

# Astrological Prediction for Profession Doctor using Classification Techniques of Artificial Intelligence

Neelam Chaplot  
Department of Computer  
Science & Engineering  
Banasthali University  
Banasthali, Rajasthan, India

Praveen Dhyani  
Executive Director  
Banasthali University  
Banasthali, Rajasthan, India

O. P. Rishi  
Department of Computer  
Science & Informatics  
University of Kota  
Kota, Rajasthan, India

## ABSTRACT

Astrology can be an interesting example of the application of various classification techniques of artificial intelligence. In astrology, predictions about different aspects of human life are done based on the planetary position of the stars at the time of birth of a person. In this research work, the positions of the planets and stars at the time of the birth of a person are utilized. This information is used to predict the possibility of person to become doctor.

Total 102 records were collected for the study out of that half of the records were of persons that were doctor and other half records of the persons that were not doctor by profession. Thereafter, various supervised classification techniques such as Logistic, Naïve Bayes, Simple Cart, Decision Stump, Decision Table and DTNB algorithm were used and results were compared for their accuracy.

## General Terms

Artificial Intelligence, Astrology, Classification Technique.

## Keywords

Artificial Intelligence, Astrological Prediction, Classification Techniques, Prediction of Profession Doctor, Weka.

## 1. INTRODUCTION

Astrology is the area that generates lots of curiosity among human beings. To find scientific theories and its relations with astrology, changes are to be done in the way astrology is looked upon. To find scientific basis to astrology changes are to be made from the traditional ways in which astrology is analyzed. In this paper, efforts have been made to identify the scientific basis using various classification techniques of artificial intelligence. These techniques are used to perform classification of data automatically. Various classification techniques used here have resemblance with the way astrology works.

First step for prediction involves performing various classification techniques of artificial intelligence that are treated as black box for already existing methods which provided good results in various fields for classifying data. Next step involves finding out the attributes that helps in improving the dynamic prediction in astrology and then updating the knowledge of the system when more records come in picture. The results in the form of rules will be generated by the classification techniques. These rules generated by classification techniques will be backtracked and verified against the rules used by astrologers. Resemblance between these rules generated by classification techniques and rules provided by astrology will provide the scientific basis for astrology.

## 2. LITERATURE REVIEW

Influences of planetary influences are plainly visible such as day and night, change of seasons and tidal effects. Planets and stars surely have influence on everyday life of humans. Astrology predicts the course of human life based on the planetary position of the stars at the time of birth of the person. There must be some substance in astrology for it has managed to survive through generations. A lot of literature is available on astrology but the scientific validity could not be proved. Although after 15th century, lot of research has been done to establish a scientific validity for astrology.

Some researchers believe that astrology is not scientific [2][3] while other feel that in depth study in this field is required to reach that conclusion[4]. There are beliefs that predictive and non-predictive are two parts of astrology and that predictive astrology is the proper subject for testing whether astrology can be used to make prediction[5].

Supervised classification techniques in artificial intelligence[6] is recently developed branch which performs well for various classification and prediction tasks. It performs well for very vague and uncertain data also. These techniques have been used for various applications such as weather forecasting[7], cancer analysis and diagnostics[8], machine translation[9] etc. Brief review of various machine learning methods and the applications is available in paper[10].

By applying different learning techniques on large amount of data available for the planetary position of the persons and their biographies, a system which can predict on various aspects of human life can be developed.

Hardly any work has been done till now in predicting the incidences of human life based on planetary positions using artificial intelligence techniques. Some preliminary work using Case Based Reasoning and nearest neighbor algorithm has been done in literature [11][12].

## 3. CLASSIFICATION TECHNIQUES

Classification techniques of artificial intelligence helps to create generalized hypothesis from a set of data so that prediction for the future can be made based on the hypothesis generated. These techniques involve developing algorithm based on prior data in which class labels are known beforehand. Algorithms are typically developed on a training dataset and then tested on an independent test data set to evaluate the accuracy of algorithms.

It is very difficult to identify the suitable classification techniques for given data because some techniques perform well on some kind of dataset but vice versa on others. This

study investigates six classification techniques. Logistic Regression, Simple Cart, Naïve Bayes Decision Stump, Decision Table and DTNB. The selection was based on the difference in the techniques for learning and level of complexity.

### 3.1 Logistic Regression

Logistic Regression[13] is a probabilistic statistical classification model. It is used to predict the class label based on one or more predictor variables, these variables are not continuous.

