

Рабочая программа дисциплины (модуля) разработана в соответствии с самостоятельно установленным МГУ образовательным стандартом (ОС МГУ) для реализуемых основных профессиональных образовательных программ высшего образования по направлению подготовки / специальности 27.04.03 «Системный анализ и управление» (программа магистратуры), утвержденным приказом МГУ от 29 мая 2023 года №697.

Год (годы) приема на обучение: 2024.

I. Цели и задачи учебной дисциплины

Целью изучения дисциплины «Машинное обучение» является формирование профессиональных компетенций в области применения технологий машинного обучения для больших данных и включает в себя разработку, оценку и адаптацию моделей больших данных в предметной среде основы анализа больших данных, например, метод ближайших соседей, байесовские классификаторы, метод опорных векторов, решающие деревья, случайный лес, градиентный бустинг, искусственные нейронные сети, а также комплекс прикладных программ для реализации данных технологий.

Задачами дисциплины являются:

- изучить современные языки программирования;
- приобрести умение применять методы машинного обучения для решения практических задач по обработке больших данных;
- приобрести умение анализировать задачи машинного обучения и осуществлять взвешенный выбор того или иного решения;
- ознакомить студентов с компьютерными технологиями обработки многомерных неструктурированных массивов разнородной статистической информации;
- научить интерпретировать полученные результаты построенных моделей.

В результате изучения данного курса обучающиеся получают знания об эволюции искусственного интеллекта, о методологии и принципах его применения для анализа данных и управления предприятием, о методах анализа и реинжиниринга бизнес-процессов, приобретут навыки и умения применения методов решающих деревьев, случайного леса, искусственных нейронных систем.

II. Место дисциплины в структуре ОПОП ВО

Дисциплина «Машинное обучение» является дисциплиной по выбору профессионального блока вариативной части программы магистратуры.

Изучение дисциплины базируется на знаниях и умениях, полученных обучающимися в процессе изучения гуманитарных, социальных и экономических дисциплин: «Интеллектуальные методы анализа бизнес-информации», «Python: введение в анализ данных», «Информационная безопасность и защита информации»

Для успешного освоения дисциплины обучающийся должен:

Знать:

- фундаментальные положения экономики, основ программирования, бизнес-анализа;
- теоретические основы информационных технологий и бизнес-аналитики;
- основные проблемы современной философии и подходов к их решению;

Уметь:

- использовать междисциплинарные системные связи наук;
- анализировать и оценивать философские проблемы при решении социальных и профессиональных задач;
- применять математический инструментарий к решению социальных и профессиональных проблем.

Владеть:

- навыками экономического и финансового анализа;
- навыками выбора наиболее актуальных направлений научных исследований, ставить задачи исследования и определять способы решения поставленных задач;
- самостоятельно приобретать и использовать в практической деятельности новые знания и умения в различных сферах деятельности.

Знания, навыки и умения, полученные при изучении дисциплины «Машинное обучение» обеспечивают успешное прохождение таких дисциплин, как, «Проектирование пользовательского опыта (UX/UI Design) «IT-системы в e-commerce», «Технологии цифрового производства» и необходимы для прохождения преддипломной практики, осуществления научно-исследовательской работы и написания выпускной квалификационной работы (магистерской диссертации). Изучается на 2 курсе (4 семестр).

III. Требования к результатам освоения дисциплины

В результате освоения дисциплины должны быть сформированы следующие компетенции:

Код и наименование компетенции	Код и наименование индикатора	Планируемые результаты
Универсальные компетенции		
<p>УК-1. Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, выработать стратегию действий, формулировать научно обоснованные гипотезы, применять методологию научного познания в профессиональной деятельности.</p>	<p>УК-1.1. Анализирует проблемную ситуацию как систему, выявляя ее составляющие и связи между ними</p>	<p>Знать:</p> <ul style="list-style-type: none"> – основные методы критического анализа; – методологию системного подхода; <p>Уметь:</p> <ul style="list-style-type: none"> – выявлять проблемные ситуации, используя методы анализа, синтеза и абстрактного мышления; – осуществлять поиск решений проблемных ситуаций на основе действий, эксперимента и опыта; – производить анализ явлений, обрабатывать полученные результаты, делать обоснованные выводы; – определять в рамках выбранного алгоритма вопросы (задачи), подлежащие дальнейшей разработке и предлагать способы их решения; <p>Владеть:</p> <ul style="list-style-type: none"> – технологиями выхода из проблемных ситуаций, навыками выработки стратегии действий; – навыками критического анализа;

		<ul style="list-style-type: none"> – навыками применения системного подхода к анализу проблемных ситуаций. – навыками интерпретации полученных данных в ходе анализа проблемной ситуации и формирования обоснованных выводов.
	<p>УК-1.2. Разрабатывает и обосновывает стратегию действий по решению проблемной ситуации на основе системного и междисциплинарных подходов.</p>	<p>Знать основные положения разработки стратегии действий по решению проблемной ситуации;</p> <p>Уметь:</p> <ul style="list-style-type: none"> – разрабатывать и обосновывать стратегию действий по решению проблемной ситуации; – использовать системный и междисциплинарные подходы к решению проблемной ситуации; <p>Владеть навыками разработки стратегии действий по решению проблемной ситуации на основе системного и междисциплинарных подходов.</p>
Общепрофессиональные компетенции		
<p>ОПК-1. Способен анализировать и выявлять естественно-научную сущность проблем управления в технических системах на основе ранее приобретенных знаний</p>	<p>ОПК-1.1. Демонстрирует знание законов, естественно-научных и математических методов для использования в профессиональной деятельности в области управления в технических системах</p>	<p>Знать: фундаментальные законы природы и основные физические математические принципы;</p> <p>Уметь: применять физические законы и математические методы для решения задач теоретического и прикладного характера в области управления в технических системах;</p> <p>Владеть: навыками использования ранее приобретенных знаний математики, естественных и</p>

