

МИНИСТЕРСТВО ОБРАЗОВАНИЯ И НАУКИ РОССИЙСКОЙ ФЕДЕРАЦИИ
федеральное государственное бюджетное образовательное учреждение
высшего образования
«Московский государственный технологический университет «СТАНКИН»

УТВЕРЖДАЮ:
Проректор по образовательной
деятельности и молодежной политике
_____ /Бильчук М.В./
«____ » _____ 202 ____ г.
M.п.

ДОПОЛНИТЕЛЬНАЯ ПРОФЕССИОНАЛЬНАЯ ПРОГРАММА

Аналитика данных и методы искусственного интеллекта

(наименование дополнительной профессиональной программы)

профессиональная переподготовка

(повышение квалификации/ профессиональная переподготовка)

Категория слушателей:

Обучающиеся по направлениям подготовки,
отнесенными к ИТ-сфере

Москва, 2024

Рабочая программа дополнительной профессиональной программы профессиональной переподготовки «Аналитика данных и методы искусственного интеллекта».

Рабочая программа предназначена для обучающихся цифровой кафедры по направлениям, отнесенных к ИТ-сфере

Образовательная программа соответствует:

- Федеральному закону от 29 декабря 2012 г. № 273-ФЗ «Об образовании в Российской Федерации»;
- Постановлению Правительства Российской Федерации от 11 октября 2023 г. № 1678 «Об утверждении Правил применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ»;
- приказу Минобрнауки России от 1 июля 2013 г. № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам»;
- постановление Правительства РФ от 11 октября 2023 г. № 1678 «Об утверждении Правил применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ»;
- Профессиональный стандарт № 06.042 «Специалист по большим данным», утвержденный приказом Министерства труда и социальной защиты Российской Федерации от 06 июля 2020 № 405н.

© ФГБОУ ВО «МГТУ «СТАНКИН», 2024

Образовательная программа повышения квалификации «Аналитика данных и методы искусственного интеллекта» включает пояснительную записку, основные требования к уровню освоения содержания дисциплины, учебный и учебно-тематический план, содержание разделов программы по модулям с кратким содержанием тематики, организационно-педагогические условия реализации программы, формы аттестации и оценочные материалы для обучения слушателей по программе, рассчитанной на 252 (двести пятьдесят два) учебных часа.

СОДЕРЖАНИЕ

Пояснительная записка

РАЗДЕЛ 1. Характеристика программы «Аналитика данных и методы искусственного интеллекта»

- 1.1. Цель реализации программы
- 1.2. Компетенции, подлежащие формированию по итогам обучения
- 1.3. Требования к уровню подготовки поступающего на обучение, необходимому для освоения программы
- 1.4. Трудоемкость обучения
- 1.5. Форма и режим обучения

РАЗДЕЛ 2. Содержание программы «Аналитика данных и методы искусственного интеллекта»

- 2.1. Учебный план программы
- 2.2. Календарный учебный график
- 2.3. Рабочие программы дисциплин (модулей)
- 2.4. Рабочая программы стажировки

РАЗДЕЛ 3. Оценка качества освоения программы «Аналитика данных и методы искусственного интеллекта»

РАЗДЕЛ 4. Организационно-педагогические условия реализации программы «Аналитика данных и методы искусственного интеллекта»

- 4.1. Материально-технические условия и организационное обеспечение реализации дополнительной профессиональной программы профессиональной переподготовки
- 4.2. Учебно-методическое обеспечение реализации дополнительной профессиональной программы профессиональной переподготовки
- 4.3. Кадровое обеспечение образовательного процесса по программе

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

1. Дополнительная профессиональная программа (программа профессиональной переподготовки) ИТ-профиля «Аналитика данных и методы искусственного интеллекта» (далее – Программа) разработана в соответствии с нормами Федерального закона РФ от 29 декабря 2012 года № 273-ФЗ «Об образовании в Российской Федерации», с учетом требований приказа Минобрнауки России от 1 июля 2013 г. № 499 «Об утверждении Порядка организации и осуществления образовательной деятельности по дополнительным профессиональным программам», с изменениями, внесенными приказом Минобрнауки России от 15 ноября 2013 г. № 1244 «О внесении изменений в Порядок организации и осуществления образовательной деятельности по дополнительным профессиональным программам, утвержденный приказом Министерства образования и науки Российской Федерации от 1 июля 2013 г. № 499»; постановления Правительства РФ от 11 октября 2023 г. № 1678 «Об утверждении Правил применения организациями, осуществляющими образовательную деятельность, электронного обучения, дистанционных образовательных технологий при реализации образовательных программ», паспорта федерального проекта «Развитие кадрового потенциала ИТ-отрасли» национальной программы «Цифровая экономика Российской Федерации»; постановления Правительства Российской Федерации от 13 мая 2021 г. № 729 «О мерах по реализации программы стратегического лидерства «Приоритет-2030»; приказа Минцифры России от 29 декабря 2023 г. № 1180 «Об утверждении методик расчета показателей федеральных проектов «Развитие кадрового потенциала ИТ-отрасли» и «Обеспечение доступа в Интернет за счет развития спутниковой связи» национальной программы «Цифровая экономика Российской Федерации», а также внесении изменений в некоторые приказы Министерства цифрового развития, связи и массовых коммуникаций Российской Федерации об утверждении методик расчета показателей федеральных проектов национальной программы «Цифровая экономика Российской Федерации» (далее – приказ Минцифры России № 1180); федерального государственного образовательного стандарта 09.03.01 Информатика и вычислительная техника (уровень бакалавриата), утвержденного приказом Минобрнауки России от 12 января 2016 г. № 5, (далее вместе – ФГОС ВО)), а также профессионального стандарта профессионального стандарта № 06.042 «Специалист по большим данным», утвержденный приказом Министерства труда и социальной защиты Российской Федерации от 06 июля 2020 № 405н.

2. Профессиональная переподготовка заинтересованных лиц (далее – Слушатели), осуществляемая в соответствии с Программой (далее – Подготовка), имеющей отраслевую направленность «Обрабатывающая промышленность», проводится в ФГБОУ ВО «Московский государственный технологический университет «СТАНКИН» (далее – Университет) в соответствии с учебным планом в очной форме обучения.

3. Разделы, включенные в учебный план Программы, используются для последующей разработки календарного учебного графика, учебно-тематического плана, рабочей программы, оценочных и методических материалов. Перечисленные документы разрабатываются Университетом самостоятельно, с учетом актуальных положений законодательства об образовании, законодательства в области информационных технологий и смежных областей знаний ФГОС ВО и профессионального стандарта 06.042 «Специалист по большим данным», утвержденный приказом Министерства труда и социальной защиты Российской Федерации от 06 июля 2020 № 405н.

4. Программа регламентирует требования к профессиональной переподготовке в области работы с данными с учетом различных условий с целью сбора, анализа, обработки и представления разнородных данных в цифровой форме, разработки алгоритмов и моделей машинного обучения и нейросетей, а также приобретение по итогам прохождения ДПП ПП новой квалификации «Специалист по искусственному интеллекту».

Срок освоения Программы составляет 252 академических часа.

К освоению Программы в рамках проекта допускаются лица:

- получающие высшее образование по очной (очно-заочной) форме, лица, освоившие основную профессиональную образовательную программу (далее – ОПОП ВО) бакалавриата – в объеме не менее первого курса (бакалавры 2-го курса), обучающихся на направлениях, относящихся к ИТ-сфере.

5. Область профессиональной деятельности «Создание и применение технологий больших данных».

В рамках реализации образовательной программы будут использованы различные организационные **формы обучения**, такие как:

1. индивидуальные формы обучения – самостоятельное выполнение заданий в рамках практических занятий, консультации с наставниками;
2. очные и дистанционные формы обучения – лекции, практические занятия.

**РАЗДЕЛ 1. ХАРАКТЕРИСТИКА ПРОГРАММЫ
«АНАЛИТИКА ДАННЫХ И МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА»**

1.1. ЦЕЛЬ РЕАЛИЗАЦИИ ПРОГРАММЫ

Целью подготовки слушателей по Программе является получение компетенции, необходимой для выполнения нового вида профессиональной деятельности в области информационных технологий обучающихся по специальностям и направлениям подготовки, отнесенными к ИТ-сфере; приобретение новой квалификации «Специалист по искусственному интеллекту».

1.2. КОМПЕТЕНЦИИ, ПОДЛЕЖАЩИЕ ФОРМИРОВАНИЮ ПО ИТОГАМ ОБУЧЕНИЯ (образовательные результаты по программе)

Получение дополнительной ИТ-квалификации «Специалист по искусственному интеллекту» для обучающихся по направлениям, отнесенных к ИТ-сфере обеспечивается формированием приведённых в таблице цифровых компетенций:

Наименование сферы	ID и наименование компетенции	Инструменты профессиональной деятельности	Целевой уровень формирования компетенций в Программе			
			Минимальный (исходный)	Базовый	Продвинутый	Экспертный
Искусственный интеллект и машинное обучение	ПК-167 Разрабатывает и реализует архитектуру ансамбля моделей	Python, Tensor Flow, Keras, Scikit-learn			Самостоятельно составляет и реализует архитектуру ансамбля моделей, прибегая к экспертной консультации.	
Искусственный интеллект и машинное обучение	ПК -170 Осуществляет сбор и подготовку данных для обучения моделей искусственного интеллекта	Python, Tensor Flow, Keras, Scikit-learn			Использует методы редукции размерности элементов набора данных и их предварительной статистической обработки, разметки структурированных и неструктурированных данных	
Искусственный интеллект и машинное обучение	ПК-171 Разрабатывает модули машинного обучения (МО) для решения задач	Python, Tensor Flow, Keras, Scikit-learn			Применяет методы и критерии оценки качества моделей МО, определяет критерии и метрики оценки результатов моделирования. Самостоятельно решает задачи анализа, прогнозирования, планирования, синтеза и принятия решений	

1.3. ТРЕБОВАНИЯ К УРОВНЮ ПОДГОТОВКИ ПОСТУПАЮЩЕГО НА ОБУЧЕНИЕ, НЕОБХОДИМОМУ ДЛЯ ОСВОЕНИЯ ПРОГРАММЫ:

К обучению по Программе допускаются обучающиеся по очной или по очно-заочной форме за счет бюджетных средств или по договорам об оказании платных образовательных услуг, освоившие программы бакалавриата в объеме не менее 1 курса (бакалавры 2 курса) – 09.00.00 Информатика и вычислительная техника, 01.00.00 Математика и механика, отнесённым к ИТ-сфере.

1.4. ТРУДОЕМКОСТЬ ОБУЧЕНИЯ

Трудоемкость учебной работы слушателя по данной программе – 252 ак. часа, включая все виды аудиторной и внеаудиторной (самостоятельной) работы.

Вид учебной работы	Всего часов
Общий объем программы	252
В том числе:	
Лекционные занятия (Л)	62
Практические занятия (ПЗ)	112
Текущие формы контроля (ТК)	8
Стажировка	16
Самостоятельная работа, включая работу по подготовке к промежуточному и итоговому контролю (С)	46
Выполнение итоговой аттестационной работы	8

1.5. ФОРМА И РЕЖИМ ОБУЧЕНИЯ

Форма обучения: *очная с применением дистанционных технологий*

Режим занятий: 6 академических часа в день.

Реализация Программы должна обеспечить получение компетенции, необходимой для выполнения нового вида профессиональной деятельности в области информационных технологий; приобретение новой квалификации Специалист по искусственному интеллекту

Учебный процесс организуется с применением инновационных технологий и методик обучения, способных обеспечить получение слушателями знаний, умений и навыков в области разработки компьютерного программного обеспечения.

Слушатель, прошедший обучение по дополнительной профессиональной программе профессиональной переподготовки «Аналитика данных и методы искусственного интеллекта» должен:

Наименование компетенции	ПК-167 Самостоятельно составляет и реализует архитектуру ансамбля моделей, прибегая к экспертной консультации
ЗНАТЬ	З 167.1 основные методы, используемые для решения задач областей искусственного интеллекта; З 167.2 современный опыт использования анализа больших данных в промышленности; З 167.3 основы моделей искусственных нейронных сетей
УМЕТЬ	У 167.1 разрабатывать и оценивать модели градиентного бустинга; У 167.2 разрабатывать и оценивать модели искусственных

	нейронных сетей; У 167.3 адаптировать и развертывать модели в промышленной среде;
ВЛАДЕТЬ	В 167.1 навыками реализации базовые архитектуры и модели искусственных нейронных сетей и градиентного бустинга
Наименование компетенции	<ul style="list-style-type: none"> ПК -170 Использует методы редукции размерности элементов набора данных и их предварительной статистической обработки, разметки структурированных и неструктурированных данных
ЗНАТЬ	<p>З 170.1 области применения технологий искусственного интеллекта в промышленности;</p> <p>З 170.2 принципы поиска и анализа информации, необходимой для решения поставленной задачи, в том числе с применением аналитики данных и машинного обучения;</p> <p>З 170.3 нейронные сети: полносвязные, свёрточные и рекуррентные нейронные сети, методы обучения нейронных сетей, нейросетевые методы понижения размерности;</p> <p>З 170.4 анализ изображений, анализ временных рядов;</p>
УМЕТЬ	<p>У 170.1 определять требования к данным из гетерогенных источников;</p> <p>У 170.2 использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников, в том числе в режиме реального времени;</p> <p>У 170.3 производить очистку данных для проведения аналитических работ;</p> <p>У 170.4 проводить интеграцию и преобразование больших объемов данных;</p> <p>У 170.5 программировать на языке высокого уровня Python: для статистической обработки данных и работы с графикой, для работы с разрозненными фрагментами данных в больших массивах, для работы с базами структурированных и неструктурированных данных</p>
ВЛАДЕТЬ	<p>В 170.1 навыками критического отбора данных и методами редукции размерности элементов набора данных</p> <p>В 170.2 навыками проверки данных на целостность и непротиворечивость.</p> <p>В 170.3 навыками использования методов поиска данных и достоверные источники данных, предварительной статистической обработки, разметки структурированных и неструктурных данных</p>
Наименование компетенции	<ul style="list-style-type: none"> ПК-171 Разрабатывает модули машинного обучения (МО) для решения задач
ЗНАТЬ	<p>З 171.1 основы языка Python и основные алгоритмические конструкции;</p> <p>З 171.2 методы интерпретации и визуализации анализа технологических данных</p>
УМЕТЬ	<p>У 171.1 осуществлять интеграцию и преобразование данных в ходе работ по анализу технологических данных;</p> <p>У 171.2 применять различные алгоритмы машинного обучения при аналитике технологических данных.</p>
ВЛАДЕТЬ	В 171.1 навыками применения методов и критериев оценки качества моделей МО, определять критерии и метрики оценки

	результатов моделирования. В 171.2 навыкам решения задач анализа, прогнозирования, планирования, синтеза и принятия решений для машиностроительных предприятий
--	---

Структура Программы

Структура Программы регулирует образовательные траектории обучающихся, последовательность освоения структурных элементов (разделов) Программы, соответственно, последовательность формирования всех образовательных результатов.

