



Программа курса «Инженер данных (Data engineer Pro)»

Номер	Название темы	Количество часов	Описание темы
1	Модуль 1. Базовый	168	Введение в большие данные. Технологии ИИ. Машинное обучение. Системы хранения и обработки данных. Базы данных. SQL. Программные компоненты извлечения, хранения и подготовки больших данных. Словари и эталонная архитектура больших данных. Обработка, удаленная, распределенная и объединенная аналитика больших данных. Управление качеством и достоверностью больших данных. Языки программирования, среды исполнения и основные фреймворки машинного обучения. Пайплайн машинного обучения. Препроцессинг данных. Работа с размерностью и структурой данных. Обучение без учителя. Кластеризация. Классификация. Регрессия. Метрики. Оценка качества алгоритма. Искусственные нейронные сети. Ансамбли алгоритмов. Повышение точности. Рекомендательные системы.
2	Модуль 2. Профессиональный	74	Решение задач компьютерного зрения. Решение задач обработки естественного языка. Решение задач распознавания и синтеза речи. Реализация проектов в области ИИ: создание приложений и их интеграция.
3	Итоговая аттестация	10	Выпускная квалификационная работа.

Министерство образования и науки Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)



Дополнительное профессиональное образование

ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА
ПРОГРАММА ПРОФЕССИОНАЛЬНОЙ ПЕРЕПОДГОТОВКИ

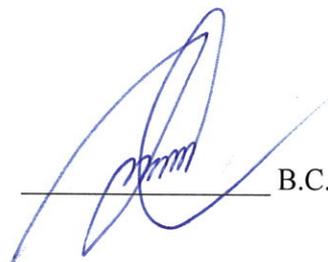
«Инженер данных (Data engineer Pro)»

Рег. № 05.22.23.01.90

Москва, 2023 г.

АВТОР ПРОГРАММЫ:

Доцент каф. ИУ12


В.С. Тынченко

СОГЛАСОВАНО:

Начальник УСП


Т.А. Гузева

Директор
Центра дополнительного образования


М.В. Стоянова



СОДЕРЖАНИЕ

1. Общая характеристика дополнительной профессиональной программы (ДПП).....	стр.	4
1.1. Цель ДПП	стр.	4
1.2. Планируемые результаты обучения	стр.	4
1.3. Дополнительные характеристики ДПП. (Характеристика новой квалификации и связанных с ней видов профессиональной деятельности, трудовых функций и (или) уровней квалификации. Профессиональные стандарты, квалификационные требования, указанные в квалификационных справочниках по соответствующим должностям, профессиям и специальностям, квалификационные требования к профессиональным знаниям и навыкам, необходимым для исполнения должностных обязанностей, которые устанавливаются в соответствии с Федеральными законами и иными нормативными правовыми актами Российской Федерации о государственной службе).....	стр.	4
1.4. Характеристика компетенций, подлежащих совершенствованию, и (или) перечень новых компетенций, формирующихся в результате освоения программы.....	стр.	4
1.5. Соответствие видов деятельности и профессиональных компетенций	стр.	5
2. Учебный план ДПП	стр.	6
2.1. Категория слушателей ДПП	стр.	6
2.2. Общая трудоёмкость программы, аудиторная и самостоятельная работа	стр.	6
2.3. Форма обучения	стр.	6
2.4. Учебный план ДПП	стр.	6
3. Календарный учебный график	стр.	6
4. Рабочие программы учебных предметов (курсов, дисциплин (модулей))	стр.	7
4.1. Рабочая программа учебного предмета (курса, дисциплины, модуля) №1 «Базовый модуль». Приложение №1.....	стр.	7
4.2. Рабочая программа учебного предмета (курса, дисциплины, модуля) №2 «Профессиональный модуль». Приложение № 2.....	стр.	7
5. Условия реализации ДПП	стр.	7
5.1. Организационные условия реализации ДПП	стр.	7
5.2. Педагогические условия реализации ДПП	стр.	7
5.3. Учебно-методическое обеспечение ДПП	стр.	7
6. Формы итоговой аттестации	стр.	9
7. Оценочные материалы итоговой аттестации	стр.	10
7.1. Комплект оценочных средств	стр.	10
7.2. Паспорт фонда оценочных средств	стр.	17

1. ОБЩАЯ ХАРАКТЕРИСТИКА ДПП

Программа подготовлена на основе:

- Федерального закона от 29 декабря 2012 года № 273-ФЗ «Об образовании в Российской Федерации»;
- требований Приказа Минобрнауки России от 01.07.2013 года № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам»;
- методических рекомендаций-разъяснений по разработке дополнительных профессиональных программ на основе профессиональных стандартов от 22 апреля 2015 года № ВК-1030/06.

1.1. Цель ДПП

Подготовить высококвалифицированных и конкурентоспособных специалистов в области профессиональной деятельности, которая включает интеллектуальные системы, биоинформатику, когнитивные информационные технологии, вычислительные технологии, компьютерные науки, технологии баз данных, компьютерную графику, теорию информации, технологии управления инфокоммуникацией и бизнес-процессами, архитектуру программного обеспечения, параллельное и распределенное программирование.

Сформировать у слушателей следующие компетенции в области профессиональной деятельности: создание информационных технологий нового поколения, обеспечивающих экономически эффективное извлечение полезной информации из больших объемов разнообразных данных путем высокой скорости их сбора, обработки и анализа, и применение этих технологий в информационно-аналитической деятельности, в системах управления и принятия решений, а также для разработки на их основе новых продуктов и услуг.

1.2. Планируемые результаты обучения

- освоение профессиональных компетенций в процессе изучения перечисленных в учебном плане профессиональных дисциплин.

- успешное освоение дисциплин программы профессиональной переподготовки, защита выпускной квалификационной работы, получение диплома о профессиональной переподготовке по программе ПП «Инженер данных (Data engineer Pro)», который дает право на ведение нового вида профессиональной деятельности - Разработка программного обеспечения.

1.3. Дополнительные характеристики ДПП.

Характеристика новой квалификации и связанных с ней видов профессиональной деятельности, трудовых функций и (или) уровней квалификации. Профессиональные стандарты, квалификационные требования, указанные в квалификационных справочниках по соответствующим должностям, профессиям и специальностям, квалификационные требования к профессиональным знаниям и навыкам, необходимым для исполнения должностных обязанностей, которые устанавливаются в соответствии с Федеральными законами и иными нормативными правовыми актами Российской Федерации о государственной службе.

Характеристики новой квалификации определены в приказе Минтруда России от 06 июля 2020 года N 405н «Об утверждении профессионального стандарта «Специалист по большим данным» (06.042)

Связанные с новой квалификацией виды:

- **профессиональной деятельности:** Создание и применение технологий больших данных.

- **трудовые функции:**

ОТФ Управление разработкой продуктов, услуг и решений на основе больших данных С
ТФ Разработка продуктов на основе встроенной аналитики больших данных (С/01.8)

1.4. Характеристика компетенций, подлежащих совершенствованию, и (или) перечень новых компетенций, формирующихся в результате освоения программы

Профессиональные компетенции базируются на основании Приказа Минобрнауки России от 23 августа 2017 г. № 811 «Об утверждении федерального государственного образовательного стандарта высшего образования по направлению подготовки 02.04.02 Фундаментальная информатика и информационные технологии (уровень магистратуры)».

Перечень компетенций:

Перечень компетенций согласно федерального государственного образовательного стандарта высшего образования по направлению подготовки/специальности:

02.04.02 Фундаментальная информатика и информационные технологии (уровень магистратуры).

УК-1. Способен осуществлять критический анализ проблемных ситуаций на основе системного подхода, вырабатывать стратегию действий.

УК-2. Способен управлять проектом на всех этапах его жизненного цикла.

УК-3. Способен организовывать и руководить работой команды, вырабатывая командную стратегию для достижения поставленной цели.

УК-4. Способен применять современные коммуникативные технологии, в том числе на иностранном(ых) языке(ах), для академического и профессионального взаимодействия.

УК-5. Способен анализировать и учитывать разнообразие культур в процессе межкультурного взаимодействия.

УК-6. Способен определять и реализовывать приоритеты собственной деятельности и способности ее совершенствования на основе самооценки.

ОПК-1. Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий.

ОПК-2. Способен применять компьютерные/ суперкомпьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности.

ОПК-3. Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования.

ОПК-4. Способен оптимальным образом комбинировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности.

ОПК-5. Способен устанавливать и сопровождать программное обеспечение информационных систем, осуществлять эффективное управление разработкой программных средств и проектов.

1.5. Соответствие видов деятельности и профессиональных компетенций

Код и наименование вида (ов) деятельности	Наименование профессиональных компетенций
ВД-1 Создание и применение технологий больших данных	ОПК-1. Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий
	ОПК-2. Способен применять компьютерные/ суперкомпьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности
	ОПК-3. Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования
	ОПК-4. Способен оптимальным образом комбинировать существующие информационно-коммуникационные технологии для решения за-

	дач в области профессиональной деятельности с учетом требований информационной безопасности
	ОПК-5. Способен устанавливать и сопровождать программное обеспечение информационных систем, осуществлять эффективное управление разработкой программных средств и проектов

2. УЧЕБНЫЙ ПЛАН ДПП

2.1. Категория слушателей ДПП

Учебный план реализуется для специалистов или бакалавров и (или) магистров, или же слушателей, обучающихся по направлению 02.04.02 Фундаментальная информатика и информационные технологии (уровень магистратуры)».

2.2. Общая трудоёмкость программы, аудиторная и самостоятельная работа

Общая трудоёмкость программы, часов: 252, из них 140 аудиторной работы и 98 самостоятельной работы, 14 академических часов итоговой аттестации.

2.3. Форма обучения: очная с применением дистанционных технологий.

2.4. Учебный план

№ п/п	Наименование модулей/тем программы	Всего, час	Виды учебных занятий				Формы контроля
			Лекции	Практ. занятия	Сам. работа	Промежут. аттестация	
1	Базовый модуль	168	32	66	68	2	Зачет
2	Профессиональный модуль	74	2	40	30	2	Зачет
3	Итоговая аттестация	10	0	0	0	-	Подготовка и защита ВКР
Всего часов		252	34	106	98	4	

3. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК

№ п/п	Наименование темы, модуля	1 неделя	2 неделя	3 неделя	4 неделя	5 неделя	6 неделя
1.	Базовый модуль						
2.	Профессиональный модуль						
3.	Итоговая аттестация						Подготовка и защита ВКР

1 неделя – 6 рабочих дней

Минимальный срок освоения ДПП – 6 недель.

4. РАБОЧИЕ ПРОГРАММЫ УЧЕБНЫХ ПРЕДМЕТОВ (КУРСОВ, ДИСЦИПЛИН (МОДУЛЕЙ))

4.1. Рабочая программа учебного предмета (курса, дисциплины, модуля) №1 «Базовый модуль». Приложение №1

4.2. Рабочая программа учебного предмета (курса, дисциплины, модуля) №2 «Профессиональный модуль». Приложение №2

5. УСЛОВИЯ РЕАЛИЗАЦИИ ДПП

5.1. Организационные условия реализации ДПП

Наименование аудитории	Вид занятия	Наименование оборудования, программного обеспечения
Аудитория для проведения лекций/семинаров	Лекции	ПК с доступом в Интернет и возможностью просмотра файлов в формате *.ppt, *.pptx, *.pdf; микрофон; колонки/наушники; камер.
Лаборатории/ производственные помещения	Практические занятия	Личный ПК/смартфон с доступом в Интернет и возможностью просмотра файлов в формате *.html, *.doc, *.docx, *.pdf; лист формата А5/А4 или блокнот; карандаш/ручка.
Коворкинги, учебные залы и т.д.	Самостоятельная работа	Личный ПК/смартфон с доступом в Интернет и возможностью просмотра файлов в формате *.html, *.doc, *.docx, *.pdf; лист формата А5/А4 или блокнот; карандаш/ручка.
Аудитория для проведения лекций/семинаров	Итоговая аттестация	ПК с доступом в Интернет и возможностью просмотра файлов в формате *.html, *.doc, *.docx, *.pdf, *.djvu, лист бумаги формата А4, ручка.

5.2. Педагогические условия реализации ДПП

№ п/п	Наименование учебной дисциплины	Преподаватель	Подпись преподавателя
1	Базовый модуль	С.В. Картамышев	
2	Профессиональный модуль	С.В. Картамышев	

5.3. Учебно-методическое обеспечение ДПП (по всей программе ДПП)

Основная литература:

1. Алпайдин Э. Машинное обучение: новый искусственный интеллект: пер. с англ. / Алпайдин Э. - М.: Фонд Развития Промышленности: Издательская группа "Точка" : Альпина Паблишер : [Интеллектуальная Литература], 2017. - 191 с. - (Завтра это будут знать все). - Библиогр.: с. 185-191. - ISBN 978-5-9908700-8-6. - ISBN 978-5-9614-6114-5.

2. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных / Силен Д., Мейсман А., Али М. ; пер. с англ. Матвеев Е. - СПб. : Питер, 2020. - 334 с. : ил. - (Библиотека программиста). - ISBN 978-5-4461-0944-9.

3. Хайкин С. Нейронные сети. Полный курс: пер. с англ. / Хайкин С.; пер. КуССуль Н. Н., Шелестов А. Ю. - 2-е изд. - М.; СПб.; Киев: Вильямс, 2006. - 1103 с.: ил. - Библиогр.: с. 996-1069. - ISBN 5-8459-0890-6.

4. Маркин, А. В. Программирование на SQL: учебное пособие для среднего профессионального образования / А. В. Маркин. — Москва: Издательство Юрайт, 2020. — 435 с. —

(Профессиональное образование). — ISBN 978-5-534-11093-7. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/456926>

5. Маркин, А. В. Программирование на SQL в 2 ч. Часть 2: учебник и практикум для вузов / А. В. Маркин. — 2-е изд., испр. и доп. — Москва: Издательство Юрайт, 2020. — 340 с. — (Высшее образование). — ISBN 978-5-534-12258-9. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/451185>.

Дополнительные материалы:

1. Белоус В. В., Пивоварова Н. В. Основы реляционных баз данных. Практикум по SQL [Электрон. ресурс]: метод. указ. к лаб. работам по курсу "Базы данных" / Белоус В. В., Пивоварова Н. В.; МГТУ им. Н. Э. Баумана, Фак. "Робототехника и комплексная автоматизация". - М.: Изд-во МГТУ им. Н. Э. Баумана, 2014. - 1 CD-ROM. - ФГУП "Информрегистр" №0321400922.

2. Басараб М. А., Коннова Н. С. Интеллектуальные технологии на основе искусственных нейронных сетей: метод. указания к выполнению лаб. работ / Басараб М. А., Коннова Н. С.; МГТУ им. Н. Э. Баумана. - М.: Изд-во МГТУ им. Н. Э. Баумана, 2017. - 53 с.: ил. - Библиогр. в конце брош. - ISBN 978-5-7038-4716-9.

3. Березкин Д. В., Рожнев А. Ю. Разработка модели оценки платежеспособности клиентов банка с применением алгоритмов машинного обучения / Березкин Д. В., Рожнев А. Ю. // Динамика сложных систем. - 2018. - Т. 12, № 4. - С. 59-66.

4. Бизли Д. М. Язык программирования Python: справочник: Пер. с англ / Бизли Д. М. - Киев: ДиаСофт, 2000. - 326 с. - ISBN 966-7393-54-2.

5. Вьюгин В. В. Элементы математической теории машинного обучения: учеб. пособие для вузов / Вьюгин В. В.; Моск. физико-техн. ин-т (гос. ун-т), РАН. Ин-т проблем передачи информации им. А. А. Харкевича. - М.: МФТИ - ИППИ РАН, 2010. - 231 с. - Библиогр.: с. 229-231. - ISBN 978-5-7417-0339-7.

6. Галушкин А. И. Нейронные сети: основы теории / Галушкин А. И. - М.: Горячая линия - Телеком, 2010. - 496 с.: ил. - Библиогр. в конце ст., с. 469-488. - ISBN 978-5-9912-0082-0.

7. Головкин В. А. Нейронные сети: обучение, организация и применение: учеб. пособие по направлению подготовки бакалавров и магистров / Головкин В. А.; общ. ред. Галушкин А. И. - М.: Издат. предприятие ред. журн. "Радиотехника" (ИПРЖР), 2001. - 256 с.: ил. - (Нейрокомпьютеры и их применение; кн. 4). - Библиогр.: с. 248-256. - ISBN 5-93108-05-8.

8. Джеймс Р., Грофф П., Вайнберг Н., Оппель Э. Дж. SQL. Полное руководство. М: Вильямс, 2014. 960 с.

9. Коннолли Т., Бегг К. Базы данных. Проектирование, реализация и сопровождение. Теория и практика. М: Вильямс, 2017. 1440 с.

10. Комашинский В. И., Смирнов Д. А. Нейронные сети и их применение в системах управления и связи / Комашинский В. И., Смирнов Д. А. - М.: Горячая линия - Телеком, 2003. - 93 с. - Библиогр.: с. 88-93. - ISBN 5-93517-094-9.

11. Осовский С. Нейронные сети для обработки информации: пер. с польск. / Осовский С.; пер. Рудинский И. Д. - М.: Финансы и статистика, 2002. - 343 с.: ил. - ISBN 5-279-02567-4.

