

CLASSIFICATION OF DOCTORS BEHAVIOR  
ON PRESCRIBING MEDICINES USING  
DATA MINING TECHNIQUES



# CLASSIFICATION OF DOCTORS BEHAVIOR ON PRESCRIBING MEDICINES USING DATA MINING TECHNIQUES

---

Deepika Singh<sup>1</sup>, Chitra Garg<sup>2</sup>,

## ABSTRACT

This paper explores data mining technique in health care. It analyses data mining task on the database of some available doctors in Jaipur city. The database shows preference of different set of doctors who perceives and give importance to the effectiveness of drug while prescribing it to the patient. In this paper, classification method analyses the result of effectiveness of drug factor using data mining tool "WEKA".

**KEYWORDS:** Data Mining, Classification, Zero-R, WEKA, KDD process

## INTRODUCTION

Data mining is a process of analyzing data from the databases and extracting hidden information from large data sets and thus summarizing into useful information. It is a tool for analysis of data from large databases and can predict future trends and behavior, allowing business to make proactive, knowledge driven decisions. Data is a raw fact and figures which can exist in variety of forms whereas information is processed data which is valuable as it can affect behavior, a decision or an outcome of the data. Information includes patterns, associations, and relationships of data and knowledge is processed information. Data mining is an advanced analysis step of Knowledge Discovery Process (KDD). Knowledge discovery process is defined with stages: selection, preprocessing, transformation, data mining, interpretation or evaluation. Data mining is mainly implemented on hardware and software platforms. Initially there are large databases stored at different places and analyzing those data becomes a tedious task. Then this evolution of data mining has taken place which can analyze these business data using various data mining techniques.

## LITERATURE SURVEY OF THE PROBLEM

Various discussions had taken place on the type of medicines or drugs which doctors provide to the patients. Quality of drug recommended by the doctor is one of the important factors for the treatment and cure of any diseases. A survey of various demographic profile of the doctors of Jaipur were studied who were asked to rate the importance attached to the effectiveness of drugs before prescribing drugs to the patient. The Likert scale was divided into five parameters ranging from highly important to not at all important (ratings: highly important-5, important-4, neutral-3, not very important-2, not at all important-1). This paper focuses on categorization of doctors who gives weightage to this factor.

## WEKA AS A DATA MINER TOOL

In this paper we have used WEKA, a data mining tool for classification. Weka stands for Waikato environment for knowledge learning, is a computer program that was developed at the University of Waikato in New Zealand. This application is completely written in Java and is compatible with almost every computing platform. Weka contains data mining tools for data pre-processing, classification, association, clustering, regression, association rules and visualization and provide outstanding results.

<sup>1,2</sup>M.Tech Scholar, Banasthali University, Rajasthan

● <sup>1</sup>dsdeepikasingh2@gmail.com

Data set can be loaded in Weka in ARFF (attribute relation file format) format. We can also convert CSV (comma delimited) format into ARFF format.

