

**ОЦЕНОЧНЫЕ МАТЕРИАЛЫ
ПО ДИСЦИПЛИНЕ**
Машинное обучение

Код модуля
1156462(1)

Модуль
Машинное обучение

Екатеринбург

Оценочные материалы составлены автором(ами):

№ п/п	Фамилия, имя, отчество	Ученая степень, ученое звание	Должность	Подразделение
1	Хлопин Дмитрий Валерьевич	кандидат физико-математических наук, без ученого звания	Доцент	прикладной математики и механики

Согласовано:

Управление образовательных программ

Ю.Д. Маева

Авторы:

- **Хлопин Дмитрий Валерьевич, Доцент, прикладной математики и механики**

1. СТРУКТУРА И ОБЪЕМ ДИСЦИПЛИНЫ **Машинное обучение**

1.	Объем дисциплины в зачетных единицах	4	
2.	Виды аудиторных занятий	Лекции Лабораторные занятия	
3.	Промежуточная аттестация	Экзамен	
4.	Текущая аттестация	Контрольная работа	2
		Домашняя работа	4

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

Индикатор – это признак / сигнал/ маркер, который показывает, на каком уровне обучающийся должен освоить результаты обучения и их предъявление должно подтвердить факт освоения предметного содержания данной дисциплины, указанного в табл. 1.3 РПМ-РПД.

Таблица 1

Код и наименование компетенции	Планируемые результаты обучения (индикаторы)	Контрольно-оценочные средства для оценивания достижения результата обучения по дисциплине
1	2	3
ОПК-4 -Способен решать задачи профессиональной деятельности с использованием информационно-коммуникационных технологий и с учетом основных требований информационной безопасности	У-1 - Обосновывать выбор современных ИТ-технологий для сбора, анализа, обработки и представления информации по профилю деятельности У-2 - Выбирать безопасные информационно-коммуникативные технологии для эффективного решения задач профессиональной деятельности	Домашняя работа № 1 Домашняя работа № 2 Домашняя работа № 3 Домашняя работа № 4 Контрольная работа № 1 Контрольная работа № 2 Лабораторные занятия Лекции Экзамен
ОПК-2 -Способен проводить под научным руководством исследования на основе современных методов в конкретной области	Д-2 - Проявлять заинтересованность в содержании и результатах исследовательской работы З-1 - Демонстрировать понимание теоретических основ методов, используемых для проведения научных	Домашняя работа № 1 Домашняя работа № 2 Домашняя работа № 3 Домашняя работа № 4 Контрольная работа № 1 Контрольная работа № 2 Лабораторные занятия Лекции

<p>профессиональной деятельности</p>	<p>исследований в профильной области П-1 - Иметь опыт выполнения стандартных исследований с использованием серийного научного и технологического оборудования, стандартной методологии и методов исследований У-1 - Соотносить цель и задачи исследования с набором методов исследования, выбирать необходимое сочетание цели и средств</p>	<p>Экзамен</p>
<p>ОПК-3 -Способен систематизировать, анализировать и обобщать результаты научных исследований на основе информационной и библиографической культуры</p>	<p>Д-1 - Демонстрировать развитие когнитивных умений З-1 - Демонстрировать понимание принципов анализа и обобщения результатов научных исследований З-2 - Формулировать требования к оформлению результатов исследований П-1 - Иметь опыт представления обобщенных результатов исследовательской деятельности и их оформления в виде текстовых, графических и иных материалов в соответствии с требованиями П-2 - Иметь опыт написания обзоров литературы, справок, методик экспериментов, описания и обсуждения результатов экспериментов на основе информационной и библиографической культуры У-1 - Систематизировать и анализировать результаты экспериментов, наблюдений, измерений У-2 - Оформлять результаты исследовательской деятельности в виде обзоров литературы, справок, методик в соответствии с принятыми в профессиональной области требованиями</p>	<p>Домашняя работа № 1 Домашняя работа № 2 Домашняя работа № 3 Домашняя работа № 4 Контрольная работа № 1 Контрольная работа № 2 Лабораторные занятия Лекции Экзамен</p>