		технических наук при решении практических задач в области управления в технических системах.
ОПК-4. Способен осуществлять оценку эффективности технических систем методами системного анализа и управления	ОПК-4.1. Выбирает и разрабатывает критерии оценки эффективности технических систем	<p>Знать:</p> <ul style="list-style-type: none"> – математические методы для оценки эффективности технических систем; – критерии и показатели оценки эффективности технических систем; – особенности разработки критериальных систем оценки; <p>Уметь:</p> <ul style="list-style-type: none"> – выбирать актуальные с учетом рабочей ситуации критерии и показатели оценки эффективности технических систем; – разрабатывать критерии и показатели оценки эффективности технических систем; – формулировать вывод об эффективности технических систем; <p>Владеть: навыками разработки критериальной системы оценки эффективности технических систем на основе современных математических методов.</p>
ОПК-6. Способен применять методы математического, функционального и системного анализа для решения задач моделирования,	ОПК-6.1. Применяет методы математического, функционального и системного анализа	<p>Знать:</p> <ul style="list-style-type: none"> – методы математического, функционального и системного анализа;

<p>исследования и синтеза автоматического управления техническими объектами</p>		<ul style="list-style-type: none"> – средства и методы автоматического управления техническими объектами; <p>Уметь:</p> <ul style="list-style-type: none"> – анализировать процессы автоматического управления техническими объектами; – моделирования, исследования и синтеза автоматического управления техническими объектами; <p>Владеть:</p> <ul style="list-style-type: none"> – навыками математического, функционального и системного анализа применения средств автоматического управления техническими объектами в организации; – навыками разработки отчета о функционировании автоматического управления техническими объектами в организации.
	<p>ОПК 6.2. Решает задачи моделирования, исследования и синтеза автоматического управления техническими объектами</p>	<p>Знать:</p> <ul style="list-style-type: none"> – методы построения моделей исследуемых процессов, явлений и объектов; – методы построения идентификации исследуемых процессов, явлений и объектов; – методы измерения, анализа и улучшения параметров процессов автоматического управления техническими объектами; <p>Уметь:</p>

		<ul style="list-style-type: none"> – исследовать и моделировать процесс применения средств автоматизации деятельности организации; – разработке мероприятий по внедрению средств автоматизации управления организацией; <p>Владеть навыками принятия управленческих решений по выбору и внедрению средств автоматизации управления организацией.</p> <p>–</p>
--	--	--

Профессиональные компетенции

<p>ПК-1. Способен выявлять и оценивать тенденции технологического развития в области ИТ и автоматизации организации, осуществлять технологическое прогнозирование</p>	<p>ПК-1.1. Выявляет и оценивает тенденции технологического развития в области ИТ и автоматизации организации</p>	<p>Знать:</p> <ul style="list-style-type: none"> – методы построения концептуальных, математических и имитационных моделей; – передовой отечественный и зарубежный опыт в области развития науки и техники; – методы прогнозирования, технико-экономических исследований научно-технических решений и нормативного проектирования инновационных видов продукции и процессов; <p>Уметь:</p> <ul style="list-style-type: none"> – анализировать научную, научно-техническую информацию в области ИТ и автоматизации организации; – выявлять и оценивать тенденции технологического развития в наукоемких сферах на основе анализа, обобщения и систематизации передового опыта в сфере инноватики по материалам ведущих научных
--	---	--

		<p>журналов и изданий, с использованием электронных библиотек и интернет-ресурсов;</p> <ul style="list-style-type: none"> – оценивать возможные результаты внедрения передовых технологических решений в области ИТ и автоматизации организации; <p>Владеть навыками подготовки предложений по повышению эффективности деятельности организации посредством внедрения ИТ и осуществления автоматизации организации.</p>
	<p>ПК-1.2. Осуществляет технологическое прогнозирование в области ИТ и автоматизации организации</p>	<p>Знать основные положения и методы технологического прогнозирования;</p> <p>Уметь:</p> <ul style="list-style-type: none"> – использовать источники информации для анализа данных, необходимых для составления прогноза в области ИТ и автоматизации организации; – применять методы анализа данных и построения математических моделей; – применять программные средства планирования, мониторинга, контроля исполнения, формирования прогнозных данных; – выполнять технико-экономический анализ проектных, конструкторских и технологических решений для выбора оптимального варианта реализации инноваций; – прогнозировать тенденции развития науки и техники в области ИТ и автоматизации организации;

		<p>Владеть навыками формирования прогноза технологического развития в области ИТ и автоматизации организации.</p>
<p>ПК-2. Способен представить (опубликовать) результат научного исследования по направлению профессиональной подготовки на конференции или в печатном издании, в том числе на иностранном языке, в виде отчета, реферата, научной статьи, оформленной в соответствии с имеющимися требованиями</p>	<p>ПК-2.1. Представляет результат научного исследования по направлению профессиональной подготовки на конференции или в печатном издании</p>	<p>Знать:</p> <ul style="list-style-type: none"> – основы представления научного результата на иностранном языке; – особенности подготовки научной публикации и выступления; <p>Уметь:</p> <ul style="list-style-type: none"> – готовить презентацию научного исследования, в том числе на иностранном языке; – представлять и аргументированно излагать результат научного исследования; – вести дискуссию с аудиторией в рамках академического и профессионального взаимодействия; <p>Владеть навыками представления результата научного исследования по направлению профессиональной подготовки на конференции или в печатном издании в рамках академического и профессионального взаимодействия.</p>
	<p>ПК-2.2. Оформляет результат научного исследования по направлению профессиональной подготовки в соответствии с имеющимися</p>	<p>Знать:</p> <ul style="list-style-type: none"> – структуру и основные этапы изложения результата научного исследования; – требования и государственные стандарты оформления научного отчета, реферата, статьи;

