

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)



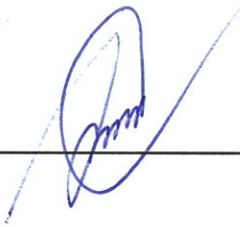
Дополнительное профессиональное образование

**ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА  
ПРОГРАММА ПРОФЕССИОНАЛЬНОЙ ПЕРЕПОДГОТОВКИ  
«Архитектор данных (Data Architect Pro)»  
Рег. 05.22.23.01.45**

Москва, 2023 г.

**АВТОР ПРОГРАММЫ:**

Доцент

  
\_\_\_\_\_ В.С. Тынченко


**СОГЛАСОВАНО:**

Начальник УСП

  
\_\_\_\_\_ Т.А. Гузева

Директор  
Центра дополнительного образования

  
\_\_\_\_\_ М.В. Стоянова

11292 

## СОДЕРЖАНИЕ

<b>1. Общая характеристика дополнительной профессиональной программы (ДПП).....</b>	<b>стр.</b>	<b>4</b>
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
<b>2. Учебный план ДПП .....</b>	<b>стр.</b>	<b>6</b>
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
<b>3. Календарный учебный график .....</b>	<b>стр.</b>	<b>6</b>
<b>4. Рабочие программы учебных предметов (курсов, дисциплин (модулей)) .....</b>	<b>стр.</b>	<b>7</b>
4.1. Рабочая программа учебного предмета (курса, дисциплины, модуля) №1 «Введение в технологии ИИ и инженерии данных». Приложение №1 .....	стр.	7
4.2. Рабочая программа учебного предмета (курса, дисциплины, модуля) №2 «Архитектура данных в области ИИ и BigData». Приложение №2.....	стр.	7
<b>5. Условия реализации ДПП .....</b>	<b>стр.</b>	<b>7</b>
5.1. Организационные условия реализации ДПП .....	стр.	7
5.2. Педагогические условия реализации ДПП .....	стр.	7
5.3. Учебно-методическое обеспечение ДПП .....	стр.	7
<b>6. Формы итоговой аттестации .....</b>	<b>стр.</b>	<b>9</b>
<b>7. Оценочные материалы итоговой аттестации .....</b>	<b>стр.</b>	<b>10</b>
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 «Архитектор данных (Data Architect 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: 266, из них 144 аудиторной работы и 88 самостоятельной работы.

### 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	
1	Введение в технологии ИИ и инженерии данных	102	32	32	38	Зачет
2	Архитектура данных в области ИИ и BigData	130	36	44	50	Зачет
3	Итоговая аттестация	34	0	0	0	Подготовка и защита ВКР
<b>Всего часов</b>		266	68	76	88	

## 3. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК

№ п/п	Наименование темы, модуля	1 неделя	2 неделя	3 неделя	4 неделя	5 неделя	6 неделя
1.	Введение в технологии ИИ и инженерии данных						
2.	Архитектура данных в области ИИ и BigData						
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 «Введение в технологии ИИ и инженерии данных». Приложение №1

4.2. Рабочая программа учебного предмета (курса, дисциплины, модуля) №2 «Архитектура данных в области ИИ и BigData». 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	Введение в технологии ИИ и инженерии данных	С.В. Картамышев	
2	Архитектура данных в области ИИ и BigData	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. Куцуль 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.



## **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 экзаменационной 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	<b>Знать:</b> Математическое моделирование <b>Уметь:</b> Осуществлять математическое и информационное моделирование <b>Владеть:</b> Проведение научно-исследовательских и опытно-конструкторских работ по созданию вариантов технического воплощения концепции продукта, создание прототипа продукта на основе встроенной аналитики больших данных

Формулировка компетенции по ФГОС	Код компетенции	Совокупные результаты освоения
<p>Способен применять компьютерные/ супер-компьютерные методы, современное программное обеспечение (в том числе отечественного производства) для решения задач профессиональной деятельности</p>	ОПК-2	<p><b>Знать:</b> Технические средства и среды сбора, хранения и обработки больших данных</p> <p><b>Уметь:</b> Осуществлять разработку программно-аппаратных компонентов и систем</p> <p><b>Владеть:</b> 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>
<p>Способен проводить анализ математических моделей, создавать инновационные методы решения прикладных задач профессиональной деятельности в области информатики и математического моделирования</p>	ОПК-3	<p><b>Знать:</b> Существующие и перспективные методы и программный инструментарий технологий больших данных</p> <p><b>Уметь:</b> Проводить аналитические и поисковые исследования по тематике информационных технологий, технологий больших данных</p> <p><b>Владеть:</b> Анализ требований к продукту, уточнение и доработка концепции, бизнес-модели и бизнес-плана создания нового продукта на основе встроенной аналитики больших данных</p>
<p>Способен оптимальным образом комбинировать существующие информационно-</p>	ОПК - 4	<p><b>Знать:</b> Принципы и методы управления защитой и обеспечением конфиденциальности больших данных</p>

Формулировка компетенции по ФГОС	Код компетенции	Совокупные результаты освоения
коммуникационные технологии для решения задач в области профессиональной деятельности с учетом требований информационной безопасности		<p><b>Уметь:</b> Проводить аналитические работы на основе технологий больших данных</p> <p><b>Владеть:</b> 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><b>Знать:</b> Современные и перспективные методы сбора, хранения и передачи данных из гетерогенных источников</p> <p><b>Уметь:</b> Осуществлять разработку программно-аппаратных компонентов и систем</p> <p><b>Владеть:</b> 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 от 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  
**«Введение в технологии ИИ и инженерии данных»**  
ДОПОЛНИТЕЛЬНОЙ ПРОГРАММЫ ПРОФЕССИОНАЛЬНОЙ  
ПЕРЕПОДГОТОВКИ **«Архитектор данных (Data Architect Pro)»**

Москва, 2023

## Оглавление

<b>1. ОБЩАЯ ХАРАКТЕРИСТИКА РП</b> .....	3
1.1. Цель .....	3
1.2. Планируемые результаты обучения .....	3
1.4. Перечень профессиональных компетенций в рамках имеющейся квалификации, качественное изменение которых осуществляется в результате обучения.....	4
1.5. Соответствие видов деятельности профессиональным компетенциям и их составляющих .....	5
<b>2. УЧЕБНЫЙ ПЛАН</b> .....	6
2.1. Категория слушателей РП.....	6
2.2. Общая трудоёмкость программы, аудиторная и самостоятельная работа.....	6
2.3. Форма обучения.....	6
2.4. Учебный план .....	6
<b>3. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК</b> .....	7
<b>4. РАБОЧАЯ ПРОГРАММА</b> .....	7
<b>5. УСЛОВИЯ РЕАЛИЗАЦИИ РП</b> .....	15
5.1. Организационные условия реализации РП .....	15
5.2. Педагогические условия реализации РП .....	15
5.3. Учебно-методическое обеспечение РП.....	15
5.4. Методические рекомендации.....	17
<b>6. ФОРМЫ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ РП</b> .....	17
<b>7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ</b> .....	18
7.1. Паспорт комплекта оценочных средств .....	18
7.2. Комплект оценочных средств.....	19

## 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 «**Введение в технологии ИИ и инженерии данных**», могут быть допущены к освоению следующего -

второго уровня дополнительной программы профессиональной переподготовки «Архитектор данных (Data Architect 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. № 808 «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
Проектирование компьютерного программного обеспечения (D/03.6)			
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 среднее профессиональное и/или высшее образование.