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

3.1. Процедуры текущей и промежуточной аттестации по дисциплине

1. Лекции: коэффициент значимости совокупных результатов лекционных занятий – 0.50		
Текущая аттестация на лекциях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>домашняя работа</i>	3	20
<i>домашняя работа</i>	9	20
<i>домашняя работа</i>	15	20
<i>контрольная работа</i>	10	20
<i>контрольная работа</i>	17	20
Весовой коэффициент значимости результатов текущей аттестации по лекциям – 0.50		
Промежуточная аттестация по лекциям – экзамен		
Весовой коэффициент значимости результатов промежуточной аттестации по лекциям – 0.50		
2. Практические/семинарские занятия: коэффициент значимости совокупных результатов практических/семинарских занятий – не предусмотрено		
Текущая аттестация на практических/семинарских занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент значимости результатов текущей аттестации по практическим/семинарским занятиям – не предусмотрено		
Промежуточная аттестация по практическим/семинарским занятиям – нет		
Весовой коэффициент значимости результатов промежуточной аттестации по практическим/семинарским занятиям – не предусмотрено		
3. Лабораторные занятия: коэффициент значимости совокупных результатов лабораторных занятий – 0.50		
Текущая аттестация на лабораторных занятиях	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
<i>Домашняя работа</i>	16	100
Весовой коэффициент значимости результатов текущей аттестации по лабораторным занятиям – 1.00		
Промежуточная аттестация по лабораторным занятиям – нет		
Весовой коэффициент значимости результатов промежуточной аттестации по лабораторным занятиям – 0.00		
4. Онлайн-занятия: коэффициент значимости совокупных результатов онлайн-занятий – не предусмотрено		
Текущая аттестация на онлайн-занятиях	Сроки – семестр,	Максимальная оценка в баллах

	учебная неделя	
Весовой коэффициент значимости результатов текущей аттестации по онлайн-занятиям -не предусмотрено		
Промежуточная аттестация по онлайн-занятиям –нет		
Весовой коэффициент значимости результатов промежуточной аттестации по онлайн-занятиям – не предусмотрено		

3.2. Процедуры текущей и промежуточной аттестации курсовой работы/проекта

Текущая аттестация выполнения курсовой работы/проекта	Сроки – семестр, учебная неделя	Максимальная оценка в баллах
Весовой коэффициент текущей аттестации выполнения курсовой работы/проекта– не предусмотрено		
Весовой коэффициент промежуточной аттестации выполнения курсовой работы/проекта– защиты – не предусмотрено		

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

Результаты обучения	Критерии оценивания учебных достижений, обучающихся на соответствие результатам обучения/индикаторам
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.
Другие результаты	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

Характеристика уровней достижения результатов обучения (индикаторов)				
№ п/п	Содержание уровня выполнения критерия оценивания результатов обучения (выполненное оценочное задание)	Шкала оценивания		
		Традиционная характеристика уровня		Качественная характеристика уровня
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

5.1. Описание аудиторных контрольно-оценочных мероприятий по дисциплине модуля

5.1.1. Лекции

Самостоятельное изучение теоретического материала по темам/разделам лекций в соответствии с содержанием дисциплины (п. 1.2. РПД)

5.1.2. Лабораторные занятия

Примерный перечень тем

1. Ближайший элемент. Сортировка чисел. Простые преобразования
2. Матричные производные. Стохастический градиент. Регуляризация
3. Основы метрик классификации и регрессии. Некоррелирующие признаки
4. Энтропия и критерий Джини. Дерево решений

- 5. X-regression. Производные для регрессии. AdaBoost classifier
 - 6. Сборка сети. Линейный слой
 - 7. Ячейка RNN. Residual Block. Языковая модель на RNN
- LMS-платформа
1. <https://ulearn.me/Course/ML/>

5.2. Описание внеаудиторных контрольно-оценочных мероприятий и средств текущего контроля по дисциплине модуля

Разноуровневое (дифференцированное) обучение.

