

Федеральное государственное образовательное  
бюджетное учреждение высшего образования  
**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ  
ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**  
(Финансовый университет)

**Краснодарский филиал Финуниверситета**

Кафедра «Математика и информатика»

СОГЛАСОВАНО

ООО «Портал-Юг»  
Генеральный директор



Е.В. Мостовой

«21» февраля 2024 г.

УТВЕРЖДАЮ

Краснодарский филиал  
Финуниверситета

Директор



Э.В.Соболев

«21» февраля 2024 г.

Калайдин Е.Н.

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ  
МАШИННОЕ ОБУЧЕНИЕ**

студентов, обучающихся по направлению подготовки

01.03.02 Прикладная математика и информатика

в соответствии с образовательными стандартами Финансового университета

(программа подготовки бакалавров)

*Рекомендовано Ученым советом Краснодарского филиала Финуниверситета  
(протокол № 12 от 20.02.2024)*

*Одобрено кафедрой «Математика и информатика»  
(протокол № 13 от 27.02.2024)*

**Краснодар 2024**

УДК: 004.94(075.8)  
ББК: 32.973-018я73  
К17

Рецензенты: Е.А. Демехин, профессор кафедры «Математика и информатика» Краснодарского филиала Финансового университета при Правительстве РФ. В.А. Кирий кандидат физико-математических наук, доцент кафедры «Математика и информатика» Краснодарского филиала Финуниверситета.

Калайдин Е.Н. Рабочая программа дисциплины машинное обучение для обучающихся по направлению 01.03.02 Прикладная математика и информатика, профиль «Анализ данных и принятие решений в экономике и финансах». – Краснодар: Краснодарский филиал Финуниверситета, кафедра «Математика и информатика», 2024 г.

Дисциплина Машинное обучение относится к модулю общепрофессиональных дисциплин по направлению подготовки 01.03.02-Прикладная математика и информатика.

В рабочей программе дисциплины определены ее цель, требования к результатам освоения дисциплины, содержание программы, тематика аудиторных занятий, формы самостоятельной работы, оценочные средства для текущего контроля и промежуточной аттестации, учебно-методическое и информационное обеспечение.

Рабочая программа дисциплины машинное обучение  
(учебно-методический семинар)

*Формат 60\*90/16. Гарнитура Times New Roman*

*Усл. п.л. 2,0. Изд. № \_от.*

*Тираж 100 экз.*

*Заказ № .*

*Отпечатано в Краснодарском филиале Финуниверситета*

© Калайдин Е.Н.  
© Краснодарский филиал Финуниверситета, 2024

## Содержание

1. Наименование дисциплины.....	4
2.Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине .....	4
3.Место дисциплины в структуре образовательной программы .....	5
4.Объем дисциплины(модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся.....	5
5.Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий .....	6
5.1. Содержание дисциплины .....	6
5.2. Учебно-тематический план.....	8
5.3. Содержание семинаров, практических занятий.....	10
6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине .....	12
6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы .....	12
6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю .....	15
7.Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине.....	19
7.3. Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний.....	32
7.4. Примеры практико-ориентированных (ситуационных заданий) .....	33
7.5. Примерные вопросы для подготовки к зачету .....	34
7.6. Примерные вопросы для подготовки к экзамену .....	35
7.7. Пример экзаменационного билета.....	36
8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины.....	36
9.Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины .....	37
10. Методические указания для обучающихся по освоению дисциплины.....	38
11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем .....	40
12.Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине .....	40

## 1. Наименование дисциплины

Дисциплина Б1.О.03.05 «Машинное обучение»

## 2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине

Дисциплина «Машинное обучение» обеспечивает формирование следующих компетенций направления подготовки 01.03.02 «Прикладная математика и информатика» (таблица 1).

Таблица 1 – Компетенции, формируемые в результате изучения дисциплины «Машинное обучение» направления подготовки 01.03.02 «Прикладная математика и информатика»

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПКН-2	Способен с помощью математической модели решать поставленную теоретическую или прикладную задачу, реализовывая алгоритм решения в виде программного модуля	Демонстрирует знание базовых математических моделей, применяемых в различных предметных областях	<b>Знать:</b> основные Математические модели, Применяемые в алгоритмах машинного обучения <b>Уметь:</b> применять алгоритмы машинного обучения для решения практических задач в различных предметных областях
		Адаптирует и применяет существующие математические модели для решения поставленной прикладной или теоретической задачи	<b>Знать:</b> математические основы алгоритмов численной оптимизации моделей машинного обучения, основные алгоритмы обучения и подбора гиперпараметров моделей <b>Уметь:</b> применять методы Обучения и подбора гиперпараметров для решения прикладных задач из различных областей
		Владеет методологией математического моделирования для решения профессиональных задач	<b>Знать:</b> методологией проведения интеллектуального анализа данных и машинного обучения <b>Уметь:</b> получать валидные и робастные оценки результативности проведенного

			анализа и моделирования и делать из них значимые прикладные выводы
--	--	--	--

### 3. Место дисциплины в структуре образовательной программы

Дисциплина «Машинное обучение» относится к модулю общепрофессиональных дисциплин направления подготовки 01.03.02 Прикладная математика и информатика, профиль «Анализ данных и принятие решений в экономике и финансах».

Программа изучения дисциплины составлена с учетом требований, установленных соответствующим ОС ВО Финуниверситета. Изучение дисциплины «Машинное обучение» основывается на сумме знаний, полученных при изучении дисциплины «Алгоритмы и структуры данных в языке Python», «Анализ данных», либо «Теория вероятностей и математическая статистика», «Дискретная математика». Для изучения данной дисциплины студент должен обладать базовыми знаниями в области информационных технологий и программирования, навыками программирования на языке Python.

В свою очередь, изучение дисциплины «Машинное обучение» позволит конкретизировать полученные знания, умения, навыки применительно к решению прикладных задач из различных предметных областей.

Знания и навыки, полученные в процессе изучения дисциплины «Машинное обучение» будут использованы студентами при изучении последующих дисциплин, предусмотренных учебным планом, при написании выпускной квалификационной (бакалаврской) работы, в процессе решения круга задач профессиональной деятельности в дальнейшем.

Таблица 2 – Междисциплинарные связи тем дисциплины с обеспечиваемыми (последующими) дисциплинами направления подготовки 01.03.02 «Прикладная математика и информатика»

№ п/п	Наименование обеспечиваемых (последующих) дисциплин	Номера разделов (тем) данной дисциплины, необходимых для изучения обеспечиваемых (последующих) дисциплин							
		1	2	3	4	5	6	7	8
1.	Эконометрика					*		*	*
2.	Технологии обработки больших данных						*	*	*

### 4. Объем дисциплины(модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся

Для направления подготовки 01.03.02 «Прикладная математика и информатика» по очной (таблица 3) и очно-заочной (таблица 4) формах обучения общая трудоёмкость дисциплины составляет 8 зач. ед., 288 часов (из

них 4 зач. ед., 144 часа в 3-м и 4 зач. ед., 144 часа в 4 семестрах.

Таблица 3 – Трудоемкость дисциплины «Машинное обучение» очная форма обучения

Вид учебной работы по дисциплине	Всего (в з.е. и часах)	Семестр 3 (в часах)	Семестр 4 (в часах)
<b>Общая трудоёмкость дисциплины (в том числе курсовой проект)</b>	<b>8/216</b>	<b>108</b>	<b>108 (курсовой проект 24 часа)</b>
Контактная работа- Аудиторные занятия	102	34	68
Лекции	50	16	34
Семинары, практические занятия	52	18	34
Самостоятельная работа	114	74	40
Вид текущего контроля	Контрольная работа	Контрольная работа	-
Вид промежуточной аттестации	Зачет, экзамен, курсовая работа	Зачет	Экзамен, курсовая работа

Таблица 4 – Трудоемкость дисциплины «Машинное обучение» очная-заочная форма обучения

Вид учебной работы по дисциплине	Всего (в з.е. и часах)	Семестр 3 (в часах)	Семестр 4 (в часах)
<b>Общая трудоёмкость дисциплины (в том числе курсовой проект)</b>	<b>8/252</b>	<b>126</b>	<b>126 (курсовой проект 24 часа)</b>
Контактная работа- Аудиторные занятия	56	28	34
Лекции	24	12	12
Семинары, практические занятия	32	16	16
Самостоятельная работа	196	80	116
Вид текущего контроля	Контрольная работа	Контрольная работа	-
Вид промежуточной аттестации	Зачет, экзамен, курсовая работа	Зачет	Экзамен, курсовая работа

## 5.Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий

### 5.1. Содержание дисциплины

#### Тема 1. Введение в машинное обучение

Основные понятия машинного обучения. Связь с другими

дисциплинами. Контекст машинного обучения как дисциплины – анализ данных и искусственный интеллект. Сферы применения машинного обучения. Типы задач машинного обучения - обучение с учителем и без учителя. Структура данных для машинного обучения. Инструментальные средства машинного обучения. Понятие модели машинного обучения.

## **Тема 2. Регрессия**

Постановка задачи регрессии. Линейная регрессия с одной переменной – функция гипотезы, функция ошибки, метод градиентного спуска. Регрессия с несколькими переменными – множественная линейная регрессия, нормализация признаков, полиномиальная регрессия. Практическое построение регрессии – загрузка и представление данных, реализация метода градиентного спуска, оценка качества регрессии, подбор скорости обучения, знакомство с библиотекой `sklearn`.