Структурные элементы (разделы Программы)	Образовательные результаты
Модуль 1. Введение в - бизнес -аналитику и искусственный интеллект с применением Python для анализа данных	Знает: З 171.1 основы языка Python и основные алгоритмические конструкции; З 171.2 методы интерпретации и визуализации анализа технологических данных Умеет: У 171.1 осуществлять интеграцию и преобразование данных в ходе работ по анализу технологических данных У 171.2 применять различные алгоритмы машинного обучения при анализе технологических данных. Владеет: В 171.1 навыками применения методов и критериев оценки качества моделей МО, определять критерии и метрики оценки результатов моделирования. В 171.2 навыкам решения задач анализа, прогнозирования, планирования, синтеза и принятия решений для машиностроительных предприятий компетенции ПК-171 Разрабатывает модули машинного обучения (МО) для решения задач
Модуль 2. Методы искусственного интеллекта для анализа табличных данных	Знает: З 170.1 области применения технологий искусственного интеллекта в промышленности; З 170.3 нейронные сети: полносвязные, свёрточные и рекуррентные нейронные сети, методы обучения нейронных сетей, нейросетевые методы понижения размерности; З 170.4 анализ изображений, анализ временных рядов Умеет: У 170.3 производить очистку данных для проведения

Структурные элементы (разделы Программы)	Образовательные результаты
	<p>аналитических работ;</p> <p>У 170.4 проводить интеграцию и преобразование больших объемов данных;</p> <p>У 170.5 программировать на языке высокого уровня Python: для статистической обработки данных и работы с графикой, для работы с разрозненными фрагментами данных в больших массивах, для работы с базами структурированных и неструктурированных данных</p> <p>Владеет:</p> <p>В 170.1 навыками критического отбора данных и методами редукции размерности элементов набора данных</p> <p>В 170.2 навыками проверки данных на целостность и непротиворечивость.</p>
	<p>компетенции</p> <p>ПК -170 Осуществляет сбор и подготовку данных для обучения моделей искусственного интеллекта</p>
	<p>Знает:</p> <p>З 167.1 основные методы, используемые для решения задач областей искусственного интеллекта;</p> <p>З 167.2 современный опыт использования анализа больших данных в промышленности;</p> <p>З 167.3 основы моделей искусственных нейронных сетей</p> <p>Умеет:</p> <p>У 167.1 разрабатывать и оценивать модели градиентного бустинга;</p> <p>У 167.2 разрабатывать и оценивать модели искусственных нейронных сетей;</p> <p>У 167.3 адаптировать и развертывать модели в промышленной среде;</p> <p>Владеет:</p> <p>В 167.1 навыками реализации базовые архитектуры и модели искусственных нейронных сетей и градиентного бустинга</p> <p>компетенции</p> <p>ПК -167 Разрабатывает и реализует архитектуру ансамбля моделей</p>

Структурные элементы (разделы Программы)	Образовательные результаты
<p>Модуль 3. Современные озера и хранилища данных, аналитика больших данных и методы искусственного интеллекта</p>	<p>Знает:</p> <p>З 170.1 области применения технологий искусственного интеллекта в промышленности;</p> <p>З 170.2 принципы поиска и анализа информации, необходимой для решения поставленной задачи, в том числе с применением аналитики данных и машинного обучения;</p> <p>З 170.3 нейронные сети: полносвязные, свёрточные и рекуррентные нейронные сети, методы обучения нейронных сетей, нейросетевые методы понижения размерности;</p> <p>Умеет:</p> <p>У 170.1 определять требования к данным из гетерогенных источников;</p> <p>У 170.2 использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников, в том числе в режиме реального времени;</p> <p>У 170.5 программировать на языке высокого уровня Python: для статистической обработки данных и работы с графикой, для работы с разрозненными фрагментами данных в больших массивах, для работы с базами структурированных и неструктурированных данных</p> <p>Владеет:</p> <p>В 170.3 навыками использования методов поиска данных и достоверные источники данных, предварительной статистической обработки, разметки структурированных и неструктурированных данных</p> <p>Компетенции</p> <p>ПК -170 Осуществляет сбор и подготовку данных для обучения моделей искусственного интеллекта</p>
<p>Модуль 4. Платформы науки о данных и машинного обучения и платформы бизнес- аналитики</p>	<p>Знает:</p> <p>З 167.1 основные методы, используемые для решения задач областей искусственного интеллекта;</p> <p>З 167.2 современный опыт использования анализа больших данных в промышленности;</p> <p>Умеет:</p> <p>У 167.3 адаптировать и развертывать модели в промышленной среде;</p>

Структурные элементы (разделы Программы)	Образовательные результаты
	<p>Владеет:</p> <p>В 167.1 навыками реализации базовые архитектуры и модели искусственных нейронных сетей и градиентного бустинга</p> <p>Компетенции</p> <p>ПК-167 Разрабатывает и реализует архитектуру ансамбля моделей</p>
Стажировка	<p>В 167.1 навыками реализации базовые архитектуры и модели искусственных нейронных сетей и градиентного бустинга</p> <p>В 170.1 навыками критического отбора данных и методами редукции размерности элементов набора данных</p> <p>В 170.2 навыками проверки данных на целостность и непротиворечивость.</p> <p>В 170.3 навыками использования методов поиска данных и достоверные источники данных, предварительной статистической обработки, разметки структурированных и неструктурированных данных</p> <p>В 171.1 навыками применения методов и критериев оценки качества моделей МО, определять критерии и метрики оценки результатов моделирования.</p> <p>В 171.2 навыкам решения задач анализа, прогнозирования, планирования, синтеза и принятия решений для машиностроительных предприятий</p>

**РАЗДЕЛ 2. СОДЕРЖАНИЕ ПРОГРАММЫ
«АНАЛИТИКА ДАННЫХ И МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА»**

**2.1. УЧЕБНЫЙ И УЧЕБНО-ТЕМАТИЧЕСКИЙ ПЛАНЫ
ДОПОЛНИТЕЛЬНОЙ ПРОФЕССИОНАЛЬНОЙ ПРОГРАММЫ
ПРОФЕССИОНАЛЬНОЙ ПЕРЕПОДГОТОВКИ
«АНАЛИТИКА ДАННЫХ И МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА»**

Цель: получение компетенции, необходимой для выполнения нового вида профессиональной деятельности в области информационных технологий обучающихся по специальностям и направлениям подготовки, отнесенным к ИТ-сфере; приобретение новой квалификации «Специалист по профессиональному интеллекту».

Категория слушателей: Обучающиеся по направлениям подготовки, отнесенными к ИТ-сфере

Срок обучения: 252 академических часа

Форма обучения: очная форма обучения, осуществляемая с применением электронного обучения и дистанционных образовательных технологий

Режим занятий: 6 академических часа в неделю (10 месяцев).

УЧЕБНЫЙ ПЛАН

№ п/ п	Наименование Модулей	Всего часов	В том числе:				Формы контроля	
			Аудиторная учебная нагрузка, часов			Самостоятельная работа		
			Лекционные занятия	Практические занятия	Контроль			
1	Модуль 1. Введение в - бизнес -аналитику и искусственный интеллект с применением Python для анализа данных	62	20	28	2	12	Практическое задание	
2	Модуль 2. Методы искусственного интеллекта для анализа табличных данных	70	18	36	2	14	Практическое задание	
2	Модуль 3. Современные озера и хранилища данных, аналитика больших данных и методы искусственного	32	8	16	2	6	Практическое задание	
4	Модуль 4. Платформы науки о данных и машинного обучения и платформы бизнес-аналитики	64	16	32	2	14	Практическое задание	
Стажировка		16				16	Отчет о стажировке	
Итоговая аттестация		8	-	-	8		Защита итогового проекта	
	ВСЕГО	252	62	112	16	62		

УЧЕБНО-ТЕМАТИЧЕСКИЙ ПЛАН

№ п/п	Наименование Модулей	Всего часов	В том числе:			Формы контроля	
			Аудиторная учебная нагрузка, часов		Самос тоятел ьная работа		
			Лекцион ные занятия	Практичес кие занятия			
1	Модуль 1. Введение в - бизнес -аналитику и искусственный интеллект с применением Python для анализа данных	62	20	28	2	12	
1.1.	Работа с электронными таблицами	12	4	6	-	2	
1.2.	Работа с Google Looker Studio	8	2	4	-	2	
1.3.	Применение машинного обучения к данным в Google Таблицах	8	2	4	-	2	
1.4.	Библиотека Pandas	16	4	8	-	4	
1.5	Библиотеки визуализации данных	12	4	6		2	
1.6	Основы информационной безопасности	4	4	-	-	-	
Промежуточная аттестация		2	-	-	2	-	
2	Модуль 2. Методы искусственного интеллекта для анализа табличных данных	70	18	36	2	14	
2.1	Машинное обучение для решения задач машиностроения	16	4	8	-	4	
2.2	Алгоритмы бустинга	14	4	8	-	2	
2.3	Кластерный анализ	8	2	4		2	
2.4	Нейронные сети	16	4	8	-	4	
2.5	Анализ временных рядов	14	4	8		2	
Промежуточная аттестация		2	-	-	2	-	
3	Модуль 3. Современные озера и хранилища данных, аналитика больших данных и методы искусственного	32	8	16	2	6	
3.1.	Облачные технологии обработки больших данных	8	2	4	-	2	
3.2.	RT.DataLake	14	4	8	-	2	
3.3.	Решение задач Data Mining в корпоративных хранилищах данных	8	2	4	-	2	
Промежуточная аттестация		2	-	-	2	-	

№ п/п	Наименование Модулей	Всего часов	В том числе:			Формы контроля	
			Аудиторная учебная нагрузка, часов		Самос- тоятел- ьная работа		
			Лекцион- ные занятия	Практичес- кие занятия			
						задание	
4	Модуль 4. Платформы науки о данных и машинного обучения и платформы бизнес- аналитики	64	16	32	2	14	
4.1.	Платформы науки о данных и машинного обучения	38	10	20	-	8	
4.2.	Платформы бизнес- аналитики	24	6	12	-	6	
Промежуточная аттестация		2	-	-	2	-	
Стажировка		16				16	
Итоговая аттестация		8	-	-	8	-	
ВСЕГО		252	62	112	16	62	

2.2. КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК

Календарный учебный план составляется при сформированной группе с учетом уровня их подготовки.

Календарный учебный график отражает периоды теоретических занятий, практик, процедур промежуточной и итоговой аттестаций и т.д.

КАЛЕНДАРНЫЙ УЧЕБНЫЙ ГРАФИК

Учебные недели	16	17	18	19	20	21	22	23	24	25	26	27	28	29	30
Итоговая аттестация (А)															

Учебные недели	31	32	33	34	35	36	37	38	39	40	41	42	43	44
Учебные занятия (У)	4	6	6	6	6	6	6	6	6	6	6	6	4	
Выездные занятия (В)														
Стажировка (Пр)													16	
Контроль (З, Э)	2											2		
Итоговая аттестация (А)														8

2.3. РАБОЧИЕ ПРОГРАММЫ ДИСЦИПЛИН/МОДУЛЕЙ

Рабочие программы дисциплин (модулей) отражают дисциплинарное содержание дополнительной профессиональной программы профессиональной переподготовки.

Модуль 1. Введение в бизнес-аналитику и искусственный интеллект с применением Python для анализа данных

Модуль 2. Методы искусственного интеллекта для анализа табличных данных

Модуль 3. Современные озера и хранилища данных, аналитика больших данных и методы искусственного

Модуль 4. Платформы науки о данных и машинного обучения и платформы бизнес-аналитики

2.3.1 РАБОЧАЯ ПРОГРАММА МОДУЛЯ 1.

«Введение в бизнес -аналитику и искусственный интеллект с применением Python для анализа данных»

Рабочая программа «Введение в бизнес-аналитику и искусственный интеллект с применением Python для анализа данных» (далее – рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки ИТ-профиля «Аналитика данных и методы искусственного интеллекта» и направлена на формирование ПК-171 Разрабатывает модули машинного обучения (МО) для решения задач

Планируемые результаты обучения по модулю 1:

По итогам освоения модуля слушатели должны:

Знать:

- З 171.1 основы языка Python и основные алгоритмические конструкции;
- З 171.2 методы интерпретации и визуализации анализа технологических данных.

Уметь:

- У 171.1 осуществлять интеграцию и преобразование данных в ходе работ по анализу технологических данных;
- У 171.2 применять различные алгоритмы машинного обучения при анализе технологических данных.

Владеть:

- В 171.1 навыками применения методов и критериев оценки качества моделей МО, определять критерии и метрики оценки результатов моделирования;
- В 171.2 навыкам решения задач анализа, прогнозирования, планирования, синтеза и принятия решений для машиностроительных предприятий.

2. Содержание модуля «Введение в бизнес -аналитику и искусственный интеллект с применением Python для анализа данных»

Тема 1.1. Работа с электронными таблицами

Будут рассмотрены следующие разделы:

- Введение в Google- таблицы, сводные таблицы Excel
- Применение сводных таблиц для маркетинговой сегментации

Формат занятия – лекция и практические занятия.

Тема 1.2. Работа с Google Looker Studio

Будут рассмотрены следующие разделы:

- Работа с Google Looker Studio
- Создание отчетов в Google Looker Studio

Формат занятия – лекция и практические занятия.

Тема 1.3. Применение машинного обучение к данным в Google Таблицах

Будут рассмотрены следующие разделы:

- Применение машинного обучения к данным в Google Таблицах
- Анализ и интерпретация моделей

Формат занятия – лекции и практические занятия.

Тема. 1.4. Библиотека Pandas

Будут рассмотрены следующие разделы:

- Особенности работы в библиотеке Pandas.
- Обработка данных в библиотеке Pandas.
- Исследовательский анализ данных (EDA) с использованием pandas
- Разведочный анализ данных с использованием библиотек автоматизации EDA

Формат занятия – лекция и практические занятия.

Тема. 1.5. Библиотеки визуализации данных

Будут рассмотрены следующие разделы:

- Библиотеки визуализации данных Matplotlib,
- Библиотеки визуализации данных Seaborn
- Библиотеки визуализации данных Altair
- Библиотеки визуализации данных Plotly Express

Формат занятия – лекция и практические занятия.

