Министерство науки и высшего образования Российской Федерации Федеральное государственное бюджетное образовательное учреждение высшего образования

ТОМСКИЙ ГОСУДАРСТВЕННЫЙ УНИВЕРСИТЕТ СИСТЕМ УПРАВЛЕНИЯ И РАДИОЭЛЕКТРОНИКИ (ТУСУР)

Кафедра автоматизации обработки информации (АОИ)

Отчёт к Лабораторной работе №7

по дисциплине «Нейронные сети и их применение»

Выполнил:

Студент группы 422–M1

\_\_\_\_Белоус Г.В.

Принял:

К.т.н., Доцент кафедры АОИ

\_\_\_\_Аксёнов С.В.

Оглавление

[1 Введение 3](#_Toc138335456)

[1.1 Цель лабораторной работы 3](#_Toc138335457)

[1.2 Задание на лабораторную работу 3](#_Toc138335458)

[2 Ход выполнения работы 5](#_Toc138335459)

[2.1 Регрессор 5](#_Toc138335460)

[2.2 Классификатор 11](#_Toc138335461)

[2.3 Итоговые метрика каждой модели 13](#_Toc138335462)

[3 Вывод 14](#_Toc138335463)

# Введение

## 1.1 Цель лабораторной работы

Получить навыки создания систем объяснений для интерпретации моделей машинного обучения, а также визуализации объяснений работы моделей для сущностей из наборов данных.

## **1.2 Задание на лабораторную работу**

Построить следующие системы объяснений для двух моделей (одного классификатора и одного регрессора), полученных при выполнении работ ранее: графики частичной зависимости (PDP – partial dependence plot), важность перестановок (permutation importance) значения Шепли (Shapley values).

Выполнить загрузку моделей, полученных ранее (классификатор и регрессор). Загрузить выборки данных, использовавшихся для обучения выбранных моделей.

Выполнить следующие задачи для каждой модели:

1. PDP: Построить графики индивидуального условного ожидания (ICE) для всех входных признаков для набора 30-50 записей из оригинальной выборки и средний график частичной зависимости. Выбрать три признака, оказывающие наибольшее влияние на целевую переменную. Для каждой комбинации из двух важных признаков, входящих в Топ-3, построить двумерный график частичной зависимости, а для двух наиболее важных – трехмерный график частичной зависимости.

2. Важность перестановок: Построить диаграммы размаха по всем входным признакам исходя из двух метрик качества, соответствующие моделям.

3. Значения Шепли: Получить итоговый график важности признаков для модели (summary\_plot) – значения Шепли для 200-500 записей в виде beeswarm (SHAP Value (impact on model output)) и barplot (mean(|SHAP|) (average impact on model output magnitude)). Вывести два графика зависимости (dependence plot): для комбинации наиболее важного признака и второго по важности признака, для комбинации наиболее важного признака и третьего по важности признака. Построить график решения (decision plot) для трех записей из набора. Построить силовой график (force plot) для двух других записей из набора.

Cделать выводы о важности признаков, влияющих на целевую переменную, исходя из результатов всех инструментов.  
Содержание отчета.

1. Описание наборов данных.

2. Параметры архитектур и обучения нейронных сетей, использованные для обучения.

3. Графики обучения для архитектур нейронных сетей с лучшими характеристиками эффективности

4. Оценки моделей на тестовых выборках в виде таблиц/ диаграмм, отображающих метрики качества.

5. Программный код с комментариями.

6. Выводы

# Ход выполнения работы

Датасет, данные из которого обрабатывались в данной работе, представляет из себя набор атрибутов для прогнозирования типа лесного покрова на основе картографических переменных в задачи классификации. Исходя из варианта все записи из столбца Cover\_Type со значением “3” (обозначает Ponderosa Pine), получили новое значение “0”, а остальные “1”. Для сокращения времени расчётов была взята выборка величиной 10000 записей. Датасет для задачи регрессии представляет из себя набор атрибутов физико-химических характеристик белого вина с конечной оценкой (quality). В наборе данных чуть меньше 5 тыс. записей, что позволяет использовать датасет полностью.

Обработанные данные были разделены на выборки для обучения и валидации (соответственно 70 и 30%). Предобработка данных осуществлялась с помощью инструментов библиотеки sclearn и метода train\_test\_split.

## Регрессор

Регрессор представляет из себя модель случайного леса с гиперпараметрами полученным из третьей лабораторной работы, подбирались параметры с помощью инструмента GridSearchCV.