## CLASSIFICATION IN WEKA

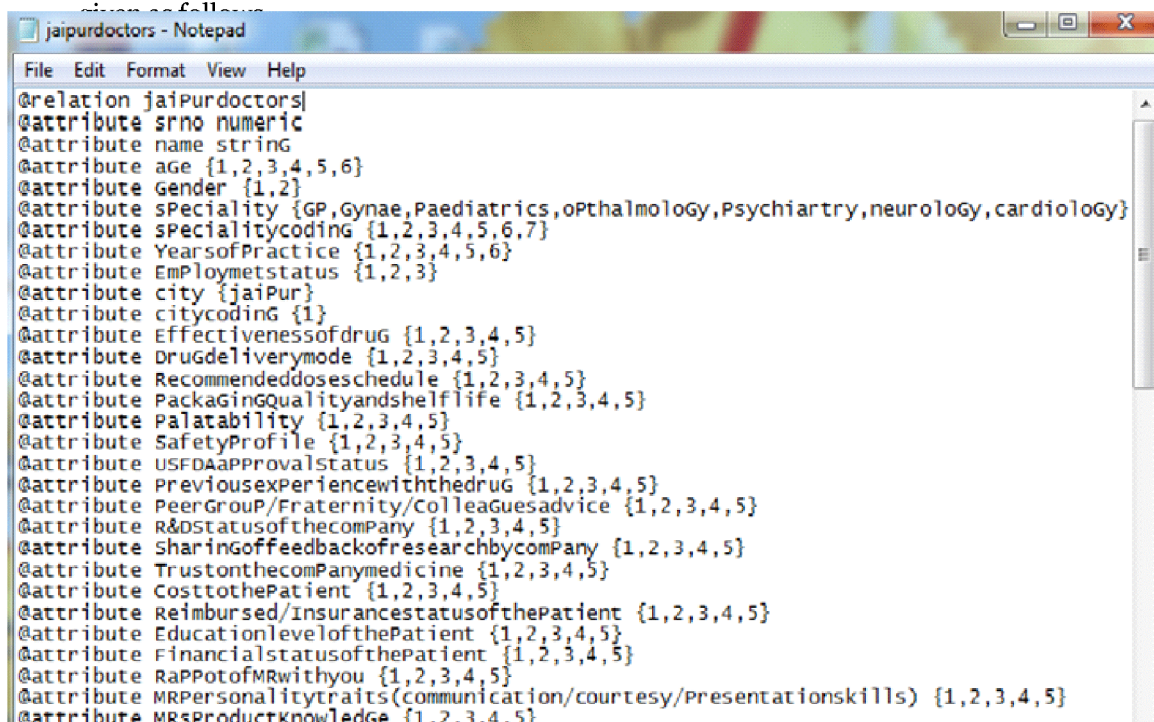
Classification is a data mining that categories item from a collection to target classes. The goal of classification is to accurately predict the target class for each case in the data. WEKA Explorer is used to perform data classification using the following classification methods with default parameters: 1) ZeroR; 2) Naïve Bayes Simple 3) Id3; and 4) J48. For each method on every data set, use the following evaluation methods (“Test options” in the “Classify” window of the WEKA Explorer): a) “Use training set”; b) “Cross validation” with 10 folds; and c) “Percentage split” set to 66%. Classification is a data mining method that categories items from a collection to target classes. The goal of classification is to accurately predict the target class for each case in the data.

## SIMPLE CLASSIFIER: ZERO-R

- It determine the most common class or the median(in the case of numeric value).
- This algorithm tests wellness of the class predicted without considering other attribute.
- It can be used as a lower bound on performance

## EXPERIMENTS AND RESULTS

- (a) Use data records as ARFF file format by converting from CSV format. jaipurdoctors.arff file is given as follows



```
@relation jaipurdoctors
@attribute srno numeric
@attribute name string
@attribute age {1,2,3,4,5,6}
@attribute gender {1,2}
@attribute speciality {GP,Gynae,Paediatrics,oPhthalmology,Psychiartry,neurology,cardiology}
@attribute specialitycoding {1,2,3,4,5,6,7}
@attribute YearsofPractice {1,2,3,4,5,6}
@attribute EmPloymetstatus {1,2,3}
@attribute city {jaiPur}
@attribute citycoding {1}
@attribute Effectivenessofdrug {1,2,3,4,5}
@attribute Drugdeliverymode {1,2,3,4,5}
@attribute Recommendeddoseschedule {1,2,3,4,5}
@attribute PackaginGQualityandshelflife {1,2,3,4,5}
@attribute Palatability {1,2,3,4,5}
@attribute SafetyProfile {1,2,3,4,5}
@attribute USFDAApprovalstatus {1,2,3,4,5}
@attribute Previousexperiencewiththedrug {1,2,3,4,5}
@attribute PeerGroup/Fraternity/Colleaguesadvice {1,2,3,4,5}
@attribute R&Dstatusofthecompany {1,2,3,4,5}
@attribute Sharingoffeedbackofresearchbycompany {1,2,3,4,5}
@attribute Trustonthecompanymedicine {1,2,3,4,5}
@attribute CosttothePatient {1,2,3,4,5}
@attribute Reimbursed/InsurancestatusofthePatient {1,2,3,4,5}
@attribute EducationlevelofthePatient {1,2,3,4,5}
@attribute FinancialstatusofthePatient {1,2,3,4,5}
@attribute RaPPotofMRwithyou {1,2,3,4,5}
@attribute MRPersonalitytraits(Communication/courtesy/Presentationskills) {1,2,3,4,5}
@attribute MRProductKnowledge {1,2,3,4,5}
```

