International Journal of Communication Networks and Information Security
2024, 16(4)
ISSN: 2073-607X, 2076-0930
https://https://ijcnis.org/ **Research Article**

An Ensemble Based Astrological Prediction Model for Profession and Marriage Using Machine Learning Strategies

S.Jaiganesh[1], Dr.P.Parameswari[2]

[1]Research Scholar, Karuppannan Mariappan College Affiliated to Bharathiar University,
[2]Principal, Palanisamy College of Arts, Perundurai.

Email: [1]prof.jaiganesh@gmail.com, [2]paramtech20@gmail.com

ARTICLE INFO

Received: 19 Apr 2024
Accepted: 28 Aug 2024

ABSTRACT

The fascination with astrology, an ancient and conventional form of prediction, continues to grow despite the absence of universal astrological prediction rules or principles globally. While accuracy is not guaranteed, astrologers prioritize offering high-quality services over establishing universal standards. In contrast, machine learning yields superior outcomes across diverse applications through its capacity to handle large, noisy, complex datasets via classification and prediction. This paper aims to present a scientific method that addresses the shortcomings of traditional astrology, identifies universal prediction rules, and employs classification techniques—Neural Network (NN), Import Vector Machine (IVM), Random Forest (RF), and Iterative Boosting—to validate the reliability of astrology in predicting profession and marriage outcomes. We computed Correctly Classified Instances (CCI), Incorrectly Classified Instances (ICI), Mean Absolute Error (MAE), Root Mean Squared Error (RMSE), and Relative Absolute Error (RAE) using cross-validation with 10, 12, and 14 folds. Additionally, we evaluated F-Measure, Precision, True Positive Rate, False Positive Rate, and area values for MCC, ROC, and PRC. For three-class labeling of professor, businessman, and doctor, we determined the true positive rates, false positive rates, accuracy, F-measure, PRC, and ROC area. We gathered birthdate, birthplace, and time of birth data from one hundred individuals across these professions, creating horoscopes using software. Data analysis involved building a datasheet in .csv format and employing the Weka tool to assess various parameters, including classifier accuracy, to identify the most effective classification method.

Keywords: Weka, horoscope, astrology, IVM, artificial intelligence and iterative boosting.

INTRODUCTION

An activity that is conventional, illogical, and indication-based is known as Writers Horoscope. Utilizing characteristics of the face, ears, iris, the palm, and footprint, fingerprints depends on a philosophy that assists in identifying humans and shows how interrelated the globe [1]. Predictions based on an individual's birth time, date, and place are made using astrology. Using this information, a horoscope with twelve components can be created, simply by saying "houses." Astrologers say this house determines a person's path. Based on the zodiac sign, each house has its own unique symbol. The first six homes symbolize private life, encompassing communities and families, while the following six houses reflect relationships, professions, and money [2]. Astrologers employ many divisional charts, Dasha, Nakshatra, nine planets, and twelve houses. to look into several facets of an individual's personality. Astrologers utilize individual horoscopes to forecast future events. Astrology may be employed to evaluate an individual's IQ, level of schooling, and professional success [3].

In the traditional method, predicting a horoscope is an extremely difficult undertaking for someone who does not know "Jyotishavidhaya," and the predictions made by astrologers (who do know horoscopes) are not guaranteed to be accurate. The results of the classification methods will be generated as rules. We will track down and compare these classification-based guidelines with astrological rules. While certain tools, like Astro Sage and Kundali, can prepare a person's horoscope and make future predictions, there are a number of issues with accuracy in forecasting, such as providing potentially inaccurate data.

A subfield of AI known as "machine learning" is responsible for the creation of a data-driven system that generates algorithmic expertise via experience. A collection of information known as training evidence is used to discover boundaries or trends with little to no human intervention, aiding in predicting and decision-making.Several fields employ machine learning techniques, including data analysis and data filtering [4].

An individual's character, personality, career, health, fortune, and destiny can all be learned a great deal about them through astrology, on the other hand. Examining artificial intelligence categorization systems—which function similarly to astrology in categorizing information used in empirical fundamentals—is the aim of this study [5]. It is possible to guarantee that the standard of instruction is upheld in astrology by following the required procedures. For this reason, in an entirely online setting, it is imperative to provide administrative or technical help as soon as possible. The major contributions of the proposed model are,

- This paper introduces a scientific way to avoid the weaknesses of traditional astrological prediction through the implementation of machine learning techniques.
- To illustrate how astrology has reliability in the prediction of professions and marriages, use Neural Network, Import Vector Machine, Random Forest, and Iterative Boosting. Compare accuracy and performance of the four.
- The research undertakes data collection and analysis on the birth date, birth place, and time of professionals from diverse backgrounds using the Weka tool in assessing and comparing the accuracy of different classifiers to determine general prediction rules.

