

## **ОЦЕНОЧНЫЕ МАТЕРИАЛЫ**

**Код модуля**  
М.1.6

**Модуль**  
Машинное обучение

**Екатеринбург, 2021**

Оценочные материалы по модулю составлены авторами:

<b>№ п/п</b>	<b>Фамилия Имя Отчество</b>	<b>Ученая степень, ученое звание</b>	<b>Должность</b>	<b>Подразделение</b>
1	Долганов Антон Юрьевич	кандидат технических наук, без учёного звания	Доцент	Кафедра радиоэлектроники и телекоммуникаций, ИРИТ-РТФ, УрФУ
2	Патракова Екатерина Сергеевна	-	Ведущий исследователь данных	ПАО Сбербанк

## 1. СТРУКТУРА И ОБЪЕМ МОДУЛЯ Машинное обучение

№ п/п	Перечень дисциплин модуля в последовательности их освоения	Объем дисциплин модуля и всего модуля в зачетных единицах и часах	Форма итоговой промежуточной аттестации по дисциплинам модуля и в целом по модулю
1.	Машинное обучение	12 з.е. / 432 ч.	Экзамен (1-3 семестры)
<b>ИТОГО по модулю:</b>		<b>12 з.е. / 432 ч.</b>	

## 2. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО МОДУЛЮ

Не предусмотрено.

**Раздел 3. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПО ДИСЦИПЛИНЕ 1**  
**МАШИННОЕ ОБУЧЕНИЕ**

**Модуль М.1.6 Машинное обучение**

Оценочные материалы составлены автором(ами):

<b>№ п/п</b>	<b>Фамилия Имя Отчество</b>	<b>Ученая степень, ученое звание</b>	<b>Должность</b>	<b>Подразделение</b>
1	Долганов Антон Юрьевич	кандидат технических наук, без учёного звания	Доцент	Кафедра радиоэлектроники и телекоммуникаций, ИРИТ-РТФ, УрФУ
2	Патракова Екатерина Сергеевна	-	Ведущий исследователь данных	ПАО Сбербанк

**1. ПЛАНИРУЕМЫЕ РЕЗУЛЬТАТЫ ОБУЧЕНИЯ (ИНДИКАТОРЫ) ПО ДИСЦИПЛИНЕ МОДУЛЯ Машинное обучение**

Таблица 1

<b>Код и наименование компетенций, формируемые с участием дисциплины</b>	<b>Планируемые результаты обучения (индикаторы)</b>	<b>Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине</b>
<b>1</b>	<b>3</b>	<b>4</b>
<p>ОПК-2 - Способен самостоятельно ставить, формализовывать и решать задачи, относящиеся к профессиональной деятельности, используя методы моделирования и математического анализа</p>	<p>З-1 - Сделать обзор основных методов моделирования и математического анализа, применимых для формализации и решения задач профессиональной деятельности</p> <p>З-2 - Характеризовать сферы применения и возможности пакетов прикладных программ для решения задач профессиональной деятельности</p> <p>У-1 - Самостоятельно сформулировать задачу области профессиональной деятельности, решение которой требует использования методов моделирования и математического анализа</p> <p>У-2 - Использовать методы моделирования и математического анализа, в том числе с использованием пакетов прикладных программ для решения задач профессиональной деятельности</p>	<p>1) Контрольная работа №1, 2, 5 2) Домашняя работа №1 3) Экзамен</p>

	<p>П-1 - Решать самостоятельно сформулированные практические задачи, относящиеся к профессиональной деятельности методами моделирования и математического анализа, в том числе с использованием пакетов прикладных программ</p> <p>Д-1 - Проявлять ответственность и настойчивость в достижении цели</p>	
<p>ОПК-5 - Способен планировать, организовывать и контролировать работы по созданию, установке и модернизации технологического оборудования и технологических процессов в сфере своей профессиональной деятельности</p>	<p>З-4 - Показать возможности использования цифровых технологий (создание цифровых двойников) для оптимизации работы по созданию, установке и модернизации технологического оборудования, технологических процессов и информационных систем</p> <p>У-4 - Использовать при необходимости техники цифрового моделирования при выполнении работ по созданию, установке и модернизации технологического оборудования, технологических процессов и информационных систем</p> <p>П-1 - Самостоятельно составить план работ в целом по этапам создания, установки и модернизации технологического оборудования, технологических процессов и</p>	<p>1) Контрольная работа №3, 4 2) Домашняя работа №2, 3 3) Экзамен</p>

