# Scientific Approach of Prediction for Professions Using Machine Learning Classification Techniques

**Snehlata Barde***
MATS School of Information Technology, MATS University Raipur, Chhattisgarh, 492004, India
E-mail: drsnehlata@matsuniversity.ac.in
ORCID iD: https://orcid.org/0000-0002-9167-5235
*Corresponding Author

**Sangeeta Tiwari**
MATS School of Information Technology, MATS University Raipur, Chhattisgarh, 492004, India
Email: sangeetatiwari4321@gmail.com
ORCID iD: https://orcid.org/0000-0002-4671-2391

**Brijesh Patel**
Department of Mechanical Engineering, National Taiwan University of Science and Technology, Taipei, Taiwan, Republic of China
Email: aero.brijesh@gmail.com
ORCID iD: https://orcid.org/0000-0003-2663-4043

**Abstract:** Astrology is a very ancient and traditional method of prediction that increases the interest of people continuously. The globe today, there are no common guidelines or principles for astrological prediction. Rather than setting universal principles and criteria for astrological prediction, astrologers focus on providing high-quality services to individuals but there is no guarantee of accuracy. Machine learning is providing the best result for analysis and prediction on many applications by the learning of computers. Prediction and classification make it possible for any learner to work on large, noisy, and complex datasets. The main motive of the paper is to introduce a scientific approach that reduces the drawback of the traditional approach and indicates the universal rules of prediction and proves the validity of astrology by the three classification techniques, Naïve Bayes, Logistic-R, and J48. It is a part of supervision learning that operates with cross-validation 10,12, and 14fold for calculating the terms 1) correctly classified instances (CCI), erroneously categorized instances (ECI), Mean absolute error (MAE), Root mean squared error (RMSE), and Relative absolute error (RAE). 2) True Positive Rate, False Positive Rate, Precision, and F-Measure values. 3) The MCC, ROC, and PRC area values. 4) To calculate the average weight of the three-class label professor, businessman, and doctor in terms of true positive rate, false-positive rate, precision, F-measure, PRC, and ROC area, 5) finally, we calculated the accuracy of each classification technique and compare which provide the better result. For this, we have collected the date of birth, place of birth, and time of birth of 100 persons who belong to different professions. 40 data of professors, 30 data of businessmen, and 30 data of doctors, prepare the horoscope of an individual with the help of software. For analysis, we create the datasheet in .csv format and apply this data sheet in the weka tool to check various parameters and the accuracy percentage of each classifier.

**Index Terms:** Machine learning, naïve Bayes, logistic -R, J48, horoscope, Astrology, Weka.

## 1. Introduction

Authors Astrology is a study that is characterized as a traditional, indication-based exercise that is logically impossible. It's based on such a belief system that illustrates the planets' interconnectedness and aids in the identification of human personalities whereas biometrics characteristics identify the person recognition by face, ear, iris, palm, and footprint [1]. Astrology is used to make predictions based on a person's birth time, date and location. This information aids in the creation of a horoscope divided into twelve parts simply says houses. According to an astrologer,

a person's entire journey is defined by this house. Every house is ruled by a different sign, which is decided by the zodiac sign. The first six houses depict a person's personal life, such as family and community, while the next six houses depict relationships, careers, and Financial [2]. 9 planets, 12 houses, Dasha, Nakshatra, and numerous divisional charts are used by astrologers. To investigate several aspects of a person's personality. Individual horoscopes are used by astrologers to make predictions. Astrology can be used to assess a person's intelligence, educational attainment, and career accomplishment [3].

In the Traditional approach, the prediction of a horoscope is a very challenging task for the person who does not have the knowledge of "Jyotishavidhaya" and astrologers (Who has the knowledge of horoscope) prediction has no guarantee of accuracy. We proposed a scientific approach to prediction with the help of classification techniques of machine learning where horoscope data is used to forecast a person's profession. The classification techniques will produce the findings in the form of rules. These classification-based rules will be traced back and cross-checked against astrological rules. Some software Astro sage and Kundali are used for preparing the horoscope and making possible the prediction of the future of the person but there are various difficulties with predicting accuracy, such as giving data that might not be accurate.

Artificial intelligence includes machine learning as a subset and it is the core notion overdue the construction of a structure based on facts that generate computer algorithm expertise through practice. Training data is the name given to this group of data, and it assists in making decisions and identifying outlines or forecasts with little hominoid interaction. Data filtering and data analytics are just a few of the disciplines where machine learning algorithms are used [4].

Astrology, on the other hand, provides a wealth of information about a person's character, personality, career, health, fortune, and destiny. The purpose of this study is to focus on artificial intelligence classification systems that automatically classify data used in scientific fundamentals, which is like astrology work [5].

By taking the necessary steps, it is possible to ensure that the quality of education is maintained in online classes [4]. Therefore, it's crucial to offer administrative or technical support early on in a totally online environment [5].

The study is comprised of five sections, which are structured as follows, an introduction is found in section 1, followed by a summary of the overall work in section 2. The methodology is stated in section 3. Section 4 includes the analysis of the obtained data and findings. Lastly, section 5 consists of the conclusion and discussion.

## 2. Background Study

A number of researchers have contributed to this field. Rishi and Dhyani (2015), described computers learn from unfamiliar, big, noisy, or complicated data sets to forecast and classify data in various applications. Using diverse scientific methodologies, they are attempting to discover universal laws and the veracity of astrology. They will use the Xero, Simple Cart, and Decision Table classification algorithms to determine the person's occupation. The data set for learning classification included 24 Singer recordings, 24 Player records, and 10 Scientist records. The Weka tool which is free and open-source issued to do analysis and prediction tasks [6]. From the instances saved in the case base, O.P. Rishi (2010), introduced the Astrological Prediction System about Profession using Case-Based Reasoning (APSAP-CBR) to learn the association between the planetary location of stars at the moment of birth of the individual and his profession. The system then utilizes this relationship to forecast how fresh cases will be handled. Once the new case has been anticipated and validated, it is either immediately maintained in the case base or modified for future usage [7].

