

Документ подписан простой электронной подписью
Информация о владельце:
ФИО: Локтионова Оксана Геннадьевна
Должность: проректор по учебной работе
Дата подписания: 10.11.2022 09:42:10
Уникальный программный ключ:
0b817ca911e6668abb13a5d426d39e5f1c11eab0173e9456fa98511ca55cd089

МИНОБРАЗОВАНИЯ РОССИИ
Федеральное государственное бюджетное
образовательное учреждение высшего образования
«Юго-Западный государственный университет»
(ЮЗГУ)

Кафедра вычислительной техники

УТВЕРЖДАЮ:

Проректор по учебной работе

О.Г. Локтионова

« 4 » 04

2022 г.



**МАШИННОЕ ОБУЧЕНИЕ И АНАЛИЗ ДАННЫХ
В ЦИФРОВОЙ ЭКОНОМИКЕ**

Методические указания к выполнению курсовой работы
для студентов направления подготовки
09.03.01 Информатика и вычислительная техника

Курск 2022 г.

УДК 004.6

Составители: С.В. Дегтярев, Е.Н. Иванова

Рецензент

Доцент кафедры программной инженерии,
кандидат технических наук

Ю.А. Халин

Машинное обучение и анализ данных в цифровой экономике: методические указания к выполнению курсовой работы для студентов направления подготовки 09.03.01 Информатика и вычислительная техника / Юго-Зап. гос. ун-т; сост.: С.В. Дегтярев, Е.Н. Иванова. – Курск, 2022. – 9 с.: табл. 1. – Библиограф.: с. 9.

Содержатся цель и задачи выполнения курсовой работы. Приводится технологическая карта выполнения курсовой работы. Указываются образовательные технологии. Приводятся примерные темы для курсовой работы. Содержатся критерии оценивания.

Методические указания соответствуют требованиям программ, утвержденным учебно-методическим объединением по направлению Информатика и вычислительная техника.

Предназначены для студентов очной и заочной форм обучения.

Текст печатается в авторской редакции

Подписано в печать *04.04.2022г.* Формат 60x84 1/16.

Усл.печ.л. *0,5* Уч.-изд.л. *0,4* Тираж 20 экз. Заказ *1 101* . Бесплатно.

Юго-Западный государственный университет.

305040, г. Курск, ул. 50 лет Октября, 94.

Введение

Одна из главных целей при подготовке специалистов, способных заниматься учебной, исследовательской деятельностью – научить самостоятельно обоснованно применять методы и способы, позволяющие находить решение.

Цели и задачи курсовой работы

Целью выполнения курсовой работы по дисциплине «Машинное обучение и анализ данных в цифровой экономике» является формирование у студентов умений и навыков подготовки данных для моделей с машинным обучением, извлечения признаков из данных, увеличивающих эффективность решения задач с машинным обучением; владения основными алгоритмами обучения с учителем и без учителя; использования основных инструментальных средств библиотек языков программирования; применения средств визуализации данных; формирования выводов из проведенного анализа.

Задачами курсовой работы являются:

- овладение умением принимать решения на основе данных;
- овладение умением формировать обучающую выборку;
- овладение умением подготавливать данные для машинного обучения;
- овладение умением формировать прогноз и формулировать выводы с помощью моделей машинного обучения;
- овладение умением оценки и оптимизации моделей машинного обучения;
- формирование навыков использовать техники визуализации данных;
- формирование навыков составления моделей для решения задач цифровой экономики.

Примерная технологическая карта выполнения курсовой работы

План выполнения курсовой работы (технологическая карта выполнения курсовой работы) – последовательность действий,

позволяющих эффективно реализовать поставленные задачи курсовой работы, развития умений и навыков студентов, формирования у них прочных знаний.

Таблица 1

Технологическая карта выполнения курсовой работы

№	Тема занятия	Реализуемые компетенции
1	2	3
1	Выбор данных для анализа в соответствии с темой курсовой работы. Необходимо описать набор данных и решаемую задачу	УК-2.1,2.2, 2.3,2.4,2.5 ПК-6.1,6.2, 6.3
2	Предварительный анализ и очистка данных. Необходимо вывести информацию о количественных характеристиках датасета, информацию об отсутствующих значениях, характеристиках и физическом смысле каждого атрибута данных, его значимости для предсказания целевой переменной, вывести несколько точек данных для иллюстрации структуры данных	УК-2.1,2.2, 2.3,2.4,2.5 ПК-6.1,6.2, 6.3
3	Преобразование атрибутов исходного датасета в числовые признаки. Данный этап может включать в себя векторизацию текста, извлечение признаков аудио и видео данных, преобразование изображений в численный массив и другие преобразования в зависимости от типа исследуемых данных	УК-2.1,2.2, 2.3,2.4,2.5 ПК-6.1,6.2, 6.3
4	Описательный анализ данных с формированием выводов. Необходимо определить шкалу измерений каждого признака, выявить аномальные значения,	УК-2.1,2.2, 2.3,2.4,2.5 ПК-6.1,6.2, 6.3

