Identifying the usage patterns in social networks with data mining tecniques – Facebook case

Ahmet Selman Bozkır, Sacide Güzin Mazman,
Ebru Akçapınar Sezer

Hacettepe University

Issues

- Data mining and methods involved in the study
- Data properties
- Identifying the factors affecting *"frequency of access to Facebook"* and *"Facebook usage time"* via different classifiers
- Association rules discovery

facebook

- More than 500 million active users
- Average user has 130 friends
- People spend over 700 billion minutes per month on Facebook
- About 70% of Facebook users are outside the United States
- More than 30 billion pieces of content (web links, news stories, blog posts, notes, photo albums, etc.) shared each month.
- There are more than 150 million active users currently accessing Facebook through their mobile devices.

Data Mining

Data mining: is the process of extracting hidden & useful information from raw data by utilizing statistics, AI and machine learning techniques and smart algorithms.


An iterative and interactive process
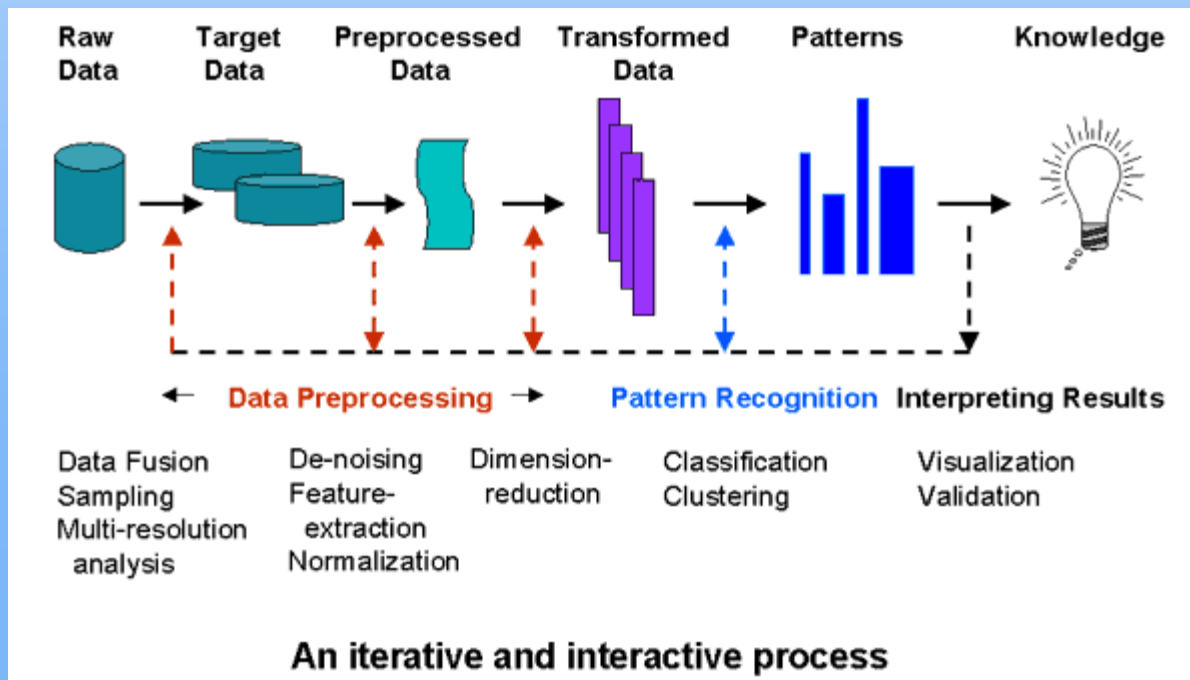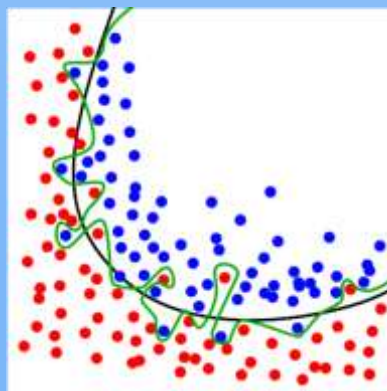
Data Mining (cont'd)