### 2.2. Общая трудоёмкость программы, аудиторная и самостоятельная работа

Общая трудоёмкость программы 102 академических часа, из них 64 академических часа аудиторной работы, 36 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 «Введение в технологии ИИ и инженерии данных» 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 ИИ и BigData. Основные понятия	Практ. задание	10	4	2	4	-
2.	Базовые инструменты специалиста в области ИИ	Практ. задание	10	4	2	4	-
3.	Основные методы ИИ, машинного обучения и Data Science	Практ. задание	28	8	12	8	-
4.	Современные принципы проектирования архитектур информационных и аналитических систем	Практ. задание	10	2	2	6	-
5.	Модели данных и СУБД	Практ. задание	14	4	6	4	-
6.	Физические особенности организации систем хранения данных на диске	Практ. задание	10	4	2	4	-
7.	Современные практики DevOps. Виртуализация и контейнеризация	Практ. задание	12	4	4	4	-



8.	Аспекты информационной безопасности при сборе, обработке и хранении данных	Практ. задание	6	2	2	2	-
9.	Промежуточная аттестация	Зачет	2	-	-	-	2
	ИТОГО	-	102	32	32	36	2

### 3. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК

№ п/п	Наименование темы, модуля	1 неделя	2 неделя	3 неделя
1	Введение в ИИ и BigData. Основные понятия			
2	Базовые инструменты специалиста в области ИИ			
3	Основные методы ИИ, машинного обучения и Data Science			
4	Современные принципы проектирования архитектур информационных и аналитических систем			
5	Модели данных и СУБД			
6	Физические особенности организации систем хранения данных на диске			
7	Современные практики DevOps. Виртуализация и контейнеризация			
8	Аспекты информационной безопасности при сборе, обработке и хранении данных			
9	Промежуточная аттестация			Зачет

**Минимальный срок освоения программы – 3 недели.**

### 4. РАБОЧАЯ ПРОГРАММА

#### 4.1. Рабочая программа модуля «Введение в технологии ИИ и инженерии данных»

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	<p><b>Знать:</b> математическое моделирование.</p> <p><b>Уметь:</b> осуществлять математическое и информационное моделирование.</p> <p><b>Владеть:</b> проведение научно-исследовательских и опытно-конструкторских работ по созданию вариантов технического воплощения концепции продукта, создание прототипа продукта на основе встроенной аналитики больших данных.</p>	<p>Формы обучения: Фронтальная.</p> <p>Методы обучения: Лекция; Практические Работы; Самостоятельная работа.</p>
ОПК-2	<p><b>Знать:</b> технические средства и среды сбора, хранения и обработки больших данных.</p> <p><b>Уметь:</b> проводить аналитические работы с использованием технологий больших данных.</p> <p><b>Владеть:</b> определение необходимых ресурсов для проведения аналитических работ.</p>	<p>Формы обучения: Фронтальная.</p> <p>Методы обучения: Лекция; Практические Работы; Самостоятельная работа.</p>
ОПК-3	<p><b>Знать:</b> существующие и перспективные методы и программный инструментарий технологий больших данных.</p> <p><b>Уметь:</b> проводить аналитические и поисковые исследования по тематике информационных технологий, технологий больших данных.</p> <p><b>Владеть:</b> 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: Фронтальная.</p> <p>Методы обучения: Лекция; Практические Работы; Самостоятельная работа.</p>
ОПК-4	<p><b>Знать:</b> принципы и методы управления защитой и обеспечением конфиденциальности больших данных.</p> <p><b>Уметь:</b> проводить аналитические работы на основе технологий больших данных.</p> <p><b>Владеть:</b> создание и оценка концепции и бизнес-модели продукта на основе встроенной аналитики</p>	<p>Формы обучения: Фронтальная.</p> <p>Методы обучения: Лекция; Практические Работы; Самостоятельная работа.</p>

	больших данных.	
ОПК-5	<p><b>Знать:</b> современные и перспективные методы сбора, хранения и передачи данных из гетерогенных источников.</p> <p><b>Уметь:</b> осуществлять разработку программно-аппаратных компонентов и систем.</p> <p><b>Владеть:</b> разработка эксплуатационной документации по продукту на основе встроенной аналитики больших данных.</p>	<p>Формы обучения: Фронтальная. Методы обучения: Лекция; Практические работы; Самостоятельная работа.</p>

#### 4.1.4 Содержание курса

##### **Тема 1. Введение в ИИ и BigData. Основные понятия (10 часов)**

Лекции (4 часа). Изучение тем:

- Введение в ИИ и BigData. Основные понятия.
- Текущее состояние отрасли, тренды ИИ и BigData, актуальные задачи.
- Природа больших данных.
- Классификация задач, решаемых ИИ и технологиями BigData.
- Специализации в области задач ИИ и BigData.
- Задачи и компетенции Архитектора данных. Стандартный пайплайн работы с данными (dataflow) в задачах ИИ.

Практическая работа (2 часа).

- Детальный разбор решений практических заданий.
- Дискуссия, обмен мнениями по темам лекций и заданий.

Самостоятельная работа (4 часа). Решение практического задания.

##### **Тема 2. Введение в ИИ и BigData. Основные понятия (10 часов)**

Лекции (4 часа). Изучение тем:

- Языки программирования, среды исполнения и операционные системы, используемые в задачах ИИ и анализа данных.
- ОС Linux. Наиболее популярные и используемые дистрибутивы.
- Языки программирования, используемые в задачах ИИ и анализа данных: Python, R, Julia, Scala, C++.
- Среда Jupyter Notebook.
- Библиотеки анализа данных и машинного обучения на Python. Библиотеки Numpy, Scipy для научных вычислений. Библиотека Pandas, как стандарт исследования данных. Визуализация в Python и срезы данных. Библиотеки Matplotlib, seaborn, plotly.

Библиотека Sklearn. Библиотеки для работы с нейросетями: TensorFlow и PyTorch.  
Библиотека компьютерного зрения OpenCV.

- Вычисление нейросетевых моделей на GPU. CUDA.

Практическая работа (2 часа).

- Детальный разбор и проверка решений практических заданий.

- Дискуссия, обмен мнениями по темам лекций и задания.

Самостоятельная работа (4 часа). Решение практического задания.

### **Тема 3. Основные методы ИИ, машинного обучения и Data Science (28 часов)**

Лекции (8 часов). Изучение тем:

- Основные понятия теории вероятности.
- Обучение с учителем и без учителя. Задачи классификация и кластеризации.
- Алгоритмы кластеризации. Проблема снижения размерности данных. Метод главных компонент (Principal Component Analysis). Метод k-means, c-means, hierarchical clustering.
- Алгоритмы классификации. Байесовский вероятностный классификатор, Метод ближайших соседей k-NN, метод опорных векторов, линейная и логистическая регрессия и нейронные сети.
- Нейронные сети. Перцептрон. Глубокое обучение. Сверточные нейронные сети. Рекуррентные нейронные сети. Генеративные модели.
- Ансамбли решателей. Бэггинг, бустинг, стэкинг. Решающие деревья (Decision tree). Случайный лес (Random forest).
- Обучение с подкреплением.
- Введение в интеллектуальный анализ данных. Предварительная обработка данных. Постобработка результатов. Задача подготовки выборок данных.