Базовый

5.2.1. Контрольная работа № 1

Примерный перечень тем

- 1. Нахождение Roc-curve
- 2. Feature Engineering и Scaling

Примерные задания

Вам на вход даны y_{real} и массив вероятностей $y_{prob} = P(y_{pred} = 1)$ необходимо реализовать функцию `roc-curve`, которая вернет 2 массива различных значений fpr и tpr , для дальнейшего построения *Roc* кривой.

Можно считать, что все вероятности ограничены $decimal = 2$ (у каждого числа не более 2-х знаков после запятой).

Sample

Input:

```
1 y_real = np.array([ 1, 1, 0, 0, 0, 1, 0, 1, 0])
2 y_prob = np.array([0.8, 0.8, 0.2, 0.2, 0.6, 0.4, 0.6, 0.6, 0.4])
```



Output:

```
1 array([0., 0., 0.4, 0.6, 1. ]), #fpr
2 array([0., 0.5, 0.75, 1., 1. ]), #tpr
```



Sample 2

Input:

```
1 y_real = np.array([ 1, 1, 0, 0, 1, 0, 1, 0])
2 y_prob = np.array([0.8, 0.8, 0.2, 0.2, 0.4, 0.4, 0.6, 0.6])
```



Output:

```
1 array([0., 0., 0.25, 0.5, 1. ]), #fpr
2 array([0., 0.5, 0.75, 1., 1. ]), #tpr
3
4 или
5
6 array([0., 0., 0.5, 1. ]), #fpr
7 array([0., 0.5, 1., 1. ]), #tpr
```



Обратите внимание на 2 пример: roc кривая, которая задается ими - одинаковая. Точка, которая уходит, находится на прямой между двумя соседними, в целом такие точки можно убирать, но будут приниматься оба варианта. Функция `sklearn.metrics.roc_curve` возвращает второй вариант.

Сразу 2 простые задачки: 1. Давайте нагенерируем несколько фич во входном фрейме **df**:

```
— TotalArea = TotalBsmtSF + 1stFlrSF + 2ndFlrSF + GrLivArea + GarageArea
— YearAverage = (YearRemodAdd + YearBuilt) / 2
— LiveAreaQual = OverallQual * GrLivArea
```

На выход отправьте **df** с тремя новыми столбиками. столбцы должны идти в том же порядке что указаны в списке в хвосте **df**.

2.

У стандартного и нормального масштабирования есть одна проблема: она учитывает все признаки, даже те, которые изначально некорректны (шум, выбросы). Чтобы избавиться от шумов и выбросов и корректно масштабировать выборку необходимо использовать [RobustScaling](#).

Ваша задача - отмасштабировать полученный фрейм с помощью [RobustScaler](#). И вернуть отмасштабированный массив (да, скалирование возвращает массив, а не DataFrame).

```
1 import pandas as pd
2
3 def feature_en(df: pd.DataFrame) -> pd.DataFrame:
4     """ ( ) -> .
5     pass
6
7 def scaling(df: pd.DataFrame) -> np.array:
8     """ ( ) -> .
9     pass
```

LMS-платформа

1. <https://ulearn.me/Course/ML/>

5.2.2. Контрольная работа № 2

Примерный перечень тем

1. Co-occurrence
2. Toxic preprocess
3. Toxic vectorizer

Примерные задания

Дано несколько предложений на русском или английском языке.

Необходимо сделать со-occurrence таблицу слов, стоящих рядом в окне длины `window_size`.

См. тесты для примера. Форматировать/изменять предложения не нужно.

```
1 def content_embeddings(list_of_strings: list) -> np.ndarray:
2     """ ( ) -> .
3
4
```

Ваша задача по комментариям определить токсичных пользователей, а также степень токсичности комментария. За большими подробностями и данными [сюда](#).

