

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

**Краснодарский филиал  
Кафедра «Математика и информатика»**

СОГЛАСОВАНО

ООО «Портал-Юг»  
Генеральный директор



Е.В. Мостовой

«20» февраля 2024 г.

УТВЕРЖДАЮ

Краснодарский филиал  
Финансового университета  
Директор



Э.В. Соболев

«20» февраля 2024 г.

**Калайдин Е.Н.**

**РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ  
ПОСТРОЕНИЕ И ОЦЕНКА МОДЕЛЕЙ МАШИННОГО ОБУЧЕНИЯ**

для студентов, обучающихся по направлению подготовки  
01.04.02 Прикладная математика и информатика,  
Направленность программы магистратуры:  
Анализ больших данных и машинное обучение  
в экономике и финансах

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

*Одобрено кафедрой «Математика и информатика»  
(протокол № 13 от 13.02.2024)*

**Краснодар 2024**

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

**Рецензенты:** Кирий В.А., доцент кафедры «Математика и информатика», канд. физ.-мат. наук,  
Коренева О.В., доцент кафедры «Математика и информатика», канд. техн. наук.

**Калайдин Е.Н. «Построение и оценка моделей машинного обучения».** Рабочая программа дисциплины для студентов, обучающихся по направлению подготовки 01.04.02 – «Прикладная математика и информатика», направленность программы магистратуры «Анализ больших данных и машинное обучение в экономике и финансах», (программа подготовки магистров) — Краснодар: Краснодарский филиал Финансового университета при Правительстве Российской Федерации, Кафедра математики и информатики, 2024.

Дисциплина **«Построение и оценка моделей машинного обучения»** относится к модулю общепрофессиональных дисциплин направления профиля «Анализ данных и принятие решений в экономике и финансах», направление подготовки 01.04.02 «Прикладная математика и информатика».

Рабочая программа содержит требования к уровню освоения содержания дисциплины, объем дисциплины и виды учебной работы, программу дисциплины и тематику практических занятий, учебно-методическое и информационное обеспечение.

*Учебное издание*

**Калайдин Евгений Николаевич**

**Построение и оценка моделей машинного обучения**

Рабочая программа дисциплины

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

Усл. п.л. 4,7. Изд. № \_\_\_\_\_ от \_\_\_\_\_. Тираж 100 экз.  
Заказ № \_\_\_\_\_

*Отпечатано в Краснодарском филиале Финуниверситета*

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

## Содержание

1. Наименование дисциплины4
2. Перечень планируемых результатов освоения образовательной программы (перечень компетенций) с указанием индикаторов их достижения и планируемых результатов обучения по дисциплине4
3. Место дисциплины в структуре образовательных программ5
4. Объем дисциплины (модуля) в зачетных единицах и в академических часах с выделением объема аудиторной (лекции, семинары) и самостоятельной работы обучающихся5
5. Содержание дисциплины, структурированное по темам (разделам) дисциплины с указанием их объемов (в академических часах) и видов учебных занятий6
6. Фонд оценочных средств для проведения промежуточной аттестации обучающихся по дисциплине13
7. Перечень основной и дополнительной учебной литературы, необходимой для освоения дисциплины19
8. Перечень ресурсов информационно-телекоммуникационной сети «Интернет», необходимых для освоения дисциплины19
9. Методические указания для обучающихся по освоению дисциплины20
10. Перечень информационных технологий, используемых при осуществлении образовательного процесса по дисциплине, включая перечень необходимого программного обеспечения и информационных справочных систем23
11. Описание материально-технической базы, необходимой для осуществления образовательного процесса по дисциплине23

## 1. Наименование дисциплины

«Построение и оценка моделей машинного обучения».

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

В совокупности с другими дисциплинами модуля учебная дисциплина «Построение и оценка моделей машинного обучения» обеспечивает формирование следующих компетенций: ПКН-4, ПКН-6.

Код компетенции	Наименование компетенции	Индикаторы достижения компетенции	Результаты обучения (умения и знания), соотнесенные с индикаторами достижения компетенции
ПКН-4	Способность разрабатывать информационные системы и алгоритмы на основе математических методов и моделей, в том числе из области искусственного интеллекта, в решении профессиональных задач	1. Демонстрирует знания в области оригинальных алгоритмов и программной реализации	<b>Знать:</b> теоретические основы в области оригинальных алгоритмов и программной реализации <b>Уметь:</b> демонстрировать знания в области оригинальных алгоритмов и программной реализации
		2. Создает оригинальные алгоритмические и программные средства в решении профессиональных задач	<b>Знать:</b> теоретические основы для создания оригинальных алгоритмических и программных средств в решении профессиональных задач <b>Уметь:</b> создавать оригинальные алгоритмические и программные средства в решении профессиональных задач
		3. Демонстрирует знания в области современных интеллектуальных технологий	<b>Знать:</b> теоретические основы в области современных интеллектуальных технологий <b>Уметь:</b> демонстрировать знания в области современных интеллектуальных технологий
		4. Выбирает интеллектуальные технологии и решает профессиональные задачи с их использованием	<b>Знать:</b> различные интеллектуальные технологии <b>Уметь:</b> выбирать интеллектуальные технологии и решать профессиональные задачи с их использованием
ПКН-6	Способность анализировать и оценивать эффективность применения	1. Демонстрирует знание основных методов прикладной математики и ин-	<b>Знать:</b> основные методы прикладной математики и информатики, применяемые в различных предметных областях

	методов прикладной математики и информатики	форматики, применяемых в различных предметных областях.	<b>Уметь:</b> различать основные методы прикладной математики и информатики, применяемые в различных предметных областях
		2. Владеет методологией математического моделирования для решения профессиональных задач	<b>Знать:</b> методологию математического моделирования для решения профессиональных задач  <b>Уметь:</b> владеть методологией математического моделирования для решения профессиональных задач

### 3. Место дисциплины в структуре образовательных программ

Дисциплина «Построение и оценка моделей машинного обучения» является дисциплиной Модуля общепрофессиональных дисциплин направления подготовки 01.04.02 - «Прикладная математика и информатика» направленности программы подготовки магистров: «Анализ больших данных и машинное обучение в экономике и финансах».

