МИНИСТЕРСТВО НАУКИ И ВЫСШЕГО ОБРАЗОВАНИЯ РОССИЙСКОЙ ФЕДЕРАЦИИ
СТАРООСКОЛЬСКИЙ ТЕХНОЛОГИЧЕСКИЙ ИНСТИТУТ ИМ. А.А. УГАРОВА
(филиал) федерального государственного автономного образовательного учреждения высшего образования
«Национальный исследовательский технологический университет «МИСИС»
СТИ НИТУ «МИСИС»

Рабочая программа утверждена
решением Ученого совета
СТИ НИТУ «МИСИС»
от «20» июня 2023 г.
протокол № 5

Рабочая программа дисциплины

Машинное обучение

Закреплена за кафедрой	**Кафедра автоматизированных и информационных систем управления**
Направление подготовки	09.03.02 Информационные системы и технологии
Профиль	Информационные системы и технологии
Квалификация	**Бакалавр**
Форма обучения	**Очная**
Общая трудоемкость	**3 ЗЕТ**

Часов по учебному плану	108	Формы контроля в семестрах:
в том числе:		зачет 5
аудиторные занятия	34	
самостоятельная работа	74	
часов на контроль	0	

Распределение часов дисциплины по семестрам

Семестр	5		Итого	
Вид занятий	УП	РП	УП	РП
Лекции				
Практические	34	34	34	34
Контактная работа	34	34	34	34
Сам. работа	74	74	74	74
Итого	108	108	108	108

Год набора 2023 г.
В редакции 2023г

Программу составил:
доцент каф. АИСУ, кандидат технических наук
Соловьев Антон Юрьевич

Должность,уч.ст., уч.зв ФИО полностью *подпись*

Рабочая программа дисциплины

Машинное обучение

наименование

Разработана в соответствии с ОС ВО НИТУ «МИСиС»:
Самостоятельно устанавливаемый образовательный стандарт высшего образования – бакалавриат Федеральное государственное автономное образовательное учреждение высшего образования «Национальный исследовательский технологический университет «МИСиС» по направлению подготовки 09.03.02 Информационные системы и технологии (приказ от 05.03.2020 г. № 95 о.в.)

Составлена на основании учебного плана 2020 года набора.
09.03.02 Информационные системы и технологии,
Профиль: Информационные системы и технологии, утвержденного Ученым советом
СТИ НИТУ «МИСиС» 22.06.2020 г., протокол № 23.

Рабочая программа одобрена на заседании кафедры
Автоматизированные и информационные системы управления

Протокол от «24» мая 2023 г. № 05

И.о. зав. кафедрой АИСУ *подпись* Д.А. Полещенко

«24» мая 2023 г.

Руководитель ОПОП ВО
И.о. зав. кафедрой АИСУ,
кандидат технических наук, доцент *подпись* Д.А. Полещенко

«24» мая 2023 г.

1. ЦЕЛИ ОСВОЕНИЯ
Целью дисциплины является формирование системных знаний о машинном обучении с учителем. О том в как и областях может применяться машинное обучение, какие задачи могут решаться и т. д. Знакомство с основными алгоритмами машинного обучения, особенностями их применения. Задачи дисциплины: • Научить современным методам классификации • Научить современным методам анализа текста • Научить метрикам алгоритмов машинного обучения Научить современным библиотекам машинного обучения

2. МЕСТО ДИСЦИПЛИНЫ В СТРУКТУРЕ ОПОП ВО	
Часть ОПОП ВО (Часть, формируемая участниками образовательных отношений)	Б1.В
2.1	**Требования к предварительной подготовке обучающихся – предшествующие дисциплины**
2.1.1	Программирование на языках высокого уровня
2.1.2	Объектно-ориентированное программирование
2.1.3	Организация эксперимента
2.1.4	Информатика
2.1.5	Учебная практика по получению первичных профессиональных умений
2.2	**Дисциплины и практики, для которых освоение данной дисциплины необходимо как предшествующее:**
2.2.1	Нейросетевые системы