	информационных систем либо отдельных этапов этой работы	
<p>ПК-2 - Способен разрабатывать и применять математические модели процессов и объектов при решении задач анализа и синтеза распределенных информационных систем и систем поддержки принятия решений</p>	<p>З-1 - Интерпретировать результаты применения математических моделей процессов и объектов при решении задач анализа и синтеза распределенных информационных систем и систем поддержки принятия решений</p> <p>З-3 - Сделать обзор методов математического моделирования и анализа данных</p> <p>У-1 - Определять оптимальные математические модели для решения задач анализа и синтеза распределенных информационных систем и систем поддержки принятия решений в зависимости от заданных условий</p> <p>П-1 - Иметь практический опыт применения математического моделирования и анализа данных</p>	<p>1) Контрольная работа №1, 2, 3, 4, 5</p> <p>2) Домашняя работа №1, 2, 3</p> <p>3) Экзамен</p>

## 2. ВИДЫ САМОСТОЯТЕЛЬНОЙ РАБОТЫ СТУДЕНТОВ, ВКЛЮЧАЯ МЕРОПРИЯТИЯ ТЕКУЩЕЙ АТТЕСТАЦИИ

### 2.1. Распределение объема времени по видам учебной работы

Таблица 2

№ п/ п	Наименование дисциплины модуля Спортивный анализ данных	Объем времени, отведенный на освоение дисциплины модуля								
		Аудиторные занятия, час.				Промежуточная аттестация (форма итогового контроля)	Контактная работа (час.)	Самостоятельная работа студента, включая текущую аттестацию (час.)	Всего по дисциплине	
		Занятия лекцион ного типа	Практиче ские работы	Лаборатор ные работы	Всего				Час.	Зач. ед.
1	2	3	4	5	6	7	8	9	10	11
1.	Машинное обучение	54	0	72	126	Экзамен (1-3 семестр)	151,89	280,11	432	12

2.1. Виды СРС, количество и объем времени на контрольно-оценочные мероприятия СРС по дисциплине

Таблица 3

№ п/п	Вид самостоятельной работы студента по дисциплине модуля	Количество контрольно-оценочных мероприятий СРС	Объем контрольно-оценочных мероприятий СРС (час.)
1.	Подготовка к аудиторным занятиям и мероприятиям текущего контроля: лекционным, практическим занятиям.		54 час.
2.	Выполнение и оформление мероприятий текущего контроля:		
2.1	Контрольная работа	5	25 час.
2.2	Домашняя работа	5	25 час.
3.	Подготовка к экзамену	3	36 час.
4.	Самостоятельное изучение материала		140,11 час.
Итого на СРС по дисциплине:			280,11 час.

3. ПРОЦЕДУРЫ КОНТРОЛЯ И ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ В РАМКАХ ТЕКУЩЕЙ И ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ ПО ДИСЦИПЛИНЕ МОДУЛЯ В БАЛЬНО-РЕЙТИНГОВОЙ СИСТЕМЕ (ТЕХНОЛОГИЧЕСКАЯ КАРТА БРС)

Процедуры текущей и промежуточной аттестации по дисциплине

1 семестр		
<b>1. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.5</b>		
Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>Домашняя работа № 1</i>	<i>1 сем., 4 нед.</i>	25
<i>Домашняя работа № 2</i>	<i>1 сем., 10 нед.</i>	25
<i>Контрольная работа № 1</i>	<i>1 сем., 6 нед.</i>	25
<i>Контрольная работа № 2</i>	<i>1 сем., 12 нед.</i>	25
<b>Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5</b>		
<b>Промежуточная аттестация по лекциям – Экзамен</b>		
<b>Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.5</b>		

<b>2. Практические/семинарские занятия: Не предусмотрены</b>		
<b>3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий – 0.5</b>		
Текущая аттестация на лабораторных занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>Выполнение и оформление лабораторных работ</i>	<i>1 сем., 1-18 нед.</i>	<i>100</i>
<b>Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям - 1</b>		
<b>Промежуточная аттестация по лабораторным занятиям – не предусмотрено</b>		
<b>Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – 0</b>		

