

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение
высшего образования
«Московский физико-технический институт
(национальный исследовательский университет)»
(МФТИ, Физтех)

УТВЕРЖДАЮ

Ректор МФТИ

д-р физ.-мат. наук, профессор

Д. В. Ливанов



2024 г.

**Дополнительная профессиональная
программа профессиональной переподготовки
«Аналитик данных»**

УГСН 09.00.00 Информатика и вычислительная техника
Направление подготовки 09.03.01 Информатика и вычислительная техника
ОКВЭД 62.01 Разработка компьютерного программного обеспечения

Москва 2024

1. Общая характеристика программы

Реализация программы профессиональной переподготовки «Аналитик данных» направлена на получение компетенции, необходимой для выполнения нового вида профессиональной деятельности. Программа профессиональной переподготовки разрабатывается на основании установленных квалификационных требований к должности «Аналитик», профессионального стандарта 06.042 «Специалист по большим данным» и требований федерального государственного образовательного стандарта высшего образования по направлению подготовки 09.03.01 «Информатика и вычислительная техника» к результатам освоения образовательной программы.

Дополнительная профессиональная программа профессиональной переподготовки является модульной. Каждый модуль представляет собой завершённую программу, направленную на формирование/совершенствование у слушателей конкретных компетенций.

1.1.Цель реализации программы

Целью реализации программы профессиональной переподготовки «Аналитик данных» является совершенствование компетенций, необходимых для профессиональной деятельности в области финансов с применением технологии ИИ для анализа данных, прогнозирования рыночных тенденций и управления рисками.

Слушателям, успешно прошедшим итоговую аттестацию по дополнительной профессиональной программе профессиональной переподготовки, выдается диплом с присвоением квалификации «Аналитик данных» установленного образца.

1.2.Совершенствуемые и/или приобретаемые компетенции

Компетенции, формируемые и совершенствуемые в результате обучения, представлены в таблицах 1 и 2.

Таблица 1

№	Компетенции в соответствии с профессиональным стандартом 06.042 Специалист по большим данным	Код компетенции
1	Способен подготовить данные для проведения аналитических работ по исследованию больших данных	ПК-1
2	Способен проводить аналитические исследования с применением технологий больших данных в соответствии с требованиями заказчика	ПК-2

Таблица 2

№	Компетенции в соответствии с направлением подготовки 09.03.01 Информатика и вычислительная техника (ФГОС)	Код компетенции
1	Способен разрабатывать алгоритмы и программы, пригодные для практического применения	ОПК-8
2	Способен осваивать методики использования программных средств для решения практических задач	ОПК-9

1.3.Планируемые результаты обучения

Планируемые результаты обучения представлены в таблице 3.

Таблица 3

№	Уметь - знать	Направление подготовки 09.03.01 Информатика и вычислительная техника ФГОС ВО
		Квалификация: бакалавр
		Код компетенции
	Знать: – принципы организации кода – понятия вектора и матрицы, векторного пространства, нормы вектора, ортогональности и гиперплоскости – как выполнять базовые операции над векторами и матрицами – основные методы обработки данных, упрощающие решение задач машинного обучения – в каких сферах применяется машинное обучение – основные понятия машинного обучения: датасет (выборка), объект, признак, таргет, матрица объект-признак, машинное обучение с учителем, таргет, модель, предсказание, функция потерь, параметр, гиперпараметр – формальную постановку задачи машинного обучения с учителем – как строится сверточная нейронная сеть – различные архитектуры, используемые в машинном обучении Уметь: – использовать сверточную нейронную сеть для обработки данных	ОПК-8
	Знать: – понятие коллекции – назначение коллекций в разработке – в каких задачах машинного обучения используются линейные модели – основные понятия теории вероятностей – понятие условной вероятности, дискретных и непрерывных случайных величин – в каких задачах можно применить наивный Байесовский классификатор – как строится нейронная сеть – проблемы затухающих и взрывающихся градиентов Уметь: – сопоставлять и выбрать необходимую структуру данных для конкретной практической задачи – формально поставить задачу линейной регрессии	ОПК-9

