

**Пермский филиал федерального государственного автономного  
образовательного учреждения высшего образования  
"Национальный исследовательский университет  
"Высшая школа экономики"**

**Программа учебной дисциплины  
«Введение в машинное обучение»**

Утверждена

Академическим советом<sup>1</sup>

Академическим советом основных образовательных программ по направлениям  
подготовки 38.03.05 Бизнес-информатика, 09.03.04 Программная инженерия, 38.04.05  
Бизнес-информатика

Протокол № от \_\_\_\_ 20\_\_

Академический руководитель образовательной программы «Бизнес-информатика»

\_\_\_\_\_ Л.В. Шестакова

Подпись

Разработчик	Бузмакова Алексей Владимирович, доцент кафедры информационных технологий в бизнесе
Число кредитов	6
Контактная работа (час.)	58
Самостоятельная работа (час.)	170
Образовательная программа, курс	Образовательная программа «Бизнес-информатика», направление подготовки 38.03.05 Бизнес-информатика, уровень бакалавриат, 3 курс
Формат изучения дисциплины	С использованием онлайн курсов

<sup>1</sup> Для ПУД из общеуниверситетского пула – Руководитель Департамента / заведующий кафедрой.

## 1. Цель, результаты освоения дисциплины и пререквизиты

Целями освоения дисциплины «Введение в машинное обучение» являются:

- приобретение навыков работы на языке Python, знание и понимание задач управления данными, в том числе, загрузка данных, преобразование данных, и предварительный анализ и визуализация данных;
- знакомство с основными задачами и моделями машинного обучения, знание методов оценки качества работы различных моделей машинного обучения;
- понимание процесса интеграции моделей машинного обучения в рамках задач стоящих перед потенциальными заказчиками;
- повышение мотивации студентов на дальнейшее углубление своих знаний в области работы с данными и машинного обучения.

Настоящая дисциплина относится к блоку дисциплин по выбору профессионального цикла.

Формат изучения дисциплины: с использованием онлайн-курсов площадки DataCamp.

В результате освоения дисциплины «Введение в машинное обучение» студенты формируют следующие компетенции: УК-3, УК-5, УК-10, ПК-3, ПК-4, ПК-31

Код	Формулировка компетенций
УК-3	Способен решать проблемы в профессиональной деятельности на основе анализа и синтеза
УК-5	Способен работать с информацией: находить, оценивать и использовать информацию из различных источников, необходимую для решения научных и профессиональных задач (в том числе на основе системного подхода)
УК-10	Способен осуществлять производственную или прикладную деятельность в международной среде
ПК-3	способен работать, используя основные методы, способы и средства получения, хранения, переработки информации
ПК-4	способен работать с информацией в глобальных компьютерных сетях
ПК-31	способен обрабатывать, анализировать и систематизировать информацию по теме исследования, используя соответствующий математический аппарат и инструментальные средства

В результате освоения учебной дисциплины, студенты должны владеть следующими знаниями, умениями и навыками:

- Знает:
  - Основные подходы к преобразованию данных;
  - Основные задачи машинного обучения;
  - Основные модели машинного обучения;
  - Основные стадии ведения проекта машинного обучения;
- Умеет:
  - работать с массивами данных
  - распознавать задачи машинного обучения в реальных бизнес-задачах
  - формализовать бизнес задачу как задачу машинного обучения
- Имеет навыки:

- загрузки, преобразования, очистки и визуализации данных на языке Python.
- обучения и применения моделей машинного обучения на языке Python
- оценки качества и интерпретации полученных результатов

## 2. Содержание учебной дисциплины

Темы, объем часов и планируемые результаты обучения представлены в таблице.

Разделы / темы дисциплины	Объем в часах				Планируемые результаты обучения (ПРО), подлежащие контролю	Формы контроля
	лк	см	ср	onl		
Раздел 1. Введение в анализ данных и язык Python	8	18	62	20	Знает основные подходы к преобразованию данных; умеет работать с массивами данных; имеет навыки загрузки, преобразования, очистки и визуализации данных на языке Python	Прохождение курсов в DataCamp, сдача лабораторных работ, выполнение проекта, экзамен
Раздел 2. Фундаментальные основы машинного обучения	8	10	40	8	Знает основные задачи машинного обучения; знает основные модели машинного обучения; умеет формализовать бизнес задачу как задачу машинного обучения; имеет навыки обучения и применения моделей машинного обучения на языке Python	Прохождение курсов в DataCamp, сдача лабораторных работ, выполнение проекта, экзамен
Раздел 3. Прикладной опыт ведения проектов машинного обучения	8	6	40		Знает основные стадии ведения проекта машинного обучения; умеет распознавать задачи машинного обучения в реальных бизнес-задачах; имеет навыки оценки качества и интерпретации полученных результатов	выполнение проекта, экзамен
<b>Часов по видам учебных занятий:</b>	<b>24</b>	<b>34</b>	<b>142</b>	<b>28</b>		

Формы учебных занятий:

лк – лекции в аудитории;

см - семинары/ практические занятия/ лабораторные работы в аудитории;

onl – лекции или иные виды работы студента с помощью онлайн-курса;

ср – самостоятельная работа студента.