Тема. 1.6. Основы информационной безопасности

Будут рассмотрены следующие разделы:

- Введение в информационную безопасность
- Криптография
- Хеш
- Стеганография

Формат занятия – лекции

Практические занятия

№ темы	Наименование раздела дисциплины	Тема занятия	Объем, час.
1.1	Работа с электронными таблицами	Введение в Google- таблицы, сводные таблицы Excel	4
		Применение сводных таблиц для маркетинговой сегментации	4
1.2	Работа с Google Looker Studio	Работа с Google Looker Studio	2
		Создание отчетов в Google Looker Studio	2
1.3	Применение машинного обучения к данным в Google Таблицах	Применение машинного обучения к данным в Google Таблицах	2
		Анализ и интерпретация моделей	2
1.4	Библиотека Pandas	Особенности работы в библиотеке Pandas.	2
		Обработка данных в библиотеке Pandas.	2
		Исследовательский анализ данных (EDA) с	2

№ темы	Наименование раздела дисциплины	Тема занятия	Объем, час.
		использованием pandas	
		Разведочный анализ данных с использованием библиотек автоматизации EDA	2
1.5	Библиотеки визуализации данных	Библиотеки визуализации данных Matplotlib,	2
		Библиотеки визуализации данных Seaborn	2
		Библиотеки визуализации данных Altair	2
		Библиотеки визуализации данных Plotly Express	2

Самостоятельная работа

№ темы	Наименование темы	Тема занятия.	Вид СРС	Объем, час.
1	2	3	4	5
1	Работа с электронными таблицами	Введение в Google- таблицы, сводные таблицы Excel Применение сводных таблиц для маркетинговой сегментации	Выполнение практического задания	2
2	Классификация и функциональное назначение основных групп информационных систем для автоматизации промышленного производства	Работа с Google Looker Studio Создание отчетов в Google Looker Studio	Выполнение практического задания	2
3	Применение машинного обучение к данным в Google Таблицах	Применение машинного обучение к данным в Google Таблицах Анализ и интерпретация моделей	Выполнение практического задания	2
4	Библиотека Pandas	Особенности работы в библиотеке Pandas. Обработка данных в библиотеке Pandas.	Выполнение практического задания	2

№ темы	Наименование темы	Тема занятия.	Вид СРС	Объем, час.
1	2	3	4	5
		Исследовательский анализ данных (EDA) с использованием pandas Разведочный анализ данных с использованием библиотек автоматизации EDA	Выполнение практического задания	2
5	Библиотеки визуализации данных	Библиотеки визуализации данных Matplotlib, Библиотеки визуализации данных Seaborn Библиотеки визуализации данных Altair Библиотеки визуализации данных Plotly Express	Выполнение практического задания	2
6	Основы информационной безопасности	Введение в информационную безопасность Криптография Хеш Стеганография	Изучение теоретических основ	2

Практическое занятие № 1 Работа в Юпитер блокноте.

Задания для практической работы:

1. Данна переменная, в которой хранится слово из латинских букв. Напишите код, который выводит на экран:
 - a. · среднюю букву, если число букв в слове нечетное;
 - b. · две средних буквы, если число букв четное.

Практическое занятие № 2 Программирование в Юпитере.

Задания для практической работы:

1. Данна строка со значениями, которые разделены запятыми:
`line = '2019-07-01,organic,4293'`
 Напишите функцию `column_count`, которая возвращает число столбцов в такой строке.
2. Создайте класс сотрудника `Employee`. При инициализации класса задается имя сотрудника `name` и его текущая зарплата `salary`. Напишите следующие методы:
 Метод `up`, который увеличивает зарплату сотрудника на 100
 Метод `print`, который выводит на экран текущую зарплату сотрудника в формате "Сотрудник Иван, зарплата 100"
3. Вам нужно помочь секретарю автоматизировать работу. Для этого нужно написать программу, которая будет на основе хранимых данных выполнять пользовательские команды.
 Исходные данные имеют следующую структуру:
 1. перечень всех документов

```

documents = [
    {'type': 'passport', 'number': '2207 876234', 'name': 'Василий Гупкин'},
    {'type': 'invoice', 'number': '11-2', 'name': 'Геннадий Покемонов'},
    {'type': 'insurance', 'number': '10006', 'name': 'Аристарх Павлов'}
]
2. перечень полок, на которых хранятся документы (если документ есть в
documents, то он обязательно должен быть и в directories)
directories = {
    '1': ['2207 876234', '11-2'],
    '2': ['10006'],
    '3': []
}

```

Общие требования к программе:

код должен быть грамотно декомпозирован (каждая функция отвечает за свою конкретную задачу, дублирующийся функционал переиспользуется, а его код не повторяется);

в коде отсутствуют глобальные переменные (за исключением documents и directories);

пользовательский ввод обрабатывается в цикле while до тех пор, пока пользователь явно не завершит программу (вводом команды “q”).

****Задание 1**

Пункт 1. Пользователь по команде “р” может узнать владельца документа по его номеру

Примеры работы:

Ведите команду:

р

Ведите номер документа:

10006

Результат:

Владелец документа: Аристарх Павлов

Ведите команду:

р

Ведите номер документа:

12345

Результат:

Документ не найден в базе

Пункт 2. Пользователь по команде “с” может по номеру документа узнать на какой полке он хранится

Примеры работы:

Ведите команду:

с

Ведите номер документа:

10006

Результат:

Документ хранится на полке: 2

Ведите команду:

с

Ведите номер документа:

12345

Результат:

Документ не найден в базе

Пункт 3. Пользователь по команде “l” может увидеть полную информацию по всем документам

Пример работы:

Введите команду:

1

Результат:

№: 2207 876234, тип: passport, владелец: Василий Гупкин, полка хранения: 1

№: 11-2, тип: invoice, владелец: Геннадий Покемонов, полка хранения: 1

№: 10006, тип: insurance, владелец: Аристарх Павлов, полка хранения: 2

Пункт 4. Пользователь по команде “ads” может добавить новую полку

Примеры работы:

Введите команду:

ads

Введите номер полки:

10

Результат:

Полка добавлена. Текущий перечень полок: 1, 2, 3, 10.

Введите команду:

ads

Введите номер полки:

1

Результат:

Такая полка уже существует. Текущий перечень полок: 1, 2, 3.

Пункт 5. Пользователь по команде “ds” может удалить существующую полку из данных (только если она пустая)

Примеры работы:

Введите команду:

ds

Введите номер полки:

3

Результат:

Полка удалена. Текущий перечень полок: 1, 2.

Введите команду:

ds

Введите номер полки:

1

Результат:

На полке есть документа, удалите их перед удалением полки. Текущий перечень полок: 1, 2, 3.

Введите команду:

ds

Введите номер полки:

4

Результат:

Такой полки не существует. Текущий перечень полок: 1, 2, 3.

Задание 2 (необязательное)

Необходимо дополнить программу из задания 1 более продвинутыми командами.

Пункт 1. Пользователь по команде “ad” может добавить новый документ в данные

Примеры работы:

Ведите команду:

ad

Ведите номер документа:

42

Ведите тип документа:

multipassport

Ведите владельца документа:

R2D2

Ведите полку для хранения:

3

Результат:

Документ добавлен. Текущий список документов:

№: 2207 876234, тип: passport, владелец: Василий Гупкин, полка хранения: 1

№: 11-2, тип: invoice, владелец: Геннадий Покемонов, полка хранения: 1

№: 10006, тип: insurance, владелец: Аристарх Павлов, полка хранения: 2

№: 42, тип: multipassport, владелец: R2D2, полка хранения: 3

Ведите команду:

ad

Ведите номер документа:

42

Ведите тип документа:

multipassport

Ведите владельца документа:

R2D2

Ведите полку для хранения:

4

Результат:

Такой полки не существует. Добавьте полку командой as.

Текущий список документов:

№: 2207 876234, тип: passport, владелец: Василий Гупкин, полка хранения: 1

№: 11-2, тип: invoice, владелец: Геннадий Покемонов, полка хранения: 1

№: 10006, тип: insurance, владелец: Аристарх Павлов, полка хранения: 2

Пункт 2. Пользователь по команде “d” может удалить документ из данных

Примеры работы:

Ведите команду:

d

Ведите номер документа:

10006

Результат:

Документ удален.

Текущий список документов:

№: 2207 876234, тип: passport, владелец: Василий Гупкин, полка хранения: 1
 №: 11-2, тип: invoice, владелец: Геннадий Покемонов, полка хранения: 1

Введите команду:

d

Введите номер документа:

123456

Результат:

Документ не найден в базе.

Текущий список документов:

№: 2207 876234, тип: passport, владелец: Василий Гупкин, полка хранения: 1

№: 11-2, тип: invoice, владелец: Геннадий Покемонов, полка хранения: 1

№: 10006, тип: insurance, владелец: Аристарх Павлов, полка хранения: 2

Пункт 3. Пользователь по команде “m” может переместить документ с полки на полку

Примеры работы:

Введите команду:

m

Введите номер документа:

11-2

Введите номер полки:

3

Результат:

Документ перемещен.

Текущий список документов:

№: 2207 876234, тип: passport, владелец: Василий Гупкин, полка хранения: 1

№: 11-2, тип: invoice, владелец: Геннадий Покемонов, полка хранения: 3

№: 10006, тип: insurance, владелец: Аристарх Павлов, полка хранения: 2

Введите команду:

m

Введите номер документа:

11-2

Введите номер полки:

10

Результат:

Такой полки не существует. Текущий перечень полок: 1, 2, 3.

Введите команду:

m

Введите номер документа:

42

Введите номер полки:

2

Результат:

Документ не найден в базе.

Текущий список документов:

№: 2207 876234, тип: passport, владелец: Василий Гупкин, полка хранения: 1

№: 11-2, тип: invoice, владелец: Геннадий Покемонов, полка хранения: 1
 №: 10006, тип: insurance, владелец: Аристарх Павлов, полка хранения: 2

Практическое занятие № 3

Задание для практической работы:

1. Создайте numpy array с элементами от числа N до 0 (например, для N = 10 это будет array ([9, 8, 7, 6, 5, 4, 3, 2, 1, 0])).
2. Создайте диагональную матрицу с элементами от N до 0. Посчитайте сумму ее значений на диагонали.
3. Решите систему уравнений:

$$\begin{cases} 4x + 2y + z = 4 \\ x + 3y = 12 \\ 5y + 4z = -3 \end{cases}$$

4. Задание и входные данные к нему находятся в материалах занятия (ноутбук NumPy_задание.ipynb, раздел "Задача 4 домашнего задания")

Практическое занятие № 4

Задания для практической работы:

1. Скачайте с сайта grouplens.org...movielens/ датасет любого размера. Определите, какому фильму было выставлено больше всего оценок 5.0.
2. По данным файла power.csv посчитайте суммарное потребление стран Прибалтики (Латвия, Литва и Эстония) категорий 4, 12 и 21 за период с 2005 по 2010 года. Не учитывайте в расчетах отрицательные значения quantity.

Практическое занятие № 5

Задания для практической работы:

1. Проанализировать вручную массив информации (функции электронных таблиц).
2. Проверить статистические гипотезы (Электронная форма, функции электронных таблиц).

Практическое занятие № 6

Задания для практической работы:

1. Реализовать документооборот небольшой компании или подразделения, и строить на основании данных автоматизировать отчётность и составление прогнозов.
2. Собрать необходимую информацию по целевой аудитории.
3. Создать дашборд для анализа данных.

3. Формы аттестации и оценочные материалы по итогам освоения Модуля 1

Промежуточный контроль знаний слушателей определяется отчета по практическому занятию (лабораторной работе).

Перечень заданий к лабораторной работе

Задание № 1

На основе полученных данных, провести предварительный анализ данных, наглядно описать их качество и классифицировать дефекты изделий

Задание №2

Напишите функцию, которая будет находить не дозагруженное оборудование по всем цехам

Задание №4

Напишите функцию, преобразующую произвольный список вида `['2018-01-01', 'yandex', 'cpc', 100]` (он может быть любой длины) в словарь `{'2018-01-01': {'yandex': {'cpc': 100}}}`

4. Учебно-методическое обеспечение и информационное обеспечение Модуля 1**Рекомендуемая литература:**

1. Скиена Стивен С. Наука о данных. Учебный курс: Пер. с англ – М: Изд-во: Диалектика (Вильямс), 2020, – 544 с. – ISBN 978-5-907144-74-3
2. Грас Джоэл. Data Science. Наука о данных с нуля: Пер. с англ – 2-ое изд., перераб. И доп. – СПб: БХВ-Петербург, 2024 – 416 с: ил. ISBN 978-5-9775-6731-2.
3. Брюс Питер, Брюс Эндрю, Гедек Питер. Практическая статистика для специалистов Data Science, – СПб: БХВ-Петербург, 2021 – 352с - ISBN: 978-5-9775-6705-3
4. Куслейка Дик. Визуализация данных при помощи дашбордов и отчетов в Excel, - М.: ДМК Пресс, 2022, - 338 с., – ISBN 978-5-9706-0966-8
5. Foster Provost, Tom Fawcett. Data Science for Business: What You Need to Know about Data Mining and Data-Analytic Thinking . O'reilly, 2013 – 414 p
6. Field Guide to Data Science . Booz, Allen, Hamilton, 2015 – 126 p.

Программное обеспечение и Интернет-ресурсы:

1. MS Excel.
2. Google-таблицы.
3. Jupyter Notebook.
4. Google Looker Studio.
5. Python и библиотека Pandas.
6. Библиотеки визуализации данных Matplotlib, Seaborn, Altair, Plotly Express.
7. Научная электронная библиотека eLIBRARY.RU.
8. Электронная образовательная среда в сети Интернет по адресу <http://edu.stankin.ru/>, функционирующая на базе программного обеспечения системы управления обучением Moodle.

5. Материально-технические условия реализации Модуля 1:

- оборудованные аудитории;
- ноутбук;
- мультимедиа проектор со звуковыми колонками;
- другие технические средства.

2.3.2 РАБОЧАЯ ПРОГРАММА МОДУЛЯ 2.