This research has five components, with the first being a summary about introduction of the research and the second is related works section. In section 3, the overall methodology of the proposed model is described. The analysis and discussions of the data gathered are presented in Section 4. Section 5 concludes with a discussion and conclusion.

RELATED WORKS

Numerous researchers have made contributions in this area. Computers may learn from fresh, huge, noisy, or complex sets of information in order to anticipate and classify information for a number of reasons, as stated by Rishi and Dhyani (2015). Astrology and universal laws are being tested using various scientific methods. For categorization, ten Scientists, twenty-four Players, and twenty-four Singing records were employed. The free and open-source Weka program is used for activities involving analysis and prediction [6].

The Astrological Prediction Program about Professional using Case-Based Reason (APSAPCBR) was suggested by O.P. Rishi (2010). The cases stored in the case database form the foundation of this architecture. The system then makes use of this relationship to predict the handling of new situations [7].

Bhandary, R. (2018) established the double-blind astrological chart comparison experiment and self-selection. Participants in a resistor group were asked to provide false profiles. Nonetheless, the resistance sampling had a much lower probability of selecting the pre-store features than the randomly assigned (p.01, significant p.05). On the other hand, the participants were chosen at randomly during the actual assessment [8].

In a rigorous empirical investigation using the birth charts from those who experienced cancer as well as those who have not, Rajopadhye, N. (2021) examined some fundamental ideas in Vedic astrology. One information had 254 birth dates of persons who acquired cancer before 60, whereas the other contained 498 birth information of those who survived until 80 and did not suffer disease. Regardless of the astronomical favorable or adverse impact of whatever entity we investigated, this was true [9].

To establish medical facilities, Ahmad A. (2020) says healthcare authorities require a realistic projection of future confirmed infections. Using historical data, machine learning systems predict future events [10]. F. Olaiya et al (2012). Climate information from Ibadan, Nigeria, was collected among 2000 and 2009. After evaluating the algorithms' efficacy using traditional metrics of performance, the best approach was selected to provide the definition criteria for the typical environmental indicators [11]. Cappozzo, A. (2020) created an astrological health forecast using data mining. They conduct an investigation by compiling data from one hundred participants into a well-known Java-based bundle called WEKA software. Accurate identification and prediction of human health can be achieved by analyzing samples and using different data mining techniques. For preparation, there are numerous instruments accessible [12].

D. O. Oyewola (2021), An attempt has been made to use case-based reasoning, a popular Artificial Intelligence technique, to create a formal approach to birth chart interpretation and astrological prediction. 450 experts in various domains were gathered for this investigation. The Nearest Neighborhood method and a standard computational approach are used to provide the fundamental foundation for astrological prediction [13]. When we know the significance levels of the prediction characteristics but not the actual amounts of the class labels, we use the resulting classification algorithm to give the trial individuals class labels [14].

In 2011, Loh, W. After recursively splitting the information space, basic forecasting algorithms are fitted for every division. As a result, a decision tree may be used to graphically illustrate the division. When dealing with dependent parameters that have a limited number of disorganized beliefs, tree-like classifications are used, and the cost of misinterpretation is used to calculate the forecasting error [15].

According to Cruz, J.A. (2007), this approach is particularly exciting since it follows the growing trend of personalized and predictive medicine. This research examined the various machine learning approaches, the kinds of details used, and their usefulness in cancer detection [16].

Decision table classifiers with an intuitive user interface (UI) are recommended for line-of-business users by Kohavi, R., and Sommerfield, D. (1998). We provide an analysis of the effectiveness of several approaches to decision table learning and showcase a visualization tool [17]. In Chaplot, N. (2016), astrology forecast trials are examined, and the Case Based Reasoning approach is employed in an attempt to determine the astrological forecast's scientific viability and fundamental beliefs. Among the AI strategies and approaches to classification utilized in this situation are DTNB, Straightforward Cart, Logistic, Nave Bayes, Decisions Stump, and Decisions Table. Astrological charts studies are utilized for predicting an individual's probability of achieving international fame [18].

J.S. Sánchez (2007) An attempt was made to explain how the k-NN ruled behaved in different situations. This work investigates the links between class interconnected, distinctive space dimension, and classifier densities and the practical effectiveness of categorizing data using different data complexity metric approaches [19]. Zhang (2011). A novel method based on rough sets with Confident Characteristics the Bagging is proposed to increase the effectiveness of classifying Chinese historical texts. It addresses traditional choosing features difficulties including cutoff filtering's loss of architecture data as well as the Bagging approach's weighting of weaker classifications. [20].