Self-selection and double-blind astrological chart matching experiment were founded by Bhandary, R. (2018), A resistor group was formed, in which participants did not give true profiles. Despite this, the resistor group selected the pre-store profiles with a much lower likelihood than the casual (p.01, significance p.05). The real testing, on the other hand, chose their profiles at random [8].

Rajopadhye, N. (2021), this study looked at a few basic concepts of Vedic astrology by conducting a systematic empirical experiment on the birth charts of persons who have had cancer and those who have never had cancer. Two data sets were created, one containing 254 birth charts of persons diagnosed with cancer before the age of 60 and the other containing 498 birth charts of those who lived to be more than 80 years old and never had cancer. In the absence of a reliable set of guidelines for predicting cancer, In terms of astrological negativity or positivity of any of the entities we investigated, however, there was no statistically significant difference between the two data sets [9]. Ahmad A. (2020), to develop medical facilities, public health policymakers require a solid projection of confirmed cases in the future. Machine learning algorithms learn from previous data and forecast the future [10].

Olaiya, F. et.al. (2012). Data mining techniques were utilized to anticipate weather reports such as rainfall, temperature, and wind speeds using Artificial Neural Network and Decision Tree algorithms, as well as meteorological data gathered between 2000 and 2009 in Ibadan, Nigeria. A data model was constructed for the meteorological data, and the classifier algorithms were then trained using this data. The algorithms' performance was evaluated using conventional performance indicators, and the best algorithm was utilized to generate classification criteria for the mean weather variables [11].

Cappozzo, A. (2020) used data mining to produce health prediction astrology. They conduct an experiment by gathering data from 100 participants and loading it into the WEKA software a prominent Java-based bundle. After examining samples and employing various data mining techniques, it is possible to correctly identify and forecast human health. Many tools are available for preparation [12].

Oyewola, D. O. (2021), An effort has been made to develop a formal method to astrological prediction and birth chart interpretation using a common Artificial Intelligence technique known as case-based reasoning. For this study, 450 experts from diverse fields were recruited. The true basis of astrological prediction utilising a normal computational approach and the Nearest Neighborhood algorithm is provided [13]. Kotsiantis, S.B., (2007), the purpose of supervised learning is to create a compact model of class label distribution in terms of predictor characteristics. When the values of the predictor characteristics are known, but the value of the class label is unknown, the resultant classifier is used to assign class labels to the testing examples [14].

Loh, W. (2011), Recursively splitting the data space and fitting a basic prediction model within each partition yields the models. As a consequence, the partitioning may be shown as a decision tree visually. For dependent variables with a finite number of unordered values, classification trees are used, with prediction error evaluated in terms of misclassification cost [15].

Cruz, J.A., (2007), this method is especially intriguing because it is part of a rising trend toward customized, predictive treatment. We did a wide study of the many types of machine learning algorithms being utilized, the types of data being incorporated, and the performance of these approaches in cancer prediction and prognosis while putting together this review [16].

Kohavi, R., & Sommerfield, D., (1998), advocate for the usage of decision table classifiers that are simple to grasp for line-of-business users. We analyze the performance of multiple methods for learning decision tables and present a visualization tool that we built in mine set [17].

Chaplot, N., (2016), this study discusses astrological prediction experiments and attempts to uncover scientific validity and principles for astrological prediction using the Case Base Reasoning approach. Simple Cart, Logistic, Nave Bayes, Decision Stump, Decision Table, and DTNB are some of the Artificial Intelligence algorithms and classification approaches utilized for this purpose. Experiments on astrological charts are used to determine whether or not a person will become renowned globally [18].

Sánchez, J.S. (2007) Attempts were made to characterize the k-NN rule's behavior in various contexts. The current study examines the usage of several data complexity metrics to define class overlapping, feature space dimensionality, and class density, as well as their relationships with the practical accuracy of this classifier [19].

Zhang, X. (2011). Improved the performance of Chinese architectural document categorization, a new algorithm based on rough set and Confidence Attribute Bagging is proposed. It addresses the problems of traditional feature selection methods, such as threshold filtering losing a lot of effective architectural information, and the shortcoming of the Bagging algorithm, where weaker classifiers of Bagging have the same weights [20].

## 3. Transformation of Traditional to Scientific Approach

### 3.1 Tradition System Process

Astrology comes under the traditional Indian science that incorporates geographical location mathematics, astronomy, and cultural transformations into its calculations. Astrology is a subject of planet study. Astrologer collects data from the individual person in terms of birth date, accurate timing when he/she was born and the place where or was he born and create the horoscope of the person that determine the whole details of the person from birth to death. Actually, In the traditional system, horoscope preparation is a very difficult task for the unknown person because everyone is not perfect in the calculation of planet position or the rashies only "Jyotishaarcharya" knows about the kundalini and they predict the person's life related to family, financial, education, profession, and health. But the accuracy of prediction is very confusing and has many challenges such as providing data that may be not correct [21]. It is a learning process. Astrologers' main focus is to satisfy the service of the person. Some hesitations to speak the truth. The traditional approach to prediction is shown in Fig.1.
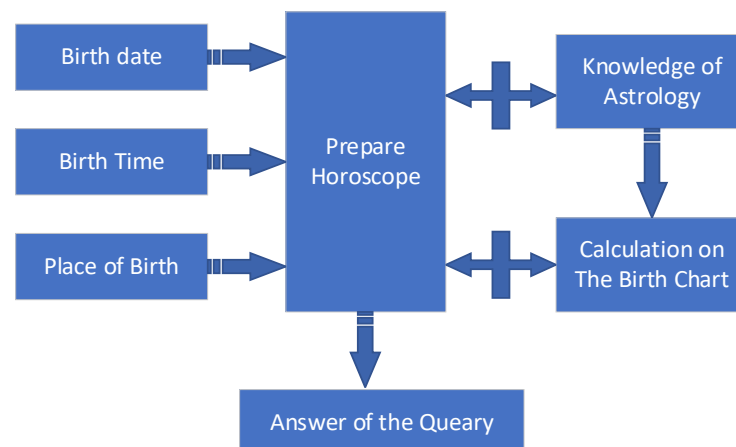