Для начала нам надо обработать комментарии и убрать из них все лишнее:

- Привести к нижнему регистру
- Убрать все кроме букв алфавита (символы и цифры)
- Убрать стоп-слова + `http`, `com`, `www`, `org`
- Простеммить слова с помощью `PorterStemmer`
- Убрать слова длины 1 и больше 20

```
1 import nltk
2 nltk.download('stopwords')
3
4 def clean_text(text):
5     """ ( ) -> .
```

Теперь наши очищенные комменты преобразуем в `tf-idf` вектора. Причем будем использовать не просто слова, а биграммы. Обучите векторайзер с параметрами `max_features = 5000` и `ngram_range= (2,2)` на текстах их трейна и теста (сконкатенируйте их предварительно) и верните его.

Также вам нужно написать функцию отбора самых важных биграм по датафрейму со значениями `tf-idf`. Для этого необходимо для каждой биграммы среднее значение в фрейме. Так мы найдем среднюю значимость данной биграммы для наших текстов.

Output `tfidf_top_features`:

	feature	tfidf
0	две хочу	0.1443
1	хочу пицца	0.1443

```
1 from sklearn.feature_extraction.text import TfidfVectorizer
2
3 def tfidf_vec(train_text, test_text):
4     """ ( ) """
5
6 def tfidf_top_features(df, top_n=25):
7     """ ( ) """
```

LMS-платформа

1. <https://ulearn.me/Course/ML/>

5.2.3. Домашняя работа № 1

Примерный перечень тем

1. Сумма цифр в массиве

2. Сортировка чисел

Примерные задания

На вход подается `np.ndarray` с натуральными числами. Надо получить массив сумм цифр в этих числах.

Sample

Input:

```
1 a = np.array([1241, 354, 121])
```

Output:

```
1 array([ 8, 12, 4])
```

```
1 import numpy as np
2
3 def num_sum(a: np.ndarray) -> np.ndarray:
4     """ ( ) """
5     pass
```

Дан одномерный массив целых чисел. Необходимо отсортировать в нем только числа, которые делятся на 2. При этом начальный массив изменять нельзя.

Sample

Input:

```
1 A = np.array([43, 66, 34, 55, 78, 105, 2])
```

Output:

```
1 array([ 43,  2, 34, 55, 66, 105, 78])
```

```
1 import numpy as np
2
3 def sort_evens(A: np.ndarray) -> np.ndarray:
4     """ ( ) -> """
5     pass
```

LMS-платформа

1. <https://ulearn.me/Course/ML/>

5.2.4. Домашняя работа № 2

Примерный перечень тем

1. Удаление Nan

Примерные задания

Серия задач в данном модуле объединена в одну большую задачу по предсказанию данных. В ходе выполнения модуля мы будем разбирать определенные техники, которые нужны для ее решения. Настоятельно рекомендуем выполнить все шаги **по порядку**, тогда в конце вы получите решение большой реальной задачи по ML.

Нам даны **данные** о домах выставленных на продажу. Нам необходимо решить задачу регрессии и предсказать цену продажи дома для X_{test} по данным X_{train} и y_{train} . В нашем случае y_{train} - это столбик `SalePrice`, X_{train} - все остальные столбики.

На вход подается 2 считанных датафрейма **df_train**, **df_test** из файлов без изменений.

Начальная подготовка:

- Разделить **df_train** на **X_train** (`pd.DataFrame`) и **y_train** (`pd.Series`).
- Сконкатенировать **X_train** и **df_test** в **df** по вертикали (можно ориентироваться по столбцу `Id` они как раз идут по-порядку). Не забудьте обновить индекс!

Задачи:

- Заменить в **df** все Nan-ы в категориальных признаках (`object`) на строку `missing`
- Заменить в **df** все Nan-ы в числовых признаках на 0.

