

# LIST OF PROJECTS

## 1. Image classification

We would like to study the classification problem on images.

- Study image classification in general
- Use necessary techniques to process images
- Apply data mining techniques to run the experiments

In the report, detail your learning steps and explain your experimental protocols.

- Comment on the results of different methods
- Compare their complexity
- Compare the obtained results with the benchmark results in the following link
  - + For the handwriting character MNIST database: <http://yann.lecun.com/exdb/mnist/>.
  - + For the Yang dataset: <https://sites.google.com/site/yangdingqi/home/publication?authuser=0>
  - + For COIL20 dataset: <https://www.cs.columbia.edu/CAVE/software/softlib/coil-20.php>
  - + etc.

## 2. Optical character recognition (OCR)

In the context of pattern recognition, detecting and recognizing characters from an image or a scanned document for a specify language is always a challenge.

Analyze the problematics then propose and design a system to accomplish this task on Vietnamese (or English).

- Study the general problem of OCR
- Collect raw data
- Pre-process data (normalize, transformation, pca)
- Choose an appropriate algorithm for the main engine of the system
- Analyze the results and propose the way of improvements

Realize experiments on the selected language character recognition and comment on the results

## 3. One-class classification

The objective is to study and understand the one-class problem which is one supervised task of classification.

- Study the principles of one class classification in literature
- Research the supervised learning algorithms to deal with one classification
- Calculate different quality measure and perform comparisons among different algorithms

Implement and build the experiments on different datasets. Compare the obtained results with the benchmark results in the link: <http://homepage.tudelft.nl/n9d04/occ/index.html>

## 4. Recommender Systems

Recommender systems have become extremely common in recent years, and are utilized in a variety of areas: some popular applications include movies, music, news, books, research articles, search queries, social tags, and products in general. Social networks or retail websites are to be integrated with this kind of systems to propose new stuffs to new customers.

The objective is to learn how a recommender system can be designed.

- A study in recommender systems is necessary, including clustering and bi-clustering techniques
- Choose an appropriate data and analyze data if possible
- Select an algorithm and perform experiments

Reference: <http://www.kdnuggets.com/2016/02/nine-datasets-investigatingrecommender-systems.html>

## 5. Data mining for weather prediction and climate change studies

The data about weather can be collected from the Internet i.e.

[https://www.meteoblue.com/en/weather/archive/export/hanoi\\_vietnam\\_1581130](https://www.meteoblue.com/en/weather/archive/export/hanoi_vietnam_1581130)

The objective is to propose a system to predict the weather in Hanoi using knowledge in Data Mining courses.

- Define the problem: problematics, context, goals
- Collect necessary weather data
- Design data mining system for predict weather information
- Run experiments and analyze the results