

Министерство науки и высшего образования Российской Федерации
федеральное государственное автономное образовательное учреждение высшего
образования

«Московский физико-технический институт
(национальный исследовательский университет)»

УТВЕРЖДАЮ

Ректор МФТИ

д-р физ.-мат. наук, профессор,

член-корреспондент РАН



И.И. Кудрявцев

«*И.И.*» 2020 г.

**Дополнительная профессиональная программа
повышения квалификации
«Математика для анализа данных»**

Москва 2020

1. Общая характеристика программы

1.1 Цель реализации программы

Целью реализации дополнительной профессиональной программы повышения квалификации (далее - Программа) «Математика для анализа данных» является формирование и совершенствование базовых компетенций в математике для Анализа Данных.

Совершенствуемые компетенции

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

Таблица 1

№	Компетенция в соответствии с направлением подготовки	«Прикладная математика и информатика (уровень бакалавриата)» код: 01.03.02
1.	Способность демонстрации общенаучных базовых знаний естественных наук, математики и информатики, понимание основных фактов, концепций, принципов теорий, связанных с прикладной математикой и информатикой	ПК-1
2.	Способность понимать и применять в исследовательской и прикладной деятельности современный математический аппарат	ПК-2
3.	Способность критически переосмысливать накопленный опыт, изменять при необходимости вид и характер своей профессиональной деятельности	ПК-3
4.	Способность применять в профессиональной деятельности современные языки программирования, библиотеки и пакеты программ, а также уверенно владеть методами машинного обучения	ПК-4

Таблица 2

№	Универсальные компетенций (УК), общекультурные компетенций (ОК) и общепрофессиональные компетенций	«Прикладная математика и информатика (уровень бакалавриата)»
---	--	--

		код: 01.03.02
1.	Способность к критическому анализу и оценке современных научных достижений, генерированию новых идей при решении исследовательских и практических задач, в том числе в междисциплинарных областях	УК-1
2.	Способность к абстрактному мышлению, анализу, синтезу, способность совершенствовать и развивать свой интеллектуальный и общекультурный уровень	ОК-2
3.	Готовность действовать в нестандартных ситуациях, нести социальную и этическую ответственность за принятые решения	ОК-3
4.	Способность использовать базовые знания естественных наук, математики и информатики, основные факты, концепции, принципы теорий, связанных с прикладной математикой и информатикой в контексте анализа данных и смежных задач	ОПК-1
5	Способность приобретать новые научные и профессиональные знания, используя современные образовательные и информационные технологии	ОПК-2

1.2. Планируемые результаты обучения

Планируемые результаты обучения представлены в таблице 3.

Таблица 3

№	Знать	Коды компетенций
1.	Принцип работы алгоритмов классификации	ПК-1 , ОПК-2
2.	Принцип работы алгоритмов регрессии	
3.	Принцип работы методов отбора признаков (одномерные, жадные, на основе моделей)	ПК-1, ОПК-2
4.	Постановка задачи обучения по прецедентам	ПК-1, ОПК-2
5.	Критерии выбора модели для построения прогноза	ПК-1, ОПК-2
6.	Язык Python и библиотеки для анализа данных (NumPy, SciPy, Pandas, Matplotlib, Sklearn)	ПК-1, ОПК-2

8.	Процедуру вычисления значения метрик качества моделей	ПК-1, ОПК-2
9.	Методы работы с естественным языком	ПК-1, ОПК-2
1.	Анализ временных рядов методами машинного обучения	ПК-1, ОПК-2
№	Уметь	Коды компетенций
1.	Решать задачу классификации данных	ПК-2, ОК-2
2.	Решать задачу регрессии на данных	ПК-2, ОК-2
3.	Выбрать алгоритм классификации или регрессии, оптимальный для конкретной задачи	ПК-2, ПК-3, УК-1
4.	Перечислить и охарактеризовать метрики качества	ПК-2, ОК-2, ОПК-1
5.	Применить средства sklearn для предобработки и процессинга объектов	ПК-2, ПК-3, ОПК-1
6.	Описать решение задачи классификации	ПК-2, ОК-2
7.	Описать решение задачи регрессии	ПК-2, ОК-2
8.	Применить на практике классификацию	ПК-2, ПК-3, УК-1, ОПК-1
9.	Применить на практике регрессию	ПК-2, ПК-3, УК-1, ОПК-1
10.	Проанализировать проблемы, возникающие при решении задачи классификации, и предложить пути их решения	ПК-3, УК-1
11.	Проанализировать проблемы, возникающие при решении задачи регрессии, и предложить пути их решения	ПК-3, УК-1
12.	Использовать STL-разложение временного ряда	ПК-1, ПК-2, ОК-2

1.3 Категории обучающихся: уровень образования – ВО, область профессиональной деятельности – специалисты, интересующиеся математическими методами в современном анализе данных .

