

МИНОБРНАУКИ РОССИИ

Федеральное государственное автономное
образовательное учреждение
высшего образования
«Новосибирский национальный
исследовательский государственный
университет»
(Новосибирский государственный
университет, НГУ)

П Р И К А З

16 ЯНВ 2023 № 0043 - 3

[о повышении квалификации по программе «Краткий курс по
основам машинного обучения»]

В целях реализации программы развития Передовой инженерной школы

ПРИКАЗЫВАЮ:

1. Утвердить программу повышения квалификации «Краткий курс по основам машинного обучения» согласно Приложению № 1.
2. Ответственность за проведение программы возложить на директора НОЦ ГПН НГУ, директора-организатора ПИШ НГУ С.В. Головина.
3. Отделу ТДиК довести до проректора по учебной работе, директора НОЦГПННГУ, декана ММФ, ИППК.

Проректор по учебной работе

С.Г. Саблина

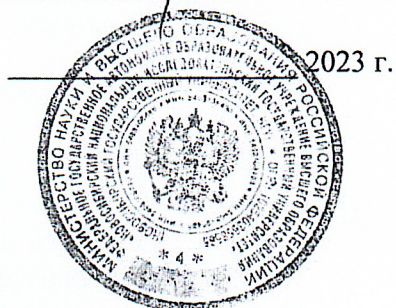
Директор НОЦ ГПН НГУ

С.В. Головин

МИНОБРНАУКИ РОССИИ
федеральное государственное автономное образовательное учреждение высшего образования
«Новосибирский национальный исследовательский государственный университет»

УТВЕРЖДАЮ
Проректор по учебной работе

 С.Г. Саблина



ПРОГРАММА

повышения квалификации

по программе «Краткий курс по основам машинного обучения»

общая трудоемкость программы **40 часов**

форма обучения **очная**

образовательные технологии **лекционное обучение, практические занятия, самостоятельная работа**

Разработали:
зав. лаб. ЛПЦТ ММЦ НГУ, д.ф.-м.н., Р.И. Мулладжанов
зав. лаб. ЛАПДиМО ММФ НГУ, к.ф.-м.н., Е.Н. Павловский,
м.н.с. ЛАПДиМО ММФ НГУ, Б.Н. Тучинов
инженер ЛАПДиМО ММФ НГУ, В. Суворов

Проверил:
Заместитель декана по развитию ММФ
к.ф.-м.н.
А.В. Карпенко

Новосибирск, 2023

1. **Цель реализации программы:** получение слушателями представления о задачах машинного обучения (далее — МО) и подходах к их решению. Успешное освоение курса позволит обучающимся выполнять такие профессиональные задачи, как:
- постановка задачи анализа данных,
 - разработка требований к проекту анализа данных,
 - обосновать применение методов машинного обучения в бизнес-задачах.

Для достижения поставленной цели выделяются следующие *задачи программы*:

- доступно донести до слушателя основные теоретические положения изучаемых тем и способы их представления;
- закрепить получаемые теоретические знания на задачах и примерах с подробным разбором их решений;
- продемонстрировать практическую значимость получаемых знаний.

Курс проводится в интерактивной форме с элементами проектной работы. Проекты выбираются самими слушателями на основе их собственных интересов и интересов компании, что позволяет достигать профессиональных целей за пределами курса.

Актуальность программы определяется ее ориентацией на практические кейсы заказчика.

Программа соответствует целям Перспективной инженерной школы НГУ в части проведения профильного обучения и распространения знаний по прикладному инжинирингу.

2. **Категория слушателей:** сотрудники компаний.
3. **Планируемые результаты обучения:** в результате освоения программы слушатели должны обладать следующими профессиональными компетенциями:
- ПК-1. Понимать возможности технологий машинного обучения при проведении бизнес-анализа.
 - ПК-2. Способность подготовить проект в сфере машинного обучения.
4. Срок обучения: 40 часов.
5. Форма обучения: очная.
6. Режим занятий: 5 дней по 8 часов с перерывами. Расписание и учебно-тематический план представлены ниже.