<b>2 семестр</b>		
<b>1. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.5</b>		
Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>Домашняя работа № 3</i>	<i>2 сем., 4 нед.</i>	<i>25</i>
<i>Домашняя работа № 4</i>	<i>2 сем., 10 нед.</i>	<i>25</i>
<i>Контрольная работа № 3</i>	<i>2 сем., 6 нед.</i>	<i>25</i>
<i>Контрольная работа № 4</i>	<i>2 сем., 12 нед.</i>	<i>25</i>
<b>Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5</b>		
<b>Промежуточная аттестация по лекциям – Экзамен</b>		
<b>Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.5</b>		
<b>3. Практические/семинарские занятия: Не предусмотрены</b>		
<b>3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий – 0.5</b>		
Текущая аттестация на лабораторных занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>Выполнение и оформление лабораторных работ</i>	<i>2 сем., 1-18 нед.</i>	<i>100</i>
<b>Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям - 1</b>		
<b>Промежуточная аттестация по лабораторным занятиям – не предусмотрено</b>		
<b>Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – 0</b>		

<b>3 семестр</b>		
<b>1. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.5</b>		
<b>Текущая аттестация на лекциях</b>	<b>Сроки – семестр, учебная неделя</b>	<b>Максимальная оценка в баллах</b>
<i>Домашняя работа № 5</i>	<i>3 сем., 6 нед.</i>	<i>50</i>
<i>Контрольная работа № 5</i>	<i>3 сем., 8 нед.</i>	<i>50</i>
<b>Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.5</b>		
<b>Промежуточная аттестация по лекциям – Экзамен</b>		
<b>Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.5</b>		
<b>4. Практические/семинарские занятия: Не предусмотрены</b>		
<b>3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий – 0.5</b>		
<b>Текущая аттестация на лабораторных занятиях</b>	<b>Сроки – семестр, учебная неделя</b>	<b>Максимальная оценка в баллах</b>
<i>Выполнение и оформление лабораторных работ</i>	<i>3 сем., 1-18 нед.</i>	<i>100</i>
<b>Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям - 1</b>		
<b>Промежуточная аттестация по лабораторным занятиям – не предусмотрено</b>		
<b>Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – 0</b>		

#### **4. КРИТЕРИИ И УРОВНИ ОЦЕНИВАНИЯ РЕЗУЛЬТАТОВ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ МОДУЛЯ**

4.1. В рамках БРС применяются утвержденные на кафедре/институте критерии (признаки) оценивания достижений студентов по дисциплине модуля (табл. 4) в рамках контрольно-оценочных мероприятий на соответствие указанным в табл.1 результатам обучения (индикаторам).

Таблица 4

##### **Критерии оценивания учебных достижений обучающихся**

<b>Результаты обучения</b>	<b>Критерии оценивания учебных достижений, обучающихся на соответствие результатам обучения/индикаторам</b>
Знания	Студент демонстрирует знания и понимание в области изучения на уровне указанных индикаторов и необходимые для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.
Умения	Студент может применять свои знания и понимание в контекстах, представленных в оценочных заданиях, демонстрирует освоение умений на уровне указанных индикаторов и необходимых для продолжения обучения и/или выполнения трудовых функций и действий, связанных с профессиональной деятельностью.

Опыт /владение	Студент демонстрирует опыт в области изучения на уровне указанных индикаторов.
Личностные качества	Студент демонстрирует ответственность в освоении результатов обучения на уровне запланированных индикаторов. Студент способен выносить суждения, делать оценки и формулировать выводы в области изучения. Студент может сообщать преподавателю и коллегам своего уровня собственное понимание и умения в области изучения.

4.2. Для оценивания уровня выполнения критериев (уровня достижений обучающихся при проведении контрольно-оценочных мероприятий по дисциплине модуля) используется универсальная шкала (табл. 5).

