

Савва Т.Ю.

Методы машинного обучения: методические указания к лабораторным работам

г. Орел

ББК 32.97

С12

Савва Т.Ю. Методы машинного обучения: методические указания к лабораторным работам / Т.Ю. Савва / Учебное электронное издание. — Орел: АНО "Центр интернет-образования", 2018. — 30 с.

Методические указания содержат описание лабораторных работ, направленных на закрепление теоретических знаний и приобретение практических навыков в области разработки моделей машинного обучения, их оценки и оптимизации.

© Савва Т.Ю., 2018

© АНО «Центр интернет образования», 2018

Содержание

Введение	4
1. Лабораторная работа «Основы анализа данных на языке Python»	5
2. Лабораторная работа «Предобработка данных и проектирование признаков»	7
3. Лабораторная работа «Визуализация зависимостей целевой переменной от входных признаков»	9
4. Лабораторная работа «Оценка качества моделей машинного обучения» ..	10
5. Лабораторная работа «Решение задач классификации с помощью байесовского классификатора и метода k-ближайших соседей»	12
6. Лабораторная работа «Градиентные методы в решении задач машинного обучения»	14
7. Лабораторная работа «Метод опорных векторов и логистическая регрессия в решении задач машинного обучения»	16
8. Лабораторная работа «Разработка и оптимизация модели машинного обучения»	18
9. Лабораторная работа «Решение задач машинного обучения с помощью логических методов классификации»	20
10. Лабораторная работа «Использование композиций алгоритмов при решении задач машинного обучения»	21
11. Лабораторная работа «Нейронные сети»	22
12. Лабораторная работа «Решение задач глубокого обучения»	23
13. Лабораторная работа «Решение задач кластеризации»	24
14. Лабораторная работа «Иерархическая кластеризация»	25
15. Лабораторная работа «Решение задач понижения размерности. Метод главных компонент»	26

16. Лабораторная работа «Нелинейные методы понижения размерности»....	27
17. Лабораторная работа «Разработка модели машинного обучения для выбранной предметной области»	28
18. Лабораторная работа «Презентация разработанной модели машинного обучения»	30

Введение

Данные методические указания содержат описание лабораторных работ, направленных на закрепление теоретических знаний и приобретение практических навыков в области разработки моделей машинного обучения, их оценки и оптимизации.

Структура лабораторного курса охватывает различные разделы машинного обучения, в частности, линейные, метрические, логические методы, композиции алгоритмов, основы нейронных сетей в задачах машинного обучения, методы обучения без учителя, такие как кластеризация, и методы понижения размерности. Особое внимание уделено отработке навыков анализа и предварительной обработки обучающей выборки, а также оценке и оптимизации моделей машинного обучения, подбору алгоритмов к решению поставленных задач.

1. Лабораторная работа «Основы анализа данных на языке Python»

Задание на лабораторную работу

Реализовать базовые операции над набором данных на языке программирования Python с применением библиотек NumPy, Pandas и Matplotlib и проанализировать значения рассчитанных показателей.

Ход работы:

1. Получить у преподавателя файл с данными для обработки и анализа.
2. Изучить навигацию в предлагаемой среде разработки программного кода. Создать новый документ для написания программного кода на языке Python. Подключить необходимые для выполнения лабораторной работы программные библиотеки: NumPy, Pandas и Matplotlib.
3. Выполнить загрузку набора данных (обучающей выборки) средствами библиотеки Pandas и вывести загруженные данные на экран. Визуально оценить структуру представленных данных по выведенному на экран фрагменту выборки: наименования столбцов (признаков), типы указанных значений, наличие пустых ячеек и т.д.
4. С помощью встроенных функций библиотек Python реализовать поэлементный доступ к данным обучающей выборки, статистические данные для выборки в целом, а также отдельно по категориальным и числовым признакам.
5. Построить диаграммы рассеяния для визуализации корреляции между признаками. Проанализировать полученные диаграммы и сопоставить сделанные выводы с данными таблицы корреляции, полученной на предыдущем шаге.
6. Предположить, какие задачи могут быть решены на основе представленной обучающей выборки. Выделить целевую переменную и признаки объектов.

Контрольные вопросы

Какие программные библиотеки языка Python используют для решения задач обработки и анализа данных?

Для чего предназначена каждая из перечисленных программных библиотек?

Как осуществляется загрузка данных для последующей обработки?

Какие основные функции из программных библиотек применяются для организации доступа к наборам данных и получения статистических данных на их основе?