Вернуть из функции измененный **df**.

```
1 import pandas as pd
2
3 def del_nan(df_train: pd.DataFrame, df_test: pd.DataFrame) -> pd.DataFrame:
4     """ ( ) -> """
5     pass
```

LMS-платформа

1. <https://ulearn.me/Course/ML/>

5.2.5. Домашняя работа № 3

Примерный перечень тем

1. Производные для регрессии

Примерные задания

1. Константный вектор $[a_0]_{i=1}^N$

$$a_0(x) = \arg \min_{c \in \mathbb{R}} \sum_{i=1}^n \mathcal{L}(c, x_i, y_i)$$

2. Градиенты функции потерь

$$g_i^t = - \left[\frac{\partial \mathcal{L}(a_t, x_i, y_i)}{\partial a_t(x_i)} \right]_{i=1}^N$$

3. Коэффициенты при композиции

$$\eta_{t+1} = \arg \min_{\eta} \sum_{i=1}^N \mathcal{L}(f_t + \eta b_{t+1}, x_i, y_i)$$

Sample 1

Input:

```
1 y = np.array([1, 2, 3])
2 f = np.array([2, 2, 2])
3 b = np.array([0, 2, 4])
```



Output:

```
1 f_0 = 2.0
2 g = [-2, 0, 2]
3 alpha = 0.2
```



```
1 def init(y_i: np.array) -> float:
2     """ ( y_i ) -> float: .
3     return a_0
4
5 def grad(a: np.array, y: np.array) -> np.array:
6     """ ( a, y ) -> np.array: .
7     return g
8
9 def eta(f: np.array, b: np.array, y: np.array) -> float:
10    """ ( f, b, y ) -> float: .
11    return eta
```

LMS-платформа

1. <https://ulearn.me/Course/ML/>

5.2.6. Домашняя работа № 4

Примерный перечень тем

1. More Features

Примерные задания

Исследовательская задача. Вам необходимо разобраться в 2-х известных фичах, для предсказания валют и реализовать. Даны предварительные ссылки на википедию, но можно юзать все что угодно.

- `RSI` - Индекс_относительной_силы
- `MACD = EMA_12 - EMA_26` - смотри описание по ссылка ниже
- `MACD_signal` - Индикатор_MACD

Подсказки:

- в `RSI` нужно делать `rolling n=14` (14 дневный стандартный период)
- В `RSI` У понижающих цен не забудьте взять `abs`
- Для подсчета `EMA_m` нужно установить `span=m` и `min_periods=m`

```
1 def extra_features(df: pd.DataFrame) -> pd.DataFrame:
2     df['RSI'] = ### ( o j o ) — ☆*: .
3     df['MACD'] = ### ( o j o ) — ☆*: .
4     df['MACD_signal'] = ### ( o j o ) — ☆*: .
5     return df
```

LMS-платформа

1. <https://ulearn.me/Course/ML/>

5.3. Описание контрольно-оценочных мероприятий промежуточного контроля по дисциплине модуля

5.3.1. Экзамен

Список примерных вопросов

1. Линейная регрессия: свойства, применимость
 2. Регуляризация моделей машинного обучения и её связь с регуляризацией некорректно поставленных задач
 3. Подходы к извлечению информативных признаков из структурированных данных
 4. Принцип работы решающего дерева. Построение и применимость
 5. Методы ансамблирования алгоритмов машинного обучения: эффект и его обоснование
 6. радиентный бустинг: принцип работы и специфика относительно прочих методов ансамблирования
 7. Нейронные сети как универсальные аппроксиматоры
 8. Алгоритм обратного распространения ошибки в нейронных сетях
 9. До-нейросетевые методы в обработке естественного языка
 10. Главные вехи в нейросетевой обработке естественного языка
- LMS-платформа – не предусмотрена

5.4 Содержание контрольно-оценочных мероприятий по направлениям воспитательной деятельности

Направление воспитательной деятельности	Вид воспитательной деятельности	Технология воспитательной деятельности	Компетенция	Результаты обучения	Контрольно-оценочные мероприятия
Профессиональное воспитание	профориентационная деятельность	Технология формирования уверенности и	ОПК-2	Д-2	Домашняя работа № 1 Домашняя работа

		<p>готовности к самостоятельной успешной профессиональной деятельности</p>			<p>№ 2 Домашняя работа № 3 Домашняя работа № 4 Контрольная работа № 1 Контрольная работа № 2 Лабораторные занятия Лекции Экзамен</p>
--	--	--	--	--	---