1.4 Форма обучения – заочному принципу (с применением электронного обучения и дистанционных образовательных технологий) в формате вебинаров.

1.5 Объем программы – 128 ак.ч:

Режим обучения: заочно – два раза в неделю по 4 академических часа на протяжении 8-ми недель, на самостоятельную подготовку отводится 64 часа , контрольное мероприятие – защита проекта.

2. Содержание программы

2.1. Учебный (тематический) план

Таблица 4

№ п/ п	Наименование модулей	Всего, час.	в том числе:					Форма контроля
			лекции	практические занятия (семинары)	самостоятель ная работа	контроль ные задания		
1	Дискретная математика.	16	4	4	8	0	тестирование	
2	Математический анализ	32,25	8	8	16	0,25	тестирование	
3	Линейная алгебра и аналитическая геометрия	31,25	8	8	15	0,25	тестирование	
4	Теория вероятностей	28,25	7	7	14	0,25	тестирование	
5	Математическая статистика и элементы аналитики	20,25	4	6	10	0,25	тестирование	
	Итого:	128	31	33	63	1		

2.2. Учебная программа

Таблица 5

№ п/п	Содержание обучения, наименование и тематика практических занятий (вебинаров), самостоятельных работы	Объем, ак.час.
1	Введение в курс. Дискретная математика	16
	Лекции: 1) Основы теории множеств 2) Основы логики 3) Основы комбинаторики 4) Основы комбинаторики 2, комбинаторная вероятность	4

	<p>Практические занятия:</p> <p>1)Задачи на простейшие операции с множествами, отображения</p> <p>2)Преобразования логических формул и формул с кванторами на модельных задачах</p> <p>3)Подсчёт мощностей простых множеств на примере "жизненных" задач</p> <p>4)Задачи на шары и перегородки и комбинаторную вероятность</p>	4
	Самостоятельная работа.	8
	Выполнение контрольных заданий.	0
2	Математический анализ	32,25
	<p>Лекции:</p> <p>1)Отображения и функции</p> <p>2)Пределы функции</p> <p>3)Производная функции одной переменной</p> <p>4)Критические точки и градиентная оптимизация</p> <p>5)Функции нескольких переменных. Градиент и оптимизация</p> <p>6)Интегралы 1. Введение в вычисление интегралов</p> <p>7)Интегралы 2. Методы интегрирования</p>	8
	<p>Практические занятия:</p> <p>1)Упражнения на базовые свойства функций</p> <p>2)Задачи на вычисление пределов</p> <p>3)Задачи на вычисление пределов функций</p> <p>4)Вычисление производной</p> <p>5)Задачи на геометрические свойства производной, освоение техники градиентного спуска</p> <p>6)Задачи на оптимизацию многомерных функций, оптимизация функций потерь</p> <p>7)Простейшие примеры вычисления интегралов. Практика интегрирования, в том числе и в геометрических задачах</p>	8
	Самостоятельная работа.	16
	Выполнение контрольных заданий.	0,25
3	Линейная алгебра и аналитическая геометрия	31,25
	<p>Лекции:</p> <p>1)Векторы и векторная алгебра</p> <p>2)Векторы и векторная алгебра 2</p> <p>3)Системы линейных уравнений</p> <p>4)Векторные пространства и подпространства</p> <p>5)Матричное исчисление 1</p> <p>6)Матричное исчисление 2</p> <p>7)Линейные отображения. Собственные значения и векторы</p> <p>8)Матричные разложения</p>	8
	<p>Практические занятия:</p> <p>1)Операции над векторами, вычисление длины вектора, проверка, что</p>	8

