

# Chapter 103 : Data Preprocessing

---

સ્ટેપોર્ન રમેશ

---

645020059-9

---

---



## Data Processing

### ▣ Data cleaning ( ការបែងចែកព័ត៌មាន )

ទំនើស ឬស្ម័គ្រ ( ឯកសារអាចមិនមេរាប់លាងដោយសារឡើង )

inconsistency ឯកសារដែលមិនមេរាប់លាង

### ❖ Data integration ( បែងចែក ឬបញ្ចប់ព័ត៌មាន )

### ❖ Data reduction ( ការបង្ការព័ត៌មាន )

ការបង្ការព័ត៌មាន ដូចជាបានការបង្ការ ( អំពីតុលាបែងចែក ឬបង្ការពីការបង្ការដែលមានព័ត៌មាន )

សំណង់ column ដូចជាបានការបង្ការដែលមានព័ត៌មាន

### ❖ Data transformation and data discretization

ការបង្ការព័ត៌មាន ដូចជាបានការបង្ការ

Normalisation ដែលបង្ការព័ត៌មានដែលមានតម្លៃខ្ពស់

ឧបាទ់របស់ការបង្ការព័ត៌មាន ដែលមានតម្លៃខ្ពស់ ដែលមានតម្លៃខ្ពស់ ដែលមានតម្លៃខ្ពស់  
ការបង្ការព័ត៌មានដែលមានតម្លៃខ្ពស់

## Why preprocessing the data

- ការបង្ការព័ត៌មាន ដូចជាអនុវត្តន៍របស់ក្រុមហ៊ុន
- ការបង្ការព័ត៌មាន ដូចជាអនុវត្តន៍របស់ក្រុមហ៊ុន
- ក្នុងព័ត៌មាន មានព័ត៌មានដែលមិនមេរាប់លាង
- ឬជាបានការបង្ការព័ត៌មានដែលមិនមេរាប់លាង

## ព័ត៌មានដែលត្រូវបានបង្ការ

ព័ត៌មានដែលត្រូវបានបង្ការ ដូចជាបានការបង្ការព័ត៌មាន ឬការបង្ការព័ត៌មាន

- incomplete ឬមានព័ត៌មានមិនត្រឹមត្រូវ

- Noisy ( errors, outliers )

- inconsistent ឬមានព័ត៌មានមិនត្រឹមត្រូវ

- intentional ឬមិនមេរាប់លាង

ទំនើស ឬស្ម័គ្រ ( outliers ) ឬការបង្ការព័ត៌មានដែលមិនមេរាប់លាង ដូចជាបានការបង្ការព័ត៌មានដែលមិនមេរាប់លាង

## Incomplete (Missing) data

- សារព័ត៌មានដែលមិនមែនពេលវេលា
  - សារព័ត៌មានទូទៅ ឬអាជីវកម្មដែលមិនមែនពេលវេលា
  - សារព័ត៌មានដែលមិនមែនពេលវេលាដែលមិនមែនពេលវេលា
  - ឯកសារព័ត៌មានដែលមិនមែនពេលវេលា
  - សារព័ត៌មានដែលមិនមែនពេលវេលា
-