Fig. 1. Traditional Approach to prediction

### 3.2 Scientific Approach

As per our knowledge, astrology is not a scientific stream, and the scientific community has rejected it due to the lack of a trustworthy explanation for presenting the entirety of the world. Many changes occur as the shifting of scenarios from astrology theories ranges from traditional to scientific [22]. Machine learning is made possible to learn astrology for everyone with the help of a computer system and increases the prediction accuracy percentage high. There are many approaches such as classification, clustering, and association rules available for analysis and action. The scientific approach to prediction is shown in Fig.2.
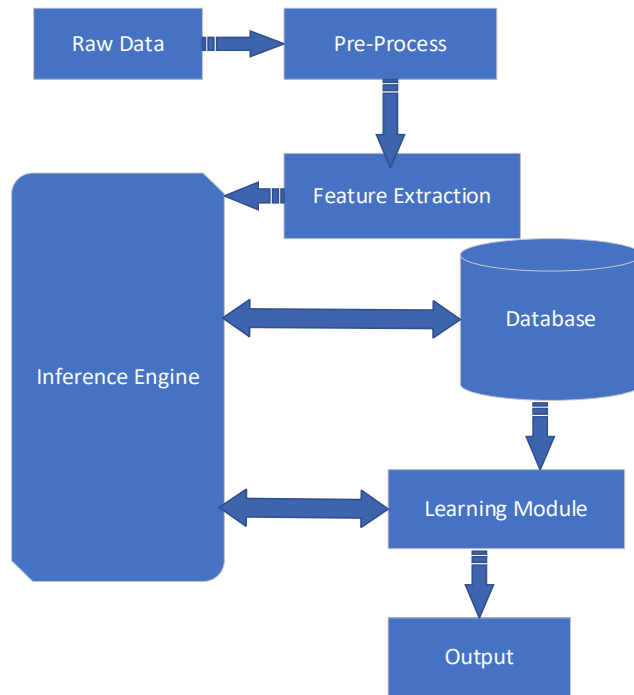


Fig. 2. Scientific Approach to the Proposed System

## 4. Methodology

Machine learning is a subfield of artificial intelligence that lies at the heart of data-driven system development that develops computer algorithm skills over time. The training data set aids in decision-making and recognizing patterns or forecasts with minimal human input [23]. After completing their education, many students are unsure of what they want to do with their lives and what they will become. Career astrology provides the best method to choose a suitable profession based on a forecast, but it is not certain and provides no evidence. This problem is solved by matching learning, which provides a scientific basis for prediction and makes learning simple for everyone [24].

Table 1. Planets and their responsible field

| Planet | Responsible Fields |
| --- | --- |
| Sun | Administrative officer, Leadership |
| Moon | Travel |
| Mercury | Media, communication |
| Jupiter | Fame, Research |
| Mars | Defense |
| Saturn | Service, Labour |
| Venus | Artist |
| Rahu | Politician |
| Ketu | Businessman |

Gathered data from people who work in various professions using a scientific manner. The second, sixth, and tenth houses of the horoscope represent a person's income and career. The placement of the planets, as well as the combination of planets and zodiac signs, determines the career.

*4.1. Raw Data Set*

We created a dataset in terms of DOB, TOB, and POB (Date of Birth, Time of Birth, and Place of Birth) of 100 people, where 40 people are professors, 30 doctors, and 30 people are businessmen.

*4.2. Preprocessing Data*

We constructed the person's horoscope chart after gathering raw data, which illustrates the twelve zodiac signs, and the location of a planet.

*4.3. Feature Extraction*

The 12 houses in a horoscope describe nine planets' positions, every house has a zodiac sign, more than one planet's combinations, and their relations based on yoge, the power of the second, sixth, and tenth houses, and other astrological words. Segment charts that look at a variety of features of a person and help to get at a highly exact result.

*4.4. Inference Engine*

An element of the system called an inference engine applies logical principles to the knowledge base to derive new knowledge. Information about the globe was kept in the knowledge base. As more rules might be triggered by new facts in the knowledge base, the process would continue to iterate. The two main ways that inference engines operate are either unique rules or facts.

*4.5. Classification Techniques*

Various categorization approaches supervised or unsupervised are available in Artificial Intelligence and can be used to generate broad theories from a data set that make possible predictions for the future [25]. However, determining whether or not this is an appropriate strategy for the data set is quite challenging. Three AI classification techniques are chosen for their ease of learning and complexity [26].

*4.5.1 Naïve Bayes Classification:*

The nave Bayes method is another common and successful probability-based classification technique. It's suitable for the prediction when the class has more option issues that the features of the class are present or absent as well as connected to or unconnected to other components [27, 28, 29].

Many applied naive Bayes classification applications rely on, despite the fact that Bayesian methods are more accurate.

