



# HUST

**ĐẠI HỌC BÁCH KHOA HÀ NỘI**  
HANOI UNIVERSITY OF SCIENCE AND TECHNOLOGY

ONE LOVE. ONE FUTURE.





ĐẠI HỌC  
BÁCH KHOA HÀ NỘI  
HANOI UNIVERSITY  
OF SCIENCE AND TECHNOLOGY

# WEB MINING

## CAPSTONE PROJECT GUIDELINE

ONE LOVE. ONE FUTURE.

# OUTLINE

---

- Topic
- Conventional project example
- Evaluation
- Contribution

- Link analysis
- Opinion mining
- Information extraction
- Recommendation systems

# Conventional project example

- Step 1: Select a problem: Sentiment classification
- Step 2: Select a method/set of methods: SVM
- Step 3: Select benchmark datasets: imdb
- Step 4: Implementation
- Step 5: Performance report and result discussion

# Conventional project example

- Step 1: Select a problem: Sentiment classification
- The problem should be well-defined
- What is the input? sentence
- What is the output? one of the two classes positive/negative

# Conventional project example

- Step 2: Select a method/set of methods: SVM
- Motivation for the method?
- Explain the methods? Input representation, training, prediction



# Conventional project example

- Step 3: Select benchmark datasets: imdb
- They should be used by previous works
- **AVAILABLE** to your group

# Conventional project example

- Step 4: Implementation
- Which language?
- Which library
- Is code already available? What is your modification?

# Conventional project example

- Step 5: Performance report and result discussion
- Use conventional metrics (P,R,F...)
- Compare to baseline models
- Strong and weak points of the methods

- Strictly following schedule
- Correct format
- Presentation including question answering
- Code
- Team work
- Final report
- “Basically, you must understand your slide, report, code, data, problem, method”

# (Possibly) outstandings

- Bringing new insight into conventional methods (CNN)
- Constructing a valuable new benchmark
- Working with an advanced method (LLM) or a cutting-edge problem (safety AI)
- (\*) Applying to an emergency domain. For example: Applying link analysis/opinion mining/information extraction to analyze a document collection about ...
  - super typhoon yagi and its impact to Vietnam
  - global environmental change and its impact to Vietnam
  - etc

(\* This project type is risky, you should be experienced)



# HUST

# THANK YOU !