## **Тема 3. Классификация**

Постановка задачи классификации как задачи машинного обучения. Отличия от задачи регрессии. Структура данных для классификации. Логистическая регрессия – функция гипотезы, граница принятия решений, функция ошибки логистической регрессии, градиентный спуск для логистической регрессии, многоклассовая классификация, алгоритм «один против всех».

## **Тема 4. Методы обучения с учителем**

Универсальность методов обучения с учителем, общая постановка задачи. Линейные модели – линейная и логистическая регрессии – как единая модель. Полиномиальные модели. Метод опорных векторов, ядра – линейное, гауссово, другие. Перцептрон. Деревья решений. К ближайших соседей. Наивная байесовская модель. Достоинства и недостатки разных типов моделей, их сравнительная характеристика, применимость. Применение этих моделей для решения задач классификации и регрессии.

## **Тема 5. Диагностика систем машинного обучения**

Метрики эффективности машинного обучения – сравнение с функциями ошибки. Типичные метрики эффективности для моделей регрессии – MAE, MSE, RMSE, MSLE, MAPE и другие. Метрики эффективности для моделей классификации – accuracy, precision, recall, F1, ROC, PR и другие. Недообучение и переобучение. Проблема bias-variance. Оценка сложности

моделей. Обобщающая способность моделей, тестовый набор, кривые обучения. Методы борьбы с недо- и переобучением. Регуляризация. Задача выбора модели – кросс-валидация, гиперпараметры моделей, поиск по сетке, валидационный набор.

### **Тема 6. Предварительный анализ и обработка данных**

Сбор данных для моделей обучения с учителем – реляционная форма данных, понятие чистых данных, оценка источников и объемов данных. Описательный (предварительный) анализ данных (EDA) – анализ репрезентативности, шкалы и типы, визуализация, проблема несбалансированности, обнаружение корреляций, аномалий в данных. Очистка и преобразование данных – удаление лишних признаков, удаление непоказательных объектов, заполнение отсутствующих значений, создание суррогатных признаков, преобразование шкал, воспроизводимость преобразования данных.

### **Тема 7. Задачи обучения без учителя**

Задача кластеризации – постановка задачи, структура датасета, результат и интерпретация. Метод К средних – формализация, гиперпараметры, применимость. Другие методы кластеризации – DBSCAN, иерархическая, агломеративная кластеризация. Задача обнаружения аномалий. Задача понижения размерности – метод главных компонент, метод независимых компонент. Обучение с подкреплением.

### **Тема 8. Практическое использование моделей машинного обучения**

Стохастический и пакетный градиентный спуск. Отбор признаков. Частичное обучение с учителем. Ансамблирование моделей – беггинг, бустинг, стакинг, случайный лес, XGBoost, CatBoost. Конвейеризация моделей машинного обучения. Основные этапы проекта по машинному обучению. Визуализация, интерпретация, представление и анализ результатов машинного обучения. Работа с разными типами данных – преобразование графической информации, методы векторизации текста.

#### **5.2. Учебно-тематический план**

Темы дисциплины и виды занятий для направления подготовки 01.03.02 «Прикладная математика и информатика» представлены в таблицах 5-6.

Таблица 5 – Распределение бюджета времени при изучении дисциплины «Машинное обучение» для направления подготовки 01.03.02 «Прикладная



математика и информатика», очная форма обучения, в часах

№ п/ п	Наименование темы (раздела) дисциплины	Трудоемкость в часах					Формы текущего контроля успеваемости
		Всего	Контактная работа - Аудиторная работа			Самостоятельная работа	
			Общая , в т.ч.:	Лекции	Семинары, практические занятия		
1	Введение в машинное обучение	27	6	2	4	14	опрос, проверка лабораторных работ
2	Регрессия	27	10	4	4	13	опрос, проверка лабораторных работ
3	Классификация	27	6	12	4	16	опрос, проверка лабораторных работ
4	Методы обучения с учителем	27	16	6	10	18	опрос, проверка лабораторных работ
5	Диагностика систем машинного обучения	27	20	10	6	16	опрос, проверка лабораторных работ
6	Предварительный анализ и обработка данных	27	18	10	6	13	опрос, проверка лабораторных работ
7	Задачи обучения без учителя	27	14	4	8	14	опрос, проверка лабораторных работ
8	Практическое использование моделей машинного обучения	27	12	2	10	10	опрос, проверка лабораторных работ
В целом по дисциплине		216	102	50	52	114	Согласно учебному плану: контрольная работа

Таблица 6 – Распределение бюджета времени при изучении дисциплины «Машинное обучение» для направления подготовки 01.03.02 «Прикладная математика и информатика», очно-заочная форма обучения, в часах

№ п/ п	Наименование темы (раздела) дисциплины	Трудоемкость в часах				Формы текущего контроля
		Всего	Контактная работа- Аудиторная работа		Самостоятельная работа	

			<b>Общая , в т.ч.:</b>	<b>Лекции</b>	<b>Семинары, практические занятия</b>		<b>успеваемость</b> <b>и</b>
1	Введение в машинное обучение	31	7	3	4	24	опрос, проверка лабораторных работ
2	Регрессия	32	8	4	3	25	опрос, проверка лабораторных работ
3	Классификация	36	6	2	5	26	опрос, проверка лабораторных работ
4	Методы обучения с учителем	31	4	2	4	22	опрос, проверка лабораторных работ
5	Диагностика систем машинного обучения	34	10	4	5	25	опрос, проверка лабораторных работ
6	Предварительный анализ и обработка данных	36	10	3	3	24	опрос, проверка лабораторных работ
7	Задачи обучения без учителя	31	4	1	3	25	опрос, проверка лабораторных работ
8	Практическое использование моделей машинного обучения	21	7	5	5	25	опрос, проверка лабораторных работ
В целом по дисциплине		252	56	24	32	196	Согласно учебному плану: контрольная работа

### 5.3. Содержание семинаров, практических занятий

Цель практических занятий по дисциплине «Машинное обучение» – закрепление теоретических знаний, освоение алгоритмов машинного обучения для решения практических задач в различных предметных областях, контроль выполнения заданий для самостоятельной работы. Занятия проводятся в активной и интерактивной формах с привлечением всех студентов к обсуждаемым вопросам, выбору оптимальных способов

решения практических задач, что способствует профессиональному развитию личности будущего бакалавра. Содержание практических занятий представлено в таблице 7.

Таблица 7 – Содержание семинаров, практических занятий по дисциплине «Машинное обучение» для направления подготовки 01.03.02 «Прикладная математика и информатика»

Наименование тем (разделов) дисциплины	Перечень вопросов для обсуждения на семинарских, практических занятиях, рекомендуемые источники из разделов 8, 9 (указывается раздел и порядковый номер источника)	Формы проведения занятий
Введение в машинное обучение	Входной контроль. Изучение технологического стека анализа данных, построенного на базе языка программирования Python. Знакомство с библиотеками numpy, pandas, matplotlib. Выполнение задания по простому статистическому анализу данных инструментальными средствами, включающем углубленное владение соответствующими инструментами. п.8, 9	Решение и обсуждение задач
Регрессия	Построение модели регрессии методом градиентного спуска своими руками. Знакомство с основами обучения параметров модели. Построение модели множественной регрессии своими руками. Построение модели библиотечными средствами. Сравнение результатов. Реализация модели регрессии на примере, близком к реальному (датасет boston). Интерпретация результатов моделирования. п.8, 9	Решение и обсуждение задач
Классификация	Реализация модели классификации (логистической регрессии) своими руками наподобие линейной регрессии. Использование библиотечных функций. п.8, 9	Решение и обсуждение задач
Методы обучения с учителем	Построение модели классификации на простых данных. Построение нескольких моделей, сравнение их эффективности. Построение модели на сложных данных с нелинейными и слабыми зависимостями. Оценка эффективности, выбор модели. Построение модели регрессии на сложных данных. Построение разных типов регрессоров. Оценка эффективности, выбор модели. Выполнение контрольной работы. п.8, 9	Решение и обсуждение задач

Диагностика систем машинного обучения	Оценка эффективности модели регрессии с помощью разных метрик. Оценка эффективности классификации с помощью разных метрик. Построение диагностических кривых (PR, ROC) для бинарной классификации. Диагностика недо-и переобучения в модели классификации. Построение кривых обучения, их интерпретация. Тестовый набор данных Демонстрация необходимости кросс-валидации. Использование кросс-валидированных оценок качества модели. Демонстрация оптимизации гиперпараметров модели, использование валидационного набора для выбора модели. Поиск модели от простых к сложным. п.8, 9	Решение и обсуждение задач
Предварительный анализ и обработка данных	Сбор и интеграция данных из разных источников. Основные численные характеристики датасета, анализ на чистоту. Нормализация признаков и заполнение отсутствующих значений на искусственном примере. Преобразование категориальных признаков в численные. Краткий алгоритм анализа данных на реальном примере. Интерпретация результатов. Отбор и инжиниринг признаков на примере, приближенном к реальному (датасет titanic). п.8, 9	Решение и обсуждение задач
Задачи обучения без учителя	Решение задачи кластеризации на искусственных данных. Метод К-средних. Метод локтя. Обнаружение аномалий на искусственных данных. Обнаружение аномалий как задача предварительного анализа данных. Понижение размерности на искусственных данных. Интерпретация результата. Понижение размерности для визуализации данных. п.8, 9	Решение и обсуждение задач
Практическое использование моделей машинного обучения	Использование простых ансамблей моделей. Оптимизация гиперпараметров моделей. Визуализация, интерпретация и представление результатов машинного обучения исходя из бизнес-задачи моделирования. Построение конвейера машинного обучения. Построение модели классификации текстов. Знакомство с разными методами векторизации текстов. Построение модели классификации изображений. Представление изображений в виде вектора. п.8, 9	Решение и обсуждение задач

## **6. Перечень учебно-методического обеспечения для самостоятельной работы обучающихся по дисциплине**

### **6.1. Перечень вопросов, отводимых на самостоятельное освоение дисциплины, формы внеаудиторной самостоятельной работы**

Основная цель самостоятельной работы студента при изучении дисциплины «Машинное обучение» — закрепить теоретические знания, полученные в ходе лекционных занятий, глубоко изучить, используя рекомендованную литературу, а также лекции по курсу, основные теоретические аспекты дисциплины, связанные с методами машинного обучения.