	<p>требованиями в виде отчета, реферата, научной статьи</p>	<p>Уметь:</p> <ul style="list-style-type: none"> – грамотно и аргументированно излагать результат научного исследования по направлению профессиональной подготовки; – грамотно оформлять научный отчет, реферат, статью в соответствии с принятыми требованиями государственными стандартами; <p>Владеть навыками изложения и грамотного оформления научного отчета, реферата, статьи в соответствии с принятыми требованиями государственными стандартами.</p>
<p>ПК-5. Способен инициировать и планировать проект в области ИТ по разработке программного продукта и ИС, контролировать его выполнение, выявлять и оценивать риски</p>	<p>ПК-5.1. Иницирует и планирует проект в области ИТ по разработке программного продукта и ИС</p>	<p>Знать:</p> <ul style="list-style-type: none"> – теорию проектного менеджмента; – методы планирования проекта; – этапы жизненного цикла проекта; <p>Уметь:</p> <ul style="list-style-type: none"> – анализировать информацию по проекту; – определять задачи проекта и оценивать степень их достижения в соответствии с фазами его жизненного цикла; – разрабатывать ИСР и расписание проекта; – разрабатывать бюджет и план финансирования проекта;

		<p>Владеть навыками разработки плана управления проектом и частных планов;</p>
	<p>ПК-5.2. Контролирует выполнение проекта в области ИТ по разработке программного продукта и ИС</p>	<p>Знать:</p> <ul style="list-style-type: none"> – основы общего менеджмента; – теорию проектного менеджмента; <p>Уметь:</p> <ul style="list-style-type: none"> – проводить мониторинг реализации проекта в области ИТ, выявлять отклонения от плана; – применять методы контроля реализации проекта; <p>Владеть навыками разработки мероприятий по компенсации отклонения от плана проекта и оценке их эффективности.</p>
	<p>ПК-5.3. Выявляет и оценивает риски проекта в области ИТ по разработке программного продукта и ИС.</p>	<p>Знать:</p> <ul style="list-style-type: none"> – теорию проектного менеджмента; – методы и средства управления рисками; <p>Уметь:</p> <ul style="list-style-type: none"> – выявлять новые риски и отслеживать существующие риски; – анализировать и оценивать риски проекта; – разрабатывать перечень рисков проекта; – выбирать способы реагирования на риски и разрабатывать мероприятия по управлению рисками;

		<ul style="list-style-type: none"> – определять стратегий и приоритетов управления рисками; <p>Владеть навыками разработки плана управления рисками.</p>
<p>ПК-9. Способен осуществлять разработку проектов совершенствования производства на основе средств автоматизации и обеспечивать функционирование автоматизированной системы управления производством.</p>	<p>ПК-9.1. Разрабатывает проекты совершенствования производства на основе средств автоматизации</p>	<p>Знать:</p> <ul style="list-style-type: none"> – национальную и международную нормативную базу в области АСУП; – особенности проектирования АСУП; – основы экономики, организации производства, труда и управления; <p>Уметь:</p> <ul style="list-style-type: none"> – прогнозировать технико-экономические показатели развития производства; – оценивать необходимость и потребности организации во внедрении средств автоматизации производством; – проектировать автоматизированные средств системы управления производства в организации; <p>Владеть навыками разработки проектов совершенствования производством на основе средств автоматизации производства.</p>
	<p>ПК-9.2. Обеспечивает функционирование автоматизированной системы управления производством в организации.</p>	<p>Знать:</p> <ul style="list-style-type: none"> – основы функционирования АСУП; – методы формирования показателей эффективности конкурентоспособности АСУП;

		<p>Уметь:</p> <ul style="list-style-type: none"> – руководить функционированием и совершенствованием действующей в организации АСУП; – координировать деятельность подразделений организации в области АСУП – контролировать функционирование АСУП в организации; <p>Владеть навыками принятия управленческих решений по контролю и координации функционирования АСУП в организации.</p>
<p>ПК-11. Способен разрабатывать методические материалы и осуществлять обучение по направлению профессиональной подготовки</p>	<p>ПК-11.1. Разрабатывает обучение методические материалы по направлению профессиональной подготовки</p>	<p>Знать:</p> <ul style="list-style-type: none"> – положения образовательной программы по направлению профессиональной подготовки; – методы разработки методических материалов. – основные методики проведения курсов обучения; <p>Уметь:</p> <ul style="list-style-type: none"> – разрабатывать презентации лекционных занятий по направлению профессиональной подготовки; – контрольные задания для проверки полученных знаний; <p>Владеть навыками разработки методических материалов по</p>

		направлению профессиональной подготовки.
ПК-12. Способен выявлять проблемы организации, связанные с информационным обеспечением и особенностями установленной базы данных, анализировать и осваивать новые информационные технологии в области баз данных.	ПК-12.1. Выявляет проблемы организации, связанные с информационным обеспечением и особенностями установленной базы данных	<p>Знать:</p> <ul style="list-style-type: none"> – принципы функционирования баз данных; – основные тенденции развития информационных технологий в области баз данных; <p>Уметь:</p> <ul style="list-style-type: none"> – исследовать рынок перспективных баз данных, их принципиальных возможностей; – осуществлять сбор и анализ нереализованных потребностей пользователей баз данных; – анализировать проблемы организации в области информационного обеспечения; – оценивать эффективность функционирования баз данных; <p>Владеть:</p> <ul style="list-style-type: none"> – навыками выявления проблем организации, связанные с информационным обеспечением и особенностями установленной базы данных; – навыками прогнозирования состояние и осуществлять планирование по развитию баз данных в организации.
ПК-13. Способен принимать управленческие решения и разрабатывать план по развитию и обновлению базы данных организации.	ПК-13.1. Выявляет альтернативы и принимает управленческие решения по разработке плана развития	Знать основные тенденции развития информационных технологий в области баз данных;

	и обновления базы данных организации	<p>Уметь выявлять альтернативы на основе заданных критериев и выбирать наилучшую;</p> <p>Владеть навыками принятия управленческих решений по разработке плана развития и обновления базы данных организации.</p>
--	--------------------------------------	--

Специализированные компетенции

<p>СПК -2. Способен применять методы анализа данных для разработки продуктовой и технологической стратегии развития организации.</p>	<p>СПК-2.1. Применяет методы анализа данных для разработки продуктовой и технологической стратегии развития организации</p>	<p>Знать:</p> <ul style="list-style-type: none"> – фундаментальные разделы математики; – методы системного анализа; – теорию продуктового менеджмента; – теорию стратегического менеджмента; <p>Уметь:</p> <ul style="list-style-type: none"> – использовать законы естественно-научных дисциплин в профессиональной деятельности и применять математический аппарат, методы оптимизации, теории вероятностей, математической статистики, системного анализа для принятия решений в области решения профессиональных задач; – применять методы анализа данных; – применять инструментальные и программные средства для анализа данных; – анализировать и оценивать инновационный потенциал организации;
---	--	---

