

УТВЕРЖДАЮ

Проректор по учебной работе и
образовательным инновациям

О.Н. Здрок

«02» июля 2021 г.

Регистрационный № УД – 10417/уч.



**ВВЕДЕНИЕ В МАШИННОЕ ОБУЧЕНИЕ ДЛЯ ПРОМЫШЛЕННЫХ
IOT-РЕШЕНИЙ**

**Учебная программа учреждения высшего образования
по учебной дисциплине для специальности:**

1-31 03 07 Прикладная информатика (по направлениям)

Направление специальности

1-31 03 07 - 01 Прикладная информатика (программное обеспечение
компьютерных систем)

Учебная программа составлена на основе образовательного стандарта высшего образования ОСВО 1-31 03 07-2013 и учебных планов G31-167/уч., G31и-198/уч. от 30.05.2013.

СОСТАВИТЕЛИ:

Кузьмик А.С., преподаватель кафедры технологий программирования факультета прикладной математики и информатики Белорусского государственного университета.

Давидовская М.И., старший преподаватель кафедры технологий программирования факультета прикладной математики и информатики Белорусского государственного университета.

РЕЦЕНЗЕНТ:

Пацей Н.В., заведующий кафедрой программной инженерии Белорусского государственного технологического университета, к.т.н., доцент.

РЕКОМЕНДОВАНА К УТВЕРЖДЕНИЮ:

Кафедрой технологий программирования Белорусского государственного университета
(протокол № 16 от 28.05.2021);

Советом факультета прикладной математики и информатики БГУ
(протокол № 11 от 22.06.2021).

Заведующий кафедрой
технологий программирования _____ А.Н. Курбацкий

ПОЯСНИТЕЛЬНАЯ ЗАПИСКА

Цели и задачи учебной дисциплины

Цель учебной дисциплины – ознакомление студентов с современными подходами в области машинного обучения и нейронных сетей для применения во встраиваемых системах и промышленных решениях «интернета вещей» (Internet of things — IoT), формирование теоретических знаний в области распознавания и сегментации изображений и навыков проектирования и разработки моделей машинного обучения.

Задачи учебной дисциплины:

1. Изучение подходов и методов машинного обучения для применения в области встраиваемых систем и промышленных IoT-решений.
2. Формирование практических умений и навыков применения методов машинного обучения для задач распознавания, сегментации и идентификации.

Место учебной дисциплины в системе подготовки специалиста с высшим образованием.

Учебная дисциплина относится к **циклу** дисциплин специализации компонента учреждения высшего образования. В современном информационном обществе существует востребованность в специалистах по машинному обучению для промышленных IoT-решений, классификации изображений с помощью свёрточных нейронных сетей, трансферному обучению.

В дисциплине уделяется достаточное внимание нейросетевым моделям, анализу изображений, семантической сегментации изображений для подготовки специалистов соответствующей предметной области.

Программа составлена с учетом **межпредметных связей** с учебными дисциплинами. Основой для изучения учебной дисциплины являются учебные дисциплины I ступени высшего образования «Программирование», «Операционные системы», «Технологии программирования».

Требования к компетенциям

Освоение учебной дисциплины «Введение в машинное обучение для промышленных IoT-решений» должно обеспечить формирование следующих академических, социально-личностных и профессиональных компетенций:

академические компетенции:

АК-1. Уметь применять базовые научно-теоретические знания для решения теоретических и практических задач.

АК-2. Владеть системным и сравнительным анализом.

ЛК-3. Владеть исследовательскими навыками.

АК-4. Уметь работать самостоятельно.

АК-5. Быть способным вырабатывать порождать новые идеи (обладать креативностью).

АК-6. Владеть междисциплинарным подходом при решении проблем.

АК-7. Иметь навыки, связанные с использованием технических устройств, управлением информацией и работой с компьютером.

социально-личностные компетенции:

СЛК-3. Обладать способностью к межличностным коммуникациям.