Самостоятельная работа студента в процессе изучения дисциплины включает:

- освоение рекомендованной преподавателем по данной дисциплине основной и дополнительной учебной литературы;
- изучение корпоративных образовательных ресурсов (электронные учебники, электронные библиотеки, электронные видеокурсы и др.);
- выполнение домашних заданий в виде решения отдельных задач;
- самостоятельный поиск информации в Интернете;
- консультации по наиболее сложным вопросам;
- участие в работе видео-клуба по кафедре и ежегодных студенческих научных конференциях;
- подготовку к зачету.

На самостоятельную работу студентов, обучающихся по направлению подготовки 01.03.02 «Прикладная математика и информатика» отводится для очной формы обучения 110 часов в третьем семестре и 40 часов в четвертом, а для очно-заочной 116 часов в третьем семестре и 110 часов в четвертом (таблица 8).

Таблица 8 – Перечень вопросов, отводимых на самостоятельное освоение обучающимися дисциплины «Машинное обучение» направления подготовки 01.03.02 «Прикладная математика и информатика»

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Введение в машинное обучение	Изучение истории машинного обучения как дисциплины, связи с другими областями знания. Изучение базовых библиотек анализа данных на языке программирования Python - numpy, pandas, matplotlib по официальной документации (при необходимости).	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Регрессия	Знакомство с нормированием данных, адаптивной скоростью обучения в ручной реализации метода градиентного спуска. Векторизация вычислений в модели множественной регрессии. Использование множественной регрессии для моделирования полиномиальных признаков и нелинейных зависимостей (корень, экспонента, синус, логарифм, обратная пропорциональность). Визуализация исходных данных и результатов моделирования в реальном случае. Анализ и представление результатов моделирования и совершение предсказания. Статистический анализ достоверности предсказания.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Классификация	Реализация алгоритма «один против всех». Оценка эффективности классификации. Сравнение двух реализаций.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Методы обучения с учителем	Визуализация результатов классификации. Построение и анализ отчета от классификации. Оценка скорости обучения и работы разных моделей машинного обучения на большом объеме данных. Визуализация результатов классификации, интерпретация полученных результатов. Совершение предсказания и оценка его надежности.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Диагностика систем машинного обучения	Автоматизация построения модели и подсчета выбранных метрик. Построение сводного отчета. Обоснование выбора метрик эффективности исходя из поставленной задачи. Визуализация диагностических кривых для множественной классификации. Использование регуляризованных моделей - ridge, lasso, elastic net. Демонстрация позитивного влияния регуляризации на точность модели. Построение и робастная оценка эффективности модели на данных, близких к реальным. Использование кросс-валидации. Построение модели и ее последующая оптимизация. Построение ряда простых моделей, диагностика и выбор класса модели, оптимизация и оценка качества.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.

Наименование тем (разделов) дисциплины	Перечень вопросов, отводимых на самостоятельное освоение	Формы внеаудиторной самостоятельной работы
Предварительный анализ и обработка данных	Анализ распределения каждого атрибута и связь с целевой переменной. Визуализация. Оценка влияния нормализации признаков на качество обучения. Заполнение отсутствующих значений на реальном примере. Группировка данных и продвинутые алгоритмы рескалирования. Полный алгоритм анализа данных на реальном примере. Анализ влияния на моделирование. Алгоритмический отбор признаков, отбор по итогам оценки важности признаков по итогам построения простых моделей.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Задачи обучения без учителя	Решение задачи кластеризации на реальных данных. Интерпретация результата. Обнаружение аномалий в реальных данных. Интерпретация результатов. Понижение размерности для задачи снижения объема данных. Понижение размерности как инжиниринг признаков.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.
Практическое использование моделей машинного обучения	Сложные ансамбли моделей - XGBoost, CatBoost. Оптимизация гиперпараметров, выбор модели. Использование различных типов графики, инфографики для оптимального представления результатов моделирования. Знакомство с разными задачами обработки текста - рубрикация, перевод, анализ тональности, генерация. Знакомство с разными задачами анализа изображений – идентификация объекта, обнаружение объекта, распознавание объекта.	Работа с учебной литературой. Разбор вопросов по теме занятия. Выполнение домашних заданий к каждому занятию.

## 6.2. Перечень вопросов, заданий, тем для подготовки к текущему контролю

Проведение аудиторной самостоятельной работы предполагает командную работу при подготовке сообщений по анализу литературных источников (книг, статей, материалов конференций) на заданную тему, подготовку ответов на контрольные вопросы по темам дисциплины, подготовку контрольных работ.

### Перечень контрольных вопросов по дисциплине

1. Дайте определение метрики эффективности и функции ошибки.

2. Дайте определение набору данных (датасета) в машинном обучении.
3. Дайте определение шкале измерения признаков.
4. Дайте определение понятию чистых данных.
5. Перечислите основные этапы проекта по машинному обучению.
6. Что представляет собой предварительный анализ данных.
7. Что представляет собой проблема отсутствующих данных.
8. Что представляет собой проблема несбалансированных классов.
9. Дайте определение параметрам и гиперпараметрам модели.
10. Дайте определение недо- и переобучения.
11. Что представляет собой диагностика модели машинного обучения.
12. Что представляет собой проблема выбора модели машинного обучения.
13. Что представляет собой измерение эффективности работы моделей машинного обучения.
14. Дайте определение метрики эффективности моделей классификации.
15. Дайте определение метрики эффективности моделей регрессии.
16. Дайте определение перекрестной проверке (кросс-валидация).
17. Что представляет собой методы векторизации текстов для задач машинного обучения.
18. Что представляет собой графическая информация в моделях машинного обучения.
19. Что представляет собой метод k средних.
20. Что представляет собой обнаружение аномалий.

### **Примеры заданий для контрольной работы**

1. Загрузить встроенный в библиотеку sklearn датасет «Ирисы». Описать основные количественные характеристики датасета. Визуализировать распределение каждого признака, входящего в датасет. Визуализировать совместное распределение каждого признака и целевой переменной. Сделать вывод о линейной разделимости классов.
2. Загрузить встроенный в библиотеку sklearn датасет «Ирисы». Построить модель парной классификации каждого признака и целевой переменной. Оценить метрики эффективности четырех моделей классификации. Сделать вывод о значимости каждого признака для классификации.
3. Загрузить встроенный в библиотеку sklearn датасет «Ирисы». Описать основные количественные характеристики датасета. Построить модель классификации объектов по всем четырем признакам, используя метод деревьев решений. Построить отчет о классификации, оценить основные метрики классификации - полноту, точность, F-метрику. Сделать вывод об эффективности классификации.
4. Загрузить встроенный в библиотеку sklearn датасет «Ирисы».



Описать основные количественные характеристики датасета. Построить модель классификации объектов по всем четырем признакам, используя метод логистической регрессии. Построить отчет о классификации, оценить основные метрики классификации - полноту, точность, F-метрику. Сделать вывод об эффективности классификации.

5. Загрузить встроенный в библиотеку sklearn датасет «Ирисы». Описать основные количественные характеристики датасета. Построить модель классификации объектов по всем четырем признакам, используя метод опорных векторов. Построить отчет о классификации, оценить основные метрики классификации - полноту, точность, F-метрику. Сделать вывод об эффективности классификации.

6. Загрузить встроенный в библиотеку sklearn датасет «Ирисы». Описать основные количественные характеристики датасета. Построить модель классификации объектов по всем четырем признакам, используя метод многослойного перцептрона. Построить отчет о классификации, оценить основные метрики классификации - полноту, точность, F-метрику. Сделать вывод об эффективности классификации.

7. Загрузить встроенный в библиотеку sklearn датасет «Ирисы». Построить модель классификации объектов по всем четырем признакам, используя любой метод классификации. Оптимизировать гиперпараметры модели при помощи поиска по сетке. Сделать вывод об улучшении эффективности классификации.

8. Загрузить встроенный в библиотеку sklearn датасет «Ирисы». Построить модель классификации объектов по всем четырем признакам, используя любой метод. Оценить эффективность модели при помощи перекрестной проверки (кросс-валидации). Сделать вывод об эффективности классификации.

9. Загрузить встроенный в библиотеку sklearn датасет «Ирисы». Понизить размерность датасета до двух измерений. Визуализировать получившийся набор данных учитывая значение целевой переменной. Сделать вывод о линейной разделимости классов.

10. Загрузить встроенный в библиотеку sklearn датасет «Бостон». Построить модель множественной регрессии. Оценить метрики эффективности регрессии. Сделать вывод об эффективности получившейся модели.