		<ul style="list-style-type: none"> – проводить маркетинговое исследование рынка и продуктов конкурентов; – применять методы разработки стратегии развития организации; <p>Владеть навыками анализа данных для решения профессиональных задач с применением современных технологий по разработке продуктовой и технологической стратегии развития организации.</p>
<p>СПК-4. Способен анализировать и оценивать влияние изменений в информационной системе на основные параметры организации и/или проекта, разрабатывать рекомендации по их учёту.</p>	<p>СПК-4.1. Анализирует и оценивает влияние изменений в информационной системе на основные параметры организации и/или проекта</p>	<p>Знать:</p> <ul style="list-style-type: none"> – методы управления организационными и технологическими изменениями; – методы анализа данных; – методы системного анализа; <p>Уметь:</p> <ul style="list-style-type: none"> – оценивать готовность организации к осуществлению изменения архитектуры информационной системы; – анализировать и оценивать влияние изменений в информационной системе на основные параметры организации и/или проекта; <p>Владеет навыками формирования отчета по влиянию изменений в информационной системе на основные параметры организации и/или проекта.</p>
	<p>СПК-4.2. Разрабатывает рекомендации по учёту изменений в</p>	<p>Знать:</p> <ul style="list-style-type: none"> – архитектуру информационной системы организации;

	информационной системе организации	<ul style="list-style-type: none"> – особенности развития экономической деятельности организации; <p>Уметь:</p> <ul style="list-style-type: none"> – учитывать изменения в экономической деятельности организации; – формировать отчеты о необходимости изменения архитектуры информационной системы организации и её элементов; <p>Владеть навыками разработки рекомендации по учёту изменений в информационной системе организации.</p>
--	------------------------------------	--

Форма обучения: очная.

IV. Формы контроля

Контроль за освоением дисциплины осуществляется в каждом дисциплинарном разделе отдельно.

Рубежный контроль: тестирование и контрольная работа по отдельным разделам дисциплины.

Итоговая аттестация в 4 семестре – зачет в устной форме собеседования.

Результаты текущего контроля и итоговой аттестации формируют рейтинговую оценку работы обучающегося. Распределение баллов по отдельным видам работ в процессе освоения дисциплины «Управление бизнес-процессами» осуществляется в соответствии с Приложением 1.

V. Объём дисциплины и виды учебной работы

Объём курса – 72 часа, 2 зачетные единицы, в том числе 30 часа – аудиторная нагрузка, из которых 8 часов – лекции, 22 часов – семинары, 42 часов – самостоятельная работа студентов. Изучается на 2 курсе (4 семестр), итоговая форма отчетности – зачет.

Вид учебной работы	Всего часов
Контактные занятия (всего)	30
В том числе:	-
Лекции	8
Практические занятия (ПЗ)	-
Семинары (С)	22
Лабораторные работы (ЛР)	-

Самостоятельная работа (всего)	42
В том числе:	-
Домашние задания	10
Реферат	10
Подготовка к тестированию	8
Подготовка к опросу	5
Подготовка к контрольной работе	5
Вид промежуточной аттестации Зачет	4
Общая трудоемкость (часы)	72
Зачетные единицы	2

VI. Структура и содержание дисциплины

п/п	Раздел	Содержание (темы)
1	Введение в анализ данных и машинного обучения	Введение в элементы text-mining и data-mining. Частотные гистограммы. Инструментальные переменные. Методы снижения размерности. Выбор функции ядра. Выбор ширины окна, переменная ширина окна. Метод парзеновского окна. Робастное оценивание плотности. Цензурирование выборки (отсев объектов-выбросов).
2	Линейные методы классификации для больших данных	Оптимальная разделяющая гиперплоскость. Понятие зазора между классами (margin). Случаи линейной разделимости и отсутствия линейной разделимости. Связь с минимизацией регуляризованного эмпирического риска. Кусочно-линейная функция потерь. Задача квадратичного программирования и двойственная задача. Понятие опорных векторов.
3	Технологии нейронных сетей для обработки больших данных	Структура многослойной нейронной сети. Функции активации. Проблема полноты. Полнота двухслойных сетей в пространстве булевских функций. Алгоритм обратного распространения ошибок. Формирование начального приближения. Проблема паралича сети. Методы оптимизации структуры сети. Выбор числа слоёв и числа нейронов в скрытом слое. Постепенное усложнение сети. Оптимальное прореживание сети (optimal brain damage).
4	Критерий выбора моделей и методы отбора признаков при анализе больших массивов многомерных данных	Внутренние и внешние критерии. Скользящий контроль, разновидности скользящего контроля. Критерий непротиворечивости. Регуляризация. Теория Вапника-Червоненкиса. Критерии, основанные на оценках обобщающей способности: Вапника-Червоненкиса, критерий Акаике (AIC), байесовский информационный

		критерий (BIC). Агрегированные и многоступенчатые критерии. Сложность задачи отбора признаков. Полный перебор. Метод добавления и удаления, шаговая регрессия. Поиск в глубину, метод ветвей и границ.
5	Кластеризация больших данных	Решающий список. Жадный алгоритм синтеза списка. Решающее дерево. Недостатки алгоритма и способы их устранения. Проблема переобучения. Редукция решающих деревьев: предредукция и постредукция. Преобразование решающего дерева в решающий список. Решающий лес и бустинг над решающими деревьями. Переключающиеся решающие деревья (alternating decision tree). Принцип голосования. Проблема различности (диверсификации) закономерностей. Методы синтеза конъюнктивных закономерностей. Алгоритм градиентного бустинга.