Predictive Methods

1. **Classification (Decision Trees, Bayesian Classification, etc…)**
2. **Regression (CART, Kernel Ridge Regression etc..)**
3. **Artificial Neural Networks**
4. **Kernel Based Methods (SVM, RVM etc…)**
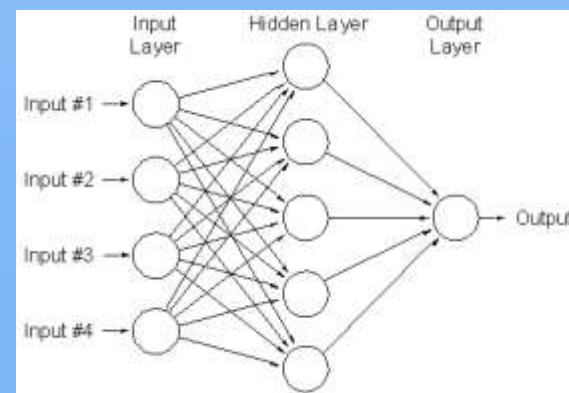
Descriptive Methods

1. **Clustering (K-Means , Hierarchical Clustering, EM etc…)**
2. **Association Rules (Apriori, GRI etc..)**
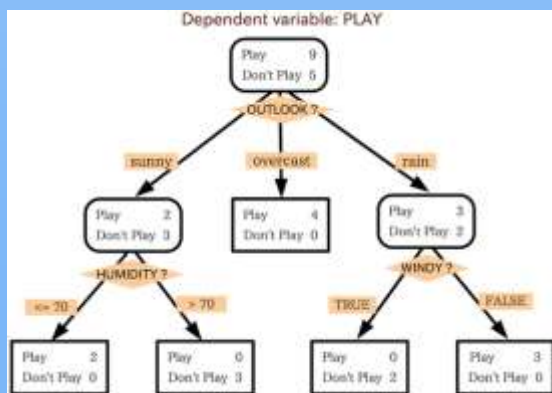
Decision Trees

SVM

ANN

Data properties

Table 1: General demographic & usage data

| Variable Name	Type	Available Answers
Sex	Discrete	Male / Female
Age	Discrete	18-25 / 26-35 / 36-40 / 41 and above
Frequency of access to Facebook	Discrete	Once in a year / Once in a month / Several times in a week / Once in a day / Several times in a day
Facebook usage time	Discrete	Less then 15 mins / Half an hour / 1 hour / 1-3 hours / More than 3 hours
Education Level	Discrete	High School / Bachelor / Master
Membership to any group	Discrete	Yes / No
Membership to student groups	Discrete	Yes / No
Membership to common interest groups	Discrete	Yes / No
Membership to internet & tech groups	Discrete	Yes / No
Membership to organizations	Discrete	Yes / No

Actual data contains **570 cases**

Data properties

**Table 2: Facebook usage in the educational perspective**

| Variable Name	Type	Available Answers
E1: Facebook contributes to the communication between classmates.	Discrete	1,2 .. 10 [Stongly disagree, - → Strongly agree]
E2: It contributes to the communication between teacher & student	Discrete	""
E3: It contributes to the discussion in class	Discrete	""
E4: It contributes to transferring course materials and resources	Discrete	""
E5: It contributes to propagation of announcements of lectures & classes	Discrete	""
E6: It's useful at assigning tasks in classes & homeworks	Discrete	""
E7: It's useful at building academic groups in the direction of common senses and requirements	Discrete	""
E8: It's useful at information sharing in classes	Discrete	""
E9: It's useful at executing the group workings	Discrete	""
E10: It's useful at accesing to the rich resources to learn	Discrete	""
E11: It's useful at providing rich multimedia contents in teaching	Discrete	""

