДОВЛЕТВОРИТЕЛЬНО	Знает:	- обучающийся не усвоил значительной части материала; - не может аргументировать научные положения; - не формулирует квалифицированных выводов и обобщений; - не владеет системой понятий.
	Умеет:	обучающийся не показал умение решать учебно-профессиональную задачу или задание.
	Владеет:	не выполнены требования, предъявляемые к навыкам, оцениваемым «удовлетворительно».

### 3.2.2. Контрольные задания и/или иные материалы для проведения промежуточной аттестации

#### Список вопросов для устных ответов

1. Постановка задачи машинного обучения. Обучение с учителем и без учителя. Типы признаков
2. Типы задач. Линейные модели
3. Примеры задач обучения без учителя (3-5 примеров)
4. Примеры задач обучения с учителем (3-5 примеров)
5. Функционал качества. Минимизация эмпирического риска
6. Переобучение. Методы валидации моделей. Кросс-валидация
7. Метрические методы классификации. Виды расстояний
8. Метод ближайшего соседа. Метод k ближайших соседей. Взвешенный метод k ближайших соседей
9. Отбор эталонных объектов. Понятие отступа. Классификация объектов
10. Регрессия по соседним объектам. Окно Парзена
11. Линейная регрессия. Постановка задачи. Матричная формулировка. Точное решение
12. Регуляризация в линейной регрессии. Виды регуляризаторов
13. Гребневая регрессия. Лассо-регрессия
14. Итеративная версия решения задачи линейной регрессии без регуляризации и с регуляризацией. Метод градиентного спуска
15. Вероятностная модель данных. Максимум правдоподобия
16. Линейный классификатор. Отступы. Функционал качества
17. Метод стохастического градиента (SGD). Преимущества и недостатки
18. Методы улучшения сходимости SGD (метод моментов, метод Нестерова)
19. Методы улучшения сходимости SGD (AdaGrad, RMSProp, Adam)
20. Логистическая регрессия. Простой байесовский классификатор. Сигмоидальная функция
21. Метод опорных векторов (SVM). Постановка задачи для линейно разделимой выборки без выбросов
22. Отступы в SVM. Учет выбросов. Формулировка теоремы Куна-Таккера. Двойственная задача
23. Классификация объектов в SVM. Постановка задачи через двойственные переменные  $\lambda$

24. Нелинейные ядра в методе SVM. Примеры ядер. Способы их построения
25. Основные метрики качества алгоритмов. ROC-кривые
26. Методы кластеризации. Типы кластерных структур
27. Функционал качества кластеризации.
28. EM-алгоритм в кластерном анализе
29. Метод k-средних
30. Иерархическая кластеризация. Формула Ланса-Уильямса
31. Быстрая агломеративная кластеризация.

### 3.2.3. Методические материалы, определяющие процедуры оценивания знаний, умений, навыков в ходе промежуточной аттестации

#### Процедура оценивания знаний (устный ответ)

Предел длительности	10 минут
Предлагаемое количество заданий	2
Последовательность выборки вопросов из каждого раздела	Случайная
Критерии оценки	<ul style="list-style-type: none"> <li>- требуемый объем и структура</li> <li>- изложение материала без фактических ошибок</li> <li>- логика изложения</li> <li>- использование соответствующей терминологии</li> <li>- стиль речи и культура речи</li> <li>- подбор примеров из научной литературы и практики</li> </ul>
«5» если	требования к ответу выполнены в полном объеме
«4» если	в целом выполнены требования к ответу, однако есть небольшие неточности в изложении некоторых вопросов
«3» если	требования выполнены частично – не выдержан объем, есть фактические ошибки, нарушена логика изложения, недостаточно используется соответствующая терминология

## 4. Учебно-методическое и материально-техническое обеспечение дисциплины (модуля)

### 4.1. Электронные учебные издания

1. Л. Рамальо Python. К вершинам мастерства: ДМК Пресс, 2016//ЭБС издательства «Лань» <https://e.lanbook.com/book/93273>.
2. Е.А. Черткова Статистика. Автоматизация обработки информации: Учебное пособие Юрайт, 2018//ЭБС Юрайт <https://www.bibli-online.ru/book/0CBA0F5B-1227-46F3-8C8E-D9BAB4AC306A>.
3. Хахаев И. А. Практикум по алгоритмизации и программированию на Python: курс М.: Национальный Открытый Университет «ИНТУИТ», 2016//ЭБС «Университетская библиотека online» <http://biblioclub.ru/index.php?page=book&id=429256>.