3. ИНДИКАТОРЫ ФОРМИРОВАНИЯ КОМПЕТЕНЦИЙ, СОВМЕЩЕННЫЕ С РЕЗУЛЬТАТАМИ ОБУЧЕНИЯ ПО ДИСЦИПЛИНЕ	
УК-3 способен: **- проектировать и разрабатывать продукцию, процессы и системы, соответствующие профилю образовательной программы;** **- выбирать и применять соответствующие методики проектирования и разработки, включая передовые методы и технологии.**	
Знать:	УК-3-З1 - возможности актуальных алгоритмов машинного обучения, которые широко используются на практике, их преимущества и недостатки
	применять методы машинного обучения при решении задач в различных прикладных областях
Уметь:	УК-3-У1 - Строить и проверять качество моделей машинного обучения с учителем; УК-3-У2 - *строить решающие алгоритмы на многомерных данных и преодолевать вычислительные проблемы, связанные с высокой размерностью данных*
Владеть:	УК-3-В1 - навыками интерпретации полученных результатов в терминах прикладной области с целью получения новых знаний и выводов
ПК-1 Способен участвовать в работах по: **- доводке и освоению информационных технологий в ходе внедрения и эксплуатации информационных технологий;** **- поддержанию работоспособности информационных систем и технологий в заданных функциональных характеристиках и в соответствии с критериями качества;** **- управлению программно-аппаратными средствами информационных служб инфокоммуникационной системы организации; - обеспечению безопасности и целостности данных информационных систем и технологий, а также условий жизненного цикла информационных систем;** **- адаптации приложений к изменяющимся условиям функционирования;**	

- администрированию сетевой подсистемы инфокоммуникационной системы организации. Использовать технологии разработки объектов профессиональной деятельности в различных областях.	
Знать:	ПК-1-З1 - основные библиотеки, применяемые в решении задач из области машинного обучения с учителем ПК-1-З2 - основные классы методов машинного обучения и области их применения
Уметь:	ПК-1-У1 - применять различные методы машинного обучения с учителем для построения моделей по данным ПК-1-У2 - использовать библиотеки языка Python для построения моделей машинного обучения
Владеть:	ПК-1-В1 - навыками формализации и построения математических моделей по массивам данных, а также их реализации в Python ПК-1-В2 - навыками использования библиотек языка Python для построения систем, обучающихся по массивам данных

4. СТРУКТУРА И СОДЕРЖАНИЕ						
Код занятия	**Наименование разделов и тем /вид занятия/**	**Семестр**	**Часов**	**Компетен-ции**	**Литература и эл. ресурсы**	**Примечание**
	Раздел 1. Введение в машинное обучение					
1.1	Предварительная обработка признаков /Пр/	5	2	ПК-1-З1 ПК-1-З2 ПК-1-У1 ПК-1-У2 ПК-1-В1 ПК-1-В2	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.3 Э 1 Э 2	
1.2	Нормализация и стандартизация признаков /Пр/	5	2	ПК-1-З1 ПК-1-З2 ПК-1-У1 ПК-1-У2 ПК-1-В1 ПК-1-В2	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.3 Э 1 Э 2	
1.3	Самостоятельная работа (подготовка к практическим работам)	5	9	ПК-1-З1 ПК-1-З2 ПК-1-У1 ПК-1-У2 ПК-1-В1 ПК-1-В2	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.3 Э 1 Э 2	
	Раздел 2. Линейные модели	5				
2.1	Линейная регрессия /Пр/	5	2	УК-3-З1 УК-3-У1 УК-3-У2 УК-3-В1	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.2 Э 1 Э 2	
2.2	Метрики линейных моделей /Пр/	5	2	УК-3-З1 УК-3-У1 УК-3-У2 УК-3-В1	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.3 Э 1 Э 2	

2.3	Самостоятельная работа (подготовка к практическим и лабораторным работам)	5	19	УК-3-З1 УК-3-У1 УК-3-У2 УК-3-В1	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.2 Л 3.3 Э 1 Э 2	
	Раздел 3. Задачи классификации	5				
3.1	Метрики классификации /Пр/	5	4		Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.3 Э 1 Э 2	
3.2	Логистическая регрессия/Пр/	5	2	УК-3-У1 ПК-1-З1 ПК-1-З2 ПК-1-У1 ПК-1-У2	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.2 Э 1 Э 2	
3.3	Метод опорных векторов/Пр/	5	2	УК-3-У1 ПК-1-З1 ПК-1-З2 ПК-1-У1 ПК-1-У2	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.2 Э 1 Э 2	
3.4	Решающие деревья/Пр/	5	2	УК-3-У1 ПК-1-З1 ПК-1-З2 ПК-1-У1 ПК-1-У2	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.2 Э 1 Э 2	
3.5	Метод ближайших соседей/Пр/	5	2	УК-3-У1 ПК-1-З1 ПК-1-З2 ПК-1-У1 ПК-1-У2	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.2 Э 1 Э 2	
3.6	Самостоятельная работа (подготовка к практическим и лабораторным работам, выполнение домашнего задания)	5	10	УК-3-З1 УК-3-У1 УК-3-У2 УК-3-В1	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.1 Л 3.2 Л 3.3 Э 1 Э 2	