Таблица 5

**Шкала оценивания достижения результатов обучения (индикаторов) по уровням**

<b>Характеристика уровней достижения результатов обучения (индикаторов)</b>				
<b>№ п/п</b>	<b>Содержание уровня выполнения критерия оценивания результатов обучения (выполненное оценочное задание)</b>	<b>Шкала оценивания</b>		
		<b>Традиционная характеристика уровня</b>		<b>Качественная характеристика уровня</b>
1.	Результаты обучения (индикаторы) достигнуты в полном объеме, замечаний нет	Отлично (80-100 баллов)	Зачтено	Высокий (В)
2.	Результаты обучения (индикаторы) в целом достигнуты, имеются замечания, которые не требуют обязательного устранения	Хорошо (60-79 баллов)		Средний (С)
3.	Результаты обучения (индикаторы) достигнуты не в полной мере, есть замечания	Удовлетворительно (40-59 баллов)		Пороговый (П)
4.	Освоение результатов обучения не соответствует индикаторам, имеются существенные ошибки и замечания, требуется доработка	Неудовлетворительно (менее 40 баллов)	Не зачтено	Недостаточный (Н)
5.	Результат обучения не достигнут, задание не выполнено	Недостаточно свидетельств для оценивания		Нет результата

**5. СОДЕРЖАНИЕ КОНТРОЛЬНО-ОЦЕНОЧНЫХ МЕРОПРИЯТИЙ ПО ДИСЦИПЛИНЕ МОДУЛЯ**

Задания по контрольно-оценочным мероприятиям в рамках текущей и промежуточной аттестации должны обеспечивать освоение и достижение результатов обучения (индикаторов) и предметного содержания дисциплины на соответствующем уровне.

**5.1. Описание контрольно-оценочных мероприятий и средств текущего контроля по дисциплине модуля**

### 5.1.1. Практические занятия

*Не предусмотрено.*

### 5.1.2. Лабораторные занятия

Номер работы	Примерный перечень тем лабораторных работ
1	Обработка сырых данных.
2	Алгоритмы обучения бинарной классификации.
3	Ансамбли. Генерация новых признаков из данных.
4	Кластеризация.
5	Работа с текстами. Векторизация текстов и классификация.
6	Бинарный поиск для текстов. Обучение собственных эмбедингов.
7	Основы TensorFlow.
8	Нейронные сети для классификации объектов.
9	Распознавание объектов на изображении.
10	Классификация на символьных сверточных сетях и рекуррентных нейронных сетях.
11	Выделение именованных сущностей из текстов.
12	Поиск синтаксических связей между выделенными сущностями.
13	Fine tuning предобученной модели на текстах другого домена.
14	Нейронные сети для генерации текстов.

### 5.1.3. Курсовая работа / Курсовой проект

*Не предусмотрено.*

### 5.1.4. Контрольная работа

**Примерная тематика** контрольных работ:

1. Классификация и кластеризация
2. Обработка естественного языка
3. Нейронные сети
4. Выделение именованных сущностей и синтаксический парсинг
5. Чат-боты и seq2seq

**Примерные задания** в составе *Контрольной работы №1*. Классификация и кластеризация:

1. Выберите **верные утверждения** про площадь под ROC-кривой (AUC-ROC)
  - a) Площадь под ROC-кривой для идеального алгоритма зависит от доли положительных объектов в выборке
  - b) ROC-кривая строится в осях "доля верных отрицательных классификаций" и "доля верных положительных классификаций"
  - c) ROC-кривая строится в осях "доля ошибочных положительных классификаций" и "доля верных положительных классификаций"
  - d) Площадь под ROC-кривой для идеального алгоритма равна единице
2. Как определяется ответ решающего дерева в задаче классификации для заданного объекта?

- a) Ответ определяется на основе значения в любой из вершин дерева
  - b) Ответ определяется на основе суммы значений из всех вершин, по которым прошел данный объект
  - c) По значению в листе, который попал в данный объект
3. Для чего отдельные деревья в случайном лесу обучаются по подвыборкам объектов?
- a) Чтобы деревья не были переобученными
  - b) Чтобы сократить время обучения
  - c) Чтобы деревья не были одинаковыми
4. В чем заключается процесс обучения алгоритма k ближайших соседей?
- a) настройка весов при объектах обучающей выборки
  - b) запоминание всех объектов обучающей выборки
  - c) подсчет попарных расстояний между объектами обучающей выборки
5. Что такое дисперсия у модели?
6. Как можно провалидировать\проверить вашу модель?
7. Что произойдет если линейной регрессии на вход подать три колонки  $x$ ,  $y$ ,  $z$ , где  $z=x+y$ ?