Практическая работа (12 часов).

- Обсуждение способов выполнения практического задания, с примерами и демонстрациями.
- Детальный разбор и проверка решений практических заданий.
- Дискуссия, обмен мнениями по темам лекций и задания.

Самостоятельная работа (8 часов). Решение практического задания.

### **Тема 4. Современные принципы проектирования архитектур информационных и аналитических систем (10 часов)**

Лекции (2 часа). Изучение тем:

- Понятие чистой и луковичной архитектуры.
- Сервис-ориентированная и микросервисная архитектура.

- Связывание компонентов сложной системы через API, очереди сообщений и шины данных.
- Уровни архитектуры. Сущности системы. Модели сущности, модели хранения, модели представления, транспортные модели.  
Практическая работа (2 часа).
- Обсуждение способов выполнения практического задания, с примерами и демонстрациями.
- Детальный разбор и проверка решений практических заданий.
- Дискуссия, обмен мнениями по темам лекций и задания.  
Самостоятельная работа (6 часов). Решение практического задания.

#### **Тема 5. Модели данных и СУБД (14 часов)**

Лекции (4 часа). Изучение тем:

- Уровни представления данных.
- Модель данных. Реляционная, колоночная, документно-ориентированная и графовая модели, модель ключ-значение.
- Форматы представления данных. CSV, XML, JSON, бинарные форматы.
- Виды СУБД. Вопросы репликации и шардирования СУБД.  
Практическая работа (6 часов).
- Обсуждение способов выполнения практического задания, с примерами и демонстрациями.
- Детальный разбор и проверка решений практических заданий.
- Дискуссия, обмен мнениями по темам лекций и задания.  
Самостоятельная работа (4 часа). Решение практического задания.

#### **Тема 6. Физические особенности организации систем хранения данных на диске (10 часов)**

Лекции (4 часа). Изучение тем:

- Типы физических систем хранения.
- Теория жестких дисков. Контроллеры и дисковые массивы.
- Протоколы СХД. Типы СХД. Достоинства и недостатки разных типов СХД.
- Особенности работы и эксплуатации физических носителей информации.  
Практическая работа (2 часа).
- Детальный разбор и проверка решений практических заданий.
- Дискуссия, обмен мнениями по темам лекций и задания.  
Самостоятельная работа (4 часа). Решение практического задания.

## **Тема 7. Современные практики DevOps. Виртуализация и контейнеризация (12 часов)**

Лекции (4 часа). Изучение тем:

- Современная методология DevOps.
- Виртуализация и контейнеризация. Docker.
- Оркестрация контейнеров. Docker Swarm. K8S.
- Мониторинг и логирование. Prometheus. Grafana. ELK.
- CI/CD.
- Конфигурационное управление. Ansible.
- Управление инфраструктурой. Terraform.

Практическая работа (4 часа).

- Обсуждение способов выполнения практического задания, с примерами и демонстрациями.
  - Детальный разбор и проверка решений практических заданий.
  - Дискуссия, обмен мнениями по темам лекций и задания.
- Самостоятельная работа (4 часа). Решение практического задания.

## **Тема 8. Аспекты информационной безопасности при сборе, обработке и хранении данных (6 часов)**

Лекции (2 часа). Изучение тем:

- Понятие модели угроз в аналитических система сбора, анализа и хранения данных.
- Нормативно-правовая база информационной безопасности.
- Информационная безопасность баз и хранилищ данных.
- Особенности сбора, хранения и обработки персональных данных.

Практическая работа (2 часа).

- Обсуждение способов выполнения практического задания, с примерами и демонстрациями.
- Детальный разбор и проверка решений практических заданий.
- Дискуссия, обмен мнениями по темам лекций и задания.

Самостоятельная работа (4 часа). Решение практического задания.

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

*Тема 1. Формулировка практического задания.*

Необходимо выбрать один из предложенный кейсов (или предложить и согласовать с преподавателем свой кейс, из личного опыта/практики/работы). Выбранную предметную область рассмотреть с точки зрения возможности применения в ней современных

возможностей ИИ и работы с данными, а также решения обозначенных проблем предметной области посредством сбора и анализа данных. Оценить примерный вид, объем и качество собираемой информации. Привести гипотетический эффект от применения указанных инструментов.

*Тема 2. Формулировка практического задания.*

1. Настроить рабочую среду (ОС Linux, Python, Jupyter Notebook).
2. Выбрать и согласовать с преподавателем в рассматриваемом кейсе любую атомарную операцию с интеллектуальной обработкой данных, которую возможно решить с помощью готовых инструментов и предобученных моделей машинного обучения (например, распознавание самолетов на ВПП по камере видеонаблюдения).
3. Установить и настроить необходимые библиотеки, загрузить предобученную модель.
4. Написать небольшую программу на Python, позволяющую решать выбранную атомарную задачу.

*Тема 3. Формулировка практического задания.*

1. Выбрать и согласовать с преподавателем в рассматриваемом кейсе любую атомарную операцию с интеллектуальной обработкой данных, которую можно классифицировать как задачу классификации, кластеризации или прогнозирования (например, сегментация клиентов по покупательскому поведению или определение тематик и эмоциональной окраски текстов в текстовых сообщениях).
2. Выбрать алгоритм/метод решения. Установить и настроить необходимые библиотеки.
3. Определить используемые структуры данных. Найти готовую выборку данных, подходящую под решаемую задачу или написать простой генератор случайных данных.
4. Провести обучение/настройку выбранной модели.
5. Написать небольшую программу на Python, позволяющую продемонстрировать решение выбранной атомарной задачи.

*Тема 4. Формулировка практического задания.*

1. Для выбранного кейса провести экзистенциальное проектирование архитектуры. Выделить основные сервисы, определить зависимости между ними и способы коммуникации.
2. Для одного из основных сервисов выполнить проектирование на уровне диаграммы компонентов в соответствии с принципами чистой архитектуры. Выделить уровни

моделей данных (модели сущности, модели хранения, модели представления, транспортные модели и т.д.).

*Тема 5. Формулировка практического задания.*

1. Для выбранного кейса разработать реляционную модель данных, куда включить основные сущности предметной области и связи между ними (не менее 7 сущностей) в 3 НФ в формате ERD.
2. Разработать отображения этой модели данных в реляционной и нереляционной СУБД (PostgreSQL и MongoDB). Для нереляционной СУБД допускается провести деморализацию модели.

*Тема 6. Формулировка практического задания.*

1. Для выбранного кейса оценить объемы принимаемых, хранимых и обрабатываемых данных, интенсивность чтения и записи, требования по надежности и долговременности хранения разных видов информации.
2. Выбрать техническое решение СХД.
3. Подобрать на рынке доступные варианты СХД, оценить бюджет.

*Тема 7. Формулировка практического задания.*

1. Запустить 3 виртуалки с помощью доступной системы виртуализации (например, VirtualBox).
2. С помощью Docker Swarm развернуть СУБД MongoDB на 3 нодах и настроить репликацию.
3. Настроить запись данных в СУБД в соответствии с выбранным кейсом.
4. С помощью Prometheus и Grafana настроить сбор и визуализацию метрик СУБД MongoDB.
5. Выключить одну из виртуальных машин, проанализировать метрики СУБД, убедиться в доступности кластера.

