

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

**федеральное государственное автономное  
образовательное учреждение высшего образования\_  
«Национальный исследовательский Нижегородский государственный университет  
им. Н.И. Лобачевского»**

Институт филологии и журналистики

---

УТВЕРЖДЕНО

решением президиума Ученого совета ННГУ

протокол № 1 от 16.01.2024 г.

**Рабочая программа дисциплины**

Машинное обучение и анализ больших данных

---

Уровень высшего образования

Магистратура

---

Направление подготовки / специальность

42.04.02 - Журналистика

---

Направленность образовательной программы

Искусственный интеллект и журналистика данных

---

Форма обучения

очная

---

г. Нижний Новгород

2024 год начала подготовки

## 1. Место дисциплины в структуре ОПОП

Дисциплина Б1.О.07 Машинное обучение и анализ больших данных относится к обязательной части образовательной программы.

## 2. Планируемые результаты обучения по дисциплине, соотнесенные с планируемыми результатами освоения образовательной программы (компетенциями и индикаторами достижения компетенций)

Формируемые компетенции (код, содержание компетенции)	Планируемые результаты обучения по дисциплине (модулю), в соответствии с индикатором достижения компетенции		Наименование оценочного средства	
	Индикатор достижения компетенции (код, содержание индикатора)	Результаты обучения по дисциплине	Для текущего контроля успеваемости	Для промежуточной аттестации
ОПК-6: Способен отбирать и внедрять в процесс медиапроизводства современные технические средства и информационно-коммуникационные технологии	ОПК-6.1: Отслеживает глобальные тенденции модернизации технического оборудования, программного обеспечения и расходных материалов, необходимых для осуществления профессиональной деятельности ОПК-6.2: Адаптирует возможности новых стационарных и мобильных цифровых устройств к профессиональной деятельности журналиста	ОПК-6.1: Знать глобальные тенденции модернизации технического оборудования, программного обеспечения и расходных материалов, необходимых для осуществления профессиональной деятельности; Уметь отслеживать глобальные тенденции модернизации технического оборудования, программного обеспечения и расходных материалов, необходимых для осуществления профессиональной деятельности; Владеть навыками практического использования знаний о глобальных тенденциях модернизации технического оборудования, программного обеспечения и расходных материалов, необходимых для осуществления профессиональной деятельности  ОПК-6.2: Знать технологии и возможности использования современных стационарных и	Контрольная работа	Экзамен: Контрольные вопросы

		<p>мобильных цифровых устройств на всех этапах создания журналистского текста и (или) продукта; специфику работы журналиста в условиях мультимедийной среды;</p> <p>Уметь применять и использовать возможности современные стационарные и мобильные цифровые устройства на всех этапах создания журналистского текста и (или) продукта;</p> <p>применять в профессиональной деятельности методы и технологии подготовки медиапродукта в разных знаковых системах (вербальной, аудио-, видео-, графика, анимация)</p> <p>Владеть навыками применения современных стационарных и мобильных цифровых устройств на всех этапах создания журналистского текста и (или) продукта;</p> <p>навыками создания журналистских текстов в условиях мультимедийной среды</p>		
<p>ПКД-1: Способен применять математические методы решения прикладных задач в профессиональной деятельности</p>	<p>ПКД-1.1: Знает математические методы решения прикладных задач и возможности их применения в профессиональной деятельности</p> <p>ПКД-1.2: Использует в процессе осуществления профессиональной деятельности современные информационные технологии, в том числе технологии искусственного интеллекта</p>	<p>ПКД-1.1:</p> <p>Знает методы комбинаторики линейной алгебры, теории вероятностей решения прикладных задач профессиональной деятельности</p> <p>ПКД-1.2:</p> <p>Умеет применять методы комбинаторики линейной алгебры, теории вероятностей решения прикладных задач профессиональной деятельности</p>	Тест	<p>Экзамен:</p> <p>Контрольные вопросы</p>