11. Загрузить встроенный в библиотеку sklearn датасет «Бостон». Построить модель регрессии по всем признакам, используя любой метод регрессии. Оптимизировать гиперпараметры модели при помощи поиска по сетке. Сделать вывод об улучшении эффективности регрессии.

12. Загрузить встроенный в библиотеку sklearn датасет «Бостон». Построить модель регрессии по всем признакам, используя любой метод. Оценить эффективность модели при помощи перекрестной проверки (кросс-валидации). Сделать вывод об эффективности регрессии.

13. Загрузить встроенный в библиотеку sklearn датасет «Бостон». Понизить размерность датасета до двух измерений. Визуализировать

получившийся набор данных учитывая значение целевой переменной. Сделать вывод.

14. Загрузить встроенный в библиотеку sklearn датасет «Цифры». Построить модель классификации объектов, используя метод многослойного перцептрона. Построить отчет о классификации, оценить основные метрики классификации - полноту, точность, F- метрику. Сделать вывод об эффективности классификации.

15. Загрузить встроенный в библиотеку sklearn датасет «Цифры». Построить модель классификации объектов, используя любой метод классификации. Оптимизировать гиперпараметры модели при помощи поиска по сетке. Сделать вывод об улучшении эффективности классификации.

16. Загрузить встроенный в библиотеку sklearn датасет «Цифры». Построить модель классификации объектов, используя любой метод. Оценить эффективность модели при помощи перекрестной проверки (кросс-валидации). Сделать вывод об эффективности классификации.

17. Загрузить встроенный в библиотеку sklearn датасет «Цифры». Понизить размерность датасета до двух измерений. Визуализировать получившийся набор данных учитывая значение целевой переменной. Сделать вывод о линейной разделимости классов.

### **Примерная тематика курсовой работы**

1. Deskриптивный анализ численного набора данных с использованием технологий визуализации.

2. Deskриптивный анализ категориального набора данных с использованием технологий визуализации.

3. Deskриптивный анализ смешанного набора данных с использованием технологий визуализации.

4. Сравнение методов регрессии на реальных наборах данных.

5. Сравнение методов классификации на реальных наборах данных.

6. Предварительный анализ данных и построение признаков в задачах обработки финансовой и экономической информации.

7. Предварительный анализ данных и построение признаков в задачах распознавания голоса.

8. Предварительный анализ данных и построение признаков в задачах распознавания текста.

9. Предварительный анализ данных и построение признаков в задачах распознавания объектов на фотографии.

10. Предварительный анализ данных и построение признаков в задачах сжатия информации.

11. Предварительный анализ данных и построение признаков в задачах повышения качества изображений.

12. Предварительный анализ данных и построение признаков в задачах идентификации личности по голосу.

13. Предварительный анализ данных и построение признаков в

задачах идентификации личности по изображению.

14. Предварительный анализ данных и построение признаков в задачах обработки текстов на естественных языках.

15. Предварительный анализ данных и построение признаков в задачах машинного перевода.

16. Предварительный анализ данных и построение признаков в задачах распознавания темы текста.

17. Предварительный анализ данных и построение признаков в задачах нормализации слов текста.

18. Предварительный анализ данных и построение признаков в задачах извлечения знаний из аудиоданных.

19. Предварительный анализ данных и построение признаков в задачах медицинской диагностики.

20. Предварительный анализ данных и построение признаков в задачах извлечения знаний из видеоданных.

Текущий контроль осуществляется в ходе учебного процесса и консультирования студентов, по результатам выполнения самостоятельных работ. Основными формами текущего контроля знаний являются:

- обсуждение вопросов тем и контрольных вопросов;
- решение задач, тестов и их обсуждение в точки зрения умения формулировать выводы, вносить рекомендации и принимать адекватные управленческие решения;
- обсуждение законодательных, правовых и нормативных актов.

О подходе к оценке знаний студентов преподаватель информирует студентов на первом семинарском (практическом) занятии. На последнем семинарском (практическом) занятии студентам сообщается оценка, которую они получают по итогам работы в семестре. Студенты могут улучшить свою оценку по итогам работы в семестре за счет отработки пропущенных занятий. Отработка пропусков, имевших место по причине работы студентов во время занятий, не допускается.

Критерии балльной оценки различных форм текущего контроля успеваемости содержатся в соответствующих методических рекомендациях кафедры «Математика и информатика».

## **7. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине**

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Машинное обучение».

Фонд оценочных средств включает контрольные материалы для проведения текущего контроля и промежуточной аттестации.

## 7.1 Описание показателей и критериев оценивания компетенций

Планируемы е результаты освоения компетенци и (индикатора достижения компетенци и)	Уровень освоения				Оценочное средство
	«неудовлетво рительно»	«удовлетво рительно»	«хорошо»	«отлично»	
ПKN-2 Способен с помощью математической модели решать поставленную теоретическую или прикладную задачу, реализовывая алгоритм решения в виде программного модуля					
Демонстрирует знание базовых математических моделей, применяемых в различных предметных областях					
<b><u>Знать:</u></b> Основные математичес кие модели, применяемы е в алгоритмах машинного обучения	Фрагментар ное представлен ие об основных математичес ких моделях, применяемы е в алгоритмах машинного обучения	Неполные представлен ия об основных математичес ких моделях, применяемы е в алгоритмах машинного обучения	Сформирова нные, но содержащие отдельные пробелы представлен ия об основных математичес ких моделях, применяемы е в алгоритмах машинного обучения	Сформирова нные систематиче ские представлен ия об основных математичес ких моделях, применяемы е в алгоритмах машинного обучения	Вопросы для оценки знаний и умений, задания в виде расчетных задач, тестовые задания
<b><u>Уметь:</u></b> применять алгоритмы машинного обучения для решения практически х задач в различных предметных областях	Фрагментар ное умение применять алгоритмы машинного обучения для решения практически х задач в различных предметных областях	Несистемат ическое применение алгоритмов машинного обучения для решения практически х задач в различных предметных областях	В целом успешное, но содержащее отдельные пробелы умение применять алгоритмы машинного обучения для решения практически х задач в различных предметных областях	Сформирова нное умение применять алгоритмы машинного обучения для решения практически х задач в различных предметных областях	Вопросы для оценки знаний и умений,
Адаптирует и применяет существующие математические модели для решения поставленной прикладной или теоретической задачи					
<b><u>Знать:</u></b>	Фрагментар	Неполные	математичес	Сформирова	Вопросы для

Планируемые результаты освоения компетенции (индикатора достижения компетенции)	Уровень освоения				Оценочное средство
	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»	
математические основы алгоритмов численной оптимизации и моделей машинного обучения, основные алгоритмы обучения и подбора гиперпараметров моделей	ное представление об математических основах алгоритмов численной оптимизации и моделей машинного обучения, основные алгоритмы обучения и подбора гиперпараметров моделей	представления об математических основах алгоритмов численной оптимизации и моделей машинного обучения, основные алгоритмы обучения и подбора гиперпараметров моделей	ких основах алгоритмов численной оптимизации и Сформированные, но содержащие отдельные пробелы представления об моделях машинного обучения, основные алгоритмы обучения и подбора гиперпараметров моделей	ные систематические представления об математических основах алгоритмов численной оптимизации и моделей машинного обучения, основные алгоритмы обучения и подбора гиперпараметров моделей	оценки знаний и умений, задания в виде расчетных задач, тестовые задания
<b>Уметь:</b> применять методы обучения и подбора гиперпараметров для решения прикладных задач из различных областей	Фрагментарное умение применять методы обучения и подбора гиперпараметров для решения прикладных задач из различных областей	Несистематическое применение методов обучения и подбора гиперпараметров для решения прикладных задач из различных областей	В целом успешное, но содержащее отдельные пробелы умение применять методы обучения и подбора гиперпараметров для решения прикладных задач из различных областей	Сформированное умение применять методы обучения и подбора гиперпараметров для решения прикладных задач из различных областей	Вопросы для оценки знаний и умений,
Владеет методологией математического моделирования для решения профессиональных задач					
<b>Знать:</b>	Фрагментар	Неполные	Сформирована	Сформирована	Вопросы для

Планируемые результаты освоения компетенции (индикатора достижения компетенции)	Уровень освоения				Оценочное средство
	«неудовлетворительно»	«удовлетворительно»	«хорошо»	«отлично»	
методологии проведения интеллектуального анализа данных и машинного обучения	ное знание методологий проведения интеллектуального анализа данных и машинного обучения	представления об методологиях проведения интеллектуального анализа данных и машинного обучения	ные, но содержащие отдельные пробелы представления об методологиях проведения интеллектуального анализа данных и машинного обучения	ные, но содержащие отдельные пробелы представления об методологиях проведения интеллектуального анализа данных и машинного обучения	оценки знаний и умений, задания в виде расчетных задач, тестовые задания
<b>Уметь:</b> получать валидные и робастные оценки результативности проведенного анализа и моделирования и делать из них значимые прикладные выводы	Фрагментарное умение получать валидные и робастные оценки результативности проведенного анализа и моделирования и делать из них значимые прикладные выводы	Несистематическое получение валидных и робастные оценки результативности проведенного анализа и моделирования и делать из них значимые прикладные выводы	В целом успешное, но содержащее отдельные пробелы умение получать валидные и робастные оценки результативности проведенного анализа и моделирования и делать из них значимые прикладные выводы	Сформированное умение получать валидные и робастные оценки результативности проведенного анализа и моделирования и делать из них значимые прикладные выводы	Вопросы для оценки знаний и умений,