Расписание

1 день	09:00 – 10:50	Проекты в сфере МО, какие задачи возможно решать (Модуль 1)
	11:00 – 12:50	Аналитический цикл CRISP-DM, примеры на реальных проектах (Модуль 1)
	14:00 – 15:50	Python для МО (Модуль 3)
	16:00 – 17:50	Практика исследования данных (Модуль 3)
2 день	09:00 – 10:50	1. Классы задач МО (классификация, регрессия, кластеризация) 2. Bias-variance. Проблема переобучения. Разбиение на train set и val set 3. Линейная регрессия (+ полиномиальная). 4. SGD 5. Логистическая регрессия (Модуль 2)
	11:00 – 12:50	Практика: линейная и логистическая регрессия (sklearn) (Модуль 2)
	14:00 – 15:50	1. Деревья решений (класс., регр.) 2. Random Forest (Модуль 2)
	16:00 – 17:50	Практика: деревья решений (Модуль 2)
3 день	09:00 – 10:50	1. AdaBoost 2. Gradient boosting (Модуль 2)
	11:00 – 12:50	Практика: бустинг (Модуль 2)
	14:00 – 15:50	1. Инжиниринг признаков 2. Критерии успеха проекта (Модуль 1)
	16:00 – 17:50	Проектная практика: формулировка идей проектов, поиск критериев успеха, определение целей анализа данных, определение критерия анализа данных (Модуль 1)
4 день	09:00 – 10:50	1. Модель нейрона 2. Многослойный перцептрон 3. Расширения (RNN, CNN, etc.) и приложения (NLP, CV, etc.) 4. Adam и AdamW (Модуль 2)
	11:00 –	Практика: PyTorch (Модуль 2)

	12:50	
	14:00 – 15:50	1. Задача кластеризации 2. K-means 3. Иерархическая кластеризация 4. Правило локтя (Модуль 2)
	16:00 – 17:50	Практика: кластеризация (Модуль 2)
5 день	09:00 – 10:50	Решение практического кейса на примере задачи заказчика или типовых задач (например, распознавание дорожных знаков, детектирование людей в кадре) (Модуль 3)
	11:00 – 12:50	Python для МО: визуализация результатов обучения (Модуль 3)
	14:00 – 15:50	Самостоятельная работа над проектом (Модуль 1)
	16:00 – 17:50	Защита постановки задачи на проект

Учебно-тематический план

Наименование разделов/модулей (тем, дисциплин)	Всего, час.	В том числе:			Форма аттестации
		Лекционное обучение	Практические занятия	Самостоятельная работа	
Модуль 1. Бизнес-анализ и управление проектами машинного обучения	10	4	2	4	
Модуль 2. Машинное обучение	20	10	10		
Модуль 3. Язык программирования Python	8	4	4		
Итоговая аттестация	2		2		Защита проекта
Итого	40	18	18	4	

7. **Условия реализации программы повышения квалификации:** Обучение проводится на территории компьютерного класса центра «Газпромнефть НГУ», ул. Пирогова 2., оборудованной средствами визуализации учебного материала (проекторы, флипчарты, маркерные доски) и необходимой вычислительной техникой.
8. **Правила аттестации** (с указанием критериев выставления оценок, в т.ч. итоговой аттестации): Освоение программы слушателями проводится в форме защиты проекта. Приобретение компетенций ПК-1 – ПК-2 слушатель подтверждает в ходе презентации и ответов на вопросы комиссии, состоящей из преподавателей курса и представителей ПИШ НГУ. Способности каждого слушателя оцениваются в соответствии с таблицей. Суммарно необходимо набрать не менее 3 баллов для аттестации.

Таблица 1. Оценка результатов обучения

Результаты обучения	Ознакомлен, 1 балл	Разбирается, 2 балла	Освоил 3 балла
Понимать возможности технологий машинного обучения при проведении бизнес-анализа.	Называет несколько конкретных случаев, где технологии машинного обучения принесли пользу бизнесу.	Объясняет, каким образом польза была получена.	Имеет целостное понимание о областях применимости той или иной технологии машинного обучения.
Способность подготовить проект в сфере машинного обучения	Может сформулировать бизнес-проблему	Идентифицирует все заинтересованные стороны и указывает их требования к результату проекта	Адекватно оценивает риски в сфере машинного обучения.

