

**Федеральное государственное автономное образовательное учреждение
высшего образования
"Национальный исследовательский университет
"Высшая школа экономики"**

Факультет экономики, менеджмента и бизнес-информатики
Кафедра информационных технологий в бизнесе

**Рабочая программа дисциплины
Машинное обучение**

для образовательной программы
«Информационная аналитика в управлении предприятием»
направления подготовки 38.04.05 Бизнес-информатика
уровень магистратура

Разработчик программы
Бузмаков А.В. avbuzmakov@hse.ru

Одобрена на заседании кафедры информационных технологий в бизнесе
«20» августа 2018 г.
И.о. зав. кафедрой информационных технологий в бизнесе
Е.Г. Плотникова _____

Утверждена Академическим советом основных образовательных программ по направлениям
подготовки 38.03.05 Бизнес-информатика, 09.03.04 Программная инженерия, 38.04.05 Бизнес-
информатика
«30» августа 2018 г., № протокола 8.2.2.1-38-09/06

Академический руководитель образовательной программы
А.И. Дерябин _____

Пермь, 2018



1 Область применения и нормативные ссылки

Настоящая программа учебной дисциплины устанавливает требования к образовательным результатам и результатам обучения студента и определяет содержание и виды учебных занятий и отчетности.

Программа предназначена для преподавателей, ведущих дисциплину «Машинное обучение», учебных ассистентов и студентов направления подготовки 38.04.05 Бизнес-информатика, обучающихся по образовательной программе «Информационная аналитика в управлении предприятием».

Программа учебной дисциплины разработана в соответствии с:

- Образовательным стандартом НИУ ВШЭ по направлению подготовки 38.04.05 Бизнес-информатика, утвержденным ученым советом НИУ ВШЭ, протокол от 06.12.2013 №50;
- Основной образовательной программой высшего образования «Информационная аналитика в управлении предприятием» направления подготовки 38.04.05 Бизнес-информатика;
- Объединенным учебным планом университета по образовательной программе «Информационная аналитика в управлении предприятием» направления подготовки 38.04.05 Бизнес-информатика, утвержденным в 2017 г.

2 Цели освоения дисциплины

Целями освоения дисциплины «Машинное обучение» являются понимание задач решаемых методами машинного обучения, знакомство с основными методами машинного обучения, приобретение навыков работы с этими методами в существующих программных средах, а также приобретение навыков постановки задач и их решения на реальных данных. Данные цели реализуются в рамках следующих общих целей образовательной программы: основ экономико-математического моделирования, методов анализа данных и процессов; углублённое изучение методов и технологий, инструментальных средств, используемых для создания информационных и аналитических систем поддержки управления; подготовка к выполнению самостоятельных научных исследований в области разработки моделей и методик, инструментальных средств создания проблемно-ориентированных систем, средств анализа бизнес-процессов и автоматизации потоков работ.

3 Компетенции обучающегося, формируемые в результате освоения дисциплины

Уровни формирования компетенций:

РБ — ресурсная база, в основном теоретические и предметные основы (знания, умения);

СД – способы деятельности, составляющие практическое ядро данной компетенции;

МЦ – мотивационно-ценностная составляющая, отражает степень осознания ценности компетенции человеком и готовность ее использовать



В результате освоения дисциплины студент осваивает компетенции:

Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
Способен вести профессиональную, в том числе научноисследовательскую деятельность в международной среде	СК-8	РБ	Способен читать документацию на иностранном языке.	Самостоятельная работа над проектом, изучения материалов в DataCamp на иностранном языке. Изучение статей на иностранном языке	Работа в DataCamp Проект
Способен проводить поиск и анализ инноваций в сфере ИКТ, а также применять их в практической деятельности	ПК-10	РБ	Владеет основными моделями машинного обучения. Понимает сильные и слабые каждой при решении конкретных практических задач.	Использование и сравнение различных моделей и методов машинного обучения при решении практических задач.	Проект
Выявлять и прогнозировать основные направления использования современных ИКТ для управления эффективностью бизнеса	ПК-11	РБ	Способен формулировать задачи на внедрение ИТ-средств на существующих программных предприятиях.	Изучение и анализ существующих методов машинного обучения и применение их при выполнении домашних заданий.	Домашнее задание Проект
Способен проводить научные исследования и готовить аналитические материалы для оценки мероприятий и выработки стратегических решений в сфере ИКТ	ПК-12	РБ	Способен поставить задачу машинного обучения и проверять степень её достижения. Способен подготовить отчёт о проделанной работе	Разбор реальных кейсов применения методов машинного обучения. Решения реальных кейсов на лабораторных работах. Обратная связь по этапам проекта.	Проект
Способен применять методы системного анализа и моделирования с целью оценки, проектирования и разработки стратегии	ПК-13	РБ	Способен поставить задачу машинного обучения по задачам бизнеса и проверять степень её достижения	Разбор реальных кейсов применения методов машинного обучения. Решения реальных кейсов на лабораторных работах. Обратная связь по этапам проекта.	Проект
Способен разрабатывать и применять экономико-математические модели для обоснования проектных решений в сфере ИКТ	ПК-14	РБ	Демонстрирует навыки решения практических задач методами машинного обучения.	Изучение, анализ и выбор моделей и методов машинного обучения на лекциях, лабораторных работах и при выполнении домашнего задания и проекта.	Домашнее задание Проект
Способен проводить исследования в области новых моделей и методов, направленных на совершенствование архитектуры предприятия, разработку и внедрение отдельных ее компонентов	ПК-15	РБ	Способен поставить задачу машинного обучения и проверять степень её достижения	Разбор реальных кейсов применения методов машинного обучения. Решения реальных кейсов на лабораторных работах. Обратная связь по этапам проекта.	Проект
Способен организовать самостоятельную и коллективную научно-исследовательскую работу на предприятии и управлять ею	ПК-16	РБ	Способен организовать и провести микроисследование по выбору признаков и моделей машинного обучения при решении реальной бизнес задачи	Разбор реальных кейсов применения методов машинного обучения. Решения реальных кейсов на лабораторных работах. Обратная связь по этапам проекта.	Проект



Компетенция	Код по ОС ВШЭ	Уровень формирования компетенции	Дескрипторы – основные признаки освоения (показатели достижения результата)	Формы и методы обучения, способствующие формированию и развитию компетенции	Форма контроля уровня сформированности компетенции
Способен организовать электронное предприятие, используя новейшие тенденции развития электронной коммерции	ПК-17	РБ	Способен поставить задачу машинного обучения по данным электронного предприятия и проверять степень её достижения	Разбор реальных кейсов применения методов машинного обучения. Решения реальных рейсов на лабораторных работах. Обратная связь по этапам проекта.	Проект
Способен совершенствовать методы ведения электронного бизнеса, работы подразделений электронного бизнеса несетевых компаний	ПК-18	РБ	Способен поставить задачу машинного обучения по данным электронного предприятия и проверять степень её достижения	Разбор реальных кейсов применения методов машинного обучения. Решения реальных рейсов на лабораторных работах. Обратная связь по этапам проекта.	Проект
Способен оценивать эффективность мероприятий относительно целей совершенствования ИТ-инфраструктуры и архитектуры предприятия и бизнес-процессов	ПК-19	РБ	Способен оценить степень достижения задачи машинного обучения	Разбор реальных кейсов применения методов машинного обучения. Решения реальных рейсов на лабораторных работах. Обратная связь по этапам проекта.	Проект
Способен применять методы, инструментарий, программное обеспечение и новейшие разработки в целях информационной безопасности на предприятии организации;	ПК-21	РБ	Знает примеры задач машинного обучения и соответствующий инструментарий в области информационной безопасности	Разбор реальных кейсов применения методов машинного обучения. Решения реальных рейсов на лабораторных работах. Обратная связь по этапам проекта.	Проект
Способен коммерциализировать бизнес-проекты в сфере ИКТ	ПК-22	РБ	Умеет оценить выигрыш от использования методов машинного обучения в бизнес-задачах	Разбор реальных кейсов применения методов машинного обучения. Решения реальных рейсов на лабораторных работах. Обратная связь по этапам проекта.	Проект
Способен внедрять инновации в сфере ИКТ-бизнеса	ПК-23	РБ	Умеет составить аналитическое приложение для решения бизнес-задач	Разбор реальных кейсов применения методов машинного обучения. Решения реальных рейсов на лабораторных работах. Обратная связь по этапам проекта.	Проект
Собирать выбирать оптимальные решения в вопросах совершенствования ИТ-инфраструктуры и архитектуры предприятия, а также его информационной безопасности	ПК-24	РБ	Знает различные модели анализа данных. Умеет выбрать оптимальную для решения поставленной задачи	Лекционный материал, разбор реальных кейсов. Работа на проектом в несколько этапов	Проект