2.1 Research gaps identified

While there have been many contributions in the area, several research gaps remain. The first one is the lack of real-world applications for comprehensive research that combines astrology with the latest, modern AI and machine learning techniques. Most empirical investigations are based on small samples and lack diversity in data type; thus, generalizing from these studies is hard. There is also a serious need for stronger validation methods to make the results from the predictive models more reliable and accurate, especially for those forecasts that relate to medical and health issues. The current studies also show a gap in the exploration of the ethical implications and potential biases inherent in using astrological data for prediction. In addition, astrologers could work more interdisciplinarily with data scientists and healthcare professionals, which would greatly enhance the practical application of the research results. Addressing these gaps can significantly improve the effectiveness and reliability of predictions derived from complex and diverse datasets.

PROPOSED ENSEMBLE MODEL: DATA & PRE-PROCESSING

In this study, 1200 records of various types are used, together with the matching planet values As seen in Table 2, one record of each type is used for testing, and the remaining 1200 1993 records of various types (sportsman (2), lawyer (4), a politician (5), performer (8), scientist (8), astrologers (9), singer (6)) along with their planets are employed for neural network system construction (i.e., training). For neural network modeling, the datasets are normalized inside the confined interval [0 1].

The main characteristics of a neural network's structural design are as follows: learning rate (α), momentum factor (μ), number of neurons by hidden layers (p), number of information vectors in the input layer (n), combinations of biases as well as learning weights (w and v), as seen in the accompanying figure 1. These parameters are known to be dynamic in response to input and network-targeted variables. Finding the ideal values for parameters such as n, p, α, and μ requires work, but it's also known that this might be the cause of the poor presentation if these parameters aren't optimized. This section discusses how these parameters were chosen and how to optimize them in this particular situation. The architectural design is displayed.

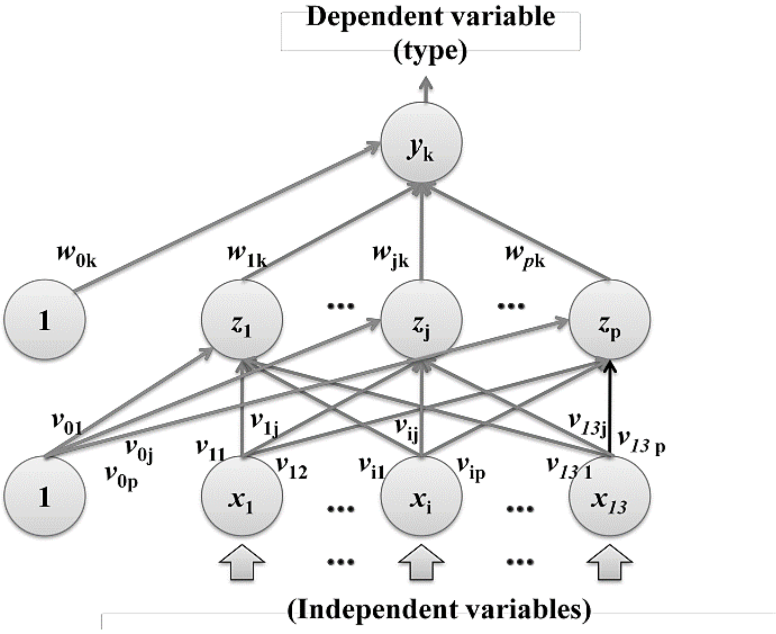

Figure 1. Developing Neural Nets for Spatial Interpolating

3.1 Optimizing the Factors

3.1.1.The quantity of variables in the input level n

There are thirteen inputs to the model in this instance: Asce, Sun, Moon, Mars, Mercury, Jupiter, Venus, Saturn, Rahu, Ketu, Urenus, Neptune, and Pluto. Take a look at Figure 1. Here, n = 13.

3.1.2.Number of outcome neurons (y)

To observe the targeted value, or type (e.g., sportsman (2), lawyer (4), politician (5), actor (8), scientist (8), astrologer (9), singer (6)), a single output neuron is utilized.

3.1.3.Number of hidden level

Almost all issues can be solved with just one hidden layer. The model is rarely advanced by using two hidden layers, and there may be a higher chance of the model convergent leading to local minima.

3.1.4.Number of neurons in hidden level (p)

Finding the lowest MSEs given matching parameter "p" involves observing their error, or mean square error, or "MSE." Following Table 1 in Fig. 2 are the system's error minimization during the training phase, together with the accompanying "p." Given that the network error MSE = 3.3237443368085606E-4 was smaller for 30 neurons, which is the optimum number in this dataset, p = 30 is chosen for this investigation.