Let R1, R2, R3……Rn be 'n' classes

All Naïve Class independent conditionally

$$P\left(\frac{X}{Rk}\right) = \prod_{i=1}^{n} P\left(\frac{Xi}{Rk}\right) \tag{1}$$

$$P\left(\frac{X}{Rk}\right) = P\left(\frac{X1}{Rk}\right) * P\left(\frac{X2}{Rk}\right) * \dots * P\left(\frac{Xn}{Rk}\right) \tag{2}$$

*4.5.2 Logistic Regression:*

Logistic-R is another most popular supervised classification approach of machine learning. It's used different class label dependent variables for prediction from the set of independent factors. Logistic-R produces the output in discrete forms like true /false, yes/no, on/off and 0/1 Instead of the exact value of 0 and 1, produces output in the range 0 to 1.

The linear regression yielded the logistic regression straight-line formulas written as:

$$Z = B_0 + B_1 X_1 + B_2 X_2 + B_3 X_3 + \dots + B_n X_n \tag{3}$$

The equation of logistic-R is in the form Z =1

$$\log\left(\frac{p}{1-P}\right) = B_0 + B_1 X \tag{4}$$

The projected probability value of X is

$$P = \frac{\ell^{(B_0 + B_1 X)}}{1 + \ell^{(B_0 + B_1 X)}} \tag{5}$$

*4.5.3 J48*:

Classifier J48. C4 generates a decision tree and uses this algorithm to generate it (an extension of ID3). A statistical classifier is another name for it. J48 produces not only correct predictions but also explains patterns data, Numeric characteristics, missing data, pruning, estimating error rates, decision tree induction difficulties, and rule generation from trees are all addressed issues.

## 5. Result and Analysis

A tabular record of relevant data using the horoscopes chart was created which contained the birth information of the person.

Table 2. Basic birth records of persons

| Person | DoB | ToB | PoB |
|--------|-----|-----|-----|
| P3 | 21/5/85 | 5:50 PM | Delhi |
| P7 | 23/10/82 | 9:25 PM | Goa |
| P17 | 4/8/87 | 9:15 AM | Agara |
| P23 | 7/6/84 | 5:45 PM | Nagpur |
| P33 | 12/4/89 | 9:25 AM | Raipur |
| P58 | 19/9/68 | 7:20 AM | Sambalpur |
| P70 | 28/5/71 | 8:36 PM | Bhopal |
| P82 | 23/8/87 | 9:45 AM | Dubai |
| P91 | 21/7/72 | 8:15 AM | Gwalior |
| P98 | 12/2/69 | 4:40 AM | Mumbai |

Following the data collection, we created a horoscope chart for each person using 'Astrosage' Kundli software to see the location of planets, their combination, and zodiac sign in the 12 houses shown in Figure 3.
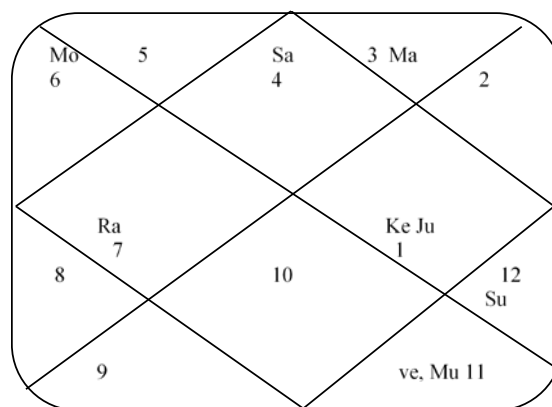


Fig. 3. Horoscope of a person

We then created a table to take the attributes such as all planets, zodiac signs with values ranging from 1 to 12, and class with the values of professor, Businessman, and doctor based on the chart shown in Table 3. The prepared dataset is loaded into the WEKA tool with CSV (comma, delimited) extension format [30]. The number of classes with labels and their weight is shown in figure 4.

Table 3. Basic details of the horoscope of a person in .csv file

| S. No. | Attribute | Value | Type |
|---|---|---|---|
| 1 | Aries | Any value 1 to 12 | Numeric |
| 2 | Taurus | Any value 1 to 12 | Numeric |
| 3 | Gemini | Any value 1 to 12 | Numeric |
| 4 | Cancer | Any value 1 to 12 | Numeric |
| 5 | Leo | Any value 1 to 12 | Numeric |
| 6 | Virgo | Any value 1 to 12 | Numeric |
| 7 | Libra | Any value 1 to 12 | Numeric |
| 8 | Scorpio | Any value 1 to 12 | Numeric |
| 9 | Sagittarius | Any value 1 to 12 | Numeric |
| 10 | Capricorn | Any value 1 to 12 | Numeric |
| 11 | Aquarius | Any value 1 to 12 | Numeric |
| 12 | Pisces | Any value 1 to 12 | Numeric |
| 13 | Sun | Any value 1 to 12 | Numeric |
| 14 | Moon | Any value 1 to 12 | Numeric |
| 15 | Jupiter | Any value 1 to 12 | Numeric |
| 16 | Mars | Any value 1 to 12 | Numeric |
| 17 | Marcury | Any value 1 to 12 | Numeric |
| 18 | Saturn | Any value 1 to 12 | Numeric |
| 19 | Venus | Any value 1 to 12 | Numeric |
| 20 | Rahu | Any value 1 to 12 | Numeric |
| 21 | Ketu | Any value 1 to 12 | Numeric |
| 22 | Class | Professor, Businessman, Doctor | Nominal |

Selected attribute
Name: Class   Type: Nominal
Missing: 0 (0%)   Distinct: 3   Unique: 0 (0%)

| No. | Label | Count | Weight |
|---|---|---|---|
| 1 | Professor | 40 | 40 |
| 2 | Business | 30 | 30 |
| 3 | Docter | 30 | 30 |