### 3. Структура и содержание дисциплины

#### 3.1 Трудоемкость дисциплины

	<b>очная</b>
<b>Общая трудоемкость, з.е.</b>	<b>3</b>
<b>Часов по учебному плану</b>	<b>108</b>
в том числе	
<b>аудиторные занятия (контактная работа):</b>	
- занятия лекционного типа	<b>14</b>
- занятия семинарского типа (практические занятия / лабораторные работы)	<b>14</b>
- КСР	<b>2</b>
<b>самостоятельная работа</b>	<b>42</b>
<b>Промежуточная аттестация</b>	<b>36</b> <b>Экзамен</b>

#### 3.2. Содержание дисциплины

(структурированное по темам (разделам) с указанием отведенного на них количества академических часов и виды учебных занятий)

Наименование разделов и тем дисциплины	Всего (часы)	в том числе			
		Контактная работа (работа во взаимодействии с преподавателем), часы из них			Самостоятельная работа обучающегося, часы
		Занятия лекционного типа	Занятия семинарского типа (практические занятия/лабораторные работы), часы	Всего	
	о ф о	о ф о	о ф о	о ф о	о ф о
Постановка задач машинного обучения	21	4	4	8	13
Обучение с учителем	21	4	4	8	13
Обучение без учителя	28	6	6	12	16
Аттестация	36				
КСР	2			2	
Итого	108	14	14	30	42

#### Содержание разделов и тем дисциплины

Вероятностная постановка задачи обучения по прецедентам. Принцип минимизации эмпирического риска. Байесовская теория решений. Принцип максимума апостериорной вероятности. Байесов классификатор. Экспериментальные методы оценки качества обучения. Метод наименьших квадратов для решения задачи восстановления регрессии. Его вывод на основе метода максимального правдоподобия. Линейная регрессионная модель. Система нормальных уравнений. Основы регрессионного анализа (проверка значимости коэффициентов, коэффициент детерминации Пирсона,

доверительные интервалы, анализ остатков). Проблема переобучения при решении задачи восстановления регрессии. Методы борьбы с переобучением: сокращение числа параметров, регуляризация (ридж-регрессия), метод лассо. Трудоемкость методов. Метод ближайших соседей для решения задачи классификации. Теорема об оценке риска в методе ближайшего соседа. Наивный байесовский классификатор. Линейный дискриминантный анализ. Квадратичный дискриминантный анализ. Логистическая регрессия.

Нейронные сети. Персептрон Розенблатта. Алгоритм обучения персептрона как метод стохастического градиентного спуска. Нейронные сети для решения задач классификации и восстановления регрессии. Обучение сети. Регуляризация как метод борьбы с переобучением. Понятие о глубоких нейронных сетях. Машина опорных векторов. Ядра и спрямляющие пространства.

Деревья решений. Метод CART (classification and regression trees) для решения задач классификации и восстановления регрессии. Отсечения ветвей и выбор финального дерева. Методы обработки пропущенных значений. Ансамбли решающих правил (классификаторов). Простое и взвешенное голосование. Бустинг. Алгоритм AdaBoost. Оценка ошибки предсказания. Бустинг и аддитивные модели. Градиентный бустинг. Алгоритм градиентного бустинга деревьев решений (MART). Баггинг. Алгоритм случайных деревьев (случайный лес.). Обучение без учителя. Кластеризация. Кластеризация методами теории графов. Метод центров тяжести. Метод медиан. Метод нечетких множеств. ЕМ-алгоритм. Иерархическая кластеризация. Агломеративные и разделяющие методы. Основы теории Вапника–Червоненкиса. Лемма Бернштейна. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечного класса решающих правил. Обоснование принципа минимизации эмпирического риска. Размерность Вапника–Червоненкиса. Лемма Зауэра. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечной размерности Вапника–Червоненкиса. Принцип структурной минимизации риска.

