**Data Mining & Business Intelligence**

**Aim:** Mini Project

**Requirements:** PC, MS Word, Python 3x, WEKA, Google Chrome

**Theory:**

**Topic:** Programming courses to be taught in educational institutions

**Problem Statement:**

Keeping up with the latest technologies is difficult and time consuming for students looking for placements in todays world. This experiment aims to solve that problem by using data mining techniques to suggest popular languages wanted by the industry to be taught in educational institutions that can include it in their curriculum.

The Stack Overflow Developers survey dataset is used to find the programming languages that users want to learn. This data is then cleaned, visualized and then Apriori association rules are generated to find other related languages a student might want to learn based on a selected one. This grouping of itemsets can be used to provide discounted fee for students or to prepare workshops for these technologies.

**Dataset:**

Name: Stack Overflow Developer Survey (2017)

Records: 51,392 (3924 used)

Attributes: 155 (6 used)

Attribute Information:

|  |  |  |
| --- | --- | --- |
| NO. | Name | Type |
| 1 | Professional | Nominal |
| 2 | Country | Nominal |
| 3 | DeveloperType | Nominal |
| 4 | NonDeveloperType | Nominal |
| 5 | HaveWorkedLanguage | Nominal |
| 6 | WantWorkLanguage | Nominal |

**Apriori Association**

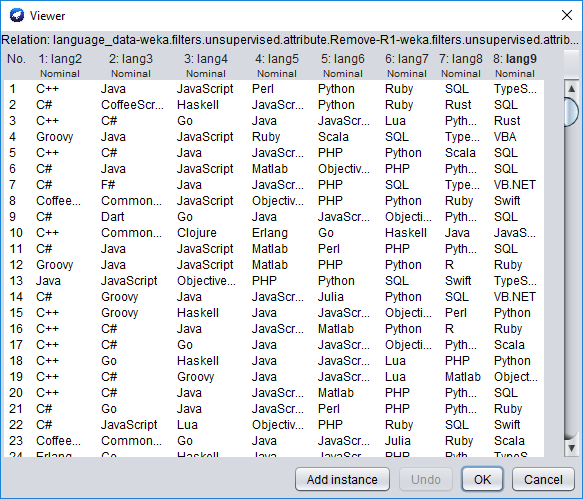


Fig 1

Dataset is imported in WEKA, selecting only 8 languages per record

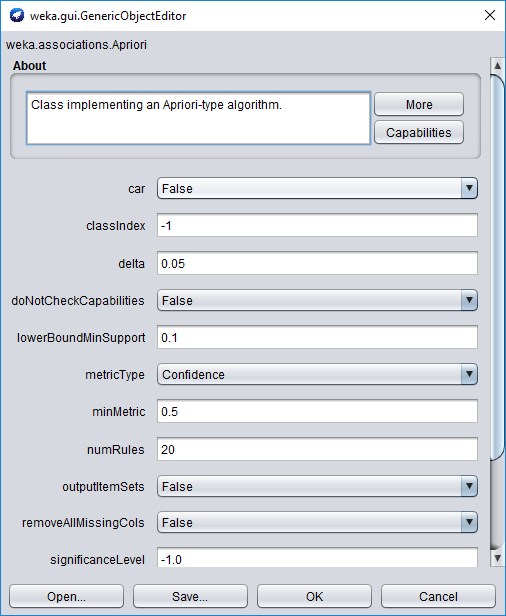


Fig 2

Apriori Association is applied to the dataset, with support 0.1 and confidence 0.5