СЛК-6. Уметь работать в команде.

профессиональные компетенции:

ПК-5. Проектировать, разрабатывать, внедрять и тестировать насыщенные Интернет приложения.

ПК-7. Применять профессиональные знания и навыки для проведения научных исследований в области прикладной информатики.

ПК-8. Разрабатывать и совершенствовать методы исследований в области информационных и телекоммуникационных систем.

ПК-9. Работать с научно-технической информацией с использованием современных информационных технологий.

ПК-11. Пользоваться глобальными информационными ресурсами.

ПК-18. Оказывать консультации по вопросам работы программного обеспечения, в том числе разработанного сторонними организациями.

ПК-22. Анализировать результаты работы насыщенных Интернет приложений и вырабатывать предложения по улучшению качества их работы.

ПК-29. Взаимодействовать со специалистами смежных профилей.

ПК-31. Готовить доклады, материалы к презентациям.

ПК-33. Осуществлять поиск, систематизацию и анализ информации по перспективам развития отрасли, инновационным технологиям, проектам и решениям.

В результате освоения учебной дисциплины студент должен:

знать:

- терминологию и понятия машинного обучения;
- алгоритмы классификации и модели машинного обучения;
- методы глубокого обучения.

уметь:

- оценивать эффективность применения методов машинного обучения;
- обеспечивать критерии оценки применения нейронных сетей.

владеть:

- навыками решения задач с помощью методов машинного обучения;
- навыками анализа и интерпретации данных на основе машинного обучения.

Структура учебной дисциплины

Дисциплина изучается в пятом семестре. Всего на изучение учебной дисциплины «Введение в машинное обучение для промышленных IoT-решений» отведено:

– для очной формы получения высшего образования – 54 часа, в том числе 34 аудиторных часа из них: лекции – 34 часа.

Трудоемкость учебной дисциплины составляет 2 зачетные единицы.

Форма текущей аттестации –зачет.

СОДЕРЖАНИЕ УЧЕБНОГО МАТЕРИАЛА

Раздел 1. Машинное обучение для промышленных IoT-решений. Простейшие алгоритмы классификации и регрессии

Тема 1.1. Введение в машинное обучение для встраиваемых систем. Первичная обработка данных

Примеры задач классификации и регрессии в реальной жизни в быту и в сфере промышленного «интернета вещей». Математическая интерпретация задач машинного обучения. Что такое модель машинного обучения. Признаковое описание объектов. Гипотеза компактности. Датасеты. Предобработка данных. Удаление выбросов. Нормализация. Кодирование признаков. Балансировка датасета. Настройка окружения для работы. Язык Python. Библиотеки keras, tensorflow, sklearn.

Тема 1.2. Классификация методом k ближайших соседей, оценка качества моделей

Метрические алгоритмы. Классификация объектов на основе расстояния между ними в многомерном пространстве. Метрики расстояния. Обучающая выборка. Тестовая выборка. Метрики качества моделей. Доля верных ответов. Точность и полнота. AUC-ROC кривая. F1-score.

Тема 1.3. Классификация решающим деревом. Кросс-валидация

Логические алгоритмы. Решающее дерево. Индекс Джини. Энтропия Шеннона. Алгоритм обучения решающего дерева. Критерий остановки обучения. Интерпретируемость результатов. K-fold разбиение. Переобучение. Как превратить дерево-классификатор в дерево-регрессор.

Тема 1.4. Обучение и виды моделей линейной регрессии

Линейная регрессия. Множественная регрессия. Веса признаков. Смещение в признаковом пространстве. Комбинирование признаков. Полиномиальная регрессия. Обучение модели. Функция потерь. Сумма квадратов ошибок. Минимизация функции потерь методом градиентного спуска. Скорость обучения. Переобучение.

Тема 1.5. Переобучение и способы борьбы с ним

Переобучение в рассмотренных ранее моделях. Ограничение на величину листьев решающих деревьев. Регуляризация линейных моделей. L1, L2 регуляризация. Валидационная выборка.