#### **4. Учебно-методическое обеспечение самостоятельной работы обучающихся**

Самостоятельная работа обучающихся включает в себя подготовку к контрольным вопросам и заданиям для текущего контроля и промежуточной аттестации по итогам освоения дисциплины приведенным в п. 5.

Для подготовки к практическим занятиям, самостоятельной работы и промежуточной аттестации рекомендуются электронные ресурсы <http://www.uic.unn.ru/~zny/ml/>

#### **5. Фонд оценочных средств для текущего контроля успеваемости и промежуточной аттестации по дисциплине (модулю)**

**5.1 Типовые задания, необходимые для оценки результатов обучения при проведении текущего контроля успеваемости с указанием критериев их оценивания:**

**5.1.1 Типовые задания (оценочное средство - Контрольная работа) для оценки сформированности компетенции ОПК-6:**

1. Дана обучающая выборка

$x_1$	0	1	1	0	0	1	1	2	6
$x_2$	3	3	1	0	1	1	2	3	1
$y$	0	0	0	0	1	1	1	1	1

Методом линейного дискриминантного анализа для каждого класса построить дискриминантную функцию и записать уравнение разделяющей поверхности.

2. Дана обучающая выборка (см. таблицу выше). Методом квадратичного дискриминантного анализа построить дискриминантные функции.
3. Дана обучающая выборка (см. таблицу выше). С помощью наивного байесова классификатора оценить вероятности  $P(Y = 1 | x_1 = 1, x_2 = 2)$
4. Дана обучающая выборка:

$x$	-1	0	0	1	2
$y$	1	-2	1	7	8

Методом наименьших квадратов построить полиномиальную модель вида  $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$ .

5. Дана обучающая выборка (см. выше). Методом ридж-регрессии построить полиномиальную модель вида  $f(x) = \beta_0 + \beta_1 x + \beta_2 x^2$ , если параметр регуляризации  $\lambda = 2$ .

### Критерии оценивания (оценочное средство - Контрольная работа)

Оценка	Критерии оценивания
зачтено	Все задачи решены полностью или решение задач обосновано, но допущены арифметические ошибки
не зачтено	Задачи не решены или сделан только первый этап в решении задач

### 5.1.2 Типовые задания (оценочное средство - Тест) для оценки сформированности компетенции ПКД-1:

- Пусть ошибка классификатора, построенного по методу  $k$  ближайших соседей, на тестовой выборке много больше ошибки на обучающей выборке. Какое из следующих действий разумно предпринять для исправления ситуации?
  - увеличить  $k$  и построить новый классификатор (+)
  - уменьшить  $k$  и построить новый классификатор
  - ничего из вышеперечисленного
- Метод  $k$ -ближайших соседей применяется для решения
  - только задач классификации
  - только задач восстановления регрессии
  - как задач классификации, так и задач восстановления регрессии (+)
- Если объекты из обучающей выборки, принадлежащие разным классам, линейно неразделимы, то
  - метод опорных векторов применять нельзя
  - можно применять метод опорных векторов только с нелинейным ядром
  - можно применять машину опорных векторов с любым ядром (+)

