**Описание проекта**

Проект посвящён исследованию интерпретируемости графовых нейронных сетей (GNN, Graph Neural Networks) на различных синтетических и реальных датасетах. В рамках работы реализованы и протестированы современные методы объяснения решений GNN, а также проведён сравнительный анализ их эффективности на задачах классификации графов и узлов.

**Цели и задачи проекта**

Основная цель проекта – анализ и сравнение методов интерпретируемости для графовых нейронных сетей. Для этого реализованы эксперименты на различных датасетах, включающих как синтетические, так и реальные графовые структуры. В ходе работы исследуются следующие задачи:

– Классификация графов (например, молекул, деревьев, синтетических структур)

– Классификация узлов в графах (например, задача Cora)

– Оценка качества объяснений, предоставляемых различными методами интерпретируемости

**Используемые технологии и библиотеки**

В проекте активно используются современные библиотеки для работы с графами и нейронными сетями:

– PyTorch и PyTorch Geometric – для построения и обучения графовых нейронных сетей

– graphxai – для применения и сравнения различных методов объяснения

– scikit–learn, pandas, numpy – для обработки данных и вычисления метрик

– networkx – для генерации и визуализации графов

– matplotlib – для построения графиков и визуализации результатов

**Описание основных экспериментов**

В каждом ноутбуке реализованы следующие этапы:

– Загрузка и подготовка датасета (реального или синтетического)

– Определение архитектур GNN (GCN, GIN, GAT, GraphSAGE)

– Обучение моделей на задаче классификации

– Применение различных методов объяснения (интерпретируемости) к обученным моделям

– Оценка качества объяснений с помощью метрик faithfulness, accuracy и др.

– Визуализация результатов и объяснений

Примеры датасетов:

– Benzene – молекулярные графы, задача бинарной классификации

– Mutagenicity – молекулярные графы, задача бинарной классификации

– Cora – научные статьи, задача классификации узлов

– ShapeGGen – синтетические графы с контролируемой структурой

– TreeDataset – синтетические деревья и недеревья для тестирования моделей и объяснений

**Реализованные модели и методы объяснения**

В проекте реализованы и протестированы следующие архитектуры GNN:

– GCN (Graph Convolutional Network)

– GIN (Graph Isomorphism Network)

– GAT (Graph Attention Network)

– GraphSAGE

Для интерпретируемости используются методы:

– GNNExplainer

– PGExplainer

– IntegratedGradExplainer

– SubgraphX

В каждом эксперименте проводится сравнение моделей и методов объяснения по различным метрикам (точность классификации, качество объяснений, устойчивость и др.). Результаты визуализируются для наглядного сравнения. Особое внимание уделяется анализу того, насколько объяснения совпадают с истинными важными структурами в графах (например, функциональными группами в молекулах или ключевыми узлами в синтетических графах).

**Выводы**

Полученные результаты могут быть полезны для дальнейших исследований в области объяснимого искусственного интеллекта, а также для практического применения GNN в задачах, где важна интерпретируемость решений (например, в химии, биоинформатике, социальных сетях и др.).