Data properties
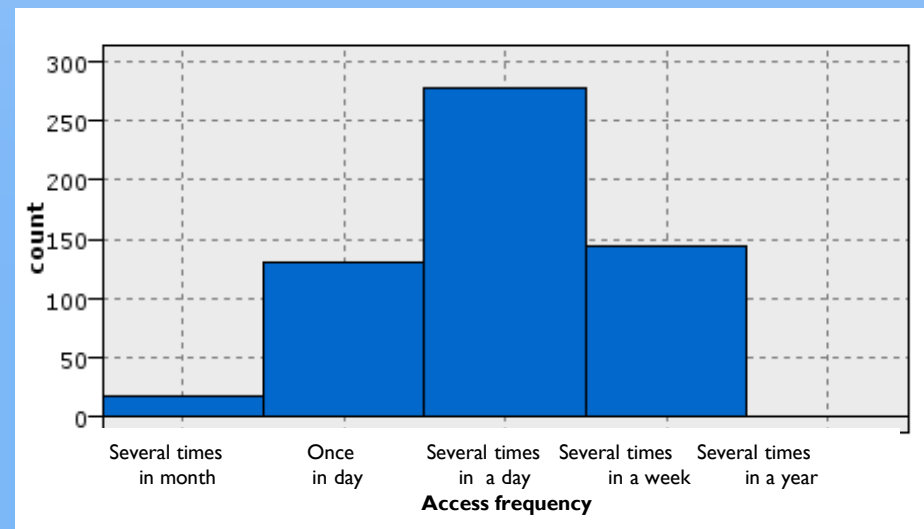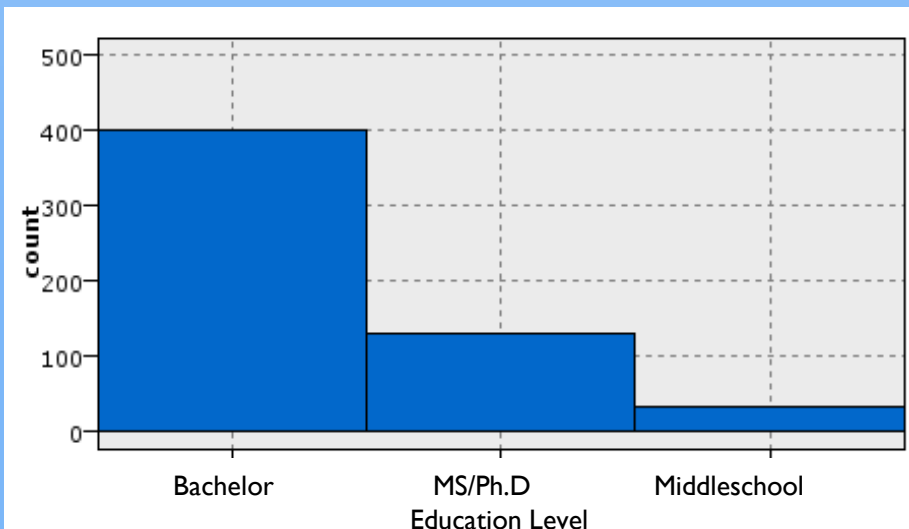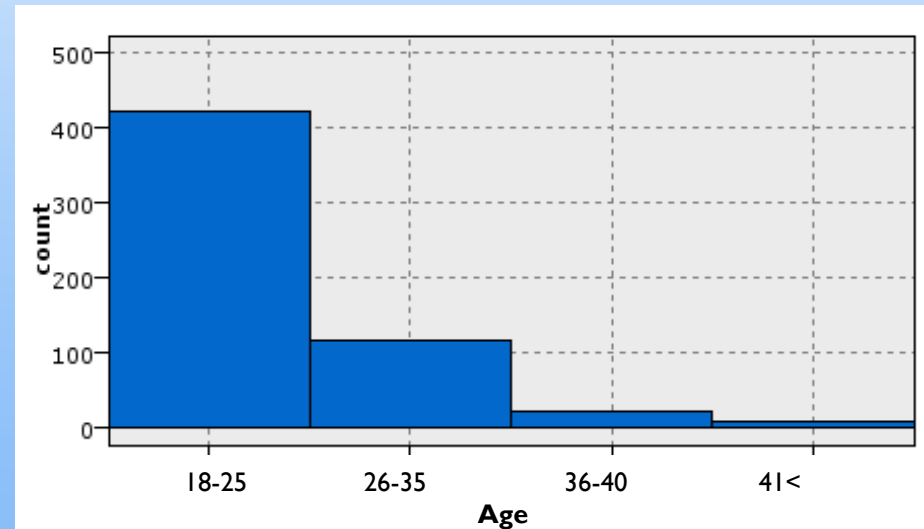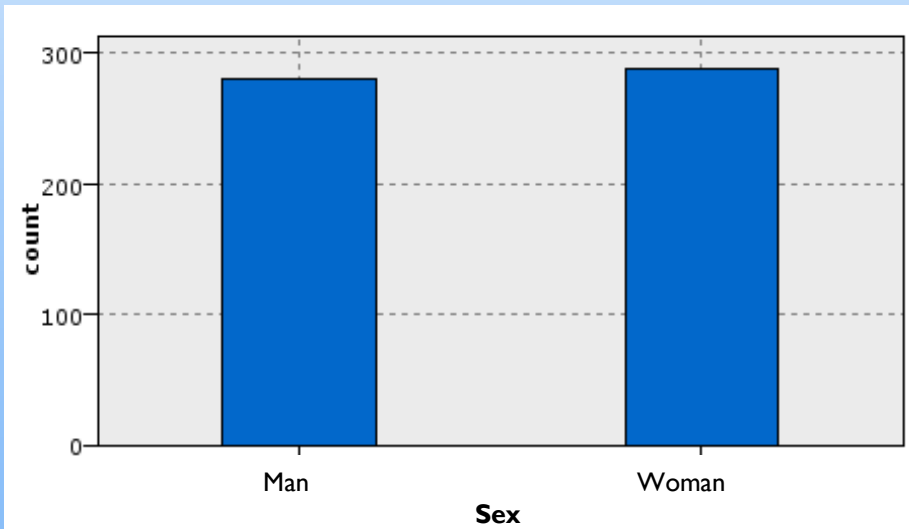