**7.2 Задания или иные материалы, необходимые для оценки знаний и умений, характеризующих формирование компетенций в процессе освоения ОП ВО**

**7.2.1 Вопросы для оценки знаний и умений, характеризующих**

## формирование компетенций

Шифр компетенции	Вопросы	Правильный ответ
ПKN - 2	1. Дайте определение понятиям «классификация» и «регрессия». Объясните разницу между ними	<p><b>Классификация</b> — это инструмент прогнозного моделирования, используемый для группировки входных переменных в предопределенные категории «метки».</p> <p><b>Регрессия</b> — это метод прогнозного моделирования, предсказывающий значение (в большинстве случаев диапазон), соответствующее входной переменной. Он часто используется для работы с потоками данных — например, для прогнозирования колебаний фондового рынка.</p>
	2. Свойства квалификации	<p>Требуется наличие как минимум двух классов, в которые будут помещены входы.</p> <p>Работает с дискретными переменными или действительными значениями.</p> <p>Точность классификации — это процент правильных предположений от общего числа классификаций.</p>
	3. Опишите разницу между «тестовым набором» и «обучающим набором»	<p>Учебный <b>набор</b> — это объем данных, используемых для обучения проекта машинного обучения.</p> <p>С другой стороны, <b>тестовый набор</b> — это данные, используемые для тестирования модели машинного обучения, которая уже прошла обучение.</p>
	4. Дайте понятие «обучение в ансамбле»	<p><b>Модель ансамбля</b> — это метод машинного обучения, который предлагает объединить несколько базовых моделей для создания единой эффективной модели. Существуют различные методы, которые инженеры используют для реализации ансамблевого обучения — давайте кратко рассмотрим их:</p> <p><b>Начальная загрузка и агрегация (BAGGing)</b> — как следует из названия, метод подразумевает начальную загрузку наборов данных и объединение нескольких загрузочных подмножеств в единое дерево решений. Затем инженер также агрегирует деревья решений для создания макромоделей.</p> <p><b>Случайный лес</b> добавляет дополнительный уровень дифференциации к разделению данных. Алгоритм загружает набор данных таким образом, что каждое дерево решений разбивается на основе набора различных функций.</p>
	5. Какие советы вы используете, чтобы избежать переобучения модели?	<p>Сравнение производительности алгоритма на наборах данных для обучения и тестирования. Если модель имеет значительно более высокую точность при обработке обучающих данных, это красный флаг переобучения.</p> <p>Перекрестная проверка модели на различных типах обучающих данных.</p> <p>Стремление к простоте. До определенного момента добавление новых функций в модель делает ее более</p>

Шифр компетенции	Вопросы	Правильный ответ
		точной и точной. Однако существует порог, после которого каждая новая функция, добавляемая разработчиком, становится избыточной. Инженеры машинного обучения определяют такой предел возможностей для каждого проекта в основном посредством <i>итеративного обучения</i> .
	6. В чем разница между «контролируемым» и «неконтролируемым» машинным обучением?	<p><b>Контролируемое машинное обучение</b> — это процесс обучения модели с учетом желаемых результатов, которые она должна производить. В процессе инженер по машинному обучению гарантирует, что реальный результат соответствует прогнозируемому.</p> <p><b>Однако неконтролируемое машинное обучение</b> — это процесс обмена немаркированными данными с моделью, что позволяет ей самостоятельно находить новую информацию. В отличие от стратегии контролируемого обучения, эта стратегия непредсказуема и может выявить закономерности, о которых инженеры не знали в начале проекта.</p>
	7. Почему инженеры по машинному обучению используют наивный байесовский классификатор?	<p>Легко понять и применить.</p> <p>Не требует больших тренировочных сетов.</p> <p>Хорошо подходит для прогнозов в реальном времени.</p> <p>Высокая скорость бега.</p> <p>Масштабируемость.</p> <p>Поддерживает как дискретные, так и непрерывные значения.</p>
	8. Хорошо ли, когда модель имеет низкое смещение и высокую дисперсию? Какими были бы ваши действия, если бы вы создали такой алгоритм?	<p>Низкое смещение и высокая дисперсия — симптомы переобучения модели. Такая модель не способна отделить «сигнал» релевантных данных от нерелевантной информации («шум»). Основная причина, по которой некоторые модели имеют низкое смещение и высокую дисперсию, заключается в избыточности функций.</p> <p>Чтобы найти способ обойти проблему необходимо найти баланс между смещением и дисперсией, который уменьшит общую ошибку.</p> <p><i>Общая ошибка = смещение <sup>2</sup> + дисперсия + неустраняемая ошибка.</i></p>
	9. Что такое уменьшение размерности?	Уменьшение размерности — это процесс сокращения ряда признаков (столбцов) в наборе данных с использованием линейных и нелинейных методов.
	10. Чем отличаются методы кластеризации KNN и k.means?	<p><b>K-means</b> — это метод кластеризации, используемый для неконтролируемых вопросов и проектов на собеседованиях по глубокому обучению. Он обрабатывает набор данных и группирует его в k кластеров.</p> <p><b>KNN — (K-Nearest Neighbours)</b> — это алгоритм кластеризации, используемый для контролируемых учебных проектов. Он использует набор помеченных</p>



Шифр компетенции	Вопросы	Правильный ответ
		значений, чтобы научить алгоритм группировать неразмеченные данные.
	11. Как анализируется производительность набора данных?	Наиболее распространенный способ анализа производительности набора данных — измерение его показателя F1. Он определяет среднее значение между точностью и полнотой модели. Оптимальная оценка F1 стремится к 1, в то время как низкоточная модель обычно приближается к 0.
	12. Как выбрать классификатор тренировочного набора?	Наиболее распространенным критерием классификации данных является размер множества. Меньшие обучающие модели с меньшей вероятностью будут иметь высокое смещение и низкую дисперсию — чтобы избежать этого, инженеру следует рассмотреть возможность использования наивного байесовского классификатора.
	13. Что такое ЛДА?	LDA (сокращение от линейного дискриминантного анализа) — это широко используемый метод уменьшения размерности. Хотя первоначальный линейный дискриминант применялся для решения задач с двумя классами, позже он был обобщен для обработки нескольких классов
	14. Как вы справляетесь с поврежденными или отсутствующими данными?	Ключевым процессом обработки поврежденных или отсутствующих данных является определение строк таблицы, содержащих такие значения, и замена их новыми. Существует множество инструментов, упрощающих обслуживание наборов данных — большинство разработчиков используют Pandas. Для обнаружения поврежденных ячеек используются методы <code>isnull()</code> и <code>dropna()</code> . Чтобы добавить значения в эти строки, используют команду <code>fillna()</code> .
	15. Объясните разницу между параметрами модели и гиперпараметрами	Основное отличие состоит в том, что не нужно задавать параметры модели — они автоматически выводятся из входных данных. Это не относится к гиперпараметрам — вам нужно указать их вручную. Параметры модели используются как: Веса в нейронных сетях Коэффициенты линейной регрессии Опорные векторные машинные векторы. Гиперпараметры используются как: k-число для кластеризации KNN. Параметры C и сигма для вспомогательных векторных машин Скорость обучения для обучения нейронных сетей.
	16. Каковы преимущества использования нейронных сетей?	Преимущества использования нейронных сетей: Высокая отказоустойчивость Распределенная память Поддержка моделей с неполными знаниями Параллельная обработка Единое хранилище данных для всей сети.

Шифр компетенции	Вопросы	Правильный ответ
	17. Каковы недостатки использования нейронных сетей?	Непредсказуемое поведение Зависимость от оборудования Проблемы управления Невозможно оценить продолжительность сети.
	18. Кратко опишите процесс создания метода машинного обучения	Определите цель модели и согласуйте технические характеристики с бизнес-целями. Накопить данные для генерации набора. Очистите набор данных. Проанализируйте данные, используя основные исследовательские методы. Создайте модель, применяя соответствующие алгоритмы машинного обучения. Протестируйте модель на тестовом наборе данных.
	19. Что вы подразумеваете под перекрестной проверкой?	Как следует из названия, перекрестная проверка – это метод проверки того, может ли данная система машинного обучения точно работать с наборами данных, отличными от того, который использовался для ее обучения.
	20. Два разных набора для перекрестной проверки	Данные для обучения – используются для обучения системы. Данные тестирования – используются для тестирования и проверки системы.
	21. Как выбрать метрики?	Метрики – это параметры, которые помогают оценить модель / систему машинного обучения. MAE, MAPE, RMSE, MSE для регрессии и Accuracy, Recall, Precision и f1 для классификации являются одними из наиболее часто используемых показателей
	22. Что такое ложные срабатывания и ложноотрицания?	<b>Ложные срабатывания</b> , как ложная тревога, когда модель предполагает наличие состояния, даже если она не существует. <b>Ложный отрицательный результат</b> представляет собой полную противоположность описанной выше ситуации, когда модель предполагает отсутствие состояния, когда он на самом деле присутствует
	23. Как проверить модель прогнозирования на основе множественной регрессии?	Чаще всего для этого используется перекрестная проверка, как описано в предыдущем вопросе. Но вы также можете использовать метод скорректированного R-квадрата. В этом методе генерируется значение r-квадрата, которое определяет соотношение между дисперсией, присутствующей в зависимых и независимых переменных набора данных. Таким образом, чем выше значение r-квадрат, тем точнее модель.
	24. Что такое полная форма НЛП	<b>NLP</b> – это сокращение от Natural Language Processing. Это дисциплина искусственного интеллекта, которая помогает машинам понимать людей и взаимодействовать с ними в более разговорной манере
	25. Что такое случайный лес?	Случайные леса – это методика обучения, основанная на концепции деревьев решений. Множественные деревья решений создаются путем случайного выбора

Шифр компетенции	Вопросы	Правильный ответ
		подмножества переменных на каждом шаге дерева решений, которое объединяется в случайный лес. Затем выбирается режим всех прогнозов в результате с наименьшей вероятностью ошибок
	26. Что делает среднеквадратическую ошибку плохим показателем производительности модели?	MSE или среднеквадратическая ошибка основана на связывании значительно более высокого веса с большими ошибками, что делает больший акцент на более широких отклонениях. Однако это хорошо работает в большинстве алгоритмов, чтобы минимизировать ошибку модели и стоимость. Иногда лучшим вариантом для MSE является MAE (средняя абсолютная ошибка) или MAPE (средняя абсолютная ошибка в процентах), что устраняет вышеуказанный недостаток и легко интерпретируется.