Тема 1.6. Классификация с помощью модели логистической регрессии

Логистическая сигмоида. Связь с линейной регрессией. Свойства логистической регрессии. Обучение модели логистической регрессии. Метод максимального правдоподобия. Выбор порогового значения при классификации.

Раздел 2. Построение ансамблей моделей

Тема 2.1. Стратегия Bootstrap aggregating (Bagging). Классификация с помощью алгоритма случайного леса

Бэггинг. Алгоритм построения модели случайного леса. Независимое обучение множества моделей решающих деревьев. Построение ансамбля деревьев-классификаторов. Разделение обучающей выборки. Бутстрэп. Обучение моделей по разным признакам. Классификация путем простого голосования. Устойчивость к переобучению.

Тема 2.2. Стратегия Boosting. Регрессия с помощью машины градиентного бустинга

Бустинг. Алгоритм обучения модели градиентного бустинга. Вычисление псевдоостатков. Отличие бэггинга от бустинга. Акцент на ошибках при обучении. Построение ансамбля деревьев-регрессоров.

Раздел 3. Нейросетевые модели. Глубокое обучение. Анализ изображений

Тема 3.1. Полносвязная нейронная сеть. Классификация с помощью многослойного персептрона.

Модель нейрона МакКаллока-Питтса. Функция активации нейрона. Персептрон Розенблатта. Композиция нейронов в слое. Связи между слоями. Нелинейность модели как инструмент для более сложного анализа признаков. Перекрестная энтропия как функция потерь для классификации.

Тема 3.2. Обучение нейронной сети. Обратное распространение ошибки. Настройка оптимизаторов

Представление нейросети в памяти. Представление нейросети как математической функции. Минимизация функции потерь. Обратное распространение ошибки как алгоритм эффективного дифференцирования функции нейронной сети. Паралич сети. Затухание градиентов. Эпоха обучения. Оптимизаторы и их виды. Скорость обучения. Другие параметры оптимизаторов. Регуляризация и дропауты как способы борьбы с переобучением нейронных сетей.

Тема 3.3. Классификация изображений с помощью свёрточных нейронных сетей. Трансферное обучение

Архитектура сверточной нейронной сети. Каналы изображений. Свёртка. Свёрточный нейрон. Фильтр. Ядро. Пулинг. Карты признаков. Полносвязная компонента сети. Трансферное обучение. ImageNet. Заморозка весов. Аугментация датасета. Многоклассовая и многозначная классификация.

Тема 3.4. Семантическая сегментация изображений

Понятие семантической сегментации. Encoder и Decoder. Архитектура нейросети для задачи семантической сегментации. Маска изображения.

УЧЕБНО-МЕТОДИЧЕСКАЯ КАРТА УЧЕБНОЙ ДИСЦИПЛИНЫ

Дневная форма получения образования с применением дистанционных образовательных технологий

Номер раздела, темы	Название раздела, темы	Количество аудиторных часов					Количество часов УСР	Форма контроля знаний
		Лекции	Практические занятия	Семинарские занятия	Лабораторные занятия	Иное		
1	2	3	4	5	6	7	8	9
1	Машинное обучение для промышленных IoT-решений. Простейшие алгоритмы классификации и регрессии	12						
1.1	Введение в машинное обучение для встраиваемых систем. Первичная обработка данных	2						устный опрос
1.2	Классификация методом k ближайших соседей, оценка качества модели	2						собеседование
1.3	Классификация решающим деревом, кросс-валидация	2						собеседование
1.4	Обучение и виды моделей линейной регрессии	2						собеседование
1.5	Переобучение и способы борьбы с ним	2						реферат
1.6	Классификация с помощью модели логистической регрессии	2						собеседование
2	Построение ансамблей моделей	6						
2.1	Стратегия Bootstrap aggregating (Bagging). Классификация с помощью алгоритма случайного леса	2						проект
2.2	Стратегия Boosting. Регрессия с помощью машины градиентного бустинга	4						устный опрос
3.	Нейросетевые модели. Глубокое обучение. Анализ изображений	16						