Table 1. MSE Associated with the amount of Neurons in the Hidden level

The hidden layer's neuron count (*p*)	MSE
2	0.001639929
3	0.001638929
4	0.001416847
5	0.001316576
6	8.835130957E-4

7	8.075992623E-4
8	7.759794687E-4
9	7.1331258051E-4
10	6.694183869E-4
11	6.218868391E-4
12	5.920673453E-4
13	5.5839961329E-4
14	5.419979919E-4
15	5.01512008626E-4
16	4.8889347944E-4
17	4.6372752796E-4
18	4.381587476E-4
19	4.360597533E-4
20	4.22659445E-4
21	4.067383331E-4
22	3.936119815E-4
23	3.8158929415E-4
24	3.78634848E-4
25	3.605922392E-
26	0.2802640426
27	0.280257533
28	3.429771353E-4
29	0.28026667705
30	3.3235606E-4

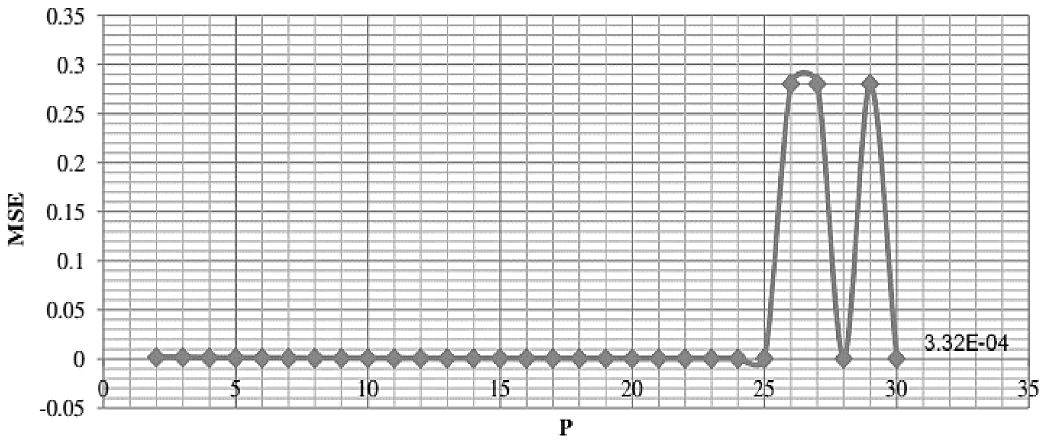

Figure 2. MSE Associated with the amount of Neurons in the Hidden Level

3.1.5.Learning rate (α)

By improving the trainable weights and biases, this parameter has a major impact on minimizing the MSEs and expediting the training process.Research indicates that a higher "α" value results in faster learning. It is imperative that researchers focus specifically towards determining an appropriate value of α to maintain a higher learning process. Through tests that explore 'α' values within a tight interval [0 1], the most favorable value of α is attempted to be identified.

Table 2. MSE Associated with Learning Rate (α)

The factor of momentum	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.6**	**0.8**	**0.9**
1	0.280	0.280	3.15E-4	2.44E-4	0.29	2.52E-4	2.7E-4	2.68E-4	2.8E-4
Suggested astro rate	0.61	**0.62**	0.63	0.64	0.65	0.66	0.67	0.68	0.69
0.6	0.28	2.52E-04	2.348E-04	0.38	2.54E-4	0.30	2.22E-4	2.43E-4	0.29