Data properties
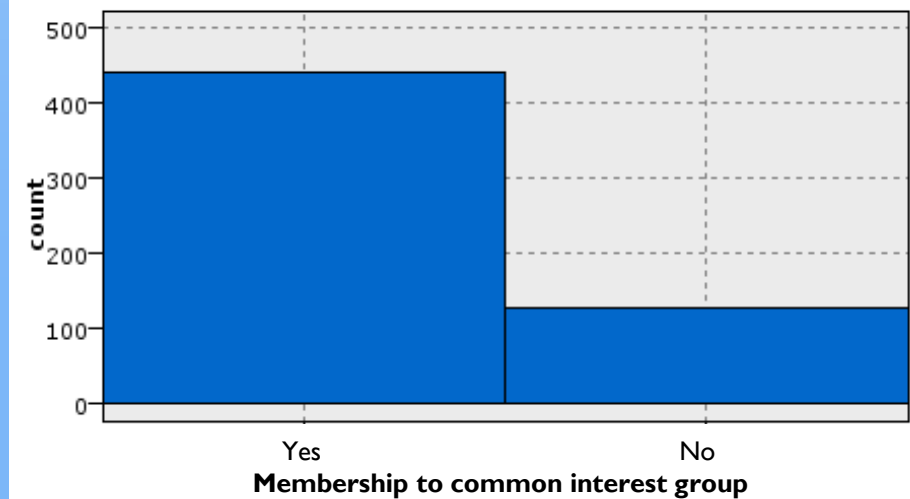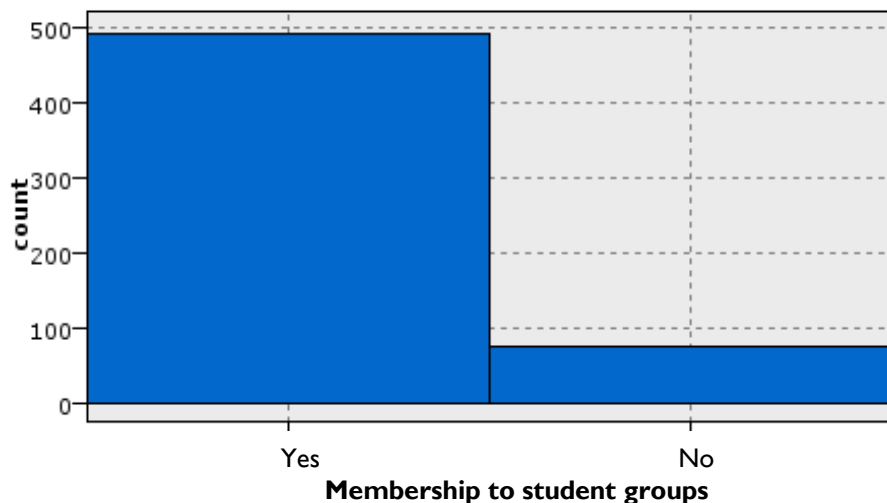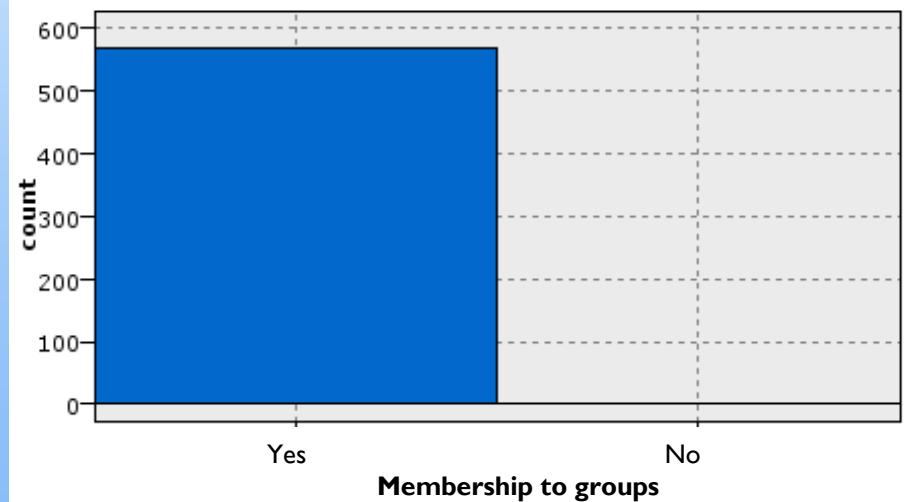