		Профессиональный стандарт 06.042 Специалист по большим данным Код компетенции
	Знать: – понятия правдоподобия в задачах машинного обучения – различные метрики оценки качества классификации – основные подходы градиентной оптимизации – основные техники регуляризации в глубоком обучении Уметь: – решать задачи линейной классификации в машинном обучении – бороться с переобучением нейросети с помощью регуляризации	ПК-1
	Знать: – метод опорных векторов, используемый для задач классификации и регрессионного анализа – способы создания нелинейного классификатора с помощью так называемого ядерного трюка (kernel trick) – как использовать решающие деревья в задаче регрессии – основные нелинейные функции активации – механизмы обратного распространения ошибки Уметь: – использовать решающие деревья в задачах машинного обучения – описывать линейные и нелинейные зависимости с помощью нейросетей	ПК-2

1.4. Категория обучающихся

Программа профессиональной переподготовки предназначена для специалистов в области решения профессиональных задач по работе с данными с помощью основных методов машинного и глубокого обучения, имеющих высшее или среднее профессиональное образование, а также лиц, получающих высшее или среднее профессиональное образование.

Требования к слушателям:

К освоению ДПП ПП допускаются лица, освоившие ОПОП ВО бакалавриата – в объеме не менее первого курса (бакалавры 2-го курса), освоившие ОПОП ВО специалитета – не менее первого и второго курсов (специалисты 3-го курса), учащиеся ОПОП ВО магистратур не ИТ-профиля.

1.5.Форма обучения

Очно-заочная с применением дистанционных образовательных технологий.

Программа может быть реализована в сетевой форме.

1.6.Объем программы

423 академических часа.

1.7.Режим обучения

13 недель (32 часа в неделю).

2. Содержание программы

2.1.Учебный (тематический) план

Учебный (тематический) план программы представлен в таблице 4.

Таблица 4

№ п/п	Наименование разделов (модулей) и тем	Всего, час.	Лекции	Практич. работа	Самост. работа	Форма контроля
УЧЕБНЫЙ ПЛАН						
1	Математические методы анализа данных	180	14	14	152	Тест
2	Методы искусственного интеллекта в анализе данных	135	10	16	109	Тест
3	Инжиниринг данных	90	12	12	66	Тест
	Итоговая аттестация	18	-	4	14	Экзамен
УЧЕБНО-ТЕМАТИЧЕСКИЙ ПЛАН						
Модуль 1. Математические методы анализа данных						
1.1	Комбинаторика и дискретная вероятность	36	3	3	30	тест
1.2	Случайные величины	36	3	3	30	тест
1.3	Основы математической статистики	36	3	3	30	тест
1.4	Методы построения оценок и статистические гипотезы	36	3	3	30	тест
1.5	Анализ временных рядов и прогнозирование	33	2	1	32	тест
	Промежуточная (итоговая) аттестация по модулю 1	1		1		зачет
	Итого по модулю 1	180	14	14	152	
Модуль 2. Методы искусственного интеллекта в анализе данных						
2.1	Введение в машинное обучение	27	2	3	22	тест
2.2	Обучение с учителем	27	2	3	22	тест
2.3	Обучение без учителя	27	2	3	22	тест
2.4	Обучение с подкреплением	27	2	3	22	тест
2.5	Основы NLP и LLM	26	2	3	21	тест
	Промежуточная (итоговая) аттестация по модулю 2	1		1		зачет
	Итого по модулю 2	135	10	16	109	
Модуль 3. Инжиниринг данных						
3.1	Основы инжиниринга данных: типы и структуры данных	15	2	2	11	тест

2.3. Учебная программа

Содержание учебной программы приведено в таблице 5.

Таблица 5

№ п/п	Наименование модуля, разделов и тем	Содержание обучения, наименование и тематика практических занятий (семинаров), самостоятельной работы	Объем, ак. час.
1	Математические методы анализа данных	Лекция 1. Комбинаторика и дискретная вероятность. 2. Случайные величины. 3. Основы математической статистики. 4. Методы построения оценок и статистические гипотезы. 5. Анализ временных рядов и прогнозирование. Практическая работа Выполнение заданий по теме лекции Самостоятельная работа Самостоятельное выполнение заданий по теме лекции. Изучение учебной и научной литературы по теме лекции.	180
2	Методы искусственного интеллекта в анализе данных	Лекция 1. Введение в машинное обучение. 2. Обучение с учителем. 3. Обучение без учителя. 4. Обучение с подкреплением. 5. Основы NLP и LLM. Практическая работа Выполнение заданий по теме лекции Самостоятельная работа Самостоятельное выполнение заданий по теме лекции. Изучение учебной и научной литературы по теме лекции.	135
3	Инжиниринг данных	Лекция 1. Основы инжиниринга данных: типы и структуры данных 2. Системы хранения данных: реляционные и NoSQL базы данных 3. Обработка и трансформация данных: инструменты и методы 4. Разработка ETL/ELT процессов: от теории к практике 5. Оптимизация производительности и масштабирование систем данных 6. Безопасность данных: принципы и лучшие практики Практическая работа Выполнение заданий по теме лекции Самостоятельная работа	90