**Примерные задания** в составе *Контрольной работы №2*. Обработка естественного языка:

1. Какой из нижеприведенных методов позволяет привести слово к его нормальной форме?
  - Стемминг
  - Сегментация
  - Лемматизация
  - Токенизация
2. 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
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. Если имеется корпус из большого количества текстов про 1000 великих людей России. У какого слова будет больший idf вес: у слова «родился» или у «Лермонтов»?
  - родился
  - Лермонтов
5. Какое из следующих утверждений верно для модели Word2Vec?
  - Архитектура word2vec состоит всего из двух слоев - continuous bag-of-words и skip-gram
  - Continuous bag-of-words (CBOW) - это модель рекуррентной нейронной сети
  - И CBOW, и Skip-gram являются моделями неглубокой нейронной сети
  - Все вышеперечисленное

6. Какое количество параметров в матрице эмбеддингов, если размер эмбеддингов 300 и размер словаря 1234?
7. Какое из следующих расстояний позволяет посчитать разницу между текстовыми строками?
  - Косинусное
  - Левенштейна
  - Евклидово
  - Жаккара
8. Кратко опишите отличие работы fasttext от word2vec.
9. Кратко опишите архитектуру нейросети на основе RNN, решающую задачу классификации
10. Опишите 3 основных этапа обучения ULMFit.

**Примерные задания** в составе *Контрольной работы №3*. Нейронные сети:

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. Что такое градиентный спуск? В чем отличие стохастического градиентного спуска от обычного градиентного спуска?
6. Что такое затухание градиента?
7. Какая функция активации позволяет решить проблему затухания градиента?
  - sigmoid
  - relu
  - tanh
8. Что будет если мы вначале зададим веса нейронной сети равными 0?
9. Какие регуляризации есть в нейронных сетях?
10. Пусть сеть принимает на вход пять признаков, содержит четыре нейрона во внутреннем слое и один нейрон в выходном слое. Сколько у нее параметров? Не забывайте про свободный член в каждом слое

**Примерные задания** в составе *Контрольной работы №4*. Выделение именованных сущностей и синтаксический парсинг:

1. Перечислите основные сложности, с которыми можно столкнуться при решении задачи NER
2. Какую проблему частично позволяет решить использование CRF в задаче NER
3. Кратко опишите архитектуру на основе LSTM, которую используют для решения задачи NER
4. Опишите 2 основных варианта символического представления токена
5. Какие 3 основные операции используются при обучении transition-based парсеров

6. Что transition-based парсеры восстанавливают в процессе обучения вместо самих деревьев
7. Кратко опишите архитектуру парсера Чена-Маннинга (нейросетевая имплементация transition-based парсера)
8. Опишите, как работает beam search по методу early updates

**Примерные задания** в составе *Контрольной работы №5*. Чат-боты и seq2seq:

1. Кратко опишите архитектуру базового seq2seq (без механизма внимания).
2. Какую проблему базового seq2seq и как решает механизм внимания?
3. Какой вид attention чаще всего используется в современных моделях?
4. Как по корпусу получается BPE и какие преимущества дает?
5. Чем отличается scaled dot product attention трансформера от обычного dot product attention?
6. Опишите устройство multi-head attention. Что такое его головы и зачем они нужны?

#### 5.1.5. Домашняя работа

**Примерная тематика** домашних работ:

1. Многоклассовая классификация интереса пользователей к объявлениям о продаже квартир
2. Классификация текстов диалогов из чатов смешанных тематик
3. Распознавание изображений с помощью сверточных нейронных сетей.
4. Классификация на символьных сверточных сетях и рекуррентных нейронных сетях.
5. Генерация текстов

**Примерные задания** в составе домашних работ:

*Домашняя работа №1:*

Соревнование на kaggle:

<https://www.kaggle.com/c/2021-homework2-multiclass-classification>

Ноутбук с заданиями:

<https://www.kaggle.com/patrakova/tasks-hw2>

1. Предобработка и генерация признаков

Для выполнения задания попробуйте придумать признаки на основе имеющихся данных, для работы с текстовыми полями вам может пригодиться LabelEncoder/OneHotEncoder, re.sub(), re.findall()

2. Тестирование различных алгоритмов

Для выполнения этой части задания разбейте размеченные данные на train и test и напишите автоматическое тестирование разных алгоритмов классификации в цикле с вычислением метрики f1\_score('macro')

3. Сделать кластеризацию данных на числовых признаках, кластер добавить к данным в качестве признака и обучить на этом классификатор. Сравнить метрику с кластером-признаком и без.
4. Загрузить свое лучшее решение в соревнование.