### Содержание разделов дисциплины:

#### Раздел 1. Введение в анализ данных и язык Python

Тема 1. Введение в анализ данных

Тема 2. Введение в язык Python

Тема 3. Основные операции работы с данными

Тема 4. Визуализация данных

#### Раздел 2. Фундаментальные основы машинного обучения

Тема 5. Статистика и машинное обучение

Тема 6. Задачи классификации и регрессии. Линейные и древесные модели

Тема 7. Бэггинг и бустинг

Тема 8. Задача кластеризации

Раздел 3. Прикладной опыт ведения проектов машинного обучения

Тема 9. Разработка и создание аналитических приложений

Тема 10. Внедрение, сопровождение и мониторинг модели

Тема 11. Экспертиза качества моделей

### 3. **Оценивание**

Текущий контроль по дисциплине «\_\_\_\_\_» включает в себя следующие элементы:

- прохождение курсов в DataCamp
- лабораторная работа 1
- лабораторная работа 2
- лабораторная работа 3
- выполнение проекта

Промежуточная аттестация по дисциплине проводится в форме экзамена. Экзамен проводится в письменном виде в системе LMS.

К блокирующим элементам контроля по дисциплине относятся следующие виды контроля: экзамен.

Пересдача по блокирующим видам контроля проводится в соответствии с Положением об организации промежуточной аттестации и текущем контроле знаний студентов НИУ ВШЭ.

Оценка по дисциплине ( $O_{\text{дисциплине}}$ ) определяется, как взвешенная сумма оценок по всем видам контроля и рассчитывается по следующей формуле:

$$O_{\text{дисциплине}} = 0.1 * O_{\text{DataCamp}} + 0.1 * O_{\text{Лаб1}} + 0.1 * O_{\text{Лаб2}} + 0.1 * O_{\text{Лаб3}} + 0.3 * O_{\text{проект}} + 0.3 * O_{\text{экзамен}}$$

где  $O_{\text{DataCamp}}$  – оценка за прохождение курсов DataCamp;

$O_{\text{Лаб1}}$ ,  $O_{\text{Лаб2}}$ ,  $O_{\text{Лаб3}}$  – оценки за лабораторные работы;

$O_{\text{проект}}$  – оценка за проект;

$O_{\text{экзамен}}$  – оценка за экзамен.

Способ округления – арифметический.

#### *Критерии оценивания*

Текущий контроль представляет из себя проект, представляющий из себя решений реальной (или приближенной к реальной задачи). Проект должен включать в себя

- формулировку решаемой реальной (или приближенной к реальной) проблемы и постановку задачи машинного обучения
- загрузку, очистку и преобразование данных
- обучение нескольких моделей машинного обучения для решения поставленной задачи
- оценка качества работы обученных моделей
- выводы о возможности применение одной из обученных моделей для решения поставленной проблемы
- хорошо откомментированный программный код на языке Python, с выполнением различных этапов проекта, в том числе с использованием функций языка Python, не проходившихся в рамках курса

Итоговый контроль является письменным экзаменом с несколькими случайными вопросами, включающие один вопрос на описание одной из моделей машинного

обучения, формулировка одной из основных концепций машинного обучения, и вопрос на интеграцию проекта по машинному обучению для решения практических задач.

#### 4. Примеры оценочных средств

Прохождение курсов в DataCamp:

- Data Types for Data Science, <https://www.datacamp.com/courses/data-types-for-data-science>
- Introduction to Data Science in Python, <https://www.datacamp.com/courses/introduction-to-data-science-in-python>
- Intermediate Python for Data Science, <https://www.datacamp.com/courses/intermediate-python-for-data-science>
- Supervised Learning with scikit-learn, <https://www.datacamp.com/courses/supervised-learning-with-scikit-learn>
- Unsupervised Learning in Python, <https://www.datacamp.com/courses/unsupervised-learning-in-python>

Лабораторная работа 1.

На основании транзакционных данных клиентов, а также их социально-демографических характеристик обучить линейную модель вероятности визита клиента в следующие полгода.

Лабораторная работа 2.

Реализовать алгоритмы обучения и предсказания в задаче классификации:

- Дерево решений
- Случайный лес
- Градиентный бустинг

Лабораторная работа 3.

Сравнить качество нескольких предсказательных моделей в задаче регрессии.

Проект

Выполняется в командах из двух человек. Тема выбирается на выбор студента. Пример возможной темы: предсказание вероятности мошеннической транзакции в финансовом секторе.