12. Сидняев Н. И., Храпов П. В. Нейросети и нейроматематика: учеб. пособие / Сидняев Н. И., Храпов П. В.; МГТУ им. Н. Э. Баумана. - М.: Изд-во МГТУ им. Н. Э. Баумана, 2016. - 83 с.: ил. - Библиогр.: с. 82. - ISBN 978-5-7038-4362-8.

13. Хливненко Л. В., Пятакович Ф. А. Практика нейросетевого моделирования: учеб. пособие / Хливненко Л. В., Пятакович Ф. А. - СПб. Лань, 2019. - 196 с. - (Учебники для вузов. Специальная литература) (Бакалавриат и магистратура). - Библиогр.: с. 182-193. - ISBN 978-5-8114-3639-2.

6. ФОРМЫ ИТОГОВОЙ АТТЕСТАЦИИ ДПП

Итоговая аттестация проводится в форме защиты выпускной квалификационной работы (ВКР)

Защита ВКР проводится в присутствии итоговой экзаменационной комиссии (ИЭК).

6.1 Процедура оценивания результатов освоения образовательной программы на защите выпускной квалификационной работы

Завершающим этапом выполнения студентом ВКР является ее защита. Защита ВКР служит элементом обязательного тестирования, проводимого в рамках итоговой аттестации выпускника, по результатам которого ИЭК выносит решение о выдаче диплома о профессиональной переподготовке, дающего право на ведение нового вида профессиональной деятельности – разработка программного обеспечения, при условии успешной защиты ВКР.

К защите ВКР допускаются слушатели, успешно завершившие в полном объеме освоение ДПП, успешно прошедшие промежуточные испытания и представившие ВКР с отзывом руководителя в установленный срок, на которую получена положительная рецензия.

Защита ВКР проводится в соответствии с утвержденным графиком, утверждаемым Директором МИЦ «Композиты России» МГТУ им. Н.Э. Баумана с участием не менее 2/3 членов ее состава.

Обязательными элементами процедуры защиты являются:

выступление слушателя – автора ВКР;

ответы студента на вопросы членов ИЭК;

оглашение отзыва руководителя;

оглашение рецензии и ответы слушателя на замечания рецензента.

Для сообщения по содержанию ВКР слушателю отводится, как правило, не более 10 минут. Для защиты слушателем могут представляться дополнительные материалы, характеризующие научную и практическую ценность выполненной работы (печатные статьи по теме, документы, указывающие на практическое применение результатов работы, акты внедрения и т.п.). На открытой защите ВКР могут присутствовать все желающие, которым председатель вправе разрешить задавать слушателю вопросы по теме, защищаемой им работы. Общая продолжительность защиты одной ВКР не должна превышать 0,5 часа.

В ходе защиты ведется протокол заседания ИЭК, в который вносятся: фамилия, имя, отчество обучающегося, название ВКР, присутствующие члены ИЭК, фамилия, имя, отчество руководителя ВКР, консультанта ВКР (если есть), перечисляются предоставленные к защите документы, заданные слушателю на защите вопросы, общая характеристика ответов слушателя, решение комиссии об оценке. Протокол подписывает председатель и члены ИЭК, участвовавшие в заседании.

Результаты защиты ВКР определяются путем открытого голосования членов ИЭК на основе оценок:

руководителя за качество ВКР, степени ее соответствия требованиям, предъявляемым к ВКР;

членов экзаменационной комиссии за содержание ВКР, ее защиту, включая доклад, ответы на вопросы членов ИЭК.

В случае возникновения спорной ситуации Председатель ИЭК имеет решающий голос.

Результат защиты ВКР слушателя оценивается по пятибалльной системе оценки знаний и проставляется в протокол заседания ИЭК, в котором расписываются председатель и члены эк-

заменационной комиссии. Оценки объявляются обучающимся в день защиты. После объявления оценок и рекомендаций комиссии защита выпускных квалификационных работ объявляется на текущий день законченной.

По результатам аттестационных испытаний обучающийся имеет право на апелляцию.

Обучающиеся, не прошедшие итоговой аттестации в связи с неявкой на аттестационное испытание по уважительной причине (временная нетрудоспособность, исполнение общественных или государственных обязанностей, вызов в суд, транспортные проблемы (отмена рейса, отсутствие билетов) или в других случаях), вправе пройти ее в течение 6 месяцев после завершения итоговой аттестации.

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ИТОГОВОЙ АТТЕСТАЦИИ

7.1.1 Комплект оценочных средств

Примерная тематика выпускных квалификационных работ:

1. Большие данные.
2. Искусственный интеллект.
3. Информационная безопасность.
4. Архитектура данных.
5. Модель данных.
6. Сбор и хранение данных.
7. Конвейер данных.
8. Представление данных.
9. Подготовка данных.
10. Облачные технологии.

7.1.2. Компетенции и критерии их оценивания

При подготовке и защите ВКР устанавливаются следующие компетенции и критерии их оценивания:

Формулировка компетенции по ФГОС	Код компетенции	Совокупные результаты освоения
Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий	ОПК-1	<p>Знать: Математическое моделирование</p> <p>Уметь: Осуществлять математическое и информационное моделирование</p> <p>Владеть: Проведение научно-исследовательских и опытно-конструкторских работ по созданию вариантов технического воплощения концепции продукта, создание прототипа продукта на основе встроенной аналитики</p>

Формулировка компетенции по ФГОС	Код компетенции	Совокупные результаты освоения
		больших данных
Способен применять компьютерные/ супер-компьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности	ОПК-2	<p>Знать: Технические средства и среды сбора, хранения и обработки больших данных</p> <p>Уметь: Осуществлять разработку программно-аппаратных компонентов и систем</p> <p>Владеть: Анализ инноваций в области информационных технологий; новых технических средств, методов и алгоритмов анализа больших данных; источников информации; технологий представления данных, методов предсказательной и предписывающей аналитики; существующих продуктов на основе встроенной аналитики больших данных</p>
Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования	ОПК-3	<p>Знать: Существующие и перспективные методы и программный инструментарий технологий больших данных</p> <p>Уметь: Проводить аналитические и поисковые исследования по тематике информационных технологий, технологий больших данных</p> <p>Владеть: Анализ требований к продукту, уточнение и доработка концепции, бизнес-модели и бизнес-плана создания нового продукта на основе встроенной аналитики</p>

Формулировка компетенции по ФГОС	Код компетенции	Совокупные результаты освоения
		тики больших данных
Способен оптимальным образом комбинировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности	ОПК - 4	<p>Знать: Принципы и методы управления защитой и обеспечением конфиденциальности больших данных</p> <p>Уметь: Проводить аналитические работы на основе технологий больших данных</p> <p>Владеть: Создание и оценка концепции и бизнес-модели продукта на основе встроенной аналитики больших данных</p>
Способен устанавливать и сопровождать программное обеспечение информационных систем, осуществлять эффективное управление разработкой программных средств и проектов	ОПК-5	<p>Знать: Современные и перспективные методы сбора, хранения и передачи данных из гетерогенных источников</p> <p>Уметь: Осуществлять разработку программно-аппаратных компонентов и систем</p> <p>Владеть: Разработка эксплуатационной документации по продукту на основе встроенной аналитики больших данных</p>

7.1.3. Шкала оценивания уровня овладения компетенциями при подготовке и защите ВКР

Оцениваемые позиции	Код компетенции	Рейтинг (в баллах)
1. Степень соответствия работы уровню квалификационных требований, предъявляемых к подготовке специалистов, а также требованиям, предъявляемых к ВКР	ОПК - 1	0-5
	ОПК - 2	0-5
	ОПК - 3	0-5
	ОПК - 4	0-5
	ОПК - 5	0-5
2. Соответствие базы источников, содержания и выводов теме, цели и задачам ВКР, достаточность и современность использованного библиографического материала и иных источников.	ОПК - 1	0-5
	ОПК - 4	0-5
3. Качество выполнения поставленных задач: - наличие в работе всех структурных элементов исследования; - использование эффективных методов исследования выбранных объектов; - наличие обоснованной авторской позиции, раскрывающей видение сущности проблемы исследователем и выбора методов ее решения; - использование в экспериментальной части исследования обоснованного комплекса методов и методик, позволяющих решить поставленные задачи; - целостность исследования, которая проявляется в связанности его теоретической и экспериментальной частей	ОПК - 1	0-5
	ОПК - 3	0-5
4. Степень самостоятельности студента при подготовке ВКР	ОПК - 1	0-5
	ОПК - 2	0-5
	ОПК - 3	0-5
	ОПК - 4	0-5
	ОПК - 5	0-5
5. Научная и практическая ценность сделанных выводов, перспективность исследования: наличие в работе материала (идей, экспериментальных данных и пр.), который может	ОПК - 1	0-5
	ОПК - 3	0-5

Оцениваемые позиции	Код компетенции	Рейтинг (в баллах)
стать источником дальнейших исследований	ОПК - 4	0-5
6. Соответствие оформления ВКР установленным требованиям	ОПК - 2	0-5
7. Выступления студента на научных конференциях по материалам ВКР, научные публикации	ОПК - 1	0-5
8. Защита ВКР: - качество доклада: композиция, полнота представления работы, ее результатов, аргументированность, убедительность; - объем и глубина знаний по теме, эрудированность, использование межпредметных связей; - культура речи, манера изложения, чувство времени, контроль над вниманием аудитории; - качество ответов на вопросы: полнота, аргументированность, использование при ответах сильных сторон работы; - деловые и волевые качества докладчика: ответственность, стремление к достижению высоких результатов, готовность к дискуссии; - наличие и качество презентации/раздаточного материала	ОПК - 1	0-5
Итого		0-100

Критерии оценивания компетенций при рейтинге от 0 до 5 баллов

5 баллов: слушатель показывает высокий уровень компетентности, знания материала программы, учебной, периодической и монографической литературы, раскрывает основные понятия и проводит их анализ на основании позиций различных авторов, в том числе иностранных авторов.

Слушатель показывает высокий уровень теоретических знаний по дисциплинам, включенным в государственный экзамен по профилю, и видит междисциплинарные связи.

Профессионально, грамотно, последовательно, хорошим языком четко излагает материал, аргументировано формулирует выводы.

Знает в рамках требований к направлению подготовки законодательно-нормативную и практическую базу.

На вопросы членов комиссии отвечает кратко, аргументировано, уверенно, по существу.

4 балла: слушатель показывает достаточный уровень компетентности, знания лекционного материала, учебной и методической литературы.

Уверенно и профессионально, грамотным языком, ясно, четко и понятно излагает состояние и суть вопроса.

Знает нормативно-законодательную и практическую базу, но при ответе допускает несущественные погрешности.

Слушатель показывает достаточный уровень профессиональных знаний, свободно оперирует понятиями, методами оценки принятия решений, имеет представление о междисциплинарных связях, увязывает знания, полученные при изучении различных дисциплин, умеет анализировать практические ситуации, но допускает некоторые погрешности.

Ответ построен логично, материал излагается хорошим языком, привлекается информативный и иллюстративный материал, но при ответе допускает некоторые погрешности.

Вопросы, задаваемые членами экзаменационной комиссии, не вызывают существенных затруднений.

от 0 до 3 баллов: слушатель показывает слабые знания лекционного материала, учебной литературы, законодательства и практики его применения, низкий уровень компетентности, неуверенное изложение вопроса.

Слушатель показывает слабый уровень профессиональных знаний, затрудняется при анализе практических ситуаций.

Не может привести примеры из реальной практики.

Неуверенно и логически непоследовательно излагает материал.

Неправильно отвечает на поставленные членами комиссии вопросы или затрудняется с ответом.

7.1.4. Требования к структуре и оформлению выпускных квалификационных работ

Выпускная квалификационная работа (ВКР) – обязательная часть итоговой аттестации выпускников. Выполнение ВКР является заключительным этапом, подводящим итоги освоения образовательной программы и служащим средством контроля приобретенных слушателем знаний, умений и компетенций за весь период обучения в Университете по образовательной программе, на основе которого экзаменационной комиссией (ЭК) принимается решение о присвоении выпускнику соответствующей квалификации.

В качестве составных частей в ВКР могут войти история исследуемого вопроса; сравнительный анализ российской и зарубежной практики; общая практика в рассматриваемой сфере, опыт решения обозначенных задач в пределах определенной территории или в рамках деятельности конкретных субъектов.

ВКР оформляется в виде текста с приложением таблиц, схем, графиков и другого иллюстративного материала.

Структура ВКР:

- титульный лист;
- задание на выполнение ВКР;

- календарный план на выполнение ВКР;
- реферат;
- содержание;
- список обозначений и сокращений;
- введение;
- основная часть (разделы) (направленность ВКР может быть: аналитическая, исследовательская, практическая, теоретическая), включающая в себя не менее двух частей (чаще всего теоретическую и практическую);
- заключение (выводы по выполненной работе);
- список использованных источников;
- приложения.

К ВКР прикладываются:

- рецензия на ВКР;
- направление на защиту ВКР.

Обязательным требованием для ВКР является логическая связь между ее частями и последовательное развитие основной идеи темы на протяжении всей работы.

Реферат. Реферат должен в кратком виде, в объеме до одной страницы, отражать данные об объеме работы, количестве разделов, иллюстраций, таблиц, приложений, использованных источников, цель и объект ВКР, методологию проведения работы, полученные результаты и новизну, область применения, степень внедрения и рекомендации по внедрению результатов исследования, экономическую эффективность или значимость работы.

Содержание. В содержании приводится перечень частей и разделов ВКР с указанием номеров страниц, на которых начинается каждый элемент работы.

Введение. Во введении обосновывается выбор темы, определяемый ее актуальностью; формулируется проблема и круг вопросов, необходимых для ее решения; определяется цель работы с ее разделением на взаимосвязанный комплекс задач, подлежащих решению, для раскрытия темы; указываются объект и предмет исследования или разработки, теоретическая и практическая значимость проведенного исследования, определяются методы исследования, дается краткий обзор базы исследования и литературных источников.

Основная часть. Содержит основные материалы ВКР (аналитические, исследовательские, практические, теоретические и др.). ВКР должна включать не менее двух глав, она может быть представлена теоретическим и практическим разделами. В основной части ВКР приводятся данные, отражающие сущность, методику и основные результаты исследования.

Содержательно главы, как правило, включают в себя:

- анализ истории вопроса и его современного состояния, обзор литературы по исследуемой теме, представление различных точек зрения и обоснование позиций автора исследования, анализ и классификацию используемого материала на базе избранной студентом методики исследования;
- описание процесса теоретических и (или) практических исследований, методов иссле-

дований, методов расчета, обоснование необходимости проведения анализа практики, ее характеристика;

- обобщение и оценку результатов исследований, включающих оценку полноты решения поставленной задачи и предложения по дальнейшим направлениям работ, оценку достоверности полученных результатов и их сравнение с аналогичными результатами отечественных и зарубежных работ.

В конце каждой главы следует обобщить материал в соответствии с целями и задачами, сформулировать выводы и достигнутые результаты.

Заключение. В заключении логически последовательно излагаются теоретические и практические выводы и предложения, к которым пришел студент в результате проделанной работы. Пишется заключение в виде тезисов (по пунктам). Выводы должны быть краткими и четкими, дающими полное представление о содержании, значимости, обоснованности и эффективности выполненной работы.

Список использованных источников. Должен включать изученную и использованную в ВКР литературу. Общее количество источников информации в списке должно содержать не менее 20-ти наименований. В списке использованных источников должны быть указаны нормативные правовые акты, учебные и научные издания, в том числе – обязательно из электронно-библиотечной системы и могут быть труды преподавателей Университета, статьи из профессиональной отечественной и зарубежной периодической печати, издания Всемирной организации интеллектуальной собственности, ведущих иностранных ВУЗов.

Список использованных источников оформляется по ГОСТ 7.1-2003 «Библиографическая запись. Библиографическое описание. Общие требования и правила составления», ГОСТ 7:82-2001 «Библиографическая запись. Библиографическое описание электронных ресурсов. Общие требования и правила составления» и ГОСТ 7.0.5-2008 «Справки по оформлению списка литературы».

Приложения. В приложения следует выносить вспомогательный материал, связанный с выполненной ВКР, который при включении в основную часть работы загромождает текст.

К вспомогательному материалу относятся: справочные материалы, промежуточные расчеты, таблицы вспомогательных цифровых данных, схемы, нормативные документы, образцы документов, инструкции, методики, распечатки компьютерных программ, иллюстрации вспомогательного характера, заполненные формы отчетности и других документов.