*Тема 8. Формулировка практического задания.*

1. Для выбранного кейса подготовить проект модели угроз (в свободной форме).
2. Предложить комплекс мер по обеспечению информационной безопасности процессов приема, обработки и хранения данных.
3. Оценить необходимость обработки персональных данных. В случае наличия такой необходимости подготовить проект регламента процессов обработки персональных данных.



## 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. / Алпай-дин 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 заверченный 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. 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. **Какие характеристики имеет большой объем данных, который можно назвать Big Data?**
  - a) Объем от нескольких гигабайт до нескольких петабайт
  - b) Объем от нескольких килобайт до нескольких гигабайт
  - c) Объем от нескольких терабайт до нескольких эксабайт
  - d) Объем от нескольких мегабайт до нескольких терабайт
  
2. **Что такое облачные технологии?**
  - a) Технологии, которые позволяют обрабатывать данные на компьютерах, расположенных в разных местах.
  - b) Методы анализа данных, основанные на математических и статистических алгоритмах.
  - c) Технологии, которые позволяют использовать вычислительные ресурсы и хранение данных через Интернет.
  
3. **Какие инструменты используются для визуализации данных в области ИИ?**

- a) Matplotlib
- b) TensorFlow
- c) OpenCV
- d) PyTorch

**4. Какое из перечисленных принципов является основным для микросервисной архитектуры?**

- a) Монолитность
- b) Распределенность
- c) Централизованность
- d) Модульность

**5. Каппа-архитектура...**

- a) Реализует пакетную и потоковую обработку данных
- b) Реализует пакетную обработку данных
- c) Реализует потоковую обработку данных
- d) Реализует обработку данных из базы данных

**6. Какие операторы SQL используются для выборки данных?**

- a) SELECT, UPDATE, DELETE.
- b) SELECT, INSERT, CREATE.
- c) SELECT, FROM, WHERE.

**7. Какой инструмент используется для оркестрации контейнеров на нескольких узлах?**

- a) Docker Compose
- b) Kubernetes
- c) VMware

**8. Каким образом можно увеличить скорость чтения данных с диска?**

- a) Увеличить объем жесткого диска
- b) Использовать RAID-массивы
- c) Подключить дополнительный процессор
- d) Оптимизировать запросы к базе данных

**9. Что такое виртуализация?**

- a) Технология, позволяющая запускать несколько виртуальных машин на одном физическом сервере.
- b) Технология, позволяющая запускать приложения в контейнерах.
- c) Технология, позволяющая запускать виртуальные машины в облаке.

**10. Что такое контейнеризация?**

- a) Технология, позволяющая управлять и развертывать приложения в контейнерах.
- b) Технология, позволяющая запускать несколько виртуальных машин на одном физическом сервере.
- c) Технология, позволяющая запускать виртуальные машины в облаке.

7.2.3. Кейсы для итогового контроля.

Кейс №1 «Демографическая статистика»

Проблема: классические методы демографической статистики, основанные на учете данных переписей, ЗАГС и МВД, не отвечают современным потребностям государства и бизнеса в достоверных, оперативных и детализированных данных о количественном и качественном составе населения, естественном движении населения, миграции населения.

Направление решения: сбор и последующий анализ анонимизированных данных со всех имеющихся источников «Больших Данных» о населении – данных сотовых операторов, данных об использовании различных сетевых ресурсов с включенной геолокацией (соц.сети, геоинформационные системы), данных с камер видеонаблюдения на дорогах, данных об использовании общественного транспорта, данных о потреблении электроэнергии, заполненности образовательных учреждений и др.

Кейс №2 «ИТ-образование»

Проблема: с развитием современных форм образовательных процессов (онлайн, смешанное образование, децентрализованное образование, «образование всю жизнь» и др.) остро встает вопрос глубокой и непрерывной диагностики обучающихся для оценки их твердых и мягких навыков, мотивации и других качеств, прогнозирования успеваемости и отчислений, профориентации на основе их цифрового следа.

Направление решения: сбор и анализ различной информации о прохождении образовательного процесса обучающимися – данные об очных посещениях образовательного учреждения, данные об онлайн-подключениях к лекциям и вебинарам, данные о работе в LMS, данные о прохождении входных, промежуточных и итоговых контролей, данные о сдаче практических и домашних заданий, данные коммуникациях в текстовой форме (через системы обмена

сообщениями и электронной почтой), данные об использовании системы контроля версий (логины, коммиты, пуши, комментарии в code review), статистика пересдач и др.

#### Кейс №3 «Производственная безопасность»

Проблема: несмотря на то, что ценность жизни и здоровья человека значительно выросла за последние 100 лет, а в промышленности и производстве все активнее используются роботизированные решения, остается немало опасных производств, где по-прежнему задействован непосредственно человек.

Направление решения: сбор и анализ данных с различных датчиков, оперативная оценка ситуации на основе собранной информации. Примеры датчиков состояния работника: датчик пульса, датчик давления, ЭКГ, дыхания, температуры, усталости, падения. Примеры датчиков окружающей среды: датчик газа, дозиметр радиации, датчик температуры, датчик давления, датчик влажности. Примеры данных по контролю помещений: данные СКУД, видеоаналитика с камер видеонаблюдения, данные от систем сигнализации.

#### Кейс №4 «Персонализация ритейла»

Проблема: в современном мире высокой конкуренции, сетевой ритейл вынужден выстраивать персонализированную коммуникацию с каждым клиентом, чтобы иметь возможность выстраивать с ним доверительные и взаимовыгодные отношения. Однако этого почти не происходит, и часто персонализация ограничивается только подсказками на основе анализа просмотренного или купленного ранее в онлайн-магазине.

Направление решения: сбор и анализ всей возможной информации о клиенте и формирование персональных предложений с учетом долгосрочной стратегии. Примеры данных для сбора и анализа - истории покупок по карте лояльности, история покупок по одним и тем же платежным картам, поиск и анализ открытой информации из соц. сетей по клиенту, история посещений офлайн-магазина по камерам видеонаблюдения, активность клиента в онлайн-сервисах магазина и др.

#### Кейс №5 «Безопасность взлетно-посадочных полос гражданского аэропорта»

Проблема: взлетно-посадочная полоса (ВПП) всегда являлась зоной повышенной опасности. К многочисленным угрозам безопасности на ВПП в последние годы добавились и новые. Пример угроз безопасности на ВПП: люди, автотранспорт, животные, мусор, птицы, другие самолеты, гражданские беспилотники, строительные работы в непосредственной близости с объектом, землетрясения.

Направление решения: интегрированная система безопасности, учитывающая данные разной природы, поступающие в режиме реального времени от разных датчиков: радарные системы



обнаружения объектов в небе, оптические системы обнаружения самолетов, беспилотников и птиц, оптические системы обнаружения и сопровождения наземных подвижных объектов, оптические системы обнаружения подозрительных предметов на поверхности ВПП, датчики посторонних предметов, системы сигнализации, радио-маячки на автотранспорте аэропорта, датчики сейсмо-активности.

Кейс №6 «Анализ и прогнозирование пассажирских потоков в гражданском аэропорту»

Проблема: аэропорты сталкиваются с необходимостью обеспечения эффективного и комфортного обслуживания пассажиров, что становится сложнее из-за роста пассажирских потоков. Важно анализировать и прогнозировать пассажирский спрос для оптимизации расписания рейсов, обеспечения безопасности и рационального использования ресурсов.

