



# Data Wrangling & Visualization

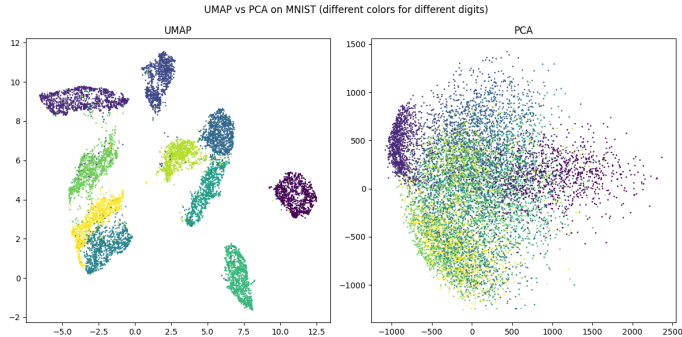
## UMAP Visualization project

Nikita Zagainov, Dmitry Tetkin, Nikita Tsukanov

Innopolis April 2025

# Background: Dimensionality Reduction

- A powerful tool to compress or simplify data
- The most advanced method nowadays - UMAP <sup>1</sup>



Different dimensionality reduction algorithms in action

<sup>1</sup><https://arxiv.org/abs/1802.03426>

- To build a web application capable of providing a useful visualization of given high-dimensional dataset in .csv format
- To show the process of UMAP algorithm fitting to provided dataset
- To provide interactive tools to for exploration of low-dimensional representation of the dataset

## Data preprocessing

We created a pipeline to preprocess the data and prepare it for visualization. This step is crucial as most of the datasets may contain non-numerical features, missing values, etc.

## UMAP fitting

We apply modified version of original UMAP algorithm <sup>2</sup> to the preprocessed data to get low-dimensional representation of the dataset for 100 consecutive iterations.

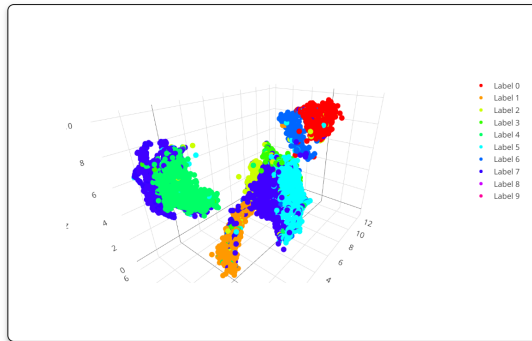
## Visualization app

We use FastAPI framework to deliver embeddings to the frontend application, where Plotly library is used to visualize the data.

---

<sup>2</sup><https://github.com/lmcinnes/umap>

Upload CSV



Frame: 85

Demo of the web application

# Useful links

Follow us on social media!



Project repository



Thank you for your attention!

**it's** **MO** *re than a*  
**UNIVERSITY**