7.2. Паспорт фонда оценочных средств

п/п	Контролируемые этапы итоговой аттестации	Код контролируемой компетенции (или ее части)	Наименование оценочного средства
1.	Подготовка и защита ВКР	ОПК-1, ОПК-2, ОПК-3, ОПК-4, ОПК-5	Требования к содержанию, структуре, оформлению и защите ВКР

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
«**Базовый модуль**»

ДОПОЛНИТЕЛЬНОЙ ПРОГРАММЫ ПРОФЕССИОНАЛЬНОЙ
ПЕРЕПОДГОТОВКИ «**Инженер данных (Data engineer Pro)**»

Москва, 2023

АВТОР ПРОГРАММЫ:

Доцент каф. ИУ12

_____ В.С. Тынченко

СОГЛАСОВАНО:

Начальник УСП

_____ Т.А. Гузева

Директор
Центра дополнительного образования

_____ М.В. Стоянова

Оглавление

1. ОБЩАЯ ХАРАКТЕРИСТИКА РП.....	4
1.1. Цель	4
1.2. Планируемые результаты обучения	4
1.3. Дополнительные характеристики РП.....	5
1.4. Перечень профессиональных компетенций в рамках имеющейся квалификации, качественное изменение которых осуществляется в результате обучения.....	5
1.5. Соответствие видов деятельности профессиональным компетенциям и их составляющих.....	6
2. УЧЕБНЫЙ ПЛАН	7
2.1. Категория слушателей РП	7
2.2. Общая трудоёмкость программы, аудиторная и самостоятельная работа	7
2.3. Форма обучения	7
2.4. Учебный план	7
3. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК.....	8
4. РАБОЧАЯ ПРОГРАММА	9
5. УСЛОВИЯ РЕАЛИЗАЦИИ РП.....	27
5.1. Организационные условия реализации РП.....	27
5.2. Педагогические условия реализации РП	27
5.3. Учебно-методическое обеспечение РП.....	28
5.4. Методические рекомендации.....	29
6. ФОРМЫ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ РП.....	30
7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ.....	31
7.1. Паспорт комплекта оценочных средств.....	31
7.2. Комплект оценочных средств	31

1. ОБЩАЯ ХАРАКТЕРИСТИКА РП

Программа подготовлена на основе:

- Федерального закона от 29 декабря 2012 года № 273-ФЗ «Об образовании в Российской Федерации»;
- требований Приказа Минобрнауки России от 01.07.2013 года № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам»;
- методических рекомендаций-разъяснений Минобрнауки России по разработке дополнительных профессиональных программ на основе профессиональных стандартов от 22 апреля 2015 года № ВК-1030/06.

Реализация программы РП направлена на получение новой компетенции, необходимой для профессиональной деятельности.

1.1. Цель

Подготовить высококвалифицированных и конкурентоспособных специалистов в области профессиональной деятельности, которая включает интеллектуальные системы, биоинформатику, когнитивные информационные технологии, вычислительные технологии, компьютерные науки, технологии баз данных, компьютерную графику, теорию информации, технологии управления инфокоммуникацией и бизнес-процессами, архитектуру программного обеспечения, параллельное и распределенное программирование.

Сформировать у слушателей следующие компетенции в области профессиональной деятельности: создание информационных технологий нового поколения, обеспечивающих экономически эффективное извлечение полезной информации из больших объемов разнообразных данных путем высокой скорости их сбора, обработки и анализа, и применение этих технологий в информационно-аналитической деятельности, в системах управления и принятия решений, а также для разработки на их основе новых продуктов и услуг.

1.2. Планируемые результаты обучения

Планируемые результаты обучения по РП:

- освоение профессиональных компетенций в процессе изучения перечисленных тем в учебном плане;
- успешное освоение программы повышения квалификации;
- успешное прохождение промежуточной аттестации (зачет).

Обучающиеся, успешно прошедшие обучение, выполнившие текущие контрольные задания и выдержавшие предусмотренный учебным планом зачет по РП «**Базовый модуль**», могут быть допущены к освоению следующего - второго уровня дополнительной программы профессиональной переподготовки «Инженер данных (Data engineer Pro)».

1.3. Дополнительные характеристики РП

Характеристики новой квалификации определены в приказе Минтруда России от Минтруда России от 06 июля 2020 года N 405н «Об утверждении профессионального стандарта «Специалист по большим данным».

Вид профессиональной деятельности:

- создание и применение технологий больших данных (Код 06.042).

Трудовые функции:

- разработка продуктов на основе встроенной аналитики больших данных (С/01.8).

1.4. Перечень профессиональных компетенций в рамках имеющейся квалификации, качественное изменение которых осуществляется в результате обучения

Получаемые компетенции базируются на основании Приказа Минобрнауки России от 23 августа 2017 г. № 811 «Об утверждении федерального государственного образовательного стандарта высшего образования по направлению подготовки 02.04.02 Фундаментальная информатика и информационные технологии (уровень магистратуры)».

Перечень компетенций:

ОПК-1. Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий.

ОПК-2. Способен применять компьютерные/ суперкомпьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности.

ОПК-3. Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования.

ОПК-4. Способен оптимальным образом комбинировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности.

ОПК-5. Способен устанавливать и сопровождать программное обеспечение информационных систем, осуществлять эффективное управление разработкой программных средств и проектов.

1.5. Соответствие видов деятельности профессиональным компетенциям и их составляющих

Профессиональные компетенции	Практический опыт	Умения	Знания
Разработка продуктов на основе встроенной аналитики больших данных (С/01.8)			
ОПК-1	Проведение научно-исследовательских и опытно-конструкторских работ по созданию вариантов технического воплощения концепции продукта, создание прототипа продукта на основе встроенной аналитики больших данных	Осуществлять математическое и информационное моделирование	Математическое моделирование
ОПК-2	Анализ инноваций в области информационных технологий; новых технических средств, методов и алгоритмов анализа больших данных; источников информации; технологий представления данных, методов предсказательной и предписывающей аналитики; существующих продуктов на основе встроенной аналитики больших данных	Осуществлять разработку программно-аппаратных компонентов и систем	Технические средства и среды сбора, хранения и обработки больших данных
ОПК-3	Анализ требований к продукту, уточнение и доработка концепции, бизнес-модели и бизнес-плана создания нового продукта на основе встроенной аналитики больших данных	Проводить аналитические и поисковые исследования по тематике информационных технологий, технологий больших данных	Существующие и перспективные методы и программный инструментарий технологий больших данных
ОПК-4	Создание и оценка концепции и бизнес-модели продукта на основе встроенной аналитики больших данных	Проводить аналитические работы на основе технологий больших данных	Принципы и методы управления защитой и обеспечением конфиденциальности больших данных
ОПК-5	Разработка эксплуатационной документации по продукту на основе встроенной аналитики больших данных	Осуществлять разработку программно-аппаратных компонентов и систем	Современные и перспективные методы сбора, хранения и передачи данных из гетерогенных источников

2. УЧЕБНЫЙ ПЛАН

2.1. Категория слушателей РП

Имеющаяся квалификация (требования к слушателям) – к освоению ДПП допускаются лица имеющие высшее образование.

2.2. Общая трудоёмкость программы, аудиторная и самостоятельная работа

Общая трудоёмкость программы 168 академических часов, из них 98 академических часов аудиторной работы, 68 академических часов самостоятельной работы и 2 академических часа промежуточной аттестации.

2.3. Форма обучения

Форма обучения по РП – очная с применением дистанционных образовательных технологий.

2.4. Учебный план

РП «Базовый модуль» реализуется одним модулем.

№ п/п	Наименование темы, модуля	Форма Контроля	Всего, час	В том числе			
				Лекции	Практические занятия	Самостоятельная работа	Промежут. аттестация
1.	Введение в большие данные. Технологии ИИ. Машинное обучение	Устный опрос	10	2	4	4	-
2.	Системы хранения и обработки данных. Базы данных. SQL	Кейс	16	4	4	8	-
3.	Программные компоненты извлечения, хранения и подготовки больших данных. Словари и эталонная архитектура больших данных	Кейс	16	2	6	8	-
4.	Обработка, удаленная, распределенная и объединенная аналитика больших данных. Управление качеством и достоверностью больших данных	Кейс	18	2	6	10	-
5.	Языки программирования, среды исполнения и основные фреймворки	Кейс	24	2	14	8	-

	машинного обучения						
6.	Пайплайн машинного обучения. Препроцессинг данных. Работа с размерностью и структурой данных	Кейс	12	2	4	6	-
7.	Обучение без учителя. Кластеризация. Классификация. Регрессия. Метрики. Оценка качества алгоритма	Кейс	28	10	14	4	-
8.	Искусственные нейронные сети	Кейс	16	4	4	8	-
9.	Ансамбли алгоритмов. Повышение точности	Кейс	16	2	6	8	-
10.	Рекомендательные системы	Кейс	10	2	4	4	-
11.	Промежуточная аттестация	Зачет	2	-	-	-	2
	ИТОГО	-	168	32	66	68	2

3. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК

№ п/п	Наименование темы, модуля	1 неделя	2 неделя	3 неделя	4 неделя
1	Введение в большие данные. Технологии ИИ. Машинное обучение				
2	Системы хранения и обработки данных. Базы данных. SQL				
3	Программные компоненты извлечения, хранения и подготовки больших данных. Словари и эталонная архитектура больших данных				
4	Обработка, удаленная, распределенная и объединенная аналитика больших данных. Управление качеством и достоверностью больших данных				
5	Языки программирования, среды исполнения и основные фреймворки машинного обучения				
6	Пайплайн машинного обучения. Препроцессинг данных. Работа с размерностью и структурой данных				
7	Обучение без учителя. Кластеризация. Классификация. Регрессия. Метрики. Оценка качества алгоритма				
8	Искусственные нейронные сети				
9	Ансамбли алгоритмов. Повышение точности				
10	Рекомендательные системы				
11	Промежуточная аттестация				Зачет

Минимальный срок освоения программы – 4 недели.

4. РАБОЧАЯ ПРОГРАММА

4.1. Рабочая программа модуля «Базовый модуль»

4.1.1. Цель изучения модуля: подготовить высококвалифицированных и конкурентоспособных специалистов в области профессиональной деятельности, которая включает интеллектуальные системы, биоинформатику, когнитивные информационные технологии, вычислительные технологии, компьютерные науки, технологии баз данных, компьютерную графику, теорию информации, технологии управления инфокоммуникацией и бизнес-процессами, архитектуру программного обеспечения, параллельное и распределенное программирование.

Сформировать у слушателей следующие компетенции в области профессиональной деятельности: создание информационных технологий нового поколения, обеспечивающих экономически эффективное извлечение полезной информации из больших объемов разнообразных данных путем высокой скорости их сбора, обработки и анализа, и применение этих технологий в информационно-аналитической деятельности, в системах управления и принятия решений, а также для разработки на их основе новых продуктов и услуг.

4.1.2. Задачи изучения модуля:

1. Сформировать у слушателей основы инженерии данных.
2. Освоение навыков проектирования архитектур, моделей, баз и конвейеров данных.

4.1.3. Планируемые результаты обучения

Процесс изучения раздела направлен на формирование следующих компетенций

Код компетенции	Перечень планируемых результатов обучения по модулю	Формы и методы обучения, способствующие формированию и развитию компетенции
ОПК-1	Знать: математическое моделирование. Уметь: осуществлять математическое и информационное моделирование. Владеть: проведение научно-исследовательских и опытно-конструкторских работ по созданию вариантов технического воплощения концепции продукта, создание прототипа продукта на основе встроенной аналитики больших данных.	Формы обучения: Фронтальная. Методы обучения: Лекция; Практические Работы; Самостоятельная работа.
ОПК-2	Знать: технические средства и среды сбора, хранения и обработки больших данных. Уметь: осуществлять разработку программно-аппаратных компонентов и систем Владеть: анализ инноваций в области информационных технологий; новых технических средств, методов и алгоритмов анализа больших данных; источников информации; технологий представления данных,	Формы обучения: Фронтальная. Методы обучения: Лекция; Практические Работы; Самостоятельная работа.

	методов предсказательной и предписывающей аналитики; существующих продуктов на основе встроенной аналитики больших данных	
ОПК-3	Знать: существующие и перспективные методы и программный инструментарий технологий больших данных. Уметь: проводить аналитические и поисковые исследования по тематике информационных технологий, технологий больших данных. Владеть: анализ требований к продукту, уточнение и доработка концепции, бизнес-модели и бизнес-плана создания нового продукта на основе встроенной аналитики больших данных.	Формы обучения: Фронтальная. Методы обучения: Лекция; Практические Работы; Самостоятельная работа.
ОПК-4	Знать: принципы и методы управления защитой и обеспечением конфиденциальности больших данных. Уметь: проводить аналитические работы на основе технологий больших данных. Владеть: создание и оценка концепции и бизнес-модели продукта на основе встроенной аналитики больших данных.	Формы обучения: Фронтальная. Методы обучения: Лекция; Практические Работы; Самостоятельная работа.
ОПК-5	Знать: современные и перспективные методы сбора, хранения и передачи данных из гетерогенных источников. Уметь: осуществлять разработку программно-аппаратных компонентов и систем. Владеть: разработка эксплуатационной документации по продукту на основе встроенной аналитики больших данных.	Формы обучения: Фронтальная. Методы обучения: Лекция; Практические работы; Самостоятельная работа.

4.1.4 Содержание курса

Тема 1. Введение в большие данные. Технологии ИИ. Машинное обучение (10 часов)

Лекции (2 часа). Изучение тем:

- Введение в предмет.
- Источники данных, взаимодействие с носителем данных. Типы структурированности данных.
- Подготовка данных. Характеристики данных, корреляция.
- Введение в статистику. Описательная статистика. Формирование и проверка гипотез.
- Техническое задание и требования к проекту Data science.

Практическая работа (4 часа). Разбор практических кейсов по применению методов обработки данных, методов машинного обучения и искусственного интеллекта в промышленной аналитике и задачах бизнес-аналитики: разбор задач при обработке «сырых» данных, примеры источников данных.

Самостоятельная работа (4 часа). Проработка материала дополнительной литературы по теме.

Наименование темы	Дидактические единицы, вынесенные на самостоятельное изучение	Формы самостоятельной работы	Учебно-методическое обеспечение	Форма контроля
Введение в большие данные. Технологии ИИ. Машинное обучение	Данные, обучение с учителем, обучение без учителя	Проработка дополнительной литературы	Алпайдин Э. Машинное обучение: новый искусственный интеллект: пер. с англ. / Алпайдин Э. - М.: Фонд Развития Промышленности: Издательская группа "Точка": Альпина Паблшер: [Интеллектуальная Литература], 2017. - 191 с. - (Завтра это будут знать все). - Библиогр.: с. 185-191	Устный опрос

Тема 2. Системы хранения и обработки данных. Базы данных. SQL (16 часов)

Лекции (4 часа). Изучение тем:

1. Системы хранения данных, базы данных
 - Системы хранения данных и их виды
 - Hadoop Distributed File System (HDFS)
 - Simple Storage Service (S3)
 - Key value storage (Key value DataBase)
 - Компоненты баз данных
 - Реляционные базы данных
 - Не реляционные базы данных
2. Реляционные базы данных и их моделирование
 - Свойства реляционных баз данных
 - Моделирование данных
 - Схема «сущность-связь» ERD
 - Схемы данных, сущности, атрибуты, связи
 - Связи «один ко многим», «многие к одному», «многие ко многим»
 - Витрины данных
 - Сравнение витрин данных и хранилищ данных
 - Бизнес-аналитика (Business intelligence, BI)
3. Функциональность PGAdmin, работа с PostgreSQL

- Назначение PGAdmin
- Системные требования
- Требования к установке
- Доступные варианты установки
- Функционал и возможности PGAdmin 4
- Установка и развертывание базы данных
- 4. Язык структурированных запросов SQL. Синтаксис
 - Общие понятия
 - Типы запросов для организации работы СУБД
 - Типы операторов
 - Преимущества и недостатки языка SQL
 - Примеры SQL-кода.

Практическая работа (4 часа).

Решение практических кейсов:

1. Установка и настройка PostgreSQL. Основные шаги:
 - a. Загрузка установщика с веб-сайта EDB.
 - b. Установка PostgreSQL (в варианте для Windows).
 - c. Запуск оболочки PostgreSQL.
2. Выполнение мета-команд PostgreSQL. Основные шаги:
 - a. Список всех баз данных.
 - b. Подключение к базе данных PostgreSQL.
 - c. Получение списка всех таблиц в базе данных.
 - d. Получение списка пользователей и ролей.
3. Выполнение SQL-операторов.
 - a. Создание новой базы данных.
 - b. Создание таблиц.
 - c. Вставка данных в таблицы PostgreSQL.
 - d. Выборка данных из SQL таблицы.
 - e. Обновление данных в PostgreSQL.
 - f. Удаление данных из SQL таблицы.
 - g. Удаление таблиц в PostgreSQL.
 - h. Удаление баз данных PostgreSQL.
4. Манипуляции с базой данных в PostgreSQL.
 - a. Создание новых ролей и пользователей.
 - b. Создание баз данных.
5. Манипуляции с таблицами.

- a. Работа с типами данных.
- b. Задание значений по умолчанию.
- c. Работа с ограничениями.
- d. Работа с ключами таблиц.
- e. Работа с внешними ключами.
- f. Ограничения внешних ключей, поддержание ссылочной целостности.

Самостоятельная работа (8 часа). Решение практического кейса.

Тема 3. Программные компоненты извлечения, хранения и подготовки больших данных.

Словари и эталонная архитектура больших данных (16 часов)

Лекции (2 часа). Изучение тем:

- Введение в анализ данных
- Введение в интеллектуальный анализ данных: основные понятия, области применения современных технологий обработки и интеллектуального анализа больших данных.
- Этапы анализа данных. Структурированные и неструктурированные данные. Сбор и подготовка данных.
- Большие данные. Организация сбора и хранения больших наборов данных. Hadoop.
- Восстановление пропущенных значений в массивах данных.
- Очистка, интеграция и преобразование данных
- Библиотека pandas.
- Структуры данных в pandas, работа со структурами данных.
- Операции над данными.
- Комбинирование данных из разных источников.
- Обработка пропущенных значений
- Программные модули и пакеты для работы с многомерными массивами данных. Визуализация данных.
- Библиотеки NumPy, SciPy: основные функции.
- Визуализация данных с matplotlib и pandas.