	Раздел 4. Ансамблевые модели	5				
4.1	Случайный лес/Пр/	5	2	ПК-1-З1 ПК-1-З2 ПК-1-У1 ПК-1-У2 ПК-1-В1 ПК-1-В2	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.2 Э 1 Э 2	
4.2	Градиентный бустинг/Пр/	5	2	ПК-1-З1 ПК-1-З2 ПК-1-У1 ПК-1-У2 ПК-1-В1 ПК-1-В2	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.2 Э 1 Э 2	
4.3	Самостоятельная работа (подготовка к лабораторным работам)	5	19	УК-3-З1 УК-3-У1 УК-3-У2 УК-3-В1	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.2 Э 1 Э 2	
	Раздел 5. Анализ текста	5				
5.1.	Лемматизация текста /Пр/	5	2	УК-3-У1 ПК-1-З1 ПК-1-З2 ПК-1-У1 ПК-1-У2	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.3 Э 1 Э 2	
5.2	Регулярные выражения /Пр/	5	2	УК-3-У1 ПК-1-З1 ПК-1-З2 ПК-1-У1 ПК-1-У2	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.3 Э 1 Э 2	
5.3	Определение тональности /Пр/	5	3	УК-3-У1 ПК-1-З1 ПК-1-З2 ПК-1-У1 ПК-1-У2	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.3 Э 1 Э 2	
5.4	Современные NLP-модели /Пр/	5	3	УК-3-У1 ПК-1-З1 ПК-1-З2 ПК-1-У1 ПК-1-У2	Л 1.1 Л 1.2 Л 1.3 Л 2.1 Л 3.2 Э 1 Э 2	
5.5	Самостоятельная работа (подготовка к практическим и	5	17	УК-3-З1 УК-3-У1	Л 1.1 Л 1.2 Л 1.3	

	лабораторным работам, выполнение домашнего задания)			УК-3-У2 УК-3-В1	Л 2.1 Л 3.1 Л 3.2 Л 3.3 Э 1 Э 2	

5. ФОНД ОЦЕНОЧНЫХ СРЕДСТВ

5.1. Вопросы для самостоятельной подготовки к зачету

УК-3-З1, ПК-1-З1, ПК-1-З2

1. Классификация алгоритмов машинного обучения.
2. Регрессия, линейная регрессия, линейные модели регрессии.
3. Целевая функция линейной регрессии.
4. Классификация. Логистическая регрессия.
5. Целевая функция логистической регрессии.
6. Регуляризация.
7. Нейронные сети. Структура нейронной сети.
8. Алгоритм обратного распространения ошибки.
9. Деревья решений. Обучение дерева решений.
10. Алгоритм Random Forest.
11. Алгоритм AdaBoost.
12. Каскад классификаторов. Обучение каскада.

5.2. Перечень работ, выполняемых по дисциплине

По дисциплине предусмотрено два домашних задания

Домашнее задание 1. Проведение экспериментов с программной реализации одного из методов классификации. Исследование влияния параметров алгоритма на точность алгоритма на обучающей и тестовой выборке.
Датасет для домашнего задания выбирается обучающимся самостоятельно

Домашнее задание 2. Анализ текстовых записей и реализация модели NLP-анализа с последующей оценкой ее эффективности
Датасет для домашнего задания выбирается обучающимся самостоятельно

Перечень вопросов для защиты домашнего задания
(УК-3-З1, ПК-1-З1, ПК-1-З2)

1. Метод стохастического градиента. Расписать градиентный шаг для квадратичной функции потерь и сигмоидной функции активации.
2. Недостатки метода SG и как с ними бороться?
3. Обоснование логистической регрессии (основная теорема), основные посылки и следствия. Как выражается апостериорная вероятность классов (надо помнить формулу).
4. Как выражается функция потерь в логистической регрессии (надо помнить формулу).
5. Две мотивации и постановка задачи метода опорных векторов.
6. Какая функция потерь используется в SVM? В логистической регрессии? Какие ещё функции потерь Вы знаете?
7. Что такое ядро в SVM? Зачем вводятся ядра? Любая ли функция может быть ядром?
8. Что такое ROC-кривая, как она определяется? Как она эффективно вычисляется?
9. В каких алгоритмах классификации можно узнать не только классовую принадлежность классифицируемого объекта, но и вероятность того, что данный объект принадлежит каждому из классов?
10. Каков вероятностный смысл регуляризации? Какие типы регуляризаторов Вы знаете?
11. В чём отличия внутренних и внешних критериев?
12. Разновидности критерия скользящего контроля.
13. Что такое критерий непротиворечивости? В чём его недостатки?
14. Основная идея отбора признаков методом полного перебора. Действительно ли это полный перебор?