Направление решения: Использование BigData для анализа и прогнозирования пассажирских потоков, сбор информации о пассажирах и их предпочтениях, а также факторах, влияющих на спрос. Данные могут быть получены из разных источников, таких как билетные системы, социальные сети, метеорологические данные, политическая и экономическая обстановка и многое другое. Машинное обучение и алгоритмы искусственного интеллекта позволяют выявлять закономерности и создавать точные прогнозы на основе анализа больших объемов данных.

Кейс №7 «Защита диких животных»

Проблема: влияние человека на окружающую среду уже привело к гибели значительной части биоразнообразия. Задача человечества – защитить оставшиеся виды, помочь им адаптироваться к изменившейся экологической обстановке и влиянию человека.

Направление решения: оснащение особо охраняемых диких животных неинвазивными датчиками, позволяющими отслеживать их состояние и быстро принимать меры в случае ухудшения здоровья и других негативных сценариев, оснащение датчиками среды обитания животных. Источники данных: неинвазивные датчики на животных (скорость и характер движения, пульс, давление, другие параметры), датчики перемещений, камеры видеонаблюдения, сейсмо-датчики для определения перемещений крупных стад животных, данные с беспилотников.

Кейс №8 «Предотвращение мошенничества в банковской сфере»

Проблема: мошенничество и кибератаки в банковской сфере представляет серьезную угрозу для клиентов и финансовых учреждений. Банки нуждаются в эффективных методах обнаружения и предотвращения мошенничества, чтобы защитить своих клиентов и свои активы.

Направление решения: Использование BigData для анализа и обработки данных транзакций, поведения пользователей, киберугроз и других аспектов, связанных с безопасностью в банковской сфере. Машинное обучение и алгоритмы искусственного интеллекта позволяют распознавать аномальные действия, выявлять сложные мошеннические схемы, предотвращать потери, связанные с мошенничеством и кибератаками. Источники данных – мессенджеры, социальные сети, электронная почта, банковские транзакции.

#### Кейс №9 «Предиктивная диагностика в медицине»

Проблема: современная медицина сталкивается с рядом проблем, таких как перегруженность медицинского персонала, неэффективное использование ресурсов и сложность прогнозирования и диагностики заболеваний на ранних стадиях развития.

Направление решения: Применение BigData в медицине позволяет собирать и анализировать данные из разных источников, таких как электронные медицинские карты, результаты диагностики, отзывы пациентов, клинические исследования и многое другое. Машинное обучение и алгоритмы искусственного интеллекта могут использоваться для оптимизации работы медицинских учреждений, улучшения качества предоставляемых услуг и создания предиктивных моделей для ранней диагностики и прогнозирования развития заболеваний. Такой подход способствует повышению эффективности медицинской помощи и улучшению качества жизни пациентов. Источники данных: истории болезней пациентов, различные виды биохимических анализов, данные неинвазивной диагностики (узи, рентген и др), измеряемые показатели различных систем организма (пульс, дыхание и др.).

#### Кейс №10 «Управление рисками»

Проблема: страховые компании сталкиваются с проблемами оценки и управления разнообразными рисками, связанные с предоставляемыми страховыми услугами. Эффективное управление рисками и создание персонализированных страховых услуг становится важным фактором конкурентоспособности страховых компаний.

Направление решения: Использование BigData для анализа и обработки данных о клиентах, страховых случаях, рисках и других факторах, связанных со страховыми услугами. Алгоритмы ИИ и машинное обучение позволяют страховым компаниям выявлять скрытые закономерности, определять риски, более точно анализировать и прогнозировать вероятность страховых случаев. Источники данных: кредитные истории клиентов, статистики доходов и расходов клиентов, информация о клиенте из социальных сетей, образование и профессия клиентов и т.д.

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  
**«Архитектура данных в области ИИ и BigData»**  
ДОПОЛНИТЕЛЬНОЙ ПРОГРАММЫ ПРОФЕССИОНАЛЬНОЙ  
ПЕРЕПОДГОТОВКИ **«Архитектор данных (Data Architect Pro)»**

Москва, 2023

## Оглавление

<b>1. ОБЩАЯ ХАРАКТЕРИСТИКА РП</b> .....	3
1.1. Цель .....	3
1.2. Планируемые результаты обучения .....	3
1.4. Перечень профессиональных компетенций в рамках имеющейся квалификации, качественное изменение которых осуществляется в результате обучения.....	4
1.5. Соответствие видов деятельности профессиональным компетенциям и их составляющих .....	5
<b>2. УЧЕБНЫЙ ПЛАН</b> .....	6
2.1. Категория слушателей РП.....	6
2.2. Общая трудоёмкость программы, аудиторная и самостоятельная работа.....	6
2.3. Форма обучения.....	6
2.4. Учебный план .....	6
<b>3. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК</b> .....	7
<b>4. РАБОЧАЯ ПРОГРАММА</b> .....	7
<b>5. УСЛОВИЯ РЕАЛИЗАЦИИ РП</b> .....	14
5.1. Организационные условия реализации РП.....	14
5.2. Педагогические условия реализации РП .....	14
5.3. Учебно-методическое обеспечение РП.....	14
5.4. Методические рекомендации.....	16
<b>6. ФОРМЫ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ РП</b> .....	17
<b>7. ОЦЕНОЧНЫЕ МАТЕРИАЛЫ ПРОМЕЖУТОЧНОЙ АТТЕСТАЦИИ</b> .....	18
7.1. Паспорт комплекта оценочных средств .....	18
7.2. Комплект оценочных средств.....	18

## 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 «**Архитектура данных в области ИИ и BigData**», могут быть допущены к освоению следующего - второго

уровня дополнительной программы профессиональной переподготовки «Архитектор данных (Data Architect 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. № 808 «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
Проектирование компьютерного программного обеспечения (D/03.6)			
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 среднее профессиональное и/или высшее образование.

### 2.2. Общая трудоёмкость программы, аудиторная и самостоятельная работа

Общая трудоёмкость программы 130 академических часа, из них 80 академических часов аудиторной работы, 48 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 «Архитектура данных в области ИИ и BigData» реализуется одним модулем.

№ п/п	Наименование темы, модуля	Форма Контроля	Всего, час	В том числе			
				Лекции	Практические занятия	Самостоятельная работа	Промежут. аттестация
1.	Проектирование архитектуры данных предприятия/проекта	Практ. задание	10	4	2	4	-
2.	Проектирование моделей данных	Практ. задание	16	6	4	6	-
3.	Проектирование процессов сбора и хранения данных	Практ. задание	16	6	4	6	-
4.	Проектирование баз, хранилищ и озер данных	Практ. задание	20	6	8	6	-
5.	Проектирование конвейеров данных	Практ. задание	18	4	8	6	-
6.	Проектирование представлений данных	Практ. задание	10	2	6	2	-
7.	MLOps	Практ. задание	10	2	2	6	-
8.	Подготовка и поддержка выборок данных для машинного обучения	Практ. задание	16	4	6	6	-
9.	Облачные технологии для больших данных и ИИ	Практ. задание	12	2	4	6	-
10.	Промежуточная аттестация	Зачет	2	-	-	-	2
	ИТОГО	-	130	36	44	48	2



### 3. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК

№ п/п	Наименование темы, модуля	1 неделя	2 неделя	3 неделя
1	Проектирование архитектуры данных предприятия/проекта			
2	Проектирование моделей данных			
3	Проектирование процессов сбора и хранения данных			
4	Проектирование баз, хранилищ и озер данных			
5	Проектирование конвейеров данных			
6	Проектирование представлений данных			
7	MLOps			
8	Подготовка и поддержка выборок данных для машинного обучения			
9	Облачные технологии для больших данных и ИИ			
10	Промежуточная аттестация			

Минимальный срок освоения программы – 3 недели.