Как реализуется построение диаграмм рассеяния, и какие выводы они позволяют сделать?

2. Лабораторная работа «Предобработка данных и проектирование признаков»

Задание на лабораторную работу

Выполнить анализ структуры и предварительную обработку данных, необходимую для подготовки обучающей выборки к моделированию.

Ход работы:

1. Получить у преподавателя файл с данными для обработки и анализа.
2. Проанализировать исходный набор данных (выборку) и оценить его репрезентативность и пригодность для последующего моделирования.
3. Выявить категориальные и численные признаки. Выполнить преобразование категориальных признаков в числовые двоичные.
4. Выполнить обработку отсутствующих значений с учетом предполагаемой степени их информативности, числа пропусков, фактора временной упорядоченности и т.д.
5. Выполнить нормализацию данных для тех признаков, для которых это целесообразно.
6. Реализовать проектирование признаков как минимум для одного признака исходного набора данных.
7. Проанализировать полученную в результате обработки выборку и оценить ее репрезентативность и пригодность для последующего моделирования.

Контрольные вопросы

Для чего выполняется предварительная обработка данных?

Что подразумевают под репрезентативностью обучающей выборки и как степень репрезентативности может повлиять на качество модели машинного обучения?

Что включает в себя предварительная обработка?

Как осуществляется преобразование категориальных признаков в ходе предварительной обработки данных?

Как выполняется обработка пропусков?

Для чего предназначена нормализация данных?

Как вы понимаете понятие важности признака для построения модели машинного обучения?

Что представляет собой процедура проектирования признаков и в чем ее важность для последующего моделирования?

Какова значимость предварительной обработки данных и проектирования признаков для качества моделирования при решении задач машинного обучения?

3. Лабораторная работа «Визуализация зависимостей целевой переменной от входных признаков»

Задание на лабораторную работу

Выполнить визуализацию зависимости целевой переменной от входных признаков посредством графиков и диаграмм и проанализировать полученные графические материалы с точки зрения отбора признаков для последующего моделирования.

Ход работы:

1. Получить у преподавателя файл с данными для обработки и анализа.
2. Рассмотреть предложенный набор данных (обучающую выборку) и оценить, между какими признаками и целевой переменной целесообразно проанализировать характер зависимостей. Определить технику визуализации (мозаичная диаграмма, диаграмма размаха, график плотности, диаграмма рассеяния) для каждого сочетания признаков в зависимости от их типа.
3. Выполнить визуализацию зависимостей между входными признаками и целевой переменной согласно плану.
4. Проанализировать полученные графические материалы с точки зрения отбора признаков для последующего моделирования.

Контрольные вопросы

Для чего выполняется визуализация зависимостей между входными признаками и целевой переменной на этапе предварительной обработки данных?

Какие техники визуализации вы знаете?

Для каких типов признаков используются данные техники визуализации?

Какие выводы позволяют сделать полученные таким образом графические материалы?

4. Лабораторная работа «Оценка качества моделей машинного обучения»

Задание на лабораторную работу

Выполнить обучение моделей классификации и регрессии и оценить их качество с помощью соответствующих методик.

Ход работы:

1. Получить у преподавателя файлы (для задач бинарной и многоклассовой классификации и для задачи регрессии) с данными для обработки и анализа.

2. Для задачи бинарной классификации:

а) выполнить предварительную обработку данных обучающей выборки;

б) выполнить обучение модели классификации посредством заданного алгоритма;

в) выполнить оценку модели и построить и проанализировать ROC-кривую; вычислить значение показателя AUC - площадь области под ROC-кривой;

г) сделать выводы относительно производительность построенной модели бинарной классификации.

3. Для задачи многоклассовой классификации:

а) выполнить предварительную обработку данных обучающей выборки;

б) выполнить обучение модели классификации посредством заданного алгоритма;

в) построить таблицу сопряженности; построить и проанализировать множество ROC-кривых и вычислить соответствующие значения показателя AUC;

г) сделать выводы относительно производительность построенной модели многоклассовой классификации.

4. Для задачи регрессии:

- а) выполнить предварительную обработку данных обучающей выборки;
- б) выполнить обучение модели регрессии посредством заданного алгоритма;
- в) выполнить оценку модели регрессии посредством методом RMSE и с помощью коэффициента детерминации;
- г) визуализировать диаграмму рассеяния для построенной модели;
- д) сделать выводы относительно производительность построенной модели регрессии.

Контрольные вопросы

В чем отличие в оценке производительности моделей классификации и регрессии?