3.1	Полносвязная нейронная сеть. Классификация с помощью многослойного персептрона	2						коллоквиум
3.2	Обучение нейронной сети. Обратное распространение ошибки. Настройка оптимизаторов	4						устный опрос
3.3	Классификация изображений с помощью свёрточных нейронных сетей. Трансферное обучение	6						реферат
3.4	Семантическая сегментация изображений	4						собеседование
	Всего	34						

ИНФОРМАЦИОННО-МЕТОДИЧЕСКАЯ ЧАСТЬ

Перечень основной литературы

1. Бринк Х. Машинное обучение. / Х. Бринк, Дж. Ричардс, Ф. Феверолф. — СПб.: Питер, 2017. — 336 с.: ил. — (Серия «Библиотека программиста»).
2. Паклин Н.Б. Бизнес-аналитика: от данных к знаниям: учеб. пособие / Н. Паклин, В. Орешков. — 2-е изд., доп. и перераб. — СПб.: Питер, 2010. — 701 с.
3. Николенко, С. Глубокое обучение. Погружение в мир нейронных сетей / С. Николенко, А. Кадури, Е. Архангельская. — СПб.: Питер, 2020. — 476 с.
4. Мэтиз, Эрик. Изучаем Python = Python Crash Course : программирование игр, визуализация данных, веб-приложения / Э. Мэтиз ; [перевел с англ. Е. Матвеев]. — 3-е изд. — СПб.: Питер, 2020. — 511 с.
5. Плас, Джейк Вандер. Python для сложных задач: наука о данных и машинное обучение / Дж. Вандер Плас ; [пер. с англ. И. Пальти]. — СПб.: Питер, 2018. — 573 с.

Перечень дополнительной литературы

1. Вьюгин В.В. Математические основы теории машинного обучения и прогнозирования. — М.: 2013. — 387 с.
2. Николенко С., Тулупьев А. Самообучающиеся системы. — М.: МЦНМО, 2009. — 288 с.
3. Жерон О. Прикладное машинное обучение с помощью Scikit-Learn и TensorFlow: концепции, инструменты и техники для создания интеллектуальных систем / пер. с англ. — М.: Диалектика, 2018. — 688 с.
4. Силен, Дэви. Основы Data Science и Big Data. Python и наука о данных / Дэви Силен, Арно Мейсман, Мохамед Али; [пер. с англ. Е. Матвеева]. — СПб.: Питер, 2017. — 334 с.
5. Christopher M. Bishop. Pattern Recognition and Machine Learning. — Springer, Information Science and Statistics series, 2006. — 738 pp.

Перечень рекомендуемых средств диагностики и методика формирования итоговой оценки

Для диагностики компетенций в рамках учебной дисциплины рекомендуется использовать следующие формы:

1. Устная форма: устный опрос, собеседование, коллоквиум.
2. Письменная форма: реферат с устной защитой и оцениванием на основе модульно-рейтинговой системы.

Формой текущей аттестации по дисциплине «Введение в машинное обучение для промышленных IoT-решений» учебным планом предусмотрен зачет.

При формировании итоговой оценки используется оценка знаний студента, дающая возможность проследить и оценить динамику процесса достижения целей обучения. Итоговая оценка предусматривает использование весовых коэффициентов для текущего контроля знаний и текущей аттестации студентов по дисциплине.

Примерные весовые коэффициенты, определяющие вклад текущего контроля знаний в итоговую оценку:

- коллоквиум – 10%;
- реферат – 20%;
- устный опрос – 30 %;
- проект – 10 %;
- собеседование – 30 %.

Итоговая оценка по дисциплине рассчитывается на основе оценки текущей успеваемости.