### 4. РАБОЧАЯ ПРОГРАММА

#### 4.1. Рабочая программа модуля «Архитектура данных в области ИИ и BigData»

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	<b>Знать:</b> математическое моделирование. <b>Уметь:</b> осуществлять математическое и информационное моделирование. <b>Владеть:</b> 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	<b>Знать:</b> технические средства и среды сбора, хранения и обработки больших данных. <b>Уметь:</b> проводить аналитические работы с использованием технологий больших данных. <b>Владеть:</b> определение необходимых ресурсов для проведения аналитических работ.	Формы обучения: Фронтальная. Методы обучения: Лекция; Практические Работы; Самостоятельная работа.
ОПК-3	<b>Знать:</b> существующие и перспективные методы и программный инструментарий технологий больших данных. <b>Уметь:</b> проводить аналитические и поисковые исследования по тематике информационных технологий, технологий больших данных. <b>Владеть:</b> 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	<b>Знать:</b> принципы и методы управления защитой и обеспечением конфиденциальности больших данных. <b>Уметь:</b> проводить аналитические работы на основе технологий больших данных. <b>Владеть:</b> 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	<b>Знать:</b> современные и перспективные методы сбора, хранения и передачи данных из гетерогенных источников. <b>Уметь:</b> осуществлять разработку программно-аппаратных компонентов и систем. <b>Владеть:</b> 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. Проектирование архитектуры данных предприятия/проекта (10 часов)

Лекции (4 часа). Изучение тем:

- Постановка задачи проектирования архитектуры данных.
- Анализ информационной инфраструктуры бизнеса/проекта.

- Разработка корпоративной/проектной бизнес-модели, глоссария и жизненного цикла данных. DDD.
  - Data Driven, Data Centric, Data Governance.
  - Базовые элементы архитектуры данных: DataBase, Data WareHouse, Data Lake, Data Platform.
- Практическая работа (2 часа).
- Детальный разбор и проверка решений практических заданий.
  - Дискуссия, обмен мнениями по темам лекций и задания.
- Самостоятельная работа (4 часа). Решение практического задания.

## **Тема 2. Проектирование моделей данных (16 часов)**

Лекции (6 часов). Изучение тем:

- Корпоративная модель данных. EDM. Стандартизация работы с данными предприятия/проектной команды.
  - Внедрение практик культуры работы с данными. Задачи и процессы.
  - Классические модели данных и их приложения к задачам ИИ и BigData. Поддержка истории изменений. BEAM (Business Event Activity Modeling). Концептуальная, логическая и физическая модели данных
  - Модели Инмона и Кимбала, нормальные формы, Data Vault, Anchor modeling.
  - Версионность и историчность (Slowly Changing Dimensions).
- Практическая работа (4 часа).
- Обсуждение способов выполнения практического задания, с примерами и демонстрациями.
  - Детальный разбор и проверка решений практических заданий.
  - Дискуссия, обмен мнениями по темам лекций и задания.
- Самостоятельная работа (6 часов). Решение практического задания.

## **Тема 3. Проектирование процессов сбора и хранения данных (16 часов)**

Лекции (6 часов). Изучение тем:

- ETL и ELT.
  - DWH, DL, Data Hub, Data Virtualisation, Data Platform.
  - Пакетная и потоковая обработка данных. Lambda architecture, Кappa architecture.
  - Парсинг и преобразование данных.
  - Источники данных.
  - Качество данных. Нормализация и стандартизация. Поиск аномалий и выбросов.
- Практическая работа (4 часа).

- Обсуждение способов выполнения практического задания, с примерами и демонстрациями.
  - Детальный разбор и проверка решений практических заданий.
  - Дискуссия, обмен мнениями по темам лекций и задания.
- Самостоятельная работа (6 часов). Решение практического задания.

#### **Тема 4. Проектирование баз, хранилищ и озер данных (20 часов)**

Лекции (6 часов). Изучение тем:

- BigData, SQL и NoSQL.
- MPP SQL базы данных (Greenplum, ClickHouse). OLAP и OLTP.
- NoSQL базы данных (Cassandra, MongoDB, Elasticsearch, Neo4J, Redis).
- Экосистема Hadoop. Map Reduce. HDFS. SQL и NoSQL на Hadoop (Hive, HBase).
- Платформы данных. Apache Atlas.

Практическая работа (8 часов).

- Обсуждение способов выполнения практического задания, с примерами и демонстрациями.
  - Детальный разбор и проверка решений практических заданий.
  - Дискуссия, обмен мнениями по темам лекций и задания.
- Самостоятельная работа (6 часов). Решение практического задания.

#### **Тема 5. Проектирование конвейеров данных (18 часов)**

Лекции (4 часа). Изучение тем:

- Пакетная и потоковая обработка данных.
- Пакетная обработка. Hadoop, Zeppelin, Hive, Spark, Luigi, Airflow.
- Потоковая обработка. Apache Nifi. Apache Kafka. Spark Streaming, Apache Storm.

Практическая работа (8 часов).

- Обсуждение способов выполнения практического задания, с примерами и демонстрациями.
  - Детальный разбор и проверка решений практических заданий.
  - Дискуссия, обмен мнениями по темам лекций и задания.
- Самостоятельная работа (6 часов). Решение практического задания.

#### **Тема 6. Проектирование представлений данных (10 часов)**

Лекции (2 часа). Изучение тем:

- Задачи BI и визуализации данных.
- PowerBI, Tableau, Qlick, Yandex DataLens.

- Визуальная аналитика многомерных данных. Снижение размерности наборов данных. Методы отображения многомерных данных.
- Визуальные языки представления данных.
- Основные принципы онтологического инжиниринга. Графовое, RDF- и OWL-представление данных. SPARQL.  
Практическая работа (6 часов).
- Обсуждение способов выполнения практического задания, с примерами и демонстрациями.
- Детальный разбор и проверка решений практических заданий.
- Дискуссия, обмен мнениями по темам лекций и задания.  
Самостоятельная работа (2 часа). Решение практического задания.

### **Тема 7. MLOps (10 часов)**

Лекции (2 часа). Изучение тем:

- Понятие MLOps (DevOps, ML, инженерия данных).
- Архитектурный дизайн MLOps.
- Специфика использования классических DevOps-инструментов в задачах MLOps.
- Работа с ML-системами в K8S (Kubeflow)
- Организация жизненного цикла ML-систем (MLFlow).

Практическая работа (2 часа).

- Обсуждение способов выполнения практического задания, с примерами и демонстрациями.
- Детальный разбор и проверка решений практических заданий.
- Дискуссия, обмен мнениями по темам лекций и задания.  
Самостоятельная работа (6 часов). Решение практического задания.