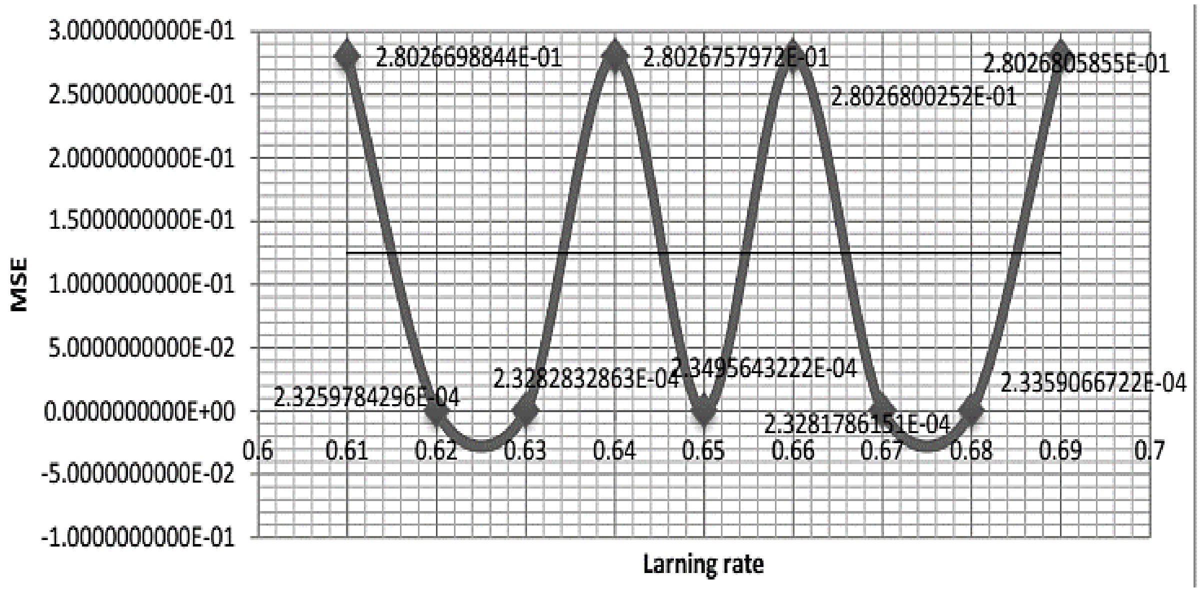

Figure 3. MSE Associated with Learning Rate (α)

3.1.6.Momentum factor (μ)

According to Karmakar et al. (2014), the forward momentum factor (μ) serves the primary function of quickening the error's convergence during the training phase of the equations. Knowing its ideal value within the near interval [0 1] is crucial. Fig. 4 and Table 3 show the values of "μ" ,which were obtained after 100 learning epochs with α = 0.62 and p = 30. The optimal value of μ = 0.96 as evident.

Table 3. MSE Correlated Momentum Amount (μ)

Rate of learning	**0.1**	**0.2**	**0.3**	**0.4**	**0.5**	**0.6**	**0.6**	**0.8**	**0.9**
0.62	0.28	0.28	0.28	0.28	3.04E-4	0.28	2.58E-4	2.34E-4	2.40E-4
Suggested momentum factor	0.95	0.96	0.94	0.95	0.96	0.98	0.99	0.97	0.98
0.6	2.38E-4	2.37E-4	0.29	2.37E-4	2.38E-4	2.37E-4	2.38E-4	2.38E-4	0.30

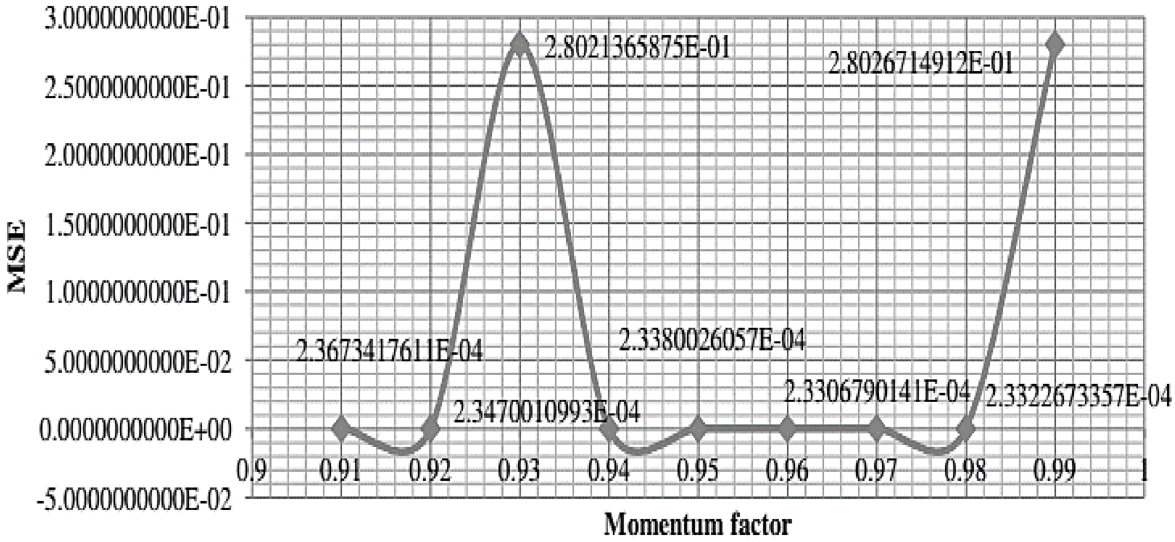

Figure 4. MSE Correlated Momentum Amount (μ)

TRAINING

According to the section 3 discussion, the parameters of the neural network determined by experiment 5 are found to be highly appropriate, with the optimal values being n = 13, p = 30, α = 0.62, and μ = 0.96. The 1200 record dataset. Of the 1200 stations, 7 records are chosen by random and utilized for testing, while 1193 records are employed for training.

