

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ

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

Дзержинский филиал ННГУ

УТВЕРЖДЕНО

решением Ученого совета ННГУ

протокол № 10 от 02.12.2024 г.

Рабочая программа дисциплины

Технологии анализа данных и машинного обучения

Уровень высшего образования

Бакалавриат

Направление подготовки / специальность

09.03.03 - Прикладная информатика

Направленность образовательной программы

ИТ-сервисы и технологии обработки данных в экономике и финансах

Форма обучения

очно-заочная

г. Дзержинск

2025 год начала подготовки

1. Место дисциплины в структуре ОПОП

Дисциплина Б1.В.19 Технологии анализа данных и машинного обучения относится к части, формируемой участниками образовательных отношений 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
УК-1: Способен осуществлять поиск, критический анализ и синтез информации, применять системный подход для решения поставленных задач	<p>УК-1.1: Демонстрирует знание принципов сбора, отбора и обобщения информации, базирующихся на системном подходе</p> <p>УК-1.2: Демонстрирует умение соотносить разнородные явления и систематизировать их в рамках избранных видов профессиональной деятельности.</p> <p>УК-1.3: Демонстрирует наличие практического опыта работы с информационными источниками, опыта научного поиска и представления научных результатов.</p>	<p>УК-1.1: Знать: принципы, методы и средства решения стандартных задач в области экономики</p> <p>Уметь: осуществлять математическую и информационную постановку задач по обработке информации.</p> <p>Владеть: информационными технологиями поиска данных и способами их использования.</p> <p>УК-1.2: Знать: математические, естественно-научные и социально-экономические методы для использования в профессиональной деятельности.</p> <p>Уметь: решать нестандартные профессиональные задачи, в том числе в новой или незнакомой среде и в междисциплинарном контексте, с применением математических, естественнонаучных, социально-экономических и профессиональных знаний</p> <p>Владеть: навыками теоретического и экспериментального исследования объектов</p>	Тест Отчет по лабораторным работам	Экзамен: Контрольные вопросы

		<p>профессиональной деятельности, в том числе в новой или незнакомой среде и в междисциплинарном контексте.</p> <p>УК-1.3: Знать: новые научные принципы и методы исследований. Уметь: применять на практике новые научные принципы и методы исследований. Владеть: навыками применения новых научных принципов и методов исследования для решения профессиональных задач.</p>		
<p>ПК-9: Способен моделировать прикладные (бизнес) процессы и объекты предметной области</p>	<p>ПК-9.1: Демонстрирует знание методических основ моделирования процессов и объектов предметной области</p> <p>ПК-9.2: Демонстрирует умение применения знаний к моделированию прикладных процессов и объектов предметной области при разработке программного обеспечения ИС</p> <p>ПК-9.3: Имеет практический опыт моделирования процессов и объектов на примере конкретной предметной области</p>	<p>ПК-9.1: Знать основные понятия, методы, модели анализа и обработки данных и наиболее перспективные прикладные сферы их применения. Уметь: применять методы и способы построения эффективных алгоритмов и моделей для решения прикладных задач анализа и обработки данных. Владеть: практическими навыками разработки и сопровождении информационных хранилищ, технологий оперативного и интеллектуального анализа данных, отражающих деятельность предприятий в различных предметных областях.</p> <p>ПК-9.2: Знать: формализацию задачи машинного обучения. Уметь: выполнять постановку задачи машинного обучения. Владеть: приемами программирования на языках анализа данных.</p>	<p>Тест</p> <p>Отчет по лабораторным работам</p>	<p>Экзамен:</p> <p>Контрольные вопросы</p>

		ПК-9.3: <i>Знать: основные принципы построения логических алгоритмов классификации.</i> <i>Уметь: выбирать методы и средства для решения задач машинного обучения.</i> <i>Владеть: методами интеллектуального анализа информации.</i>		
--	--	--	--	--

3. Структура и содержание дисциплины

3.1 Трудоемкость дисциплины

	очно-заочная
Общая трудоемкость, з.е.	5
Часов по учебному плану	180
в том числе	
аудиторные занятия (контактная работа):	
- занятия лекционного типа	14
- занятия семинарского типа (практические занятия / лабораторные работы)	28
- КСР	2
самостоятельная работа	100
Промежуточная аттестация	36
	Экзамен

3.2. Содержание дисциплины

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

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 обеспечения самостоятельной работы обучающихся используются:

Электронные курсы, созданные в системе электронного обучения ННГУ:

Прикладные методы статистического обучения, <https://e-learning.unn.ru/course/view.php?id=11685>.