Рекомендуемая тематика коллоквиума:

1. Коллоквиум «Машинное обучение. Классические алгоритмы классификации и регрессии».
2. Коллоквиум «Нейронные сети и анализ изображений»

Текущий контроль знаний проводится в соответствии с учебно-методической картой дисциплины.

Описание инновационных подходов и методов к преподаванию учебной дисциплины

При организации образовательного процесса используются **метод анализа конкретных ситуаций (кейс-метод) и метод проектного обучения.**

Кейс-метод предполагает:

- приобретение студентом знаний и умений для решения практических задач в области машинного обучения и применения в сфере IoT;
- анализ ситуации, используя профессиональные знания, собственный опыт, дополнительную литературу и иные источники по теме дисциплины.

Метод проектного обучения обеспечивает:

- способ организации учебной деятельности студентов, развивающий актуальные для учебной и профессиональной деятельности навыки планирования, самоорганизации, сотрудничества и предполагающий создание собственного продукта на основе методов машинного обучения;
- приобретение навыков для решения исследовательских, творческих, социальных, предпринимательских и коммуникационных задач на основе методов машинного обучения.

Методические рекомендации по организации самостоятельной работы обучающихся

Для организации самостоятельной работы студентов по учебной дисциплине следует использовать современные информационные ресурсы: разместить на образовательном портале комплекс учебных и учебно-методических материалов (учебно-программные материалы, учебное издание для теоретического изучения дисциплины, методические указания к лабораторным занятиям, материалы текущего контроля и текущей аттестации, позволяющие определить соответствие учебной деятельности обучающихся требованиям образовательных стандартов высшего образования и учебно-программной документации, в т.ч. вопросы для подготовки к зачету, задания, тесты, вопросы для самоконтроля, тематика рефератов и др., список рекомендуемой литературы, информационных ресурсов и др.).

Темы реферативных работ

1. Методы машинного обучения (не рассматриваемых в рамках учебной дисциплины).
2. Машинное обучение в сфере промышленных IoT-решений.
3. Алгоритмы обучения с учителем.

Примерный перечень вопросов к зачету

1. Введение в машинное обучение. Первичная обработка данных.
2. Классификация методом k ближайших соседей, оценка качества модели
3. Классификация решающим деревом, кросс-валидация, переобучение
4. Обучение и виды моделей регрессии
5. Методы борьбы с переобучением
6. Классификация с помощью логистической регрессии.
7. Стратегия Bootstrap aggregating (Bagging).
8. Классификация с помощью алгоритма случайного леса.
9. Стратегия Boosting.
10. Регрессия с помощью машины градиентного бустинга.
11. Полносвязная нейронная сеть.

12. Классификация с помощью многослойного персептрона.
13. Обучение нейронной сети.
14. Обратное распространение ошибки.
15. Настройка оптимизаторов.
16. Классификация изображений с помощью свёрточных нейронных сетей.
17. Трансферное обучение.
18. Семантическая сегментация изображений.

ПРОТОКОЛ СОГЛАСОВАНИЯ УЧЕБНОЙ ПРОГРАММЫ УВО

Название учебной дисциплины, с которой требуется согласование	Название кафедры	Предложения об изменениях в содержании учебной программы учреждения высшего образования по учебной дисциплине	Решение, принятое кафедрой, разработавшей учебную программу (с указанием даты и номера протокола)
Информационная безопасность мобильных приложений	Технологий программирования	Нет	Оставить содержание учебной дисциплины без изменения, протокол № 16 от 28.05.2021

**ДОПОЛНЕНИЯ И ИЗМЕНЕНИЯ К УЧЕБНОЙ ПРОГРАММЕ ПО
ИЗУЧАЕМОЙ УЧЕБНОЙ ДИСЦИПЛИНЕ**

на ____ / ____ учебный год

№ п/п	Дополнения и изменения	Основание

Учебная программа пересмотрена и одобрена на заседании кафедры
_____ (протокол № ____ от _____ 201_ г.)

Заведующий кафедрой

УТВЕРЖДАЮ
Декан факультета