Практическая работа (6 часов).

Решение практических кейсов:

1. Работа с большими данными при помощи библиотеки Pandas.
 - a. Обзор данных.
 - b. Фильтрация и сортировка данных.
 - c. Группировки данных.
 - d. Функция Apply.
 - e. Комбинирование нескольких массивов данных.

- f. Оценка статистики набора данных.
 - g. Визуализация данных.
 - h. Создание фрейма данных.
 - i. Обработка временных рядов.
 - j. Удаление данных.
 - 2. Работа с большими данными с использованием библиотек NumPy и SciPy.
 - a. Импорт модуля NumPy.
 - b. Другие способы создания массивов.
 - c. Математика массивов.
 - d. Итерация массива.
 - e. Основные операции с массивом.
 - f. Операторы сравнения и проверка значений.
 - g. Выбор элемента массива и манипулирование им.
 - h. Векторная и матричная математика.
 - i. Полиномиальная математика.
 - j. Статистика.
 - k. Случайные числа.
 - l. Другие функции.
 - m. Модули, доступные в SciPy.
 - 3. Визуализация данных matplotlib.
 - a. Рисунки и подграфики.
 - b. Цвет, маркеры и стили линий.
 - c. Подписи к осям, масштаб и легенда.
 - d. Сохранение рисунков в файл.
 - 4. Построение графиков с помощью pandas и seaborn.
 - a. Линейные графики.
 - b. Столбчатые диаграммы.
 - c. Гистограммы и графики плотности распределения.
 - d. Диаграммы рассеяния или точечные графики.
 - e. Категориальные данные.
- Самостоятельная работа (8 часа). Решение практического кейса.

Тема 4. Обработка, удаленная, распределенная и объединенная аналитика больших данных. Управление качеством и достоверностью больших данных (18 часов)

Лекции (2 часа). Изучение тем:

- Обработка, удаленная, распределенная и объединенная аналитика больших данных

- Введение в анализ больших данных. Как алгоритмы для больших данных отличаются от обычных?
- Базы данных и СУБД. SQL и NoSQL.
- Модель MapReduce. Поток данных.
- Основы систем Hadoop, Spark и других.
- Алгоритмы на больших данных: Кластеризация, понижение размерности, популярные предметные наборы и ассоциативные правила.
- Алгоритмы на больших данных: Рекомендательные системы и интернет-реклама
- Алгоритмы на больших данных: анализ и обработка данных из социальных сетей
- Применения алгоритмов обработки больших данных в задачах принятия решений.
- Архитектура систем обработки больших данных
- Управление качеством и достоверностью больших данных
- Требования к качеству больших данных, понятие Data Quality
- Метрики качества больших данных
- Процесс управления данными (Data Management)
- Пирамида процессов Data Management.

Практическая работа (6 часов).

Разбор практических кейсов:

1. Параллельные вычисления больших данных с использованием модели MapReduce.
 - a. Разбиение кода на независимо выполняемые фрагменты.
 - b. Распараллеливание выполнения.
 - c. Анализ эффективности архитектуры MapReduce на больших данных.
2. Реализация кластеризации данных по методу K средних с использованием Map-Reduction.
 - a. Просмотр и очистка данных.
 - b. Фильтрация данных.
 - c. Загрузка файлов в hdfs.
 - d. Подготовка, обработка текста и вычисление вектора признаков после хеширования.
 - e. итеративный расчет новых средних векторов (центроидов каждого кластера).
 - f. отнесение записей к кластерам в соответствии с текущими значениями итерационных центроидов.
 - g. Просмотр результатов, анализ быстродействия алгоритма.
3. Создание системы рекомендаций с помощью Hadoop.
 - a. Подключение к данным потребительской активности.
 - b. Группировка по пользователю, расчёт объединенного списка всех элементов и получение матрицы рейтинга для пользователей.

- c. Подсчёт списка комбинаций предметов и построение матрицы совместного появления предметов.
 - d. Объединение матрицы совместной встречаемости и матрицы баллов.
 - e. Расчёт списка рекомендуемых результатов.
 - 4. Оценка качества данных с применением библиотек Python.
 - a. Оценка качества данных. Метрики качества.
 - b. Цикл оценки датасета с помощью библиотеки GE (Great Expectations).
 - c. Проверка статистических параметров датасета с помощью объекта Schema библиотеки Pandera.
 - d. Подготовка и выполнение модульных тестов для данных с применением библиотек Deequ/PyDeequ.
- Самостоятельная работа (10 часов). Решение практического кейса.

Тема 5. Языки программирования, среды исполнения и основные фреймворки машинного обучения (24 часа)

Лекции (2 часа). Изучение тем:

- Языки программирования и среды разработки, которые используются в задачах машинного обучения. Фреймворки для машинного обучения.
- Введение в язык программирования Python, основные среды исполнения (IDE). Типы данных в Python. Базовые конструкции языка: условные конструкции, циклы.
- Объявление и вызов функций. Функции без параметров, с параметрами, с параметрами по умолчанию. Объявление классов и создание объектов в языке программирования Python. Работа с текстовыми файлами.
- Запуск командной строки. Основные команды при работе с командной строкой, работа с переменными операционной системы.
- Основные библиотеки при работе с данными: Numpy, Pandas, Matplotlib, Scipy. Срезы данных, агрегация данных, загрузка и запись данных, оптимизация в Python и основные графики.

Практическая работа (14 часов).

Решение практического кейса: «Визуализация данных с использованием основных библиотек языка программирования Python в среде Jupyter Notebook». В работе необходимо загрузить данные из csv файла, посмотреть распределения данных в каждой колонке с помощью гистограмм. Кроме этого, рассчитать описательную статистику данных: оценка математического ожидания, расчет стандартного отклонения, медианы. Визуализировать связи между колонками данных с помощью парных графиков.

Самостоятельная работа (8 часов). Решение практического кейса.

Тема 6. Пайплайн машинного обучения. Препроцессинг данных. Работа с размерностью и структурой данных (12 часов)

Лекции (2 часа). Изучение тем:

- Основные шаги при разработке систем машинного обучения.
- Постановка задачи и анализ предметной области.
- Разработка и анализ требований, составление спецификаций.
- Проектирование архитектуры приложения и моделей машинного обучения.
- Обучение и тестирование моделей.
- Разработка и тестирование приложения.
- Введение в эксплуатацию.
- Библиотека Scikit-learn.
- Знакомство с библиотекой: основные возможности и классы решаемых задач.
- Использование класса Pipeline для создания пайплайна обработки данных при решении задач машинного обучения.
- Виды задач машинного обучения. Обучение на размеченных и неразмеченных данных.
- Основы предобработки данных.
- Основные задачи подготовки наборов данных.
- Очистка данных и заполнение пропусков. Поиск выбросов.
- Разбиение выборки на обучающее, тестовое и валидационное множество.
- Методы балансировки наборов данных.
- Нормализация и стандартизация.
- Обработка категориальных признаков.
- Проблема снижения размерности данных. Метод главных компонент (Principal Component Analysis).

Практическая работа (4 часа).

Решение практических кейсов:

1. «Применение методов препроцессинга данных для подготовки датасета к решению задач машинного обучения».

Практический кейс представляет собой написание программы на языке программирования Python в среде разработки Jupyter Notebook. Основные шаги:

- a. Чтение данных, визуализация.
- b. Поиск некорректных значений и очистка данных.
- c. Исправление пропусков и выбросов.
- d. Обработка категориальных признаков.
- e. Разбиение выборки на обучающее, тестовое и валидационное множество.

2. «Создание пайплайна обработки данных для проекта по машинному обучению с использованием библиотеки Sklearn».

Практический кейс представляет собой написание программы на языке программирования Python в среде разработки Jupyter Notebook. Основные шаги:

a. Установка библиотеки scikit-learn и знакомство с основными возможностями.

b. Создание объекта класса Pipeline, включающего в себя шаги (все или часть), рассмотренные в практическом кейсе №1.

c. Запуск пайплайна на предложенном наборе данных.

Самостоятельная работа (6 часов). Решение практического кейса.

Тема 7. Обучение без учителя. Кластеризация. Классификация. Регрессия. Метрики. Оценка качества алгоритма (28 часов)

Лекции (10 часов). Изучение тем:

- Проблема снижения размерности данных. Метод главных компонент (Principal Component Analysis) и стохастическое вложение соседей с t-распределением (t-SNE).
- Алгоритмы кластеризации. Метод k-means, c-means, hierarchical clustering. Выбор алгоритма кластеризации.
- Алгоритмы кластеризации. Поиск гиперпараметров алгоритма классификации с помощью поиска по сетке. Метод k-means, c-means, hierarchical clustering. Выбор алгоритма кластеризации. Метрики оценки качества алгоритмов.
- Задача классификации. Поиск гиперпараметров алгоритма восстановления регрессии с помощью случайного поиска. Оценка точности работы классификатора, матрица ошибок. Модель анализа ROC-AUC.
- Методы классификации. Метод ближайших соседей k-NN, метод машины опорных векторов (SVM), дерево решений, логистическая регрессия.
- Задача восстановления регрессии. Оценка точности работы модели: средняя квадратичная ошибка модели, средняя ошибка по модулю, относительная ошибка модели.
- Методы восстановления регрессии. Линейная регрессия, полиномиальная регрессия, метод k-ближайших соседей.

Практическая работа (14 часов).

Будут рассмотрены следующие темы:

Программный код для решения задачи снижения размерности. Выбор количество компонент в методе главных компонент;

Запуск основных методов решения задачи кластеризации. Подбор количества кластеров в методе k-means и в иерархической кластеризации, оценка качества полученной кластеризации;

Запуск, настройка и оценка качества классификации некоторых из методов классификации: k-NN, дерево решений, SVM и k-NN;

Запуск, настройка и оценка качества восстановленной регрессии из методов восстановления регрессии: k-NN, дерево решений, линейная регрессия и полиномиальная регрессия.

Самостоятельная работа (4 часа). Решение практического кейса.

Тема 8. Искусственные нейронные сети (16 часов)

Лекции (4 часа). Изучение тем:

- Базовые элементы ИНС. Функции базовых элементов.
- Синапс.
- Сумматор.
- Нелинейный преобразователь.
- Виды функций активации.
- Общая логика решение задачи нейросетью.
- Архитектуры и классификация ИНС.
- Топологии нейронных сетей.
- Статические и динамические сети.
- Сети прямого распространения и рекуррентные.
- Классификация сетей по функциям активации.
- Сложности в выборе архитектур нейросети.
- Алгоритмы обучения ИНС.
- Постановка задачи обучения с учителем. Формирование обучающей и тестовой выборки.
- Метод обратного распространения ошибки.
- Проблемы процесса обучения, использование валидационной выборки.
- Обучение без учителя.
- Реализация нейронных сетей в Keras.

Практическая работа (4 часа).

Решение практического кейса:

1. Реализации полносвязной ИНС-регрессора на языке Python с использованием фреймворка Keras.
 - a. Основные положения фреймворка Keras.
 - b. Виды моделей Keras.
 - c. Виды слоев Keras.

- d. Цикл работы с моделью в Keras.
- e. Разбор решения задач аппроксимации функций нейросетью в Keras.
- 2. Реализация полносвязной и сверточной ИНС-классификатор на языке Python с использованием фреймворка Keras.
 - a. Основные шаги создания и использования сверточной сети с помощью Keras.
 - b. Графики обучения сетей.
 - c. Сохранение и считывание сети в процессе работы.
 - d. Разбор решения задачи распознавания рукописных цифр (датасет MNIST) нейросетью в Keras.

Самостоятельная работа (8 часов). Решение практического кейса.

Тема 9. Ансамбли алгоритмов. Повышение точности (16 часов)

Лекции (2 часа). Изучение тем:

- Понятие ансамбля методов.
- Виды ансамблевых методов
- Стекинг.
- Структура модели. Понятие метамоделей (смесителя)
- Обработка выборки методом удерживаемого набора
- Бэггинг
- Общая идея обучения на случайных подвыборках
- Понятие бутстрап-выборки, алгоритмы формирования бутстрап-выборок.
- Алгоритм выработки итогового решения
- Бустинг.
- Алгоритм последовательного улучшения результата путем добавления новых моделей (бустинга)
- Преимущества и недостатки, ограничения метода
- Адаптивный бустинг (AdaBoost)
- Алгоритм повышения значимости ошибочно решенных примеров (адаптивного бустинга)
- Правила комбинирования (бэггинга) моделей при адаптивном бустинге
- Градиентный бустинг
- Градиентный алгоритм оптимизации весов при выполнении бустинга
- Преимущества и ограничения метода.

Практическая работа (6 часов).

Решение практического кейса:

1. Реализация ансамблевых регрессоров на языке Python с использованием библиотек `sklearn.model_selection` и `sklearn.ensemble`
 - a. Реализация регрессора методом стекинга
 - b. Реализация регрессора методом бэггинга
 - c. Реализация регрессора методом обычного бустинга
 - d. Реализация регрессора методом адаптивного бустинга
 - e. Реализация регрессора методом градиентного бустинга
 - f. Сравнение эффективности ансамблевых методов.
- Самостоятельная работа (8 часов). Решение практического кейса.

Тема 10. Рекомендательные системы (10 часов)

Лекции (2 часа). Изучение тем:

- Поиск ассоциативных правил.
- Постановка задачи поиска ассоциативных правил. Задача анализа рыночной корзины.
- Характеристики и значимость ассоциативных правил.
- Методы поиска ассоциативных правил.
- Рекомендательные системы и их типы.
- Коллаборативная фильтрация (*collaborative filtering*).
- Основанные на контенте (*content-based*).
- Основанные на знаниях (*knowledge-based*).
- Гибридные (*hybrid*). Виды задач машинного обучения. Обучение на размеченных и размеченных данных.
- Разработка рекомендательных систем.
- Алгоритмы, модели и методы, используемые для создания рекомендательных систем.
- Метрики качества.
- Применение рекомендательных систем для решения различных задач.

Практическая работа (4 часа).

Решение практического кейса:

1. «Разработка рекомендательной системы»

Практический кейс представляет собой написание программы на языке программирования Python в среде разработки Jupyter Notebook. Основные шаги:

- a. Постановка задачи.
- b. Чтение данных, визуализация.
- c. Проектирование и разработка моделей для получения рекомендаций различного типа.
- d. Тестирование и выбор наилучшей модели.

Самостоятельная работа (4 часа). Решение практического кейса.

4.1.5. Оценочное средство для текущего контроля (примерные формулировки практических заданий):

Тема 1. Вопросы устного опроса.

1. Какие источники данных вам известны?
2. Как необходимо выполнять подготовку данных?
3. Приведите характеристики данных.

Тема 2. Формулировка практического кейса.

Рассматривается практический кейс «Разработка структуры базы данных для хранения диагностической информации болезни Паркинсона»

Описание: Решение практического кейса предполагает получение обучающимися описание целевой задачи диагностирования болезни Паркинсона, наборов исходной диагностической информации. По описанию наборов данных необходимо обучающимся необходимо сформировать структуру сущностей и их параметрическое описание, которые далее должны быть имплементированы в виде логической и физической модели базы данных. Следующим этапом является развертывание базы данных в соответствии с разработанными моделями с инициализацией таблиц и связей между ними. После этого осуществляется формирование базовых запросов для наполнения, поиска и получения информации из базы данных, а также запросов на редактирование, администрирование пользователей и резервное копирование базы данных.

Гипотезы задачи: Разработка структуры базы данных и ее развертывание позволят сформировать эффективный инструмент для формирования информационного обеспечения системы интеллектуальной обработки медицинской информации для построения моделей поддержки принятия решений.

Актуальность: Разработка корректных и эффективных структур баз данных для размещения, поиска, хранения и выдачи информации целевой практической задачи является важным этапом решения практических задач на основе дата-ориентированных технологий, таких как технологии машинного обучения и технологии искусственного интеллекта.

Тема 3. Формулировка практического кейса.

Рассматривается практический кейс «Работа с разнотипными источниками данных диагностической информации, разработка структуры агрегированных данных, размещение в хранилище данных и обеспечение доступ к ним»

Описание: Решение практического кейса предполагает формирование озера данных («озера сырых данных») полученных из существенно различающихся каналов сбора информации. Для рассматриваемого в рамках практического кейса набора дата-сетов - это изображения траекторий движения пера, фиксируемые при попытках диагностируемого изобразить фигуры (линии) требуемой формы с использованием графического планшета, регистрирующего координаты пера, силу нажатия, продолжительность и угол нажатия; записи, сделанные при произношении диагностируемым лицом слов и фраз с фиксацией параметров акустического ряда. Данные должны быть загружены в базу данных озера данных с использованием сформированных запросов. Должны быть проверены на корректность поисковые запросы, запросы на редактирование и выгрузку из озера данных требуемых срезов (стратификаций).

Гипотезы задачи: реализация методов формирования наборов данных для последующей обработки может быть выполнена на основе концепции data fusion - смешения данных из разных источников, полученных с использованием существенно различающихся процедур оценки параметрического состояния объектов аналитического исследования.

Актуальность: рациональное проектирование наборов данных может быть выполнено на основе data fusion - смешения данных из разных источников и реализовано средствами баз данных под управлением современных СУБД.

Тема 4. Формулировка практического кейса.