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 - Тест) для оценки сформированности компетенции УК-1:

1. С чем, из ниже перечисленного сравнивают линейный классификатор?

А) с аксоном;

Б) с правилом Хебба;

В) с генетическим алгоритмом;

Г) с нейроном.

2. Что называют задачей восстановления регрессии?

А) Задача обучения по прецедентам при $Y=R$.

Б) Задача поиска вектора параметров α^*

В) Задача поиска минимума среднего квадрата ошибки.

Г) Задача модели регрессии.

3. В чём заключается задача кластеризации?

А) разбить обучающую выборку на непересекающиеся подмножества, так, чтобы каждое множество состояло из объектов, близких по метрике ρ , а объекты разных множеств существенно отличались.

Б) распределить все обучающее множество на несколько классов.

В) упростить обработку данных, т.е. воссоединить множества в один класс.

4. Что, из ниже перечисленного, относится к обучающей выборке?

А) классификация данных

Б) объекты с известными ответами

В) алгоритм решающий функцию

5. Какие алгоритмы лучше работают на больших обучающих выборках?

А) Бустинг;

Б) Беггинг;

В) RSM.

5.1.2 Типовые задания (оценочное средство - Тест) для оценки сформированности компетенции ПК-9:

1. Что называется переобучением?

А) описание искомого алгоритма как суперпозиции некоторых элементарных функций;

Б) поиск преобразования исходящего пространства признаков в новое пространство существенно меньшей размерности;

В) когда по мере увеличения числа используемых признаков средняя ошибка на обучающей выборке монотонно убывает;

Г) когда средняя оценка на независимых контрольных данных сначала уменьшается, затем проходит через точку минимума и далее только возрастает.

2. Как называется алгоритм, который добавляет к набору G по одному признаку, каждый раз выбирая тот признак, который приводит к наибольшему уменьшению внешнего критерия?

А) Add;

Б) Bootstrap;

В) FullSearch;

Г) Поиск в глубину;

Д) Поиск в ширину.

3. Что называют обучением с подкреплением?

А) обучение основанное на собственном опыте;

Б) обучение с n-количеством учителей;

В) обучение к контролирующим параметрами.

4. Выберите неверные утверждения:

А) Наивный байесовский классификатор может быть только параметрическим.

Б) Наивный байесовский классификатор будет оптимальным, если признаки действительно независимы.

В) При классификации объекта заодно оцениваются априорные вероятности его принадлежности каждому из классов.

Г) Наивный байесовский классификатор может быть только непараметрическим.

5. Что, из ниже перечисленного, служит целями кластеризации?

- А) упрощение дальнейшей обработки данных;
- Б) сокращение объема хранимых данных;
- В) построение иерархии множества объектов;
- Г) распределение обучающего множества на несколько классов;
- Д) упрощение входных параметров.

Критерии оценивания (оценочное средство - Тест)

Оценка	Критерии оценивания
превосходно	Процент выполнения тестовых заданий 100 %
отлично	Процент выполнения тестовых заданий 95 – 99 %
очень хорошо	Процент выполнения тестовых заданий 94 %
хорошо	Процент выполнения тестовых заданий 75 – 93 %
удовлетворительно	Процент выполнения тестовых заданий 50 – 74 %
неудовлетворительно	Процент выполнения тестовых заданий 20 - 50 %
плохо	Процент выполнения тестовых заданий < 20 %

5.1.3 Типовые задания (оценочное средство - Отчет по лабораторным работам) для оценки сформированности компетенции УК-1:

Задание 1. Метрики качества задач классификации

1. Скачайте данные.
2. Обучите 4 классификатора, чтобы предсказать поле (биологический ответ молекулы) из набора данных "bioresponse.csv":
 - мелкое дерево решений;
 - глубокое дерево решений;
 - случайный лес на мелких деревьях;
 - случайный лес на глубоких деревьях;
3. Рассчитайте следующие метрики, чтобы проверить качество ваших моделей:

- доля правильных ответов (accuracy);
- точность;
- полнота;
- log-loss.

4. Постройте precision-recall и ROC-кривые для ваших моделей.

5. Обучите классификатор, который избегает ошибок второго рода и рассчитайте для него метрики качества.

Задание 2.Предобработка данных. Отбор признаков.

1. Скачать данные для исследования.
2. Реализовать функции one-hot-encoding и softmax средствами базового Python 3.6. Разделить переменные на численные и категориальные, масштабировать и нормировать данные.
3. Реализовать модель логистической регрессии средствами базового Python для решения задачи бинарной классификации для полного набора признаков из предобработанных данных и для непредобработанных данных, а также только для количественных признаков.
- 4) Сделать вывод об изменении качества работы модели в зависимости от применения предобработки данных и объёма признаков, на которых обучалась модель.