Изучение дисциплины «Построение и оценка моделей машинного обучения» базируется на знаниях, полученных в рамках изучения дисциплин «Введение в науки о данных», «Проектирование архитектур информационных систем», а также использует знания, полученные в ходе дисциплины «Методы визуализации данных».

Освоение дисциплины «Построение и оценка моделей машинного обучения» играет важную роль в дальнейшем успешном изучении дисциплин Модуля направленности программы магистратуры.

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

Вид учебной работы по дисциплине	Всего (в з/е и часах)	Модуль 3 (в часах)
<b>Общая трудоёмкость дисциплины</b>	<b>3/108</b>	<b>108</b>
<i>Контактная работа - Аудиторные занятия</i>	<i>32</i>	<i>32</i>
<i>Лекции</i>	<i>8</i>	<i>8</i>
<i>Семинары, практические занятия</i>	<i>24</i>	<i>24</i>
<b>Самостоятельная работа</b>	<b>76</b>	<b>76</b>
Вид текущего контроля	Контрольная работа	Контрольная работа
Вид промежуточной аттестации	Зачет	Зачет

## **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.**

Введение в машинное обучение. Искусственный интеллект, машинное обучение и глубокое обучение. История анализа данных. Основные понятия: обучение с учителем, обучение без учителя, обучение с подкреплением, модель алгоритмов, метод обучения, функция потерь и функционал качества, принцип минимизации эмпирического риска, обобщающая способность, скользящий контроль, обучающая и тестовая выборка. Кросс-валидация. Основные типы данных: таблицы, тексты, изображения, звук. Современные достижения.

Постановки задач в машинном обучении: классификация, регрессия, ранжирование, кластеризация, прогнозирование, сокращение размерности. Примеры прикладных задач.

#### **Тема 2. Подготовка и предварительный анализ данных.**

Табличные данные. Бинарные данные. Номинальные данные. Номинальные упорядоченные данные. Количественные данные. Трансформация данных. Борьба с пропусками. Источники данных. Импорт табличных данных в Pandas. Экспорт табличных данных в Pandas.

Описательная статистика. Корреляционный анализ. Фильтрация данных. Визуализация данных. Гистограммы. Линейные графики. Диаграммы рассеяния. Секторные диаграммы. Столбчатые диаграммы. Ящики с усами. Plotly. Matplotlib. Признаковое пространство. Целевая переменная.

#### **Тема 3. Модели регрессии.**

Теорема Гаусса-Маркова. Метод наименьших квадратов. Модель линейной регрессии. Парная регрессия. Множественная регрессия. Параметрическая регрессия. Оценка качества модели: множественный и парный коэффициент корреляции, индекс детерминации. Оценка значимости регрессии. Средне-квадратическая ошибка RMSE. Средняя абсолютная относительная ошибка MAPE. Нелинейные модели регрессии. Линеаризирующие преобразования функций. Непараметрическая регрессия. Локально взвешенный метод наименьших квадратов. Робастная непараметрическая регрессия. Алгоритм LOWESS.

#### **Тема 4. Логистическая регрессия.**

Предсказание вероятности принадлежности к классу. Логистическая функция или сигмоида. Переход от линейной регрессии к логистической. Понятие наблюдаемой вероятности и шансов. Функция потерь логистической регрессии. Градиентный спуск. Пакетный градиентный спуск. Стохастический градиентный спуск. Мини-пакетный градиентный спуск. Label encoding. One-hot encoding. Логистическая регрессия в задаче классификации.

## **Тема 5. Байесовский подход.**

Наивный Байес. Байесовский классификатор. Параметрическое восстановление плотности распределения вероятностей. Непараметрическое восстановление плотности распределения вероятностей. Метод окна Парзена с постоянной и переменной шириной окна. Ядерное сглаживание. Оценка Надарая-Ватсона с постоянной и переменной шириной окна. Выбор функции ядра. Проблема выбросов. Задача отсева выбросов. Мультиколлинеарность признаков. Методы устранения мульт-коллинеарности. Регуляризация корреляционной матрицы.

## **Тема 6. Метрические алгоритмы.**

Обобщённый метрический классификатор. Метод ближайших соседей kNN и его обобщения. Понятие выступа объекта. Алгоритм отбора эталонов. Проблема размерности. Выбор метрики. Обобщённый метрический регрессор. Метод потенциальных функций и его связь с линейной моделью классификации. Матрица ошибок, ошибка первого рода (ложно отрицательная), ошибка второго рода (ложно положительная). Критерии качества классификации: чувствительность и специфичность, ROC-кривая и AUC-кривая, доля верных распознаваний (Accuracy), точность (Precision) и полнота (Recall), F-мера, AUC-PR.

## **Тема 7. Методы сокращения размерности.**

Методы отбора признаков. Метод главных компонент. Собственное значение. Собственный вектор. Варимакс вращение. Факторные нагрузки. Многомерное шкалирование. Сингулярное разложение матриц. Сокращение размерности для визуализации.

## **Тема 8. Линейные алгоритмы классификации.**

Классификация линейной функцией. Линейная модель нейрона МакКаллока-Питтса. Градиентный метод численной минимизации. Персептрон Розенблатта. Дельта-правило ADALINE. Правило Хэбба. Понятие выступа для линейного классификатора. Непрерывные функции потерь: кусочно-линейная (SVM), кусочнолинейная (Hebb's rule), логарифмическая, квадратичная, сигмоидная, экспоненциальная, пороговая функция потерь. Проблема переобучения. Валидационная выборка. Разные штрафы за ошибки. Критерий минимума среднего риска.