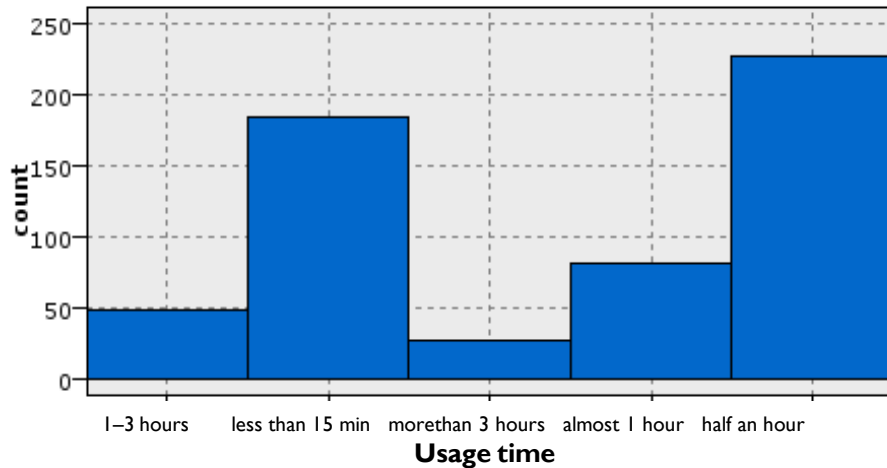
Data properties
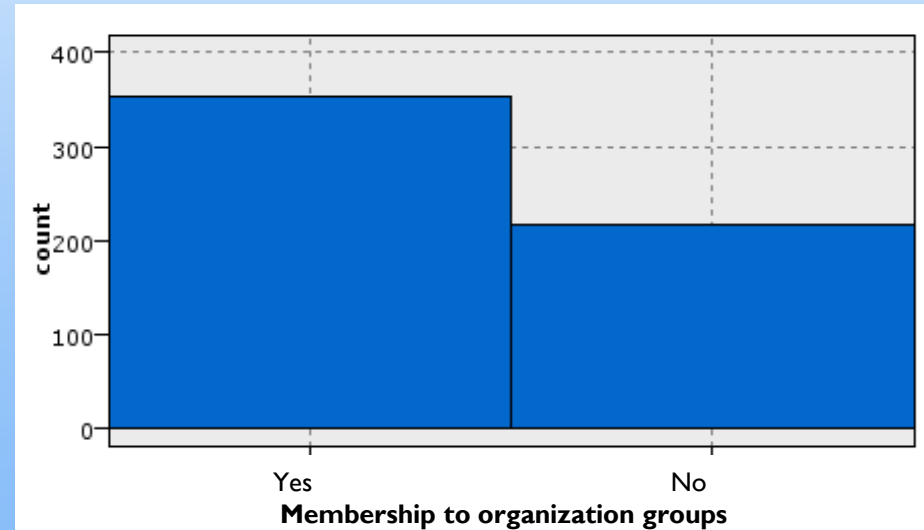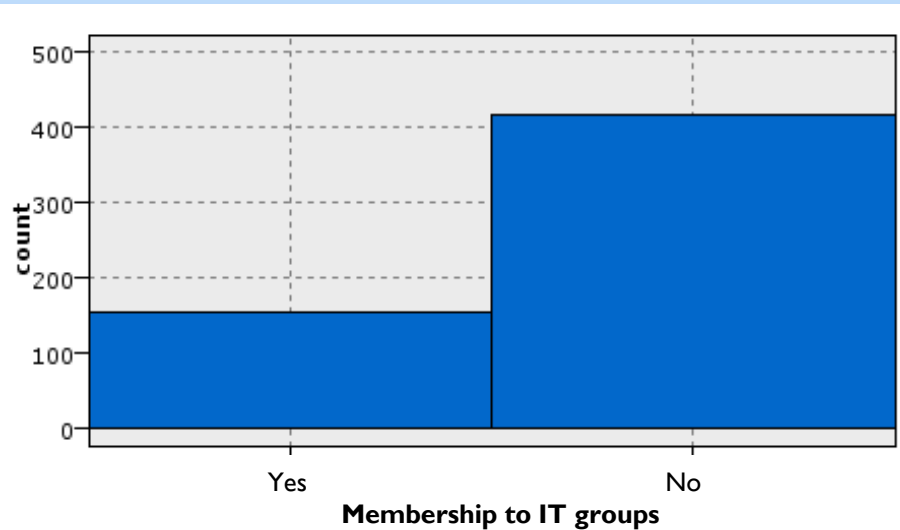
Classification Phase