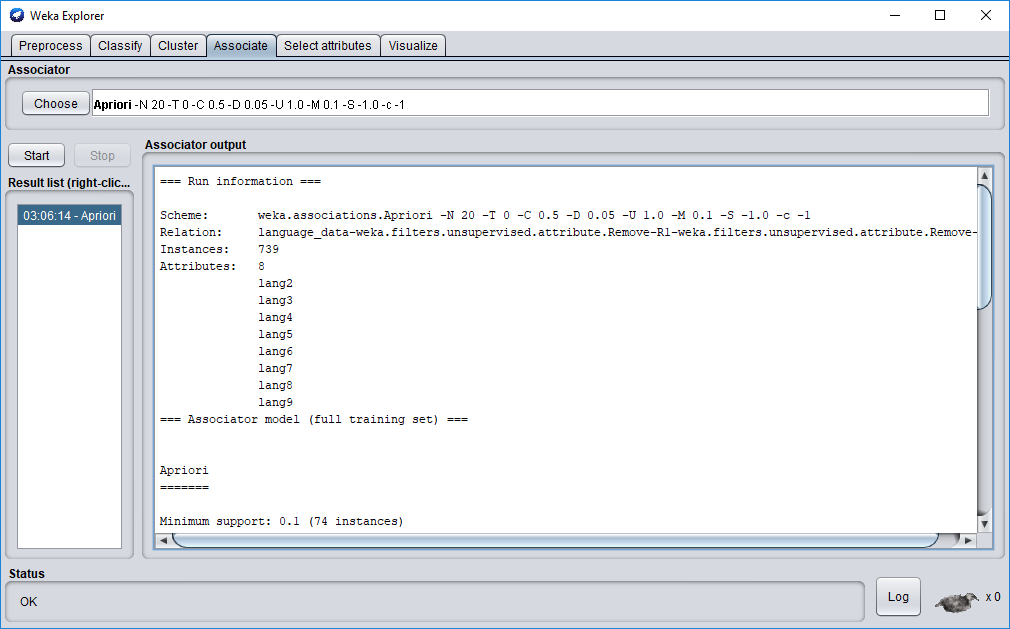


Fig 3

Results of association are shown

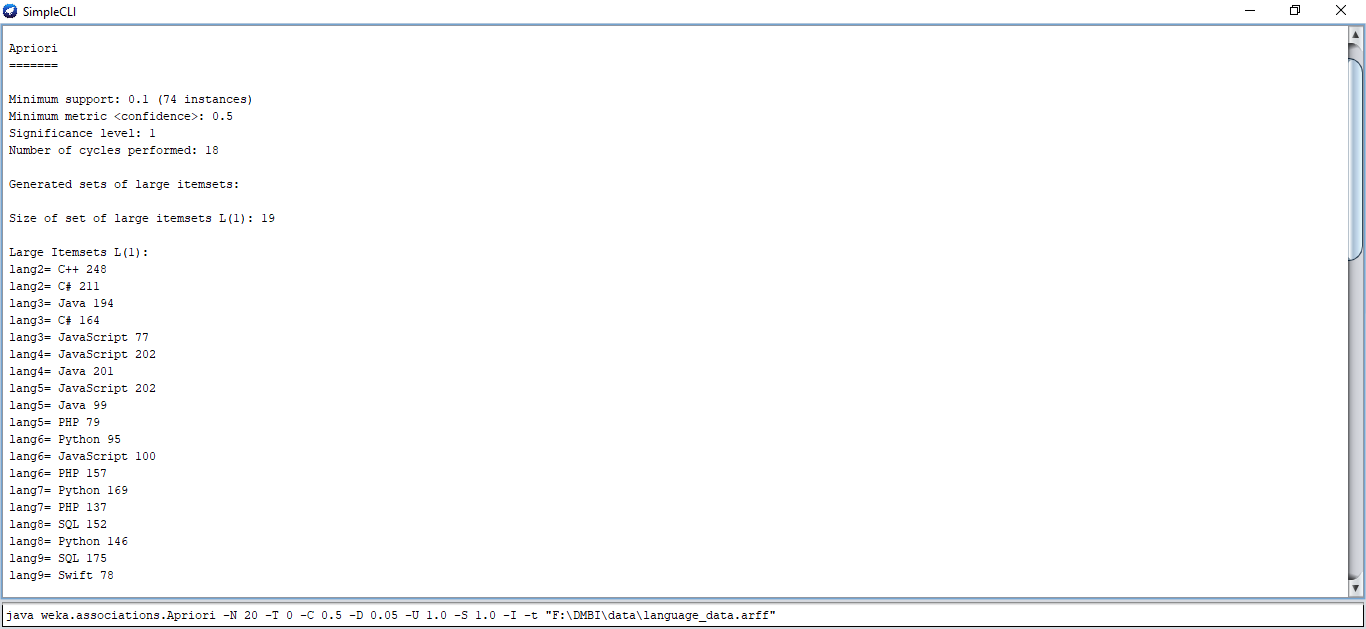




Fig 4

Apriori Association shows frequent itemsets in L1, L2 and L3

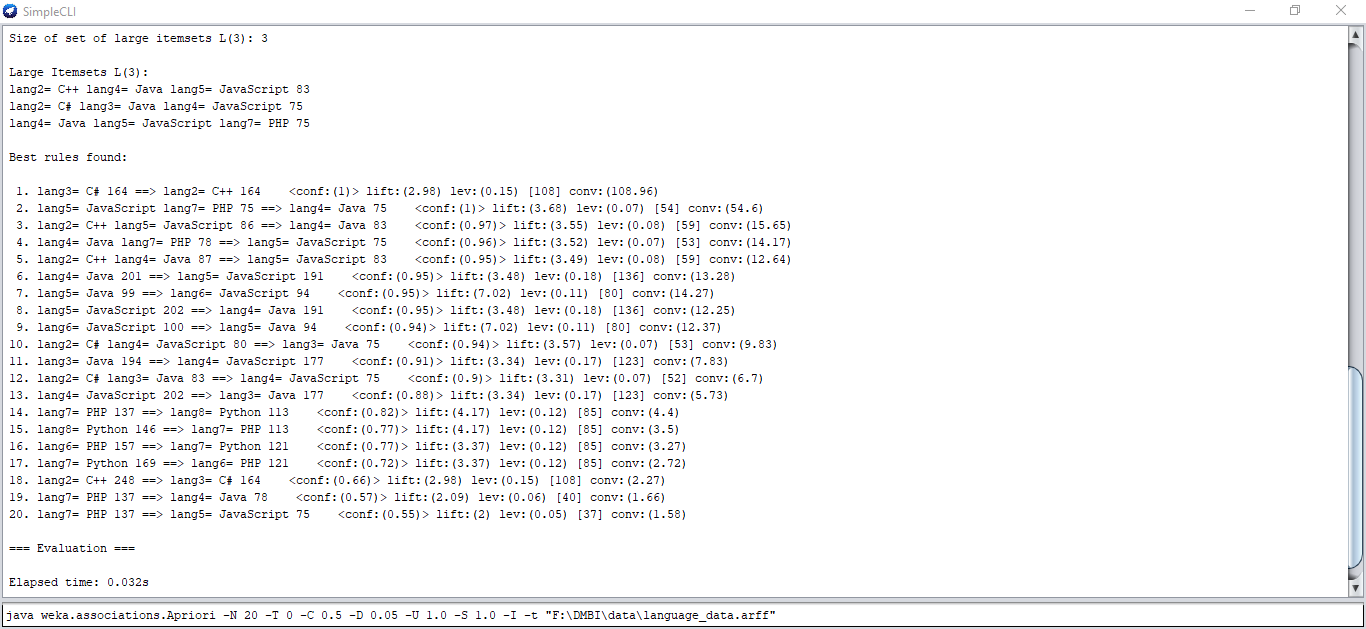


Fig 5

Apriori Association Mining generates frequent itemsets and association rules that show frequently grouped together Programming languages