

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

**Программа учебной дисциплины
«Машинное обучение»**

Утверждена

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 __ _____ 2019

Академический руководитель ОП

_____ Дерябин А.И.

Подпись

ФИО

Разработчик	Бузмаков Алексей Владимирович, к.т.н., доцент кафедры информационных технологий в бизнесе, avbuzmakov@hse.ru
Число кредитов	7
Контактная работа (час.)	66
Самостоятельная работа (час.)	186
Образовательная программа, курс	Информационная аналитика в управлении предприятием, направление подготовки 38.04.05 Бизнес-информатика, уровень магистратура, 1 курс
Формат изучения дисциплины	без использования онлайн курса

1. Цель, результаты освоения дисциплины и пререквизиты

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.

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.

Настоящая дисциплина относится к Вариативной части. Обязательные дисциплины специализаций.

Формат изучения дисциплины: без использования онлайн курса.

В результате освоения дисциплины «Машинное обучение» студенты формируют следующие компетенции:

Код	Формулировка компетенций
УК-8	Способен вести профессиональную, в том числе научно- исследовательскую деятельность в международной среде
ПК-10	Способен проводить исследования и поиск новых моделей и методов совершенствования архитектуры предприятия
ПК-11	Способен проводить аналитические и поисковые исследования в сфере экономики, управления и ИКТ для выявления продуктовых, технологических, организационных, маркетинговых инноваций.
ПК-12	Способен проводить научные исследования для выработки стратегических решений в области ИКТ
ПК-13	Способен согласовывать с заказчиком, планировать и выполнять самостоятельную и коллективную научно-исследовательскую работу
ПК-14	Способен готовить демонстрационные материалы, проводить консультации и разрабатывать рекомендации для потенциальных заказчиков по вопросам разработки и совершенствования архитектуры предприятия
ПК-15	Способен проводить консультации и разрабатывать рекомендации для потенциальных заказчиков по вопросам развития ИТ-инфраструктуры предприятия
ПК-16	Способен отбирать новшества в сфере ИКТ, формировать и обосновывать предложения по созданию продуктов и услуг на их основе, разрабатывать стратегию и

	планы реализации инноваций на их основе
ПК-17	Способен совершенствовать, развивать и преобразовывать архитектуру предприятия на основе внедрения инноваций в сфере ИКТ
ПК-18	Способен разрабатывать образовательные программы и учебно-методические материалы по управленческим и ИТ-дисциплинам
ПК-19	Способен проводить лекционные и практические занятия по управленческим и ИТ-дисциплинам

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

знать: методы машинного обучения;

уметь: осуществлять постановку задач;

иметь навыки: практического применения методов машинного обучения при решении задач в конкретных экономических условиях.

2. Содержание учебной дисциплины

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

Разделы дисциплины	Объем в часах				Планируемые результаты обучения (ПРО), подлежащие контролю	Формы контроля
	лк	см	сп	onl		
Введение в машинное обучение	2	2	4	0	Знает методы машинного обучения	Домашнее задание Проект Самостоятельная работа
Задача регрессии	4	8	40	0	Осуществляет постановку и решение задачи регрессии	Домашнее задание Проект Самостоятельная работа
Задача классификации	4	8	40	0	Осуществляет постановку и решение задачи классификации	Домашнее задание Проект Самостоятельная работа
Задача кластеризации	4	6	36	0	Осуществляет постановку и решение задачи кластеризации	Домашнее задание Проект Самостоятельная работа
Задача выявления эффекта от воздействия	4	4	18	0	Осуществляет постановку и решение задачи выявления эффекта от воздействия	Домашнее задание Проект Самостоятельная работа
Рекомендационные системы	4	4	18	0	Знает задачи, решаемые рекомендационными системами и методы их решения	Домашнее задание Проект Самостоятельная работа
Временные ряды	2	4	16	0	Знает задачи, решаемые временными рядами и методы их решения	Домашнее задание Проект Самостоятельная работа
Поиск паттернов	2	4	14	0	Знает задачи поиска паттернов и методы их решения	Домашнее задание Проект Самостоятельная работа
				0		Письменный экзамен
Часов по видам учебных занятий:	26	40	186	0		

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;

сп – самостоятельная работа студента.

Содержание тем дисциплины:

Раздел 1 Введение в машинное обучение

Тема 1. Введение.

Раздел 2. Задача регрессии

Тема 2. Постановка задачи регрессии.

Тема 3. Основные методы решения задачи регрессии

Раздел 3. Задача классификации

Тема 4. Постановка задачи классификации.

Тема 5. Основные методы решения задачи классификации

Раздел 4. Задача кластеризации

Тема 6. Постановка задачи кластеризации.

Тема 7. Основные методы решения задачи кластеризации

Раздел 5. Задача выявления эффекта от воздействия

Тема 8. Постановка задачи выявления эффекта от воздействия.

Тема 9. Основные методы решения задачи выявления эффекта от воздействия

Раздел 6. Рекомендационные системы

Тема 10. Задачи, решаемые рекомендационными системами и методы их решения.

Раздел 7. Временные ряды

Тема 12. Предсказание временных рядов. Основные подходы.

Раздел 8. Поиск паттернов

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

3. Оценивание

Текущий контроль по дисциплине «Машинное обучение» включает в себя следующие элементы: домашнее задание, проект, самостоятельная работа, выполняемые на протяжении всего курса.

Промежуточная аттестация по дисциплине проводится в форме экзамена. Экзамен проводится в письменном виде. Блокирующие элементы отсутствуют.

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

$$O_{\text{дисциплине}} = 0,2 * O_{\text{дом.зад.}} + 0,2 * O_{\text{проект}} + 0,2 * O_{\text{сам.раб.}} + 0,4 * O_{\text{экзамен}}$$

где O_i – оценка за определенный элемент контроля.

Способ округления – арифметический.

Критерии оценивания

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 количеству и качеству вопросов, заданных аудиторией, по её вовлеченности. 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 окончанию 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 непосредственно на занятиях и личных встречах.

Особенности выполнения самостоятельной работы представлены в Приложении 1.

4. Примеры оценочных средств

- 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.

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.

5. Ресурсы

5.1. Рекомендуемая основная литература

№п/п	Наименование
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

5.2. Рекомендуемая дополнительная литература

№п/п	Наименование
1.	Статистический анализ данных в MS Excel: Учебное пособие / А.Ю. Козлов, В.С. Мхитарян, В.Ф. Шишов. - М.: ИНФРА-М, 2012. - 320 с.: 60x90 1/16. - (Высшее образование). (переплет) ISBN 978-5-16-004579-5 - Режим доступа: http://znanium.com/catalog/product/238654
2.	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

5.3. Программное обеспечение

№п/п	Наименование	Условия доступа
1.	Adobe Acrobat Reader	Свободное лицензионное соглашение
2.	Google Chrome Enterprise	Свободное лицензионное соглашение
3.	Windows 10	Договор

5.4. 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 установленным программным обеспечением.

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.