	<p>векторы образуют базис, нахождение точки пересечения двух прямых</p> <p>2)Решение задач с помощью скалярного произведения и метода координат, вычисление расстояния от точки до прямой и плоскости</p> <p>3)Решение конкретных систем линейных уравнений</p> <p>4)Решение задач на связь размерности подпространств и СЛУ, им соответствующих. Поиск базисов подпространств</p> <p>5)Перемножение матриц в природе, вычисление ранга</p> <p>6)Вычисление определителя. Решение на практике переопределённой системы</p> <p>7)Задачи на свойства линейных отображений. Нахождение собственных значений и векторов. Интуиция за этими объектами</p>	
	Самостоятельная работа. Изучение литературы.	15
	Выполнение контрольных заданий.	0,25
4.	Теория вероятностей	28,25
	<p>Лекции:</p> <p>1)Основы теории вероятностей</p> <p>2)Формула Байеса</p> <p>3)Дискретные случайные величины и их характеристики</p> <p>4)Абсолютно непрерывные случайные величины</p> <p>5)Независимость случайных величин</p> <p>6)Математическое ожидание абсолютно непрерывных случайных величин</p> <p>7)Законы больших чисел и предельные теоремы. Нормальное распределение</p>	7
	<p>Практические занятия:</p> <p>1)Решение разнообразных задач по классической и геометрической вероятности</p> <p>2)Решение задач на условную вероятность</p> <p>3)Поиск математических ожиданий и дисперсий дискретных случайных величин</p> <p>4)Применение формулы свёртки</p> <p>5)Вычисление математических ожиданий в абсолютно непрерывном случае</p> <p>6)Применения ЦПТ на практике</p>	7
	Самостоятельная работа. Изучение литературы	14
	Выполнение контрольных заданий.	0,25
5.	Математическая статистика и элементы аналитики	20,25

Лекции: 1)Задача математической статистики и точечные оценки 2)Доверительные интервалы 3)Проверка гипотез и статистические критерии 4)Корреляционный анализ 5)Линейная регрессия	4
Практическое занятие. 1)Построение оценок по методу моментов и методу максимального правдоподобия 2)Построение доверительных интервалов в теории и на практике 3)Построение статистических критериев, измерение качества критериев. 4)Применение p-value для отбора признаков 5)Вычисление корреляций. Связь с метриками машинного обучения 6)Построение оценок в линейной регрессионной модели	6
Самостоятельная работа	10
Выполнение контрольных заданий	0,25
Итого	128

3. Формы аттестации и оценочные материалы

Оценка качества освоения программы проводится по двухбалльной системе: «зачтено», «не зачтено» по результатам промежуточного контроля (тестирование, проверочные задания на взаимную оценку), контроля посещаемости практических занятий (вебинаров) и результатам итоговой аттестации.

Слушатель считается аттестованным в случае положительных результатов работы (не менее 70% баллов от итоговой оценки) в процессе обучения и успешной сдачи экзамена. При этом баллы за экзамены начисляются только при достижении 50% порога при прохождении каждого экзаменационного испытания. После аттестации слушатель получает оценку “отлично”, если набрано не меньше 80% баллов от возможного максимума, “хорошо”, если набрано не меньше 65%, “удовлетворительно”, если набрано не меньше 40%, в противном случае слушатель курс не сдает.

Результат тестирования, решения проверочных заданий и написания кода проверяется автоматически системой на образовательной платформе. Экзаменационная работа проверяется преподавателем.

Составляющие процесса обучения, которые оцениваются в ходе обучения, и их вклад в итоговую оценку представлены в таблице 6.

Таблица 6 – Составляющие процесса обучения

	Основные показатели оценки	Вклад в итоговую оценку
1	Основной курс обучения на образовательной платформе	50%

2	Практические занятия	10%
3	Итоговая аттестация	40%

Оценочные материалы

Пример тестового задания

1. Даны две независимые выборки объема 11 и 14, извлеченные из нормальных совокупностей X, Y . Известны также исправленные дисперсии, равные соответственно 0,75 и 0,4. Необходимо проверить нулевую гипотезу о равенстве генеральных дисперсий при уровне значимости $\gamma=0,05$. Конкурирующую гипотезу выбрать по желанию.
2. Применяя метод замены переменной, найти неопределенный интеграл $\int \frac{1}{x} \cdot -x dx \times 6 \cos 3 \cdot 2 \sin$.
3. Найти неопределенный интеграл $e^x dx \times \sin 4 \cdot 12 \cdot \int$ методом интегрирования по частям.
4. Вычислить площадь фигуры, ограниченной линиями, заданными уравнениями в полярных координатах $r = 6 \cdot 2 \cos \phi, r = 2 \cdot 6 \sin \phi$.