В чем отличие в оценке производительности моделей бинарной и многоклассовой классификации?

Что представляет собой оценка модели на методом отложенных данных и с помощью k-блоков?

Что показывает ROC-кривая и показатель AUC?

Что представляет собой таблица сопряженности?

В чем заключается оценка модели методом RMSE и посредством коэффициента детерминации?

Для чего выполняют анализ распределения величин остатков (ошибок) после оценки модели регрессии?

Как осуществляется сетчатый поиск для повышения качества моделей?

5. Лабораторная работа «Решение задач классификации с помощью байесовского классификатора и метода k-ближайших соседей»

Задание на лабораторную работу

Решить предложенные задачи классификации с помощью байесовского классификатора (Гауссова и полиномиального) и метрического метода k-ближайших соседей.

Ход работы:

1. Получить у преподавателя файлы с данными (№1, №2, №3) и постановку задач классификации.
2. Сформировать обучающую выборку №1 и визуализировать данные выборки. Реализовать Гауссов наивный байесовский классификатор и визуализировать результаты моделирования. Выполнить оценку качества модели классификации.
3. Загрузить обучающую выборку №2. Реализовать полиномиальный наивный байесовский классификатор и выполнить оценку качества модели классификации.
5. Загрузить обучающую выборку №4. Построить модели классификации с помощью метрического алгоритма k-ближайших соседей (для заданного k).
6. Подобрать оптимальный параметр k для той же задачи классификации.
7. Сравнить качество классификации для различных значений параметра k.

Контрольные вопросы

В чем состоит байесовский подход к классификации?

В каких случаях целесообразно применять наивную байесовскую классификацию к решению задач машинного обучения?

В чем состоит специфика метрических методов машинного обучения?

Как осуществляется классификация методом k-ближайших соседей?

Как определить оптимальное число соседей при реализации соответствующего метода классификации?

Почему при решении практических задач нецелесообразно применять классификацию методом ближайшего соседа ($k=1$)?



6. Лабораторная работа «Градиентные методы в решении задач машинного обучения»

Задание на лабораторную работу

Решить предложенные задачи с применением методов линейной и логистической регрессии, самостоятельно реализовав алгоритм градиентного спуска.

Ход работы:

1. Получить у преподавателя файлы с данными для обработки и анализа.
2. Решить задачу с применением линейной регрессии:
 - а) загрузить предложенный набор данных;
 - б) выполнить предварительную обработку данных и проектирование признаков (при необходимости);
 - в) сформировать обучающую и тестовую выборки;
 - г) обучить выборку методом линейной регрессии;
 - д) выполнить предсказание для тестовой выборки;
 - е) оценить качество классификации построенной модели путем построения ROC-кривой и вычисления показателя AUC-ROC;
 - ж) проанализировать полученные результаты.
3. Решить задачу с применением логистической регрессии:
 - а) загрузить необходимые файлы и проанализировать структуру представленных в них данных;
 - б) выполнить предварительную обработку данных и проектирование признаков при необходимости;
 - в) выделить из множества признаков существенные для моделирования, сформировать обучающую выборку и разделить ее на тренировочную и тестовую часть для последующего применения оценки моделирования методом отложенных данных;

- г) обучить модель методом логистической регрессии на тренировочных данных, реализовав при этом метод градиентного спуска самостоятельно;
- д) выполнить предсказание на тестовой выборке;
- е) построить ROC-кривую и рассчитать значение показателя AUC, сделать выводы относительно качества модели;
- ж) добавить в обучающую выборку новый признак и повторить обучение и оценку на новой выборке, сравнить и проанализировать результаты.

Контрольные вопросы

- Что представляет собой метод градиентного спуска?
- Как работает метод линейной регрессии?
- В чем преимущества и недостатки метода линейной регрессии?
- Что представляет собой метод логистической регрессии?
- Что такое регуляризация?
- Какие виды регуляризации вы знаете?
- Как сказывается наличие и отсутствие регуляризации на величине AUC-ROC?

7. Лабораторная работа «Метод опорных векторов и логистическая регрессия в решении задач машинного обучения»

Задание на лабораторную работу

Решить предложенные задачи с применением метода опорных векторов и логистической регрессии.

