

# Data Mining Application in Classification Scheme of Human Subjects According to Ayurvedic Prakruti – Temperament

Murtaza M. Junaid Farooque<sup>1</sup>, Mohammed Aref<sup>1</sup>, Mohammed Imran Khan<sup>2</sup> and Shareque Mohammed<sup>2</sup>

<sup>1</sup>Department of MIS, Dhofar University, Salalah, Sultante of Oman; mfarooque@du.edu.om, mohammed\_aref@du.edu.om

<sup>2</sup>Department of Accounting and Finance, Dhofar University, Salalah, Sultante of Oman; Imran@du.edu.om, smohammed@du.edu.om

## Abstract

**Objective:** To identify, and apply data mining technique in classification of human subjects according to ayurvedic principles of dosha prakruti. **Method/Analysis:** The data of 67 healthy subjects of both genders has been collected using traditional prakruti questionnaire. There were 37 parameters. The collected data was analyzed using data mining techniques. Various Classification algorithms were applied and to develop a model to predict the human Prakruti. **Finding:** The classification model built and tested was found to predicting the prakruti with great deal of accuracy. This model can be integrated in health information system. Although this will not replace the human judgment but will definitely reduce the effort and time. **Applications/Improvements:** Modern technique of artificial intelligence has been applied to ancient science of Ayurveda, which can help medical informatics researcher, alternative medicine Practitioners to adopt appropriate strategy in clinical decision making and Personalized medication.

**Keywords:** Ayurveda, Classification, Data mining, Prakruti, Tridosha

## 1. Introduction

Data mining aims to discover meaningful insights and knowledge from data. Discoveries are often expressed as models, and we often describe data mining as the process of building models. A model captures, in some formulation, the essence of the discovered knowledge. A model can be used to assist in our understanding of the world. Models can also be used to make prediction or diagnosis<sup>1</sup>. The diagnosis are has significant role in medicine. Data mining offers methodological and technical solutions to deal with the analysis of medical data and construction of prediction models<sup>2</sup>.

Ayurveda is an Indian traditional system of medicine considers every individual as unique and the treatment is

based on the defined constitution types. A basic feature of the theory and practice of Ayurveda is the classification of human beings into three doshas (vata, pitta, and kapha) and their mixed constitutional types called prakriti. In Ayurveda, the fundamental concept is the tridosha theory, the three doshas i.e. Vatta, Pitta and Kapha are present in human body, their balance or harmony is health while imbalance or disharmony is disease state. The dominance of any Dosha in a person determines his or her prakruti. Prakruti is an important criterion for evaluating an individual's status of physical constitution as well as for diagnosis, epidemiology, clinical decision-making, and prognostication. The prakriti or the Ayurvedic constitutional type of an individual is fixed at conception and is articulated through physical, psychological, immunologi-

\*Author for correspondence

cal, and behavioral traits. The, Prakruti is an important criterion for evaluating an individual's status of physical constitution, which in turn becomes useful in diagnosis, clinical decision-making, and prognostication<sup>3</sup>.

<sup>4</sup>Chien and Chen conducted study in a semiconductor company to support their hiring decision for indirect labors using data mining approach. They found that the approach was practically viable. Markus Brameier and Wolfgang Banzhaf<sup>5</sup> compared the application of linear genetic programming and neural networks in medical data mining and found that genetic programming performs better in classification and generalization.

Riccardo Bellazzi and Blaz Zupan<sup>6</sup> concluded that the data mining methodological and technical solutions is useful in analysis of medical data and construction of medical prediction models. Determination of prakruti is important step in ayurvedic concept of personalized medication. Many researcher have developed different various methodologies. These include pulse detection, psychometric scaling, development of questionnaires and software like AyuSoft<sup>7</sup>. Tripathi Y.B.<sup>8</sup> Proposed free radical theory of diseases based on tri dosha theory of Ayurveda.

Alex hankey<sup>9</sup> provided scientific evidence for the proposed identification of the doshas on the basis of systems analysis and suggested that its relates to coenzyme. Joshi R.R.<sup>10</sup> showed that the concept of tridosha has a sound empirical basis that could be used for the scientific establishment of Ayurveda in a new light.

Farooque et al<sup>11</sup> adapted decision tree approach for classification of human subjects according to unani mizaj while chee et al<sup>12</sup> adopted SOM Algorithm for classification of subjects according to Sasang typology. Srideivanai N and R. M. Chandrasekaran demonstrated creation of expert clinical system for the diagnosis of the diabetic mellitus using clustering and classification techniques<sup>13</sup>. P. Suganya and C. P. Sumathi<sup>14</sup> used data mining algorithm for classification of parkinson diseases while C. Kalaiselvi and G. M. Nasira<sup>15</sup> used these technique in prediction of heart diseases and cancer.