The formula can be stated as the probability of occurrence of an event  $Y = 1$

$$\ln\left(\frac{p}{1-p}\right) = B_0 + B_1X$$

$\ln$  is natural logarithm and  $B_0 + B_1X$  is line regression equation

Expected probability that  $Y = 1$  for a given value of  $X$ .

$$p = \frac{e^{(B_0+B_1X)}}{1 + e^{(B_0+B_1X)}}$$

### 3.2 Simple Cart Algorithm

It is one of the decision tree algorithms. It does the attribute selection, based on implementing minimal cost-complexity pruning. It constructs binary search tree. For more information see [14].

CART stands for classification and regression trees. Cart can be applied as classification tree or as a regressive tree based on whether response variable is categorical or continuous. It constructs binary trees hence each internal node has exactly two outgoing edges. The splits are selected by identifying that which variable should be used for splitting and determining precise rules for splitting[3]

### 3.3 Naïve Bayes Classification Algorithm

Naïve Bayes classifier is a classification algorithm based on probabilities. It is based on the concept of independence of the variable that is a Naive Bayes classifier assumes that the presence or absence of features of a class is unrelated to the presence or absence of any other features.

Naive Bayes classifiers can be trained very efficiently in a supervised learning environment. In many practical applications, parameter estimation for Naive Bayes models uses the method of maximum likelihood. In other words, one can work with the Naive Bayes model without believing in Bayesian probability or using any Bayesian methods.

Let there be 'm' Classes:  $C_1, C_2, C_3 \dots C_m$

Naïve assumes that class is conditionally independent [15]

$$P(X|C_i) = \prod_{k=1}^n P(X_k|C_i)$$

$$P(X|C_i) = P(X_1|C_i) * P(X_2|C_i) * \dots * P(X_n|C_i)$$

### 3.4 Decision Table Algorithm

Decision Table consist of two major parts [16]. First is schema i.e. lists of attributes and second is the body i.e. labeled instances from the space described by the features in

schema. Decision table classifier is based conceptually on a simple lookup table. If no matching instances are found than decision table returns the majority class of the training set otherwise majority class of all matching instances is returned. To build a Decision Table Majority (DTM), induction algorithm must decide which features to include in the schema and which instances to store in the body [17].

### 3.5 Decision Stump Algorithm

Decision Stump [18] is one-level decision tree. It is a decision tree with one root node connected to leaf node. A decision stump makes a prediction based on the value of just a single input feature hence are also called 1-rules.

Based on input features, different types of trees are created. For nominal features, a decision tree which contains a leaf for each possible feature value is constructed or a decision tree with two leafs is created. One leaf corresponds to some chosen category and the other leaf for all other categories. For continuous features, usually, some threshold feature value is selected, and the decision tree containing two leafs are created for values below and above the threshold.

### 3.6 DTNB Algorithm

DTNB algorithm [19] is a simple Bayesian network in which decision table represents a conditional probability table. Learning algorithm splits the attributes into two disjoint subsets, one for decision table and the other for Naive Bayes. Forward selection method is used, at each step selected attributes are modeled by Naïve Bayes and remainder by decision table and all attributes are modeled by decision table initially.

The class probability estimates of the Decision Table and Naïve Bayes must be combined to generate overall class probability estimates. Assuming  $X_T$  is the set of attributes in the Decision Table and  $X^{\perp}$  the one in Naive Bayes, the overall class probability is computed as

$$Q(Y/X) = \alpha \times Q_{DT(Y/X^T)} \times Q_{NB(Y/X^{\perp})}/Q(Y)$$

Where  $Q_{DT(Y/X^T)}$  and  $Q_{NB(Y/X^{\perp})}$  are the class probability estimates obtained from the Decision Table and Naive Bayes respectively,  $\alpha$  is a normalization constant and  $Q(Y)$  is prior probability of class.

## 4. METHDOLOGY

Data of 102 persons were collected out of that 51 records were of doctor and 51 records were of persons those are not doctor by profession. Data were collected from reliable local persons on paper and through emails. These persons were from a certain geographical area of India. The information collected was Date of Birth, Time of Birth, Place of Birth, Profession and brief biography of the person.

Data relevant to the analysis was generated and retrieved by finding the planetary positions of zodiac and planets based on the details of place of birth, date of birth and time of birth of the person. Then these records were stored in tabular form for future use.

The attributes used by classification techniques of artificial intelligence are as follows.