15. Что такое «проблема мультиколлинеарности» в задачах многомерной линейной регрессии? Какие есть три подхода к её устранению?
16. Сравнить гребневую регрессию и лассо. В каких задачах предпочтительнее использовать лассо?

Также по дисциплине предусмотрено 16 практическихи 8 лабораторных работ

Перечень вопросов для прохождения устного опроса при защите практических работ (УК-3-31, ПК-1-31, ПК-1-32)

1. Что такое объект, целевая переменная, признак, модель, функционал ошибки и обучение?
2. Запишите формулы для линейной модели регрессии и для среднеквадратичной ошибки. Запишите среднеквадратичную ошибку в матричном виде.
3. Что такое коэффициент детерминации? Как интерпретировать его значения?
4. Чем отличаются функционалы MSE и MAE?
5. Что такое градиент? Какое его свойство используется при минимизации функций?
6. Запишите формулу для одного шага градиентного спуска. Какие способы оценивания градиента вы знаете? Почему не всегда можно использовать полный градиентный спуск?
7. Методы Momentum, AdaGrad, RMSProp.
8. Что такое кросс-валидация? На что влияет количество блоков в кросс-валидации?
9. Чем гиперпараметры отличаются от параметров? Что является параметрами и гиперпараметрами в линейных моделях и в решающих деревьях?
10. Что такое регуляризация? Запишите L1- и L2-регуляризаторы.
11. Почему L1-регуляризация отбирает признаки?
12. Запишите формулу для линейной модели классификации. Что такое отступ? Как обучаются линейные классификаторы и для чего нужны верхние оценки пороговой функции потерь?
13. Что такое точность, полнота и F-мера?
14. Что такое AUC-ROC? Опишите алгоритм построения ROC-кривой.
15. Что такое AUC-PRC? Опишите алгоритм построения PR-кривой.
16. Запишите функционал логистической регрессии. Как он связан с методом максимума правдоподобия?
17. Запишите задачу метода опорных векторов для линейно неразделимого случая. Как функционал этой задачи связан с отступом классификатора?
18. В чём заключаются one-vs-all и all-vs-all подходы в многоклассовой классификации?
19. Как измеряется качество в задаче многоклассовой классификации? Что такое микро- и макро-усреднение?
20. В чём заключается преобразование категориальных признаков в вещественные с помощью счётчиков? Почему использование счётчиков может привести к переобучению? Какие методы борьбы с этой проблемой счётчиков вам известны?
21. Какие способы калибровки вероятностей вы знаете? В чём они заключаются?
22. Как определить для линейной модели, какие признаки являются самыми важными?
23. Опишите жадный алгоритм обучения решающего дерева.
24. Почему с помощью бинарного решающего дерева можно достичь нулевой ошибки на обучающей выборке без повторяющихся объектов?
25. Как в общем случае выглядит критерий информативности? Как он используется для выбора предиката во внутренней вершине решающего дерева? Как вывести критерий Джини и энтропийный критерий?
26. Для какой ошибки строится разложение на шум, смещение и разброс? Запишите формулу этой ошибки.
27. Запишите формулы для шума, смещения и разброса метода обучения для случая квадратичной функции потерь.
28. Приведите пример семейства алгоритмов с низким смещением и большим разбросом; семейства алгоритмов с большим смещением и низким разбросом. Поясните примеры.
29. Что такое бэггинг?
30. Что такое случайный лес? Чем он отличается от бэггинга над решающими деревьями?
31. Что такое out-of-bag оценка в бэггинге?
32. Запишите вид композиции, которая обучается в градиентном бустинге. Как выбирают количество базовых алгоритмов в ней?
33. Что такое сдвиги в градиентном бустинге? Как они вычисляются и для чего используются?

34. Как обучается очередной базовый алгоритм в градиентном бустинге? Что такое сокращение шага?
35. В чём заключается переподбор прогнозов в листьях решающих деревьев в градиентном бустинге?
36. Как в xgboost выводится функционал ошибки с помощью разложения в ряд Тейлора?
37. Какие регуляризации используются в xgboost?

