

МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
ФЕДЕРАЛЬНОЕ ГОСУДАРСТВЕННОЕ АВТОНОМНОЕ
ОБРАЗОВАТЕЛЬНОЕ УЧРЕЖДЕНИЕ ВЫСШЕГО ОБРАЗОВАНИЯ
«НАЦИОНАЛЬНЫЙ ИССЛЕДОВАТЕЛЬСКИЙ ЯДЕРНЫЙ УНИВЕРСИТЕТ «МИФИ»
(НИЯУ МИФИ)

«УТВЕРЖДАЮ»
Первый проректор НИЯУ МИФИ
Нагорнов О.В.
и.ю.л. 2020 г.



**ПРОГРАММА ПОВЫШЕНИЯ
КВАЛИФИКАЦИИ**

«Практическое применение методов машинного обучения в data-driven
подходе»

(наименование программы)

1. Цель реализации программы

Цель: получение новых компетенций, необходимых для профессиональной деятельности:

- Извлечение сырых данных, написание запросов к разнородным СУБД
- Обработка данных — очистка, обогащение, проверка качества
- Построение моделей машинного обучения (прогноз, классификация, кластеризация и т.п.)
- Внедрение построенных моделей в виде интеллектуального компонента для микросервисов, выполняющих бизнес-задачи
- Выявление закономерностей в данных и описание их в виде аналитических отчетов
- Проверка прикладных гипотез на данных
- Презентация результатов работы бизнесу

2. Требования к результатам обучения

В результате освоения программы слушатель должен приобрести следующие знания и умения, необходимые для получения компетенции, указанной в п.1:

слушатель должен знать:

- Основные принципы извлечения данных из структурированных и слабоструктурированных файлов, а также реляционных баз данных
- Основные методы подготовки данных: чистку, трансформацию, восстановление пропущенных значений, обработку аномалий
- Основные методы визуальной аналитики
- Основные методы машинного обучения, предназначенные для решения задачи прогнозирования
- Основные методы машинного обучения, предназначенные для решения задачи классификации без учителя
- Основные методы машинного обучения, предназначенные для решения задачи классификации с учителем

слушатель должен уметь:

- создавать новые data-driven проекты
- работать с источниками данных
- строить модели машинного обучения
- оценивать качество моделей машинного обучения
- строить визуальные представления данных

3. Содержание программы

Учебный план

программы повышения квалификации
«Практическое применение методов машинного обучения в data-driven подходе»

Категория слушателей - наличие высшего образования, область профессиональной деятельности. любая требующая аналитического подхода

Срок обучения — 16 час.

Форма обучения — с отрывом от работы

п/п	Наименование разделов	Всего, час.	В том числе	
			лекции	практич. и лаборат. занятия
1	Основы машинного обучения	4		
2	Изучение и подготовка данных	4	2	2
3	Создание и оценка качества моделей машинного обучения	6	2	4
4	Основы визуальной аналитики	2		2
Итоговая аттестация		Зачет		

Учебно-тематический план

программы повышения квалификации

«Практическое применение методов машинного обучения в data-driven подходе»

№ п/п	Наименование разделов и тем	Всего час.	В том числе	
			лекции	практич. и лаборат. занятия
	2	3	4	5
1	Основы машинного обучения			
1.1	Основные определения и термины, подходы в машинном обучении. Роль машинного обучения в современных прикладных задачах. Основные этапы работы аналитика с методами машинного обучения Data-driven подход.	2	2	

1.2	Современные аналитические платформы. Основы работы с инструментальными средствами машинного обучения. Создание нового проекта, работа с элементами меню, понятие интеллектуального рабочего процесса -«пайплайна» обработки данных	2	1	1
2	Изучение и подготовка данных			
2.1	Основные задачи извлечения данных, Работа со структурированными файлами, подключение к реляционным БД, написание “..-запросов.			1
2.2	Обогащение данных. Очистка, восстановление пропущенных значений, обработка аномалий	2	1	1
2.3	Понятие качества данных и методы его проверки	1	1	
3	Создание и оценка качества моделей машинного обучения			
3.1	Методы классификации с учителем и без учителя, методы прогнозирования, Понятие обучающей выборки. Обучение и использование моделей.	4	2	2
3.2	Основные метрики качества моделей машинного обучения. Матрица запутанности (ошибок). Точность, полнота, MSE, MAE и др. метрики. Оценка качества моделей.	2		2
4	Основы визуальной аналитики			
4.1	Работа с простыми визуализациями. Столбиковые диаграммы, круговые диаграммы, диаграммы рассеяния, тепловые карты, «ящик- с усами».	2		2