**Table 1. Attributes used for performing classification.**

| S.No. | Attribute | Type    | Value              |
|-------|-----------|---------|--------------------|
| 1     | Aries     | Nominal | Value from 1 to 12 |

|    |             |         |                       |
|----|-------------|---------|-----------------------|
| 2  | Taurus      | Nominal | Value from 1 to 12    |
| 3  | Gemini      | Nominal | Value from 1 to 12    |
| 4  | Cancer      | Nominal | Value from 1 to 12    |
| 5  | Leo         | Nominal | Value from 1 to 12    |
| 6  | Virgo       | Nominal | Value from 1 to 12    |
| 7  | Libra       | Nominal | Value from 1 to 12    |
| 8  | Scorpio     | Nominal | Value from 1 to 12    |
| 9  | Sagittarius | Nominal | Value from 1 to 12    |
| 10 | Capricorn   | Nominal | Value from 1 to 12    |
| 11 | Aquarius    | Nominal | Value from 1 to 12    |
| 12 | Pisces      | Nominal | Value from 1 to 12    |
| 13 | Sun         | Nominal | Value from 1 to 12    |
| 14 | Moon        | Nominal | Value from 1 to 12    |
| 15 | Mars        | Nominal | Value from 1 to 12    |
| 16 | Mercury     | Nominal | Value from 1 to 12    |
| 17 | Venus       | Nominal | Value from 1 to 12    |
| 18 | Jupiter     | Nominal | Value from 1 to 12    |
| 19 | Saturn      | Nominal | Value from 1 to 12    |
| 20 | Rahu        | Nominal | Value from 1 to 12    |
| 21 | Ketu        | Nominal | Value from 1 to 12    |
| 22 | Gender      | Nominal | Male or Female        |
| 23 | Class       | Nominal | Doctor and Non Doctor |

## 5. EXPERIMENTS AND EVALUATIONS

Data created in the above process was loaded in the Weka tool by creating the file in ARRF format. Various classification

techniques such as Simple Cart, Logistic, Naïve Bayes, Decision Stump, Decision Table and DTNB were used on the dataset with 10, 12, and 14 fold cross validation. The results of various techniques are shown through graph in Fig. 1 for correctly and incorrectly classified data for identification of profession doctor of a person.

Better results were produced by Simple Logistic with 12 fold cross validation with an accuracy of 54.902 %. Decision Stump algorithm with 14 fold classification gave results with an accuracy of 50%.

## 6. CONCLUSION AND FUTURE WORK

With limited number of sample cases and different training methods the prediction that person will become a doctor was made. The methods that gave better results are Simple Logistic with 12 fold cross validation with an accuracy of 54.902%. Decision Stump algorithm with 14 fold classification gave results with an accuracy of 50%.

Although, the results are not as per expectations but are not discouraging also. Few methods are performing better than others so if remaining methods are tried or hybrid methods are created than accuracy of the system can be increased. Also data set consisted of only 102 records so by increasing the number of cases the accuracy of the results are expected to increase further. Results generated by working on different classifiers can be used to analyze and identify the weakness and strength of different classification methods so that new classifiers can be developed which utilizes the strength of various classifiers and thus in turn increases the accuracy in prediction. For the accuracy purpose, other aspects of person information like knowledge level, family background etc. must be included as attributes along with birth charts, astrological charts, and planetary positions in different houses in horoscope. Few more attributes like strength of the planets, aspect ratio of the planets and interrelationship among planets can also be included for increasing the accuracy of results.

Predictions are performed to find that a person will become doctor or not similarly predictions can be performed to identify basic nature of the person, the attitude of the person, money and other aspect of person's life.

## 7. ACKNOWLEDGMENTS

Research work of this scale is not possible without direct and indirect support of many persons. We would like to thank them all. Also, we extend our special thanks to the persons who have provided us with personal details to carry forward our research. Lastly, we thank family and friends to provide unconditional moral support to carry forward the research.