### **Тема 8. Подготовка и поддержка выборок данных для машинного обучения (16 часов)**

Лекции (4 часа). Изучение тем:

- Сбор, подготовка, разметка и валидация выборок (датасетов).
- Хранение выборки и улучшение качества выборки.
- Нормализация и стандартизация. Поиск аномалий и выбросов.
- Методы обработки и визуализации.
- Разбалансированные датасеты и методы их балансировки.
- Доверительные интервалы. Обработка категориальных признаков (Sklearn).
- Оценка качества данных выборки (Spark Deequ)
- Формирование и поддержание каталог данных выборок (Amundsen).

Практическая работа (6 часов).

- Обсуждение способов выполнения практического задания, с примерами и демонстрациями.
- Детальный разбор и проверка решений практических заданий.
- Дискуссия, обмен мнениями по темам лекций и задания.

Самостоятельная работа (6 часов). Решение практического задания.

## **Тема 9. Облачные технологии для больших данных и ИИ (14 часов)**

Лекции (2 часа). Изучение тем:

- Сбор, подготовка, разметка и валидация выборок (датасетов).
- Хранение выборки и улучшение качества выборки.
- Нормализация и стандартизация. Поиск аномалий и выбросов.
- Методы обработки и визуализации.
- Разбалансированные датасеты и методы их балансировки.
- Доверительные интервалы. Обработка категориальных признаков (Skilearn).
- Оценка качества данных выборки (Spark Deequ)
- Формирование и поддержание каталог данных выборок (Amundsen).

Практическая работа (4 часа).

- Обсуждение способов выполнения практического задания, с примерами и демонстрациями.
- Детальный разбор и проверка решений практических заданий.
- Дискуссия, обмен мнениями по темам лекций и задания.

Самостоятельная работа (8 часов). Решение практического задания.

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

*Тема 1. Формулировка практического задания.*

1. Для выбранного кейса провести анализ задачи сбора, анализа и хранения данных.
2. Разработать глоссарий и модель жизненного цикла данных.
3. Оценить какие базовые элементы архитектуры данных могут применяться в рассматриваемом кейсе и для каких видов сведений.

*Тема 2. Формулировка практического задания.*

1. Для выбранного кейса выбрать подход моделирования данных и согласовать с преподавателем.
2. Разработать модель данных выбранного кейса. При этом покрыть:

- a. Сущности предметной области
- b. Собираемые данные от разных источников (не менее 5 видов данных).

*Тема 3. Формулировка практического задания.*

1. Для выбранного кейса выбрать архитектуру процесса сбора и хранения
2. Спроектировать процесса сбора и хранения в зависимости от выбранной архитектуры и с учетом всех видов сведений.
3. Разработать уточненные модели данных для каждого этапа процесса.

*Тема 4. Формулировка практического задания.*

1. Выбрать технические решения для реализации разработанной ранее модели данных для выбранного кейса, согласовать его с преподавателем.
2. Для каждого технического решения разработать схему данных.
3. Настроить схемы данных в выбранных технических решениях.

*Тема 5. Формулировка практического задания.*

1. Провести выбор технического решения для пакетной или потоковой обработки данных, согласовать его с преподавателем.
2. Развернуть и настроить выбранное решение.
3. Настроить один поток данных (при необходимости использовать генераторы данных вместо реальных данных), с любыми двумя операциями обработки (например, нормализация и кластеризация данных).

*Тема 6. Формулировка практического задания.*

1. Для выбранного кейса спроектировать представление обработанных данных (как минимум, для одного вида сведений).
2. Выбрать техническое решение для визуализации представления.
3. Настроить представление и визуализацию.

*Тема 7. Формулировка практического задания.*

1. Для выбранного кейса, на основе уже разработанных программных модулей спроектировать жизненный цикл разрабатываемой системы.
2. Развернуть и настроить MLFlow (возможно использование облачного решения).

*Тема 8. Формулировка практического задания.*

1. Выбрать один из видов сведений из выбранного кейса, на основе которого сформировать выборку данных (датасет).
2. Создать каталог данных выборок.
3. Оценить качество тестовой выборки.

*Тема 9. Формулировка практического задания.*

1. Провести выбор облачной платформы для реализации решений выбранного кейса (выбрать платформу с бесплатным пробным периодом).
2. Перенести на облачную платформу часть разработанных программных модулей, по согласованию с преподавателем.

## **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. / Алпай-дин 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 заверченный 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. 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. Что такое конфиденциальность данных?

- a) Это уровень доступа к данным, который ограничивает их использование только специальными пользователями.
- b) Это характеристика данных, которая гарантирует их защиту от несанкционированного доступа.
- c) Это способность данных сохранять свои свойства и качество при передаче и обработке.

### 2. Что такое шифрование данных?

- a) Это процесс обнаружения и устранения вредоносных программ на компьютере.
- b) Это комплекс мер, направленных на предотвращение угроз безопасности, связанных с вредоносными программами.
- c) Это процесс преобразования информации в нечитаемый для посторонних вид.

### 3. ETL..

- a) Entity Transaction Layer
- b) Extraction Transformation Loading
- c) Extract Transform Load
- d) Extract Typed Layer

### 4. Что такое ER-модель данных?

- a) Модель, описывающая бизнес-процессы организации
- b) Модель, описывающая процессы сбора данных
- c) Модель, описывающая сущности и их связи в некоторой предметной области
- d) Модель, описывающая процессы аналитики данных

### 5. Что такое озеро данных?

- a) Система хранения данных, используемая для операционной обработки данных в режиме реального времени.
- b) Система хранения данных, используемая для резервного копирования информации.
- c) Централизованное хранилище данных, используемое для хранения неструктурированных и полуструктурированных данных.

**6. Что такое репликация данных?**

- a) Процесс сохранения данных на нескольких серверах для повышения отказоустойчивости и быстрого доступа к данным.
- b) Процесс копирования данных в новую базу данных для обеспечения ее безопасности.
- c) Процесс удаления устаревших данных из базы данных.

**7. Какие преимущества имеет использование конвейера данных?**

- a) Увеличение скорости обработки данных, снижение затрат на разработку.
- b) Увеличение объема хранимых данных, улучшение качества данных.
- c) Упрощение работы с данными, повышение безопасности.

**8. Какие задачи входят в MLOps?**

- a) Разработка моделей машинного обучения
- b) Создание датасетов
- c) Тестирование, деплой и мониторинг моделей машинного обучения

**9. Что такое облачные технологии для больших данных?**

- a) Это способность обработки данных в большом масштабе с помощью вычислительных мощностей, предоставляемых облачными провайдерами.
- b) Это способность обработки данных на локальных вычислительных машинах без использования облачных провайдеров.
- c) Это способность обработки данных только в масштабах малого и среднего бизнеса.

**10. Что включает в себя подготовка данных для машинного обучения?**

- a) Сбор, очистка, трансформация и анализ данных
- b) Выбор алгоритма машинного обучения и настройка его параметров
- c) Создание визуализаций данных и отчетов

**7.2.3. Кейсы для итогового контроля.**

**Кейс №1 «Анализ социальных медиа»**

Проблема: В современном мире социальные медиа играют значительную роль в формировании общественного мнения, а также могут оказывать влияние на рынок и бизнес. Оценка общественного мнения и изучение различных социальных процессов более невозможна без анализа больших объемов данных, генерируемых пользователями социальных сетей.