Разделы дисциплин и виды занятий (ак. часы)

п/п	Наименование раздела дисциплины	Лекция	Практические занятия	Лабораторные занятия	Семинар	СРС	Форма текущего контроля
1	Методы кодирования и обработки неструктурированной информации	1	-	-		6	Опрос
2	Линейные методы классификации для больших данных	1	-	-	4	8	Тест Домашнее задание
3	Технологии нейронных сетей для обработки больших данных	2	-	-	4	8	Опрос Домашнее задание
4	Критерий выбора моделей и методы отбора признаков при анализе больших массивов многомерных данных	2	-	-	6	8	Тест
5	Логические методы классификации. Методы кластеризации	2	-	-	8	8	КР

	при работе с большими данными						
	Промежуточная аттестация (зачет)					4	
	Итого	8	-	-	22	42	

Разделы дисциплины и междисциплинарные связи

№ п/п	Наименование обеспечиваемых (последующих) дисциплин	№ № разделов данной дисциплины, необходимых для изучения обеспечиваемых (последующих) дисциплин				
		1	2	3	4	5
1.	Системы искусственного интеллекта		+	+	+	
2.	Разработка программного обеспечения	+				+
3.	Python: введение в анализ данных	+				

VII. Образовательные технологии

В процессе освоения дисциплины «Машинное обучение» используются следующие образовательные технологии:

1. Стандартные методы обучения:

- лекции;
- семинары;
- письменные или устные домашние задания;
- консультации преподавателей;
- самостоятельная работа студентов, в которую входит освоение теоретического материала, подготовка к семинарам, выполнение указанных выше письменных работ.

2. Методы обучения с применением интерактивных форм образовательных технологий:

- интерактивные лекции;
- анализ деловых ситуаций на основе кейс-метода и имитационных моделей;
- круглые столы;
- групповые дискуссии и проекты;
- обсуждение результатов работы студенческих исследовательских групп.

VIII. Учебно-методическое, информационное и материально-техническое обеспечение дисциплины

Учебно-методическое и информационное обеспечение дисциплины

а) Основная литература:

1. Цифровой бизнес : учебник / под науч. ред. О. В. Китовой. — Москва : ИНФРА-М, 2021. — 418 с. — (Высшее образование: Магистратура). - ISBN 978-5-16-013017-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1659834> (дата обращения: 03.12.2021). – Режим доступа: по подписке.
2. Шёнталер, Ф. Бизнес-процессы: языки моделирования, методы, инструменты : практическое руководство / Франк Шёнталер, Готфрид Фоссен, Андреас Обервайс, Томас Карле ; пер. с нем. - Москва : Альпина Паблишер, 2019. - 264 с. - ISBN 978-5-96142-482-9. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1078471> (дата обращения: 02.12.2021). – Режим доступа: по подписке.
3. Программные продукты и системы : международный научно-практический журнал. - Тверь : НИИ Центрпрограммсистем, 2018. - Т. 31, № 2. - 420 с. - ISSN 0236-235X. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1016281> (дата обращения: 02.12.2021). – Режим доступа: по подписке.
4. Артяков, В. В. Управление инновациями. Методологический инструментарий : учебник / В.В. Артяков, А.А. Чурсин. — Москва : ИНФРА-М, 2022. — 206 с. — (Высшее образование: Магистратура). — DOI 10.12737/textbooks_1013514.Chursin. - ISBN 978-5-16-014965-3. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1851664> (дата обращения: 02.12.2021). – Режим доступа: по подписке.
5. Цифровизация: практические рекомендации по переводу бизнеса на цифровые технологии. - Москва : Альпина Паблишер, 2019. - 252 с. - ISBN 978-5-9614-2849-0. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1222514> (дата обращения: 02.12.2021). – Режим доступа: по подписке.

б) Дополнительная литература:

6. Землянский, А. А. Управление информационными ресурсами в научно-исследовательской работе : учебное пособие / А. А. Землянский, И. Е. Быстренина. - 2-е изд. - Москва : Дашков и К, 2021. - 110 с. - ISBN 978-5-394-04149-5. - Текст : электронный. - URL: <https://znanium.com/catalog/product/1232484> (дата обращения: 02.12.2021). – Режим доступа: по подписке.

Перечень ресурсов информационно-телекоммуникационной сети «Интернет» и информационных справочных систем

Перечень профессиональных баз данных и информационных справочных систем

1. ЭБС «Юрайт» [раздел «ВАША ПОДПИСКА: учебники и учебные пособия издательства «Юрайт»]: сайт. – URL: <https://www.biblio-online.ru/catalog/>
2. ЭБС издательства «Лань» [учебные, научные издания, первоисточники, художественные произведения различных издательств; журналы] : сайт. – URL: <http://e.lanbook.com>

3. <https://www.econ.msu.ru/elibrary> – электронная библиотека Экономического факультета МГУ
4. <https://www.nbmgu.ru> – Научная библиотека МГУ

Перечень ресурсов информационно-телекоммуникационной сети «Интернет»

1. Машинное обучение и анализ данных | Coursera – <https://ru.coursera.org/specializations/machine-learning-data-analysis>
2. Python 3 для начинающих – <https://pythonworld.ru/>
3. Официальный сайт разработчиков интерпретатора языка Python - <https://www.python.org/>
4. Визуальный исполнитель кода для языков программирования Python, Java, C, C++, JavaScript, and Ruby – <http://pythontutor.com/>
5. www.budget.ru – Финансовое казначейство РФ
6. www.economy.gov.ru – Министерство экономического развития и торговли РФ.
7. www.gks.ru – Федеральная служба государственной статистики РФ.
8. www.mcsx.ru – Министерство сельского хозяйства РФ
9. www.minfin.ru – Министерство финансов РФ
10. www.mon.gov.ru – Министерство образования РФ
11. www.who.int/en/ - Всемирная Организация Здравоохранения
12. www.un.org/ - Организация Объединенных Наций
13. www.worldbank.org – Всемирный Банк Реконструкции и Развития
14. www.cbr.ru Центральный Банк Российской Федерации Макроэкономическая статистика
15. stat.hse.ru Базы данных экономической статистики РФ
16. www.levada.ru – Аналитический Центр Юрия Левады (Левада-Центр) – российская негосударственная исследовательская организация.
17. <http://www.rbc.ru/> - РосБизнесКонсалтинг
18. <http://quote.rbc.ru/shares/> - Курсы акций, облигаций, валют и т.п.
19. <http://www.forex.ru/index.html> - Форекс.
20. <http://www.rts.ru/> - Биржа РТС.
21. www.micex.ru/ - Биржа ММВБ.