## **Тема 9. Модели классификации Pandas и Scikit-Learn.**

Библиотека Scikit-Learn. Линейные классификаторы. Метод опорных векторов. Оптимальная разделяющая гиперплоскость. Понятие зазора между классами. Случай линейной разделимости и отсутствия линейной разделимости.

Классификаторы на базе деревьев решений. Энтропия. Глубина дерева. Алгоритм CART. Бинарная классификация. Мультиклассовая классификация.

Мультиклассовая логистическая регрессия. Многослойный персептрон.

## **Тема 10. Композиции алгоритмов.**

Понятие композиции (ансамбля) алгоритмов. Теоретические предпосылки композиционных методов. Одиночные модели и ансамбли моделей. Простое голосование. Взвешенное голосование. Аппроксимация функции потерь: сигмоидная, логарифмиче-

ская, кусочно-линейная, экспоненциальная, квадратичная. Последовательные методы: бустинг - алгоритмы AdaBoost, AnyBoost, категориальный бустинг (CatBoost), градиентный бустинг (модели xgBoost, LightGBM). Стохастические методы построения композиций (параллельные методы комитетов): бутстреп, бэггинг, метод случайных подпространств. Случайный лес (Random Forest). Оптимизация моделей. Прунинг.

### **Тема 11. Задача кластеризации.**

Постановка задачи кластеризации. Примеры прикладных задач. Типы кластерных структур. Оптимизационные постановки задач кластеризации и частичного обучения. Алгоритмы кластеризации: распространение сходства (Affinity Propagation), агломеративная (иерархическая) кластеризация (Agglomerative Clustering), сбалансированное итеративное сокращение и кластеризация с помощью иерархий (BIRCH), основанная на плотности пространственная кластеризация для приложений с шумами (DBSCAN), дивизивный алгоритм К-средних (K-means), мини-пакетный К-средних (Mini-Batch K-Means), сдвиг среднего значения (Mean Shift), упорядочение точек для обнаружения кластерной структуры (OPTICS), спектральная кластеризация (Spectral Clustering), модель гауссовой смеси (Mixture of Gaussians). EM-алгоритм. Построение дендрограммы. Оценка числа кластеров: метод локтя, коэффициент силуэта, критерий Акаике (AIC), Байесовский информационный критерий (BIC). Оценка качества кластеризации: индексы ARI и AMI, гомогенность, полнота, V-мера.

### **Тема 12. Прогнозирование временных рядов.**

Понятие временного ряда. Тренд и сезонность. Авторегрессия (AR). Скользящее среднее (MA). Авторегрессионное скользящее среднее (ARMA). Авторегрессионное интегрированное скользящее среднее (ARIMA). Сезонная авторегрессия с интегрированным скользящим средним (SARIMA). Сезонное авторегрессионное интегрированное скользящее среднее с экзогенными регрессорами (SARIMAX). Векторная авторегрессия (VAR). Векторная авторегрессия скользящего среднего (VARMA). Скользящее среднее векторной авторегрессии с экзогенными регрессорами (VARMAX). Простое экспоненциальное сглаживание (SES). Экспоненциальное сглаживание Хольта-Винтерса (HWES). Подбор параметров моделей. Прогнозирование и оценка качества.



## 5.2. Учебно-тематический план

№ п/ п	Наименование тем (разделов) дисциплины	Трудоёмкость в часах						Формы текуще- го контроля успеваемости
		Все- го	Контактная работа - Ауди- торная работа			Самостоятель- ная работа		
			Об- щая, в т.ч.:	Лек- ции	Семинары, практиче- ские занятия			
1.	Введение в ма- шинное обуче- ние.	9	2	1	1		7	Самостоятель- ные работы. Участие в реше- нии практиче- ских заданий на семинарских за- нятиях. Обсуж- дение решенных заданий.
2.	Подготовка и предваритель- ный анализ дан- ных.	9	2	1	1		7	
3.	Модели регрес- сии.	9	2	1	1		7	
4.	Логистическая регрессия.	8	2	1	1		6	
5.	Байесовский подход.	8	4	1	3		4	Самостоятель- ные работы. Участие в реше- нии практиче- ских заданий на семинарских за- нятиях.
6.	Метрические алгоритмы.	8	2	1	1		6	
7.	Методы сокра- щения размер- ности.	10	4	1	3		6	
8.	Линейные алго- ритмы класси- фикации.	10	2	-	2		8	
9.	Модели класси- фикации. Pandas и Scikit- Learn.	9	2	-	2		7	
10.	Композиции ал- горитмов.	9	2	1	1		7	
11.	Задача кластеризации.		10	4	-	4	6	
12.	Прогнозирование вре- менных рядов.		9	4	-	4	5	
	В целом по дисциплине		108	32	8	24	76	Согласно учеб- ному плану: контрольная работа
	Итого в %		100	30	8	22	70	