Ход работы:

1. Получить у преподавателя файлы с данными для обработки и анализа.
2. Решить задачу с применением метода опорных векторов:
 - а) загрузить данные из файла, содержащие тексты по определенным тематикам;
 - б) вычислить TF-IDF-признаки для всех текстов;
 - в) подобрать минимальный лучший параметр C для метода опорных векторов с линейным ядром при помощи кроссвалидации по k -блокам, а качестве меры качества использовать долю верных ответов;
 - г) обучить модель с помощью метода опорных векторов по всей выборке с лучшим параметром C , найденным на предыдущем шаге;
 - д) Найдите 10 слов с наибольшим по модулю весом. Они являются ответом на это задание.
3. Решить задачу с применением логистической регрессии:
 - а) загрузить предложенный набор данных;
 - б) выполнить предобработку данных и проектирование признаков;
 - в) сформировать обучающую и тестовую выборки;
 - г) обучить выборку методом логистической регрессии;
 - д) выполнить предсказание для тестовой выборки;
 - е) построить ROC-кривую и выполнить оценку AUC-ROC для построенной модели;
 - ж) проанализировать полученные результаты и предложить способы повышения качества модели. Попробовать их реализовать.

Контрольные вопросы

В чем особенности применения метода опорных векторов?

Какие вектора называют опорными?

Как влияет на метод опорных векторов выбор вида ядра?

Что представляет собой метод логистической регрессии?

Для каких задач целесообразно применять метод опорных векторов с линейным ядром?

Для каких задач целесообразно применять логистическую регрессию?

8. Лабораторная работа «Разработка и оптимизация модели машинного обучения»

Задание на лабораторную работу

Разработать модель машинного обучения для предложенной обучающей выборки и выполнить оценку качества прогнозирования модели.

Ход работы:

1. Получить у преподавателя файл с данными для обработки и анализа и описание цели моделирования (постановку задачи).

2. Изучить предложенную выборку. Сделать предобработку данных и, при необходимости, выполнить проектирование признаков.

3. Выполнить визуализацию зависимости целевой переменной от признаков объектов. Сделать выводы относительно важности рассмотренных признаков и провести отбор существенных для моделирования столбцов выборки.

3. Выбрать алгоритм для решения задачи и реализовать его. При необходимости сделать соответствующие настройки параметров алгоритма.

4. Выполнить оценку прогностической точности модели. При необходимости вернуться к этапам предобработки данных и выбора реализации алгоритма для повышения точности.

5. Сделать выводы о результатах моделирования, потенциальных возможностях повышения точности модели.

Контрольные вопросы

Какие этапы включает в себя процесс разработки модели машинного обучения?

Какие этапы во многом требуют от разработчика применения эвристических подходов?

Какие признаки были отобраны из всей предложенной выборки для построения модели и почему?

Почему некоторые признаки были исключены до реализации этапа обучения?

Почему для решения поставленной задачи вами был выбран тот или иной алгоритм?

Какие параметры алгоритмы вы установили, и как это отразилось на результате моделирования?

Каким образом выполнялась оценка точности построенной модели?

Является ли, по вашему мнению, достигнутая точность модели достаточной для данной задачи?

9. Лабораторная работа «Решение задач машинного обучения с помощью логических методов классификации»

Задание на лабораторную работу

Решить предложенную задачу классификации с применением решающих деревьев.

Ход работы:

1. Получить у преподавателя файл с данными для обработки и анализа.
2. Выполните при необходимости предварительную обработку данных.
3. Постройте модель классификации методом решающих деревьев заданной глубины.
4. С помощью построенной модели определите вероятности отнесения объектов к заданным классам.
5. С учетом заданного порога преобразуйте вероятности к одному вектору, содержащему бинарные предсказания.
6. Выполните оценку качества классификации построенной модели.

Контрольные вопросы

Какие этапы включает в себя алгоритм принятия решений на основе бинарного решающего дерева?

Как размер обучающей выборки влияет на формируемую алгоритмом структуру решающего дерева?

Для чего используется критерий информативности?

Какие критерии информативности чаще всего используют при построении решающих деревьев?

Как осуществляется обработка пропусков данных при использовании решающих деревьев классификации?

Для чего выполняется редукция?

10. Лабораторная работа «Использование композиций алгоритмов при решении задач машинного обучения»

Задание на лабораторную работу

Решить предложенную задачу классификации с применением алгоритмов «Случайный лес» и «Градиентный бустинг» и сравнить полученные результаты.