«Методы искусственного интеллекта для анализа табличных данных»

Рабочая программа «Методы искусственного интеллекта для анализа табличных данных» (далее – рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки ИТ-профиля «Аналитика данных и методы искусственного интеллекта» и направлена на формирование ПК -170 Осуществляет сбор и подготовку данных для обучения моделей искусственного интеллекта, ПК -167 Разрабатывает и реализует архитектуру ансамбля моделей

Планируемые результаты обучения по модулю 2:

По итогам освоения модуля слушатели должны:

Знать:

- З 167.1 основные методы, используемые для решения задач областей искусственного интеллекта;
- З 167.2 современный опыт использования анализа больших данных в промышленности;
- З 167.3 основы моделей искусственных нейронных сетей
- З 170.1 области применения технологий искусственного интеллекта в промышленности;
- З 170.3 нейронные сети: полносвязные, свёрточные и рекуррентные нейронные сети, методы обучения нейронных сетей, нейросетевые методы понижения размерности;
- З 170.4 анализ изображений, анализ временных рядов

Уметь:

- У 167.1 разрабатывать и оценивать модели градиентного бустинга;
- У 167.2 разрабатывать и оценивать модели искусственных нейронных сетей;
- У 167.3 адаптировать и развертывать модели в промышленной среде;
- У 170.3 производить очистку данных для проведения аналитических работ;
- У 170.4 проводить интеграцию и преобразование больших объемов данных;
- У 170.5 программировать на языке высокого уровня Python: для статистической обработки данных и работы с графикой, для работы с разрозненными фрагментами данных в больших массивах, для работы с базами структурированных и неструктурированных данных

Владеть:

- В 167.1 навыками реализации базовые архитектуры и модели искусственных нейронных сетей и градиентного бустинга
- В 170.1 навыками критического отбора данных и методами редукции размерности элементов набора данных
- В 170.2 навыками проверки данных на целостность и непротиворечивость.

2. Содержание модуля «Методы искусственного интеллекта для анализа табличных данных»

Тема 2.1. Машинное обучение для решения задач машиностроения

Будут рассмотрены следующие разделы:

- Ознакомление с основными методами и моделями машинного обучения.

- Применение машинного обучения в промышленности
- Машинное обучение для решения задач Data Mining.
- Линейные модели и градиентный спуск в машинном
- Алгоритмы построения деревьев решений, критерии разделения.
- Бэггинг, Random Forest,
- Стакинг
- Ансамблевые методы машинного обучения

Формат занятия – лекция и практические задание.

Тема 2.2. Алгоритмы бустинга

Будут рассмотрены следующие разделы:

- Бустинг.
- AdaBoost и градиентный бустинг над решающими деревьями
- Фреймворки машинного обучения

Формат занятия – лекция и практическое занятие.

Тема 2.3. Кластерный анализ

Будут рассмотрены следующие разделы:

- Кластерный анализ,
- Алгоритм k-means
- Поиск ассоциативных правил

Формат занятия – лекция и практическое занятие.

Тема 2.4. Нейронные сети.

Будут рассмотрены следующие разделы:

- Введение в нейронные сети
- Глубокие нейронные сети

Формат занятия – лекция и практические задание.

Тема 2.5. Анализ временных рядов.

Будут рассмотрены следующие разделы:

- Модели временных рядов;
- Применение различных моделей для анализа массивных данных;
- Автоматическое машинное обучение (AutoML)

Формат занятия – лекции и практические задания.

Практические занятия

№ темы	Наименование раздела дисциплины	Тема занятия	Объем, час.
2.1	Машинное обучение для решения задач машиностроения	Ознакомление с основными методами и моделями машинного обучения. Применение машинного обучения в промышленности Машинное обучение для решения задач Data Mining	2
		Линейные модели и градиентный спуск в машинном Алгоритмы построения деревьев решений, критерии разделения.	2
		Бэггинг, Random Forest,	2
		Стакинг Анбасмблиевые методы машинного обучения	2
2.2	Алгоритмы бустинга	Бустинг.	2
		AdaBoost и градиентный бустинг над решающими деревьями	4
		Фреймворки машинного обучения	2
2.3	Кластерный анализ	Кластерный анализ	2
		Алгоритм k-means Поиск ассоциативных правил	2
2.4	Нейронные сети.	Введение в нейронные сети	4
		Глубокие нейронные сети	4
2.5	Анализ временных рядов	Модели временных рядов;	2
		Применение различных моделей для анализа массивных данных;	2
		Автоматическое машинное обучение (AutoML)	4

Самостоятельная работа

№ темы	Наименование темы	Тема занятия.	Вид СРС	Объем, час.
1	2	3	4	5
1	Машинное обучение для	Ознакомление с основными методами и моделями	Выполнение практического	2

№ темы	Наименование темы	Тема занятия.	Вид СРС	Объем, час.
1	2	3	4	5
	решения задач машиностроения	машинного обучения. Применение машинного обучения в промышленности Машинное обучение для решения задач Data Mining Линейные модели и градиентный спуск в машинном Алгоритмы построения деревьев решений, критерии разделения.	задания	
		Бэггинг, Random Forest, Стакинг Ансамблевые методы машинного обучения	Выполнение практического задания	2
2	Алгоритмы бустинга	Бустинг. AdaBoost и градиентный бустинг над решающими деревьями Фреймворки машинного обучения	Выполнение практического задания	2
3	Кластерный анализ	Кластерный анализ Алгоритм k-means Поиск ассоциативных правил	Выполнение практического задания	2
4	Нейронные сети.	Введение в нейронные сети	Выполнение практического задания	2
		Глубокие нейронные сети	Выполнение практического задания	2
5	Анализ временных рядов	Модели временных рядов; Применение различных моделей для анализа массивных данных; Автоматическое машинное обучение (AutoML)	Выполнение практического задания	2

Практическое занятие № 1

Возьмите датасет

1. Оставьте два признака
2. Разделите данные на выборку для обучения и тестирования
3. Постройте модель машинного обучения
4. Визуализируйте предсказания для тестовой выборки и центры классов
5. Отбросьте целевую переменную и оставьте только два признака
6. Подберите оптимальное число кластеров для алгоритма kmeans и визуализируйте полученную кластеризацию

Практическое занятие № 2

Задание 1. Сгенерировать наборы данных:

Для каждого набора данных:

- Обучить и протестировать KNN
- Построить графики точности (train/test) при разных k (1...8)
- Построить границы классов

Подсказка: можно использовать sklearn.datasets

Задание 2

Переписать класс KNN, заменим операции в цикле на векторные вычисления (используйте библиотеку numpy)

Задание 3

25 объектов классифицированы на 3 класса

C="Cat"

F="Fish"

H="Hen"

истинные значения:

y_true = [C,C,C,C,C,C, F,F,F,F,F,F,F,F,F,F, H,H,H,H,H,H,H,H]

результат системы:

y_pred = [C,C,C,C,H,F, C,C,C,C,C,H,H,F,F, C,C,C,H,H,H,H,H,H]

Практическое занятие № 3

Задание 1.

Написать код на Python, который воспроизводит вывод, представленный на картинке ниже (не используя сторонние библиотеки, кроме numpy):

1. Взять датасет homework.csv
2. Описание датасета доступно тут - www.kaggle.com/rhardness
3. Предсказываем значение столбца MEDV на основе других признаков
4. Решить задачу регрессии, используя алгоритм линейной регрессии (понадобиться библиотека scikit-learn)
5. Оценить качество регрессии при помощи метрики MSE, используя возможности библиотеки scikit-learn

Практическое занятие № 4

Исходные данные

Датасет Mortality and Water Hardness (www.kaggle.com/rhardness)

В датасете содержатся данные по средней годовой смертности на 100000 населения и концентрации кальция в питьевой воде для 61 большого города в Англии и Уэльсе. Города дополнитель но поделены на северные и южные.

Задание 1.

1. Ответить на вопрос: есть ли связь между жёсткостью воды и средней годовой смертностью?

2. Построить точечный график
3. Рассчитать коэффициенты корреляции Пирсона и Спирмена
4. Построить модель линейной регрессии
5. Рассчитать коэффициент детерминации
6. Вывести график остатков

Задание 2.

1. Ответить на вопрос: сохраняется ли аналогичная зависимость для северных и южных городов по отдельности?
2. Разделить данные на 2 группы
3. Повторить аналогичные шаги из пункта 1 для каждой группы по отдельности

Практическое задание №5

Исходные данные

Датасет Титаник

Задание

1. Проведите разведочный анализ датасета.
2. С помощью корреляционной матрицы определите взаимосвязи данных.
3. Сделайте выводы.
4. Обучите модель линейной регрессии для определения выживаемости пассажиров.

Практическое задание №6

Исходные данные

Датасет состоит из нескольких медицинских (независимых) переменных и одной целевой (зависимой) переменной Outcome. Независимые переменные включают количество беременностей у пациента, уровень инсулина, возраст и т. д.

Задание

Решите задачу кластеризации

3. Формы аттестации и оценочные материалы по итогам освоения Модуля 2

Промежуточный контроль знаний слушателей определяется формированием отчета по практическому занятию (лабораторной работе) и выполнением практико-ориентированного кейса.

Пример практико-ориентированных кейсов.

Кейс №1

Реализация модели прогнозирования средних цен и объемов потребления отдельных групп товаров. Горизонт прогнозирования – 3 месяца.

Основным источником информации являются наборы данных по целевым продуктовым группам за исторический период с начала 2019 года. Однако, для повышения точности прогнозирования возможно использовать для обучения модели дополнительные данные, например, об основных демографических показателях территорий.

Допустимо использовать социально-экономические данные из открытых источников с указанием источника.

Исходные данные (пример):

Датасет ФНС, содержащий сведения о ценах и объемах реализации 32

наименований продуктовых товаров, 4 видов ГСМ и потребительских корзин. Каждая строка наблюдения – это данные по субъекту РФ за определенный день. Всего набор содержит 5 577 874 фактов наблюдений за 821 день по 86 регионам по 79 атрибутам за период с 1 января 2019 года по 31 марта 2021 года.

Формат данных:

date – Дата наблюдения, в формате ДД.ММ.ГГГГ
 region – Код региона в кодировке ФНС России
 okato – ОКАТО региона РФ
 oktmo – ОКТМО региона РФ
 pasta – макароны (средняя взвешенная цена), руб/кг
 legumes – бобовые, руб/кг
 bread – хлеб, руб/кг
 flour – мука, руб/кг
 rice – рис, руб/кг
 groats – другие крупы, руб/кг
 potatoes – картофель, руб/кг
 cucumbers_tomatoes – огурцы и помидоры, руб/кг
 vegetables – прочие овощи, руб/кг
 roots – столовые корнеплоды, руб/кг
 cabbage – капуста, руб/кг
 fruit – фрукты, руб/кг
 sugar – сахар, руб/кг
 candy – конфеты, руб/кг
 biscuits – печенье, руб/кг
 mutton – баранина, руб/кг
 beef – говядина, руб/кг
 chicken – мясо птицы, руб/кг
 pork – свинина, руб/кг
 fish – рыба свеж, руб/кг
 herring – сельдь, руб/кг
 curd – творог, руб/кг
 sour_cream – сметана, руб/кг
 butter – масло сливочное, руб/кг
 milk_kefir – молоко кефир, руб/л
 cheese – сыр, руб/кг
 egg – яйца, руб/шт
 margarine – маргарин и другие жиры, руб/кг
 oil – масло растительное, руб/л
 salt – соль, руб/кг
 spice – специи, руб/шт
 tea – чай, руб/шт
 CPI_1 – стоимость потребительской корзины для трудоспособных граждан, руб
 CPI_2 – стоимость потребительской корзины для пенсионеров, руб
 CPI_3 – стоимость потребительской корзины для детей до 18 лет, руб
 Pasta_value – макароны (объем проданной продукции), кг
 legumes_value – бобовые (объем проданной продукции), кг
 bread_value – хлеб (объем проданной продукции), кг
 flour_value – мука (объем проданной продукции), кг
 rice_value – рис (объем проданной продукции), кг

groats_value – другие крупы (объем проданной продукции), кг
 potatoes_value – картофель (объем проданной продукции), кг
 cucumbers_tomatoes_value – огурцы и помидоры (объем проданной продукции), кг
 vegetables_value – прочие овощи (объем проданной продукции), кг
 roots_value – столовые корнеплоды (объем проданной продукции), кг
 cabbage_value – капуста (объем проданной продукции), кг
 fruit_value – фрукты (объем проданной продукции), кг
 sugar_value – сахар (объем проданной продукции), кг
 candy_value – конфеты (объем проданной продукции), кг
 biscuits_value – печенье (объем проданной продукции), кг
 mutton_value – баранина (объем проданной продукции), кг
 beef_value – говядина (объем проданной продукции), кг
 chicken_value – мясо птицы(объем проданной продукции), кг
 pork_value – свинина (объем проданной продукции), кг
 fish_value – рыба свеж (объем проданной продукции), кг
 herring_value – сельдь (объем проданной продукции), кг
 curd_value – творог (объем проданной продукции), кг
 sour_cream_value – сметана (объем проданной продукции), кг
 butter_value – масло сливочное (объем проданной продукции), кг
 milk_kefir_value –молоко кефир (объем проданной продукции), л
 cheese_value – сыр (объем проданной продукции), кг
 egg_value – яйца (объем проданной продукции), шт
 margarine_value – маргарин и другие жиры (объем проданной продукции), кг
 oil_value – масло растительное (объем проданной продукции), л
 salt_value – соль (объем проданной продукции) , кг
 spice_value – специи (объем проданной продукции), шт
 tea_value – чай (объем проданной продукции), шт
 ai92 – бензин марки АИ-92, руб/литр
 ai95 – бензин марки АИ-95, руб/литр
 ai98 – бензин марки АИ-98, руб/литр
 dt – дизельное топливо, руб/литр
 ai92_value – бензин марки АИ-92 (объем проданной продукции), л
 ai95_value – бензин марки АИ-95 (объем проданной продукции), л
 ai98_value – бензин марки АИ-98 (объем проданной продукции), л
 dt_value – дизельное топливо (объем проданной продукции), л

Уникальной комбинацией является комбинация полей: {date, region} или {date, oktmo}, или {date, akato}.

Формат решения:

Необходимо предсказать цены и объемы за период с 01.04.2021 по 30.06.2021 в разрезе регионов (см. файл test.csv).

Расчет метрики

Для каждого прогнозируемого показателя (столбца) рассчитывается его среднее значение $v_{mean\ j}$ и значение метрики MAE: $v_{mae\ j}$

Для каждого показателя вычисляется отношение

$$E\ j = v_{mae\ j} / v_{mean\ j}$$

Рассчитывается среднее значение Е среди всех столбцов

$$E_{mean} = 1 / n * sum(E\ j)$$

Берется обратная величина и делится на константу 1000

$$score = 1 / (1000 * E_mean)$$

Кейс №2

Имеются данные первичного медицинского осмотра. Задача: определить вероятность наличия сердечно-сосудистых заболеваний (поле *cardio*). Использовать бинарную классификацию.

Кейс №3

Предсказать рейтинг энергопотребления (Energy Star Score) здания и понять, какие признаки оказывают на него сильнейшее влияние.