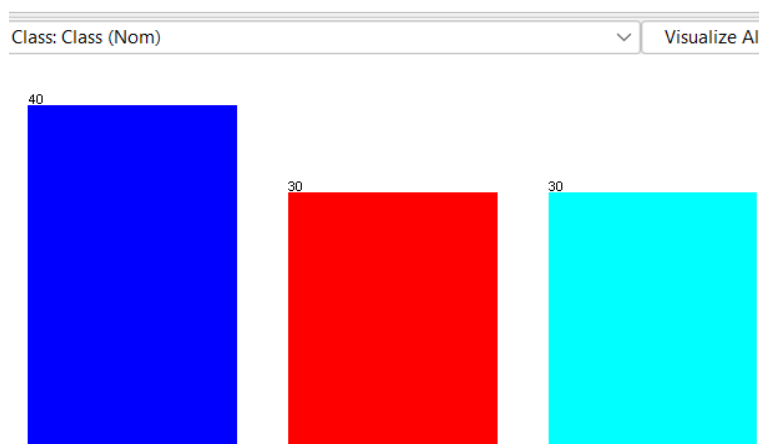Class: Class (Nom)    ∨    Visualize All



Fig. 4. Different classes of a person

We worked on three different classification techniques naïve Bayes, logistic regression, and J48 with cross-validation 10,12, and 14-fold. The value in terms of correctly classified instances (CCI) and erroneously categorized instances (ECI), Relative absolute error (RAE), Mean absolute error (MAE), and Root mean squared error (RMSE) are calculated as a result shown in table 4.

Table 4. Three classifiers result using 10,12, and 14-fold

| Classifier | Fold | CCI | ICI | MAE | RMSE | RAE% |
|---|---|---|---|---|---|---|
| Naïve Bayes | 10 | 41 | 59 | 0.40 | 0.52 | 90.9 |
| Naïve Bayes | 12 | 42 | 58 | .39 | .51 | 88.8 |
| Naïve Bayes | 14 | 43 | 57 | .38 | .50 | 88.4 |
| Logistic -R | 10 | 50 | 50 | .36 | .49 | 83.4 |
| Logistic -R | 12 | 51 | 49 | .37 | .50 | 83.4 |
| Logistic -R | 14 | 48 | 52 | .38 | .51 | 86.1 |
| J48 | 10 | 49 | 51 | .34 | .53 | 76.7 |
| J48 | 12 | 57 | 43 | .29 | .49 | 66.1 |
| J48 | 14 | 53 | 47 | .32 | .52 | 72.8 |

Table 5 indicates the three-class labels professor(P), businessman(B), and doctor(D), calculation in terms of Ture Positive Rate, False Positive Rate, Precision, and F-Measure values.

Table 5. TP, FP, precision, and f-measure for the classes

| Classification | Class | fold | TP Rate | FP Rate | precision | F-measure |
|---|---|---|---|---|---|---|
| Naïve Bayes | P | 10 | .475 | .267 | .543 | .507 |
| | B | | .233 | .286 | .259 | .246 |
| | D | | .500 | .329 | .395 | .441 |
| | P | 12 | .525 | .267 | .568 | .545 |
| | B | | .233 | .286 | .259 | .246 |
| | D | | .467 | .314 | .389 | .424 |
| | P | 14 | .600 | .233 | .632 | .615 |
| | B | | .233 | .286 | .259 | .246 |
| | D | | .400 | .329 | .343 | .369 |
| Logistic -R | P | 10 | .675 | .350 | .563 | .614 |
| | B | | .400 | .186 | .480 | .436 |
| | D | | .367 | .229 | .407 | .386 |
| | P | 12 | .650 | .267 | .619 | .634 |
| | B | | .367 | .229 | .407 | .386 |
| | D | | .467 | .243 | .452 | .459 |
| | P | 14 | .600 | .233 | .632 | .615 |
| | B | | .400 | .243 | .414 | .407 |
| | D | | .400 | .300 | .364 | .381 |
| J48 | P | 10 | .750 | .317 | .612 | .674 |
| | B | | .367 | .186 | .458 | .407 |
| | D | | .267 | .271 | .296 | .281 |
| | P | 12 | .800 | .283 | .653 | .719 |
| | B | | .400 | .200 | .462 | .429 |
| | D | | .433 | .171 | .520 | .473 |
| | P | 14 | .775 | .267 | .660 | .713 |
| | B | | .333 | .171 | .455 | .385 |
| | D | | .400 | .271 | .387 | .393 |

Table 6 indicates the MCC, ROC, and PRC area value for the three-class labels professor, Businessman, and Doctor and Table 7 indicates the average weight of the three-class label professor, businessman, and doctor in terms of true positive rate, false-positive rate, precision, and F-measure value.

Table 6. MCC, ROC, and PRC area for the classes

| Classification | Class | fold | MCC | ROC area | PRC area |
|---|---|---|---|---|---|
| Naïve Bayes | P | 10 | .214 | .660 | .509 |
| | B | | -0.054 | .478 | .369 |
| | D | | .162 | .596 | .399 |
| | P | 12 | .214 | .660 | .509 |
| | B | | -.054 | .478 | .369 |
| | D | | .162 | .596 | .399 |
| | P | 14 | .370 | .691 | .556 |
| | B | | -.054 | .541 | .358 |
| | D | | .069 | .605 | .406 |
| Logistic -R | P | 10 | .319 | .698 | .512 |
| | B | | .227 | .599 | .425 |
| | D | | .143 | .611 | .405 |
| | P | 12 | .380 | .696 | .521 |
| | B | | .143 | .611 | .433 |
| | D | | .222 | .609 | .397 |
| | P | 14 | .370 | .537 | .699 |
| | B | | .159 | .393 | .584 |
| | D | | .097 | .344. | .559 |
| J48 | P | 10 | .674 | .781 | .624 |
| | B | | .407 | .587 | .423 |
| | D | | .281 | .560 | .333 |
| | P | 12 | .506 | .801 | .613 |
| | B | | .209 | .621 | .448 |
| | D | | .277 | .684 | .458 |
| | P | 14 | .499 | .770 | .624 |
| | B | | .179 | .671 | .422 |
| | D | | .127 | .625 | .393 |