The starting weights and biases are chosen at random from nearby intervals [0 1]. Next As shown in the accompanying Fig.5, the model minimizes the MSE as well as the model predicted values for each epoch. Up to MSE = 1.2864E-04, the model error MSE within the actual and predicted results was minimized during the training phase.

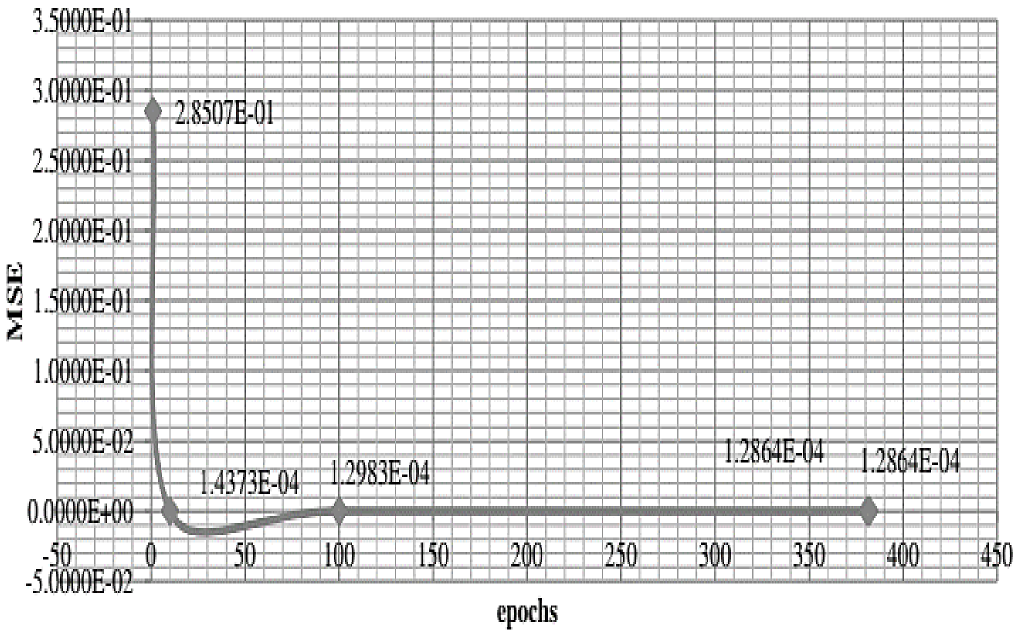

Figure 5. Reduction of error during the training phase

4.1 Trained the NN model

One could consider MSEG = 4.76E-04 to be the ideal value and consider the model to be fully trained. The model has now displayed its peak performance. The trainable biases and weights are adjusted throughout training. The modified weights and biases at global minimum MSEG are displayed in the following Table. To interpolate the average rainfall in our scenario, geo-coordinates (that is, latitude, longitude, & altitude) for an unknown station may now be entered.

4.2 Performance in training

Figure 6 below shows the model's performance during the training phase. It is also clear from Fig. 7 that there is a very little relative deviation among expected and actual numbers. Consequently, the model gains acceptance.

