ECSP 2015 Geneva 1-3 July
2nd European Congress for Social Psychiatry

# Symptom Checklist - 90 - Revised Rating Scale: A Data Mining Approach

Evgenia Gkintoni[1], Magdalena Nikiel[2], Sampson Fytros[2], Constantinos Halkiopoulos[2,3], Gerasimos Antzoulatos[2,3]

[1]*Dept. of Psychology, University of Crete*

[2]*Dept. of Business Administration, Technological Educational Institute of Western Greece*

[3]*Dept. of Mathematics, Univ. of Patras Artificial Intelligence Research Centre*

## Introduction

In this paper Machine Learning and Data Mining methods were applied to capture the psychological state of students between 18-26 years old. For recording, tracking and evaluation of the psychological condition, was used the standardized scale "*Symptom Checklist-90 (SCL-90)*", which examines a wide range of psychological problems and symptoms of psychopathology.

## Symptom Checklist – 90 - Revised

SCL-90 consists of 90 questions and identifies 9 clinical points (sub-scales) and 3 answer sets enclosing the rating of individual scales. The sub-scales are defined as follows: Somatization[a], Obsessive-Compulsive[b], Interpersonal Sensitivity[c], Depression[d], Anxiety[e], Anger-Hostility[f], Phobic Anxiety[g], Paranoid Ideation[h], Psychotism[i].

a. **Somatization**: reflects the discomfort (anxiety, restlessness) derived from the perception of physical dysfunction.
b. **Obsessive-Compulsions**: reflects behaviors that are closely identical to the clinical syndrome described as obsessive compulsive disorder.
c. **Interpersonal Sensitivity**: focuses on feelings of personal inadequacy and inferiority, particularly in comparison with other people.
d. **Depression**: reflects an apparent (main, central, lively) area of the resulting data of clinical depressive syndrome.
e. **Anxiety**: includes a variety of symptoms and experiences which are commonly associated with clinically manifest high levels of anxiety.
f. **Anger-Hostility**: the consistent observation that the presence of anger and aggressive behavior serve as important determinants.
g. **Phobic Anxiety**: reflects symptoms that have been observed with high frequency in conditions defined, such as phobic disorder or agoraphobia.
h. **Paranoid Ideation**: comes from the notion that paranoid behavior is considered more like a syndrome.
i. **Psychotism**: describes a full continuum of psychotic behavior.

## Methodology

1. **Design** and **creation** of the electronic questionnaires and posted through the website http://www.cicos.gr.
2. **Collection** of the questionnaires and **preprocessing** the answers, in order to transform them in an appropriate format (raw data) for analysis.
3. **Exploration** of the dataset using tools from the field of Descriptive Statistics.
4. **Data Mining Analysis** of the dataset so as to extract useful, hidden knowledge, in the form of patterns, rules and clusters.
5. **Interpret** and **evaluate** the knowledge.

## Sample

- A total of 101 students (57.4% female, 42.6% male) were recruited from the Technological Institute of Western Greece.
- The majority of students were born in urban areas (75.9% of females and 65.1% of males).
- These rates have increased, as students have moved to urban areas for studies. Thus, the 89,1% of them live in urban areas.

## Measurement Tools

- The 9 aforementioned sub-scales of SCL-90 test is measured.
- The Data Mining analysis of the dataset is conducted using the R-project.

## Mining Association Rules

- **Association rules**: rules presenting association or correlation between itemsets. An association rule has the form of A →B, where A and B are two disjoint itemsets.
- **Goal**: studies whether the occurrence of one feature is related to the occurrence of others.
- Three most widely used measures for selecting interesting rules are:
  - ✓ **Support** is the percentage of cases in the data that contains both A and B,
  - ✓ **Confidence** is the percentage of cases containing A that also contain B, and
  - ✓ **Lift** is the ratio of confidence to the percentage of cases containing B.

## Association Rules

- By using Apriori algorithm, 85 rules are emerged (Fig. 1)
- After the remove of the redundant rules and choose the most significance rules (lift >1.2) the following 17 rules are emerged (Fig. 2)