Table 7. The average weight of the three-class

| Classification | Cross-validation folds | TP Rate | FP Rate | Precision |
|---|---|---|---|---|
| Naïve Bayes | 10 | .410 | .291 | .413 |
| | 12 | .420 | .287 | .421 |
| | 14 | .430 | .278 | .433 |
| Logistic -R | 10 | .500 | .264 | .491 |
| | 12 | .510 | .248 | .505 |
| | 14 | .480 | .256 | .480 |
| J48 | 10 | .490 | .264 | .471 |
| | 12 | .570 | .225 | .556 |
| | 14 | .530 | .240 | .516 |

J48 provide the highest score of TP rate at 12-fold while naïve Bayes gives the minimum TP rate. The graphical representation shown in figure5.
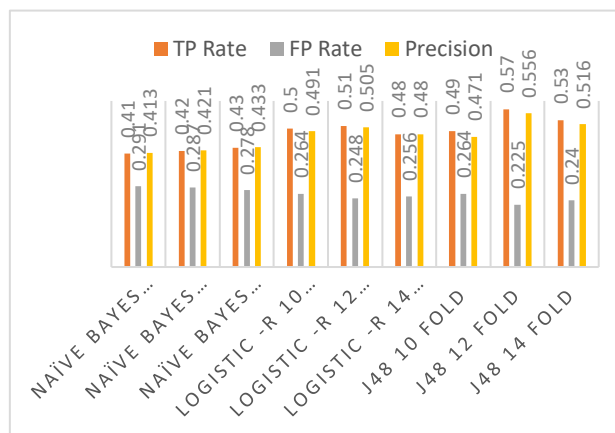


Fig. 5. Average weight rate of TP, FP, and Precision of Three class

Table 8 indicates the average weight of the three-classes professor, businessman and doctor in terms of F-measure, PRC, and ROC area, and its graphical representation is shown in Figure 6.

Table 8. The average weight of three classessses professor, businessman, and doctor

| Classification | Cross-validation folds | F-Measure | PRC Area | ROC Area |
|---|---|---|---|---|
| Naïve Bayes | 10 | .409 | .434 | .660 |
| Naïve Bayes | 12 | .419 | .463 | .625 |
| Naïve Bayes | 14 | .431 | .452. | .620 |
| Logistic -R | 10 | .492 | .238 | .642 |
| Logistic -R | 12 | .507 | .457 | .644 |
| Logistic -R | 14 | .482 | .436 | .622 |
| J48 | 10 | .476 | .476 | .657 |
| J48 | 12 | .558 | .517 | .712 |
| J48 | 14 | .518 | .494 | .697 |
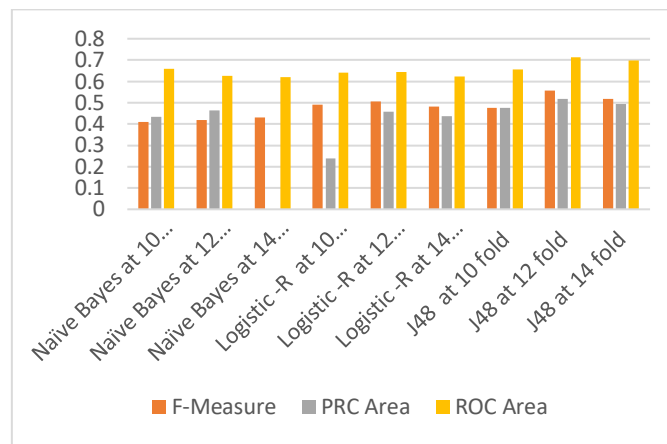


Fig. 6. Average weight in term F-Measure, PRC, and ROC Area of three classes

Cross-validation at 12 folds, classifier J48 generates the better accuracy results of 88.76%, while naïve Bayes generate minimum accuracy of 62.56%, all classifiers' results are shown in figure 7.
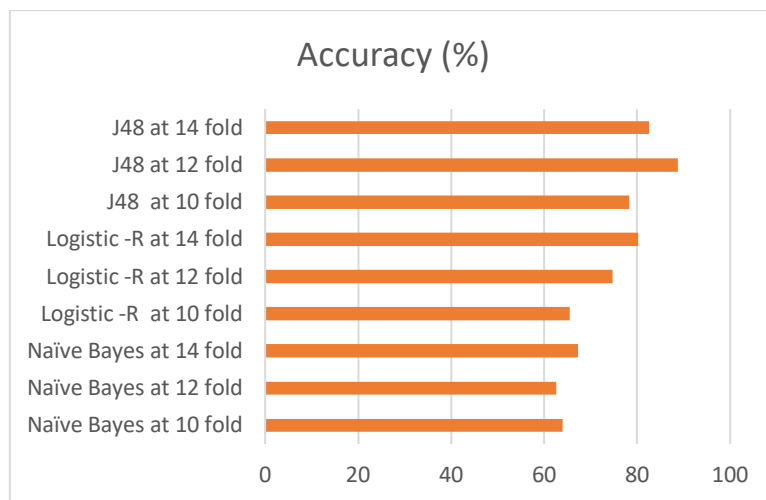


Fig. 7. Graph for Naïve Bayes, Logistic Regression, and J48 classification technique

## 6. Conclusion and Discussion

For the same reason, good astrological software is available nowadays, but it does not calculate any data for any event. We used some limited/ restricted data and training methods in this study article. We gathered information from 100 people to predict their career area, which is Professor, Businessman, and Doctor, based on their birth information, in order to forecast their future. We computed the result and found that the high percentage of accuracy in 12-fold J48 is

88.76% and the low percentage of accuracy 12 fold naive Bayes is 62.56 %, respectively, using three classification techniques: nave bays, logistic-R, and J48 with cross-validation 10, 12, and 14 fold.