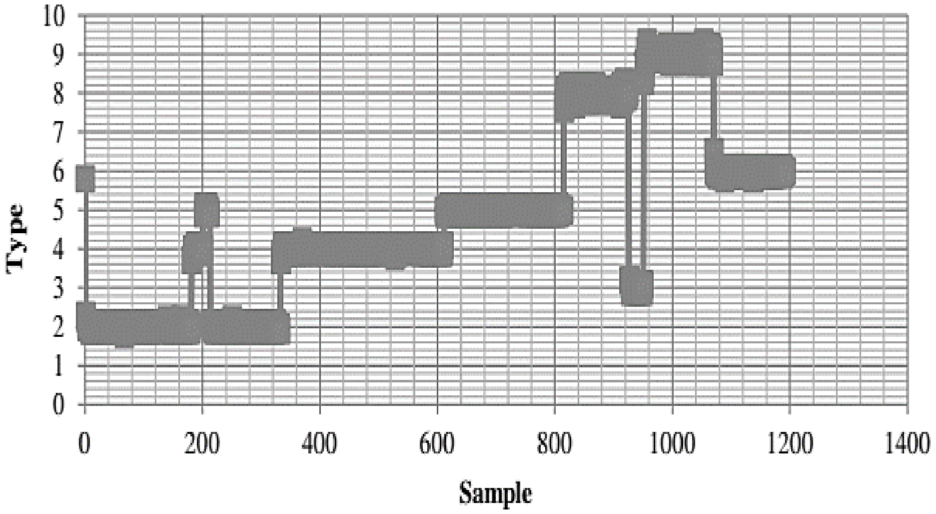

Figure 6. Performance During the Instruction Process while prediction phase

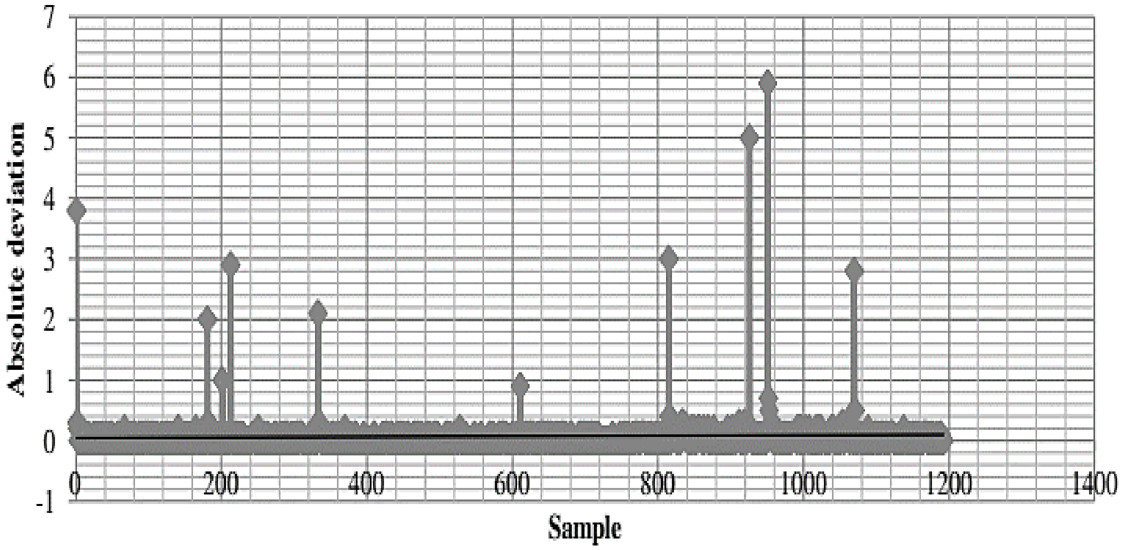

Figure 7. Absolute Difference between Expected and Real Values with actual range

4.3 IVM, RF and Iterative boosting integration

In this work, we used 2-layer stacking to build an ensemble model. The learning algorithms utilized in the first and second layers, respectively.

The following steps can be utilized to explain how a prediction framework is constructed.

Phase 1: Separate the information collection into both training and testing data;

Phase 2: Create an additional information sets by utilizing Base-learners' output;

Phase 3: Teach the Meta-learner to use the freshly created collection of data to produce the ultimate forecast outcome.

To increase prediction accuracy, machine learning algorithms can be combined using the 2-layer stacking approach. Model performance is impacted by the choice of Base- and Meta-learner while building their model.

The Base-learner should have strong prediction performance, which is the foundation of employing stacking to enhance the categorization impact. Generally speaking, stacking functions best when the combining methods are very different from each other. Accordingly, the more the disparities in categorizing concepts that exist amongst the multiple Base-learners, the more beneficial they may be in an ensembles operation. The outcomes from classification will be improved accordingly. In the first layer, these techniques were used to choose the Base-learners. In this investigation, the Base-learners selected based on their superior classification performances: the Random forest (RF) [21-23], IVM [24–26], and Iterative boosting [27, 28] classifiers. These days, they are frequently used as effective predictive models for predicting student accomplishment. Since both the RF & the AdaBoost techniques are superior ensemble decision tree-based methods, their were employed for classification. In contrast, RF generates several decision tree structures using a randomly chosen subset on training samples & variables.

RF [29] can accommodate thousands of input parameters without variable elimination and doesn't need the assumption of a distribution of the data. However, RF's accuracy in generalization is poor since its tree structure is unreliable and might overfit the the data from training. When classifying difficult data from small- to medium-sized data sets, the SVM performs exceptionally well. While linear IVM classifiers exhibit high efficiency and operate well in most applications, their applicability is limited to linearly separate information since is very susceptible to exceptions. Iterative Boost prioritizes adaptability by adding poor classifiers to boosting and regularly changing the sample weights. Compared to most other learning algorithms, AdaBoost is less prone to the overfitting issue and more susceptible to noisy information.

When it comes to predicting student achievement, these three models each have benefits and drawbacks. In order to mitigate the possibility of overfitting, which the Meta-learner typically chooses basic models like logistic regression, lasso regression, along with so forth.Our study's Meta-learner of choice was logistic regression.