Рассматривается практический кейс «Верификация и fusion аналитика данных медицинской диагностики, полученных из различных источников»

Описание: Решение практического кейса предполагает выполнение манипуляций с наборами данных по диагностированию наличие болезни Паркинсона, направленных на формирование унифицированных представлений паттернов обучения с объединением (слиянием) источников данных и информационных таблиц. Рассматривается также вопрос агрегирования данных, структуризации данных для их последующей комбинированной стратификации и визуализации. Выполняется оценка прироста информации в наборе данных при использовании дополнительных факторов в диагностическом исследовании.

Гипотезы задачи: Распределенная и объединенная аналитика данных позволяет сформировать наборы данных комплексно описывающих предметную область анализируемой задачи.

Актуальность: рациональное проектирование наборов данных может быть выполнено на основе data fusion - смешения данных из разных источников и реализовано средствами баз данных под управлением современных СУБД.

Тема 5. Формулировка практического кейса.

Рассматривается практический кейс «Разработка программного модуля предобработки и визуализации больших данных различного типа»

Описание: Решение практического кейса предполагает разработку на языке Python программного кода, исполнение которого предполагает выгрузку из базы данных необходимого количества записей датасетов, формирование различных вариантов визуализации исходных данных, в том числе, в форме диаграмм рассеяния (парных графиков) и срезов с учетом различных стратификаций. Также выполняется простая статистическая обработка данных с визуализацией гистограмм по атрибутам и в привязке к информации о выходных параметрах, характеризующих наличие или отсутствие болезни Паркинсона у диагностируемого. Рассчитываются статистические характеристики наблюдений признаков объектов.

Гипотезы задачи: Гибкая инструменты визуализации данных позволяют идентифицировать и классифицировать задачу интеллектуальной обработки данных, сформулировать предположения о предпочтительных к использованию методах и моделях машинного обучения.

Актуальность: Разработка инструментов предобработки и визуализации данных позволяет сформировать интерактивные инструменты формирования данных для построения моделей машинного обучения, применения технологий искусственного интеллекта, позволяет провести предварительный экспертный анализ данных для рационального выбора типов таких технологий, идентификации и классификации задачи анализа данных.

Тема 6. Формулировка практического кейса.

Рассматривается практический кейс «Сокращение размерности и определение главных и составных признаков на основе сырых наборов данных медицинского диагностирования»

Описание: Решение практического кейса предполагает следующие операции препроцессинга данных: основные задачи подготовки наборов данных, очистка данных и заполнение пропусков, поиск выбросов, разбиение выборки на обучающее, тестовое и валидационное множество, балансировка набора данных. Операции препроцессинга выполняются над множеством структурированных данных диагностической задачи, связанной с определением наличия болезни Паркинсона у диагностируемых. Также предполагается проверка признаков на значимость, оценка снижения количества информации выборке при удалении наименее значимых признаков. отдельным этапом работы с данными является построение и анализ вариантов сокращения размерности проецирующими методами: методом главных компонент, методом факторного анализа и нелинейными проекционными методами.

Гипотезы задачи: Размерность исходного набора данных задачи может быть сокращено без существенной потери точности путем отсеивания малозначимых признаков либо преобразованием исходного набора признаков в пространство меньшей размерности без значимой потери точности решения целевой задачи анализа данных и ухудшения точности моделей машинного обучения.

Актуальность: Формирование рациональной структуры данных, сокращение размерности и формирование составных признаков (компонент) позволяет получить более компактные обучающие множества для дальнейшего применения методов машинного обучения и интеллектуальной обработки данных.

Тема 7. Формулировка практического кейса.

Рассматривается практический кейс «Подготовка данных для кластеризации, классификации и регрессионного анализа в целях определения и прогнозирования развития болезни Паркинсона»

Описание: Решение практического кейса предполагает рассмотрение набора данных по диагностированию наличия болезни Паркинсона у диагностируемых в аспекте решения базовых задач анализа данных: кластеризации, классификации и регрессии. Набор данных для задачи кластеризации представляется неразмеченным, применяются методы кластеризации, анализируется сравнительное качество кластеризации по метрике. Сравняется полученная автоматическая разметка с разметкой набора данных. Процесс повторяется при условии сокращения размерности ранее рассмотренными методами, сравнивается качество кластеризации, как характеристика делимости исходного множества. По аналогичной схеме, но в условиях обучения с учителем, решаются задачи регрессии и классификации. Рассматривается ситуация случайного перераспределения исходного набора на обучающее и тестовое подмножества. Оценивается статистическая устойчивость результатов для сравнения альтернативных подходов машинного обучения.

Гипотезы задачи: Для построения классификатора, регрессора и выполнения кластеризации на наборе медицинской диагностики могут быть использованы методы машинного обучения и интеллектуального анализа данных, в том числе, нейронные сети с получением приемлемых оценок точности получаемых моделей решения задач.

Актуальность: Построение приемлемо-точных моделей машинного обучения, оценка качества которых подтверждена статистически, является существенным основанием для формирования алгоритмического и модельного обеспечения системы поддержки принятия решений.

Тема 8. Формулировка практического кейса.

Рассматривается практический кейс «Применения искусственных нейронных сетей для обработки данных для кластеризации информации медицинского назначения»

Описание: В ходе решения практического кейса обучающиеся на практике знакомятся с процедурами подготовки данных для построения нейросетевых моделей и классификатор, позволяющих спрогнозировать и выявить у диагностируемого наличие болезни Паркинсона, в том числе на ранних стадиях. Выполняется подготовка данных для построения нейросетевых моделей, включая нормализацию данных, кодирование (различного типа) для категориальных данных. Далее в автоматизированном режиме выполняется построение решателей на основе искусственных нейронных сетей различного типа, рассматриваются результаты обучения нейросетевых моделей с использованием различных структур слоев и обучающих алгоритмов.

Гипотезы задачи: Искусственные нейронные сети являются эффективным инструментом обработки данных различного типа, однако для их использования требуется предобработка данных, для некоторых типов необходимо преобразование данных к числовым значениям, а также требуется нормализация данных.

Актуальность: Искусственные нейронные сети, фактически, являются сквозной технологией, широко применяемой для анализа диагностических данных, в том числе, в медицинской диагностике. Общность аппарата позволяет экстраполировать результаты кейса на многие задачи интеллектуального анализа данных.

Тема 9. Формулировка практического кейса.

Рассматривается практический кейс «Повышение точности обработки медицинской диагностической информации на основе ансамблирования моделей данных и комбинации источников данных»

Описание: В ходе решения практического кейса в условиях практического применения рассматриваются методы ансамблирования: boosting, bagging, stacking для построения регрессоров и классификаторов для набора данных, содержащих информацию о диагностических признаках болезни Паркинсона. Рассматриваются в автоматизированном и ручном режиме варианты разбиения, переразбиения и подготовки данных для различных схем ансамблирования. Проводится статистическое исследование и сравнительный анализ эффективности нескольких методов построения ансамблей.

Гипотезы задачи: Точность решения задач классификации и регрессии для выявления и прогнозирования наличия болезни Паркинсона может быть повышена при использовании ансамблей (коллективов) моделей и/или классификаторов, сформированных специальным образом и на наборах данных, подготовленных для построения таких множеств.

Актуальность: Ансамблирование моделей и алгоритмов является эффективным инструментом повышения точности базовых методов машинного обучения и моделей искусственного интеллекта, однако требует специфической процедуры подготовки данных и построения классификаторов/регрессоров.

Тема 10. Формулировка практического кейса.

Рассматривается практический кейс «Проработка проекта формирования информационного обеспечения системы рекомендательной системы»

Описание: В ходе выполнения практического кейса проектируется схема информационного обеспечения рекомендательной медицинской системы для диагностирования и обнаружения признаков болезни Паркинсона. Разрабатывается схема интерфейсных модулей с предполагаемым типовым программным обеспечением, определяется модульный состав подключения и работы с выбранной СУБД. Элементы выполненных проектов реализуются на языке Python.

Актуальность: рекомендательные системы и системы поддержки принятия решений являются одним из наиболее востребованных практикоориентированных решений, формируемых в настоящее время на основе методов машинного обучения и технологий искусственного интеллекта.

5. УСЛОВИЯ РЕАЛИЗАЦИИ РП

5.1. Организационные условия реализации РП

Наименование аудитории	Вид занятия	Наименование оборудования, программного обеспечения
Аудитория для проведения лекций/семинаров	Лекции	ПК с доступом в Интернет и возможностью просмотра файлов в формате *.ppt, *.pptx, *.pdf; микрофон; колонки/наушники; камер.
Лаборатории/ производственные помещения	Практические занятия	Личный ПК/смартфон с доступом в Интернет и возможностью просмотра файлов в формате *.html, *.doc, *.docx, *.pdf; лист формата А5/А4 или блокнот; карандаш/ручка.
Коворкинги, учебные залы и т.д.	Самостоятельная работа	Личный ПК/смартфон с доступом в Интернет и возможностью просмотра файлов в формате *.html, *.doc, *.docx, *.pdf; лист формата А5/А4 или блокнот; карандаш/ручка.
Аудитория для проведения лекций/семинаров	Итоговая аттестация	ПК с доступом в Интернет и возможностью просмотра файлов в формате *.html, *.doc, *.docx, *.pdf, *.djvu, лист бумаги формата А4, ручка.

5.2. Педагогические условия реализации РП

Реализация программы обеспечивается преподавательским составом, удовлетворяющим следующим условиям:

- наличие высшего профессионального образования, соответствующее профилю программы, из числа штатных преподавателей, или привлеченных на условиях почасовой оплаты труда;
- значительный опыт практической деятельности в соответствующей сфере из числа штатных преподавателей или привлеченных на условиях почасовой оплаты труда

5.3. Учебно-методическое обеспечение РП

Основная литература:

1. Алпайдин Э. Машинное обучение: новый искусственный интеллект: пер. с англ. / Алпайдин Э. - М.: Фонд Развития Промышленности: Издательская группа "Точка" : Альпина Паблишер : [Интеллектуальная Литература], 2017. - 191 с. - (Завтра это будут знать все). - Библиогр.: с. 185-191. - ISBN 978-5-9908700-8-6. - ISBN 978-5-9614-6114-5.
2. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных / Силен Д., Мейсман А., Али М. ; пер. с англ. Матвеев Е. - СПб. : Питер, 2020. - 334 с. : ил. - (Библиотека программиста). - ISBN 978-5-4461-0944-9.
3. Хайкин С. Нейронные сети. Полный курс: пер. с англ. / Хайкин С.; пер. Кусскуль Н. Н., Шелестов А. Ю. - 2-е изд. - М.; СПб.; Киев: Вильямс, 2006. - 1103 с.: ил. - Библиогр.: с. 996-1069. - ISBN 5-8459-0890-6.
4. Маркин, А. В. Программирование на SQL: учебное пособие для среднего профессионального образования / А. В. Маркин. — Москва: Издательство Юрайт, 2020. — 435 с. — (Профессиональное образование). — ISBN 978-5-534-11093-7. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/456926>
5. Маркин, А. В. Программирование на SQL в 2 ч. Часть 2: учебник и практикум для вузов / А. В. Маркин. — 2-е изд., испр. и доп. — Москва: Издательство Юрайт, 2020. — 340 с. — (Высшее образование). — ISBN 978-5-534-12258-9. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/451185>.

Дополнительные материалы:

1. Белоус В. В., Пивоварова Н. В. Основы реляционных баз данных. Практикум по SQL [Электрон. ресурс]: метод. указ. к лаб. работам по курсу "Базы данных" / Белоус В. В., Пивоварова Н. В.; МГТУ им. Н. Э. Баумана, Фак. "Робототехника и комплексная автоматизация". - М.: Изд-во МГТУ им. Н. Э. Баумана, 2014. - 1 CD-ROM. - ФГУП "Информрегистр" №0321400922.
2. Басараб М. А., Коннова Н. С. Интеллектуальные технологии на основе искусственных нейронных сетей: метод. указания к выполнению лаб. работ / Басараб М. А., Коннова Н. С.; МГТУ им. Н. Э. Баумана. - М.: Изд-во МГТУ им. Н. Э. Баумана, 2017. - 53 с.: ил. - Библиогр. в конце брош. - ISBN 978-5-7038-4716-9.
3. Березкин Д. В., Рожнев А. Ю. Разработка модели оценки платежеспособности клиентов банка с применением алгоритмов машинного обучения / Березкин Д. В., Рожнев А. Ю. // Динамика сложных систем. - 2018. - Т. 12, № 4. - С. 59-66.
4. Бизли Д. М. Язык программирования Python: справочник: Пер. с англ / Бизли Д. М. - Киев: ДиаСофт, 2000. - 326 с. - ISBN 966-7393-54-2.
5. Вьюгин В. В. Элементы математической теории машинного обучения: учеб. пособие для вузов / Вьюгин В. В.; Моск. физико-техн. ин-т (гос. ун-т), РАН. Ин-т проблем передачи информации им. А. А. Харкевича. - М.: МФТИ - ИППИ РАН, 2010. - 231 с. - Библиогр.: с. 229-231. - ISBN 978-5-7417-0339-7.

6. Галушкин А. И. Нейронные сети: основы теории / Галушкин А. И. - М.: Горячая линия - Телеком, 2010. - 496 с.: ил. - Библиогр. в конце ст., с. 469-488. - ISBN 978-5-9912-0082-0.
7. Головкин В. А. Нейронные сети: обучение, организация и применение: учеб. пособие по направлению подготовки бакалавров и магистров / Головкин В. А.; общ. ред. Галушкин А. И. - М.: Издат. предприятие ред. журн. "Радиотехника" (ИПРЖР), 2001. - 256 с.: ил. - (Нейрокомпьютеры и их применение; кн. 4). - Библиогр.: с. 248-256. - ISBN 5-93108-05-8.
8. Джеймс Р., Грофф П., Вайнберг Н., Оппель Э. Дж. SQL. Полное руководство. М: Вильямс, 2014. 960 с.
9. Коннолли Т., Бегг К. Базы данных. Проектирование, реализация и сопровождение. Теория и практика. М: Вильямс, 2017. 1440 с.
10. Комашинский В. И., Смирнов Д. А. Нейронные сети и их применение в системах управления и связи / Комашинский В. И., Смирнов Д. А. - М.: Горячая линия - Телеком, 2003. - 93 с. - Библиогр.: с. 88-93. - ISBN 5-93517-094-9.
11. Осовский С. Нейронные сети для обработки информации: пер. с польск. / Осовский С.; пер. Рудинский И. Д. - М.: Финансы и статистика, 2002. - 343 с.: ил. - ISBN 5-279-02567-4.
12. Сидняев Н. И., Храпов П. В. Нейросети и нейроматематика: учеб. пособие / Сидняев Н. И., Храпов П. В. ; МГТУ им. Н. Э. Баумана. - М.: Изд-во МГТУ им. Н. Э. Баумана, 2016. - 83 с.: ил. - Библиогр.: с. 82. - ISBN 978-5-7038-4362-8.
13. Хливненко Л. В., Пятакович Ф. А. Практика нейросетевого моделирования: учеб. пособие / Хливненко Л. В., Пятакович Ф. А. - СПб. Лань, 2019. - 196 с. - (Учебники для вузов. Специальная литература) (Бакалавриат и магистратура). - Библиогр.: с. 182-193. - ISBN 978-5-8114-3639-2.

5.4. Методические рекомендации

РП построена по тематическому принципу, каждый раздел представляет собой логически завершённый материал.

Преподавание программы основано на личностно-ориентированной технологии образования, сочетающей два равноправных аспекта этого процесса: обучение и учение. Личностно-ориентированный подход развивается при участии слушателей в активной работе на практических занятиях. Личностно-ориентированный подход направлен, в первую очередь, на развитие индивидуальных способностей обучающихся, создание условий для развития творческой активности слушателя и разработке инновационных идей, а также на развитие самостоятельности мышления при решении учебных задач разными способами, нахождение рационального варианта решения, сравнения и оценки нескольких вариантов их решения и т.п. Это способствует формированию приемов умственной деятельности по восприятию новой информации, ее запоминанию и осознанию, созданию образов для сложных понятий и процессов, приобретению навыков поиска решений в условиях неопределенности.

Практические занятия проводятся для приобретения навыков решения практических задач в предметной области модуля. Задания, выполняемые на практических занятиях, выполняются с использованием активных и интерактивных методов обучения.

Самостоятельная работа слушателей предназначена для проработки дополнительной литературы. Результаты практических заданий слушателей учитываются на итоговой аттестации.

При изучении курса предусмотрены следующие методы организации и осуществления учебно-познавательной деятельности:

- объяснительно-иллюстративный метод;
- репродуктивный метод;
- частично-поисковый метод.

6. ФОРМЫ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ РП

Итоговая аттестация проводится в форме зачета. Зачет проводится в формате тестирования и решения кейса. Результатом зачета служат правильные ответы на вопросы теста, состоящего из 10 (двадцати) вопросов.

По результатам итоговой аттестации слушателю выставляется оценка «ЗАЧТЕНО/НЕ ЗАЧТЕНО»:

Оценка «ЗАЧТЕНО» выставляется слушателю, который:

- правильно ответил не менее чем на 60% вопросов в тесте;
- практическое задание (кейс от индустриального партнера) выполнено полностью, нет ошибок в логических рассуждениях. Возможно наличие одной неточности или описки, не являющихся следствием незнания или непонимания учебного материала. Слушатель показал полный объем знаний, умений в освоении, пройденных тем и применение их на практике.
- продемонстрировал необходимые систематизированные знания и достаточную степень владения принципами предметной области программы, понимание их особенностей и взаимосвязь между ними в течение всего срока обучения по ДПП.

