# EARIN project Final report

Krzysztof Rudnicki
Jakub Kliszko

June 11, 2023

## 1 Introduction

The goal of our project was to create a model for anime reccomender
After entering anime name from the database model should output recommended animes

## 2 Used data and algorithms

### 2.1 Data

We used different dataset from originally specified in the project description
We decided to use Anime Recommendation Database from Kaggle: LINK
Main reasons why we decided to use this database was that it was bigger than original one, was more recent, it was described as being 100% usable by Kaggle and still had decent amount of code examples
We are mostly interested in rating_complete.csv file which contains information about anime ratings from users who completed the anime

### 2.2 Algorithms

We decided to use collaborative filtering to develop our model, It makes personalized recommandations based on preferences of similar users
We represent anime data-set as embedding vector
We use K-nearest neighbors model and decided to test it out with different

metrics, neighbors and algorithms

### 2.2.1 Algorithms

We decided to test our model with 2 algorithms:

1. Brute
2. Auto

Ball Tree and KD Tree do not work on sparse input (as is the case with our input) so we decided to omit them

### 2.2.2 Neighbor number

We decided to test our model with 5 different neighbor amount:

1. 5 - Popular starting point for small-medium datasets
2. square root of available data - Usually helps to balance between underfitting and overfitting
3. half of available data - Usually usefull for checking overall trend than specific nuances
4. logarithm of available data - Used for very large datasets
5. n-1 neighbors - Usually leads to overgeneralization as we use all instances excepct one for prediciton

### 2.2.3 Metrics

For brute algorithm we tested it will all possible metrics:

1. Cityblock
2. Cosine
3. Euclidean
4. l1
5. l2
6. Manhattan