№ п/п	Наименование модуля, разделов и тем	Содержание обучения, наименование и тематика практических занятий (семинаров), самостоятельной работы	Объем, ак.час.
		Самостоятельное выполнение заданий по теме лекции. Изучение учебной и научной литературы по теме лекции.	
	Итоговая аттестация	экзамен	18
	Итого:		423

3. Формы аттестации и оценочные материалы

3.1.Формы аттестации

Оценка качества освоения программы включает текущую, промежуточную 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.

Контроль качества знаний осуществляется преподавателем.

Текущий контроль успеваемости слушателей проводится в форме тестирования на платформе дистанционного обучения.

Объектом оценивания выступают: формируемые знания и умения.

Промежуточная аттестация слушателей включает сдачу зачетов в форме тестирования на платформе дистанционного обучения.

Итоговая аттестация обучающихся проходит в форме экзамена в формате тестирования, в ходе которого обучающиеся используют и демонстрируют необходимые компетенции, предусмотренные в процессе освоения программы. Доступ к прохождению итоговой аттестации предоставляется слушателям, успешно прошедшим промежуточную аттестацию по всем модулям программы.

Итоговая аттестация по программе является обязательной и проводится в сроки, предусмотренные учебным планом и графиком учебного процесса.

Самостоятельная работа слушателей подразделяется на аудиторную и внеаудиторную. Аудиторную самостоятельную работу составляют практические задания, которые выполняются слушателями во время учебных занятий, результаты ее выполнения проверяются и оцениваются преподавателем в учебном процессе.

Внеаудиторная самостоятельная работа включает формы: изучение дополнительной литературы, подготовка итоговых проектов по модулям, подготовка проекта к итоговой аттестации.

Основными критериями качества организации самостоятельной работы служит наличие контроля результатов самостоятельной работы.

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

Оценка «зачтено» выставляется обучающемуся, если он показал всесторонние, систематизированные, глубокие знания учебной программы курса и умение уверенно применять их на практике при решении конкретных задач, свободное и правильное обоснование принятых решений.

Оценка «не зачтено» выставляется обучающемуся, который не знает большей части основного содержания учебной программы курса, допускает грубые ошибки в формулировках основных понятий дисциплины и не умеет использовать полученные знания при решении типовых практических задач.

Максимальная сумма, которую можно набрать, успешно выполнив все контрольные мероприятия, составляет 100 баллов. Для получения положительной оценки «зачтено» необходимо набрать не менее 70% за итоговые задания каждого модуля и итоговую аттестацию.

В ведомость итоговой аттестации оценка выставляется в соответствии с нижеприведенной таблицей 6.

Таблица 6

Сумма баллов	Оценка
50 - 100	Зачтено
менее 49	Не зачтено

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

Апелляция принимается в день сдачи итогового экзамена сразу после объявления оценок.

В ходе рассмотрения апелляции проверяется правильность и аргументированность выставленной оценки. При необходимости ИАК 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».

ИАК 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 ИАК по апелляции является окончательным.

3.2.Оценочные материалы

Таблица 7

Наименование модуля, разделов и тем	Основные показатели оценки	Формы и методы контроля и оценки	Вес задания, %
1. Математические методы анализа данных	ПК-1, ПК-2, ОПК-8, ОПК-9	тестирование на платформе дистанционного обучения	20
2. Методы искусственного интеллекта в анализе данных	ПК-1, ПК-2, ОПК-8, ОПК-9	тестирование на платформе дистанционного обучения	20
3. Инжиниринг данных	ПК-1, ПК-2, ОПК-8, ОПК-9	тестирование на платформе дистанционного обучения	20
Итоговая аттестация	ПК-1, ПК-2, ОПК-8, ОПК-9	Итоговое тестирование	40

Примеры заданий

Примеры тестовых вопросов:

1. С помощью каких приложений чаще всего разрабатываются ИТ-продукты?
2. Выберите верное утверждение о взаимосвязи регрессионного- и юнит-тестирования.
3. Какой из двух видов тестирования (юнит и регрессионного) выполняется дольше?
4. Какой из перечисленных инструментов не может быть использован для написания документации?