- Target:
  Identifying the factors affecting *"frequency of access to Facebook"* and *"Facebook usage time"* via different classifiers and finding best performance giving technique.

- Applied Techniques:
  C5, CART, CHAID, Bayesian Network, ANN and SVM.

- Software Platform:
  Spss Clementine 12

Classification Phase

- Target Variable: <u>Access Frequency</u>

  SVM: 69.65% true classification / 30.35 % false classification

  C5:   55.79% / 44.21%
  Chaid: 50.35% / 49.65%
  C&RT: 52.81% / 47.19%
  Bayes:53.51% / 46.49%
  ANN: 48.77% / 51.23%

Variable Importance
Targets: Access frequency
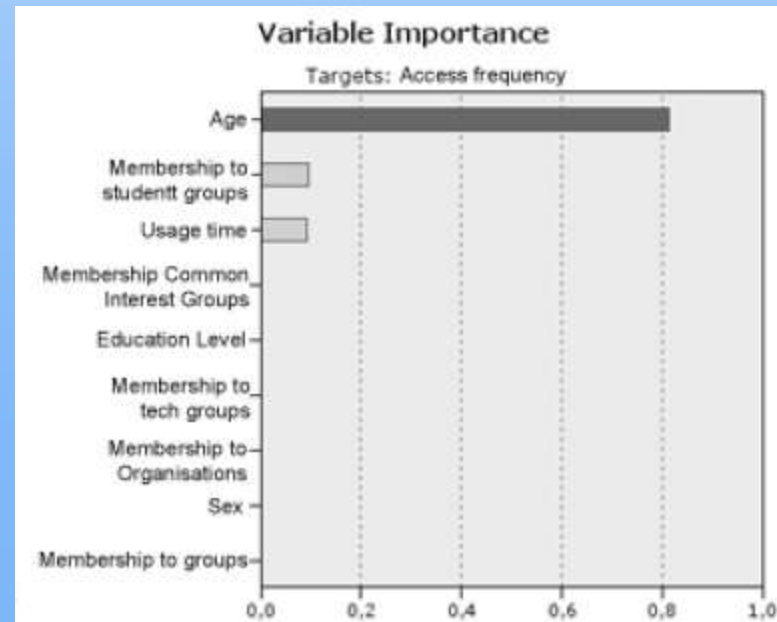
Classification Phase