Рекомендуемые обучающие, справочно-информационные, контролирующие и прочие компьютерные программы, используемые при изучении дисциплины

№ п/п	Название рекомендуемых по разделам и темам программы технических и компьютерных средств обучения	Номера тем
1.	Python 3	1-5
2.	R	1-5

Методические указания для обучающихся по освоению дисциплины

В процессе изучения курса обучающиеся обязаны соблюдать дисциплину, вовремя приходить на занятия, делать домашние задания, осуществлять подготовку к семинарам и контрольным работам, проявлять активность на занятиях.

При этом важное значение имеет самостоятельная работа, которая направлена на формирование у учащегося умений и навыков правильного оформления конспекта и работы с ним, работы с литературой и электронными источниками информации, её анализа, синтеза и обобщения. Для проведения самостоятельной работы обучающимся предоставляется список учебно-методической литературы.

Материально-техническое обеспечение дисциплины

– Для проведения образовательного процесса необходима аудитория, оборудованная компьютером и проектором, необходимыми для демонстрации презентаций. Обязательное программное обеспечение Microsoft Office, Python 3, R;

Темы курсовых работ

Курсовая работа по дисциплине «Машинное обучение» не предусмотрена.

Вопросы для текущего контроля студентов

1. Что такое решающее дерево?
2. Какие критерии информативности используются при синтезе решающего дерева и почему?
3. Достоинства и недостатки решающих деревьев.
4. Зачем делается редукция решающих деревьев?
5. Какие есть два основных типа редукции решающих деревьев?
6. Как преобразовать решающее дерево в решающий список, и зачем это делается?
7. Что такое ADT (alternating decision tree)? Как происходит построение ADT?
8. Каковы основные цели кластеризации?
9. Основные типы кластерных структур. Приведите для каждой из этих структур пример алгоритма кластеризации, который для неё НЕ подходит.
10. В чём заключается алгоритм кратчайшего незамкнутого пути? Как его использовать для кластеризации? Как с его помощью определить число кластеров? Всегда ли это возможно?
11. Основные отличия алгоритма k-средних и EM-алгоритма. Кто из них лучше и почему?
12. Что такое дендрограмма? Всегда ли её можно построить?

Пример теста для контроля знаний обучающихся

Выберите правильные ответы (правильных ответов может быть несколько):

1. Большие данные – это:
 - а) Данные объемом более 1Тб
 - б) Данные объемом более 10Тб
 - в) Данные объемом более 100Тб
 - г) Нет ограничений на минимальный объем
2. Наиболее редко на практике применяются методы машинного обучения, основанные на:
 - а) Алгоритмах обучения без учителя
 - б) Алгоритмах обучения с учителем
 - в) Алгоритмах обучения с подкреплением
 - г) Свёрточных нейронных сетях
3. Как называется ряд задач, направленных на предсказание численного значения некоторой величины по входных данных?
 - а) Кластеризация б) Регрессия в) Прогноз г) Классификация
4. Реализация метода обучения с учителем не нуждается в:
 - а) Обучающей выборке

- б) Тестовой выборке
- в) Оценочной выборке
- г) Проверочной выборке

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

- а) классификация данных
- б) объекты с известными ответами
- в) алгоритм решающий функцию

6. Задача регрессии - это:

- а) множество объектов, разделенных на классы
- б) исследование влияние одного или нескольких признаков на объект
- в) определение порядка признака согласно рангу

7. Задача ранжирования - это:

- а) множество объектов, разделенных на классы
- б) исследование влияние одного или нескольких признаков на объект
- в) определение порядка признака согласно рангу

8. Какой тип экспериментального исследования имеет цель - понимание, на что влияют параметры метода обучения?

- а) исследование задач ранжирования
- б) исследование задач классификации
- в) исследование на модельных данных

9. Какой тип экспериментального исследования имеет цель - либо решение конкретной прикладной задачи, либо выявление «слабых мест»?

- а) исследование задач ранжирования
- б) исследование на реальных данных
- в) исследование на модельных данных

10. Что, из ниже перечисленного, не относится к типу экспериментального исследования?

- а) исследование задач ранжирования
- б) исследование на реальных данных
- в) исследование на модельных данных

11. Для машинного обучения подходят данные

- а) Бинарные
- б) Любых форматов в цифровом виде
- в) Предварительно подготовленные, очищенные от ошибок, пропусков и выбросов, а также нормализованные и представленные в виде числовых векторов
- г) Числовые типа int

12. Перечислите четыре основных характеристики Big Data:

- а) Virtualization, Volume, Variability, Vehicle
- б) Variety, Velocity, Volume, Value
- в) Verification, Volume, Velocity, Visualization
- г) Video, Value, Variety, Volume

13. Выберите неверное высказывание:

- а) большие объёмы данных приводят к слабой их структуризации, поэтому появляется такое разнообразие данных
- б) увеличившаяся производительность телекоммуникационных каналов привела к росту объёмов передаваемой информации
- в) удешевление систем хранения на единицу информации привело к росту рынка больших данных

14. Выберите верное утверждение::

- а) Data Warehouse создаются для проверки гипотез при анализе больших данных
- б) "Песочница" используется для снижения нагрузки на основной Data Warehouse
- в) каждый Data Warehouse должен содержать "песочницу"
- г) "Песочница" необходима для любого процесса аналитики

15. Отметьте те из вариантов, в которых данные структурированы:

- а) данные о продажах компании, представленные в виде ежемесячных отчётов в формате MS Word
- б) таблица с ежедневными показаниями температуры помещения за год в файле формата csv
- в) текст педагогической поэмы А.С. Макаренко, представленный в формате PDF
- г) библиотека фильмов, представленных в формате mp4 на одном жестком диске

16. Выберите неверное высказывание:

- а) большие объёмы данных приводят к слабой их структуризации, поэтому появляется такое разнообразие данных
- б) увеличившаяся производительность телекоммуникационных каналов привела к росту объёмов передаваемой информации
- в) удешевление систем хранения на единицу информации привело к росту рынка больших данных

17. Отметьте верное понимание Variety в контексте характеристик Big Data:

- а) высокая скорость генерирования данных
- б) разные типы данных в колонках таблиц реляционных СУБД
- в) разнообразие отраслей, являющихся источниками данных
- г) разнообразие типов данных, включающих в себя структурированные, полуструктурированные и неструктурированные

18.: Объекты состоят из признаков?