Оценка «НЕ ЗАЧТЕНО» ставятся слушателю, который:

- правильно ответил менее чем на 60% вопросов в тесте;
- практическое задание (кейс от индустриального партнера) выполнено не полностью. Допущены грубые ошибки. Задание выполнено не самостоятельно. Задание не сдано.
- имеет крайне слабые теоретические и практические знания, обнаруживает неспособность к построению самостоятельных заключений

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

7.1. Паспорт комплекта оценочных средств

Предметы оценивания	Объекты оценивания	Показатели оценки
ОПК-1. Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий	Тестовое задание	Количество правильных ответов
ОПК-2. Способен применять компьютерные/ суперкомпьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности	Тестовое задание	Количество правильных ответов
ОПК-3. Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования	Тестовое задание	Количество правильных ответов
ОПК-4. Способен оптимальным образом комбинировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности	Тестовое задание	Количество правильных ответов
ОПК-5. Способен устанавливать и сопровождать программное обеспечение информационных систем, осуществлять эффективное управление разработкой программных средств и проектов	Тестовое задание	Количество правильных ответов

7.2. Комплект оценочных средств

7.2.1. Темы для подготовки к зачету:

1. Большие данные.
2. Искусственный интеллект.
3. Системы хранения данных.
4. Программные компоненты извлечения, хранения и подготовки больших данных.
5. Языки программирования, среды исполнения и основные фреймворки машинного обучения.
6. Обучение без учителя.
7. Искусственные нейронные сети.
8. Ансамбли алгоритмов.
9. Повышение точности.
10. Рекомендательные системы.

7.2.2. Примерные тесты для проведения зачета:

- 1) Как называется свойство алгоритма, означающее, что он всегда приводит к результату через конечное, возможно, очень большое, число шагов?
- а) Понятность
 - б) Дискретность
 - в) Массовость
 - г) Результативность
- 2) Алгоритм называется линейным, если:
- а) Он предполагает многократное повторение одних и тех же действий
 - б) Ход его выполнения зависит от истинности тех или иных условий
 - в) Он представлен в табличной форме
 - г) Его команды выполняются в порядке следования друг за другом
- 3) Алгоритм включает в себя ветвление, если:
- а) Ход его выполнения зависит от истинности тех или иных условий
 - б) Его команды выполняются в порядке следования друг за другом
 - в) Его команды записаны в табличной форме
- 4) Наибольшей наглядностью обладают следующие формы записи алгоритмов:
- а) Словесные
 - б) Рекурсивные
 - в) Построчные
 - г) Графические
- 5) Для групповых операций с файлами используются маски имен файлов. Маска представляет собой последовательность букв, цифр и прочих допустимых в именах файлов символов, в которых также могут встречаться следующие символы: символ «?» (вопросительный знак) означает ровно один произвольный символ; символ «*» (звездочка) означает любую последовательность символов произвольной длины, в том числе «*» может задавать и пустую последовательность. В каталоге находится 6 файлов: adobe.xls; idol.xlsx; london.xls; adobe.xml; odor.xlsx; sdoba.xls. Определите, по какой из масок из них будет отобрана указанная группа файлов: adobe.xls; idol.xlsx; odor.xlsx; sdoba.xls.
- а) ?do*.xls

б) ?do?*.xls*

в) *do*.x*

г) ?do?.xls*

6) **Запишите десятичное число 100 в системе счисления с основанием 9.**

а) 121

б) 99

в) 139

г) 131

7) **Закон больших чисел утверждает, что:**

а) чем больше единиц охвачено статистическим наблюдением, тем лучше проявляется общая закономерность;

б) чем больше единиц охвачено статистическим наблюдением, тем хуже проявляется общая закономерность;

в) чем меньше единиц охвачено статистическим наблюдением, тем лучше проявляется общая закономерность.

8) **В статистике значение признака, находящееся в середине ранжированной (упорядоченной по возрастанию или убыванию) совокупности, называется:**

а) модой

б) медианой

в) средним выборочным

г) статистической нормой

9) **Среднеквадратическое отклонение характеризует:**

а) взаимосвязь данных;

б) разброс данных;

в) динамику данных.

10) **Термин корреляция в статистике понимается как:**

а) связь, зависимость;

б) отношение, соотношение;

в) функцию, уравнение.

7.2.3. Кейсы для итогового контроля.

Кейс №1. Машинное обучение для распознавания траекторий квантовых измерений и улучшения считывания данных.

Способность выполнять точные измерения важна для максимизации информации, которую можно извлечь из физической системы. Это особенно верно в отношении экспериментальной квантовой обработки информации, поскольку квантовые системы очень чувствительны к шумовым эффектам, а частота ошибок квантовых операций и измерений должна быть небольшой, чтобы отказоустойчивые квантовые вычисления стали реальностью. Цель — предоставить методы диагностики ошибок измерения и повышения точности с помощью алгоритмов классификации и кластеризации в машинном обучении.

Сверхпроводящие квантовые биты (кубиты) стали многообещающим кандидатом для создания отказоустойчивого квантового компьютера благодаря их длительному времени когерентности и высокой точности многокубитных операций. Заметный прогресс был достигнут в снижении количества ошибок при выполнении этих операций, однако еще предстоит проделать значительную работу для реализации отказоустойчивых квантовых вычислений в больших сетях кубитов. В схемной квантовой электродинамике (сQED) сверхпроводящий ангармонический осциллятор, такой как трансмон, связан с резонатором, производя зависящий от состояния сдвиг частоты резонатора. Это позволяет измерять кубиты, управляя резонатором и записывая выходную траекторию в фазовом пространстве. На практике существуют значительные источники случайного шума и систематических воздействий, таких как T1-процессы, когда система спонтанно переходит в свое основное состояние, из-за которых однократные траектории (сингл-шоты) могут показаться сложными и трудноразличимыми.

Таким образом, актуальность обусловлена необходимостью снижения ошибок логических операций кубитов (гейтов). Подобные ошибки являются следствием низких времен когерентности кубита, которые в свою очередь снижаются за счет влияния внешних шумовых эффектов, таких как тепловые эффекты или эффекты релаксации кубитов.

Имеются данные о состоянии кубитов, представляющие собой наборы сингл-шотов: множества последовательных точек – измерений (X,Y). Измерения проведены в двух состояниях кубитов – возбужденном и невозбужденном, что представлено в двух файлах.

Целью является модель, позволяющая разделять измерения сингл-шотов на группы в зависимости от уровня шума каждого отдельного измерения, что позволит фильтровать группы измерений и отсеивать наиболее зашумленные.

Необходимо выполнить кластерный анализ: сгруппировать траектории в естественные подмножества данных, отличающихся влиянием шума, что позволит диагностировать систематические ошибки.

Кейс основан на реальных производственных задачах компании ООО «БаумИнформ», применяющей технологии интеллектуальной обработки данных для решения своих производственных задач, что подтверждается соответствующим письмом (в приложении).

Кейс №2. Прогнозирование размеров сварного шва при электронно-лучевой сварке тонкостенных конструкций аэрокосмического назначения.

В качестве исходных данных были взяты результаты экспериментальных исследований, проводимых в целях улучшения технологического процесса электронно-лучевой сварки изделия, сборка которого состоит из элементов, состоящих из разнородного материала. Установка электронно-лучевой сварки, на которой проводились исследования, предназначена для сварки электронным лучом в глубоком вакууме деталей сборочных единиц из нержавеющей сталей, титановых, алюминиевых и специальных сплавов. Существующая установка электронно-лучевой сварки обеспечивает повторяемость режимов в рамках возможностей реализованной системы управления. Работы по сварке выполнялись на образцах-имитаторах, соответствующих технологическому изделию. Для уменьшения вложения энергии при сварке:

1. Снижалась величина сварочного тока (IW);
2. Увеличивался ток фокусировки электронного пучка (IF);
3. Увеличивалась скорость сварки (VW);
4. Менялось расстояние от поверхности образцов до электронно-оптической системы (FP).

По совокупности параметров технологических режимов обеспечивались минимально возможные размеры сварных швов: глубина шва ($Depth$) и ширина шва ($Width$).

В процессе выполнения работ была произведена электронно-лучевая сварка 18-ти единиц образцов. Результаты металлографического контроля по размерам сварочного шва для каждого образца проводились в 4-х поперечных сечениях сварочного шва. Ускоряющее напряжение было постоянным в диапазоне 19,8 – 20 кВ. Набор полученных данных собраны в составе режимов сварки, размеров сварочных швов в поперечных сечениях всех образцов.

Для процесса электронно-лучевой сварки формальная постановка задачи поддержки принятия технологических решений представляет собой задачу регрессии, в которой необходимо найти математическую зависимость между набором исходных параметров процесса ЭЛС и результатом процесса при этих параметрах.

Математическая постановка задачи в данном случае будет выглядеть следующим образом.

Пусть имеются набор параметров технологического процесса:

- 1) IW – величина сварочного тока;
- 2) IF – ток фокусировки электронного луча;
- 3) VW – скорость перемещения луча по поверхности изделия;

4) FP – расстояние от электронно-оптической системы до поверхности детали

Также имеется набор результатов технологического процесса:

1) depth – глубина сварного шва;

2) width – ширина сварного шва.

Необходимо построить такую модель машинного обучения которая обладает наилучшей мерой качества аппроксимации представленных данных.

Кейс основан на реальных производственных задачах компании ООО «Ниагара», применяющей технологии интеллектуальной обработки данных для решения своих производственных задач в области композиционных материалов, что подтверждается соответствующим письмом (в приложении).

Кейс №3. Прогнозирование конечных свойств новых материалов (композиционных материалов).

Композиционные материалы — это искусственно созданные материалы, состоящие из нескольких других с четкой границей между ними. Композиты обладают теми свойствами, которые не наблюдаются у компонентов по отдельности. При этом композиты являются монолитным материалом, т. е. компоненты материала неотделимы друг от друга без разрушения конструкции в целом. Яркий пример композита - железобетон. Бетон прекрасно сопротивляется сжатию, но плохо растяжению. Стальная арматура внутри бетона компенсирует его неспособность сопротивляться сжатию, формируя тем самым новые, уникальные свойства. Современные композиты изготавливаются из других материалов: полимеры, керамика, стеклянные и углеродные волокна, но данный принцип сохраняется. У такого подхода есть и недостаток: даже если мы знаем характеристики исходных компонентов, определить характеристики композита, состоящего из этих компонентов, достаточно проблематично. Для решения этой проблемы есть два пути: физические испытания образцов материалов, или прогнозирование характеристик. Суть прогнозирования заключается в симуляции представительного элемента объема композита, на основе данных о характеристиках входящих компонентов (связующего и армирующего компонента).

На входе имеются данные о начальных свойствах компонентов композиционных материалов (количество связующего, наполнителя, температурный режим отверждения и т.д.). На выходе необходимо спрогнозировать ряд конечных свойств получаемых композиционных материалов.

Актуальность: Созданные прогнозные модели помогут сократить количество проводимых испытаний, а также пополнить базу данных материалов возможными новыми характеристиками материалов, и цифровыми двойниками новых композитов.

Датасет со свойствами композитов. Объединение делать по индексу тип объединения INNER

Кейс основан на реальных производственных задачах компании ООО «Ниагара», применяющей технологии интеллектуальной обработки данных для решения своих производственных задач в области композиционных материалов, что подтверждается соответствующим письмом (в приложении).

Кейс №4. Классификация госконтрактов по предметам закупки.

Задача: необходимо на основе данных с ftp.zakupki.gov.ru научиться определять группу, к которой относится контракт с кодом ОКПД-2 41, 42, 43, 71.1.

Группы могут быть следующими:

1. Строительно-монтажные работы (СМР)
2. Проектно-изыскательские работы (ПИР)
3. Строительный надзор
4. Подключение коммуникаций
5. Прочее.

По ОКПД-2 контракты в общем случае должны разделяться так:

Строительно-монтажные работы (СМР) - 41, 42, 43 (кроме нижеперечисленных)

Проектно-изыскательские работы (ПИР) - 41.1, 71.1

Подключение коммуникаций - 43.22

Строительный надзор – четкой группы нет.

Проблема: Далеко не всегда контракты указываются с нужным кодом, поэтому есть проблема как такие контракты «отловить» и определить в нужную группу.

Поэтому задача предполагает классификацию контрактов на основе объекта закупки, который сформулирован естественным языком. Также предполагаем, что могут иметь значение цена контракта и его длительность.

На основе этого на входе данные о контрактах. На выходе необходимо получить группу для каждого контракта.

Иногда контракт может относиться одновременно в несколько групп.

Кейс основан на реальных производственных задачах компании ООО «СК-СХД», работающей в области искусственного интеллекта, разработки систем хранения больших данных, что подтверждается соответствующим письмом (в приложении).

Кейс №5. Анализ импорта/экспорта промышленных станков в Российскую Федерацию.

Исходные данные: статистика импорта/экспорта промышленных станков за период с 2017 по 2022 гг. Основные параметры: дата таможенных операций, код ТН ВЭД, наименование, производитель, страна происхождения, стоимость единицы, вес единицы, количество. В исходном датасете 66165 записей, представлены 223 кода ТН ВЭД, которые сгруппированы в

44 класса машин. Постановка задачи от заказчика: «Необходимо выделить обоснованные кластеры, на которые был спрос в течение 5 лет. По выделенным кластерам заказчик выделит станки-представители, для которых подберет типовые комплектующие и рассчитает количественную потребность в этих комплектующих с учетом объемов станков, импортированных согласно базе. Общая цель исследования – создать алгоритм для принятия решения по импортозамещению наиболее популярных кластеров станков».

Подход к исследованию

В связи с тем, что имеющаяся классификация (44 класса машин) не позволяет выделить обоснованные/информативные кластеры станков, а также в связи с тем, что код ТН ВЭД и стоимость могут быть определены некорректно, с заказчиком было согласовано использовать для кластеризации только 2 поля: наименование станка и вес единицы. С помощью методов анализа текста, необходимо провести процедуры токенизации (разбиение на отдельные слова), лемматизации (неопределенная форма для различных частей речи), и векторизация текста. При векторизации полезно использовать список стоп-слов, или шумовых слов: предлоги, союзы, междометия и т.п.

Непосредственно для кластеризации предлагается использовать несколько методов с различными гиперпараметрами. Для оценки качества кластеризации предлагается использовать метрики S_Dbw , коэффициент силуэта. Настройку гиперпараметров методов кластеризации произвести с использованием алгоритмов оптимизации гиперпараметров (GridSearch, RandomSearch, Bayesian optimization, HyperBand, Genetic algorithm и т.п.).

В качестве результата ожидается размеченный датасет, то есть исходный датасет с добавлением поля «Метка кластера».

Для интерпретации результатов кластеризации предлагается использовать процедуру Subgroup Discovery, чтобы найти правила/паттерны формирования кластеров в зависимости от наличия определенных слов в названии станка.

Наиболее мощные кластеры (первые 10 по популярности) необходимо проанализировать по признаку наличия определенных слов (ключевых слов) в названии станка, а также по весу единицы оборудования. Сделать вывод об информативности и специфичности полученных кластеров.

Кейс основан на реальных производственных задачах компании ООО «СК-СХД», работающей в области искусственного интеллекта, разработки систем хранения больших данных, что подтверждается соответствующим письмом (в приложении).

Министерство науки и высшего образования Российской Федерации
Федеральное государственное бюджетное образовательное учреждение высшего образования
«Московский государственный технический университет имени Н.Э. Баумана
(национальный исследовательский университет)»
(МГТУ им. Н.Э. Баумана)

РАБОЧАЯ ПРОГРАММА ДИСЦИПЛИНЫ
«Профессиональный модуль»
ДОПОЛНИТЕЛЬНОЙ ПРОГРАММЫ ПРОФЕССИОНАЛЬНОЙ
ПЕРЕПОДГОТОВКИ **«Инженер данных (Data engineer Pro)»**

Москва, 2023

АВТОР ПРОГРАММЫ:

Доцент каф. ИУ12

_____ В.С. Тынченко

СОГЛАСОВАНО:

Начальник УСП

_____ Т.А. Гузева

Директор
Центра дополнительного образования

_____ М.В. Стоянова

Оглавление

1. ОБЩАЯ ХАРАКТЕРИСТИКА РП.....	4
1.1. Цель	4
1.2. Планируемые результаты обучения	4
1.3. Дополнительные характеристики РП.....	5
1.4. Перечень профессиональных компетенций в рамках имеющейся квалификации, качественное изменение которых осуществляется в результате обучения.....	5
1.5. Соответствие видов деятельности профессиональным компетенциям и их составляющих.....	6
2. УЧЕБНЫЙ ПЛАН	7
2.1. Категория слушателей РП	7
2.2. Общая трудоёмкость программы, аудиторная и самостоятельная работа	7
2.3. Форма обучения	7
2.4. Учебный план	7
3. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК.....	7
4. РАБОЧАЯ ПРОГРАММА	8
5. УСЛОВИЯ РЕАЛИЗАЦИИ РП.....	15
5.1. Организационные условия реализации РП.....	15
5.2. Педагогические условия реализации РП	16
5.3. Учебно-методическое обеспечение РП.....	16
5.4. Методические рекомендации.....	17
6. ФОРМЫ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ РП.....	19
7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ.....	20
7.1. Паспорт комплекта оценочных средств.....	20
7.2. Комплект оценочных средств	20

1. ОБЩАЯ ХАРАКТЕРИСТИКА РП

Программа подготовлена на основе:

- Федерального закона от 29 декабря 2012 года № 273-ФЗ «Об образовании в Российской Федерации»;
- требований Приказа Минобрнауки России от 01.07.2013 года № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам»;
- методических рекомендаций-разъяснений Минобрнауки России по разработке дополнительных профессиональных программ на основе профессиональных стандартов от 22 апреля 2015 года № ВК-1030/06.