Задания:

Сравнение работы моделей ML и DL для задач NLP.

Вы использовали модели классического машинного обучения (ML) и глубокие языковые модели (DL) — трансформеры для обработки текстовых данных. Проанализируйте сильные и слабые стороны моделей ML и DL.

За это задание можно получить максимум 4 балла, по одному баллу за полный и непротиворечивый ответ:

- о сильных сторонах ML;
- о слабых сторонах ML;
- о сильных сторонах DL;
- о слабых сторонах DL.

```
import ipywidgets as widgets
conclusion = widgets.Textarea(
    value='',
    placeholder="""
Сильные стороны моделей ML
Слабые стороны моделей ML
Сильные стороны моделей DL
Слабые стороны моделей DL
""",
    description='Ваш ответ',
    disabled=False
)
conclusion
```

Ваш ответ

```
[ ] def task5():
    print(conclusion.value)
    return {"conclusion":conclusion.value }
task5()

{'conclusion': ''}
```

Кейсы для самоконтроля:

1. Откройте демо проекта: <https://huggingface.co/spaces/vivien/clip>. Попробуйте найти информацию по базе. Похожи ли будут изображения? Важно — необходимо искать информацию на английском языке.
2. Прочитайте статью с сайта Яндекс Метрики: <https://yandex.ru/support/metrika/general/how-it-works.html>. Выясните, какие дополнительные метрики может предоставить нам Яндекс Метрика для решения проблемы холодного старта.

Задачи:

1. Используйте данные <https://finance.yahoo.com/>.
 - Проверьте на стационарность ряд цены на нефть.
 - Проверьте на стационарность ряд из лог-приростов цены на нефть.
 - Постройте линейную регрессию курса доллара США к рублю к цене на нефть и проверить значимость параметров.
 - Постройте линейную регрессию логарифма курса доллара США к рублю к логарифму цены на нефть и проверить значимость параметров.
 - Исследовать автокорреляцию лог-приростов цены на нефть.
2. Установите локальные клиентские приложения Docker и GitHub. Разместите заготовку кода в GitHub, предварительно создав репозиторий. Создайте локальный Docker-образ проекта и отправить его в DockerHub. Настройте автоматическое развертывание кода с использованием GitHub → DockerHub → Server (SSH).

4. Организационно-педагогические условия реализации программы

4.1. Учебно-методическое обеспечение и информационное обеспечение программы

4.1.1. Список литературы

Основная литература:

1. Нейронные сети: основы теории / А.И. Галушкин. - М.: Гор. линия-Телеком, 2012. - 496 с.
2. Гифт Н. Прагматичный ИИ. Машинное обучение и облачные технологии : науч. изд./ Н. Гифт ; пер. с англ. И. Пальти. - СПб : Питер, 2019, -300 с.
3. Евсеенко С. М. Этапы развития технологий искусственного интеллекта и уточнение терминологии / С. М. Евсеенко // Инновации. - 2021. - № 4. - С. 39-48.
4. А.И. Панов. Введение в методы машинного обучения с подкреплением. Учебное пособие. МОСКВА, МФТИ, 2019.
5. Орлов, А.И. Организационно-экономическое моделирование .Ч.3 Статистические методы анализов данных. / А.И. Орлов. - М.: МГТУ , 2012. - 623 с.
6. Боровиков, В.П. Популярное введение в современный анализ данных в системе Statistica: Учебное пособие / В.П. Боровиков. - М.: ГЛТ, 2013. - 288 с.
7. Бородин Е.А., Даценко Н.В., Никитин Б.Е., Мачтаков С.Г., Хромых Е.А. Проектирование баз данных. Учебное пособие для подготовки обучающихся по направлениям 09.03.02 - «Информационные системы и технологии», 09.03.03 «Прикладная информатика» / Воронеж, 2023.
8. Григорьев Ю.А., Плутенко А.Д., Бурдаков А.В., Ермаков О.Ю. Сравнение стратегий выборки данных для приближенной обработки запросов к большой базе данных // Информационные технологии. 2022. Т. 28. № 5. С. 240-249.