Ход работы:

1. Получить у преподавателя файл с данными для обработки и анализа.
2. Выполните при необходимости предварительную обработку данных и проектирование признаков.
3. Обучить модели с помощью алгоритма «Случайный лес» с различным значением числа деревьев.
4. Определить, при каком числе деревьев качество модели на кросс-валидации максимально.
5. Определить, какое минимальное число деревьев дает качество на кросс-валидации выше 60%.
6. Обучить модель с помощью алгоритма «Градиентный бустинг».
7. Построить ROC-кривые для реализованных моделей и сравнить результаты для разных алгоритмов.

Контрольные вопросы

Как реализуется классификация методом «Случайный лес»?

Какие параметры алгоритма можно настраивать?

Как изменяется качество модели, построенной с помощью алгоритма «Случайный лес» при различном числе деревьев для данной задачи?

Как реализуется классификация методом «Градиентный бустинг»?

Какая из реализованных в ходе данной лабораторной работы моделей дает наибольшую оценку качества?

11. Лабораторная работа «Нейронные сети»

Задание на лабораторную работу

Получить базовые навыки по созданию и обучению искусственных нейронных сетей с использованием встроенных средств языка Python и его программных модулей.

Ход работы:

1. Получить исходные данные у преподавателя и согласовать задание на лабораторную работу
2. Изучить материал по использованию встроенных возможностей модулей, реализованных в Python для работы с искусственными нейронными сетями.
3. Создать нейронную сеть в соответствии с предложенным заданием, установив требуемые параметры и функции активации.
4. Выполнить предсказания на построенной нейронной сети и оценку качества модели.
5. Проанализировать полученные результаты и сделать выводы о проделанной работе.

Контрольные вопросы

Что представляют собой искусственные нейронные сети?

Что такое функция активации нейронной сети?

Какие функции можно использовать для активации нейронов?

Почему линейные функции активации не желательно применять в искусственных нейронных сетях?

Чем отличается функция активации для бинарной и многоклассовой классификаций?

Какие параметры используют для настройки обучения искусственных нейронных сетей?

12. Лабораторная работа «Решение задач глубокого обучения»

Задание на лабораторную работу

Получить базовые навыки по созданию и обучению глубоких нейронных сетей средствами языка Python и модуля Keras.

Ход работы:

1. Получить исходные данные у преподавателя и согласовать задание на лабораторную работу
2. Изучить функциональные возможности библиотеки Keras (через Tensorflow или Theano) для работы с глубокими сетями.
3. Создать нейронную сеть в соответствии с предложенным заданием, установив требуемые параметры и функции активации.
4. Выполнить предсказания на построенной нейронной сети и оценку качества модели.
5. Проанализировать полученные результаты и сделать выводы о проделанной работе.

Контрольные вопросы

Чем глубокая сеть отличается от простой искусственной нейронной сети?

Какие типы глубоких сетей вы знаете, и в чем заключаются особенности каждой из них?

Какое практическое применение находят глубокие сети разных типов?

Какие фреймворки могут быть использованы для работы с глубокими сетями?

Какие параметры используют для настройки обучения глубоких сетей?

13. Лабораторная работа «Решение задач кластеризации»

Задание на лабораторную работу

Решить задачу кластеризации посредством алгоритма K-Means.

Ход работы:

1. Получить у преподавателя файл с исходными данными.
2. Загрузить файл с обучающей выборкой. Проанализировать имеющиеся данные. При необходимости выполнить предобработку данных и проектирование признаков.
3. Выполнить кластеризацию на основе алгоритм K-Means с заданными параметрами.
4. После выделения кластеров все пиксели, отнесенные в один кластер, заполнить одним цветом.
5. Выполнить оценку качества получившейся сегментации и проанализировать полученные результаты.

Контрольные вопросы

К какому типу задач относят задачу кластеризации?

Для решения каких практических задач он наиболее часто применяется в настоящее время?

В чем состоит алгоритм кластеризации K-Means?

Какие метрики используются для оценки качества кластеризации?

14. Лабораторная работа «Иерархическая кластеризация»

Задание на лабораторную работу

Получить навыки применения иерархической кластеризации к решению задач машинного обучения.

Ход работы:

1. Получить исходные данные у преподавателя и согласовать задание на лабораторную работу.
2. Изучить материал по программным средствам языка Python, позволяющим реализовать иерархическую кластеризацию.
3. Реализовать иерархическую кластеризацию для предложенной обучающей выборки.
4. Проанализировать полученные результаты и сделать выводы о проделанной работе.

Контрольные вопросы

Чем отличается иерархическая кластеризация от кластеризации на основе K-Means алгоритма?

Какие подходы к реализации иерархической кластеризации существуют?