Реализация программы РП направлена на получение новой компетенции, необходимой для профессиональной деятельности.

1.1. Цель

Подготовить высококвалифицированных и конкурентоспособных специалистов в области профессиональной деятельности, которая включает интеллектуальные системы, биоинформатику, когнитивные информационные технологии, вычислительные технологии, компьютерные науки, технологии баз данных, компьютерную графику, теорию информации, технологии управления инфокоммуникацией и бизнес-процессами, архитектуру программного обеспечения, параллельное и распределенное программирование.

Сформировать у слушателей следующие компетенции в области профессиональной деятельности: создание информационных технологий нового поколения, обеспечивающих экономически эффективное извлечение полезной информации из больших объемов разнообразных данных путем высокой скорости их сбора, обработки и анализа, и применение этих технологий в информационно-аналитической деятельности, в системах управления и принятия решений, а также для разработки на их основе новых продуктов и услуг.

1.2. Планируемые результаты обучения

Планируемые результаты обучения по РП:

- освоение профессиональных компетенций в процессе изучения перечисленных тем в учебном плане;
- успешное освоение программы повышения квалификации;
- успешное прохождение промежуточной аттестации (зачет).

Обучающиеся, успешно прошедшие обучение, выполнившие текущие контрольные задания и выдержавшие предусмотренный учебным планом зачет по РП «**Профессиональный модуль**», могут быть допущены к освоению следующего - второго уровня дополнительной программы профессиональной переподготовки «Инженер данных (Data engineer Pro)».

1.3. Дополнительные характеристики РП

Характеристики новой квалификации определены в приказе Минтруда России от Минтруда России от 06 июля 2020 года N 405н «Об утверждении профессионального стандарта «Специалист по большим данным».

Вид профессиональной деятельности:

- создание и применение технологий больших данных (Код 06.042).

Трудовые функции:

- разработка продуктов на основе встроенной аналитики больших данных (С/01.8).

1.4. Перечень профессиональных компетенций в рамках имеющейся квалификации, качественное изменение которых осуществляется в результате обучения

Получаемые компетенции базируются на основании Приказа Минобрнауки России от 23 августа 2017 г. № 811 «Об утверждении федерального государственного образовательного стандарта высшего образования по направлению подготовки 02.04.02 Фундаментальная информатика и информационные технологии (уровень магистратуры)».

Перечень компетенций:

ОПК-1. Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий.

ОПК-2. Способен применять компьютерные/ суперкомпьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности.

ОПК-3. Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования.

ОПК-4. Способен оптимальным образом комбинировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности.

ОПК-5. Способен устанавливать и сопровождать программное обеспечение информационных систем, осуществлять эффективное управление разработкой программных средств и проектов.

1.5. Соответствие видов деятельности профессиональным компетенциям и их составляющих

Профессиональные компетенции	Практический опыт	Умения	Знания
Разработка продуктов на основе встроенной аналитики больших данных (С/01.8)			
ОПК-1	Проведение научно-исследовательских и опытно-конструкторских работ по созданию вариантов технического воплощения концепции продукта, создание прототипа продукта на основе встроенной аналитики больших данных	Осуществлять математическое и информационное моделирование	Математическое моделирование
ОПК-2	Анализ инноваций в области информационных технологий; новых технических средств, методов и алгоритмов анализа больших данных; источников информации; технологий представления данных, методов предсказательной и предписывающей аналитики; существующих продуктов на основе встроенной аналитики больших данных	Осуществлять разработку программно-аппаратных компонентов и систем	Технические средства и среды сбора, хранения и обработки больших данных
ОПК-3	Анализ требований к продукту, уточнение и доработка концепции, бизнес-модели и бизнес-плана создания нового продукта на основе встроенной аналитики больших данных	Проводить аналитические и поисковые исследования по тематике информационных технологий, технологий больших данных	Существующие и перспективные методы и программный инструментарий технологий больших данных
ОПК-4	Создание и оценка концепции и бизнес-модели продукта на основе встроенной аналитики больших данных	Проводить аналитические работы на основе технологий больших данных	Принципы и методы управления защитой и обеспечением конфиденциальности больших данных
ОПК-5	Разработка эксплуатационной документации по продукту на основе встроенной аналитики больших данных	Осуществлять разработку программно-аппаратных компонентов и систем	Современные и перспективные методы сбора, хранения и передачи данных из гетерогенных источников

2. УЧЕБНЫЙ ПЛАН

2.1. Категория слушателей РП

Имеющаяся квалификация (требования к слушателям) – к освоению ДПП допускаются лица имеющие высшее образование.

2.2. Общая трудоёмкость программы, аудиторная и самостоятельная работа

Общая трудоёмкость программы 74 академических часа, из них 42 академических часа аудиторной работы, 30 академических часов самостоятельной работы и 2 академических часа промежуточной аттестации.

2.3. Форма обучения

Форма обучения по РП – очная с применением дистанционных образовательных технологий.

2.4. Учебный план

РП «Профессиональный модуль» реализуется одним модулем.

№ п/п	Наименование темы, модуля	Форма Контроля	Всего, час	В том числе			
				Лекции	Практические занятия	Самостоятельная работа	Промежут. аттестация
1.	Решение задач компьютерного зрения	Кейс	16	-	10	6	-
2.	Решение задач обработки естественного языка	Кейс	18	-	10	8	-
3.	Решение задач распознавания и синтеза речи	Кейс	18	-	10	8	-
4.	Реализация проектов в области ИИ: создание приложений и их интеграция	Кейс	20	2	10	8	-
5.	Промежуточная аттестация	Зачет	2	-	-	-	2
ИТОГО			74	2	40	30	2

3. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК

№ п/п	Наименование темы, модуля		
		1 неделя	2 неделя
1	Решение задач компьютерного зрения		
2	Решение задач обработки естественного языка		
3	Решение задач распознавания и синтеза речи		
4	Реализация проектов в области ИИ: создание приложений и их интеграция		
5	Промежуточная аттестация		

Минимальный срок освоения программы – 2 недели.

4. РАБОЧАЯ ПРОГРАММА

4.1. Рабочая программа модуля «Профессиональный модуль»

4.1.1. Цель изучения модуля: подготовить высококвалифицированных и конкурентоспособных специалистов в области профессиональной деятельности, которая включает интеллектуальные системы, биоинформатику, когнитивные информационные технологии, вычислительные технологии, компьютерные науки, технологии баз данных, компьютерную графику, теорию информации, технологии управления инфокоммуникацией и бизнес-процессами, архитектуру программного обеспечения, параллельное и распределенное программирование.

Сформировать у слушателей следующие компетенции в области профессиональной деятельности: создание информационных технологий нового поколения, обеспечивающих экономически эффективное извлечение полезной информации из больших объемов разнообразных данных путем высокой скорости их сбора, обработки и анализа, и применение этих технологий в информационно-аналитической деятельности, в системах управления и принятия решений, а также для разработки на их основе новых продуктов и услуг.

4.1.2. Задачи изучения модуля:

1. Сформировать у слушателей основы инженерии данных.
2. Освоение навыков проектирования архитектур, моделей, баз и конвейеров данных.

4.1.3. Планируемые результаты обучения

Процесс изучения раздела направлен на формирование следующих компетенций

Код компетенции	Перечень планируемых результатов обучения по модулю	Формы и методы обучения, способствующие формированию и развитию компетенции
ОПК-1	Знать: математическое моделирование. Уметь: осуществлять математическое и информационное моделирование. Владеть: проведение научно-исследовательских и опытно-конструкторских работ по созданию вариантов технического воплощения концепции продукта, создание прототипа продукта на основе встроенной аналитики больших данных.	Формы обучения: Фронтальная. Методы обучения: Лекция; Практические Работы; Самостоятельная работа.
ОПК-2	Знать: технические средства и среды сбора, хранения и обработки больших данных. Уметь: осуществлять разработку программно-аппаратных компонентов и систем Владеть: анализ инноваций в области информационных технологий; новых технических средств, методов и алгоритмов анализа больших данных; источников информации; технологий представления данных, методов предсказательной и предписывающей	Формы обучения: Фронтальная. Методы обучения: Лекция; Практические Работы; Самостоятельная работа.

	аналитики; существующих продуктов на основе встроенной аналитики больших данных	
ОПК-3	<p>Знать: существующие и перспективные методы и программный инструментарий технологий больших данных.</p> <p>Уметь: проводить аналитические и поисковые исследования по тематике информационных технологий, технологий больших данных.</p> <p>Владеть: анализ требований к продукту, уточнение и доработка концепции, бизнес-модели и бизнес-плана создания нового продукта на основе встроенной аналитики больших данных.</p>	<p>Формы обучения:</p> <p>Фронтальная.</p> <p>Методы обучения:</p> <p>Лекция;</p> <p>Практические Работы;</p> <p>Самостоятельная работа.</p>
ОПК-4	<p>Знать: принципы и методы управления защитой и обеспечением конфиденциальности больших данных.</p> <p>Уметь: проводить аналитические работы на основе технологий больших данных.</p> <p>Владеть: создание и оценка концепции и бизнес-модели продукта на основе встроенной аналитики больших данных.</p>	<p>Формы обучения:</p> <p>Фронтальная.</p> <p>Методы обучения:</p> <p>Лекция;</p> <p>Практические Работы;</p> <p>Самостоятельная работа.</p>
ОПК-5	<p>Знать: современные и перспективные методы сбора, хранения и передачи данных из гетерогенных источников.</p> <p>Уметь: осуществлять разработку программно-аппаратных компонентов и систем.</p> <p>Владеть: разработка эксплуатационной документации по продукту на основе встроенной аналитики больших данных.</p>	<p>Формы обучения:</p> <p>Фронтальная.</p> <p>Методы обучения:</p> <p>Лекция;</p> <p>Практические работы;</p> <p>Самостоятельная работа.</p>

4.1.4 Содержание курса

Тема 1. Решение задач компьютерного зрения (16 часов)

Практическая работа (10 часов).

Практический кейс: «Решение задачи классификации изображений». Для решения кейса будут выполнены следующие этапы:

- Загрузка изображений;
- Первичная обработка изображений
- Изменение формата;
- Работа с цветом изображений;
- Фильтрация изображений;
- Обучение моделей машинного обучения для решения задачи классификации – случайный лес, глубокие нейронные сети, сверточные нейронные сети;
- Портирование и сохранение моделей, а также работа с предобученными моделями.
- Оценка полученных результатов.

Самостоятельная работа (6 часов). Решение практического кейса.

Тема 2. Решение задач обработки естественного языка (18 часов)

Практическая работа (10 часов).

Решение практических кейсов:

1. Задачи обработки естественного языка (Natural Language Processing, NLP)
2. Уровень сигналов
 - a. распознавание текста,
 - b. распознавание речи,
 - c. синтез речи,
 - d. синхронный перевод.
3. Уровень слов и словосочетаний
 - a. морфологический анализ слова,
 - b. лемматизация,
 - c. исправление грамматических ошибок,
 - d. выявление отношений между словами,
 - e. распознавание именованных сущностей
4. Уровень предложений
 - a. синтаксический разбор предложений,
 - b. генерация предложений,
 - c. распознавание именованных сущностей,
 - d. распознавание именных групп,
 - e. анализ тональности
5. Уровень документов
 - a. выделение ключевых слов (keywords extraction),
 - b. автоматическое аннотирование текста (text summarization),
 - c. автоматическое заглавие текста,
 - d. перевод с одного языка на другой,
 - e. тематическое моделирование – определение причастности документа к одной из тем (topic modeling).
6. Уровень корпусов
 - a. выявление и устранение дубликатов,
 - b. выявление плагиата.
7. Конвейер обработки естественного языка
 - a. Выделение предложений
 - b. Токенизация
 - c. Определение частей речи
 - d. Лемматизация

- e. Обработка стоп-слов
- f. Парсинг зависимостей
- g. Поиск групп существительных
- h. Распознавание именованных сущностей (Named Entity Recognition, NER)
- i. Разрешение кореференции
- 8. Реализация NLP на Python. Применение библиотек
- 9. Базовые методы NLP на Python с применением библиотеки NLTK
 - a. Установка библиотеки NLTK
 - b. Использование токенизаторов NLTK
 - c. Выполнение стемминга и лемматизации текста при помощи NLTK
 - d. Использование корпусов текстов в NLTK
 - e. Морфологическая разметка (тэггинг) текстов
 - f. Синтаксический анализ текстов
- 10. Понятия токенизации, стемминга, лемматизации
 - a. Токенизация русского текста при помощи функций NLTK
 - b. Очистка текста от стоп-слов
 - c. Лемматизация и приведение слов к начальной форме
- 11. Модель «Мешка слов» (Bag of Words), ее применение для классификации текста
 - a. Подготовка текста к созданию модели Bag-Of-Words
 - b. Построение частотной модели текста
 - c. Использование естественного текста для формирования рекомендаций
- 12. Применение машинного обучения для sentiment-анализа предложений
 - a. Очистка текстовых данных для sentiment-анализа
 - b. Подготовка текста к созданию модели Bag-Of-Words
 - c. Разделение данных и обучение классификатора
 - d. Тестирование и интерпретация ML-модели sentiment-анализа
- 13. Тематическое моделирование
 - a. Тематическое моделирование, основные понятия
 - b. Подготовка текстовых данных для тематического моделирования
 - c. Анализ сходства тем при помощи скрытого распределения Дирихле (LDA) в Python
 - d. Анализ сходства тем при помощи неотрицательной матричной факторизации (NMF) в Python.

Самостоятельная работа (8 часов). Решение практического кейса.

Тема 3. Решение задач распознавания и синтеза речи (18 часов)

Практическая работа (10 часов).

1. Задачи систем обработки речи:
 - a. Распознавание речи
 - b. Обработка естественного языка
 - c. Синтез речи
2. Трудности в разработке системы распознавания речи
 - a. Размер словаря
 - b. Характеристики канала
 - c. Режим разговора
 - d. Стилль речи
 - e. Зависимость от громкоговорителя (динамика)
 - f. Тип шумов в канале
 - g. Характеристики микрофона
3. Визуализация звуковых сигналов
 - a. Запись
 - b. Выборка
 - c. Отображение
4. Характеристика аудиосигнала: преобразование в частотную область
 - a. преобразование сигнала во временной области в частотную область
 - b. понимание его частотных характеристик сигнала с использованием БПФ (быстрых преобразований Фурье)
5. Генерация монотонного аудиосигнала
6. Извлечение информации о речи из аудиосигнала
 - a. Методы извлечения информации – MFCC, PLP, PLP-RASTA
 - b. Извлечение информации из аудиосигнала с помощью Python, используя технику MFCC и библиотеку `python_speech_features`
7. Распознавание произнесенных слов с помощью Google Speech API на Python
8. Синтез речи
 - a. Задача синтеза речи. Область применения
 - b. Использование модуля gTTS для преобразования текста в речь в Python
9. Анализ и синтез звуковой информации с помощью Python
 - a. Установка библиотек
 - b. Чтение аудиофайлов
 - c. Анализ параметров аудиосигнала
 - d. Анализ частотных характеристик сигнала с помощью БПФ
 - e. Генерация монотонного аудиосигнала с заданными параметрами
 - f. Извлечение информации из аудиосигнала с помощью метода MFCC

- g. Распознавание голоса с помощью Google API
 - 10. Синтез речи при помощи модуля gTTS
 - a. Установка библиотек
 - b. Реализация объекта, распознающего речь, на базе функции gTTS
 - c. Настройка параметров функции распознавания.
- Самостоятельная работа (8 часов). Решение практического кейса.

Тема 4. Реализация проектов в области ИИ: создание приложений и их интеграция (20 часа)

Лекции (2 часа). Изучение тем:

- Управление проектами по разработке приложений в области ИИ.
- Анализ требований, оценка трудоемкости и сроков разработки системы.
- Специфика создания приложений, использующих модели и методы ИИ.
- Основные инструменты управления проектами в области ИИ.
- Системы контроля версий.
- Инструменты для управления версиями программного обеспечения.
- Версионирование моделей и данных. Основные подходы.
- Версионирование экспериментов при разработке систем ИИ.
- Работа с предобученными моделями. Специфика внедрения моделей машинного обучения в мобильные, настольные и веб-приложения.
- Создание веб-приложений, использующих модели машинного обучения.
- Введение в веб-разработку.
- Знакомство с библиотеками языка Python для создания веб-приложений.
- Основы развёртывания и запуска веб-приложений.

Практическая работа (10 часов).