- а) Да
- б) Нет

19. Задача классификации - это:

- а) множество объектов, разделенных на классы
- б) исследование влияние одного или нескольких признаков на объект
- в) определение порядка признака согласно рангу

20. Верно ли утверждение? Всякая оптимизация по неполной информации и избыточная сложность параметров приводит в переобучению.

- а) Да
- б) Нет

21. С каким типом задачи синтеза связаны проблемы, возникающие у генеративных сетей?

- а) текстурный
- б) глубинный
- в) объемный
- г) цветной

22. Какая задача на данный момент не решена применением нейросетей?

- а) стилизация фото
- б) описание фото
- в) ретушь дефектов фото
- г) коррекция оптических деформаций

23. Какие задачи из ниже перечисленных относятся к задачам классификации?

- а) определение наиболее целесообразного способа лечения;
- б) определение длительности и исхода заболевания;
- в) оценивание кредитоспособности заёмщика;
- г) задачи поискового вывода

24. Какие задачи, из ниже перечисленных, являются задачами прогнозирования?

- а) математический прогноз даты сильных землетрясений;
- б) определение длительности и исхода заболевания;
- в) обнаружение спама;
- г) задачи поискового вывода.

25. Алгоритм k-средних предназначен для решения задачи:

- а) Классификации
- б) Кластеризации
- в) Прогнозирования
- г) Снижения размерности

26. Какой алгоритм не подходит для решения задачи, объекты в которой нужно разделить на классы?

- а) Случайный лес
- б) Дерево принятия решений
- в) Линейная регрессия
- г) Логистическая регрессия

27. Neo4j – это:

- а) База данных
- б) Архитектура нейронной сети
- в) Платформа распределенных вычислений
- г) Компилятор языка

28. Целью поиска ассоциативных правил является ...

- а) нахождение частых зависимостей между объектами или событиями;
- б) определение класса объекта по его характеристиками;
- в) определение по известным характеристиками объекта значение некоторого его параметра;
- г) поиск независимых групп и их характеристик во всем множестве анализируемых данных.

29. Выберите неверный ответ:

- а) большие данные – это данные объёма свыше 1 Тб
- б) проблема больших данных – это проблема, когда при существующих технологиях хранения и обработки существенная обработка данных затруднена или невозможна
- в) большие данные – это тренд в области ИТ, подогреваемый маркетинговыми кампаниями крупных вендоров
- г) большие данные как правило не структурированы

30. Принцип MapReduce не состоит в том, чтобы

- а) производить вычисления на узлах, где информация изначально была сохранена
- б) использовать вычислительные мощности систем хранения
- в) использовать функциональное программирование для решения задач массивно-параллельной обработки

Вопросы к зачету

31. Что такое оценка плотности Парзена-Розенблатта (надо помнить формулу). Выписать формулу алгоритма классификации в методе парзеновского окна.
32. На что влияет ширина окна, а на что вид ядра в методе парзеновского окна?
33. Многомерное нормальное распределение (надо помнить формулу). Вывести формулу квадратичного дискриминанта. При каком условии он становится линейным?
34. Что такое «смесь распределений» (надо помнить формулу)?
35. Что такое «выбросы»? Как осуществляется фильтрация выбросов?
36. Что такое обобщённый алгоритм классификации (надо помнить формулу)? Какие вы знаете частные случаи?
37. Как определяется понятие отступа в метрических алгоритмах классификации?
38. Что такое окно переменной ширины, в каких случаях его стоит использовать?
39. Метод стохастического градиента. Расписать градиентный шаг для квадратичной функции потерь и сигмоидной функции активации.
40. Что такое «сокращение весов»?
41. Обоснование логистической регрессии (основная теорема). Как выражается апостериорная вероятность классов (надо помнить формулу).
42. Как выражается функция потерь в логистической регрессии (надо помнить формулу).
43. Две мотивации и постановка задачи метода опорных векторов. Уметь вывести постановку задачи SVM (рекомендуется помнить формулу постановки задачи).
44. Какая функция потерь используется в SVM? В логистической регрессии? Какие ещё функции потерь Вы знаете?
45. Что такое ядро в SVM? Зачем вводятся ядра? Любая ли функция может быть ядром?
46. Что такое ROC-кривая, как она определяется? Как она эффективно вычисляется?
47. В каких алгоритмах классификации можно узнать не только классовую принадлежность классифицируемого объекта, но и вероятность того, что данный объект принадлежит каждому из классов?