Кейс №4

Важно, чтобы компании, выпускающие кредитные карты, могли распознавать мошеннические транзакции по кредитным картам, чтобы с клиентов не взималась плата за товары, которые они не покупали.

Набор данных содержит транзакции, совершенные по кредитным картам в сентябре 2013 года держателями карт из Европы. Этот набор данных представляет транзакции, которые произошли за два дня, где у нас есть 492 мошенничества из 284 807 транзакций. Набор данных сильно несбалансирован, на положительный класс (мошенничество) приходится 0,172% всех транзакций. Обучить модель для выявления мошеннических транзакций.

Кейс №5

На основе набора данных выяснить, кто лучшие клиенты. Для этого используйте методику матричного принципа RFM. RFM означает - Новизна, Частота и Денежный. RFM – это, по сути, метод сегментации клиентов, которая работает, используя прошлое поведение ваших клиентов на основе их покупок, чтобы разделить клиентов на группы

Кейс №6

На основе имеющегося набора данных необходимо распознать письма со спам рассылкой.

Кейс №7

Реализовать модель прогнозирования эффективности использования топлива с помощью машинного обучения. Для этого использовать один из известных наборов данных среди практиков машинного обучения, набор данных Auto MPG, чтобы создать модель для прогнозирования топливной экономичности транспортных средств в конце 1970-х и начале 1980-х годов

4. Учебно-методическое обеспечение и информационное обеспечение Модуля 2

Рекомендуемая литература:

1. Скиена Стивен С. Наука о данных. Учебный курс: Пер. с англ – М: Изд-во: Диалектика (Вильямс), 2020, – 544 с. – ISBN 978-5-907144-74-3
2. Грас Джоэл. Data Science. Наука о данных с нуля: Пер. с англ – 2-ое изд., перераб. И доп. – СПб: БХВ-Петербург, 2024 – 416 с: ил. ISBN 978-5-9775-6731-2.
3. Брюс Питер, Брюс Эндрю, Гедек Питер. Практическая статистика для специалистов Data Science, – СПб: БХВ-Петербург, 2021 – 352с - ISBN: 978-5-9775-6705-3
4. Дж. Плас: Python для сложных задач. Наука о данных и машинное обучение – СПб: Питер, 2022 – 576 с. – ISBN 978-5-4461-0914-2

5. Груздев А.В. Предварительная подготовка данных в Python. Том 1. Инструменты и валидация – М.: ДМК Пресс, 2023 – 816 с. – ISBN 978-5-93700-156-6
6. Груздев А.В. Предварительная подготовка данных в Python. Том 2. План, примеры и метрики качества – М.: ДМК Пресс, 2023 – 814 с. – ISBN 978-5-93700-177-1

Программное обеспечение и Интернет-ресурсы:

1. Google-таблицы.
2. Jupyter Notebook.
3. Google Looker Studio.
4. Google Colaboratory
5. Python и библиотека Pandas, NumPy, scikit-learn
6. Библиотеки визуализации данных Matplotlib, Seaborn, Altair, Plotly Express.
7. Библиотека размеченных данных www.kaggle.com
8. Научная электронная библиотека eLIBRARY.RU.
9. Электронная образовательная среда в сети Интернет по адресу <http://edu.stankin.ru/>, функционирующая на базе программного обеспечения системы управления обучением Moodle.

5. Материально-технические условия реализации Модуля 2:

- оборудованные аудитории;
- ноутбук;
- мультимедиа проектор со звуковыми колонками;
- другие технические средства.

2.3.3 РАБОЧАЯ ПРОГРАММА МОДУЛЯ 3.

«Современные озера и хранилища данных, аналитика больших данных и методы искусственного интеллекта»

1. Цель и задачи освоения Модуля 3.

Рабочая программа «Современные озера и хранилища данных, аналитика больших данных и методы искусственного интеллекта» (далее – рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки ИТ-профиля «Аналитика данных и методы искусственного интеллекта» и направлена на формирование ПК -170 Осуществляет сбор и подготовку данных для обучения моделей искусственного интеллекта,

Планируемые результаты обучения по модулю 3:

По итогам освоения модуля слушатели должны:

Знать:

- З 170.1 области применения технологий искусственного интеллекта в промышленности;
- З 170.2 принципы поиска и анализа информации, необходимой для решения поставленной задачи, в том числе с применением аналитики данных и машинного обучения;
- З 170.3 нейронные сети: полносвязные, свёрточные и рекуррентные нейронные сети, методы обучения нейронных сетей, нейросетевые методы понижения размерности;

Уметь:

- У 170.1 определять требования к данным из гетерогенных источников;
- У 170.2 использовать инструментальные средства для извлечения, преобразования, хранения и обработки данных из разнородных источников, в том числе в режиме реального времени;
- У 170.5 программировать на языке высокого уровня Python: для статистической обработки данных и работы с графикой, для работы с разрозненными фрагментами данных в больших массивах, для работы с базами структурированных и неструктурированных данных

Владеть:

- В 170.3 навыками использования методов поиска данных и достоверные источники данных, предварительной статистической обработки, разметки структурированных и неструктурированных данных

2. Содержание модуля «Современные озера и хранилища данных, аналитика больших данных и методы искусственного интеллекта»

Тема 3.1. Облачные технологии обработки больших данных

Будут рассмотрены следующие разделы:

- Большие данные
- Инструменты работы с большими данными
- Apache Spark
- Отечественные платформы аналитики больших данных
- RT.WideStore
- Технологии Amazon Web Services для обработки больших данных

Формат занятия – лекция и практическое занятие.

Тема 3.2. RT.DataLake

Будут рассмотрены следующие разделы:

- RT.DataLake
- Маркетинговая аналитика в RT.Warehouse

Формат занятия – лекция и практическое занятие.

Тема 3.3. Решение задач Data Mining в корпоративных хранилищах данных

Будут рассмотрены следующие разделы:

- Решение задач Data Mining в корпоративных хранилищах данных

Формат занятия – лекция и практическое занятие.

Практические занятия

№ темы	Наименование раздела дисциплины	Тема занятия	Объем, час.
3.1	Облачные технологии обработки больших данных	Большие данные Инструменты работы с большими данными	2
		Apache Spark	2
3.2	Алгоритмы бустинга	RT.DataLake	4
		Маркетинговая аналитика в RT.Warehouse	4
3.3	Решение задач Data Mining в корпоративных хранилищах данных	Решение задач Data Mining в корпоративных хранилищах данных	4

Самостоятельная работа

№ темы	Наименование темы	Тема занятия.	Вид СРС	Объем, час.
1	2	3	4	5
1	Облачные технологии обработки больших данных	Большие данные Инструменты работы с большими данными Линейные модели и градиентный спуск в машинном Алгоритмы построения деревьев решений, критерии разделения.	Выполнение практического задания	2
2	Алгоритмы бустинга	RT.DataLake Маркетинговая аналитика в RT.Warehouse	Выполнение практического задания	2
3	Решение задач Data Mining в корпоративных	Решение задач Data Mining в корпоративных хранилищах данных	Выполнение практического задания	2

№ темы	Наименование темы	Тема занятия.	Вид СРС	Объем, час.
1	2	3	4	5
	хранилищах данных			

Кейс №1

На основе методов искусственного интеллекта, а также считая метрикой качества построенной модели MAE (mean absolute error), предстоит предсказать уровень заражения вирусом в условиях пандемии на примере COVID-19 в городах России.

Вирус COVID-19 начал распространяться внутри России в 2020 году. Вирусы подобного типа по словам ученых рискуют стать частыми из-за глобального распространения авиации и других быстрых способов перемещения людей по планете. Для наиболее полного анализа распространения вируса правильно рассмотреть случаи заражения по каждому отдельному городу. Необходимо предсказать уровень заражения вирусом в городах, считая метрикой качества построенной модели MAE (mean absolute error).

Кейс №2

На основе методов искусственного интеллекта предстоит определить уровень стресса человека по зарегистрированным при помощи датчиков полиграфа (пьезо- и фотоплетизмограмма) реакциям на вопросы. Такая классификация производится при помощи поиска закономерностей в размеченных вручную экспертами-полиграфологами реакциями.

Разработанное урешение поможет сделать возможным создание системы экспресс-диагностики уровня стресса человека в реакциях на различные стимулы. Такой скрининг может превентивно выявлять негативные намерения, девиантное поведение и депрессивные состояния.

3. Формы аттестации и оценочные материалы по итогам освоения Модуля 3

Промежуточный контроль знаний слушателей определяется формированием отчета по практическому занятию (лабораторной работе) и выполнением практико-ориентированного кейса.

Пример практико-ориентированных кейсов.

Кейс №1

При помощи методов искусственного интеллекта предстоит предсказать, будет ли совершено определенное действие пользователем или нет в зависимости от его кликстрима.

Задача построена на неструктурированных текстовых источниках данных. Каждый объект выборки характеризуется набором интернет-сессий (непрерывные последовательности посещенных пользователем сайтов).

Предложенное решение позволит сформировать спектр услуг под приоритеты и потребности отдельного пользователя VK.

Решение задачи осложняется низкой долей положительного класса, наличием пропусков и зашумленности в данных, необходимостью выделения признаков из

не структурированного текстового источника.

Кейс №2

В Ростелекоме ежедневно используется множество моделей машинного обучения. При этом для построения таких моделей в основном задействованы внутренние данные (возраст, пол, история платежей и т. п.), и в меньшей степени внешние данные, такие как экономическая статистика, социальные и демографические показатели, курсы валют и фондовых рынков и пр.

Построить модель машинного обучения на предоставленном датасете, а также, по возможности, обогатить его внешними данными, чтобы улучшить качество модели.

Кейс №3

Разработать модель, которая в режиме квази-реального времени сможет формировать прогноз инфляции в текущем месяце. Такой прогноз будет опираться на открытые микроданные о ценах, собранных посредством веб-скрейпинга с сайтов интернет-магазинов. Использование ежедневной информации об онлайн-ценах позволит более точно и оперативно прогнозировать месячный показатель инфляции.

Особенность задачи – необходимость работы с большим количеством временных рядов. База данных содержит ежедневные цены о нескольких миллионах позиций, собранных с сайтов более 200 интернет-магазинов. При этом прогнозирование усложняется тем, что состав позиций постоянно изменяется. Отслеживаемые товары периодически выпадают из выборки, а новые появляются.

Разработка поможет усовершенствовать инструментарий прогнозирования инфляции, дополнив его методами на основе альтернативных данных. Точность инструментов прогнозирования во многом определяет эффективность и своевременность принимаемых решений в сфере денежно-кредитной политики и поддержания финансовой стабильности.

3. Формы аттестации и оценочные материалы по итогам освоения Модуля 3

Промежуточный контроль знаний слушателей определяется собеседованием по материалам лекции и практических занятий, а также формированием отчета по практическому занятию (лабораторной работе) и выполнением контрольной работы (по материалам семинара).

Перечни вопросов для собеседования вариативны и могут модифицироваться в зависимости от изменения и дополнения лекционного материала.

4. Учебно-методическое обеспечение и информационное обеспечение Модуля 3

Рекомендуемая литература:

1. Лакшманан В., Тайджани Д. Google BigQuery. Всё о хранилищах данных, аналитике и машинном обучении, СПб: Питер, 2021. – 496с. ISBN: 978-5-4461-1707-9
2. Петров А. Распределенные данные. Алгоритмы работы современных систем хранения информации – СПб.: Питер, 2022. – 336 с.
3. Стренхольт П. Масштабируемые данные. Лучшие шаблоны высоконагруженных архитектур. – СПб.: Питер, 2022 – 368с.: ил. – ISBN 987-5-4461-1461-0.
4. Железнов, М. М. Методы и технологии обработки больших данных: учебно-методическое пособие / М. М. Железнов. – Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. – URL:

<https://www.iprbookshop.ru/101802.html> (дата обращения: 15.06.2024). – Режим доступа: для авторизованных пользователей.

5. Дружинин, Д. В. Высокопроизводительные вычисления и облачные технологии: учебное пособие / Д. В. Дружинин. – Томск: Издательство Томского государственного университета, 2020. – 93 с. – ISBN 978-5-94621-921-1. – Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. – URL: <https://www.iprbookshop.ru/116813.html> (дата обращения: 15.06.2024). – Режим доступа: для авторизованных пользователей
6. Рамсундар Бхарат. TensorFlow for Deep Learning: Пер. с англ. - СПб: БХВ-Санкт-Петербург, 2022, – 256 с., – ISBN 978-5-9775-4014-8
7. Брюс Питер, Брюс Эндрю, Гедек Питер. Практическая статистика для специалистов Data Science, – СПб: БХВ-Петербург, 2021 – 352с - ISBN: 978-5-9775-6705-3
8. Дж. Плас: Python для сложных задач. Наука о данных и машинное обучение – СПб: Питер, 2022 – 576 с. – ISBN 978-5-4461-0914-2

Программное обеспечение и Интернет-ресурсы:

1. Jupyter Notebook.
2. Apache Spark
3. RT.DataLake
4. RT.WarehouseGoogle Looker Studio.
5. Python и библиотека Pandas, NumPy, scikit-learn
6. Библиотеки визуализации данных Matplotlib, Seaborn, Altair, Plotly Express.
7. Библиотека размеченных данных www.kaggle.com
8. Научная электронная библиотека eLIBRARY.RU.
9. Электронная образовательная среда в сети Интернет по адресу <http://edu.stankin.ru/>, функционирующая на базе программного обеспечения системы управления обучением Moodle.

5. Материально-технические условия реализации Модуля 3:

- оборудованные аудитории;
- ноутбук;
- мультимедиа проектор со звуковыми колонками;
- другие технические средства.

2.3.4 РАБОЧАЯ ПРОГРАММА МОДУЛЯ 4.

«Платформы науки о данных и машинного обучения и платформы бизнес-аналитики»

Рабочая программа «Платформы науки о данных и машинного обучения и платформы бизнес-аналитики» (далее – рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки ИТ-профиля «Аналитика данных и методы искусственного интеллекта» и направлена на формирование ПК-167 Разрабатывает и реализует архитектуру ансамбля моделей

Планируемые результаты обучения по модулю 4:

По итогам освоения модуля слушатели должны:

Знать:

- З 167.1 основные методы, используемые для решения задач областей искусственного интеллекта;
- З 167.2 современный опыт использования анализа больших данных в промышленности;

Уметь:

- У 167.3 адаптировать и развертывать модели в промышленной среде;

Владеть:

- В 167.1 навыками реализации базовые архитектуры и модели искусственных нейронных сетей и градиентного бустинга

2. Содержание модуля «Платформы науки о данных и машинного обучения и платформы бизнес-аналитики»

Тема 4.1. Платформы науки о данных и машинного обучения

Будут рассмотрены следующие разделы:

- Платформа H2O.ai
- Платформа RapidMiner
- Аналитические технологии отечественной платформы Loginom
- Платформа Knime
- Исследование и визуализация данных в RT.DataVision

Формат занятия – лекция и практические задания.