### 7.2.2 Тесты

Шифр компетенции	Тестовые задания	Правильный ответ
ПКН - 2 2	Какой метод представляет собой итерационный процесс смены поколений? (1) RSM; (2) Беггинг; (3) CCEL.	3
	2. Что из ниже перечисленного не относится к корректирующим операциям? (1) простое голосование; (2) взвешенное голосование; (3) голосование по старшинству; (4) метод стохастического градиента; (5) обучение по Хеббу; (6) правило мягкой конкуренции.	4, 5, 6
	3. Какой алгоритм строит набор конъюнктивных закономерностей? (1) алгоритм простого голосования; (2) алгоритм взвешенного голосования; (3) алгоритм КОРА; (4) алгоритм ТЭМП.	3
	4. Что, из ниже перечисленного, не подается на вход в алгоритме IRLS? (1) матрица "объекты-признаки" и вектор ответов; (2) обучающая выборка; (3) обучающая выборка и дискриминантная функция.	2,3

Шифр компетенции	Тестовые задания	Правильный ответ
	<p>5. Какой алгоритм имеет такое условие, что пока есть путь между двумя вершинами разных классов, то удалить самое длинное ребро на этом пути?</p> <p>(1) алгоритм кратчайшего незамкнутого пути;  (2) алгоритм кластеризации;  (3) алгоритм частичного обучения;  (4) алгоритм co-training.</p>	3
	<p>6. Что называют многослойной сетью?</p> <p>(1) сеть, которая имеет больше одного слоя;  (2) сети, которые способны решить задачу XOR;  (3) сеть, которая имеет 3 слоя.</p>	1
	<p>7. Верно ли утверждение? Всякая оптимизация по неполной информации и избыточная сложность параметров приводит в переобучению.</p> <p>(1) Да  (2) Нет</p>	1
	<p>8. При каком <math>n</math> в карте сходства отображается результат многомерного шкалирования в виде плоского точечного графика?</p> <p>(1) <math>n = 0</math>  (2) <math>n = 1</math>  (3) <math>n = 2</math>  (4) <math>n = 3</math>  (5) <math>n = 4</math></p>	3
	<p>9. Что называется вероятностями перехода?</p> <p>(1) задание обучения с подкреплением, удовлетворяющее марковскому свойству;  (2) пространства состояний и действий, которые являются конечными;  (3) при заданном состоянии <math>s</math> и действии <math>a</math> вероятность каждого возможного следующего состояния <math>s'</math>.</p>	3
	<p>10. Какой пример, из ниже перечисленных, является примером смеси алгоритмов?</p> <p>(1)</p> $b(x) = F(b_1(x), g_1(x), \dots, b_r(x), g_r(x)) = \sum_{t=1}^T g_t(x) b_t(x) ;$ <p>(2)</p> $b(x) = F(b_1(x), \dots, b_r(x)) = \sum_{t=1}^T \alpha_t b_t(x), x \in X, \alpha_t \in R ;$	1

Шифр компетенции	Тестовые задания	Правильный ответ
	$b(x) = F(b_1(x), \dots, b_r(x)) = \frac{1}{T} \sum_{t=1}^T b_t(x), x \in X$ (3) ; (4) $a(x) = F(b_1(x), \dots, b_r(x))$ .	
	11. В задачах классификации признаки могут быть строковыми, вещественными, числовыми. (1) Да (2) Нет	1
	12. Чтобы использовать расстояние Махаланобиса в задаче определения принадлежности заданной точки одному из N классов, нужно найти матрицы ковариации всех классов. (1) Да (2) Нет	1
	13. Верно ли утверждение? Метод SG позволяет настраивать веса на избыточно больших выборках, за счет того, что случайной подвыборки может оказаться достаточно для обучения. (1) Да (2) Нет	1
	14. Как будет называться модель, в которой учитывается 11 слов? (1) униграммной; (2) биграммной; (3) триграммной; (4) n-граммной.	4
	Что из ниже перечисленного не относится к недостаткам тривиальной рекомендующей системой? (1) не учитываются интересы конкретного пользователя; (2) проблема "холодного старта"; (3) надо хранить всю матрицу F; (4) нечего рекомендовать новым пользователям.	4
	Какие задачи, из ниже перечисленных, являются задачами прогнозирования? (1) математический прогноз даты сильных землетрясений; (2) определение длительности и исхода заболевания; (3) обнаружение спама; (4) прогнозирование вероятности летального исхода; (5) задачи поискового вывода.	1,4
	В зависимости от значений отступа обучающие объекты условно делятся на: (1) 2 типа;	4

Шифр компетенции	Тестовые задания	Правильный ответ
	(2) 3 типа; (3) 4 типа; (4) 5 типов; (5) 6 типов.	
	Действительно ли, что RSM выполняется строго последовательно не допуская эффективного распространения? (1) Да (2) Нет	2
	Что называют в теории нейронных сетей сокращением весов? (1) квадратичную регуляризацию; (2) нормализацию признаков; (3) стохастический градиент; (4) распределение Лапласа.	1
	Что называют в теории нейронных сетей сокращением весов? (1) квадратичную регуляризацию; (2) нормализацию признаков; (3) стохастический градиент; (4) распределение Лапласа.	1
	Действительно ли, что метод INCAS позволяет решать задачи, в которых нет линейной делимости? (1) Да (2) Нет	1
	Верно ли что, если обучающая выборка имеет большой объем, то можно использовать метод Хебба? (1) Да (2) Нет	2
	Верно ли, что оценить вероятность $P(y b_j(x) = 1)$ можно, только с помощью эмпирической оценкой по различным данным? (1) Да (2) Нет	2
	Что, из ниже перечисленного, не относится к типу экспериментального исследования? (1) исследование задач ранжирования (2) исследование на реальных данных (3) исследование на модельных данных	1
	Являются ли вероятностные модели типом латентных моделей? (1) Да (2) Нет	1

Шифр компетенции	Тестовые задания	Правильный ответ
	Какой алгоритм на каждом шаге отбирает целые популяции? (1) процедура стабилизации; (2) процедура редукции; (3) генетический алгоритм синтеза конъюнкций; (4) поиск информативных конъюнкций.	3
	Верно ли утверждение, что ёмкость семейства линейных решающих правил $A$ равна размерности пространства $n$ ? (1) Да (2) Нет	1
	Отрицательные отступы и классифицирующиеся неверно имеют: (1) ошибочные объекты; (2) шумовые объекты; (3) пограничные объекты; (4) неинформативные объекты.	1
	Процесс упрощения сети, в алгоритме OBD, останавливается, когда: (1) внутренний критерий стабилизируется; (2) заданный внешний критерий начинает возрастать; (3) градиентный алгоритм находит новый локальный минимум $Q$ ; (4) информация, накопленная в сети, является полезной и не теряется при добавлении новых нейронов.	1,2
	Какие данные не используют в картах Кохонена в качестве входных? (1) количество побед $m$ ; (2) обучающая выборка; (3) характеристика $C=0.1$ ; (4) темп обучения; (5) количество слоёв.	1-5
	Что, из ниже перечисленного является недостатком решающих списков? (1) простота классификации; (2) обученное по выборке правило классификации можно выполнять "вручную"; (3) возможность обработки разнотипных данных с пропусками; (4) если множество правил $\Phi$ выбрано неудачно, список может не построится; (5) каждый объект классифицируется только одним правилом.	4,5
	Верно ли, что при $n=3$ многомерное шкалирование позволяет отобразить выборку в виде множества точек на плоскости? (1) Да (2) Нет	2

Шифр компетенции	Тестовые задания	Правильный ответ
	Какие модели, из ниже перечисленных относятся к многомодальным тематическим моделям? (1) коллаборативная фильтрация; (2) персонализация рекламы в Интернете; (3) модель с непрерывным временем; (4) модель с медленно меняющейся тематикой; (5) регуляризация матрицы перевода слов; (6) марковские модели синтаксиса языка.	1,2

Перечень компетенций с указанием индикаторов их достижения в процессе освоения образовательной программы содержится в разделе 2. «Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине».

### 7.3. Типовые контрольные задания или иные материалы, необходимые для оценки индикаторов достижения компетенций, умений и знаний

Примеры оценочных средств для проверки каждой компетенции, формируемой дисциплиной в таблице 9.