### 5.3. Содержание семинаров, практических занятий

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.	<ol style="list-style-type: none"> <li>1. Место машинного обучения относительно искусственного интеллекта и глубокого обучения.</li> <li>2. Типы машинного обучения.</li> <li>3. Методы машинного обучения.</li> <li>4. Постановки задач в машинном обучении: классификация, регрессия, ранжирование, кластеризация, прогнозирование. Примеры прикладных задач.</li> <li>5. Настройка среды Python. Импорт библиотек NumPy, Pandas, Plotly, Matplotlib, Scikit-Learn.</li> <li>6. Облачный сервис Google-Colab.</li> </ol> <p><i>Рекомендуемые источники литературы:</i> Основная литература: 8. [1,2,3] Дополнительная литература: 8. [1,3]</p>	<p>-работа с текстом лекции, разбор вопросов по теме занятия;</p> <p>-изучение рекомендованных к занятию литературных источников;</p> <p>- настройка среды программирования</p>
Подготовка и предварительный анализ данных.	<ol style="list-style-type: none"> <li>1. Импорт табличных данных в Pandas.</li> <li>2. Трансформация данных.</li> <li>3. Борьба с пропусками.</li> <li>4. Описательная статистика данных.</li> <li>5. Построение корреляционных матриц.</li> <li>6. Визуализация данных.</li> </ol> <p><i>Рекомендуемые источники литературы:</i> Основная литература: 8. [1,2,3] Дополнительная литература: 8. [1,4]</p>	<p>-работа с текстом лекции, разбор вопросов по теме занятия;</p> <p>-изучение рекомендованных к занятию литературных источников;</p> <p>-подготовка к семинарскому занятию;</p> <p>- выполнение практического задания по теме.</p>
Модели регрессии.	<ol style="list-style-type: none"> <li>1. Парная регрессия.</li> <li>2. Множественная регрессия.</li> <li>3. Оценка качества модели: множественный и парный коэффициент корреляции, индекс детерминации. Оценка значимости регрессии. Среднеквадратическая ошибка RMSE. Средняя абсолютная относительная ошибка MAPE.</li> <li>4. Локально взвешенный метод наименьших квадратов. Робастная непараметрическая регрессия. Алгоритм LOWESS.</li> </ol> <p><i>Рекомендуемые источники литературы:</i> Основная литература: 8. [1,2,3] Дополнительная литература: 8. [2, 3]</p>	<p>-работа с текстом лекции, разбор вопросов по теме занятия;</p> <p>-изучение рекомендованных к занятию литературных источников;</p> <p>-подготовка к семинарскому занятию;</p> <p>- выполнение практического задания по теме.</p>

<p>Логистическая регрессия.</p>	<ol style="list-style-type: none"> <li>1. Предсказание вероятности принадлежности к классу.</li> <li>2. Логистическая функция или сигмоида.</li> <li>3. Стохастический градиентный спуск.</li> <li>4. Label encoding. One-hot encoding.</li> </ol> <p>Логистическая регрессия для задачи классификации. Кросс-валидация.</p> <p><i>Рекомендуемые источники литературы:</i> Основная литература: 8. [1,2,3]. Дополнительная литература: 8. [1]</p>	<p>-работа с текстом лекции, разбор вопросов по теме занятия; -изучение рекомендованных к занятию литературных источников; -подготовка к семинарскому занятию; - выполнение практического задания по теме.</p>
<p>Байесовский подход.</p>	<ol style="list-style-type: none"> <li>1. Наивный Байес. Байесовский классификатор.</li> <li>2. Параметрическое восстановление плотности распределения.</li> <li>3. Непараметрическое восстановление плотности распределения.</li> <li>4. Метод окна Парзена с постоянной и переменной шириной окна.</li> <li>5. Ядерное сглаживание.</li> </ol> <p><i>Рекомендуемые источники литературы:</i> Основная литература: 8. [1,2,3] Дополнительная литература: 8. [1,3]</p>	<p>-работа с текстом лекции, разбор вопросов по теме занятия; -изучение рекомендованных к занятию литературных источников; -подготовка к семинарскому занятию; - выполнение практического задания по теме.</p>
<p>Метрические алгоритмы.</p>	<ol style="list-style-type: none"> <li>1. Обобщённый метрический классификатор.</li> <li>2. Метод ближайших соседей kNN и его обобщения.</li> <li>3. Понятие выступа объекта.</li> <li>4. Алгоритм отбора эталонов. Проблема размерности. Выбор метрики.</li> <li>5. Матрица ошибок, ошибка первого рода (ложно отрицательная), ошибка второго рода (ложно положительная).</li> <li>6. Критерии качества классификации: чувствительность и специфичность, ROC-кривая и AUC-кривая, доля верных распознаваний (Accuracy), точность (Precision) и полнота (Recall), F-мера.</li> </ol> <p><i>Рекомендуемые источники литературы:</i> Основная литература: 8. [1] Дополнительная литература: 8. [1,3,4]</p>	<p>-работа с текстом лекции, разбор вопросов по теме занятия;  -изучение рекомендованных к занятию литературных источников;  -подготовка к семинарскому занятию;  - выполнение практического задания по теме.</p>

