

អំពី Pandas dataframe



Dataframe

- Dataframe ជាទិន្នន័យ យកចេញពីកម្មវិធីគ្រប់គ្រងទិន្នន័យទូទៅ ដែលមិនទាន់បានរៀបចំ។
- Pandas ជាបណ្ណាល័យទាញយកឯកសារទិន្នន័យមករៀបចំជា Dataset ឱ្យម៉ាស៊ីនរៀន។

ឯកសារ CSV

The screenshot shows a spreadsheet application interface. The menu bar includes File, Edit, View, Insert, Format, Styles, Sheet, Data, and Tools. The toolbar contains icons for file operations, editing, and formatting. The font settings are Liberation Sans, size 10. The active cell is A1, and the formula bar shows 'sepal.length'. The spreadsheet contains a table with 14 rows and 6 columns. The first row is the header, and the subsequent rows contain numerical data for sepal and petal dimensions, along with the variety name 'Setosa'.

	A	B	C	D	E
1	sepal.length	sepal.width	petal.length	petal.width	variety
2	5.1	3.5	1.4	0.2	Setosa
3	4.9	3	1.4	0.2	Setosa
4	4.7	3.2	1.3	0.2	Setosa
5	4.6	3.1	1.5	0.2	Setosa
6	5	3.6	1.4	0.2	Setosa
7	5.4	3.9	1.7	0.4	Setosa
8	4.6	3.4	1.4	0.3	Setosa
9	5	3.4	1.5	0.2	Setosa
10	4.4	2.9	1.4	0.2	Setosa
11	4.9	3.1	1.5	0.1	Setosa
12	5.4	3.7	1.5	0.2	Setosa
13	4.8	3.4	1.6	0.2	Setosa
14	4.8	3	1.4	0.1	Setosa

រៀបចំទិន្នន័យឆ្នាំក្នុងឯកសារ iris.csv

- ```
field = ['sepal.length', 'sepal.width',
 'petal.length', 'petal.width',
 'variety']
```
- ប្រើអញ្ញត  
  - `skipinitialspace=True`
  - `usecols=fields`

# សរសេរកូដ

```
data_frame = pd.read_csv("data/iris.csv",
 skipinitialspace=True,
 usecols=fields)

បំបែក data frame ទៅជា data & label
data = data_frame[['sepal.length', 'sepal.width',
 'petal.length', 'petal.width']]
label = data_frame[['variety']]
```

# បំប្លែងទៅ numpy array

- `data_array = np.asarray(data)`
- `label_array = np.asarray(label)`

#សាកសរសេរកូដបង្រៀនម៉ាស៊ីន និងព្យាករណ៍