## 2. Formal Theory of Ayurveda Prakruti

Every Individual  $I$  has 3 dosha ie. Vata ,Pitta, and Kalpha denoted by

$$I = \{ \varpi, \pi, \kappa \} \quad (1)$$

For every  $I$  there is a unique combination of  $\varpi$ ,  $\pi$  and  $\kappa$ , every individual will have a degree of membership in all three dosha and no dosha is absent therefore we say that the degree of membership is fuzzy:

$$\{\varpi\} + \{\pi\} + \{\kappa\} = 1.0 \quad (2)$$

The prakruti of an individual can be determined on the basis of dominance of the dosha. No two dosha will be equal in any case. The most dominant dosha will be the temperament of the individual. i.e. temperament (Pr) can be determined as:

$$\Pi\rho = \max(\varpi, \pi, \kappa) \quad (3)$$

$\varpi$ ,  $\pi$  and  $\kappa$ , can have any values between 0 and 1.0

Every individual  $I$  has degree of membership of all three sets, i.e.  $\varpi$ ,  $\pi$  and  $\kappa$ , the maximum belonging will determined by maximum membership.

If there is any change in the value of  $\varpi$ ,  $\pi$  or  $\kappa$ , the temperament  $\Pi\rho$  is distort to  $\Pi\rho'$ , there is condition of instability and health is distorted and it should be brought back to the stable condition.

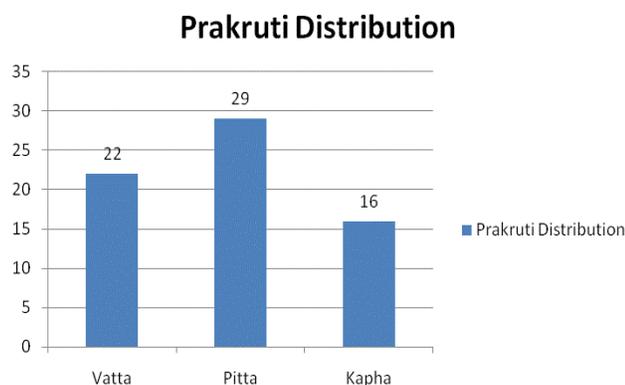
Personalized medicine can increase efficacy and safety with customized diagnosis and treatment. If identification Temperament is included in hospital Management system it could lead to better decisions, right diagnosis and treatment<sup>16</sup>.

## 3. Methodology

Data of 67 healthy subjects of both genders was collected using questionnaire and physical observation. At first, the subjects were manually classified by experienced Ayurvedic physician using 37 Parameter. The Class wise distribution was as follows

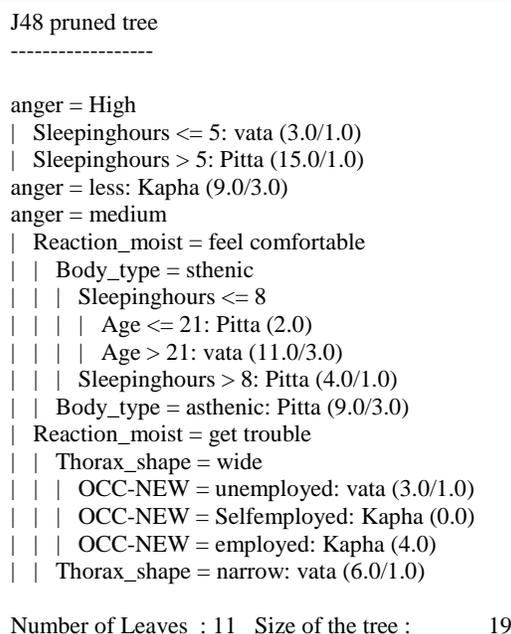
The collected data was used to generate decision tree in data mining software weka. The analysis was carried out on DELL I7 machine with windows7 operating System. The decision tree was generated using J48 algorithm as in Figure 2

As it can observed from the tree only 7 out of 37 parameter are decisive for predicting the prakruti of a subject. This Parameter are Anger, Sleeping hours, Body Type, Reaction to moisture, Age, shape of Thorax, and occupation. In the figure there are two number next to each leaf. The first number is the total number of instances (weight of instances) reaching the leaf. The second number is the number (weight) of those instances that are misclassified.



**Figure 1.** Distribution of subject according to prakruti.

Different other data mining algorithm like naive bayes classifier, logistic regression and ANN (Multi-layer perceptron was used) were applied on the same Data. The error statistic was recorded to compare with the decision tree. The result of comparison is given in Table 1.