Методы сокращения размерности.	<ol style="list-style-type: none"> <li>1. Методы отбора признаков.</li> <li>2. Метод главных компонент.</li> <li>3. Собственное значение. Собственный вектор.</li> <li>4. Сокращение размерности для визуализации.</li> </ol> <p><i>Рекомендуемые источники литературы:</i> Основная литература: 8. [1,2,3] Дополнительна литература: 8. [2,3]</p>	<ul style="list-style-type: none"> <li>-работа с текстом лекции, разбор вопросов по теме занятия;</li> <li>-изучение рекомендованных к занятию литературных источников;</li> <li>-подготовка к семинарскому занятию;</li> <li>- выполнение практического задания по теме</li> </ul>
Линейные алгоритмы классификации.	<ol style="list-style-type: none"> <li>1. Персептрон Розенблатта.</li> <li>2. Дельта-правило ADALINE.</li> <li>3. Правило Хэбба.</li> <li>4. Понятие выступа для линейного классификатора.</li> <li>5. Непрерывные функции потерь.</li> </ol> <p><i>Рекомендуемые источники литературы:</i> Основная литература: 8. [2,3] Дополнительная литература: 8. [3,4]</p>	<ul style="list-style-type: none"> <li>-работа с текстом лекции, разбор вопросов по теме занятия;</li> <li>-изучение рекомендованных к занятию литературных источников;</li> <li>-подготовка к семинарскому занятию;</li> <li>- выполнение практического задания по теме.</li> </ul>
Модели классификации. Pandas и Scikit-Learn.	<ol style="list-style-type: none"> <li>1. Библиотека Scikit-Learn.</li> <li>2. Линейные классификаторы. Метод опорных векторов. Оптимальная разделяющая гиперплоскость.</li> <li>3. Классификаторы на базе деревьев решений. Энтропия. Глубина дерева.</li> <li>4. Алгоритм CART.</li> <li>5. Бинарная классификация.</li> <li>6. 6. Мультиклассовая классификация.</li> </ol> <p><i>Рекомендуемые источники литературы:</i> Основная литература: 8. [1,2,3] Дополнительная литература: 8. [1]</p>	<ul style="list-style-type: none"> <li>-работа с текстом лекции, разбор вопросов по теме занятия;</li> <li>-изучение рекомендованных к занятию литературных источников;</li> <li>-подготовка к семинарскому занятию;</li> <li>- выполнение практического задания по теме.</li> </ul>
Композиции алгоритмов.	<ol style="list-style-type: none"> <li>1. Понятие композиции (ансамбля) алгоритмов.</li> <li>2. Последовательные методы: бустинг - алгоритмы AdaBoost, AnyBoost, категориальный бустинг (CatBoost), градиентный бустинг (модели xgBoost, LightGBM).</li> <li>3. Стохастические методы построения Композиций (параллельные методы комитетов): бутстреп, бэггинг.</li> <li>4. Случайный лес (Random Forest).</li> </ol> <p><i>Рекомендуемые источники литературы:</i> Основная лите-</p>	<ul style="list-style-type: none"> <li>-работа с текстом лекции, разбор вопросов по теме занятия;</li> <li>-изучение рекомендованных к занятию литературных источников;</li> <li>-подготовка к семинарскому занятию;</li> <li>- выполнение прак-</li> </ul>

	ратура: 8. [1,2,3] Дополнительная литература: 8. [1,2]	тического задания по теме.
Задача кластеризации.	<ol style="list-style-type: none"> <li>1. Постановка задачи кластеризации.</li> <li>2. Агломеративная (иерархическая) кластеризация (Agglomerative Clustering).</li> <li>3. Сбалансированное итеративное сокращение и кластеризация с помощью иерархий (BIRCH).</li> <li>4. Основанная на плотности пространственная кластеризация для приложений с шумами (DBSCAN).</li> <li>5. Дивизивный алгоритм К-средних (K- means),</li> <li>6. Модель гауссовой смеси (Mixture of Gaussians).</li> <li>7. Оценка числа кластеров: метод локтя, коэффициент силуэта.</li> </ol> <p><i>Рекомендуемые источники литературы:</i> Основная литература: 8. [1,2,3] Дополнительная литература: 8. [1,2,4]</p>	<p>-работа с текстом лекции, разбор вопросов по теме занятия;</p> <p>-изучение рекомендованных к занятию литературных источников;</p> <p>-подготовка к семинарскому занятию;</p> <p>- выполнение практического задания по теме.</p>
Прогнозирование временных рядов.	<ol style="list-style-type: none"> <li>1. Понятие временного ряда. Тренд и сезонность.</li> <li>2. Авторегрессия (AR).</li> <li>3. Скользящее среднее (MA).</li> <li>4. Авторегрессионное скользящее среднее (ARMA).</li> <li>5. Авторегрессионное интегрированное скользящее среднее (ARIMA).</li> <li>6. Векторная авторегрессия (VAR).</li> <li>7. Векторная авторегрессия скользящего среднего (VARMA).</li> <li>8. Простое экспоненциальное сглаживание (SES).</li> <li>9. Прогнозирование и оценка качества.</li> </ol> <p><i>Рекомендуемые источники литературы:</i> Основная литература: 8. [1,2,3] Дополнительная литература: 8. [2,3,4]</p>	<p>-работа с текстом лекции, разбор вопросов по теме занятия;</p> <p>-изучение рекомендованных к занятию литературных источников;</p> <p>-подготовка к семинарскому занятию;</p> <p>- выполнение практического задания по теме.</p>

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

Оценочные средства предназначены для контроля и оценки образовательных достижений обучающихся, освоивших программу учебной дисциплины «Построение и оценка моделей машинного обучения».

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

***Вопросы для оценки знаний и умений, характеризующих формирование компетенций***

Шифр компе-	Вопросы	Правильный ответ
-------------	---------	------------------

тенции		
ПКН-4	1. В каком году был создан первый искусственный интеллект?	1950
	2. Какие модели и методы используются для обработки естественного языка?	Нейронные сети, статистические модели, методы машинного обучения
	3. Какие модели и методы используются для компьютерного зрения?	CNN, GAN, RNN
	4. Какая модель обучения с подкреплением основана на методе Монте-Карло?	Модель Монте-Карло
	5. Как называется функция, которая используется для оценки качества модели в задачах классификации?	Функция качества классификации
	6. Что такое гиперпараметры?	Параметры, определяющие сложность модели
	7. С чем может быть связано переобучение модели?	Низкая обобщающая способность модели
	8. Какая проблема может возникнуть при работе с данными в машинном обучении?	Проблема несбалансированных классов
	9. Какой из методов предобработки наиболее эффективен для решения задачи классификации?	Стандартизация
	10. К какой категории методов оптимизации относится метод градиентного спуска?	Нелинейным
ПКН-6	11. Что позволяет теорема Байеса?	Учитывать априорные распределения вероятностей при обучении модели
	12. Какие задачи решаются с помощью глубокого обучения?	Классификация и регрессия
	13. Какую библиотеку можно использовать для анализа данных и выполнения статистических операций?	Pandas
	14. В чем состоит особенность модели дерева решений?	Построение иерархической структуры решений
	15. Какая модель используется для предсказания непрерывных целевых переменных?	Модель линейной регрессии
	16. Для чего используются ядерные методы в машинном обучении?	Аппроксимация нелинейных функций
	17. Для чего используется метод DBSCAN (Density-Based Spatial Clustering of Applications with Noise)?	Нахождение K-ближайших соседей и выделение кластеров
	18. Как выбираются локальные веса в локально-взвешенном сглаживании?	На основе расстояния до соседей
	19. Что такое бэггинг?	Метод ансамблевого обучения
	20. Для чего используется ARIMA (Autoregressive Integrated Moving Average)?	Прогнозирование временных рядов