### Критерии оценивания (оценочное средство - Тест)

Оценка	Критерии оценивания
зачтено	51-100%
не зачтено	0-50%

## 5.2. Описание шкал оценивания результатов обучения по дисциплине при промежуточной аттестации

### Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатор достижения компетенций)	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	не зачтено			зачтено			
<u>Знания</u>	Отсутствие знаний теоретического материала. Невозможность оценить полноту знаний вследствие отказа обучающегося от ответа	Уровень знаний ниже минимальных требований. Имели место грубые ошибки	Минимально допустимый уровень знаний. Допущено много негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько негрубых ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Допущено несколько несущественных ошибок	Уровень знаний в объеме, соответствующем программе подготовки. Ошибок нет.	Уровень знаний в объеме, превышающем программу подготовки.
<u>Умения</u>	Отсутствие минимальных умений. Невозможность оценить наличие умений вследствие отказа обучающегося от ответа	При решении стандартных задач не продемонстрированы основные умения. Имели место грубые ошибки	Продemonстрированы основные умения. Решены типовые задачи с негрубыми ошибками. Выполнены все задания, но не в полном объеме	Продemonстрированы все основные умения. Решены все основные задачи с негрубыми ошибками. Выполнены все задания в полном объеме, но некоторые с недочетами	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания в полном объеме, но некоторые с недочетами.	Продemonстрированы все основные умения. Решены все основные задачи с отдельным и несущественными недочетами, выполнены все задания в полном объеме	Продemonстрированы все основные умения. Решены все основные задачи. Выполнены все задания, в полном объеме без недочетов
<u>Навыки</u>	Отсутствие базовых навыков. Невозможность оценить наличие навыков вследствие отказа обучающегося от	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки	Имеется минимальный набор навыков для решения стандартных задач с некоторым	Продemonстрированы базовые навыки при решении стандартных задач с некоторым и	Продemonстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продemonстрированы навыки при решении нестандартных задач без ошибок и	Продemonстрирован творческий подход к решению нестандартных задач

	ответа		и недочетами	недочетами		недочетов	
--	--------	--	-----------------	------------	--	-----------	--

### Шкала оценивания при промежуточной аттестации

Оценка		Уровень подготовки
зачтено	превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно», продемонстрированы знания, умения, владения по соответствующим компетенциям на уровне выше предусмотренного программой
	отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично».
	очень хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «очень хорошо»
	хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо».
	удовлетворительно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «удовлетворительно», при этом хотя бы одна компетенция сформирована на уровне «удовлетворительно»
не зачтено	неудовлетворительно	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно».
	плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

### 5.3 Типовые контрольные задания или иные материалы, необходимые для оценки результатов обучения на промежуточной аттестации с указанием критериев их оценивания:

#### 5.3.1 Типовые задания (оценочное средство - Контрольные вопросы) для оценки сформированности компетенции ОПК-6

1. Вероятностная постановка задачи обучения по прецедентам. Принцип минимизации эмпирического риска. Байесовская теория решений. Принцип максимума апостериорной вероятности. Байесов классификатор.
2. Экспериментальные методы оценки качества обучения.
3. Метод наименьших квадратов для решения задачи восстановления регрессии. Его вывод на основе метода максимального правдоподобия. Линейная регрессионная модель. Система нормальных уравнений. Основы регрессионного анализа (проверка значимости коэффициентов, коэффициент детерминации Пирсона, доверительные интервалы, анализ остатков).
4. Линейный дискриминантный анализ. Квадратичный дискриминантный анализ. Логистическая регрессия.
5. Размерность Вапника– Червоненкиса. Лемма Зауэра. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечной размерности Вапника–Червоненкиса. Принцип структурной минимизации риска.