Although the results fell short of expectations, they are also not depressing. Few approaches outperform others, thus the accuracy of the system can be improved if the remaining ways are attempted or hybrid methods are developed.

Additionally, the data set only contained 100 records, therefore it is believed that if the number of cases increases, so will the accuracy of the results. In order to create new classifiers that make use of the strengths of other classifiers and hence improve prediction accuracy, results from working on several classifiers can be utilized to examine and identify the strengths and weaknesses of various classification methods.

Along with birth charts, astrological charts, and planetary positions in various houses in the horoscope, additional parts of a person's information, such as knowledge level, family background, etc., must be added as qualities for the purpose of accuracy. To improve the accuracy of the results, a few other factors could be added, such as the planets' strengths, aspect ratios, and relationships with one another.

Predictions are made to determine if a person will become a doctor, professor, or businessman or not. They can also be made to determine a person's basic character, attitude, financial situation, and other aspects of their life.

## 7. Limitations of the Study

A prognosis for a person who has decided whether or not to pursue the careers of Professor, Businessman, or Doctor. However, it is dependent on a variety of factors, such as a person's surroundings. His qualifications are other important factors in determining the type of work he gets, and money is another factor to consider.

## 8. Future Scope

In this research, the results focused on working on three classifiers that can be utilized to examine and identify the strengths and weaknesses of classification methods. Along with birth charts, and planetary positions in various houses in the horoscope. But in the future to improve the accuracy of the results, a few other factors would be added, such as the planets' strengths, aspect ratios, and relationships with one another. Additional parts of a person's information, such as knowledge level, and family background, are also important to know before the prediction process.

## References

[1] S. Barde, A.S. Zadgaonkar, (2014, Feb.). G.R. Sinha, "Multimodal biometrics using face, ear and iris modalities," International Journal of Computer Applications, NCRAIT no. 2, pp.9-15,

[2] O. P. Rishi and N. Chaplot, (2010, Dec.). "Archetype of astrological prediction system about profession of any persons' using case based reasoning," 2010 International Conference on Communication and Computational Intelligence (INCOCCI), pp. 373-377.

[3] S. Tiwari and S. Barde,( 2021, Aug.). "Machine Learning Make Possible to Astrological Prediction for Government Job Using Classification Techniques," 2021 Asian Conference on Innovation in Technology (ASIANCON), pp. 1-4, doi: 10.1109/ASIANCON51346.2021.9544586.

[4] S.B. Kotsiantis, I.D. Zaharakis, & P.E. Pintelas, "Machine learning: a review of classification and combining techniques," Artif Intell Rev., vol. 26, pp.159–190, 2006.

[5] R.P. Bhandary, P.S. Sharma, & H. Tharoor, "Prediction of Mental Illness Using Indian Astrology: Cross-Sectional Findings from a Prospective Study," Journal of Scientific Exploration., vol.32, no.3, pp. 555-578, Jan 2018.

[6] N. Chaplot, P. Dhyani and O. P. Rishi, (2015, May). "Astrological prediction for profession using classification techniques of artificial intelligence," International Conference on Computing,Communication & Automation, 2015, pp. 233-236,

[7] O. P. Rishi and N. Chaplot, (2010, Dec.)."Predictive role of case based reasoning for astrological predictions about profession: System modeling approach," 2010 International Conference on Communication and Computational Intelligence (INCOCCI), pp. 313-317.

[8] R.P. Bhandary, P.S.Sharma, & H. Tharoor, "Prediction of Mental Illness Using Indian Astrology: Cross-Sectional Findings from a Prospective Study," Journal of Scientific Exploration, vol.32.no.3, pp, 555-578, Sep.2018.

[9] N.R. Rajopadhye, A. Rajopadhye, M. Rajopadhye, & P. Kulkarni, "Empirical testing of few fundamental principles of Vedic astrology through comparative analysis of astrological charts of cancer diseased persons versus persons who never had it," International journal of applied research, vol.7, pp.74-85, May 2021.

[10] A. Ahmad, S. Garhwal, S.K. Ray, "The Number of Confirmed Cases of Covid-19 by using Machine Learning: Methods and Challenges," Arch Computat Methods eng, vol. 28, pp. 2645–2653, Jun 2021.

[11] F. Olaiya, & A.B. Adeyemo, "Application of Data Mining Techniques in Weather Prediction and Climate Change Studies," International Journal of Information Engineering and Electronic Business, vol.4, pp.51-59.Feb 2012.

[12] A. Cappozzo, F. Greselin, & T.B. Murphy, "A robust approach to model-based classification based on trimming and constraints," Adv Data Anal Classif, vol. 14, pp.327–354, Jun 2020.

[13] DO. Oyewola, EG. Dada, S. Misra, R. Damaševičius, "Predicting COVID-19 Cases in South Korea with All K-Edited Nearest Neighbors Noise Filter and Machine Learning Techniques," Information. vol.12, no.12, pp.528-536, Dec 2021.

[14] S.B. Kotsiantis, I.D. Zaharakis, & P.E. Pintelas, "Machine learning: a review of classification and combining techniques," Artif Intell Rev, vol. 26, pp.159–190, Nov 2006.

[15] W. Loh, Classification and regression trees. Wiley Interdisciplinary Reviews: Data Mining and Knowledge Discovery, vol.1, no.1, pp.14-23, Jan 2011.

[16] JA. Cruz, DS. Wishart, "Applications of machine learning in cancer prediction and prognosis," Cancer Inform, vol.11, no. 2, pp.59-77, Feb 2007.