## Практико-ориентированные задания

Шифр компетенции	Практико-ориентированные задания	Правильный ответ
ПКН-4	1. В базе данных имеется 10 000 записей. Сколько информации (в битах) необходимо для кодирования одной записи, если на каждый атрибут выделяется по 3 бита?	29
	2. В задаче регрессии имеется 500 точек данных. Сколько параметров требуется настроить в линейной регрессии, если каждая точка представлена двумя координатами (x, y)?	1000
	3. Имеется набор данных для классификации, состоящий из 50 000 примеров. Сколько раз необходимо выполнить итерацию обучения алгоритма k-ближайших соседей (k-NN), если в каждой итерации используется 10% набора данных для проверки?	10
	4. Ваша задача разработать систему рекомендаций для электронной коммерции. На основе какой фильтрации будет оптимально это сделать?	Коллаборативной
	5. Перед вами стоит задача построить модель прогнозируемых временных рядов. Какие модельные формы вы будете для этого использовать?	ARIMA, LSTM
	6. Дан набор данных, состоящий из последовательностей действий пользователей на веб-сайте. Требуется классифицировать эти действия, то есть определить, какие из них являются целевыми, а какие – нет. Что для этого нужно сделать?	Определить цели, собрать данные, создать модель классификации и внедрить её на веб-сайте
	7. Доктор хочет построить модель, прогнозирующую возможность развития осложнений у пациента после операции - все ли будет благополучно в течение нескольких следующих месяцев или нет. К какому типу относится данная задача?	Классификация
	8. Астроном хочет построить модель, которая сможет разбить известные науке звезды на группы по их характеристикам, чтобы лучше изучить их особенности. К какому типу относится данная задача?	Кластеризация
	9. Заполните пропуск: «Для выпуклой функции .... указывает сторону наискорейшего убывания»	Градиент
ПКН-6	10. Что может являться критерием останова в градиентном спуске?	Достижение определенной точности

	11. Рассмотрим два объекта: у первого отступ линейного классификатора равен -10, у второго -1000. Как соотносятся два этих объекта?	Второй объект находится дальше от разделяющей поверхности, чем первый
	12. Рассмотрим пользователя социальной сети как объект в задаче машинного обучения. Что может являться задачей классификации?	Предсказание пола пользователя
	13. Ансамблевые методы обычно делают более качественные предсказания, чем отдельные алгоритмы (у ансамблей ниже ошибка на тестовой выборке). А какой показатель всегда ухудшается при ансамблировании?	Скорость выполнения предсказаний
	14. Алгоритм бэггинг подразумевает выбор случайных частей данных, обучение алгоритма (например, решающего дерева) на каждой части и составление ансамбля из обученных алгоритмов. Для чего нужно обучать алгоритмы на разных частях данных?	Для выполнения несовпадающих предсказаний
	15. Метеоролог хочет построить модель, прогнозирующую температуру воздуха на завтра. К какому типу относится данная задача?	Прогнозирование
	16. Верно ли суждение, что элементами обучающей выборки являются объекты, характеристики которых являются значениями признаков	Да
	17. Рассмотрим признак “Образовательная программа” при анализе данных по студентам университета. Этот признак может принимать три значения: “Экономика”, “Математика”, “Философия”. Воспользуемся one-hot кодированием и заменим этот признак на три бинарных, которые будут соответствовать категориям в том порядке, в котором они перечислены выше. Как будет закодирован признак со значением “Философия”?	(0,0,1)
	18. Чему будет равен корень из среднеквадратичной ошибки для набора из 3 наблюдений, где отклонение предсказания линейной регрессии от реальных значений равны: -1, 2, 2?	2

## Тесты

Шифр компетенции	Тестовые задания	Правильный ответ
ПКН-4	1. Определите какой из методов машинного обучения	с)



	используется для прогнозирования временных рядов. а) Метод опорных векторов б) Деревья решений с) Метод k-ближайших соседей д) Регрессия	
	2. Переобучение в контексте машинного обучения — это ... а) Процесс подбора оптимальных параметров модели машинного обучения б) Ситуация, когда модель идеально подходит под обучающую выборку, но плохо работает на новых данных с) Потеря обобщающей способности модели д) Невозможность подобрать оптимальные параметры модели	б)
	3. Выберите метрику оценки качества модели машинного обучения для задачи классификации, которая отражает долю верно классифицированных объектов обучающей выборки: а) Точность б) Полнота с) F1-мера д) ROC-AUC	а)
	4. Выберите какие из перечисленных характеристик могут быть использованы для оценки качества модели машинного обучения. а) Количество деревьев в ансамблевой модели б) Время обучения модели с) Точность модели на обучающей и тестовой выборке д) Среднее время обработки одного объекта	с)
	5. Определите, что из перечисленного может являться причиной переобучения модели машинного обучения. а) Недостаточное количество данных для обучения б) Избыточное количество гиперпараметров модели с) Недостаточное количество слоев в нейронной сети д) Избыточная сложность модели	д)
	6. Метод регуляризации — это ... а) Добавление случайного шума в обучающую выборку б) Отбрасывание некоторых объектов из обучающей выборки с) Ограничение сложности модели для предотвращения переобучения д) Увеличение объема обучающей выборки	с)
ПКН-6	7. Определите метод, который используется для обнаружения переобучения модели машинного обучения. а) Кросс-валидация б) Регуляризация	а)

