

# Knowledge extraction from TB patients- A hybrid model with clustering and decision tree

F.Firouzi Jahantigh

*Assistant Professor*

*Department of Industrial Engineering, University of Sistan and Baluchestan, Zahedan, Iran*

[firouzi@eng.usb.ac.ir](mailto:firouzi@eng.usb.ac.ir)

Hakimeh Ameri

*MSc in information technology*

*Department of Industrial Engineering, Khaje Nasir Toosi University of Technology*

*Tehran, Iran*

*Ha.amery@gmail.com*

**Abstract—** According to the World Health Organization report, 9 million people have been infected by active TB and about 1.5 to 2 million people lose their lives annually due to this disease. The study was conducted on 600 patients from Masih-e-Daneshvari tuberculosis research center. The K-Means clustering data mining algorithms and decision tree are used to perform the categorization and determining common indicators among patients. 3 clusters according to Dunn index were chosen as the optimal clusters. The cluster field added to data set and different decision trees used to find the highest accuracy. The C 5.0 tree has 97.6% accuracy. According the results of this study the most important factors identified are hemoglobin, age, sex, smoking, alcohol consumption and Creatinine. C 5.0 rules by 50% confidence are extracted.

**Keywords—** *Tuberculosis, clustering, decision trees, data mining*

## BIOGRAPHY

**Dr. Farzad Firouzi Jahantigh** is Assistant Professor in Department of Industrial Engineering, University of Sistan and Baluchestan, Zahedan, of Iran. His main research interests and lectures are in the fields of Operations Research in Medicine and Healthcare, Healthcare Systems Engineering, supply chain management, Quality Evaluation, Health Information Technology, Data analysis, Fuzzy Logic and Queuing Theory. He has published more than 20 papers in various ISI, international, and internal ElmiPajoheshi journals as well as conferences. Email: [firouzi@eng.usb.ac.ir](mailto:firouzi@eng.usb.ac.ir)

**Hakimeh Ameri** is graduated in M.S at K.N.Toosi Univerity of science and technology in information technology. Her main research interest is on bioinformatics, Data analysis and big data. She has 7 published papers in this filed. She now teaching Artificial Intelligence, data mining, Information technology, programming languages and data structure in University. Email: [ha.amery@gmail.com](mailto:ha.amery@gmail.com)