5.1.4 Типовые задания (оценочное средство - Отчет по лабораторным работам) для оценки сформированности компетенции ПК-9:

Задание 1. Функции ошибок в машинном обучении

1. Скачать данные.
2. Реализовать модель логистической регрессии со следующими функциями потерь:
 - a) Logistic loss
 - б) Adaboost loss
 - в) binary crossentropy
3. Визуализировать кривые обучения модели бинарной классификации в виде динамики изменения каждой из функций ошибок п.2 на тренировочной и тестовой выборках.
4. Сравнить качество классификации по метрике accuracy в каждом из трёх модификаций алгоритма.

Задание 2. Алгоритмы кластеризации

1. Скачать данные (MNIST): <http://yann.lecun.com/exdb/mnist/>
2. Реализовать в виде набора функций алгоритм k-means и алгоритм иерархической кластеризации (функции linkage модуля scipy.cluster) для разделения набора данных без меток на кластеры.
3. Построить кривую зависимости интеркластерного расстояния от числа кластеров для алгоритма иерархической кластеризации, выбрать оптимальный порог разделения.
4. Сравнить результаты двух выбранных алгоритмов по выбранной метрике оценки качества кластеризации.

Критерии оценивания (оценочное средство - Отчет по лабораторным работам)

Оценка	Критерии оценивания
превосходно	если выполнены 100% условий и требований, сформулированных в задании.
отлично	если выполнены 95% – 99% условий и требований, сформулированных в задании.
очень хорошо	если выполнены 94% условий и требований, сформулированных в задании.
хорошо	если выполнены 70% – 93% условий и требований, сформулированных в задании.
удовлетворительно	если выполнены 40% – 70% условий и требований, сформулированных в задании.
неудовлетворительно	если выполнены 20% - 40% условий и требований, сформулированных в задании.
плохо	если выполнены менее 20% условий и требований, сформулированных в задании.

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

Шкала оценивания сформированности компетенций

Уровень сформированности компетенций (индикатор достижения компет	плохо	неудовлетворительно	удовлетворительно	хорошо	очень хорошо	отлично	превосходно
	не зачтено		зачтено				

компетенций)							
<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 отдельными 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 ответа	При решении стандартных задач не продемонстрированы базовые навыки. Имели место грубые ошибки	Имеется минимальный набор навыков для решения стандартных задач с некоторым и недочетами	Продemonстрированы базовые навыки при решении стандартных задач с некоторым и недочетами	Продemonстрированы базовые навыки при решении стандартных задач без ошибок и недочетов	Продemonстрированы навыки при решении нестандартных задач без ошибок и недочетов	Продemonстрирован творческий подход к решению нестандартных задач

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 УК-1

1. Какие величины используют в качестве оценок вектора средних и матрицы ковариаций многомерной нормальной случайной величины?
2. Какая оценка является несмещенной оценкой матрицы ковариаций многомерной случайной величины?
3. Какая оценка матрицы ковариаций является оценкой метода максимального правдоподобия в случае нормального распределения генеральной совокупности?
4. Как строится доверительная область для вектора средних многомерной нормальной случайной величины, если матрица ковариаций многомерной величины известна (неизвестна)?
5. Как строится критерий для проверки гипотез: о равенстве векторов средних; о равенстве матриц ковариаций; о сферичности распределения; о однородности выборок; о независимости выборок (в случае нормального распределения).
6. В чем суть метода главных компонент, какие условия накладываются на главные компоненты?
7. Опишите принцип построения главных компонент.
8. Какие существуют критерии отбора главных компонент?
9. В чем суть факторного анализа? Какие условия накладываются на обобщенные и характерные факторы?
10. Опишите метод главных компонент, как метод факторного анализа.
11. Что такое каноническая модель факторного анализа? В чем ее принципиальное отличие от модели, используемой в методе главных компонент?
12. Как строится итерационная процедура поиска матрицы факторных нагрузок в методе максимального правдоподобия?