Тема 4.2. Платформы бизнес- аналитики

Будут рассмотрены следующие разделы:

- Создание интерактивной отчетности в Tableau
- Аналитические технологии Power BI
- Визуализация данных в Yandex DataLens

Формат занятия – лекция и практическое занятие.

Практические занятия

№ темы	Наименование раздела дисциплины	Тема занятия	Объем, час.
4.1	Платформы науки о данных и	Платформа H2O.ai	4

№ темы	Наименование раздела дисциплины	Тема занятия	Объем, час.
	машинного обучения	Платформа RapidMiner	4
		Аналитические технологии отечественной платформы Loginom	4
		Платформа Knime	4
		Исследование и визуализация данных в RT.DataVision	4
4.2	Платформы бизнес-аналитики	Создание интерактивной отчетности в Tableau	4
		Аналитические технологии Power BI	4
		Визуализация данных в Yandex DataLens	4

Самостоятельная работа

№ темы	Наименование темы	Тема занятия.	Вид СРС	Объем, час.
1	2	3	4	5
1	Платформы науки о данных и машинного обучения	Платформа H2O.ai Платформа RapidMiner	Выполнение практического задания	2
		Аналитические технологии отечественной платформы Loginom	Выполнение практического задания	2
		Платформа Knime	Выполнение практического задания	2
		Исследование и визуализация данных в RT.DataVision	Выполнение практического задания	2
2	Платформы бизнес-аналитики	Создание интерактивной отчетности в Tableau	Выполнение практического задания	2
		Аналитические технологии Power BI	Выполнение практического задания	2
		Визуализация данных в Yandex DataLens	Выполнение практического задания	2

Практическое занятие № 1

Датасет **moons**

Сгенерируйте датасет

Задание 1. Создайте тензоры с обучающими и тестовыми данными

Задание 2. Реализация логистической регрессии

Необходимо написать модуль на PyTorch реализующий $logits = XW + b$, где W и b – параметры (`nn.Parameter`) модели. Иначе говоря, здесь реализуем своими руками модуль `nn.Linear` (в этом пункте его использование запрещено). Инициализируйте веса нормальным распределением (`torch.randn`).

Дайте ответ на вопрос 1: Сколько обучаемых параметров у получившейся модели?

Задание 3. Реализация цикла обучения

Дайте ответ на вопрос 2: Сколько итераций потребовалось, чтобы алгоритм сошелся?

Задание 4. Реализуйте **predict** и посчитайте **accuracy** на **test**.

Дайте ответ на вопрос 3: Какое accuracy получается после обучения?

Практическое занятие № 2

Датасет **MNIST**

Датасет MNIST содержит рукописные цифры. Загрузим датасет и создадим `DataLoader`-ы. Пример можно найти в семинаре по полносвязным нейронным сетям.

Полносвязные нейронные сети

Сначала решим MNIST с помощью полносвязной нейронной сети.

Задание 1. Простая полносвязная нейронная сеть

Создайте полносвязную нейронную сеть с помощью класса `Sequential`. Сеть состоит из:

Уплощения матрицы в вектор (`nn.Flatten`);

Двух скрытых слоёв из 128 нейронов с активацией `nn.ELU`;

Выходного слоя с 10 нейронами.

Задайте лосс для обучения (кросс-энтропия).

Задание 2. Дополните цикл обучения.

Задание 3. Протестируйте разные функции активации.

Попробуйте разные функции активации. Для каждой функции активации посчитайте массив **validation accuracy**. Лучше реализовать это в виде функции, берущей на вход активацию и получающей массив из **accuracies**.

Построим график **accuracy/epoch** для каждой функции активации.

Задание 4.

Дайте ответ на вопрос: Какая из активаций показала наивысший **accuracy** к концу обучения?

Практическое занятие № 3

Сверточные нейронные сети

Ядра. Сначала немного поработаем с самим понятием ядра свёртки.

Задание 1. Реализуйте LeNet

Если мы сделаем параметры сверток обучаемыми, то можем добиться хороших результатов для задач компьютерного зрения. Реализуйте архитектуру LeNet, предложенную еще в 1998 году! На этот раз используйте модульную структуру (без помощи класса `Sequential`).

Наша нейронная сеть будет состоять из

Свёртки 3x3 (1 карта на входе, 6 на выходе) с активацией ReLU;
MaxPooling-a 2x2;

Свёртки 3x3 (6 карт на входе, 16 на выходе) с активацией ReLU;
 MaxPooling-a 2x2;
 Уплощения (nn.Flatten);
 Полносвязного слоя со 120 нейронами и активацией ReLU;
 Полносвязного слоя с 84 нейронами и активацией ReLU;
 Выходного слоя из 10 нейронов.

Задание 2.

Дайте ответ на вопрос: Какое accuracy получается после обучения с точностью до двух знаков после запятой?

3. Формы аттестации и оценочные материалы по итогам освоения Модуля 4

Промежуточный контроль знаний слушателей определяется формированием отчета по практическому занятию (лабораторной работе) и выполнением практико-ориентированного кейса.

Пример практико-ориентированных кейсов.

Кейс №1

Цель: Применение искусственного интеллекта для поиска изображения воздушного судна на снимках местности при проведении поисково-спасательной операции путем обработки большого объёма отснятого материала в районе возможного местонахождения пропавшего воздушного судна.

Задание

Необходимо определить по снимкам, присутствует ли на них самолет.

Данные для обучения

31080 снимков: на 7899 из них есть самолёт, при этом 40 из них синтезированы

Тестовый набор:

1000 изображений: 200 из них содержат признак самолёта, при этом 60 изображений с признаком самолёта синтезированы

Данные:

avia-train.zip - изображения обучающей выборки

train.csv - разметка обучающей выборки.

признак 'sign': 0 - нет самолета, 1 - есть самолет

avia-test.zip - изображения тестовой выборки

test.csv - тестовая выборка (без разметки)

sample_submission.csv - пример решения

Метрика:

ROC AUC

Кейс №2

Разработать программный модуль учета популяции моржей на основе данных, полученных с беспилотных летательных аппаратов на территории заповедника «Ненецкий». Каждое лето около тысячи атлантических моржей создают крупнейшее лежбище на территории заповедника. Наблюдение за этой территорией дает ученым важные сведения о состоянии подвида, занесенного в Красную книгу России. Разработанное решение позволит оптимизировать время оценки численности по снимкам с большой территории и распознать численность стада моржей на лежбище или в прибрежной полосе по данным фотоснимков без подключения к сети интернет в случае скученности стада, а также с учетом особенностей местности.

Кейс №3

На основе методов искусственного интеллекта и обучения с подкреплением предлагается разработать стратегию планирования маршрута для группы роботов (агентов). Исследуемая среда представляет собой множество интеллектуальных агентов, которые действуют сообща для решения поставленных задач. В качестве решения нужно разработать не фиксированный подход, а скорее стратегию, которая поможет агентам достичь целей, избегая столкновений друг с другом и при этом минимизировать количество выполняемых действий. Предполагается, что решение кейса имеет широкую область применения, например: построение маршрутов поездов на железных дорогах, перемещение групп роботов по складам и их совместная кооперация в различных условиях, порядок прохождения трудных перекрестков для беспилотных автомобилей и т. д. При успешном решении такой задачи, создание автономной системы взаимодействия роботов становится на шаг ближе.

Кейс №4

На основе методов искусственного интеллекта предлагается разработать прототип системы поиска, классификации и визуализации аномалий в информации между базами вылова и переработки рыбопродуктов внутри РФ с выводом результатов в машиночитаемом и человекочитаемом виде. Раньше биологическому разнообразию почти ничего не угрожало. Но уже и тогда из-за отсутствия регулирования и непонимания некоторых биологических процессов бесследно исчезали целые виды живых существ. С развитием цивилизации стали появляться защитные механизмы в виде квот, а вместе с ними и механизмы контроля их выполнения. Однако сегодня суда все же могут уйти в другой порт, перегрузить вылов на плавбазу и использовать множество других возможностей «сокрытия» и обхода биологически обоснованных ограничений (на объемы, периоды, методы ловли и т. д.). Разработанное решение позволит контролировать несостыковки в механизмах ведения журналов вылова и переработки рыбопродуктов, которые предъявляются контролеру в случае посещения им судна. Это позволит сократить прецеденты незаконного улова, выявить контрафакты и нарушения технологических норм.

Кейс №5

Разработать систему для автоматического определения классификационного кода товара, который перемещается через таможенную границу стран-членов Евразийского экономического союза (ЕАЭС). Создание системы позволит декларантам сократить время, необходимое на определение классификационного кода товара, а также исключить возможные негативные издержки в виде административного и уголовного правонарушения. Кроме того, разработанное решение позволит должностным лицам таможенных органов ускорить процесс проведения таможенного оформления и выявить субъективные ошибки при классификации товара, что обеспечит увеличение уровня полноты и своевременности поступления денежных средств в бюджет РФ.

Кейс №6

Разработать удобную систему описания архива и поиска фото/видео файлов различного качества и объема, в котором фотографии автоматически описываются согласно таблице тегов по событиям.

В ходе своей деятельности отряд собирает уникальный архив фото/видео материала с поисков пропавших людей. Эти материалы помогают освещать проблему пропажи людей в социальных сетях и СМИ, проводить профилактику

как для детей, так и для взрослых на тему: что делать, чтобы не потеряться, что делать, если потерялся в городе и в лесу, и показывать жизнь отряда с целью привлечения к поискам новых добровольцев.

Разработанное решение позволит упростить описание материала, сделать удобным доступ и использование материала в хранилищах, а также решение позволит искать необходимые материалы по описанию.

Поисково-спасательный отряд «ЛизаАлерт» занимается поиском пропавших людей в лесу и в городе. Все поисковые мероприятия организуются силами добровольцев «ЛизаАлерт» на безвозмездной основе. За почти 12 лет отряд участвовал в поиске и спасении живыми более 115 000 человек.

Кейс №7

Необходимо помочь молодым пациентам стоматологов. Многие из детей с раннего возраста боятся врачей, лечащих зубы. Поэтому особо важно, чтобы работа с ними велась в игровой и увлекательной манере. Для этих целей в Правительстве Ханты-Мансийского автономного округа разрабатывается система по автоматическому определению состояния зубов маленьких пациентов при помощи фотоснимка и технологии искусственного интеллекта. Фотоснимки будут осуществляться посредством использования специального робота-помощника.

На основе фотоизображений ротовой полости, с применением технологий искусственного интеллекта, необходимо будет создать MVP-системы детекции с удобным web – интерфейсом по определению больных зубов у детей согласно заданного классификатора. При этом, важно обратить внимание на то, что использовать систему будут именно врачи, поэтому она должна быть легко адаптируема к любым возможностям ПК и понятна в использовании. Система должна носить рекомендательный характер.

Кейс №8

На основе представленных данных аэрофотосъемки, сформированных датасетов и материалов из открытых источников обучить нейросеть поиску белых медведей, а также сформировать интерфейс загрузки данных и представления результатов распознавания с учетом требования по автономности решения (без использования сети «Интернет»).

Распознавание белых медведей в Арктике на основании аэрофотосъемки – трудозатратная задача, учитывая площади аэрофотосъемки и особенности местности (белый медведь на белом снегу). Основная сложность задачи заключается в малом объеме выборки с объектами. Фотографии аэрофотосъемки с медведями в Арктике получить сложно. При подобной съемке собирается большое количество пустых фотографий, среди которых лишь незначительное число содержит объект поиска.

4. Учебно-методическое обеспечение и информационное обеспечение Модуля 4

Рекомендуемая литература:

1. Брюс Питер, Брюс Эндрю, Гедек Питер. Практическая статистика для специалистов Data Science, – СПб: БХВ-Петербург, 2021 – 352с - ISBN: 978-5-9775-6705-3
2. Рашка С. Машинное обучение с PyTorch и Scikit-Learn: Пер. с анг./ С. Рашка, Ю. Лю, В. Мирджалили. – Астана: Фолиант, 2024 – 688с., ISBN 978-601-11-0034-2
3. Рамсундар Бхарат. TensorFlow for Deep Learning: Пер. с англ. - СПб: БХВ-Санкт-Петербург, 2022, – 256 с., – ISBN 978-5-9775-4014-8
4. Брюс Питер, Брюс Эндрю, Гедек Питер. Практическая статистика для специалистов Data Science, – СПб: БХВ-Петербург, 2021 – 352с - ISBN: 978-5-9775-6705-3
5. Дж. Плас: Python для сложных задач. Наука о данных и машинное обучение – СПб:

Питер, 2022 – 576 с. – ISBN 978-5-4461-0914-2

6. Скиена Стивен С. Наука о данных. Учебный курс: Пер. с англ – М: Изд-во: Диалектика (Вильямс), 2020, – 544 с. – ISBN 978-5-907144-74-3
7. Грас Джоэл. Data Science. Наука о данных с нуля: Пер. с англ – 2-ое изд., перераб. И доп. – СПб: БХВ-Петербург, 2024 – 416 с: ил. ISBN 978-5-9775-6731-2.

Программное обеспечение и Интернет-ресурсы:

1. Jupyter Notebook.
2. Python и библиотека Pandas, NumPy, scikit-learn,
3. PyTorch, Tensorflow, Keras
4. Библиотеки визуализации данных Matplotlib, Seaborn, Altair, Plotly Express.
5. Платформа H2O.ai
6. Платформа RapidMiner
7. Loginom
8. Платформа Knime
9. RT.DataVision
10. Tableau
11. Yandex DataLens
12. Научная электронная библиотека eLIBRARY.RU.
13. Электронная образовательная среда в сети Интернет по адресу <http://edu.stankin.ru/>, функционирующая на базе программного обеспечения системы управления обучением Moodle.

5. Материально-технические условия реализации Модуля 4:

- оборудованные аудитории;
- ноутбук;
- мультимедиа проектор со звуковыми колонками;
- другие технические средства.