- Target Variable: <u>Usage Time</u>

SVM: 62.63% true classification / 37.37 % false classification

C5:   47.54% / 52.46%

Chaid: 41.4% / 58.6%

C&RT: 43.68% / 56.32%

Bayes: 45.09% / 54.91%

ANN: 47.72% / 52.28%

Variable Importance
Targets: Usage time
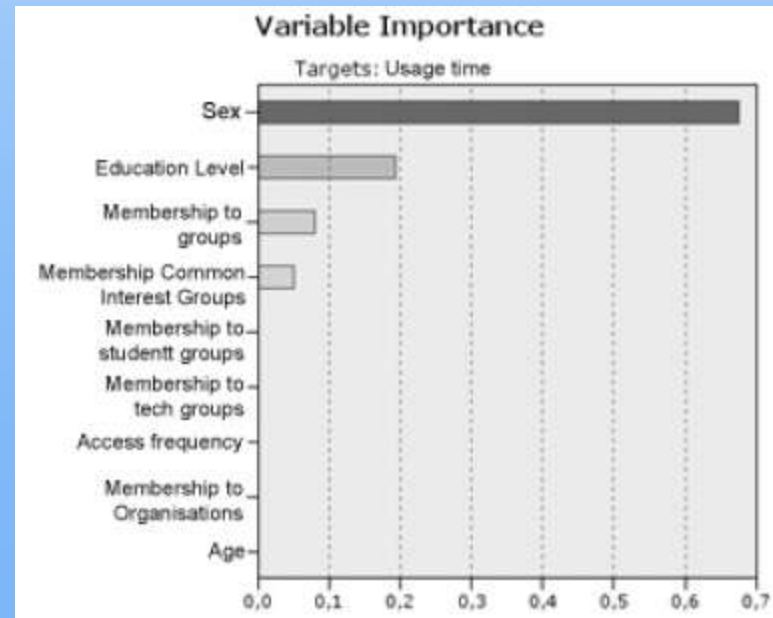
Association Analysis

- In association rules mining study, it's intented to discover hidden relations in the dataset by merging two tables introduced at startup.

- SPSS Clementine and <u>Apriori</u> technique is employed with *%5 support* and *%15 confidence* settings.

Apriori Analysis

Ancedent	Consequent	Confidence	Rule Support	Lift
"It contributes to the communication between teacher & student" = 7	"It's useful at accessing to the rich resources to learn" = 9	11.32%	1.03%	3.07
"Facebook contributes to the communication between classmates" = 8 **and** "It contributes to the communication between teacher & student" = 8	"It contributes to transferring course materials and resources" = 6	20.69%	1.05%	2.7
"It contributes to the communication between teacher & student" =3	"Facebook contributes to the communication between classmates" =6	17.02%	1.40%	2.48
"Facebook contributes to the communication between classmates."=6	"It's useful at providing rich multimedia contents in teaching"= 3	20.51%	1.40%	2.65
"Facebook contributes to the communication between classmates"=7 **and** "It contributes to the communication between teacher & student"=5	"It contributes to propagation of announcements of lectures & classes" = 2	12.5 %	0.52%	2.03

As a consequence...

- Social network (Facebook in particular) usage patterns can be identified by the help of data mining techniques.

- According to classification sucees rates, SVM is the leading classifier among the others. A decision tree method C5 is the next one.

- Apriori algorithm can be used to discover associations among the ideas of students in social network usage.

The End

- Thanks for listening...

- Questions?