### 5.3.2 Типовые задания (оценочное средство - Контрольные вопросы) для оценки сформированности компетенции ПКД-1

1. Метод ближайших соседей для решения задачи классификации. Теорема об оценке риска в методе ближайшего соседа.
2. Нейронные сети. Персептрон Розенблатта. Алгоритм обучения персептрона как метод стохастического градиентного спуска. Нейронные сети для решения задач классификации и восстановления регрессии. Обучение сети. Регуляризация как метод борьбы с переобучением. Понятие о глубоких нейронных сетях.
3. Ансамбли решающих правил (классификаторов). Простое и взвешенное голосования. Бустинг. Алгоритм AdaBoost. Оценка ошибки предсказания. Бустинг и аддитивные модели. Градиентный бустинг. Алгоритм градиентного бустинга деревьев решений (MART). Баггинг. Алгоритм случайных деревьев (случайный лес.)
4. Основы теории Вапника–Червоненкиса. Лемма Бернштейна. Теорема о равномерной сходимости эмпирического риска к ожидаемому риску в случае конечного класса решающих правил. Обоснование принципа минимизации эмпирического риска
5. Проблема переобучения при решении задачи восстановления регрессии. Методы борьбы с переобучением: сокращение числа параметров, регуляризация (ридж-регрессия), метод лассо. Трудоемкость методов.
6. Наивный байесовский классификатор.
7. Машина опорных векторов. Ядра и спрямляющие пространства.
8. Деревья решений. Метод CART (classification and regression trees) для решения задач классификации и восстановления регрессии. Отсечения ветвей и выбор финального дерева. Методы обработки пропущенных значений.
9. Обучение без учителя. Кластеризация. Кластеризация методами теории графов. Метод центров тяжести. Метод медиан. Метод нечетких множеств. ЕМ-алгоритм.
10. Иерархическая кластеризация. Агломеративные и разделяющие методы.

### Критерии оценивания (оценочное средство - Контрольные вопросы)

Оценка	Критерии оценивания
превосходно	свободное владение основным и дополнительным материалом без ошибок и погрешностей
отлично	свободное владение основным материалом с незначительными ошибками и погрешностями
очень хорошо	достаточное владение основным материалом с незначительными погрешностями
хорошо	владение основным материалом с рядом заметных погрешностей
удовлетворительно	владение минимальным материалом, необходимым по данному предмету, с рядом ошибок

Оценка	Критерии оценивания
неудовлетворительно	владение материалом недостаточно, необходима дополнительная подготовка. Работу за время семестра можно оценить как неудовлетворительную
плохо	отсутствие владения материалом. Работа за время семестра была оценена на «плохо»

## 6. Учебно-методическое и информационное обеспечение дисциплины (модуля)

Основная литература:

1. Платонов А. В. Машинное обучение : учебное пособие / А. В. Платонов. - Москва : Юрайт, 2022. - 85 с. - (Высшее образование). - URL: <https://urait.ru/bcode/508804> (дата обращения: 14.08.2022). - ISBN 978-5-534-15561-7 : 319.00. - Текст : электронный // ЭБС "Юрайт"., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=821948&idb=0>.

Дополнительная литература:

1. Флах П. Машинное обучение. Наука и искусство построения алгоритмов, которые извлекают знания из данных : монография / Флах П. - Москва : ДМК-пресс, 2023. - 401 с. - ISBN 978-5-89818-300-4., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=878863&idb=0>.
2. Python и машинное обучение: крайне необходимое пособие по новейшей предсказательной аналитике, обязательное для более глубокого понимания методологии машинного обучения / Рашка С. - Москва : ДМК-пресс, 2017., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=659331&idb=0>.

Программное обеспечение и Интернет-ресурсы (в соответствии с содержанием дисциплины):

1. The R Project for Statistical Computing <https://www.r-project>
2. Welcome to Python.org <https://www.python.org/>
3. scikit-learn: machine learning in Python [scikit-learn.org/](https://scikit-learn.org/)

## 7. Материально-техническое обеспечение дисциплины (модуля)

Учебные аудитории для проведения учебных занятий, предусмотренных образовательной программой, оснащены мультимедийным оборудованием (проектор, экран), техническими средствами обучения.

Помещения для самостоятельной работы обучающихся оснащены компьютерной техникой с возможностью подключения к сети "Интернет" и обеспечены доступом в электронную информационно-образовательную среду.

Программа составлена в соответствии с требованиями ОС ННГУ по направлению подготовки/специальности 42.04.02 - Журналистика.

Автор(ы): Золотых Николай Юрьевич, доктор физико-математических наук, доцент.

Заведующий кафедрой: Золотых Николай Юрьевич, доктор физико-математических наук.

Программа одобрена на заседании методической комиссии от 12.01.2024, протокол № 12.