4. Организационно-педагогические условия реализации программы

4.1 Учебно-методическое обеспечение и информационное обеспечение программы

1. Hastie T., Tibshirani R., Friedman J. The Elements of Statistical Learning. Springer, 2009.
2. Bishop C.M. Pattern Recognition and Machine Learning. Springer, 2006.
3. Murphy K.P. Machine Learning: A Probabilistic Perspective. The MIT Press, 2012.
4. Boyd S., Vandenberghe L. Convex Optimization. Cambridge University Press, 2009.
5. Sheldon A. Linear Algebra Done Right. Springer, 2015.
8. DasGupta A. Probability for statistics and machine learning. Springer, 2011.
10. Беклемишев Д.В. Курс аналитической геометрии и линейной алгебры - 10-е изд. - М. Наука, 2003.
11. Иванов Г.Е. Лекции по математическому анализу. Ч.1 - М.:МФТИ, 2004, 2011.
12. Тер - Крикоров А.М., Шабунин М.И. Курс математического анализа - М.Ж: МФТИ, 1997, М.: Физматлит, 2003-2007.
14. Умнов. Аналитическая геометрия и линейная алгебра (2011) — МФТИ.
15. Деммель. Вычислительная линейная алгебра. Теория и приложения (2001)
16. Ильин, Позняк, Основы математического анализа (2005) — МГУ.
17. Тер-Крикоров, Шабунин. Курс математического анализа (2001) — МФТИ

4.2 Материально-технические условия реализации программы

Таблица 8

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
---	-------------	---

Система дистанционного обучения провайдера массовых открытых онлайн курсов	Лекции	Слушателю необходимо наличие доступа в сеть интернет, компьютер. Преподавателю курса необходимо наличие доступа администратора курса.
Информационно-коммуникационная платформа дистанционных семинаров	Практические занятия (дистанционные семинары)	Слушателю необходимо наличие доступа в сеть интернет, компьютер. Преподавателю курса необходимо оборудование для проведения дистанционных семинаров (вебинаров), качественный отказоустойчивый доступ в сеть интернет.
Система дистанционного обучения провайдера массовых открытых онлайн курсов	Самостоятельная работа	Наличие компьютера и доступа в сеть интернет.
Система дистанционного обучения провайдера массовых открытых онлайн курсов	Рубежный контроль, Итоговая аттестация	Наличие компьютера и доступа в сеть интернет.

5. Организация образовательного процесса

Слушатели получают доступ к электронным учебным материалам посредством ресурсов поддержки электронного обучения ЦИОТ МФТИ и партнерских образовательных площадок. Форматы представления электронных учебных материалов: в виде массовых онлайн курсов (МООС) в системе дистанционного обучения провайдера массовых открытых онлайн курсов.

Преподаватель проводит практические занятия дистанционно в форме вебинаров с использованием платформы ZOOM (или аналогичной).

Итоговая аттестация проводится в форме защиты индивидуального задания.

Самостоятельная работа выполняется слушателем в удобном для него режиме.

В таблице 9 описаны образовательные технологии.

Таблица 9

№ п/п	Вид занятия	Форма проведения занятий	Цель
1	Лекция	Самостоятельный просмотр видеолекций	Ознакомление слушателей с базовым материалом по тематике курса

2	Практические занятия	Выполнение практических заданий, получение обратной связи от преподавателя. Обсуждение вопросов, возникших в результате просмотра видеолекций и изучения литературы.	Практическое освоение теоретических знаний, а также углубление знаний по курсу
3	Самостоятельная работа	Самостоятельное изучение дополнительных материалов и литературы. Выполнение тренировочных тестов и заданий.	Углубление знаний по курсу.
4	Выполнение контрольных заданий	Выполнение тестов, проверочных заданий.	Практическое освоение теоретических знаний, контроль освоения материалов.
5	Итоговая аттестация	Защита индивидуальной работы.	Практическое освоение теоретических знаний, контроль освоения материалов.

6. Составители программы:

Райгородский Андрей Михайлович

Доктор физико-математических наук, профессор, директор ФПМИ МФТИ

Волков Никита Васильевич

руководитель проектов лаборатории инноватики ФПМИ

Яровиков Юрий Николаевич

руководитель проектов лаборатории инноватики ФПМИ

Согласовано

Согласовано

Зам. директора ЦДПО

Директор ФПМИ, д.ф.-м.н., профессор

У.Б. Вещезерова




А.М. Райгородский