Экзамен

Экзамен представляет из себя тест состоит из двух частей

1. Первая часть теста – множественный выбор ответов. Пример возможных вопросов теста находится по ссылке: <https://yadi.sk/d/Atd3mRi60jpnmQ>
2. Вторая часть тест – выполнение практического задания. Пример данных также находится по ссылке: [https://yadi.sk/d/W4aNAvcw\\_PkHfg](https://yadi.sk/d/W4aNAvcw_PkHfg)

#### 5. Ресурсы

##### 5.1. Рекомендуемая основная литература

№п/п	Наименование
1.	T. Hastie, R. Tibshirani, J. Friedman. <i>The Elements of Statistical Learning (2nd edition)</i> . Springer, 2017. Доступно online: <a href="https://proxylibrary.hse.ru:2184/book/10.1007/978-0-387-84858-7">https://proxylibrary.hse.ru:2184/book/10.1007/978-0-387-84858-7</a> (требуется предварительно войти в <a href="https://library.hse.ru">https://library.hse.ru</a> ), также на сайте автора <a href="https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print12.pdf">https://web.stanford.edu/~hastie/ElemStatLearn/printings/ESLII_print12.pdf</a>
2.	J.P. Mueller. <i>Beginning Programming with Python for Dummies</i> , 2nd Edition, 2018. Доступно online: <a href="https://proxylibrary.hse.ru:2251/toc.aspx?bookid=139144">https://proxylibrary.hse.ru:2251/toc.aspx?bookid=139144</a> (требуется предварительно войти в <a href="https://library.hse.ru">https://library.hse.ru</a> )
3.	L. Massaron and J.P. Mueller. <i>Python for Data Science For Dummies</i> , 2015.

	Доступно online: <a href="https://proxylibrary.hse.ru:2251/toc.aspx?bookid=82524">https://proxylibrary.hse.ru:2251/toc.aspx?bookid=82524</a> (требуется предварительно войти в <a href="https://library.hse.ru">https://library.hse.ru</a> )
--	--

## 5.2. Рекомендуемая дополнительная литература

№п/п	Наименование
4.	Anna Manning, <i>Databases for Small Business: Essentials of Database Management, Data Analysis, and Staff Training for Entrepreneurs and Professionals.</i> – Apress, 2015, 319 pp. Доступно online: <a href="https://proxylibrary.hse.ru:2251/toc.aspx?bookid=104365">https://proxylibrary.hse.ru:2251/toc.aspx?bookid=104365</a> (требуется предварительно войти в <a href="https://library.hse.ru">https://library.hse.ru</a> )

## 5.3. Программное обеспечение

№п/п	Наименование	Условия доступа
1.	MS Office 2010	Из внутренней сети НИУ ВШЭ - Пермь (договор)
		<i>Например, из внутренней сети университета (договор)/ свободное лицензионное соглашение</i>

## 5.4. Профессиональные базы данных, информационные справочные системы, интернет-ресурсы (электронные образовательные ресурсы)

№п/п	Наименование	Условия доступа
1.	Электронно-библиотечные ресурсы	По подписке НИУ ВШЭ
2.		Из внутренней сети университета (договор)

## 5.5. Материально-техническое обеспечение дисциплины

Например, для проведения лекций и семинаров по дисциплине необходимо наличие ноутбука (компьютера) с установленным пакетом Microsoft® PowerPoint, мультимедийного проектора и аудиооборудования. Для выполнения самостоятельной работы необходим компьютер с подключением к сети Интернет.

## 6. Особенности организации обучения для лиц с ограниченными возможностями здоровья и инвалидов

В случае необходимости, обучающимся из числа лиц с ограниченными возможностями здоровья (по заявлению обучающегося), а для инвалидов также в соответствии с индивидуальной программой реабилитации инвалида, могут предлагаться следующие варианты восприятия учебной информации с учетом их индивидуальных психофизических особенностей, в том числе с применением электронного обучения и дистанционных технологий:

6.1.1. для лиц с нарушениями зрения: в печатной форме увеличенным шрифтом; в форме электронного документа; в форме аудиофайла (перевод учебных материалов в аудиоформат); индивидуальные консультации с привлечением тифлосурдопереводчика; индивидуальные задания и консультации.

6.1.2. для лиц с нарушениями слуха: в печатной форме; в форме электронного документа; видеоматериалы с субтитрами; индивидуальные консультации с привлечением сурдопереводчика; индивидуальные задания и консультации.

6.1.3. для лиц с нарушениями опорно-двигательного аппарата: в печатной форме; в форме электронного документа; в форме аудиофайла; индивидуальные задания и консультации.

## 7. Дополнительные сведения

Особенности самостоятельной работы по курсу отражены в Приложении 1.