Дополнительная литература:

1. Дэвид, М. Х. Цифровая схемотехника и архитектура компьютера / М. Х. Дэвид, Л. Х. Сара. — Москва : ДМК Пресс, 2017. — 792 с.
2. Козлов, А. Статистический анализ данных в MS Excel: Учебное пособие / А. Козлов. - М.: ИНФРА-М, 2012. - 320 с.
3. Маккинли, У. Python и анализ данных / У. Маккинли. - М.: ДМК, 2015. - 482 с.
4. Тюрин, Ю.Н. Анализ данных на компьютере: Учебное пособие / Ю.Н. Тюрин, А.А. Макаров; Науч. ред. В.Э. Фигурнов. - М.: ИД ФОРУМ, 2013. - 368 с.
5. Аль Мусави О.А.Р., Кравец О.Я. Исследование алгоритмов повторной оптимизации запросов в облачных базах данных // Решение. 2022. Т. 1. С. 168-171.
6. В. Baum, Р. Petrie. Statistical inference for probabilistic functions of finite state Markov chains, Ann. Math. Statist. (ISSN: 0003-4851) 37 (1966), p. 1554–1563.
7. J.D. Hamilton. A new approach to the economic analysis of nonstationary time series and the business cycle, Econometrica, 57 (1989), p. 357–384.
8. Huisman R., Mahieu R. Regime Jumps in Electricity Prices (August 2001). ERIM Report Series Reference No. ERS-2001-48-FA. Available at SSRN: <https://ssrn.com/abstract=271910>.
9. Поиск взаимосвязей на примере Нефть-Рубль. Доступна на <https://habr.com/post/253285/>.
10. Полбин А.В., Шумилов А.В., Бедин А.Ф., Куликов А.В. Модель реального обменного курса рубля с марковскими переключениями режимов

// Прикладная эконометрика. 2019. Т. 55. No 3. С. 32-50.
<https://doi.org/10.24411/1993-7601-2019-10009>

4.1.2. Ресурсы информационно-телекоммуникационной сети "Интернет"

1. Обучение с подкреплением, второе издание: Введение (серия «Адаптивные вычисления и машинное обучение»)

<https://www.amazon.com/Reinforcement-Learning-Introduction-Adaptive-Computation/dp/0262039249>

2. Reinforcement Learning Lecture 1a: May 2, 2018

<https://cs.uwaterloo.ca/~ppoupart/teaching/cs885-spring18/slides/cs885-lecture1a.pdf>

4.2. Описание материально-технической базы, необходимой для осуществления образовательного процесса по программе

Таблица 8

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
Система дистанционного обучения	Лекция	Персональные компьютеры/ноутбуки, доступ в Интернет, проектор, программное обеспечение Microsoft Visual Studio 2010 или выше
Система дистанционного обучения	Практическая работа	Персональные компьютеры/ноутбуки, доступ в Интернет, принтеры, проектор, программное обеспечение Microsoft Visual Studio 2010 или выше
-	Самостоятельная работа	Персональные компьютеры/ноутбуки, доступ в Интернет
Система дистанционного обучения	Итоговая аттестация	Персональные компьютеры/ноутбуки, доступ в Интернет, принтеры, проектор, программное обеспечение Microsoft Visual Studio 2010 или выше Членам аттестационной комиссии необходимо оборудование для проведения дистанционных семинаров (вебинаров), качественный отказоустойчивый доступ в сеть интернет.

5. Организация образовательного процесса

В таблице 9 описаны образовательные технологии.

Таблица 9

№ п/п	Вид занятия	Форма проведения занятий	Цель
1	Лекция	Ознакомление с теоретическими основами анализа данных	актуализация и систематизация

			теоретических знаний по дисциплине
2	Практическая работа	Выполнение заданий	осознание связей между теорией и практикой, повышение степени понимания материала
3	Самостоятельная работа	Самостоятельное изучение дополнительных материалов и литературы.	получение дополнительных теоретических знаний
4	Промежуточная (итоговая) аттестация	Выполнение заданий.	контроль освоения программы

6. Составители программы

Вахранев Антон Юрьевич, Finrule Finance, Москва, преподаватель, ведущий разработчик.

Согласовано,
Эксперт ОСОП

 Ж. И. Зубцова

Согласовано
Руководитель проектов (Центр
дополнительного, дополнительного
профессионального и онлайн образования
“Пуск”)

 И. В. Тихмянова