Федеральное государственное бюджетное образовательное учреждение высшего образования

**«ФИНАНСОВЫЙ УНИВЕРСИТЕТ ПРИ ПРАВИТЕЛЬСТВЕ РОССИЙСКОЙ ФЕДЕРАЦИИ»**

**Департамент информационных технологий**

**и анализа больших данных**

Пояснительная записка(план) к курсовой работе

по дисциплине “Машинное обучение”

на тему:

«Машинное обучение в задачах повышения качества изображений»

Выполнил(а):

студент(ка) группы ПМ22-2 факультета информационных технологий и анализа больших данных

Дмитриев Михаил Андреевич

Научный руководитель:

доцент. Остроухова Н.Г.

2024 г

12.05.2024

# ***План по выполненной работе***

**Введение**

* Введение в проблематику повышения качества изображений.
* Обзор существующих подходов и технологий.
* Значимость проблемы для различных областей: от медицины до астрономии.

**Выбор и описание набора данных**

* Обоснование выбора конкретного набора данных для анализа.
* Детальное описание набора данных, включая источник, размер, типы изображений.
* Описание проблем, которые предполагается решить с помощью машинного обучения.

**Предварительный анализ и очистка данных**

* Вывод информации о количественных характеристиках датасета.
* Анализ отсутствующих значений и предварительная очистка данных.
* Иллюстрация структуры данных с помощью нескольких примеров изображений.

**Преобразование атрибутов данных**

* Преобразование изображений в численные признаки.
* Обзор используемых методов преобразования: векторизация, нормализация и т.д.
* Описание выбранных методов предобработки и их влияние на качество модели.

**Описательный анализ данных**

* Визуализация распределений признаков.
* Анализ аномальных значений и их обработка.
* Выявление и обработка коррелированных признаков.

**Применение методов обучения без учителя**

* Понижение размерности для улучшения визуализации и эффективности обучения.
* Кластеризация изображений для выявления основных групп.
* Поиск аномалий для выявления и исключения испорченных изображений.

**Разделение данных на обучающую и тестовую выборки**

* Обоснование выбранного метода разделения.
* Количественные характеристики выборок.

**Обучение моделей**

* Выбор и обоснование применяемых алгоритмов машинного обучения.
* Обучение не менее 7 различных моделей.
* Анализ результатов, выбор наиболее перспективной модели.

**Усовершенствование моделей**

* Оптимизация гиперпараметров с использованием Grid Search.
* Применение техник улучшения модели: регуляризация, ансамблирование, нормализация данных.
* Использование понижения размерности для создания суррогатных признаков.

**Анализ эффективности модели**

* Сравнение моделей по различным метрикам эффективности.
* Визуализация результатов моделирования.
* Сравнение с существующими решениями, обсуждение перспектив улучшения.

**Заключение**

* Подведение итогов выполненной работы.
* Обзор достигнутых результатов и их значимости.
* Предложения по дальнейшим исследованиям в области повышения качества изображений с помощью машинного обучения.

**Приложение**

* Полный код выполнения работы.
* Ссылки на используемые ресурсы и инструменты.

# Описание проблемы

Проблема в области машинного обучения в задачах повышения качества изображений заключается в том, что существующие методы и модели не всегда обеспечивают удовлетворительные результаты в различных условиях и для разнообразных типов изображений. Вот несколько основных аспектов этой проблемы:

1. Обработка различных типов исходных данных: Исходные изображения могут иметь различные характеристики, такие как разрешение, контрастность, яркость, наличие шума и артефактов. Существующие модели могут плохо справляться с такими разнообразиями, что приводит к недостаточному качеству результата.

2. Переносимость: Многие модели обучаются на конкретных наборах данных или для определенных типов изображений, и их производительность может снижаться при применении к новым данным или условиям. Переносимость моделей между различными задачами и наборами данных остается сложной задачей.

3. Эффективность и скорость обработки: Некоторые методы повышения качества изображений требуют значительных вычислительных ресурсов и времени для обработки даже одного изображения. Это ограничивает их применимость в реальных временных задачах или при работе с большими объемами данных.

4. Оценка качества: Оценка качества обработанных изображений является сложной задачей и может быть субъективной. Требуется разработка объективных метрик оценки качества, которые хорошо коррелируют с восприятием человека.

5. Обучение на размеченных данных: Для некоторых методов требуется большое количество размеченных данных для обучения. Подготовка таких данных может быть трудоемкой и дорогостоящей задачей.

Решение этих проблем требует разработки новых эффективных алгоритмов и моделей машинного обучения, улучшения существующих методов, а также совершенствования процессов оценки качества и обучения на размеченных данных.

Ссылка на гитхаб: https://github.com/C0pReI/ML\_Kursavaya