4 Место дисциплины в структуре образовательной программы

Настоящая дисциплина относится к циклу дисциплин программы и блоку дисциплин, обеспечивающих вариативную подготовку.

Изучение данной дисциплины базируется на следующих дисциплинах:

- Технологии анализа данных

Для освоения учебной дисциплины студенты должны владеть следующими знаниями и компетенциями:

- Знание основ функционирования персональных компьютеров.



- Знание основ организации обработки данных с помощью компьютеров.
- Базовые навыки работы с персональным компьютером в среде Microsoft Windows.
- Базовые знания и навыки работы с офисными приложениями (текстовым процессором и электронными таблицами).
- Знание основных понятий математического анализа, теории вероятности и математической статистики.
- Знание основ построения баз данных.
- Базовые навыки работы на языке программирования R.

5 Тематический план учебной дисциплины

№	Название раздела	Всего часов	Аудиторные часы				Самостоятельная работа
			Лекции	Семинары	Практические занятия	Другие виды работы	
1	Введение в машинное обучение	6	2		2		2
2	Задача регрессии	44	4		8		32
3	Задача классификации	44	4		8		32
4	Задача кластеризации	42	4		6		32
5	Задача выявления эффекта от воздействия	24	4		4		16
6	Рекомендационные системы	24	4		4		16
7	Временные ряды	22	2		4		16
8	Поиск паттернов	22	2		4		16
ВСЕГО:		228	26		40		162

6 Формы контроля знаний студентов

Тип контроля	Форма контроля	2 год		Параметры
		1	2	
Текущий	Домашнее задание	*	*	Подготовка и презентация темы по выбору студента. Проходит на протяжении всего курса.
	Проект		*	Постановка задачи (желательно по интересной для студента тематике) и её решение методами машинного обучения. Постановка задачи сдаётся в конце первого модуля, решение задачи – в конце второго модуля. Форма сдачи – презентация и отчёт.
Итоговый	Экзамен		*	Письменная работа

7 Критерии оценки знаний, навыков

Домашнее задание заключается в подготовке студентами доклада по одной из тем курса. Работа может быть выполнена в группе не более из двух человек. Каждый студент должен учувствовать в не менее чем двух докладах. Оценка доклада состоит из

- Полноты доклада: базовая часть (4 балла) и дополнительный материал (2 балла);
- Понятность доклада (4 балла).

Полнота доклада оценивается как по списку затронутых понятий, так и по степени уверенности человека в рассказываемом материале. Понятность доклада оценивается по количе-



ству и качеству вопросов, заданных аудиторией, по её вовлечённости. Оценка уменьшается на 40%, если доклад не состоится в назначенное время. Данная домашняя задание может быть сдано только во время лекций или семинаров.

Проект заключается в решении реальной задачи методами машинного обучения, по возможности представляющей интерес для студентов. Работа может быть выполнена в команде из не более чем двух человек. Оценка состоит из:

- Постановка задачи, в т.ч. введение в тему, актуальность, противоречие и проблема, цель работы (3 балла), описание используемых данных;
- Адекватность выбора метода машинного обучения для решения поставленной задачи, настройки параметров используемого метода (3 балла);
- Качество предлагаемого решения (программы для её решения, описания порядка действий);
- Понятность текста и доклада.

По окончании половины курса, студенты обязаны предоставить постановку задачи на проверку. При отсутствии постановки задачи в срок, снимается 2 балла. Комментарии по постановке могут быть исправлены в окончательном тексте проекта.

Оценки по всем формам текущего контроля выставляются по 10-ти балльной шкале.

