


☒ Repository ☐ Web [View ALL Data Sets](#)

Dishonest Internet users Dataset Data Set

Download: [Data Folder](#), [Data Set Description](#)

Abstract: The dataset was used to test an architecture based on a trust model capable to cope with the evaluation of the trustworthiness of users interacting in pervasive environments.

Data Set Characteristics:	Multivariate	Number of Instances:	322	Area:	Computer
Attribute Characteristics:	N/A	Number of Attributes:	5	Date Donated	2018-03-20
Associated Tasks:	Classification, Clustering	Missing Values?	N/A	Number of Web Hits:	9828

Source:

Gianni D'Angelo, Ph.D.
Department of Law, Economics, Management and Quantitative Methods (DEMM)
University of Sannio
Benevento, Italy.

Email: dangelo '@' unisannio.it

Data Set Information:

In pervasive computing the interacting users are not able to obtain information about the trustworthiness of each other. Thus, unfair users can act maliciously towards others. The proposed solution enables to evaluate the trustworthiness of each user by monitoring the behavior of each other during their interaction on the network. These behaviors are represented by tuples including significant parameters. Based on these tuples, the architecture combines some artificial intelligence-based technologies to implement a decision making system.

The tuples are as follows:

$e_{ij} =$

where:

e_{ij} - i -th entity interacting with j -th entity.

EID $_j$ - j -th entity Identification

CT - Counting Trust. It is used to count how many trustworthy transactions (belonging to a specific context) occur after the last untrustworthy transaction.

CU - Counting Un-trust. It is used to count how many untrustworthy transactions (belonging to a specific context) occur after the last trustworthy transaction.

LT - Last Time. It is used to take into account of the date at which the last experience in a specific context took place.

TC - Transactions Context. It is used to identify the type of transaction, such as game, e-commerce, social network and others.

TS - Trust Score. It is the score that an entity gives to another entity at the end of each direct interaction.

The data set was obtained by a Java simulator which implemented the proposed architecture.

It includes data for the three most popular types of attack, namely:

- Counting-based attack. The user tries to gain a good reputation by alternating the honest and dishonest behavior.
- Time-based attack. User again tries to gain a good reputation by alternating the honest and dishonest behavior, but acts in different time.
- Context-based attack. R tries to gain a good reputation by acting honestly for a type of transaction and dishonestly for another one.

Because EID $_j$ parameters are not relevant for the decision-making process, only the following parameters were reported in the dataset:

- CT
- CU
- LT
- TC
- TS

Because, there could be situation in which users have not historical data (tuples) for interacting with another one, it may get data (tuples) from third-parties who previously have had interaction with the inquired user.

Nevertheless, the trustworthiness of such third party entities (recommenders) needs to be evaluated also. Indeed, they may act through attacks, such as: Ballot Stuffing (BS), Bad mouthing, and Random opinion (RO).

Changing of the TS parameter for a number of rows in the dataset, and in according to a specific attack, allows to obtain different datasets useful for the recommenders trustworthiness evaluation.

According to this, the following datasets are also provided:

- BM $_x$ %.txt x is the percentage of unfair recommendations obtained by a BM attack. It ranges from 10 to 50.
- BS $_x$ %.txt x is the percentage of unfair recommendations obtained by a BS attack. It ranges from 10 to 50.
- RO $_x$ %.txt x is the percentage of unfair recommendations obtained by a RO attack. It ranges from 10 to 50.

Attribute Information:

- 1) CT {CT_range_1, CT_range_2, CT_range_3, CT_range_4}
- 2) CU {CU_range_1, CU_range_2, CU_range_3, CU_range_4}
- 3) LT {LT_range_1, LT_range_2, LT_range_3, LT_range_4}
- 4) TC {sport, game, ECommerce, holiday}
- 5) TS {trustworthy, untrustworthy}

The numerical attributes (CT, CU, LT) was discretized.

Several of the papers listed below contain detailed descriptions of how these attributes were discretized.

Relevant Papers:

G. D'Angelo, S. Rampone, F. Palmieri, "Developing a Trust Model for Pervasive Computing Based on Apriori Association Rules Learning and Bayesian Classification", SOCO " Soft Computing Journal, Vol.21, n.21, pp. 6297-6315, 2017. DOI: 10.1007/s00500-016-2183-1

Citation Request:

If you intend to use this dataset on your research, please cite the following works:

1. G. D'Angelo, S. Rampone, F. Palmieri, "Developing a Trust Model for Pervasive Computing Based on Apriori Association Rules Learning and Bayesian Classification", SOCO " Soft Computing Journal, Vol.21, n.21, pp. 6297-6315, 2017. DOI: 10.1007/s00500-016-2183-1
2. G. D'Angelo, S. Rampone and F. Palmieri, 'An Artificial Intelligence-Based Trust Model for Pervasive Computing,' 2015 10th International Conference on P2P, Parallel, Grid, Cloud and Internet Computing (3PGCIC), Krakow, 2015, pp. 701-706. DOI: 10.1109/3PGCIC.2015.94

Supported By:



In Collaboration With:



[About](#) || [Citation Policy](#) || [Donation Policy](#) || [Contact](#) || [CML](#)