### 4.2. Электронные образовательные ресурсы

1. Электронная библиотечная система «ЭБС ЮРАЙТ» Biblio-online.ru (ЭБС «Юрайт») [Электронный ресурс]. – URL: <https://urait.ru/>.
2. Электронно-библиотечная система ZNANIUM [Электронный ресурс]. – URL: <https://znanium.com/>.
3. Электронная библиотечная система «Консультант студента» [Электронный ресурс]. – URL: <https://www.studentlibrary.ru/>.
4. e-Library.ru: Научная электронная библиотека [Электронный ресурс]. – URL: <http://elibrary.ru/>.
5. Научная электронная библиотека «КиберЛенинка» [Электронный ресурс]. – URL: <http://cyberleninka.ru/>.
6. Информационная система «Единое окно доступа к образовательным ресурсам» [Электронный ресурс]. – URL: <http://window.edu.ru/>.
7. Федеральный центр информационно-образовательных ресурсов [Электронный ресурс]. – URL: <http://fcior.edu.ru/>.



#### 4.3. Современные профессиональные базы данных и информационные справочные системы

Обучающимся обеспечен доступ (удаленный доступ) к ниже следующим современным профессиональным базам данных и информационным справочным системам:

1. Словари и энциклопедии на Академике [Электронный ресурс]. – URL: <http://dic.academic.ru>.
2. Система информационно-правового обеспечения «Гарант» [Электронный ресурс]. – <http://www.garant.ru/>.
3. База данных Института философии РАН: Философские ресурсы: Текстовые ресурсы: <https://iphras.ru/page52248384.htm>.

#### 4.4. Комплект лицензионного и свободно распространяемого программного обеспечения, в том числе отечественного производства

1. Лицензионное программное обеспечение: операционная система Microsoft Windows, пакет офисных приложений Microsoft Office.
2. Свободно распространяемое программное обеспечение: свободные пакеты офисных приложений Apache Open Office, LibreOffice.
3. Программное обеспечение отечественного производства: справочно-правовая система «Гарант» (Электронный периодический справочник «Система ГАРАНТ»), образовательная платформа ЮРАЙТ (Электронная библиотечная система «ЭБС ЮРАЙТ» Biblio-online.ru (ЭБС «Юрайт»)), электронно-библиотечная система ZNANIUM, электронная библиотечная система «Консультант студента».

#### 4.5. Оборудование и технические средства обучения

Для реализации дисциплины (модуля) используются учебные аудитории для проведения учебных занятий, которые оснащены оборудованием и техническими средствами обучения, и помещения для самостоятельной работы обучающихся, которые оснащены компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечены доступом в электронную информационно-образовательную среду РХТУ им. Д.И. Менделеева. Допускается замена оборудования его виртуальными аналогами.

Наименование учебных аудиторий для проведения учебных занятий и помещений для самостоятельной работы*	Оснащенность учебных аудиторий для проведения учебных занятий и помещений для самостоятельной работы оборудованием и техническими средствами обучения
Учебные аудитории для проведения учебных занятий	Учебная аудитория укомплектована специализированной мебелью, отвечающей всем установленным нормам и требованиям, оборудованием и техническими средствами обучения (мобильное мультимедийное оборудование).
Помещение для самостоятельной работы	Помещение оснащено компьютерной техникой с возможностью подключения к сети «Интернет» и обеспечением доступа в электронную информационно-образовательную среду РХТУ им. Д.И. Менделеева и к ЭБС.

\* Номер конкретной аудитории указан в приказе об аудиторном фонде, расписании учебных занятий и расписании промежуточной аттестации.