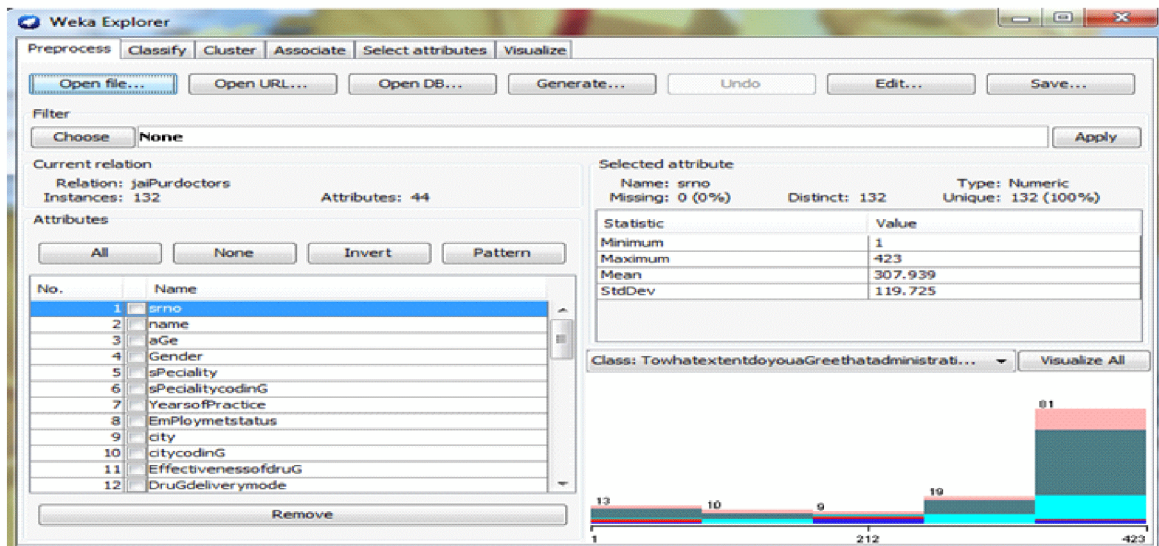
**Fig: 1 Attributes of doctors in jaipurdoctors.arff file**

```

@attribute FinancialstatusofthePatient {1,2,3,4,5}
@attribute RaPPotofMRwithyou {1,2,3,4,5}
@attribute MRPersonalitytraits(communication/courtesy/Presentationskills) {1,2,3,4,5}
@attribute MRsProductKnowledge {1,2,3,4,5}
@attribute PhysicianssamplesProvidedbyMR {1,2,3,4,5}
@attribute Distributionofbrouchersandleaflets {1,2,3,4,5}
@attribute AcademicSPonsorshIpservices {1,2,3,4,5}
@attribute Availabilityofthedruginthemarket {1,2,3,4,5}
@attribute InformationreceivedthroughM.R.s {1,2,3,4,5}
@attribute ThrouGhScientificJournals {1,2,3,4,5}
@attribute ThrouGhScientificMedicalSocieties {1,2,3,4,5}
@attribute ThrouGhInternet {1,2,3,4,5}
@attribute ThrouGhColleagues {1,2,3,4,5}
@attribute HowoftendoyouPrescribeadruGwhIchomesunderDrugsPriceControlbyNationalDrugPriceA
@attribute HowoftendoyouPrescribeanOTCdruG? {1,2,3,4,5}
@attribute HowoftendoyouPrescribeabrandrequestedbythePatient? {1,2,3,4,5}
@attribute TowhatextentdoyouaGreetthatsomeofthedoctorsareinfluencedbythePersonalised/custom
@attribute TowhatextentdoyouaGreetthatgenericdrugsareequallyeffectiveandhavesamequalityasco
@attribute TowhatextentdoyouaGreetthatadministrationshouldnotrestrictavailabilityofbrandsyou

@data
1."tr_joshi",6,1,Gynae,2,6,1,jaiPur,1,5,4,5,4,4,5,4,4,3,4,3,3,4,4,4,4,4,4,2,2,2,3,4,4,4,3,
2,"surendra",5,1,GP,1,5,1,jaiPur,1,5,5,5,5,5,5,4,4,3,5,4,4,4,4,4,2,2,4,2,2,4,4,2,5,5,5,5
3,"Pc_kaPri",5,1,Gynae,2,5,1,jaiPur,1,5,5,5,5,5,5,5,4,4,5,4,4,4,3,4,4,2,2,4,2,2,2,2,4,5,5,
4,"m",4,2,Gynae,2,4,1,jaiPur,1,5,5,5,5,3,5,4,3,1,3,3,2,2,3,2,2,2,1,3,1,5,1,4,3,5,5,5,3,4,2
12,"rohit_bansal",1,1,GP,1,1,2,jaiPur,1,5,4,4,5,4,5,5,4,3,4,4,4,4,5,5,5,3,4,4,3,3,4,5,4,5,
13,"vinod",4,1,GP,1,3,1,jaiPur,1,5,5,5,4,4,4,5,5,4,3,5,4,5,2,5,5,3,2,2,2,2,5,5,4,5,5,5,4,5
14,"dayal",5,1,OPhtalmology,4,4,2,jaiPur,1,5,4,4,4,5,5,4,4,3,4,4,4,4,4,4,4,2,2,2,2,2,4,4,4,4
15,"sk",5,1,Paediatrics,3,5,1,jaiPur,1,5,4,5,4,4,4,4,4,3,4,4,4,5,4,4,4,2,2,2,2,2,4,4,4,4,4
    
```