	<ul style="list-style-type: none"> <li>c) Градиентный бустинг</li> <li>d) Ансамблирование моделей</li> </ul>	
	<p>8. Определите инструмент, который используется для визуализации процесса обучения нейронной сети и анализа ее производительности.</p> <ul style="list-style-type: none"> <li>a) TensorBoard</li> <li>b) Jupyter Notebook</li> <li>c) PyCharm</li> <li>d) Visual Studio Code</li> </ul>	a)
	<p>9. Определите метод, который используется для определения оптимального числа кластеров при кластеризации данных.</p> <ul style="list-style-type: none"> <li>a) Метод локтя (Elbow Method)</li> <li>b) Алгоритм DBSCAN</li> <li>c) Алгоритм k-ближайших соседей</li> <li>d) Алгоритм опорных векторов (SVM)</li> </ul>	a)
	<p>10. Определите метод машинного обучения, который обычно используется для решения задач регрессии.</p> <ul style="list-style-type: none"> <li>a) Метод опорных векторов</li> <li>b) Деревья решений</li> <li>c) Метод k-ближайших соседей</li> <li>d) Линейная регрессия</li> </ul>	d)
	<p>11. Определите, что показывает ROC-кривая.</p> <ul style="list-style-type: none"> <li>a) Чувствительность и специфичность модели</li> <li>b) Вероятность ложного срабатывания и пропуска события</li> <li>c) Точность и полноту модели</li> <li>d) Вероятность ошибки модели</li> </ul>	b)
	<p>12. Выберите один или несколько вариантов ответа, в которых используются ансамблевые модели машинного обучения.</p> <ul style="list-style-type: none"> <li>a) Если модель работает хорошо на обучающих данных, но имеет низкую обобщающую способность</li> <li>b) Если задача имеет высокие размерности входных данных и малое количество обучающих образцов</li> <li>c) Если требуется повысить точность модели</li> <li>d) Если результаты модели неустойчивы и подвержены влиянию случайных флуктуаций</li> </ul>	c)

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

### **7.1. Основная литература**

1. Кондрашов, Ю. Н., Анализ данных и машинное обучение на платформе MS SQL Server : учебное пособие / Ю. Н. Кондрашов. — Москва : Русайнс, 2024. — 303 с. — ISBN 978-5-466-01955-1. — URL: <https://book.ru/book/947076>. — Текст : электронный.

2. Андриянов, Н. А. Построение и оценка моделей машинного обучения. 01.04.02 «Прикладная математика и информатика», всех профилей (программы подготовки магистров) : учебное пособие / Н. А. Андриянов, П. В. Никитин. — Москва : Финансовый университет, 2024. — 140 с. — Текст : электронный // Лань : электронно-библиотечная система. — URL: <https://e.lanbook.com/book/345374> — Режим доступа: для авториз. пользователей.

3. Сорокин, А. Б. Технологии обучения: кластеризация и классификация : учебное пособие / А. Б. Сорокин, Л. М. Железняк. — Москва : РТУ МИРЭА, 2021. — 49 с. — ЭБС Лань. — URL: <https://e.lanbook.com/book/182493>. — Текст : электронный.

### **7.2. Дополнительная литература**

1. Платонов, А. В. Машинное обучение : учебное пособие для вузов / А. В. Платонов. — Москва : Издательство Юрайт, 2024. — 85 с. — (Высшее образование). — ISBN 978-5-534-15561-7. — Текст : электронный // Образовательная платформа Юрайт [сайт]. — URL: <https://ezpro.fa.ru:2058/bcode/520544> .

2. Андрейчиков А. В. Интеллектуальные информационные системы и методы искусственного интеллекта : учебник / А. В. Андрейчиков, О. Н. Андрейчикова. — Москва : ИНФРА-М, 2021. — 530 с. — ЭБС ZNANIUM.com. — URL: <https://znanium.com/catalog/product/1009595>. - Текст : электронный.

3. Григорьев, А. А. Методы и алгоритмы обработки данных : учебное пособие / А. А. Григорьев, Е. А. Исаев. — 2-е изд., перераб. и доп. — Москва : ИНФРА-М, 2021. — 383 с. — ЭБС ZNANIUM.com. — URL: <https://znanium.com/catalog/product/1032305>. - Текст : электронный.

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

1. Личный кабинет обучающегося <https://org.fa.ru>
2. Сайт библиотеки Scikit-Learn <https://scikit-learn.org/stable/>
3. Электронная библиотека Финансового университета (ЭБ) <http://elib.fa.ru/>
4. Электронно-библиотечная система BOOK.RU <http://www.book.ru>
5. Электронно-библиотечная система «Университетская библиотека ОНЛАЙН» <http://biblioclub.ru/>
6. Электронно-библиотечная система Znanium <http://www.znanium.com>

7. Электронно-библиотечная система издательства «ЮРАЙТ» <https://urait.ru/>
8. Электронно-библиотечная система издательства Проспект <http://ebs.prospekt.org/books>
9. Электронно-библиотечная система издательства «Лань» <https://e.lanbook.com/>
10. Деловая онлайн-библиотека Alpina Digital <http://lib.alpinadigital.ru/>
11. Научная электронная библиотека eLibrary.ru <http://elibrary.ru>
12. Национальная электронная библиотека <http://нэб.пф/>
13. СПАРК <https://spark-interfax.ru/>
14. Academic Reference <http://ar.cnki.net/ACADREF>
15. Пакет баз данных компании EBSCO Publishing, крупнейшего агрегатора научных ресурсов ведущих издательств мира <http://search.ebscohost.com>
16. Электронные продукты издательства Elsevier <http://www.sciencedirect.com>
17. Emerald: Management eJournal Portfolio <https://www.emerald.com/insight/>
18. Scopus <https://www.scopus.com>
19. Электронная коллекция книг издательства Springer: Springer eBooks <http://link.springer.com/>

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

Цель методических рекомендаций - обеспечить студенту магистратуры (далее - студенту) оптимальную организацию процесса изучения дисциплины, а также выполнения различных форм самостоятельной работы.