Вопросы по домашним заданиям могут задаваться как удалённо (через почту) так и на непосредственно на занятиях и личных встречах.

8 Содержание дисциплины

Раздел 1 Введение в машинное обучение

Тема 1. Введение.

Задача регрессии

Тема 2. Постановка задачи регрессии.

Тема 3. Основные методы решения задачи регрессии

Тема 4. Задача классификации

Тема 5. Постановка задачи классификации.

Тема 6. Основные методы решения задачи классификации

Тема 7. Постановка задачи кластеризации.

Тема 8. Основные методы решения задачи кластеризации

Задача выявления эффекта от воздействия

Тема 9. Постановка задачи выявления эффекта от воздействия.

Тема 10. Основные методы решения задачи выявления эффекта от воздействия

Рекомендационные системы

Тема 11. Задачи, решаемые рекомендационными системами и методы их решения.

Временные ряды

Тема 12. Предсказание временных рядов. Основные подходы.

Поиск паттернов

Тема 13. Задача поиска интересных паттернов. Некоторые виды паттернов и методы их выявления

Формы и методы проведения занятий по разделу, применяемые учебные технологии:

Чтение лекций, проведение практических занятий.

9 Образовательные технологии

Используется «проблемное» чтение лекций по дисциплине с использованием компьютерного мультимедийного оборудования, предусматривающее разбор практических задач, моделирование типовых ситуаций, возникающих при решении задач анализа данных. На



практике используются инструментальные средства (системы программирования, аналитические платформы, офисный пакет), позволяющие получить базовые навыки работы.

Текущий и итоговый контроль предусматривает организацию «защит» выполненных заданий с оформлением и публичным представлением результатов. В качестве основы для оформления документов рекомендуются стандарты оформления программной документации (ЕСПД), отчетов о НИР и НИОКР. Для обмена материалами по дисциплине, проведения экспресс-опросов (тестирования), сдачи контрольных работ используются возможности LMS.

9.1 Методические рекомендации преподавателю

На лекциях рекомендуется использовать мультимедийные презентации, иллюстрирующие материал. При этом студенты должны активно участвовать в обсуждении вопросов. Многие из тем курса тесно связаны с уже пройденными студентами курсами: «Базы данных», «Имитационное моделирование», «Эконометрика», «Хранилища данных». Рассмотрение подобных тем следует начинать с повторения студентами того, как рассматриваемые вопросы освещались в рамках пройденных курсов.

9.2 Методические указания студентам

Студентам рекомендуется следующая схема работы над практическим заданием:

- 1) проработать конспект лекций;
- 2) проанализировать рекомендованную литературу;
- 3) проанализировать варианты решений, предложенные преподавателем;
- 4) при затруднениях сформулировать вопросы к преподавателю.

10 Оценочные средства для текущего контроля и аттестации студента

10.1 Оценочные средства для оценки качества освоения дисциплины в ходе текущего контроля

- Что относят к методам машинного обучения?
- Процесс выявления знаний в базах данных (KDD)
- Постановка задачи классификации. Метрики оценки качества работы соответствующих методов.
- Постановка задачи регрессии. Метрики оценки качества работы соответствующих методов.
- Постановка задачи кластеризации. Метрики оценки качества работы соответствующих методов.
- Постановка задачи определения эффекта от воздействия. Метрики оценки качества работы соответствующих методов.
- Постановка задачи, решаемой рекомендационными системами. Метрики оценки качества работы соответствующих методов.
- Постановка задачи предсказания временных рядов. Метрики оценки качества работы соответствующих методов.
- Примеры постановки задачи поиска интересных паттернов. Что называется интересным паттерном?
- Методы:
 - C4.5
 - CART
 - Лес решений



- Наивный Байес
- Градиентный бустинг
- Регрессии (линейная, логистическая)
- K-means
- Иерархический кластеринг
- Casual Trees
- Сведение задачи оценки эффекта от воздействия к задаче регрессии.
- Подходы к построению рекомендационных систем.
- Модели ARIMA.
- Поиск ассоциативных правил.

10.2 Примеры заданий промежуточной аттестации

Пример домашней работы №1: Подготовьте дома и доложите постановку задачи классификации, метрики, используемые для оценки качества работы методов классификации, а также расскажите как работает метод C4.5 для построения дерева решений.