5.3.2 Типовые задания (оценочное средство - Контрольные вопросы) для оценки сформированности компетенции ПК-9

1. Как оценивается значимость модели факторного анализа?
2. Для чего в факторном анализе используют процедуру вращения факторов?
3. Как оцениваются значения факторов по методам Томпсона и Бартлетта?
4. В чем суть метода канонических корреляций? Каким условиям должны удовлетворять канонические переменные?
5. Опишите процедуру построения канонических переменных.
6. Как оценивается значимость канонических корреляций?
7. В чем заключается задача классификации объектов без обучения?
8. Как определяется в кластерном анализе мера близости объектов?
9. Как определяется расстояние между кластерами?
10. Какие методы кластерного анализа относятся к иерархическим агломеративным методам?
11. В чем суть итеративных методов? Опишите метод к-средних и метод поиска сгущений.
12. Что такое функционалы качества разбиения? Приведите примеры функционалов разбиения при известном числе кластеров и неизвестном числе кластеров.
13. Какие статистические критерии используются для проверки значимости различия кластеров.
14. В чем суть дискриминантного анализа? Что такое дискриминантная функция?
15. Опишите процедуру линейного дискриминантного анализа.
16. Как используется отношение правдоподобия в качестве решающего правила в дискриминантном анализе? Опишите данный метод дискриминантного анализа для нормально распределенных совокупностей.
17. Как статистически оценить значимость результатов дискриминантного анализа?

Критерии оценивания (оценочное средство - Контрольные вопросы)

Оценка	Критерии оценивания
превосходно	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «превосходно»
отлично	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «отлично», при этом хотя бы одна компетенция сформирована на уровне «отлично»
очень хорошо	Все компетенции (части компетенций), на формирование которых направлена

Оценка	Критерии оценивания
	дисциплина, сформированы на уровне не ниже «очень хорошо», при этом хотя бы одна компетенция сформирована на уровне «очень хорошо»
хорошо	Все компетенции (части компетенций), на формирование которых направлена дисциплина, сформированы на уровне не ниже «хорошо», при этом хотя бы одна компетенция сформирована на уровне «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»
неудовлетворительно	Хотя бы одна компетенция сформирована на уровне «неудовлетворительно», ни одна из компетенций не сформирована на уровне «плохо»
плохо	Хотя бы одна компетенция сформирована на уровне «плохо»

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

Основная литература:

- Макшанов А. В. Современные технологии интеллектуального анализа данных : учебное пособие для спо / Макшанов А. В., Журавлев А. Е., Тындыкарь Л. Н. - Санкт-Петербург : Лань, 2020. - 228 с. - Книга из коллекции Лань - Информатика. - ISBN 978-5-8114-5451-8., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=880007&idb=0>.
- Митина О. А. Языки программирования для статистической обработки данных: Практикум / Митина О. А. - Москва : РТУ МИРЭА, 2021. - 139 с. - Библиогр.: доступна в карточке книги, на сайте ЭБС Лань. - Книга из коллекции РТУ МИРЭА - Информатика., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=754924&idb=0>.

Дополнительная литература:

- Нестеров С. А. Основы интеллектуального анализа данных. Лабораторный практикум : учебное пособие / Нестеров С. А. - Санкт-Петербург : Лань, 2020. - 40 с. - Библиогр.: доступна в карточке книги, на сайте ЭБС Лань. - Книга из коллекции Лань - Информатика. - ISBN 978-5-8114-4509-7., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=707838&idb=0>.
- Митина О. А. Технологии организации, обработки и хранения статистических данных : учебное пособие / Митина О. А., Юрченков И. А. - Москва : РТУ МИРЭА, 2019. - 163 с. - Библиогр.: доступна в карточке книги, на сайте ЭБС Лань. - Книга из коллекции РТУ МИРЭА - Информатика., <https://e-lib.unn.ru/MegaPro/UserEntry?Action=FindDocs&ids=754884&idb=0>.
- Железнов М. М. Методы и технологии обработки больших данных : учебно-методическое пособие / Железнов М. М. - Москва : МИСИ – МГСУ, 2020. - 46 с. - Библиогр.: доступна в карточке книги, на сайте ЭБС Лань. - Книга из коллекции МИСИ – МГСУ - Информатика. - ISBN

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

1. Электронная библиотека учебников [Электронный ресурс]. - Режим доступа: <http://studentam.net> — Загл. с экрана.
2. Российская государственная библиотека [Электронный ресурс]. - Режим доступа: <http://www.rsl.ru> — Загл. с экрана.
3. Научная электронная библиотека [Электронный ресурс]. - Режим доступа: <http://elibrary.ru/> — Загл. с экрана.
4. Python. URL: <https://www.python.org/>
5. Программное обеспечение: менеджер пакетов Anaconda, Spyder (open-source IDE для Python).

7. Материально-техническое обеспечение дисциплины (модуля)

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

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 09.03.03 - Прикладная информатика.

Автор(ы): Кечкина Наталия Игоревна, кандидат технических наук.

Заведующий кафедрой: Поляков Евгений Артурович, кандидат педагогических наук.

Программа одобрена на заседании методической комиссии от 28.12.2024, протокол № 21.