2.4 РАБОЧАЯ ПРОГРАММА СТАЖИРОВКИ

Рабочая программа стажировки (далее – рабочая программа) является частью дополнительной профессиональной программы профессиональной переподготовки ИТ-профиля «Аналитика данных и методы искусственного интеллекта» и направлена на формирование практических навыков по компетенциям: ПК-167 Разрабатывает и реализует архитектуру ансамбля моделей, ПК-171 Разрабатывает модули машинного обучения (МО) для решения задач, ПК -170 Осуществляет сбор и подготовку данных для обучения моделей искусственного интеллекта

Планируемые результаты стажировки:

По итогам освоения модуля слушатели должны:

Владеть:

- В 167.1 навыками реализации базовые архитектуры и модели искусственных нейронных сетей и градиентного бустинга
- В 170.1 навыками критического отбора данных и методами редукции размерности элементов набора данных
- В 170.2 навыками проверки данных на целостность и непротиворечивость.
- В 170.3 навыками использования методов поиска данных и достоверные источники данных, предварительной статистической обработки, разметки структурированных и неструктурированных данных
- В 171.1 навыками применения методов и критериев оценки качества моделей МО, определять критерии и метрики оценки результатов моделирования.
- В 171.2 навыкам решения задач анализа, прогнозирования, планирования, синтеза и принятия решений для машиностроительных предприятий

2. Содержание стажировки

Практический опыт применения Python для анализа технологических данных

Практический опыт применения методов искусственного интеллекта для анализа табличных технологических данных

Практический опыт использования хранилищ данных

Практический опыт аналитики больших данных и методов искусственного интеллекта на машиностроительных предприятиях

Практический опыт применения платформ науки о данных и машинного обучения и платформ бизнес- аналитики для решения задач машиностроительных предприятий

Задание для стажировки

Для получения положительной оценки задание на стажировку должно быть выполнено не менее 60%.

Цель: Разработка и тестирование приложения, основанного на машинном обучении/искусственного интеллекте.

Задачи:

- Запуск базовых моделей
- Оценка качества результата по релевантным для задачи метрикам
- Получение отчетов по результатам

Форма работы: индивидуальная / групповая (2-3 человека в группе)

Набор технологий:

- Python, Pandas, Numpy
- Pytorch / Tensorflow / Keras
- Google Colaboratory
- Другие подходящие для задачи библиотеки

План работы:

- Необходимо разработать базовую программную реализацию модели (допускается использование существующей реализации)
- Необходимо провести эксперименты с одним или более наборами данных и представить результаты в виде отчета в формате PDF

Требования к результатам

Код должен быть выложен на [github](#) / Google Colaboratory и удовлетворять следующим критериям:

Оценка за код задания будет распределена между следующими аспектами:

- функциональность,
- структура и организация кода, ○ инструкция для запуска моделей.

Оценка отчета и презентации состоит из следующих компонент:

- качество отчета,
- качество документации по наборам данных,
- качество слайдов с постановкой задачи, выбранным подходом и результатами

Результаты аттестационных испытаний определяются отметками «зачет» / «незачет» и объявляются после оформления в установленном порядке протоколов заседаний итоговой аттестационной комиссии.

Примеры заданий на стажировку:

Кейс 1: Оптимизация производственного процесса на станке с ЧПУ

Описание: Компания, занимающаяся производством деталей различного типа, использует станки с числовым программным управлением (ЧПУ) для обработки заготовок. Цель состоит в улучшении эффективности производства и сокращении времени цикла обработки на основе использования элементом машинного обучения. Необходимо на основе датасета найти узкие места производства (сокращение времени смены инструмента, изменения логистических цепочек)

Задачи:

- Разработка и внедрение алгоритма оптимального распределения инструментов и траекторий обработки.
- Предсказание времени работы оборудования и прогнозирование необходимости технического обслуживания.
- Разработка рекомендаций по оптимизации производственного процесса на основе данных аналитики.

Кейс 2: Управление качеством в производстве машин

Описание: Производитель машин сталкивается с проблемой нестабильного качества готовой продукции из-за различий в производственных процессах и неправильной настройки

оборудования. Необходимо на использовании датасетов первоначальных настроек оборудование попытаться обучить модель для подбора оптимальных параметров производства

Задачи:

- Анализ производственных процессов и выявление основных факторов, влияющих на качество продукции.
- Разработка модели прогнозирования качества продукции на основе данных об оборудовании, сырье и технологических параметрах.
- Автоматизация процесса контроля качества с использованием методов машинного обучения и компьютерного зрения.
- Внедрение системы управления качеством, которая позволит быстро реагировать на отклонения и предотвращать брак.

Кейс 3: Оптимизация процесса лазерной резки металла

Описание: Компания, специализирующаяся на лазерной резке металла, сталкивается с проблемой неравномерности качества резки и излишних затрат на материалы. На основе предлагаемого датасета подобрать режимы работы лазера

Задачи:

- Анализ и оптимизация параметров лазерной резки для различных типов металла и толщин.
- Разработка алгоритма автоматического выбора оптимальных параметров резки на основе характеристик заготовки и требуемых качественных характеристик.
- Создание модели прогнозирования качества резки на основе данных об оборудовании и характеристик металла.
- Реализация системы контроля качества резки с использованием датчиков и оптических систем.

Кейс 4: Разработка автономного мобильного робота для промышленных складов

Описание: Компания хочет разработать и внедрить автономного мобильного робота для автоматизации складских операций и оптимизации логистических процессов. Использую датасеты, которые описывают режимы работы роботов-тележек из перемещение. Необходимо составить оптимальные маршруты, сократить время загрузки и разгрузки складов

Задачи:

- Обучение и настройка алгоритмов машинного обучения для навигации, обнаружения препятствий и управления роботом.
- Тестирование и оптимизация работы робота в реальных условиях склада.

Кейс 5: Оптимизация производственных процессов с использованием индустриального интернета вещей (ПоТ)

Описание: Крупный завод хочет внедрить технологии Интернета вещей для мониторинга и оптимизации производственных процессов. Датасеты описывают данные, собираемые с датчиков ПоТ

Задачи:

- Разработка системы сбора, хранения и анализа данных о производственных процессах.
- Реализация алгоритмов машинного обучения для прогнозирования отказов оборудования и оптимизации его работы.

Кейс 6: Дефектоскопия сварных швов на производстве

Описание: Компания, занимающаяся производством металлических конструкций, сталкивается с проблемой выявления дефектов на сварных швах. Необходимо разработать систему компьютерного зрения для автоматического обнаружения дефектов на сварных соединениях.

Задачи:

- Использование алгоритмов компьютерного зрения для анализа изображений сварных швов.
- Обучение нейронных сетей на большом объеме данных с различными типами дефектов.
- Интеграция камер и системы компьютерного зрения на производственной линии.
- Разработка системы оповещения и классификации дефектов для операторов производства.

Кейс 7: Прогнозирование срока службы оборудования

Описание: Крупная промышленная компания хочет оптимизировать процесс замены и обслуживания оборудования, предсказывая срок его службы с помощью анализа данных и машинного обучения. Датасеты для этой задачи собирались с использованием ПоТ датчиков и API систем управления

Задачи:

- Сбор и анализ данных о параметрах работы оборудования, таких как температура, вибрация, нагрузка и т.д. (можно использовать готовый датасет)
- Применение методов машинного обучения, таких как регрессия и временные ряды, для прогнозирования срока службы оборудования.
- Разработка модели, учитывающей различные факторы, влияющие на износ оборудования.
- Внедрение системы мониторинга и оповещения, предупреждающей о необходимости замены или технического обслуживания оборудования.

Кейс 8: Анализ и прогнозирование потребления энергии на производстве

Описание: Производственная компания хочет оптимизировать расход энергии на производстве, путем анализа данных и прогнозирования потребления энергии. На производстве используется большое количество оборудования, которое потребляет сотни киловатт энергии, при этом зачастую оборудование работает не оптимизировано, не в соответствии с планом

Задачи:

- Сбор данных о потреблении энергии на различных участках производства.

- Применение методов временных рядов и прогнозирования для предсказания будущего потребления энергии.
- Использование алгоритмов машинного обучения для выявления закономерностей и факторов, влияющих на потребление энергии.
- Разработка модели оптимального управления потреблением энергии для снижения затрат и повышения энергоэффективности производства.

Кейс 9: Анализ и оптимизация производственного цикла

Описание: Производственная компания хочет улучшить эффективность производственных процессов, анализируя данные о производственных циклах и оптимизируя их.

Задачи:

- Сбор и анализ данных о продолжительности и последовательности операций в производственном цикле.
- Применение методов анализа данных и машинного обучения для выявления узких мест и оптимизации производственного процесса.
- Разработка модели предсказания продолжительности производственных циклов на основе исторических данных и текущих условий производства.
- Внедрение системы мониторинга и управления производственными циклами, позволяющей быстро реагировать на изменения и оптимизировать процесс.

Кейс 10: Оптимизация распределения ресурсов на производственной линии

Описание: Крупное производство сталкивается с проблемой неэффективного распределения ресурсов на производственной линии, что приводит к задержкам и увеличению времени цикла производства.

Задачи:

1. Анализ производственных процессов и потоков материалов.
2. Определение оптимальной последовательности операций и распределения ресурсов.
3. Разработка модели оптимизации, учитывающей ограничения и приоритеты производства.
4. Внедрение системы мониторинга и контроля выполнения задач на производственной линии.
5. Анализ эффективности и корректировка стратегии оптимизации в соответствии с изменяющимися условиями.

Кейс 11: Автоматизация контроля качества продукции на линии упаковки

Описание: Компания, упаковывающая продукцию на производственной линии, сталкивается с проблемой неэффективного контроля качества упаковки, что приводит к возможным ошибкам и браку.

Задачи:

1. Разработка системы компьютерного зрения для автоматического контроля качества упаковки.
2. Обучение нейронных сетей на образцах корректной и некорректной упаковки.
3. Интеграция системы контроля качества в производственную линию.
4. Оптимизация процесса упаковки с учетом результатов контроля качества.

Кейс 12: Автоматизация складской логистики с использованием IoT

Описание: Компания с большим складом сталкивается с проблемой неэффективного управления запасами и перемещением товаров на складе, что приводит к задержкам и ошибкам в отгрузках. Все данные или заносятся вручную или используют простые системы ведения учета производственных процессов

Задачи:

1. Оптимизация маршрутов перемещения товаров с использованием алгоритмов машинного обучения (на основе датасетов карт перемещения и загрузки обоурдования)
2. Анализ эффективности и корректировка стратегии логистики в соответствии с изменяющимися потребностями и условиями рынка.

Кейс 13: Прогнозирование спроса на продукцию

Описание: Производственная компания сталкивается с проблемой неоптимального планирования производства из-за неэффективного прогнозирования спроса на продукцию. В какие-то моменты времени спрос превышает предложения, в какие-то – наоборот.

Задачи:

1. Анализ данных о прошлых продажах и факторах, влияющих на спрос
2. Разработка модели прогнозирования спроса с использованием методов временных рядов и анализа данных.
3. Оптимизация производственных планов и запасов с учетом прогнозов спроса (на основе представляемых датасетов)
4. Анализ эффективности прогнозирования и корректировка моделей в соответствии с изменениями на рынке и внутренними факторами компании.

Кейс 14: Анализ и оптимизация энергопотребления в промышленном комплексе

Описание: Крупный промышленный комплекс сталкивается с проблемой высоких затрат на энергию и неэффективного использования энергоресурсов, особенно это связано с системами охлаждения подсистем высоконагруженного оборудования и системами кондиционирования

Задачи:

1. Анализ данных о потреблении энергии на различных участках производства.

2. Использование методов машинного обучения для выявления закономерностей и оптимизации потребления энергии на основе датасетов о потреблении энергия в определенные промежутки времени с указанием задействованного оборудования.

3. Разработка модели оптимального управления энергопотреблением с учетом производственных процессов и целей компании.

4. Анализ результатов оптимизации и корректировка стратегии управления энергоресурсами в соответствии с изменяющимися условиями и требованиями.

Кейс 15

Необходимо построить модель и обучив ее на имеющихся данных выявлять дефекты в готовой продукции.

Описание

Данные по выпуску продукции на экструдере. Экструдер состоит из трех отдельных установок, которые принимают гранулированное сырье, а затем растягивают его в пластиковую пленку путем нагрева и вытягивания в три слоя. Качество выпускаемой продукции контролируется по двум аспектам: параметрами продукции и ее дефектам. Такие параметры, как толщина пленки и растяжение, контролируются датчиками, но дефекты наблюдаются оператором только визуально, и не всегда в реальном времени (существует практика определения времени дефекта по длине пленки и маркерам закладки). Технологи производства связывают эти дефекты с несколькими факторами, среди которых неисправность оборудования и низкое качество поступающего сырья.

Самый важный вид дефекта – обрыв пленки, и именно его заказчик просит свести к минимуму. Это происходит по разным причинам, некоторые из которых имеют механическую природу, например, сор, попавший в «головку» экструдера. Некоторые из них происходят из-за других типов дефектов. Три самых распространенных дефекта – это дырки на полотне, недоплавка сырых гранул и выход окалины. В настоящее время практика наблюдения этих дефектов на данных установки отсутствует; лабораторные испытания образцов максимальны. Подобные дефекты не обязательно приводят к обрывам.

РАЗДЕЛ 3. ОЦЕНКА КАЧЕСТВА ОСВОЕНИЯ ПРОГРАММЫ «АНАЛИТИКА ДАННЫХ И МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА»

Оценка качества освоения программы включает *текущую и итоговую аттестацию* обучающихся.

Текущая аттестация – служит для оценки объёма и уровня усвоения слушателем учебного материала одного модуля программы и применяется в рамках текущего контроля успеваемости слушателя. Учебные достижения обучающихся по всем видам учебных заданий в ходе текущего контроля.

Итоговая аттестация выполняется сдачей итоговой аттестационной работы.

К итоговой аттестации допускаются лица, выполнившие требования, предусмотренные программой, и успешно прошедшие все оценочные процедуры, предусмотренные программами профессиональных модулей.

Форма итоговой аттестации по программе: защита итогового проекта

Слушателям после успешного окончания обучения (выполнившим все требования учебного плана) выдаются документы установленного образца о профессиональной переподготовке (диплом о профессиональной переподготовке).

Итоговая аттестация слушателей проводится в сроки, предусмотренные учебным планом и графиком учебного процесса.

Форма и условия аттестационных испытаний определяются программой

Задание для итоговой аттестации.

Для получения положительной оценки итоговое задание должно быть выполнено не менее 60%.