**Fig: 2 Data of doctors in jaipurdoctors.arff file**



**Fig: 3 Loading of Data into WEKA**

(b) Load the jaipurdoctors.arff file into WEKA

- (c) Select the classification tab in WEKA explorer and with the help of ZeroR (Majority Predictor) algorithm following output will be

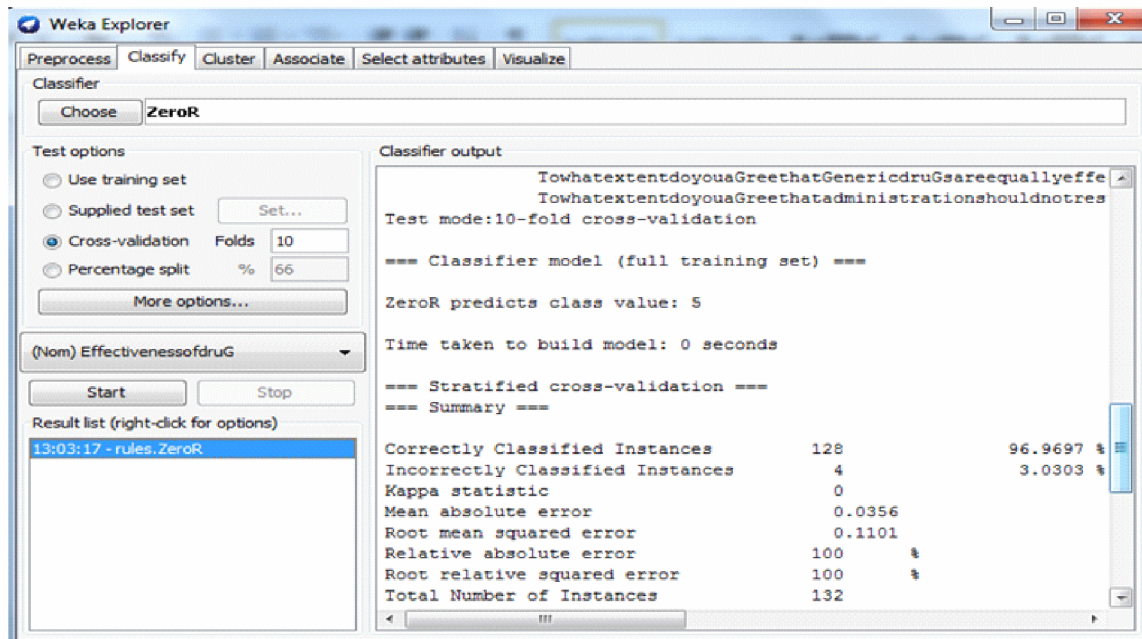


Fig: 4 (a) Classification Output of WEKA

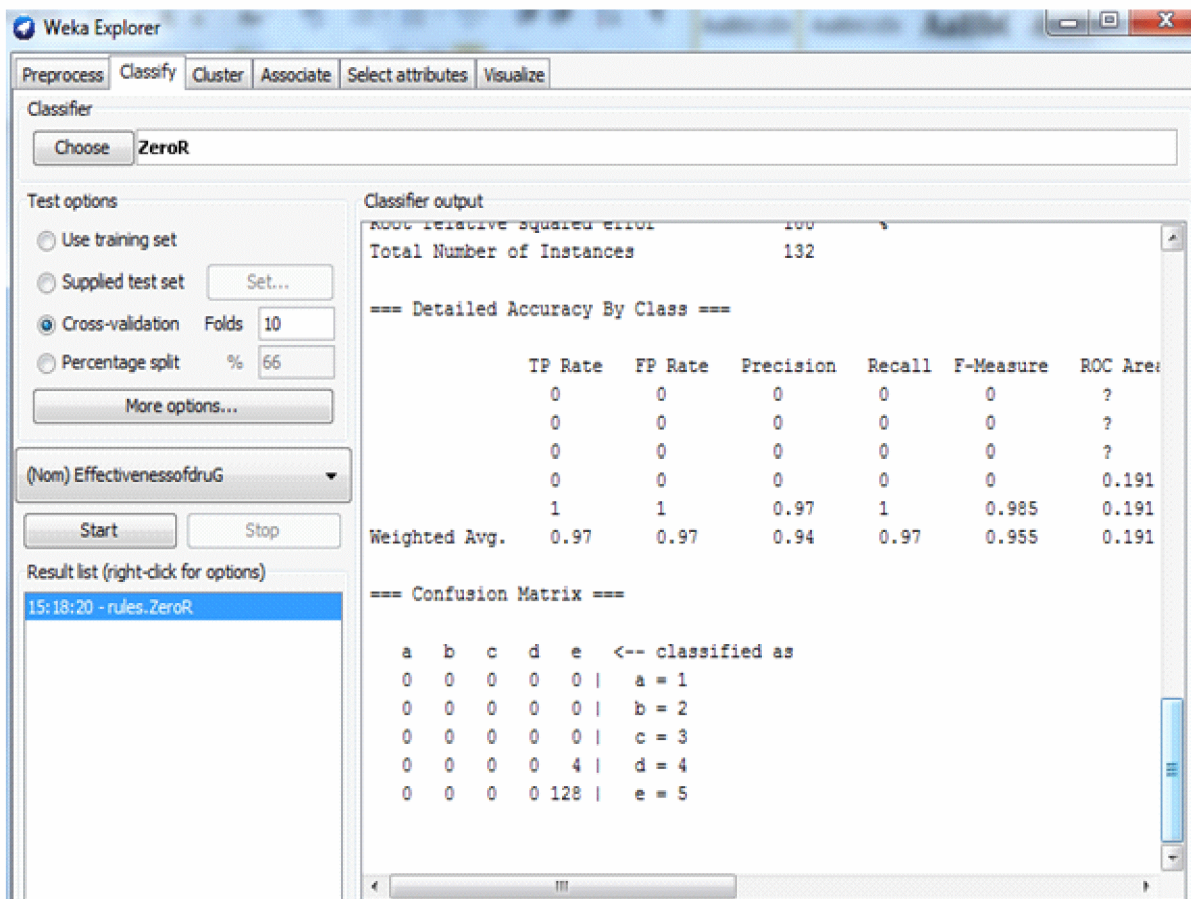


Fig: 4(b) Classification Output

**(d) Analysis of Results**

We employed four performance measures: Precision, Recall Measure and ROC Space, F-Measure. Confusion matrix is obtained to calculate the four measures.

These are calculated as follows:

$$\text{Recall} = \text{TP} / (\text{TP} + \text{FN})$$

$$\text{Precision} = \text{TP} / (\text{TP} + \text{FP})$$

$$\text{F-Measure} = (2 * \text{TP}) / (2 * \text{TP} + \text{FP} + \text{FN})$$

TP: cell denoting number of samples classified as true while they are true

TN: cell denoting number of samples classified as false while they are false

FN: number of samples classified as false while they are true

FP: number of samples classified as true while they were actually false

In this analysis precision rate is 0.94 and correctly classified instances are 96.97%.

**CONCLUSION**

In this paper, we have discussed Data Mining classification method which classifies different sets of doctors who gives importance to effectiveness of drugs prescribed to the patients. In this analysis, it has been observed that precision rate is high and high percentage of correctly classified instances. Thus this parameter is highly important for doctors while prescribing drug.