В результате вычислений для модели случайного леса, оптимальными гиперпараметрами стали {'max\_depth': None, 'max\_features': 3, 'min\_samples\_leaf': 1, 'min\_samples\_split': 2, 'n\_estimators': 150}.

В итоге модель обучалась на учебной и тестовой выборке, графики PDP, важности перестановок и значения Шелли показаны ниже:

Изображение выглядит как текст, диаграмма, линия, График

Автоматически созданное описание

Рисунок 2.1 – График частичной зависимости для наиболее важных признаков на качество вина

Изображение выглядит как текст, диаграмма, График, снимок экрана

Автоматически созданное описание

Рисунок 2.2 – Трехмерный график, показывающий влияние признаков alcohol и density на оценку вина

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Рисунок 2.3 – График важности перестановок

Значения Шепли (Shapley values) являются концептом из теории кооперативных игр и используются в контексте объяснимости моделей машинного обучения. Они представляют собой меру вклада каждого признака в предсказание модели.

Значения Шепли основаны на идее справедликого распределения выгоды между игроками в кооперативной игре. В контексте объяснимости моделей, модель рассматривается как "игра", в которой признаки являются "игроками". Значения Шепли определяют, какую долю вклада в предсказание каждый признак вносит, учитывая взаимодействия с другими признаками.

Применительно к машинному обучению, Значения Шепли могут быть использованы для объяснения предсказаний моделей, например, для классификации или регрессии. Они позволяют определить, какие признаки вносят наибольший вклад в предсказание модели и в какой степени.

Для вычисления Значений Шепли существует несколько подходов, их выбор зависит от конкретного алгоритма модели и требований. Одним из распространенных методов вычисления Значений Шепли является метод "coalitional game", который основан на переборе всех возможных комбинаций признаков и расчете их вклада в предсказание модели.

Значения Шепли могут быть представлены как абсолютные значения (показывающие вклад каждого признака в предсказание независимо от остальных) или как относительные значения (показывающие вклад каждого признака в предсказание в сравнении с базовым уровнем, например, средним предсказанием модели).

Использование Значений Шепли позволяет более полно и интерпретируемо объяснить предсказания моделей машинного обучения, выявить наиболее важные признаки и понять, как они взаимодействуют друг с другом. Это может быть полезным для принятия решений на основе моделей, проверки их справедливости и повышения доверия к моделям в контексте объяснимости.

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Рисунок 2.4 – График среднего вклада признаков в целевое значение

Изображение выглядит как текст, снимок экрана, диаграмма, линия

Автоматически созданное описание

Рисунок 2.5 – График вклада признаков в целевое значение с цветовым обозначением

Изображение выглядит как текст, диаграмма, снимок экрана, График

Автоматически созданное описание

Рисунок 2.6 – График влияние двух наиболее важных признаков друг на друга

Изображение выглядит как текст, снимок экрана, линия, число

Автоматически созданное описание

Рисунок 2.7 – Пример построения прогноза основываясь на значениях признаков

## Классификатор

Классификатор представляет из себя модель случайного леса, гиперпараметры так же были подобраны с помощью инструмента GridSearchCV, представляют из себя: {'max\_depth': None, 'max\_features': 7, 'min\_samples\_leaf': 1, 'min\_samples\_split': 2, 'n\_estimators': 200}.

Результирующие графики обучения классификатора обученной на учебной на рисунках ниже:

Изображение выглядит как текст, снимок экрана, число, Шрифт

Автоматически созданное описание

Рисунок 2.8 – График важности перестановок

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Рисунок 2.9 – График среднего вклада признаков в целевое значение

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Рисунок 2.10 – График вклада признаков в целевое значение с распространением и цветовым обозначением

Изображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описаниеИзображение выглядит как текст, снимок экрана, Шрифт, число

Автоматически созданное описание

Рисунок 2.11 – Пример построения прогноза основываясь на значениях признаков

## Итоговые метрика каждой модели

Результирующие метрики разобранных моделей находятся в отчётах по второй и третьей лабораторных работах.

# Вывод

В ходе выполнения лабораторной работы были получить навыки создания систем объяснений для интерпретации моделей машинного обучения, а также визуализации объяснений работы моделей для сущностей из наборов данных.

С помощью представленных в работе график можно лучше разобраться в обрабатываемых данных и нагляднее представить результаты построенной предсказательной модели.

Листинг программы находится в соответствующей папке Lab7 по ссылке на репозиторий: https://github.com/peremichka256/NSiIP