### **Методические рекомендации по изучению дисциплины**

Студентам необходимо ознакомиться:

-с содержанием рабочей программы дисциплины (далее - РПД), с целями и задачами дисциплины, ее связями с другими дисциплинами образовательной программы, методическими разработками по данной дисциплине, имеющимися на образовательном портале и сайте департамента, с графиком консультаций преподавателей.

### **Рекомендации по подготовке к лекционным занятиям (теоретический курс)**

Изучение дисциплины требует систематического и последовательного накопления знаний, следовательно, пропуски отдельных тем не позволяют глубоко освоить предмет. Именно поэтому контроль над систематической работой студентов всегда находится в центре внимания департамента.

Студентам рекомендуется:

-перед каждой лекцией просматривать рабочую программу дисциплины, что позволит сэкономить время на записывание темы лекции, ее основных вопросов, рекомендуемой литературы;

-на отдельные лекции приносить соответствующий материал на бумажных или электронных носителях, представленный лектором на портале. Данный материал будет охарактеризован, прокомментирован, дополнен непосредственно на лекции;

-перед очередной лекцией необходимо просмотреть по конспекту материал предыдущей лекции. При затруднениях в восприятии материала следует обратиться к

основным литературным источникам, если разобраться в материале опять не удалось, то обратитесь к лектору (по графику его консультаций) или к преподавателю на практических занятиях. Не оставляйте «белых пятен» в освоении материала.

### **Рекомендации по подготовке к практическим (семинарским) занятиям**

Студентам следует:

- приносить с собой рекомендованную преподавателем литературу к конкретному занятию;
- до очередного практического занятия по рекомендованным литературным источникам проработать теоретический материал, соответствующей темы занятия;
- при подготовке к практическим занятиям следует обязательно использовать не только лекции, но и другую учебную литературу;
- в начале занятий задать преподавателю вопросы по материалу, вызвавшему затруднения в его понимании и освоении, при решении задач, заданных для самостоятельного решения;
- в ходе семинара давать конкретные, четкие ответы по существу вопросов;
- на занятии доводить каждую задачу до окончательного решения, демонстрировать понимание проведенных расчетов (анализов, ситуаций), в случае затруднений обращаться к преподавателю.

### **Методические рекомендации по выполнению различных форм самостоятельных домашних заданий**

Самостоятельная работа студентов включает в себя выполнение различного рода заданий, которые ориентированы на более глубокое усвоение материала изучаемой дисциплины. К выполнению заданий для самостоятельной работы предъявляются следующие требования: задания должны выполняться самостоятельно и представляться в установленный срок, а также соответствовать установленным требованиям по оформлению.

Студентам следует:

- руководствоваться графиком самостоятельной работы, определенным РПД; выполнять все плановые задания, выдаваемые преподавателем для самостоятельного выполнения, и разбирать на семинарах и консультациях неясные вопросы;
- использовать при подготовке нормативные документы Финансового университета;
- при подготовке к экзамену параллельно прорабатывать соответствующие теоретические и практические разделы дисциплины, фиксируя неясные моменты для их обсуждения на плановой консультации.

### **Методические рекомендации по работе с литературой**

Любая форма самостоятельной работы студента (подготовка к семинарскому занятию, выполнение практического задания для самостоятельной работы, начинается с изучения соответствующей литературы как в библиотеке, так и дома.

К каждой теме учебной дисциплины подобрана основная и дополнительная литература.

При работе с литературой рекомендуется делать записи. Записи в той или иной форме не только способствуют пониманию и усвоению изучаемого материала, но и помогают вырабатывать навыки явного изложения в письменной форме тех или иных теоретических вопросов.

### **Методические указания по проведению практических занятий**

По структуре практические занятия следует разделить на учебные и контрольные.

**Учебные практические занятия** структурно состоят из следующих компонент:

- проверка наличия выполненного задания самостоятельной работы каждого студента;
- выборочная проверка корректности выполнения домашнего задания;
- разбор типичных ошибок, возникших в самостоятельной работе;
- рассмотрение теоретических вопросов, связанных с текущим практическим занятием;
- разбор методов выполнения практических заданий и решения задач;
- корректировка заданий для самостоятельной работы студентов.

**Контрольные практические занятия** структурно состоят из следующих компонент:

- проведение аудиторных самостоятельных работ;
- подведение итогов и разбор типичных ошибок, возникших при выполнении самостоятельных работ.

Студенты должны обратить внимание на перечень основных контрольных мероприятий, которые проводятся в соответствии с рабочей программой дисциплины.

Конкретные сроки проведения этих мероприятий своевременно доводятся до сведения студентов.

Контрольные работы проводятся в форме письменного экспресс-опроса, открытых и закрытых тестовых заданий, решения ситуационных и практических задач по темам учебной дисциплины

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

### **10.1. Комплект лицензионного программного обеспечения:**

- ОС Astra Linux
- Libre Office
- Антивирус Kaspersky

### **10.2. Современные профессиональные базы данных и информационные справочные системы:**

- Информационно-правовая система «Консультант Плюс»;
- Информационно-правовая система «Гарант»;
- Система комплексного раскрытия информации «СКРИН»: <https://skrin.ru>

### **10.3. Сертифицированные программные и аппаратные средства защиты информации - не предусмотрены.**

## **11. 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 Краснодарского филиала Финансового университета.