Состав итогового проекта

1. Постановка задачи

- подробно описать производственную задачу, ее актуальность для заказчика
- определить круг стейкхолдеров по задаче, провести сбор бизнес-требований по задаче, сформулировать гипотезы для проверки
- определить метрики для проверки гипотез, приведены источники информации для сбора данных для расчета таких метрик

2. Анализ

- провести исследование аналогичных решений, описать, привести результаты их применения (если доступно) и ссылки на материалы
- провести предварительный анализ данных, наглядно описать их качество, аномалии, зависимости, приведены ссылки на код
- описать и обосновать алгоритмы и техники, применяемые для решения задачи, их требования к исходным данным
- сформировать дизайн отчетности по бизнес-задаче, выбрана форма представления итоговых результатов

3. Методика решения

- описать шаги преобразования и очистки данных, привести ссылки на рабочий код, воспроизводящий все описанные шаги
- в результате очистки данных найти ранее аномалии и зависимости устраниены или существенно снижены, качество данных повышено
- обосновать выбор методов работы с данными и метрик для решения бизнес-задачи, построить систему метрик, соответствующая содержанию исходной задачи
- описать сценарий проверки гипотез, возникших при анализе исходных бизнес-требований

- описать модель данных и преобразования для построения итогового отчета/дашборда по бизнес-задаче

4. Результаты

- для сформулированных гипотез по бизнес-задаче реализовать и описать результаты проверок (опционально - в формате цикла HADI)

- результаты работы визуализировать, подготовить отчет/презентация для представления стейкхолдерам

- привести сравнение качества итогового решения бизнес-задачи с аналогичными решениями (если доступно)

- описать область и результаты внедрения решения исходной задачи (если применимо)

5. Выводы и заключение

- привести основные выводы по проделанной работе, ключевые пункты визуализированы и четко описаны

- перечислить пути развития и улучшения решения

Результаты аттестационных испытаний определяются отметками «зачтено», «незачтено», «отлично», «хорошо», «удовлетворительно», «неудовлетворительно» и объявляются после оформления в установленном порядке протоколов заседаний итоговой аттестационной комиссии.

При проведении итоговой аттестации в форме зачета предлагается следующая шкала оценок:

Отметка «**незачтено**» выставляется слушателю, не показавшему освоение планируемых результатов (знаний, умений, компетенций) предусмотренных программой, допустившему серьезные ошибки в выполнении предусмотренных программой заданий, не справившемуся с выполнением аттестационной работы.

Отметку «**зачтено**» заслуживает слушатель, показавший освоение планируемых результатов (знаний, умений, компетенций) предусмотренных программой, сформированность новых компетенций и профессиональных умений для осуществления профессиональной деятельности, знакомый с литературой, публикациями по программе.

Отметку «**отлично**» заслуживает слушатель, чей ответ отличается полнотой в раскрытии содержания вопросов билета, свободным владением материала, оперированием категориями, понятиями, технологиями, связью теории с практической деятельностью, даны четкие и правильные ответы на дополнительные вопросы.

На «**хорошо**» оценивается ответ не в достаточно свободной форме раскрывающий содержание вопросов билета, допущены погрешности при формулировке определений, небольшие неточности при ответах на дополнительные вопросы.

На «**удовлетворительно**» оценивается ответ, не полностью раскрывающий содержание вопросов билета или ответ дан на один вопрос билета, допущены неточности при ответах на дополнительные вопросы.

«**Неудовлетворительно**» оценивается ответ, не раскрывающий содержание вопросов билета и не раскрывающий дополнительные вопросы.

В случае несогласия с результатами итоговой аттестации слушатель имеет право подать апелляцию в аттестационную комиссию на имя председателя.

Апелляция принимается в день сдачи устного итогового экзамена сразу после объявления оценок. Апелляция по письменным итоговым экзаменам принимается в день объявления оценки по письменному испытанию.

В ходе рассмотрения апелляции проверяется правильность и аргументированность выставленной оценки. При необходимости АК принимает решение о возможности пересдачи итогового экзамена или повторной защиты выпускной квалификационной работы. Итоги рассмотрения апелляции вносятся в соответствующий протокол в раздел «Особое мнение комиссии».

АК рассматривает апелляцию по итоговым экзаменам и сообщает свое решение не позднее, чем через сутки с момента подачи апелляции. При принятии решения о возможности пересдачи итогового экзамена комиссией анализируется лист ответа слушателя и при необходимости ведется беседа по дополнительным вопросам. Решение АК по апелляции является окончательным.

Самостоятельная работа слушателей подразделяется на аудиторную и внеаудиторную.

Аудиторную самостоятельную работу составляют отчеты, оформленные по результатам проведения практических занятий.

Внеаудиторная самостоятельная работа включает формы: *написание контрольных работ посредством ЭОС* (электронной образовательной среды) по результатам проведения практических занятий.

Основными критериями качества организации самостоятельной работы служит наличие контроля результатов самостоятельной работы и технических условий выполнения заданий.

Основными современными формами организации самостоятельной работы слушателей являются творческие работы и работа с информационными компьютерными технологиями.

РАЗДЕЛ 4. ОРГАНИЗАЦИОННО-ПЕДАГОГИЧЕСКИЕ УСЛОВИЯ РЕАЛИЗАЦИИ ПРОГРАММЫ «АНАЛИТИКА ДАННЫХ И МЕТОДЫ ИСКУССТВЕННОГО ИНТЕЛЛЕКТА»

4.1. Материально-технические условия и организационное обеспечение реализации дополнительной профессиональной программы профессиональной переподготовки

Для эффективного осуществления образовательного процесса необходимы:

1. Мультимедийная аудитория, оснащенная оборудованием (стационарным или переносным): проектор, экран, компьютер (имеющий доступ в Интернет и ЭИОС университета), доской (маркерной), партами, стульями, местом для преподавателя – стол и стул.

2. Компьютерный класс, оснащенный компьютерами (имеющими доступ в Интернет и ЭИОС университета), организованных в локальную сеть, позволяющий проводить занятия с использованием программных продуктов, указанных в разделе 4.7, доской (меловой или маркерной), партами, стульями, местом для преподавателя – стол и стул.

3. Помещения для самостоятельной работы, оснащенные компьютерами (имеющими доступ в Интернет и ЭИОС университета), организованные в локальную сеть, доской (меловой или маркерной), партами, стульями, местом для преподавателя – стол и стул.

4. Компьютеры, имеющие доступ в Интернет и ЭИОС Университета, с использованием программных продуктов, указанных в разделе 4.2.

При реализации программы дополнительного образования:

1. Слушателям предлагается научная и учебно-методическая помощь квалифицированных преподавателей.
2. Все взаимодействие со слушателями (рассылка новостей, информационные материалы, техническая поддержка и пр.) осуществляется посредством ЭОС (электронной образовательной среды).

4.2. Учебно-методическое обеспечение реализации дополнительной профессиональной программы профессиональной переподготовки

Рекомендуемая литература:

1. Горбаченко В.И. Машинальное обучение: учебное пособие / Горбаченко В.И., Савенков К.Е., Малахов М.А. – Москва: Ай Пи Ар Медиа, 2023. – 217 с. – ISBN 978-5-4497-1860-0. – Текст: электронный // IPR SMART: [сайт]. – URL: <https://www.iprbookshop.ru/125886.html> (дата обращения: 25.03.2023). – Режим доступа: для авторизованных пользователей. - DOI: <https://doi.org/10.23682/125886>

2. Зуев С.В. Методы анализа данных: учебное пособие / Зуев С.В.. – Москва: Ай Пи Ар Медиа, 2023. – 132 с. – ISBN 978-5-4497-2107-5. – Текст: электронный // IPR SMART: [сайт]. – URL: <https://www.iprbookshop.ru/129059.html> (дата обращения: 25.03.2023). – Режим доступа: для авторизованных пользователей. - DOI: <https://doi.org/10.23682/129059>

3. Сузи Р.А. Язык программирования Python: учебное пособие / Сузи Р.А.. – Москва: Интернет-Университет Информационных Технологий (ИНТУИТ), Ай Пи Ар Медиа, 2020. – 350 с. – ISBN 978-5-4497-0705-5. – Текст: электронный // IPR SMART: [сайт]. – URL: <https://www.iprbookshop.ru/97589.html> (дата обращения: 25.03.2023). – Режим доступа: для авторизованных пользователей

4. Меликов, П. И. Изучаем основы Python. Практический курс для data-аналитиков / П. И. Меликов. – Москва, Алматы: Ай Пи Ар Медиа, EDP Hub (Идиши Хаб), 2023. – 480 с. – ISBN 978-5-4497-2162-4, 978-601-81002-1-5. – Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. – URL: <https://www.iprbookshop.ru/130920.html> (дата обращения: 25.03.2023). – Режим доступа: для авторизованных пользователей

5. Маккини, У. Python и анализ данных / У. Маккини ; перевод А. А. Слинкин. – 2-е

изд. – Москва: ДМК Пресс, 2020. – 540 с. – ISBN 978-5-97060-590-5. – Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. – URL: <https://www.iprbookshop.ru/125361.html> (дата обращения: 25.03.2023). – Режим доступа: для авторизированных пользователей

6. Бюиссон, Ф. Анализ поведенческих данных на R и Python / Ф. Бюиссон ; перевод А. В. Логунов. – Москва: ДМК Пресс, 2022. – 368 с. – ISBN 978-5-97060-992-7. – Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. – URL: <https://www.iprbookshop.ru/125286.html> (дата обращения: 25.03.2023). – Режим доступа: для авторизированных пользователей

7. Хилл, К. Научное программирование на Python / К. Хилл ; перевод А. В. Снастин. – Москва: ДМК Пресс, 2021. – 646 с. – ISBN 978-5-97060-914-9. – Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. – URL: <https://www.iprbookshop.ru/125125.html> (дата обращения: 25.03.2023). – Режим доступа: для авторизированных пользователей

8. Железнов, М. М. Методы и технологии обработки больших данных: учебно-методическое пособие / М. М. Железнов. – Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. – URL: <https://www.iprbookshop.ru/101802.html> (дата обращения: 15.06.2024). – Режим доступа: для авторизированных пользователей.

9. Дружинин, Д. В. Высокопроизводительные вычисления и облачные технологии: учебное пособие / Д. В. Дружинин. – Томск: Издательство Томского государственного университета, 2020. – 93 с. – ISBN 978-5-94621-921-1. – Текст: электронный // Цифровой образовательный ресурс IPR SMART: [сайт]. – URL: <https://www.iprbookshop.ru/116813.html> (дата обращения: 15.06.2024). – Режим доступа: для авторизированных пользователей

10. Скиена Стивен С. Наука о данных. Учебный курс: Пер. с англ – М: Изд-во: Диалектика (Вильямс), 2020, – 544 с. – ISBN 978-5-907144-74-3

11. Грас Джоэл. Data Science. Наука о данных с нуля: Пер. с англ – 2-ое изд., перераб. И доп. – СПб: БХВ-Петербург, 2024 – 416 с: ил. ISBN 978-5-9775-6731-2.

12. Брюс Питер, Брюс Эндрю, Гедек Питер. Практическая статистика для специалистов Data Science, – СПб: БХВ-Петербург, 2021 – 352с - ISBN: 978-5-9775-6705-3

13. Куслейка Дик. Визуализация данных при помощи дашбордов и отчетов в Excel, - М.: ДМК Пресс, 2022, - 338 с., – ISBN 978-5-9706-0966-8

14. Дж. Плас: Python для сложных задач. Наука о данных и машинное обучение – СПб: Питер, 2022 – 576 с. – ISBN 978-5-4461-0914-2

15. Груздев А.В. Предварительная подготовка данных в Python. Том 1. Инструменты и валидация – М.: ДМК Пресс, 2023 – 816 с. – ISBN 978-5-93700-156-6

16. Груздев А.В. Предварительная подготовка данных в Python. Том 2. План, примеры и метрики качества– М.: ДМК Пресс, 2023 – 814 с. – ISBN 978-5-93700-177-1

17. Лакшманан В., Тайджани Д. Google BigQuery. Всё о хранилищах данных, аналитике и машинном обучении, СПб: Питер, 2021. – 496с. ISBN: 978-5-4461-1707-9

18. Петров А. Распределенные данные. Алгоритмы работы современных систем хранения информации – СПб.: Питер, 2022. – 336 с.

19. Стренгхольт П. Масштабируемые данные. Лучшие шаблоны высоконагруженных архитектур. – СПб.: Питер, 2022 – 368с.: ил. – ISBN 987-5-4461-1461-0.

20. Рамсундар Бхарат. TensorFlow for Deep Learning: Пер. с англ. - СПб: БХВ-Санкт-Петербург, 2022, – 256 с., – ISBN 978-5-9775-4014-8

21. Рашка С. Машинное обучение с PyTorch и Scikit-Learn: Пер. с англ./ С. Рашка, Ю. Лю, В. Мирджалили. – Астана: Фолиант, 2024 – 688с., ISBN 978-601-11-0034-2

22. Рамсундар Бхарат. TensorFlow for Deep Learning: Пер. с англ. - СПб: БХВ-Санкт-Петербург, 2022, – 256 с., – ISBN 978-5-9775-4014-8

Программное обеспечение и Интернет-ресурсы:

1. ЭБС «Университетская библиотека онлайн» издательства «Директ-Медиа»// Режим

- доступа URL: www.biblioclub.ru;
2. Электронные издания научно-технической библиотеки, размещенные в разделе университета в ЭБС «Университетская библиотека онлайн»;
 3. Электронная образовательная среда в сети Интернет // Режим доступа URL: <http://edu.stankin.ru/>;
 4. Сайт университета в сети Интернет по адресу // Режим доступа URL: <http://www.stankin.ru/>;
 5. Электронная библиотека научных публикаций «Российский индекс научного цитирования» // Режим доступа URL: <https://elibrary.ru/defaultx.asp>;
 6. ЭБС «IPRBOOKS» группа компаний IPR MEDIA// Режим доступа URL: www.iprbookshop.ru;
 7. MS Excel.
 8. Google-таблицы.
 9. Jupyter Notebook.
 10. Google Looker Studio.
 11. Google Colaboratory
 12. Python и библиотека Pandas, NumPy, scikit-learn
 13. Библиотеки визуализации данных Matplotlib, Seaborn, Altair, Plotly Express.
 14. Apache Spark
 15. RT.DataLake
 16. RT.WarehouseGoogle Looker Studio.
 17. RT.DataVision
 18. PyTorch, Tensorflow, Keras
 19. Платформа H2O.ai
 20. Платформа RapidMiner
 21. Loginom
 22. Платформа Knime
 23. Tableau
 24. Библиотека размеченных данных www.kaggle.com
 25. Yandex DataLens

Форма дополнений и изменений в рабочей программе**Дополнения и изменения в рабочей программе на 20__/20__ уч.г.**

Внесенные изменения на 20__/20__ учебный
год

УТВЕРЖДАЮ
Ректор/курирующий проректор

(подпись, расшифровка подписи)

“ ____ ” _____ 20.... г

В рабочую программу вносятся следующие изменения:

- 1);
- 2);
- 3)

или делается отметка о нецелесообразности внесения каких-либо изменений на данный
учебный год

Рабочая программа пересмотрена и одобрена на заседании МИРИТ

(дата, номер протокола заседания кафедры, подпись зав. кафедрой).