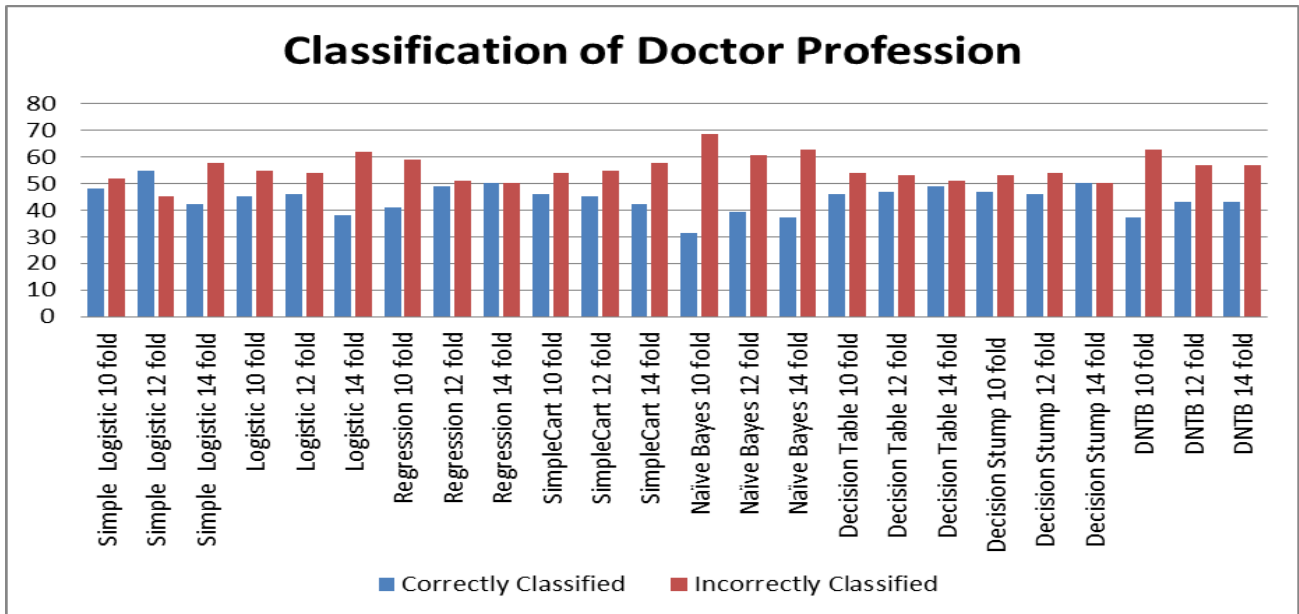


Fig 1: Graph for Classification of Profession Doctor using various Supervised Classification Techniques

## 8. REFERENCES

- [1] Hall, M., Frank, E., Holmes G., Pfahringer, B., Reutemann, P., & Witten, I. H. 2009. The WEKA data mining software. ACM SIGKDD Explorations Newsletter.
- [2] Ivan W. Kelly. 1997. A Concept of Modern Astrology a Critique. Article in Psychological Reports.
- [3] John H. Mcgrew and Richard M. Mcfall. 1990. A Scientific Inquiry Into the Validity of Astrology. Journal of Scientific Exploration.
- [4] Ken McRitchie. August 2011. Support for Astrology from the Carlson Double-blind Experiment. ISAR International Astrologer.
- [5] Penny Seator. 2008/2009 Astrological Prediction and Statistical Tests. The British Astrological Association.
- [6] S. B. Kotsiantis, I. D. Zaharakis, P. E. Pintelas. 2007. Machine learning: a review of classification and combining techniques. Springer Science Business Media
- [7] B Folorunsho Olaiya, Adesesan Barnabas Adeyemo. 2012. Application of Data Mining Techniques in Weather Prediction and Climate Change Studies. I. J. Information Engineering and Electronic Business.
- [8] Joseph A. Cruz and David S. Wishart. 2006. Applications of Machine Learning in cancer prediction and prognosis cancer. Informatica
- [9] Harold Somers. June 1999. Review Article: Example-based Machine Translation. Machine Translation.
- [10] Neelam Chaplot, Praveen Dhyani, O.P.Rishi. March 2013. A Review on Machine Learning Concepts for Prediction Based Application, International Journal of Computational Science, Engineering & Technology.
- [11] O.P. Rishi and Neelam Chaplot. Dec 2010. Predictive role of case based reasoning for astrological predictions about profession: System modeling approach, International Conference on Communication and Computational Intelligence.
- [12] O.P. Rishi and Neelam Chaplot. Dec 2010. Archetype of astrological prediction system about profession of any persons' using case based reasoning, in International Conference on Communication and Computational Intelligence.
- [13] le Cessie S., Van Houw ngen J.C. 1992. Ridge Estimators in Logistic Regression. Applied Statistics.
- [14] Leo Breiman, Jerome H. Friedman, Richard A. Olshen, Charles J. Stone. 1984. Classification and Regression Trees. Wadsworth International Group, Belmont, California.
- [15] Harry Zhang, Jiang Su. 2004. Naive Bayesian Classifiers for Ranking. Lecture Notes in Computer Science Volume 3201.
- [16] Ron Kohavi and Daniel Sommer.1998. Targeting Business Users with Decision Table Classifiers. KDD-98 Proceedings American Association of Artificial Intelligence.
- [17] Kohavi R, 1995. The Power of Decision Tables. In Proceeding European Conference on Machine Learning.
- [18] Iba, Wayne and Langley, Pat. July 1992. Induction of One-Level Decision Trees. ML92: Proceedings of the Ninth International Conference on Machine Learning, Aberdeen, Scotland, San Francisco, CA, Morgan Kaufmann.
- [19] Mark Hall, Eibe Frank. 2008. Combining Naive Bayes and Decision Tables. In Proceedings of the 21st Florida Artificial Intelligence Society Conference.