**Figure 2.** Decision tree of classification of subject according to prakruti.

**Table 1.** Comparison of Four classification schemes

Scheme	Decision Tree	Naive Bayes	Logistic Regression	ANN (Multi-layer perceptron)
Kappa statistic	0.1511	-0.0039	0.1128	0.1232
Mean absolute error	0.3793	0.413	0.3841	0.3837
Root mean squared error	0.5477	0.563	0.6053	0.569

## 4. Results and Discussion

The widespread availability of new computational methods and tools for data analysis and predictive modeling requires medical informatics researchers and practitioners to systematically select the most appropriate strategy to cope with clinical prediction problems. Mean absolute error is highest in naive bayes classification i.e. 0.413 and lowest in decision tree i.e. 0.3793. The mean absolute error of decision tree, logistic regression and Multi-layer perceptron is almost near 0.38. The root mean squared error is highest in case of logistic regression (0.6053) and lowest in case of decision tree (0.5477). The lower value of mean absolute error and root mean square error indicated the level of accuracy of the model. The kappa statistic is lowest and negative in case of naive bayes (-0.0039) and highest in case of decision tree (0.1511).

## 5. Conclusion

In this study four different classification schemes were compared. The decision tree generated as shown in Figure 1 reflects that 8 attributes can be used decisively to classify the subjects according to Prakruti (Temperament). While comparing with other scheme i.e. Naive bayes, Neural network and logistic regression, the values of Root mean square Error, Mean Absolute Error and Kappa statistic are almost similar as depicted in Table 1. This indicates that the temperament classification is independent of the scheme applied. The only exception is in the case of naive bayes classification wherein the kappa statistic is very low and negative as compared to other methods. Henceforth, there is a great scope of research in this area to bring synthesis in the field of alternative medicine and artificial intelligence.

## 6. References

1. Tang PN, Steinbach M, Kumar V. Introduction to Data Mining. Pearson Addison: Wesley Bostan; 2005.
2. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: Current issues and guidelines. International Journal of medical informatics. 2008; 77(2):81–97.
3. Rotti H, Raval R, Bellampalli R, et al. Determinants of Prakriti, the Human Constitution Types of Indian Traditional Medicine and its Correlation with Contemporary Science. Journal of Ayurveda Integrative Medicine. 2014; 5(3):167–75.
4. Chien CF, Chen LF. Data mining to improve personnel selection and enhance human capital: A case study in high-technology industry. Expert Systems with Applications. 2008; 34(1):280–90.
5. Brameier M, Banzhaf W. A Comparison of Linear Genetic Programming and Neural Networks in Medical Data Mining. IEEE Transactions on Evolutionary Computation. 2001; 5(1):17–26.
6. Bellazzi R, Zupan B. Predictive data mining in clinical medicine: Current issues and guidelines. International Journal of medical informatics. 2008; 77(2):81–97.
7. Rotti H, Raval R, Bellampalli R, et al. Determinants of Prakriti, the Human Constitution Types of Indian Traditional Medicine and its Correlation with Contemporary Science. Journal of Ayurveda Integrative Medicine. 2014; 5(3):167–75.
8. Tripathi YB. Molecular approach to ayurveda. Indian Journal of Experimental Biology. 2000; 38(5): 409–14.
9. Hankey A. A test of the systems analysis underlying the scientific theory of Ayurveda's Tridosha. J. Altern Complement Medicine. 2005; 11(3): 385–90.
10. Joshi RR. A Biostatistical Approach to Ayurveda: Quantifying the Tridosha. Journal of Alternative and Complement Medicine. 2004; 10(5):879–89.
11. Murtaza M, Junaid Farooque, Sayyed Abidurrahman, and Farhana Sarkhawas. Identification of Mizaj (Temprament) based on Tibbi Fundamentals using Classification as Tool. Proceeding of National conference on data mining (NCDM2011). 2011.p. 95–97.
12. Chee H, Kim BC, Kim Y, Eom K. Sasang Typology classification with data reduction and SOM Algorithm. IEEE 13<sup>th</sup> international conference on e-health networking, applications and services; Columbia. 2011. p. 371–74
13. Nagarajan S, Chandrasekaran RM. Design and Implementation of Expert Clinical System for Diagnosing Diabetes Using Data Mining Techniques. Indian Journal of Science and Technology. 2015
14. Suganya P, Sumathi CP. A Novel Metaheuristic Data Mining Algorithm for the Detection and Classification of Parkinson Disease. Indian Journal of Science and Technology. 2015; 8(14):1–9.
15. Kalaiselvi C, Nasira GM. Prediction of Heart Diseases and Cancer in Diabetic Patients Using Data Mining Techniques. Indian Journal of Science and Technology. 2015; 8(14):1–7.
16. Afrough S, Murtaza M, Junaid Farooque. Integrating the process of Determining Mizaj (Temprament) in a Hospital Management System. Allana Management Journal of Research. 2012; 11(2):82–87.