[17] R. Kohavi, and D. Sommerfield, (1998). "Targeting Business Users with Decision Table Classifiers," KDD-98 Proceedings.

[18] O. P. Rishi and N. Chaplot, (2010, Dec.). "Predictive role of case-based reasoning for astrological predictions about the profession: A system modeling approach," 2010 International Conference on Communication and Computational Intelligence (INCOCCI), pp. 313-317.

[19] J.S. Sánchez, R.A. Mollineda, & J.M. Sotoca, "An analysis of how training data complexity affects the nearest neighbor classifiers," Pattern Anal Applic, vol. 10, pp.189–201, Feb. 2007.

[20] X. Li, Zhang, C., Dong, L., "Rough Sets and Confidence Attribute Bagging for Chinese Architectural Document Categorization," Journal of Software, vol. 6, no. 4, pp.628-635, Apr 2011.

[21] R.P. Bhandary, P.S. Sharma, & H. Tharoor, "Prediction of Mental Illness Using Indian Astrology: Cross-Sectional Findings from a Prospective Study," Journal of Scientific Exploration, vol. 32, no.3, pp.555-578, Sep. 2018.

[22] P. Bhattacharjee, Scientific astrological prediction of human life. International Journal of Jyotish Research, vol.5, no.1, pp.12-16, Aug. 2020.

[23] M. Singla, D.Ghosh and K.K.Shukla, A survey of robust optimization-based machine learning with special reference to support vector machines. Int. J. Mach. Learn. & Cyber, vol.11, pp.1359–1385, July 2020.

[24] D. Faccini, F. Maggioni, & F.A. Potra, "Robust and Distributionally Robust Optimization Models for Linear Support Vector Machine," pre-print submitted to Elsevier, vol.5, no.1, pp. 1-40, Jun 2021.

[25] J. Liu, C. Zhong, M. Seltzer, & C. Rudin,(2022, Feb). "Fast Sparse Classification for Generalized Linear and Additive Models," Proceedings of The 25th International Conference on Artificial Intelligence and Statistics in Proceedings of Machine Learning Research, vol.151, pp. 9304-9333.

[26] R.C. Prati, J. Luengo, & F. Herrera, "Emerging topics and challenges of learning from noisy data in nonstandard classification: a survey beyond binary class noise," Knowl Inf Syst, vol.60, pp. 63–97, July 2019.

[27] H.Zhang, J. Su,(2004). "Naive Bayesian Classifiers for Ranking. In: Boulicaut, (eds) Machine Learning: ECML, Lecture Notes in Computer Science, Springer, Berlin, Heidelberg, vol. 3201, 2004.

[28] M.R. Keyvanpour, & F. Baesi, (2015, Feb). "IG-NB: A Hybrid Method for Data Classification," International Conference Data Mining, Civil and Mechanical Engineering (ICDMCME'2015), Bali (Indonesia), pp.263-291.

[29] M.A. Hall, & E. Frank, (2008). "Combining Naive Bayes and Decision Tables," Proceedings of the Twenty-First International FLAIRS Conference.pp.318-319.

[30] M.A. Hall, E. Frank, G. Holmes, B. Pfahringer, P. Reutemann, & I.H. Witten, "The WEKA data mining software: an update," ACM SIGKDD Exploration Newsletter, vol. 11, no.1, pp. 10-18, Jun. 2009.

## Authors' Profiles

**Dr. Snehlata Barde** is working as a Professor in the MATS school of Information technology at MAT'S University, Raipur, (C.G.). She holds a Ph.D. in IT &CA, from C.V.R. University, Bilaspur. MCA from Pt.Ravi Shankar Shukla University, Raipur, and MSc. (Maths) from Devi Ahilya University Indore. Her research interest includes Digital Image Processing and its Applications in Biometric Security, Forensic Science, Pattern Recognition, Segmentation, Simulation and Modulation, Fuzzy logic, Soft Computing Techniques, Cyber Crime, and IoT. She has published 75 research papers in various International and National Journals and Conferences. She has attained 36 seminars, workshops, and training programs, and she has published 6 Book Chapters and 1 Book in Springer and Elsevier. She got a certificate from CMI (Chartered Management Institute) Level 5 in "Management and Leadership", UKIERI AICTE TECHNICAL LEADERSHIP PROGRAMME UK-India Education and Research Initiatives, from the British Council, and an NPTEL Certificate for reviewing the translated files of the course CLOUD COMPUTING offered by IIT Kharagpur. She got the "Manav Gaurav" award from the Human Wellness & Social Development Organization Chhattisgarh and Best Paper Award from AMET Deemed University Chennai. She is a member of IAENG (International Association of Engineers), Hong Kong, Universal Association of Computer and Electronics Engineers (UACEE), International Computer Science and Engineering Society (ICSES), and International Economics Development and Research Center (IEDRC). Indian Academicians and Researchers Association(IARA), Chartered Management Institute (CMI) in Management and Leadership with Dudley College of Technology UK. She has around 23 years of experience in teaching at technical institutes and universities.

**Sangeeta Tiwari** is a research scholar of MSIT, MATS University Raipur and working as an Assistant professor (Guest Faculty) in the computer science department at GOVT.N.P.G. College of Science Raipur. She has completed a Master of Computer Application from MATS University Raipur, Her research interests include Software Engineering, Machine Learning, and Data Mining. Currently, she is doing research in her desired fields by combining her work experiences and research interests.

**Brijesh Patel** received a Ph.D. degree in Mechanical Engineering from MATS University, India in 2018. Currently, he works as a post-doctoral researcher at the Department of Mechanical Engineering, National Taiwan University of Science and Technology, Taiwan. His research interests are robotics and drones, industrial aerodynamics, and composite material.