48. Каков вероятностный смысл регуляризации? Какие типы регуляризаторов Вы знаете?
49. Что такое принцип максимума совместного правдоподобия данных и модели (надо помнить формулу)?
50. Постановка задачи многомерной линейной регрессии. Матричная запись.
51. Что такое сингулярное разложение? Как оно используется для решения задачи наименьших квадратов?
52. Что такое «проблема мультиколлинеарности» в задачах многомерной линейной регрессии? Какие есть три подхода к её устранению?
53. Сравнить гребневую регрессию и лассо. В каких задачах предпочтительнее использовать лассо?
54. Какую проблему решает метод главных компонент в многомерной линейной регрессии? Записать матричную постановку задачи для метода главных компонент.
55. Как свести задачу многомерной нелинейной регрессии к последовательности линейных задач?
56. Метод настройки с возвращениями (backfitting): постановка задачи и основная идея метода.
57. Какие методы построения логистической регрессии Вы знаете?
58. Метод обратного распространения ошибок. Основная идея. Основные недостатки и способы их устранения.
59. Как можно выбирать начальное приближение в градиентных методах настройки нейронных сетей?
60. Как можно ускорить сходимость в градиентных методах настройки нейронных сетей?
61. Что такое «паралич» сети, и как его избежать?
62. Как выбирать число слоёв в градиентных методах настройки нейронных сетей?
63. Как выбирать число нейронов скрытого слоя в градиентных методах настройки нейронных сетей?
64. В чём заключается метод оптимального прореживания нейронной сети? Какие недостатки стандартного алгоритма обратного распространения ошибок позволяет устранить метод ODB?
65. Основная идея отбора признаков с помощью генетического алгоритма.
66. Основная идея отбора признаков с помощью случайного поиска.
67. В чём отличия случайного поиска от случайного поиска с адаптацией?
68. Основная идея метода bagging.
69. Основная идея метода случайных подпространств.
70. Что такое смесь экспертов (помнить формулу)?
71. Приведите примеры выпуклых функций потерь. Почему свойство выпуклости помогает строить смеси экспертов?
72. Что такое решающий список?
73. Какие критерии информативности используются при синтезе решающего списка и почему?
74. Достоинства и недостатки решающих списков.
75. Что такое решающее дерево?
76. Какие критерии информативности используются при синтезе решающего дерева и почему?
77. Достоинства и недостатки решающих деревьев.
78. Зачем делается редукция решающих деревьев?
79. Какие есть два основных типа редукции решающих деревьев?
80. Как преобразовать решающее дерево в решающий список, и зачем это делается?
81. Что такое ADT (alternating decision tree)? Как происходит построение ADT?
82. Каковы основные цели кластеризации?

83. Основные типы кластерных структур. Приведите для каждой из этих структур пример алгоритма кластеризации, который для неё НЕ подходит.
84. В чём заключается алгоритм кратчайшего незамкнутого пути? Как его использовать для кластеризации? Как с его помощью определить число кластеров? Всегда ли это возможно?
85. Основные отличия алгоритма k-средних и EM-алгоритма. Кто из них лучше и почему?
86. Что такое дендрограмма? Всегда ли её можно построить?
87. Какой функционал качества оптимизируется сетью Кохонена? (помнить формулу)

Примеры контрольной работы

Комплект заданий №1

Задача 1

Разработать модель кредитного скоринга по существующей распределенной базе данных клиентов и их кредитных историй, используя алгоритм логистической регрессии с/без регуляризации и наивный байесовский классификатор.

Задача 2

Сравнить и проанализировать точность получаемых моделей.

Комплект заданий №2

Задача 1

Разработать модель вероятности банкротства компании по самостоятельно собранной распределенной базе данных, используя элементы парсинга, алгоритм градиентного бустинга и метод опорных векторов.

Задача 2

Сравнить и проанализировать точность получаемых моделей. Провести их интерпретацию. Сделать вывод относительно применимости того или иного алгоритма для анализируемого набора данных.

Комплект заданий №3

Задача 1

Разработать модель кредитного скоринга по существующей распределенной базе данных клиентов и их кредитных историй, используя алгоритм логистической регрессии с/без регуляризации.

Задача 2

Сравнить и проанализировать точность получаемых моделей.

Комплект заданий №4

Задача 1

Разработать модель вероятности банкротства компании по самостоятельно собранной распределенной базе данных, используя алгоритм градиентного бустинга. Проанализировать точность получаемых моделей.

Задача 2

Разработать модель кредитного скоринга по существующей распределенной базе данных клиентов и их кредитных историй, используя алгоритм логистической регрессии с/без регуляризации. Проанализировать точность получаемых моделей.

Примеры домашнего задания

Задание 1

Разработать модель кредитного скоринга по существующей распределенной базе данных клиентов и их кредитных историй, используя алгоритм логистической регрессии с/без регуляризации и наивный байесовский классификатор. Сравнить и проанализировать точность получаемых моделей. Провести их интерпретацию. Сделать вывод относительно применимости того или иного алгоритма для анализируемого набора данных.

Задание 2

Разработать модель кредитного скоринга по существующей распределенной базе данных клиентов и их кредитных историй, используя алгоритм искусственной нейронной сети и случайный лес. Сравнить и проанализировать точность получаемых моделей. Провести их интерпретацию. Сделать вывод относительно применимости того или иного алгоритма для анализируемого набора данных.

Задание 3

Разработать модель вероятности банкротства компании по самостоятельно собранной распределенной базе данных, используя элементы парсинга, алгоритм градиентного бустинга и метод опорных векторов. Сравнить и проанализировать точность получаемых моделей. Провести их интерпретацию. Сделать вывод относительно применимости того или иного алгоритма для анализируемого набора данных.

Приложение 1.

СИСТЕМА РЕЙТИНГОВОЙ ОЦЕНКИ И КОНТРОЛЯ ЗНАНИЙ СТУДЕНТОВ

№ п/п	СТРУКТУРА	Баллы по каждому модулю
1.	Оценка за активное участие в учебном процессе и посещение занятий: Всех занятий Не менее 75% Не менее 50% Не менее 25% Итого:	5 4 3 2 до 5
2.	устный опрос в форме собеседования (УО-1) письменный опрос в виде теста (ПР-1) письменная контрольная работа (ПР-2) устный опрос в форме коллоквиума (УО-2) письменная работа в форме реферата (ПР-4) Итого:	5 10 15 15 45
3.	Зачет	50
	ВСЕГО:	100

Пересчет на 5 балльную систему

2 (неудовлетворительно)	3 (удовлетворительно)	4 (хорошо)	5 (отлично)
----------------------------	--------------------------	---------------	----------------

< 50	50-64	65-84	85-100
----------------	--------------	--------------	---------------

Язык преподавания: русский.

Автор (авторы) программы: к.т.н., доцент Высшей школы управления и инноваций МГУ имени М.В. Ломоносова Н.А. Моисеев

Преподаватель (преподаватели) программы: д.э.н., профессор Высшей школы управления и инноваций МГУ имени М.В. Ломоносова Н.А. Моисеев