Учебная программа**

повышения квалификации

«Практическое применение методов машинного обучения в data-driven подходе»

Раздел 1. Основы машинного обучения (4 часа)

Тема 1.1 Основные определения и термины, подходы в машинном обучении. Роль машинного обучения в современных прикладных задачах. Основные этапы работы аналитика с методами машинного обучения Data-driven подход (2 часа)

Тема 1.2 Основы работы с инструментальными средствами машинного обучения. Создание нового проекта, работа с элементами меню, понятие интеллектуального рабочего процесса - «пайплайна» обработки данных (2 час.)

Раздел 2. Изучение и подготовка данных (4 часа)

Тема 2.1 Основные задачи извлечения данных. Работа со структурированными файлами, подключение к реляционным БД, написание SQL-запросов. Извлечение данных из текстовых файлов. (1 час.)

Тема 2.2 Обогащение данных. Очистка, восстановление пропущенных значений, обработка аномалий. Создание новых переменных. Кодирование и биннинг. (2 час.)

Тема 2.3 Понятие качества данных и методы его проверки. Техническое и бизнес качество данных. (1 час.)

Раздел 3. Создание и оценка качества моделей машинного обучения (6 час.)

Тема 3.1 Методы классификации с учителем и без учителя, методы прогнозирования. Понятие обучающей выборки. Обучение и использование моделей. Метод k-средних, иерархическая агломерация. Деревья решений. Регрессионные модели. (4 час.)

Тема 3.2 Основные метрики качества моделей машинного обучения. Матрица запутанности (ошибок). Точность, полнота, MSE, MAE и др. метрики. Оценка качества моделей. (2 час.)

Раздел 4. Основы визуальной аналитики (2 час.)

Тема 4.1 Работа с простыми визуализациями. Столбиковые диаграммы, круговые диаграммы, диаграммы рассеяния, тепловые карты, «ящик- с усами». (2 час)

Перечень лабораторных работ

Номер темы	Наименование лабораторной работы
1	«Работа с инструментальными средствами» (1 час.)
2	«Основы подготовки данных. Работа с источниками данных» (1 час.)
2	«Основы подготовки данных. Обогащение данных» (1 час.)
3	«Основы применения методов машинного обучения» (4 час.)
3	«Оценка качества методов машинного обучения» (2 час.)
4	«Визуальная аналитика» (2 час)

4. Материально-технические условия реализации программы

Наименование специализированных аудиторий, кабинетов, лабораторий	Вид занятий	Наименование оборудования, программного обеспечения
1	2	
Аудитория	лекции	компьютер, мультимедийный проектор, экран, доска

Лаборатория Компьютерный класс	лабораторные работы	КММЕ, Ofбсе, SQLite
-----------------------------------	------------------------	---------------------

5. Учебно-методическое обеспечение программы

1. Бринк Х., Ричардс Дж, Феверолф М. Машинное обучение. — СПб.: Питер, 2017. — 336 с.: ил. — (Серия «Библиотека программиста»). ISBN 978-5-496-02989-6
2. Hastie T., Tibshirani R, Friedman J. The Elements of Statistical Learning (2nd edition). Springer, 2009.
3. Bishop C. M. Pattern Recognition and Machine Learning. Springer, 2006.
4. Mohri M., Rostamizadeh A., Talwalkar A. Foundations of Machine Learning. MTT Press, 2012.
5. Mohammed .J. Zaki, Wagner Meira Jr. Data Mining and Analysis. Fundamental Concepts and Algorithms. CambridgeUniversityPress, 2014.

6. Оценка качества освоения программы

Оценка качества освоения программы осуществляется аттестационной комиссией в виде междисциплинарного зачета в электронной форме на основе пятибалльной системы оценок по основным разделам программы. Перечень разделов и вопросов, выносимых на междисциплинарный экзамен, приведен в приложении А. Слушатель считается аттестованным, если имеет положительные оценки (3,4 или 5) по всем разделам программы, выносимым на зачет.

7. Составители программы

Киреев В.С., канд. техн. наук, доцент (раздел 1, темы 1.1- 1.12)