Таблица 9 - Типовые оценочные средства для проверки каждой компетенции, формируемой дисциплиной «Машинное обучение» направления 01.03.02 «Прикладная математика и информатика»

Наименование компетенции	Наименование индикаторов достижения компетенции	Результаты обучения (умения и знания), Соотнесенные с индикаторами достижения компетенции	Типовые контрольные задания
<b>ПКН-2</b> Способен с помощью математической модели решать поставленную теоретическую или прикладную задачу, реализовывая алгоритм решения в виде	Демонстрирует знание базовых математических моделей, применяемых в различных предметных областях.	<b>Знать:</b> основные математические модели, применяемые в алгоритмах машинного обучения <b>Уметь:</b> применять алгоритмы машинного обучения для решения практических задач в различных предметных областях	Загрузить встроенный в библиотеку sklearn датасет «Бостон». Построить модель множественной регрессии. Оценить метрики эффективности регрессии. Сделать вывод об эффективности получившейся модели. Логистическая регрессия в задачах классификации.



программного модуля	Адаптирует и применяет существующие математические модели для решения поставленной прикладной или теоретической задачи.	<p><b>Знать:</b> математические основы алгоритмов численной оптимизации моделей Машинного обучения, Основные алгоритмы Обучения и подбора гиперпараметров моделей</p> <p><b>Уметь:</b> применять методы обучения и подбора гиперпараметров для решения прикладных задач из различных областей</p>	Загрузить встроенный в библиотеку sklearn датасет «Цифры». Построить Модель классификации объектов, используя любой метод классификации. Оптимизировать гиперпараметры модели при помощи поиска по сетке. Сделать вывод об улучшении эффективности классификации. Метод градиентного спуска для парной линейной регрессии.
	Владеет методологией математического моделирования для решения профессиональных задач.	<p><b>Знать:</b> методологией проведения интеллектуального анализа данных и машинного обучения</p> <p><b>Уметь:</b> получать валидные и робастные оценки результативности проведенного анализа и моделирования и делать из них значимые прикладные выводы</p>	Основные этапы проекта по машинному обучению. Перекрестная проверка (кросс-валидация). Назначение, схема работы.

#### 7.4. Примеры практико-ориентированных (ситуационных заданий)

Практическая часть представляет собой задачу по анализу конкретного датасета. Для каждого задания будет необходимо:

1. Загрузить датасет в Python.
2. Описать набор данных и решаемую задачу.
3. Выделить целевую переменную и факторные переменные.
4. Удалить ненужные данные, проанализировать отсутствующие значения.
5. Прокомментировать количественные параметры датасета.
6. Разбить выборку на обучающую и тестовую.
7. Работа по вариантам.

**Вариант 1.** Очистка данных и обучение моделей.

Данный вариант предполагает фокусировку на обучении нескольких видов моделей обучения с учителем. В зависимости от набора данных, может предполагаться задача классификации и регрессии. Необходимо после

минимальной подготовки датасета к обучению обучить несколько моделей и сравнить их эффективность.

**Вариант 2.** Описательный анализ и визуализация данных.

Данный вариант предполагает фокусировку на исследовании данных и визуализации. При решении этого варианта следует провести как можно более подробный описательный анализ данных с использованием максимального спектра средств визуализации. При этом следует делать значимые выводы об обнаруженных в данных закономерностях.

**Вариант 3.** Построение модели и оптимизация гиперпараметров.

Данный вариант предполагает фокусировку на процессе улучшения эффективности модели обучения с учителем. Студенту следует подготовить датасет к обучению, обучить одну из моделей с учителем со значениями гиперпараметров по умолчанию, получить значение эффективности. После этого вручную или автоматически подобрать значения гиперпараметров таким образом, чтобы получить максимальный прирост эффективности.

**Вариант 4.** Выбор признаков.

Данный вариант предполагает фокусировку на улучшении модели путем ввода новых признаков в модель. Следует подготовить модель к обучению, обучить модель и зафиксировать начальный уровень эффективности. Затем следует исследовать влияние исключения существующих и введения новых признаков в модель на эффективность. Как вариант можно рассматривать введение полиномиальных признаков. Следует стремиться к максимальному увеличению эффективности модели.

**Вариант 5.** Исследование влияния обучения без учителя на эффективность обучения.

Данный вариант предполагает фокусировку на использовании методов обучения без учителя для ускорения или повышения эффективности обучения с учителем. Следует подготовить модель к обучению, обучить модель и зафиксировать начальный уровень эффективности. Затем следует попробовать применить понижение размерности, обнаружение аномалий или кластеризацию (в любой комбинации) для трансформации исходного датасета. В конце работы следует сделать значимый вывод об изменении скорости и эффективности обучения с учителем.

## **7.5. Примерные вопросы для подготовки к зачету**

1. Понятие машинного обучения. Отличие машинного обучения от других областей программирования.
2. Классификация задач машинного обучения. Примеры задач из различных классов.
3. Основные понятия машинного обучения: набора данных, объекты, признаки, атрибуты, модели, параметры.
4. Структура и представление данных для машинного обучения.
5. Инструментальные средства машинного обучения.
6. Задача регрессии: постановка, математическая формализация.

7. Метод градиентного спуска для парной линейной регрессии.
8. Понятие функции ошибки: требования, использование, примеры.
9. Множественная и нелинейная регрессии.
10. Нормализация признаков в задачах регрессии.
11. Задача классификации: постановка, математическая формализация.
12. Метод градиентного спуска для задач классификации.
13. Логистическая регрессия в задачах классификации.
14. Множественная и многоклассовая классификация. Алгоритм «один против всех».
15. Метод опорных векторов в задачах классификации.
16. Понятие ядра и виды ядер в методе опорных векторов.
17. Метод решающих деревьев в задачах классификации.
18. Метод k ближайших соседей в задачах классификации.
19. Однослойный перцептрон в задачах классификации.

## **7.6. Примерные вопросы для подготовки к экзамену**

1. Метрики эффективности и функции ошибки: назначение, примеры, различия.
2. Понятие набора данных (датасета) в машинном обучении. Требования, представление. Признаки и объекты.
3. Шкалы измерения признаков. Виды шкал, их характеристика.
4. Понятие чистых данных. Определение, очистка данных.
5. Основные этапы проекта по машинному обучению.
6. Предварительный анализ данных: задачи, методы, цели.
7. Проблема отсутствующих данных: причины, исследование, пути решения.
8. Проблема несбалансированных классов: исследование, пути решения.
9. Понятие параметров и гиперпараметров модели. Обучение параметров и гиперпараметров. Поиск по сетке.
10. Понятие недо- и переобучения. Определение, пути решения.
11. Диагностика модели машинного обучения. Методы, цели.
12. Проблема выбора модели машинного обучения. Сравнение моделей.
13. Измерение эффективности работы моделей машинного обучения. Метрики эффективности.
14. Метрики эффективности моделей классификации. Виды, характеристика, выбор.
15. Метрики эффективности моделей регрессии. Виды, характеристика, выбор.
16. Перекрестная проверка (кросс-валидация). Назначение, схема работы.
17. Методы векторизации текстов для задач машинного обучения.

18. Представление графической информации в моделях машинного обучения.
19. Задачи без учителя. Кластеризация. Метод k средних.
20. Задачи без учителя. Обнаружение аномалий.
21. Задачи без учителя. Понижение размерности. Метод PCA.
22. Конвейеры в библиотеке sklearn. Назначение, использование.
23. Использование методов визуализации данных для предварительного анализа.
24. Исследование коррелированности признаков: методы, цели, выводы.
25. Решкалирование данных. Виды, назначение, применение. Нормализация и стандартизация данных.
26. Преобразование категориальных признаков в числовые.
27. Ансамблевые модели машинного обучения. Виды ансамблирования.
28. Конвейеризация моделей машинного обучения.
29. Методы визуализации данных для машинного обучения.
30. Задача выбора модели. Оценка эффективности, валидационный набор.

## **7.7. Пример экзаменационного билета**

### **Экзаменационный билет №**

1. Представление графической информации в моделях машинного обучения. (20 баллов)
2. Метрики эффективности и функции ошибки: назначение, примеры, различия. (20 баллов)
3. Вариант 3. Построение модели и оптимизация гиперпараметров. Датасет: <https://www.kaggle.com/nandvard/microsoft-data-science-cupstone> (20 баллов)

## **8. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины**

### **Основная литература:**

1. Колдаев, В. Д. Структуры и алгоритмы обработки данных : учебное пособие / В. Д. Колдаев. - Москва : РИОР : ИНФРА-М, 2021. - 296 с. - ЭБС ZNANIUM.com. - URL: <https://znanium.com/catalog/product/1230215> (дата обращения: 06.02.2022). - Текст: электронный.
2. Флах, П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : учебник / П. Флах. - 2-е изд. - Москва.: ДМК Пресс, 2023. - 401 с. - ISBN 978-5-89818-300-4. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2085038>

### **Дополнительная литература:**

3. Нагаева, И. А. Основы алгоритмизации и программирования: практикум : учебное пособие / И. А. Нагаева, И. А. Кузнецов. - Москва : Берлин : Директ- Медиа, 2021. - 169 с. - ЭБС Университетская библиотека ONLINE. - URL: <https://biblioclub.ru/index.php?page=book&id=598404> (дата обращения: 06.02.2022). - Текст : электронный.