Направление решения: Использование машинного обучения для анализа информации из социальных медиа и социальных сетей. Выявление трендов при помощи алгоритмов ИИ,

определение влияния социальных тенденций на рынок и потребителей. Источники данных: посты в группах и каналах социальных сетей и мессенджеров, комментарии пользователей, описания пользователей, изображения, загружаемые пользователями, комментарии пользователей на новостных ресурсах.

#### Кейс №2 «Анализ рынка труда»

Проблема: Понимание текущих и будущих потребностей рынка труда в кадрах является ключевым фактором для формирования образовательных программ, развития инфраструктуры для обучения и переобучения, а также принятия решений со стороны работодателей и соискателей работы. Традиционные методы анализа рынка труда могут быть недостаточно точными и оперативными для прогнозирования изменений на рынке.

Направление решения: Сбор и анализ больших данных о вакансиях, зарплате, навыках и спросе на различные профессии с использованием источников, таких как онлайн-биржи труда, социальные сети, базы данных кадровых агентств и данные корпоративных систем управления персоналом. Использование алгоритмов машинного обучения и искусственного интеллекта для анализа данных, выявления трендов и прогнозирования будущих потребностей на рынке труда может помочь государственным структурам, образовательным учреждениям и работодателям принимать более обоснованные решения.

#### Кейс №3 «Оптимизация логистики»

Проблема: Логистические процессы являются важной частью многих отраслей. Оптимизация процессов, связанных с логистикой, может привести к снижению затрат, повышению эффективности и уровня обслуживания клиентов.

Направление решения: Сбор и анализ BigData о перемещении транспортных средств, товаров, маршрутов, времени доставки, погодных условий и других факторов, влияющих на логистические процессы. Использование алгоритмов ИИ и Машинного обучения для оптимизации маршрутов и распределения ресурсов, а также балансировки логистических маршрутов. Источники данных: логистические маршруты, информация о трафике из геоинформационных систем и систем видеонаблюдения, информация о логистических и экономических связях между регионами, информация от транспондеров, геометки в социальных сетях и др.

#### Кейс №4 «Анализ данных о транспортных потоках на дорогах»

Проблема: В современном мире автомобили заняли прочное место в повседневной жизни человека, и при проектировании дорог и движения часто возникает проблема оптимизации временных интервалов, на которых основана работа светофоров. Интервалы переключения сигналов зачастую фиксированные и не меняются вне зависимости от загруженности трасс. Для создания оптимальных режимов работы светофоров могут тратиться большие ресурсы.

Направление решения: Сбор и анализ BigData о трафике с учетом режимов работы светофоров. Использование алгоритмов ИИ и Машинного обучения для оптимизации интервалов переключения сигналов светофоров в зависимости от загруженности участка дороги в конкретный момент времени. Источники данных: данные о трафике из геоинформационных систем, данные с камер видеонаблюдения о трафике и количестве пешеходов, статистика загрузки дорог и переходов в зависимости от времени суток и времени года.

#### Кейс №5 «Расписание в университете»

Проблема: Перед началом учебного года/семестра деканаты при участии кафедр занимаются составлением расписания для обучающихся, но часто бывает так, что расписание проектируется таким образом, что у одних студентов дни бывают более нагруженными, чем другие, а у других появляются «окна» в расписании. Также важной проблемой является доступность преподавателей-совместителей, которые могут вести занятия не в любое время.

Направление решения: сбор и анализ различной информации об учебных программах для каждой кафедры и каждого курса, которые позволят в дальнейшем спроектировать максимально «эффективное» расписание, равномерно распределяющее нагрузку по дням и предусматривающее наименьшее возможное количество «окон» в нем, учитывающее интересы студентов, преподавателей и учебного процесса. Источники данных: учебные планы, данные о типах занятий (теоретические, практические, физкультура), данные о загрузке преподавателей, статистика о эффективности определенных конфигураций занятий в расписании, данные о расположении аудиторий и корпусов.

#### Кейс №6 «Поиск внеземной жизни»

Проблема: В последние десятилетия человечество все больше ищет планеты в различных системах, которые могут быть пригодны для жизни и на которых потенциально может быть жизнь в простейшей форме. Однако, данные, поступающие с телескопов и позволяющие заглядывать в самые дальние уголки вселенной, очень избыточны и сложны в интерпретации, а их анализ занимает много времени

Направление решения: Использование машинного обучения для анализа информации, поступающей с радиотелескопов. Выявление небесных тел, которые потенциально могут иметь



на своей поверхности воду в жидком агрегатном состоянии, и анализ данных об атмосфере, близости к светилам и др. с целью упрощения поиска планет, пригодных для формирования жизни. Источники информации: данные с телескопов в различных радио-диапазонах.

#### Кейс №7 «Безопасность в общественных местах»

Проблема: Преступники, совершившие преступления почти всегда скрываются с места преступления, что отяжеляет их дальнейший поиск с целью ареста и расследования преступления. Несмотря на наличие в городской среде большого количества датчиков в виде камер наружного наблюдения, ручной поиск правонарушителя может быть очень трудоемким и, в случае небольшого правонарушения, нецелесообразным.

Направление решения: Использование машинного обучения для анализа информации с камер наружного наблюдения и других источников для поиска правонарушителей. Источники данных: камеры видеонаблюдения на улицах, в зданиях, данные из социальных сетей, данные банковских транзакций, данные мобильных сетей связи.

#### Кейс №8 «Рекомендательная система для сервиса прослушивания музыки»

Проблема: Сервисы для прослушивания музыки в последние 10 лет набрали большую популярность, причем популярными становятся те сервисы, которые не только содержат в себе большую медиатеку, но и предлагают пользователям лучшую рекомендательную систему, основанную на их вкусах.

Направление решения: сбор и анализ информации о всех клиентах сервисов, включая интересующие их жанры, которые они указали при регистрации, и ту музыку, которую они чаще всего слушают, и формирование рекомендательной системы на основе проанализированных данных. Источники данных: история прослушиваний, предпочтения клиентов, схожесть клиентов, данные из социальных сетей.

#### Кейс №9 «Самые интересные места для посещения»

Проблема: В крупных городах существует множество мест и заведений, которые, если судить по различным туристическим отзывам в интернете, стоит посетить. И туристу, приехавшему в новый город, приходится доверяться этим инструкциям при планировании своего времяпрепровождения.

Направление решения: сбор информации о перемещениях и продолжительности нахождения жителей города с целью создания максимально релевантных инструкций для путешественников. Так как местные жители лучше знают, куда лучше сходить и чем заняться, данная информация будет наиболее востребованной при выполнении данной работы.

Источники данных: информация о трафике горожан из геоинформационных систем и систем видеонаблюдения, отзывы в социальных сетях и сайтах агрегаторов.

Кейс №10 «Поиск подверженных износу деталей у автомобилей»

Проблема: Каждая модель любого автомобиля имеет характерные дефекты, которые потом приводят к износу его деталей.

Направление решения: сбор и анализ информации о самых частых выходящих из строя деталей для каждой модели различных марок автомобилей. Полученная система позволит при диагностике и прохождении технического обслуживания с большей вероятностью находить дефекты и устранять их. Источники данных: информация о комплектующих автомобилей, данные из дилерских и ремонтных центров, отзывы в социальных сетях, информация от производителей.