15. Лабораторная работа «Решение задач понижения размерности. Метод главных компонент»

Задание на лабораторную работу

Решить задачу составления фондового индекса с применением метода главных компонент.

Ход работы:

1. Получить у преподавателя файл с данными для обработки и анализа.
2. На загруженных данных необходимо обучить преобразование PCA с заданным числом компонент.
3. Определить, сколько компонент хватит, чтобы объяснить 90% дисперсии?
4. Применить построенное преобразование к исходным данным и выбрать значения первой компоненты.
5. Загрузить информацию об индексе Доу-Джонса из файла.
6. Определить, чему равна корреляция Пирсона между первой компонентой и индексом Доу-Джонса? Какая компания имеет наибольший вес в первой компоненте?

Контрольные вопросы

В чем заключается метод главных компонент?

Для каких целей он применяется?

Как определить число новых признаков при реализации данного метода?

Как и для чего выполняется оценка корреляции Пирсона при решении задач машинного обучения?



16. Лабораторная работа «Нелинейные методы понижения размерности»

Задание на лабораторную работу

Изучить и опробовать на практике нелинейные методы понижения размерности при решении задач машинного обучения.

Ход работы:

1. Получить у преподавателя файл с данными для обработки и анализа.
2. Выполнить загрузку набора данных (обучающей выборки).
Визуально оценить структуру представленных данных и выполнить при необходимости предобработку данных и проектирование признаков.
3. Применить к подготовленной обучающей выборке один из нелинейных алгоритмов понижения размерности.
4. Выполнить оценку качества и визуализацию результатов.
Проанализировать достигнутое решение.

Контрольные вопросы

Какие нелинейные методы понижения размерности вы знаете?

В чем состоит отличие этих методов от метода главных компонент?

Какие прикладные задачи машинного обучения чаще всего решаются нелинейными методами понижения размерности?

17. Лабораторная работа «Разработка модели машинного обучения для выбранной предметной области»

Задание на лабораторную работу

Разработать модели машинного обучения для выбранной предметной области на основе предложенной обучающей выборки.

Ход работы:

1. Получить или согласовать с преподавателем задание на лабораторную работу, в том числе предметную область и обучающую выборку.

2. Изучить обучающую выборку. Сделать предобработку данных и, при необходимости, выполнить проектирование признаков.

3. Выполнить отбор существенных для решения задачи признаков на основе, в том числе на основе визуализации зависимости целевой переменной от признаков объектов.

4. Выбрать алгоритм для решения задачи и реализовать его. При необходимости сделать соответствующие настройки параметров алгоритма.

5. Выполнить оценку прогностической точности модели. При необходимости вернуться к этапам предобработки данных и выбора реализации алгоритма для повышения точности.

6. Сделать выводы о результатах моделирования, потенциальных возможностях повышения точности модели.

Контрольные вопросы

Какие этапы включает в себя процесс разработки модели машинного обучения?

Какие признаки были отобраны из всей предложенной выборки для построения модели и почему?

Почему некоторые признаки были исключены до реализации этапа обучения?

Почему для решения поставленной задачи вами был выбран тот или иной алгоритм?

Какие параметры алгоритмы вы установили, и как это отразилось на результате моделирования?

Каким образом выполнялась оценка точности построенной модели?

Является ли, по вашему мнению, достигнутая точность модели достаточной для данной задачи?

18. Лабораторная работа «Презентация разработанной модели машинного обучения»

Задание на лабораторную работу

Подготовить и представить доклад с презентационными материалами, содержащий описание процесса разработки модели машинного обучения.

Ход работы:

1. Разработать структуру презентации в соответствии с порядком выполнения этапов разработки модели машинного обучения.
2. Для каждого этапа разработки модели машинного обучения разработать один или несколько слайдов, раскрывающих назначение и особенности выполнения данного этапа, трудности, связанные с его реализацией. Сопроводить слайды графическим материалом и фрагментами таблиц, иллюстрирующих каждый этап разработки.
3. Подготовить доклад в соответствии с составленными презентационными материалами.
4. Представить доклад устно и ответить на предлагаемые вопросы по теме лабораторной работы.

Контрольные вопросы

Какие этапы включает в себя процесс разработки модели машинного обучения?

Какие сложности сопровождают реализацию каждого из этапов?

Какие этапы во многом требуют от разработчика применения эвристических подходов?

Как осуществляется выбор алгоритма машинного обучения при решении конкретной задачи?

Как определить, что разработанная модель готова к внедрению и эксплуатации?