*Домашняя работа №2:*

Соревнование на kaggle:

<https://www.kaggle.com/c/2021-homework3-texts>

1. Разработать классификатор для текстов из чатов знакомств, чатов по python и чатов по ML.
2. Напишите свой векторизатор на основе word2vec. Чтобы получить вектор текста, нужно брать word2vec-вектора отдельных слов, суммировать в один вектор и затем этот вектор нормализовать.
3. Сделать классификацию текстов с использованием предобученных, построенных по новостных заголовкам, эмбедингах в рамках проекта RusVectōrēs.

*Домашняя работа №3:*

Разработайте сверточную нейронную сеть для распознавания объектов на изображениях из набора данных CIFAR-10. Доля правильных ответов (метрика accuracy) на тестовом наборе данных должна быть не менее 0,87.

Соревнование на kaggle:

<https://www.kaggle.com/t/5c22e3a159a546f085be12ec9b265245>

Ноутбук с заданиями:

<https://www.kaggle.com/lkatran/base-line-4-4>

Можно использовать любые предварительно обученные нейронные сети.

*Домашняя работа №4:*

Разработать два варианта нейронных сетей посимвольно и пословесно для классификации твитов на позитивные и негативные. Можно использовать предобученные эмбединги.

*Домашняя работа №5:*

Найти массив текстов и разработать нейронную сеть для генерации подобных текстов на символьных или словесных сверточных сетях и рекуррентных нейронных сетях. Для улучшения качества сгенерированного текста использовать attention.

1.

**5.1.6. Расчетная работа / Расчетно-графическая работа**

*Не предусмотрено.*

**5.1.7. Реферат / эссе / творческая работа**

*Не предусмотрено.*

**5.1.8. Проектная работа**

*Не предусмотрено.*

**5.1.9. Деловая (ролевая) игра / Дебаты / Дискуссия / Круглый стол**

*Не предусмотрено.*

**5.1.10. Кейс-анализ**

*Не предусмотрено.*

**5.2. Описание контрольно-оценочных мероприятий промежуточного контроля по дисциплине модуля**

### 5.2.1. Экзамен в форме независимого тестового контроля (НТК)

НТК по дисциплине модуля не проводится.

**5.2.2. Экзамен в традиционной форме** (устные ответы на вопросы экзаменационных билетов):

1. Чем отличается задача классификации от кластеризации?
2. Какие метрики используются в задаче классификации. Что такое precision и recall?
3. В каких случаях лучше использовать precision и recall?
4. Что такое не сбалансированная выборка?
5. Как обычно работают с многоклассовой классификацией? Как применить Логистическую регрессию к многоклассовой классификации? One vs rest, one vs all, что дольше\больше?
6. Алгоритмы машинного обучения.
7. Ансамбли: стекинг, бустинг, блэндинг.
8. В чем будет разница в метриках? Как считаются micro и macro усреднение?
9. Какие методы есть для решения задачи кластеризации? Метрики для задач кластеризации?
10. Этапы построения модели.
11. Методы отбора фичей? В каких ситуациях применяется?
12. Что такое регуляризация и чем она полезна. Чем отличается L1 от L2 регуляризации?
13. Что такое overfitting? Как с ним бороться?
14. Обучение нейронной сети. Алгоритм обратного распространения ошибки.
15. Модификации градиентного спуска: градиентный спуск с инерцией.
16. Модификации градиентного спуска: ускоренный градиентный спуск Нестерова.
17. Модификации градиентного спуска: адаптивный градиентный спуск.
18. Модификации градиентного спуска: метод адаптивного скользящего среднего.
19. Модификации градиентного спуска: метод адаптивного шага обучения.
20. Обработка текста.
21. Методы токенизации текста.
22. Методы векторизации текста.
23. Алгоритм векторизации текста tf-idf.
24. Определение дистрибутивной гипотезы
25. Какие метрики чаще всего используют для вычисления близости между векторами текстов?
26. Что такое эмбединги? Виды эмбедингов.
27. Отличия в обучении алгоритмов cbow и skip gram для word2vec.
28. Что такое negative sampling?
29. Плюсы и минусы word2vec. В чем разница word2vec и fasttext?