Решение практического кейса:

«Разработка веб-приложения для отображения предсказаний модели машинного обучения»

Практический кейс представляет собой написание программы на языке программирования Python с последующим развёртыванием и запуском. Основные шаги:

- a. Создание удаленного репозитория.
- b. Проектирование архитектуры системы.
- c. Разработка и тестирование веб-приложения.
- d. Разработка, обучение и тестирование моделей машинного обучения. Выбор наилучшей модели.
- e. Интеграция модели в веб-приложение.

f. Сборка и запуск приложения (локально).

Самостоятельная работа (8 часов). Решение практического кейса.

4.1.5. Оценочное средство для текущего контроля (примерные формулировки практических заданий):

Тема 1. Формулировка практического кейса.

Рассматривается практический кейс «Обработка диагностических изображений для выявления болезни Паркинсона»

Описание: Выполняется сегментация изображения для выявления траектории формирования рисунков диагностируемым, проводится комплекс предобработки данных, очистки, нормализации данных, разметка дата сета. Далее выполняется построение индивидуального и ансамблевого классификатора различными (не менее 3-х) методами интеллектуального анализа данных и машинного обучения. Выполняется статистическое исследование точности построенных классификаторов, оценивается статистическая устойчивость результатов и определяется наиболее эффективный вариант решения целевой задачи. Полученный классификатор интегрируется в программную систему обработки данных по диагностике оценке прогнозных признаков болезни Паркинсона.

Гипотезы задачи: На основе полученного набора данных выстроить приемлемо-точную систему.

Тема 2. Формулировка практического кейса.

Рассматривается практический кейс «Обработка текстовой информации отзывов»

Описание: В рамках решаемого практического кейса выполняется полный цикл обработки текстовой информации.

Гипотезы задачи: Обработка текстовой информации может быть эффективно выполнена методами машинного обучения и интеллектуального анализа данных, но для этого требуется специальные процедуры структурирования и предобработки текстов.

Тема 3. Формулировка практического кейса.

Рассматривается практический кейс «Обработка данных речевой медицинской диагностики для выявления признаков болезни Паркинсона»

Описание: Классификатор строится на основе набора данных записи речевой информации, фиксируемой в ходе начитки специального текста диагностируемым. Выполняется сегментация речевой информации и обработка записей для выявления целевых параметров произношения диагностируемым, проводится комплекс предобработки данных, очистки, нормализации данных, разметка дата сета.

Гипотезы задачи: На основе полученного набора данных необходимо выстроить приемлемо-точечную систему.

Тема 4. Формулировка практического кейса.

Рассматривается практический кейс «Разработка приложения для подготовки и обработки данных различного типа, полученных средствами медицинского диагностирования болезни Паркинсона»

Описание: В рамках решаемого практического кейса выполняется программная реализация и сборка в единое программное приложение всех разработанных ранее процедур подготовки, предобработки данных и моделей машинного обучения и моделей искусственного интеллекта. Собранное приложение проверяется на работоспособность, отлаживается и подключается к базе данных под управление СУБД. оценивается работоспособность и эффективность интегрированных в приложение решателей: классификаторов, регрессоров, кластеризирующих методов. Реализуется процедура выгрузки предобработанных структурированных данных, а также сохранение для последующего использования аналитических моделей: регрессоров, классификаторов и кластеров.

Гипотезы задачи: Создание приложения, комплексующего рассмотренные методы и средства хранения, получения, поиска, предобработки и очистки данных, сокращения размерности и структурирования данных, построения моделей с использованием методов машинного обучения и на основе технологий искусственного интеллекта, выгрузки данных и моделей, является эффективным вариантом решения практических датаориентированных задач.

5. УСЛОВИЯ РЕАЛИЗАЦИИ РП

5.1. Организационные условия реализации РП

Наименование аудитории	Вид занятия	Наименование оборудования, программного обеспечения
Аудитория для проведения лекций/семинаров	Лекции	ПК с доступом в Интернет и возможностью просмотра файлов в формате *.ppt, *.pptx, *.pdf; микрофон; колонки/наушники; камер.
Лаборатории/ производственные помещения	Практические занятия	Личный ПК/смартфон с доступом в Интернет и возможностью просмотра файлов в формате *.html, *.doc, *.docx, *.pdf; лист формата А5/А4 или блокнот; карандаш/ручка.
Коворкинги, учебные залы и т.д.	Самостоятельная работа	Личный ПК/смартфон с доступом в Интернет и возможностью просмотра файлов в формате *.html, *.doc, *.docx, *.pdf; лист формата А5/А4 или блокнот; карандаш/ручка.
Аудитория для проведения лекций/семинаров	Итоговая аттестация	ПК с доступом в Интернет и возможностью просмотра файлов в формате *.html, *.doc, *.docx, *.pdf, *.djvu, лист бумаги формата А4, ручка.

5.2. Педагогические условия реализации РП

Реализация программы обеспечивается преподавательским составом, удовлетворяющим следующим условиям:

- наличие высшего профессионального образования, соответствующее профилю программы, из числа штатных преподавателей, или привлеченных на условиях почасовой оплаты труда;
- значительный опыт практической деятельности в соответствующей сфере из числа штатных преподавателей или привлеченных на условиях почасовой оплаты труда

5.3. Учебно-методическое обеспечение РП

Основная литература:

1. Алпайдин Э. Машинное обучение: новый искусственный интеллект: пер. с англ. / Алпайдин Э. - М.: Фонд Развития Промышленности: Издательская группа "Точка" : Альпина Паблишер : [Интеллектуальная Литература], 2017. - 191 с. - (Завтра это будут знать все). - Библиогр.: с. 185-191. - ISBN 978-5-9908700-8-6. - ISBN 978-5-9614-6114-5.
2. Силен Д., Мейсман А., Али М. Основы Data Science и Big Data. Python и наука о данных / Силен Д., Мейсман А., Али М. ; пер. с англ. Матвеев Е. - СПб. : Питер, 2020. - 334 с. : ил. - (Библиотека программиста). - ISBN 978-5-4461-0944-9.
3. Хайкин С. Нейронные сети. Полный курс: пер. с англ. / Хайкин С.; пер. Куссуль Н. Н., Шелестов А. Ю. - 2-е изд. - М.; СПб.; Киев: Вильямс, 2006. - 1103 с.: ил. - Библиогр.: с. 996-1069. - ISBN 5-8459-0890-6.
4. Маркин, А. В. Программирование на SQL: учебное пособие для среднего профессионального образования / А. В. Маркин. — Москва: Издательство Юрайт, 2020. — 435 с. — (Профессиональное образование). — ISBN 978-5-534-11093-7. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/456926>
5. Маркин, А. В. Программирование на SQL в 2 ч. Часть 2: учебник и практикум для вузов / А. В. Маркин. — 2-е изд., испр. и доп. — Москва: Издательство Юрайт, 2020. — 340 с. — (Высшее образование). — ISBN 978-5-534-12258-9. — Текст: электронный // ЭБС Юрайт [сайт]. — URL: <https://urait.ru/bcode/451185>.

Дополнительные материалы:

1. Белоус В. В., Пивоварова Н. В. Основы реляционных баз данных. Практикум по SQL [Электрон. ресурс]: метод. указ. к лаб. работам по курсу "Базы данных" / Белоус В. В., Пивоварова Н. В.; МГТУ им. Н. Э. Баумана, Фак. "Робототехника и комплексная автоматизация". - М.: Изд-во МГТУ им. Н. Э. Баумана, 2014. - 1 CD-ROM. - ФГУП "Информрегистр" №0321400922.
2. Басараб М. А., Коннова Н. С. Интеллектуальные технологии на основе искусственных нейронных сетей: метод. указания к выполнению лаб. работ / Басараб М. А., Коннова Н. С.; МГТУ им. Н. Э. Баумана. - М.: Изд-во МГТУ им. Н. Э. Баумана, 2017. - 53 с.: ил. - Библиогр. в конце брош. - ISBN 978-5-7038-4716-9.
3. Березкин Д. В., Рожнев А. Ю. Разработка модели оценки платежеспособности клиентов банка с применением алгоритмов машинного обучения / Березкин Д. В., Рожнев А. Ю. // Динамика сложных систем. - 2018. - Т. 12, № 4. - С. 59-66.
4. Бизли Д. М. Язык программирования Python: справочник: Пер. с англ / Бизли Д. М. - Киев: ДиаСофт, 2000. - 326 с. - ISBN 966-7393-54-2.

5. Вьюгин В. В. Элементы математической теории машинного обучения: учеб. пособие для вузов / Вьюгин В. В.; Моск. физико-техн. ин-т (гос. ун-т), РАН. Ин-т проблем передачи информации им. А. А. Харкевича. - М.: МФТИ - ИППИ РАН, 2010. - 231 с. - Библиогр.: с. 229-231. - ISBN 978-5-7417-0339-7.
6. Галушкин А. И. Нейронные сети: основы теории / Галушкин А. И. - М.: Горячая линия - Телеком, 2010. - 496 с.: ил. - Библиогр. в конце ст., с. 469-488. - ISBN 978-5-9912-0082-0.
7. Головкин В. А. Нейронные сети: обучение, организация и применение: учеб. пособие по направлению подготовки бакалавров и магистров / Головкин В. А.; общ. ред. Галушкин А. И. - М.: Издат. предприятие ред. журн. "Радиотехника" (ИПРЖР), 2001. - 256 с.: ил. - (Нейрокомпьютеры и их применение; кн. 4). - Библиогр.: с. 248-256. - ISBN 5-93108-05-8.
8. Джеймс Р., Грофф П., Вайнберг Н., Оппель Э. Дж. SQL. Полное руководство. М: Вильямс, 2014. 960 с.
9. Коннолли Т., Бегг К. Базы данных. Проектирование, реализация и сопровождение. Теория и практика. М: Вильямс, 2017. 1440 с.
10. Комашинский В. И., Смирнов Д. А. Нейронные сети и их применение в системах управления и связи / Комашинский В. И., Смирнов Д. А. - М.: Горячая линия - Телеком, 2003. - 93 с. - Библиогр.: с. 88-93. - ISBN 5-93517-094-9.
11. Осовский С. Нейронные сети для обработки информации: пер. с польск. / Осовский С.; пер. Рудинский И. Д. - М.: Финансы и статистика, 2002. - 343 с.: ил. - ISBN 5-279-02567-4.
12. Сидняев Н. И., Храпов П. В. Нейросети и нейроматематика: учеб. пособие / Сидняев Н. И., Храпов П. В.; МГТУ им. Н. Э. Баумана. - М.: Изд-во МГТУ им. Н. Э. Баумана, 2016. - 83 с.: ил. - Библиогр.: с. 82. - ISBN 978-5-7038-4362-8.
13. Хливненко Л. В., Пятакович Ф. А. Практика нейросетевого моделирования: учеб. пособие / Хливненко Л. В., Пятакович Ф. А. - СПб. Лань, 2019. - 196 с. - (Учебники для вузов. Специальная литература) (Бакалавриат и магистратура). - Библиогр.: с. 182-193. - ISBN 978-5-8114-3639-2.

5.4. Методические рекомендации

РП построена по тематическому принципу, каждый раздел представляет собой логически завершенный материал.

Преподавание программы основано на личностно-ориентированной технологии образования, сочетающей два равноправных аспекта этого процесса: обучение и учение. Личностно-ориентированный подход развивается при участии слушателей в активной работе на практических занятиях. Личностно-ориентированный подход направлен, в первую очередь, на развитие индивидуальных способностей обучающихся, создание условий для развития творческой активности слушателя и разработке инновационных идей, а также на развитие самостоятельности мышления при решении учебных задач разными способами, нахождение рационального варианта решения, сравнения и оценки нескольких вариантов их решения и т.п. Это способствует формированию приемов умственной деятельности по восприятию новой информации, ее запоминанию и осознанию, созданию образов для сложных понятий и процессов, приобретению навыков поиска решений в условиях неопределенности.

Практические занятия проводятся для приобретения навыков решения практических задач в предметной области модуля. Задания, выполняемые на практических занятиях, выполняются с использованием активных и интерактивных методов обучения.

Самостоятельная работа слушателей предназначена для проработки дополнительной литературы. Результаты практических заданий слушателей учитываются на итоговой аттестации.

При изучении курса предусмотрены следующие методы организации и осуществления учебно-познавательной деятельности:

- объяснительно-иллюстративный метод;
- репродуктивный метод;
- частично-поисковый метод.

6. ФОРМЫ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ РП

Итоговая аттестация проводится в форме зачета. Зачет проводится в формате тестирования и решения кейса. Результатом зачета служат правильные ответы на вопросы теста, состоящего из 10 вопросов.

По результатам итоговой аттестации слушателю выставляется оценка «ЗАЧТЕНО/НЕ ЗАЧТЕНО»:

Оценка «ЗАЧТЕНО» выставляется слушателю, который:

- правильно ответил не менее чем на 60% вопросов в тесте;
- практическое задание (кейс от индустриального партнера) выполнено полностью, нет ошибок в логических рассуждениях. Возможно наличие одной неточности или описки, не являющихся следствием незнания или непонимания учебного материала. Слушатель показал полный объем знаний, умений в освоении, пройденных тем и применение их на практике.
- продемонстрировал необходимые систематизированные знания и достаточную степень владения принципами предметной области программы, понимание их особенностей и взаимосвязь между ними в течение всего срока обучения по ДПП.

Оценка «НЕ ЗАЧТЕНО» ставятся слушателю, который:

- правильно ответил менее чем на 60% вопросов в тесте;
- практическое задание (кейс от индустриального партнера) выполнено не полностью. Допущены грубые ошибки. Задание выполнено не самостоятельно. Задание не сдано.
- имеет крайне слабые теоретические и практические знания, обнаруживает неспособность к построению самостоятельных заключений

7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ

7.1. Паспорт комплекта оценочных средств

Предметы оценивания	Объекты оценивания	Показатели оценки
ОПК-1. Способен находить, формулировать и решать актуальные проблемы прикладной математики, фундаментальной информатики и информационных технологий	Тестовое задание	Количество правильных ответов
ОПК-2. Способен применять компьютерные/ суперкомпьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности	Тестовое задание	Количество правильных ответов
ОПК-3. Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования	Тестовое задание	Количество правильных ответов
ОПК-4. Способен оптимальным образом комбинировать существующие информационно-коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности	Тестовое задание	Количество правильных ответов
ОПК-5. Способен устанавливать и сопровождать программное обеспечение информационных систем, осуществлять эффективное управление разработкой программных средств и проектов	Тестовое задание	Количество правильных ответов

7.2. Комплект оценочных средств

7.2.1. Темы для подготовки к зачету:

1. Большие данные.
2. Искусственный интеллект.
3. Информационная безопасность.
4. Архитектура данных.
5. Модель данных.
6. Сбор и хранение данных.
7. Конвейер данных.
8. Представление данных.
9. Подготовка данных.
10. Облачные технологии.

7.2.2. Примерные тесты для проведения зачета:

- 1) **Основным отличием реляционной базы данных является:**
 - а) организация данных в виде отношений
 - б) строго древовидная структура
 - в) представление в виде графов

- 2) **База данных – это:**
 - а) совокупность файлов на жестком диске
 - б) пакет пользовательских программ
 - в) совокупность сведений, характеризующих объекты, процессы или явления реального мира

- 3) **Последовательность операций над базой данных, переводящих ее из одного непротиворечивого состояния в другое непротиворечивое состояние, называется:**
 - а) транзитом
 - б) циклом
 - в) транзакцией

- 4) **Предметная область – это:**
 - а) часть реального мира, представляющая интерес для данного исследования
 - б) база данных, разработанная для решения конкретной задачи
 - в) ER-диаграмма, отражающая заданную область внешнего мира

- 5) **Фильтрация записей в таблицах выполняется с целью:**
 - а) выборки необходимых данных
 - б) группировки данных
 - в) сортировки данных

- 6) **Множество объектов, отобранное с целью изучения поведения всех возможных объектов такого типа называется:**
 - а) Генеральной совокупностью
 - б) Стратифицированной выборкой
 - в) Репрезентативной выборкой
 - г) Выборкой объектов

7) Разность максимального и минимального значения в выборке называется:

- а) Медиана
- б) Размах
- в) Дисперсия
- г) Корреляция

8) Имеется множество наблюдений параметра исследуемого объекта:

{2; 1; 2; 3; 8; 4; 2; 5; 4; 7; 6; 2; 1; 4; 5}

Чему равна Мода (Mode) выборки?

9) Сверточная нейронная сеть имеет аббревиатуру:

- а) LSTM
- б) CNN
- в) Transformer
- г) YOLO

10) Задача распознавания речи имеет название:

- а) Semantic Segmentation
- б) Instance Segmentation
- в) Object Detection
- г) Speech Recognition

7.2.3. Кейсы для итогового контроля.

Кейс №1. Прогнозирование безопасной навигации судоходных рек при осуществлении «северного завоза».

Краткое описание:

По территории Красноярского края протекает огромное количество рек. Однако многие из них являются судоходными лишь в короткий период половодья. Точно предсказать срок навигации на таких реках – важнейшая задача.

Описание:

По территории Красноярского края протекает огромное количество рек, многие из этих рек являются судоходными и являются важнейшими транспортными путями. Однако навигация в енисейском бассейне крайне сложна. Многие реки являются судоходными лишь в короткий период половодья. Точно предсказать срок навигации на таких реках – важнейшая задача.

Ежегодно в Красноярском крае происходит «северный завоз» - это комплекс мероприятий по доставке речным транспортом необходимых запасов, оборудования и материалов в населенные