	выполнить визуализацию распределения каждого признака, поверить на нормальность, построить кореллограммы и совместные распределения каждого признака с целевой переменной, выявить коррелированные признаки и признаки, не несущие информации для данной задачи	
5	Применение алгоритмов машинного обучения без учителя: кластеризацию, понижение размерности и поиск аномалий. По результатам необходимо сформировать выводы	УК-2.1,2.2, 2.3,2.4,2.5 ПК-6.1
6	Разделение набора данных на обучающую и тестовую выборки. Необходимо обосновать количественные характеристики и метод разделения (временной, случайный, последовательный)	УК-2.1,2.2, 2.3,2.4,2.5 ПК-6.1,6.2
7	Обучение моделей для решения выбранной задачи. Необходимо проанализировать результаты, сформулировать выводы	УК-2.1,2.2, 2.3,2.4,2.5 ПК-6.1,6.2
8	Выбор перспективной модели для решения задачи. Необходимо изменить гиперпараметры модели, найти оптимальные гиперпараметры	УК-2.1,2.2, 2.3,2.4,2.5 ПК-6.1,6.2, 6.3
9	Проведение усовершенствования моделей. Необходимо применить различные способы: ввести регуляризацию, изменить параметры модели, ввести суррогатные признаки, отобрать признаки, нормализовать данные, ансамбировать модели, изменить алгоритм предварительной обработки данных. По результатам сформировать выводы.	УК-2.1,2.2, 2.3,2.4,2.5 ПК-6.1,6.2, 6.3

10	Изменение порядка предобработки данных для повышения эффективности модели. Необходимо применить понижение размерности для создания суррогатных признаков. По результатам сформировать выводы	УК-2.1,2.2, 2.3,2.4,2.5 ПК-6.1,6.2, 6.3
11	Визуализация результатов моделирования (графики, линии обучения, таблицы сравнения моделей, таблицы классификации и т.п.). Сформировать выводы, спрогнозировать перспективное решение проблемы	УК-2.1,2.2, 2.3,2.4,2.5 ПК-6.1,6.2, 6.3

Критерии оценки

1. Структурированность отчета.

В работе должны прослеживаться четкая структура:

- подготовительный этап по выбору данных;
- анализ данных;
- построение простых моделей;
- сравнение и анализ моделей;
- выводы, прогноз;
- корректировка моделей.

2. Наличие обоснованных выводов.

Работа должна содержать текстовые замечания, поясняющие каждый этап выполнения работы, обоснование выбора того или иного метода, способа и т.п. Важна полнота и адекватность выводов.

3. Временная оценка работы алгоритмов.

Все инструкции, запускающие цикл обучения модели должны содержать замер времени обучения (программный). При сравнении моделей необходимо учитывать и время обучения.

4. Визуализация.

Работа должна содержать визуализацию результатов. Необходимо не только выбрать тот или иной способ визуализации, но обосновать свой выбор. Оценивается разнообразие, наглядность и информативность визуализации.

5. Разнообразие моделей.

В работе необходимо продемонстрировать умение работать с разнообразными моделями обучения, применимыми к одной задаче.

6. Улучшение модели.

В работе необходимо продемонстрировать умение анализировать обученную модель и искать пути для ее совершенствования. Оценивается количество итераций совершенствования модели и их эффективность.

7. Предобработка данных.

Работа должна содержать исчерпывающий алгоритм предварительной обработки данных. Это необходимо, чтобы исправить все несовершенства в данных и сделать набор данных как можно более пригодным для машинного обучения. Оценивается сложность и воспроизводимость процедуры предварительной обработки данных.

8. Использование метрик эффективности.

Оценивается разнообразие и адекватность задаче примененных метрик эффективности (включая время обучения), а также полнота сравнения и правильность выводов из сравнения моделей по разным метрикам.

9. Валидность результатов.

В работе должно быть продемонстрировано умение оценивать достоверность измерения метрик моделей и повышать ее с использованием перекрестной проверки. Если происходит выбор модели, то ее итоговая эффективность должна измеряться не чистом наборе данных.

Примерные темы курсовой работы

1. Применение механизма внимания в методах машинного обучения.

2. Анализ мультимодальных данных с помощью машинного обучения

3. Оптимизация параметров методов машинного обучения с помощью байесовских методов

4. Определение лучшего метода машинного обучения для прикладной задачи в области анализа изображений и текстов

5. Анализ параметров распределения результатов метода

машинного обучения

6. Решение задачи классификации объектов с помощью градиентного бустинга решающих деревьев

7. Система машинного обучения для задачи классификации объектов в условиях ограниченной памяти

8. Интеграция системы машинного обучения для анализа данных с помощью контейнеризации

9. Анализ применимости методов машинного обучения в условиях малого количества данных

10. Анализ параметров выходных нейронов глубоких нейронных сетей

11. Система аугментации данных для задачи классификации объектов

12. Определение лучшей архитектуры глубоких нейронных сетей для прикладной задачи

13. Анализ применимости методов машинного обучения к мультимодальным данным

14. Интеграция генеративно-согласительной модели машинного обучения для анализа данных

15. Сравнение градиентного и адаптивного бустинга для задачи классификации объектов

16. Система машинного обучения для задачи классификации объектов в условиях ограниченной памяти

17. Анализ мультимодальных данных с помощью машинного обучения

18. Определение лучшего метода машинного обучения для прикладной задачи в области анализа изображений и текстов

19. Применение механизма внимания в методах машинного обучения

20. Определение лучшей архитектуры глубоких нейронных сетей для прикладной задачи

21. Определение лучшей архитектуры глубоких нейронных сетей для прикладной задачи

22. Интеграция генеративно-согласительной модели машинного обучения для анализа данных

23. Анализ параметров выходных нейронов глубоких нейронных сетей

24. Анализ параметров выходных нейронов глубоких
нейронных сетей

Список использованных источников

1. Бринк Хенрик. Машинное обучение / Бринк Хенрик, Ричардс Джозеф, Феверолф Марк. – СПб : Питер, 2017. – 336 с. – Текст : непосредственный
2. Гифт Ной. Пракматичный ИИ. Машинное обучение и облачные технологии / Гифт Ной. – СПб : Питер, 2019. – 304 с. – Текст : непосредственный