5.4. Методика оценки освоения дисциплины

№ п/п	Вид оценочного средства	Критерий	Оценка
1	Выполнение и защита домашних заданий	Все задачи домашнего задания выполнены без ошибок, либо с непринципиальными ошибками, не влияющими на физическую суть результата	«Зачтено»
		Задание не выполнено полностью, либо выполнены не все задачи, либо в решении допущены существенные ошибки, неисправленные после собеседования с преподавателем	«Не зачтено»
3	Устный опрос по материалам лабораторных и практических занятий	Обучающийся уверенно и правильно отвечает на вопросы	«Зачтено»
		Обучающийся не верно отвечает на вопросы	«Не зачтено»

6. УЧЕБНО-МЕТОДИЧЕСКОЕ И ИНФОРМАЦИОННОЕ ОБЕСПЕЧЕНИЕ

6.1. Рекомендуемая литература

6.1.1. Основная литература

Обозначение	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л 1.1	Б. Шарден, Л. Массарон, А. Боскетти	Крупномасштабное машинное обучение вместе с Python	НТБ СТИ НИТУ «МИСиС»	М. : ДМК, 2018
Л 1.2	Андреас Мюллер, Сара Гвид	Введение в машинное обучение с помощью Python	НТБ СТИ НИТУ «МИСиС»	СПб. : ООО "Альфа-книга", 2018.
Л 1.3	Дж. Грас	Data Science. Наука о данных с нуля	НТБ СТИ НИТУ «МИСиС»	СПб. : БХВ-Петербург, 2018

6.1.2. Дополнительная литература

Обозначение	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л 2.1	Т. Геддис	Начинаем программировать на Python	НТБ СТИ НИТУ «МИСиС»	СПб. : БХВ-Петербург, 2019

6.1.3. Методические разработки

Обозначение	Авторы, составители	Заглавие	Библиотека	Издательство, год
Л 3.1	Соловьев А.Ю.	Методические указания к выполнению домашней работы по курсу «Машинное обучение с учителем»	https://lms.misis.ru/enroll/6MMCH4	
Л 3.2	Соловьев А.Ю	Методические указания	https://lms.misis.ru/enroll/6MMCH4	

		к выполнению лабораторных работ по курсу «Машинное обучение с учителем»		
Л 3.3	Соловьев А.Ю	Методические указания к выполнению практических занятий «Машинное обучение с учителем»	https://lms.misis.ru/enroll/6MMCH4	

6.2. Перечень ресурсов информационно-телекоммуникационной сети «Интернет»	
Э 1	Открытое образование [Электронный ресурс]: https://openedu.ru/
Э 2	LMS Canvas [Электронный ресурс]: https://lms.misis.ru
6.3. Перечень программного обеспечения	
П 1	Microsoft Windows
П 2	Microsoft Office
П3	Дистрибутив Anaconda
6.4. Перечень информационных справочных систем и профессиональных баз данных	
И 1	eLIBRARY.RU - НАУЧНАЯ ЭЛЕКТРОННАЯ БИБЛИОТЕКА [Электронный ресурс]: https://elibrary.ru/

7. МАТЕРИАЛЬНО-ТЕХНИЧЕСКОЕ ОБЕСПЕЧЕНИЕ	
7.1	**Аудитория №406** **«Лаборатория прикладного программирования»** Перечень основного оборудования, учебно-наглядных пособий: • Монитор - 9шт. • Персональный компьютер - 9шт. • Проектор • Экран настенный • Усилитель-распределитель • Комплект учебной мебели на 25 посадочных мест.
7.2	**Аудитория №306** **«Кабинет для самостоятельной работы»** Перечень основного оборудования, учебно-наглядных пособий: • проектор; • доска; • экран настенный; • компьютер – 6 шт.; • комплект учебной мебели на 20 человек. В помещении для самостоятельной работы обучающихся имеется подключение к сети «Интернет» и доступ в электронную информационно-образовательную среду организации.

8. МЕТОДИЧЕСКИЕ УКАЗАНИЯ ДЛЯ ОБУЧАЮЩИХСЯ
Промежуточная аттестация по дисциплине предусмотрена в виде зачета Обучение проводится в один семестр и организуется в соответствии с настоящей программой. Самостоятельная работа студентов осуществляется и контролируется с помощью: – индивидуального опроса студентов при защите практических работ, – сдачи домашних заданий Зачет проставляется при условии выполнения учебного плана дисциплины, и по результатам ответов на вопросы к зачету