9. Оценочные материалы

9.1. Перечень вопросов по программе повышения квалификации:

Темы для дискуссий по бизнес-кейсам:

1. Как способствовать тому, чтобы компании активнее обменивались данными?
2. Какие данные компании могут принести дополнительную выгоду?

Задания для самостоятельной работы:

1. Поставить задачу; провести анализ применимости изученных методов и инструментов к выбранным проектам.
2. Сгенерировать гипотезы по своим проектам, получить данные.

10. Литература

(а) основная литература

1. Thomas Frisendal. Design Thinking Business Analysis / Business Concept Mapping Applied // Springer-Verlag Berlin Heidelberg 2012. URL: <https://doi.org/10.1007/978-3-642-32844-2>.
2. C. B. B. D. Manyika, Big Data: The Next Frontier for Innovation, Competition, and Productivity [Electronic resource] // McKinsey Global Institute, 2011. – URL: http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx.
3. CRISP-DM 1.0 / Step-by-step data mining guide. Pete Charman and others. [Электронный ресурс] URL: <https://www.the-modeling-agency.com/crisp-dm.pdf>
4. Deep Learning, Ian Goodfellow and Yoshua Bengio and Aaron Courville, MIT Press, Goodfellow-et-al-2016
Доступна по адресу <http://www.deeplearningbook.org/>
5. Introduction to machine learning, Nils J. Nilsson, Robotics Laboratory, Department of Computer Science Stanford University 1998
Доступна по адресу <http://robotics.stanford.edu/~nilsson/MLBOOK.pdf>
6. Swamynathan M. Mastering Machine Learning with Python in Six Steps: A Practical Implementation Guide to Predictive Data Analytics Using Python. – Apress, 2017. URL: <https://doi.org/10.1007/978-1-4842-2866-1>.
7. Kwangjo Kim, Muhamad Erza Aminanto, Harry Chandra Tanuwidjaja. Deep learning https://link.springer.com/chapter/10.1007/978-981-13-1444-5_4
8. Santanu Pattanayak. Pro Deep Learning with TensorFlow. <https://link.springer.com/book/10.1007/978-1-4842-3096-1>
9. Taweh Beysolow II. Applied Natural Language Processing with Python. <https://link.springer.com/book/10.1007/978-1-4842-3733-5>
10. Redmond E., Wilson J. R. Seven Databases in Seven Weeks. A Guide to Modern Databases and the NoSQL Movement. ISBN-13: 978-1934356920, ISBN-10: 1934356921, 2012, Release P2.0 2013, 354 p.
11. Wu L., Yuan L., You J. Survey of Large-Scale Data Management Systems for Big Data Applications. JOURNAL OF COMPUTER SCIENCE AND TECHNOLOGY 30(1): 163–183 Jan. 2015. DOI 10.1007/s11390-015-1511-8
12. Kroenke D. M. Database Processing. ISBN 0-13-064839-6, Prentice Hall PTR, 8-th ed. 2002.
13. Banker K., Bakkum P., Verch S., Garrett D., Hawkins T. MongoDB in Action, Second Edition - 2016
14. Tiwari S. Professional NoSQL. John Wiley & Sons, Inc. Indianapolis, ISBN: 978-0-470-94224-6, USA. 2011
15. Wallace Jackson - JSON Quick Syntax Reference (The Expert's Voice in Web Development), 2016
16. Gaurav Vaish. Getting Started with NoSQL. ISBN 978-1-84969-4-988, Packt Publishing Ltd. Birmingham, UK. 2013

(б) дополнительная литература

17. Тельнов Ю.В. Реинжиниринг бизнес-процессов (Учебное пособие). / Московский международный институт эконометрики, информатики, финансов и права. - М., 2003. – 99с.