**ACKNOWLEDGEMENTS**

We express our gratitude and deep regards to faculty member, Dr. Neeti Kasliwal, Assistant professor, FMS. Wisdom, Banasthali University, Dr. S.Taruna, Associate Professor, Rajasthan for their exemplary guidance, monitoring, guidance and constant encouragement throughout this paper.

**REFERENCES**

- Balasubramanian, T., & Umarani, R. (2012, March). An analysis on the impact of fluoride in human health (dental) using clustering data mining technique. In *Pattern Recognition, Informatics and Medical*
- Baradwaj, B. K., & Pal, S. (2012). Mining educational data to analyze students' performance. *arXiv preprint arXiv:1201.341*
- Dunham, M. H. (2006). *Data mining: Introductory and advanced topics*. Pearson Education India. *Engineering (PRIME), 2012 International Conference on* (pp. 370-375). IEEE.
- Gangwar, A., & Bhardwaj, M. (2012). An overview: Peak to average power ratio in OFDM system & its effect. *International Journal of Communication and Computer Technologies*, 1(2), 22-25
- Gupta, G. K. (2014). *Introduction to data mining with case studies*. PHI Learning Pvt. Ltd.
- JIAWEI, H., MICHELINE, K., & DATA, M. (2007). CONCEPTS AND TECHNIQUES.
- Madeira, M., & Joshi, A. (2013, September). Analyzing close friend interactions in social media. In *Social Computing (SocialCom), 2013 International Conference on* (pp. 932-935). IEEE.