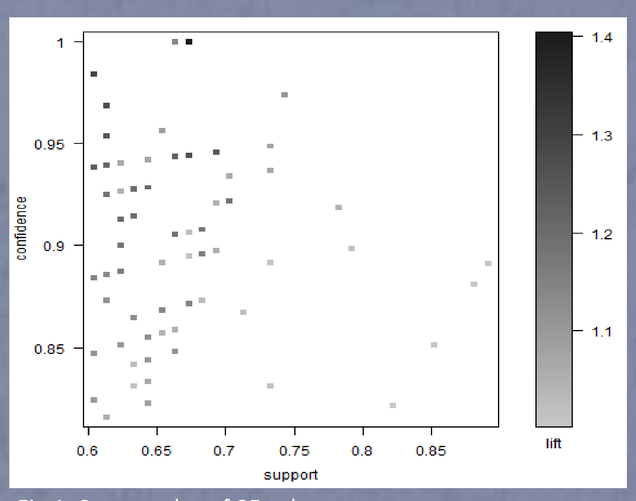


Fig 1: Scatter-plot of 85 rules

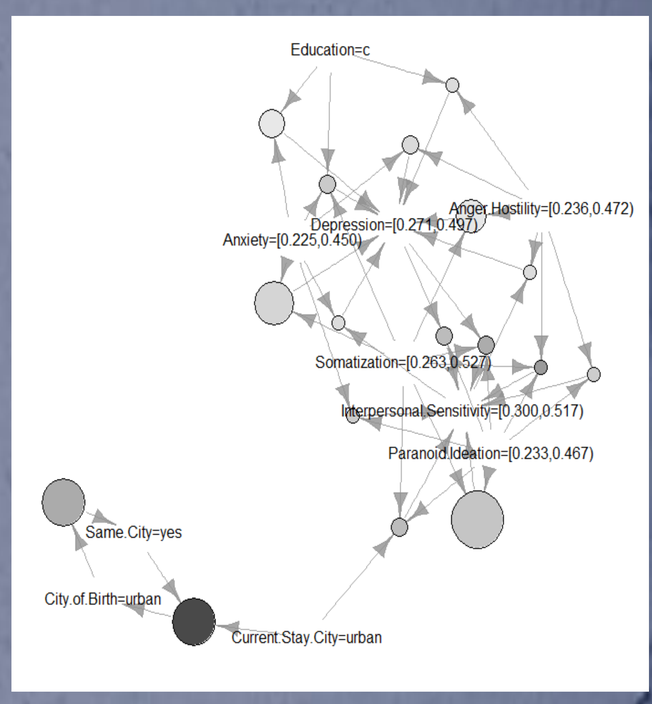


Fig 2: 17 rules with Lift > 1.2

The first six (6) most important Association Rules with Confidence ≥ 0,95 and Lift ≥ 1.2

| | lhs | | rhs | support | confidence | lift |
|---|---|---|---|---|---|---|
| 1 | {Current.Stay.City=urban, Same.City=yes} | => | {City.of.Birth=urban} | 0.6732673 | 1.0000000 | 1.402778 |
| 2 | {Somatization=[0.263,0.527), Anger.Hostility=[0.236,0.472), Paranoid.Ideation=[0.233,0.467)} | => | {Interpersonal.Sensitivity=[0.300,0.517)} | 0.6039604 | 0.9838710 | 1.290532 |
| 3 | {City.of.Birth=urban} | => | {Same.City=yes} | 0.6732673 | 0.9444444 | 1.271852 |
| 4 | {Somatization=[0.263,0.527), Depression=[0.271,0.497), Paranoid.Ideation=[0.233,0.467)} | => | {Interpersonal.Sensitivity=[0.300,0.517)} | 0.6138614 | 0.9687500 | 1.270698 |
| 5 | {Depression=[0.271,0.497), Paranoid.Ideation=[0.233,0.467)} | => | {Interpersonal.Sensitivity=[0.300,0.517)} | 0.6138614 | 0.9538462 | 1.251149 |
| 6 | {Current.Stay.City=urban, Somatization=[0.263,0.527), Paranoid.Ideation=[0.233,0.467)} | => | {Interpersonal.Sensitivity=[0.300,0.517)} | 0.6138614 | 0.9538462 | 1.251149 |

## Data Clustering

- **Goal**: ranking of respondents in clusters using clustering techniques (i.e. k-means), according to 9 clinical points (sub-scales) of the SCL-90.
- **Results**: the best clustering outcome, according to Silhouette measure, is the one that separates data into two (k=2) clusters (Fig. 3).
- Studying further the results of clustering we conclude:
  - ✓ 64% of females that belong into second cluster tend to present lower clinical points than the 36% of females that belong into first cluster. The corresponding rates for males are 55% and 45%.

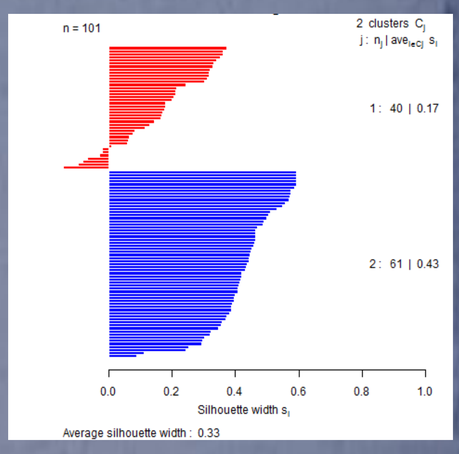


Fig 3: Silhouette for k-means with k=2

## Conclusions

The results indicate among others, that the use of Data Mining methods is an important tool to export and receive the conclusions and decisions especially in the field of psychological assessment and in neuroscience.

## References

Oded Maimon and Lior Rokach (Editors), "*Data Mining and Knowledge Discovery Handbook*", 2nd ed., Springer, 2010

Pang-Ning Tan, Michael Steinbach and Vipin Kumar, "*Introduction to Data Mining*", Addison-Wesley, 2006

Derogatis, L.R., '*The SCL-90 Manual I. Scoring, Administration, and Procedures for the SL-90.Baltimore'*, MD: John Hopkins University School of Medicine, Clinical Psychometrics Unit., 1977

***Contact:*** *Evgenia Gkintoni, e-mail: evigintoni@yahoo.gr*