18. Майкл Хаммер, Джеймс Чампи. Реинжиниринг корпорации. Манифест революции в бизнесе. «Манн, Иванов и Фербер» — М., 2006. — 287 с. (Библиотека НГУ: книгохранение – 1997 г. изд.: №1800161969, читальный зал — 2006 г. изд.: №1800330250)
19. Шматалюк А, Ферапонтов М., Громов А., Каменнова М. Моделирование бизнеса. Методология ARIS. — М.: Весть-Метатехнология, 2001. — С. 8–14.
20. Understanding Machine Learning, © 2014 by Shai Shalev-Shwartz and Shai Ben-David
Published 2014 by Cambridge University Press
PDF of manuscript posted by permission of Cambridge University Press.
<http://www.cs.huji.ac.il/~shais/UnderstandingMachineLearning/copy.html>
21. Trevor Hastie, Elements of statistical learning, Springer [Electronic resource], 2009. – URL: <http://statweb.stanford.edu/~tibs/ElemStatLearn/>.
22. Крутиков, В.Н. Анализ данных : учебное пособие / В.Н. Крутиков, В.В. Мешечкин ; Министерство образования и науки Российской Федерации, Федеральное государственное бюджетное образовательное учреждение высшего профессионального образования «Кемеровский государственный университет». - Кемерово : Кемеровский государственный университет, 2014. - 138 с. : ил. - Библиогр. в кн. - ISBN 978-5-8353-1770-7 ; То же [Электронный ресурс]. - URL: <http://biblioclub.ru/index.php?page=book&id=278426>.
23. С. Aggarwal. Neural Networks and Deep Learning. <https://link.springer.com/book/10.1007/978-3-319-94463-0>
24. Slav Petrov. Coarse-to-Fine Natural Language Processing. <https://link.springer.com/book/10.1007/978-3-642-22743-1>
25. Chris Biemann, Alexander Mehler. Text Mining. <https://link.springer.com/book/10.1007/978-3-319-12655-5>

(в) Интернет-ресурсы

Таблица 10.1

№ п/п	Наименование Интернет-ресурса	Краткое описание
1	Innovations for Requirement Analysis. From Stakeholders' Needs to Formal Designs [Electronic resource]. – URL: https://link.springer.com/book/10.1007/978-3-540-89778-1	Полнотекстовые электронные копии статей в области анализа требований (2007 год).
2	Thomas Frisendal. Design Thinking Business Analysis [Electronic resource] // Springer-Verlag Berlin Heidelberg 2012. URL: https://doi.org/10.1007/978-3-642-32844-2	Электронная копия книги Томаса Фрисендала (2012).
3	Петров, А.В. Моделирование процессов и систем [Электронный ресурс] : учебное пособие / А.В. Петров. — Электрон. дан. — Санкт-Петербург : Лань, 2015. — 288 с. — Режим доступа: https://e.lanbook.com/book/68472 . — Загл. с экрана.	Моделирование процессов и систем (2015).
4	Coursera.org Machine Learning by Stanford University https://www.coursera.org/learn/machine-learning/	

5	Fast.AI online courses https://course.fast.ai/	
6	Воронцов К.В. Математические методы обучения по прецедентам [Электронный ресурс]. – Режим доступа: http://www.machinelearning.ru/wiki/images/6/6d/Voron-ML-1.pdf . – Загл. с экрана	Известный курс Воронцова К.В. (2016). На русском языке.
7	C. B. B. D. Manyika, Big Data: The Next Frontier for Innovation, Competition, and Productivity [Electronic resource] // McKinsey Global Institute, 2011. – URL: http://www.mckinsey.com/~media/McKinsey/dotcom/Insights%20and%20pubs/MGI/Research/Technology%20and%20Innovation/Big%20Data/MGI_big_data_full_report.ashx	Отчёт МакКинзи о Больших Данных (2011).
8	Berikov V.B., Litvinenko A.G. Methods for statistical data analysis with decision trees. URL: http://www.math.nsc.ru/AP/datamine/eng/decisiontree.htm .	Описание алгоритмов построения деревьев решений
9	The Keras Blog. https://blog.keras.io	