

Knowledge Discovery and Data Mining

Mid-term project

General requirements:

- Using pure python in python notebook to solve the following tasks (Note: only use some common libs in python such as numpy, pandas...)
- Each group, only team lead submits the python notebook to Google classroom.

1. The Eclat algorithm (4 points)

Input: dataset with many transactions, minimum support.

Output: list of frequent itemset.

```
// Initial Call:  $\mathcal{F} \leftarrow \emptyset, P \leftarrow \{ \langle i, \mathbf{t}(i) \rangle \mid i \in \mathcal{I}, |\mathbf{t}(i)| \geq \text{minsup} \}$ 
Eclat ( $P, \text{minsup}, \mathcal{F}$ ):
1 foreach  $\langle X_a, \mathbf{t}(X_a) \rangle \in P$  do
2    $\mathcal{F} \leftarrow \mathcal{F} \cup \{ (X_a, \text{sup}(X_a)) \}$ 
3    $P_a \leftarrow \emptyset$ 
4   foreach  $\langle X_b, \mathbf{t}(X_b) \rangle \in P$ , with  $X_b > X_a$  do
5      $X_{ab} = X_a \cup X_b$ 
6      $\mathbf{t}(X_{ab}) = \mathbf{t}(X_a) \cap \mathbf{t}(X_b)$ 
7     if  $\text{sup}(X_{ab}) \geq \text{minsup}$  then
8        $P_a \leftarrow P_a \cup \{ \langle X_{ab}, \mathbf{t}(X_{ab}) \rangle \}$ 
9   if  $P_a \neq \emptyset$  then Eclat ( $P_a, \text{minsup}, \mathcal{F}$ )
10
```

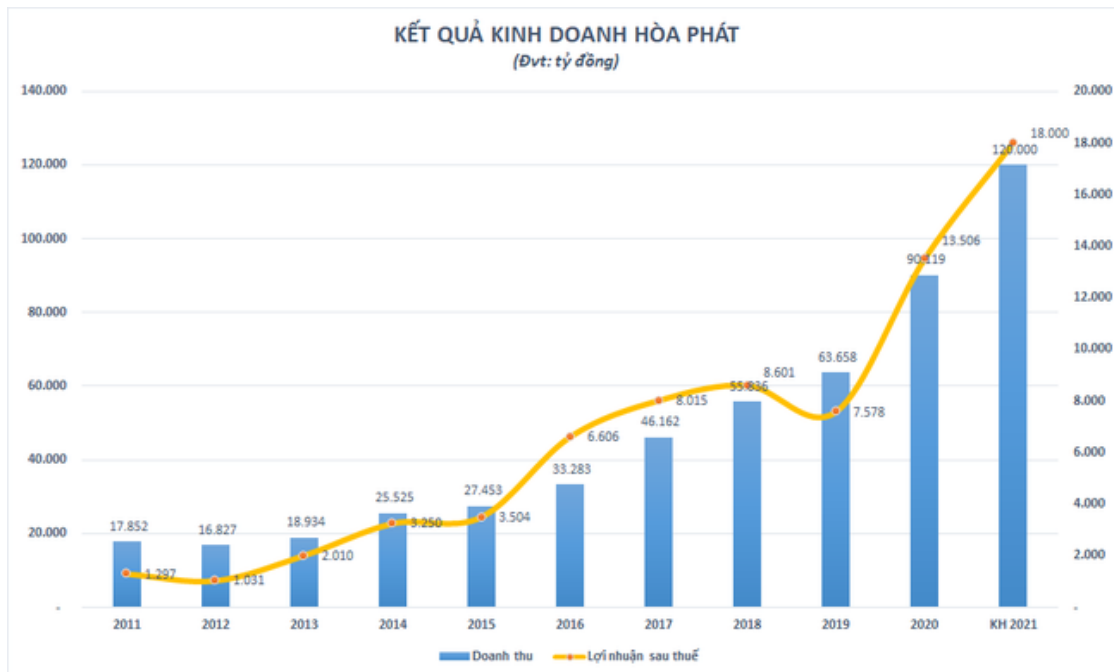
Algorithm explanation: \mathcal{F} is result; P is 1-itemsets (whose supports are larger than minsup) with their tidsets.

2. Implement the crawling task (3 points)

Input: stock symbol (e.g. HPG)

Process: Program goes to <https://dstock.vndirect.com.vn/> to crawl the information to draw the output graph.

Output: Revenue and profit in each year.



3. Implement the crawling task (3 points)

Input: stock symbol (e.g. NLG)

Process: Program goes to <https://dstock.vndirect.com.vn/> to crawl the information to draw the output graph.

Output: Structure of total assets of this company in each quarter.