Пример домашней работы №2: реферат на тему Предсказание стоимости квартиры на основании существующих объявлений на Avito.

11 Порядок формирования оценок по дисциплине

Преподаватель оценивает работу студентов на практических занятиях: оценивается активность студента, в т.ч. вопросы, которые он задаёт. Оценки за работу на семинарских и практических занятиях преподаватель выставляет в рабочую ведомость. Оценка по 10-ти балльной шкале за работу на практических занятиях определяется перед итоговым контролем - $O_{\text{контактная}}$.

Преподаватель оценивает самостоятельную работу студентов: оценивается полнота и своевременность выполнения лабораторных работ. Оценки за самостоятельную работу студента преподаватель выставляет в рабочую ведомость. Оценка по 10-ти балльной шкале за самостоятельную работу определяется перед завершающим контролем - $O_{\text{сам.}}$.

Накопленная оценка по дисциплине рассчитывается по формуле:

$$O_{\text{накопленная}} = 0.6 * O_{\text{текущий}} + 0.2 * O_{\text{контактная}} + 0.2 * O_{\text{сам. работа}}$$

где $O_{\text{текущий}}$ рассчитывается как взвешенная сумма домашних заданий предусмотренных в ОУП

$$O_{\text{текущий}} = 0.3 * O_{\text{дз1}} + 0.7 * O_{\text{дз2}}.$$

В диплом выставляется результирующая оценка по учебной дисциплине.

$$O_{\text{результ}} = 0.7 * O_{\text{накопл}} + 0.3 * O_{\text{экз}}$$

Способ округления накопленной оценки завершающего контроля: арифметический.

Примеры расчёта оценки приведены в Приложении 1.

12 Учебно-методическое и информационное обеспечение дисциплины

12.1 Основная литература

1. Компьютерные технологии анализа данных в эконометрике / Д.М. Дайитбегов. - М.: ИНФРА-М: Вузовский учебник, 2008. - 578 с.: 70x100 1/16. - (Научная книга). (переплет) ISBN 978-5-16-003380-8 - Режим доступа: <http://znanium.com/catalog/product/143137>



2. Основы алгоритмизации и программирования на Python : учеб. пособие / С.Р. Гуриков. — М. : ФОРУМ : ИНФРА-М, 2018. — 343 с. — (Высшее образование: Бакалавриат). - Режим доступа: <http://znanium.com/catalog/product/924699>

12.2 Дополнительная литература

3. Статистический анализ данных в MS Excel: Учебное пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. - М.: ИНФРА-М, 2012. - 320 с.: 60х90 1/16. - (Высшее образование). (переплет) ISBN 978-5-16-004579-5 - Режим доступа: <http://znanium.com/catalog/product/238654>

4. Экономическая информатика: Введение в экономический анализ информационных систем [Электронный ресурс] : Учебник. - М.: Инфра-М, 2005. - 965 с.: - (Учебники экономического факультета МГУ им. М.В. Ломоносова). - ISBN 5-16-002009-8. - Режим доступа: <http://znanium.com/catalog/product/534301>

12.3 Программные средства

Для успешного освоения дисциплины, студент использует следующие программные средства:

- язык R
- среда разработки Rstudio

12.4 Дистанционная поддержка дисциплины

Часть лекций и упражнений берётся с платформы DataCamp:

1. Introduction to Machine Learning, DataCamp, <https://www.datacamp.com/courses/introduction-to-machine-learning-with-r>
2. Supervised Learning in R: Regression, DataCamp, <https://www.datacamp.com/courses/supervised-learning-in-r-regression>
3. Supervised Learning in R: Classification, DataCamp, <https://www.datacamp.com/courses/supervised-learning-in-r-classification>
4. Unsupervised Learning in R, DataCamp, <https://www.datacamp.com/courses/unsupervised-learning-in-r>

13 Материально-техническое обеспечение дисциплины

Для проведения лекционных занятий используется компьютер с установленным программным обеспечением для демонстрации презентаций и проектор.

Практические занятия проводятся в компьютерных классах с установленными R, RStudio, Word, Excel, PowerPoint, и с возможностью выхода в интернет