

# Отчет по лабораторной работе №1 по курсу «Искусственный интеллект»

Студент группы 8О-306: Киреев А. К.

Работа выполнена: 20.05.2022

Отчет сдан:

Итоговая оценка:

## 1. Тема работы

Базовые алгоритмы машинного обучения

## 2. Постановка задачи

- 1) реализовать следующие алгоритмы машинного обучения: Linear/ Logistic Regression, SVM, KNN, Naïve Bayes в отдельных классах
- 2) Данные классы должны наследоваться от BaseEstimator и ClassifierMixin, иметь методы fit и predict
- 3) Вы должны организовать весь процесс предобработки, обучения и тестирования с помощью Pipeline
- 4) Вы должны настроить гиперпараметры моделей с помощью кросс валидации, вывести и сохранить эти гиперпараметры в файл, вместе с обученными моделями
- 5) Прodelать аналогично с коробочными решениями
- 6) Для каждой модели получить оценки метрик: Confusion Matrix, Accuracy, Recall, Precision, ROC\_AUC curve
- 7) Проанализировать полученные результаты и сделать выводы о применимости моделей
- 8) Загрузить полученные гиперпараметры модели и обученные модели в формате pickle на гит вместе с jupyter notebook ваших экспериментов

## 3. Ход выполнения работы

- 1) Подготовил датасет из прошлой ЛР для работы.
- 2) Реализовал свой класс Logistic Regression с параметрами lr, регуляризацией. Организовал процесс предобработки при помощи pipeline, настроил гиперпараметры при помощи кросс валидации. Получил оценки метрик: Confusion Matrix, Accuracy, Recall, Precision, ROC\_AUC curve при помощи встроенных инструментов sklearn.
- 3) Аналогично реализовал и протестировал SVM, kNN, Naïve Bayes.
- 4) Сравнил мои реализации с реализациями из коробки.

## 4. Выводы

В данной ЛР я познакомился с sklearn и базовыми алгоритмами машинного обучения. Также я попробовал реализовать свои версии этих алгоритмов. Как показали результаты, ручные реализации базовых классификаторов не сильно хуже тех, что есть в sklearn (а некоторые показали себя даже лучше). Линейные модели (логистическая регрессия и линейный SVM) на данном наборе показали себя хорошо. Нелинейный SVM показал себя тоже хорошо, но не сильно лучше линейных моделей. Хуже же всего себя показали kNN и Naïve Bayes. kNN даже

при небольшом числе соседей быстро переобучился, получив ассигасу = 1 на тренировочной выборке, а Naive Bayes просто плохо подошел под задачу в силу своих допущений.