4. Машинное обучение в финансах : учебник / С. Ю. Богатырев, А. А. Помулев, А. В. Затевахина [и др.] ; под. ред. С. Ю. Богатырева. - Москва : Прометей, 2024. - 224 с. - ISBN 978-5-00172-572-5. - Текст : электронный. - URL: <https://znanium.ru/catalog/product/2144368>

5. Златопольский, Д. М. Программирование: типовые задачи, алгоритмы, методы: учебное пособие / Д. М. Златопольский. — 4-е изд. (эл.). — Москва : Лаборатория знаний, 2020.— 226 с.: ил. — ЭБС Университетская библиотекаONLINE.—

URL:<https://biblioclub.ru/index.php?page=book&id=222873> ( дата обращения 06.02.2022). - Текст: электронный.

## **9.Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины**

1. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>

2. Электронно-библиотечная система BOOK.RU <http://www.book.ru>

3. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>

4. Электронно-библиотечная система Znanium <http://www.znanium.com>

5. Электронно-библиотечная система издательства «ЮРАЙТ» <https://urait.ru/>

6. Электронно-библиотечная система издательства Проспект <http://ebs.prospekt.org/books>

7. Электронно-библиотечная система издательства Лань <https://e.lanbook.com/>

8. Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>

9. Электронная библиотека Издательского дома «Гребенников» <https://grebennikon.ru/>

10. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>

11. Национальная электронная библиотека Б.Ир://нэб.рф/

12. Финансовая справочная система «Финансовый директор» <http://www.1fd.ru/>

13. Pyru 1.0.9 [Электронный ресурс]: сайт. - Режим доступа: <https://pypi.python.org/pypi/pyru>

14. Python Data Analysis Library [Электронный ресурс]: сайт. - Режим доступа: <http://pandas.pydata.org/>

15. Python Documentation [Электронный ресурс]: сайт. - Режим доступа: <http://python.org/doc/>

16. Python Standard Library [Электронный ресурс]: сайт. - Режим доступа: <https://docs.python.org/2/library/>
17. Scikit-learn Machine Learning in Python [Электронный ресурс]: сайт. - Режим доступа: <http://scikit-learn.org>
18. The Python Tutorial // <https://docs.python.org/3/tutorial/index.html>
19. NumPy User Guide // <http://docs.scipy.org/doc/numpy/user/index.html>
20. Pandas User Guide <http://pandas.pydata.org/pandas-docs/stable/>

## **10. Методические указания для обучающихся по освоению дисциплины**

При изучении теоретического материала необходимо опираться на рабочую программу дисциплины, материалы лекций и литературу из основного списка. Кроме этого, необходимо активно работать с Интернет-источниками и пособиями других авторов, помогающими усвоить материал отдельных разделов программы.

Необходимо конспектировать лекции, помечая сложные и непонятные моменты с тем, чтобы задать вопросы лектору в конце лекции или же на консультации.

При подготовке к семинарским занятиям необходимо изучить вопросы, вынесенные на самостоятельное изучение, так как семинарские занятия предполагают их обсуждение и дискуссию по теме; кроме того, задания для самостоятельной работы необходимы для того, чтобы успешно выполнить самостоятельные задания на семинарах.

Индивидуальные задания для работы на компьютере, файлы с выполненными заданиями необходимо хранить в личной сетевой папке в компьютерной сети вуза.

Для выполнения кейса задачи, рекомендуется использовать язык программирования Python и специализированные библиотеки для анализа данных и машинного обучения.

Студент должен прислать рабочий notebook с решением кейса задачи, комментариями кода и аналитическими выводами до конца экзамена на электронную почту преподавателя.

Критерии оценки практической экзаменационной работы:

1. Структурированность отчета. В работе должна прослеживаться четкая структура - подготовительный этап, анализ данных, построение простых моделей, сравнение и анализ моделей, выводы, построение моделей с учетом выводов, итоговый результат.
2. Наличие выводов. Работа должна содержать текстовые замечания, поясняющие каждый шаг работы студента: что делается, зачем и какую информацию это нам дает. Оценивается полнота и адекватность выводов.
3. Визуализация. Работа должна демонстрировать навыки студента визуализировать информацию. Особенно на этапах описательного анализа и анализа обучаемости модели. Оценивается разнообразие, наглядность и информативность визуализации.

4. Использование метрик эффективности. Оценивается разнообразие и адекватность задаче примененных метрик эффективности (включая время обучения) а также полнота сравнения и правильность выводов из сравнения моделей по разным метрикам.

5. Валидность результатов. Студент должен продемонстрировать умение оценивать достоверность измерения метрик моделей и повышать ее с использованием перекрестной проверки (кроссвалидации). Использование k-fold cross validation является предпочтительным методом измерения эффективности модели. Если происходит выбор модели, то ее итоговая эффективность должна измеряться на чистом наборе данных.

Список датасетов, использующихся на экзамене:

1. <https://www.kaggle.com/uciml/mushroom-classification>
2. <https://www.kaggle.com/lodetomasi1995/income-classification>
3. <https://www.kaggle.com/uciml/glass>
4. <https://www.kaggle.com/uciml/german-credit>
5. <https://www.kaggle.com/zaurbegiev/my-dataset>
6. <https://www.kaggle.com/kaushiksuresh147/customer-segmentation>
7. <https://www.kaggle.com/deepu1109/star-dataset>
8. <https://www.kaggle.com/vinesmsuic/star-categorization-giants-and-dwarfs>
9. <https://www.kaggle.com/shebrahimi/financial-distress>
10. [https://www.kaggle.com/teej\\_mahal20/airline-passenger-satisfaction](https://www.kaggle.com/teej_mahal20/airline-passenger-satisfaction)
11. <https://www.kaggle.com/amir75/caesarean-section-classification>
12. <https://www.kaggle.com/sachinsharma1123/performance-prediction>
13. <https://www.kaggle.com/ninzaami/loan-predication>
14. <https://www.kaggle.com/caparrini/beatsdataset>
15. <https://www.kaggle.com/zhiruo19/covid19-symptoms-classification>
16. <https://www.kaggle.com/kunalvsingh93/banking-modelclassification>
17. <https://www.kaggle.com/mansoordaku/ckdisease>
18. <https://www.kaggle.com/mnassrib/telecom-churn-datasets>
19. <https://www.kaggle.com/akshayksingh/kidney-disease-dataset>
20. <https://www.kaggle.com/henriqueyamahata/bank-marketing>
21. <https://www.kaggle.com/maajdl/yeh-concret-data>
22. <https://www.kaggle.com/hellbuoy/car-price-prediction>
23. <https://www.kaggle.com/rhuebner/human-resources-data-set>
24. <https://www.kaggle.com/loveall/appliances-energy-prediction>
25. <https://www.kaggle.com/elikplim/forest-flres-data-set>
26. <https://www.kaggle.com/nandvard/microsoft-data-science-capstone>
27. <https://www.kaggle.com/shebrahimi/financial-distress>
28. <https://www.kaggle.com/aungpyaeap/beauty>
29. <https://www.kaggle.com/vbmokin/ammonium-prediction-in-river-water>
30. <https://www.kaggle.com/veer06b/marrket-mix-dataset>

## **11. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем**

1. Комплект программного обеспечения, антивирус;
  - 1) Браузер.
  - 2) Дистрибутив языка Python 3.6 (или более поздней версии) Anaconda
  - 3) Облачная среда разработки Google Colab или аналогичная
  - 4) Среда разработки Jupyter Notebook
2. Современные профессиональные базы данных и информационные справочные системы
  - не используются
3. Сертифицированные программные и аппаратные средства защиты информации
  - не используются

Каждый обучающийся в течение всего обучения обеспечивается индивидуальным неограниченным доступом к электронно-библиотечной системе и электронной информационно-образовательной среде Краснодарского филиала Финансового университета.

Электронная информационно-образовательная среда Финансового университета обеспечивает:

– доступ к учебным планам, рабочим программам дисциплин, практик, к указаниям электронных библиотечных систем и электронным образовательным ресурсам, указанным в рабочих программах. По дисциплине «Машинное обучение» в электронной информационно-образовательной среде Финансового университета представлены следующие виды информационных ресурсов:

- аннотации дисциплины;
- мультимедийные презентации по всем темам курса;
- методический материал;
- рабочие программы дисциплины.
- фиксацию хода образовательного процесса, результатов промежуточной аттестации;
- проведение всех видов занятий, процедур оценки результатов обучения, реализация которых осуществляется с применением электронного обучения.

## **12. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине**

Материально-техническая база Краснодарского филиала Финансового университета соответствует действующим противопожарным правилам и нормам, обеспечивает проведение всех видов дисциплинарной и



междисциплинарной подготовки, практической и научно-исследовательской работ обучающихся, предусмотренных учебным планом.

Образовательный процесс обеспечивается специальными помещениями, которые представляют собой аудитории для проведения занятий лекционного типа, занятий семинарского типа, выполнения курсовых работ, групповых и индивидуальных консультаций, текущего контроля и промежуточной аттестации, помещения для самостоятельной работы студентов и помещения для хранения и профилактического обслуживания учебного оборудования. Специальные помещения укомплектованы специализированной мебелью и техническими средствами обучения, обеспечивающими представление учебной информации большой аудитории, демонстрационным оборудованием.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой, обеспечивающей доступ к сети Интернет и электронной информационно-образовательной среде Финансового университета.