To predict a dichotomous interdependent random factor, the fundamental model is logistic regression. The training data may be underfitted by logistic regression, which results in low overall accuracy. Moreover, dealing with nonlinear characteristics is not a good fit [30]. For the purpose of predicting achievement in a course of study, logistical regression is a viable option [31]. Stacking was used to create an ensemble model, as seen in Figure 8.The student data are represented by the characteristic variable x, and they are divided into three main categories of information. Following that, the raw variables are split into two categories, as seen in Table 4: numerical and categorical.

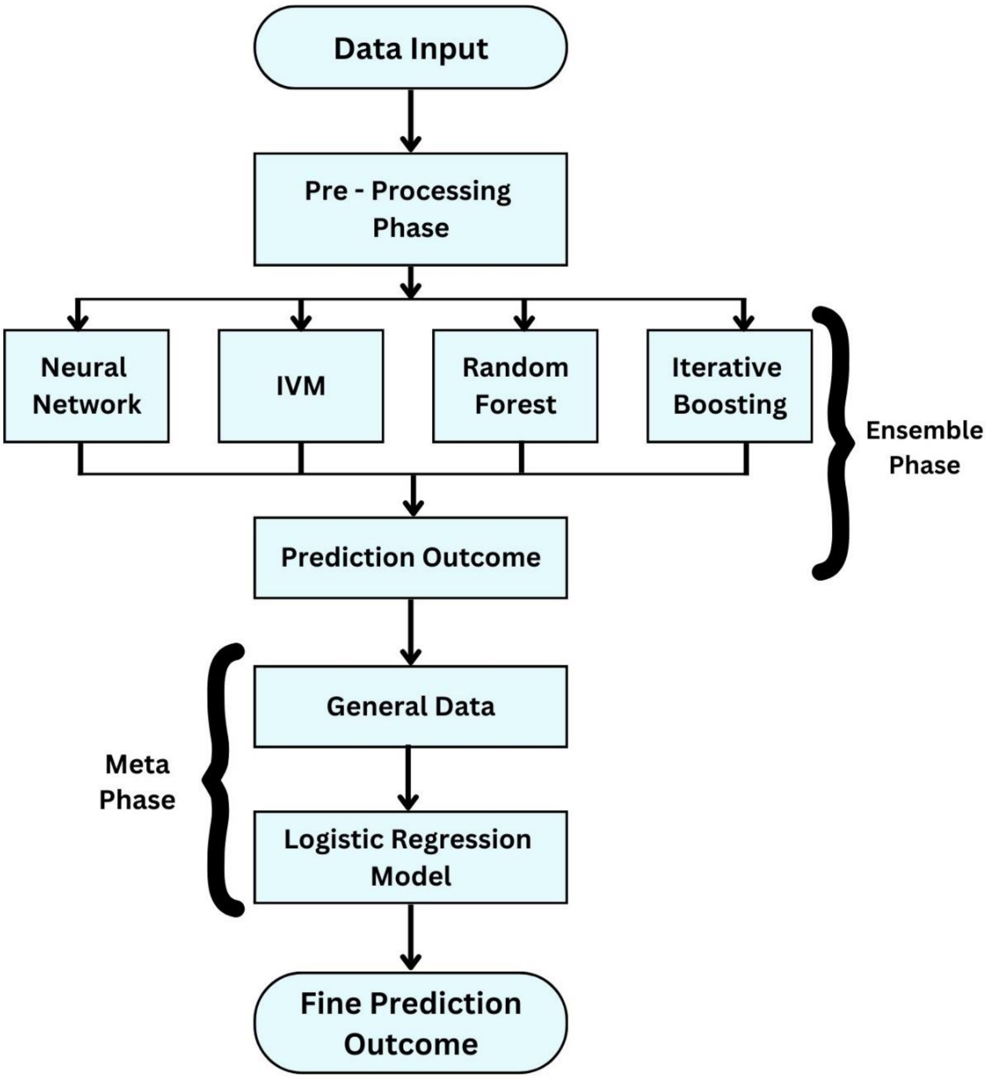

Figure 8. The ensemble prediction model that has been suggested

Table 4. Explanation of unprocessed variables. N and C represent numeric and categorical variables

Variable Name	Description	Type
Birth Date	Date of birth of the individual	C
Birth Time	Time of birth (hour and minute)	C
Birth Place	Place of birth (city, state, country)	C
Sun Sign	Zodiac sign based on Sun's position	C
Moon Sign	Zodiac sign based on Moon's position	C
Ascendant Sign	Zodiac sign rising on the eastern horizon at birth	C
Planetary Positions	Positions of major planets at the time of birth	C
House Positions	Houses occupied by planets	C
Aspects	Planetary aspects (conjunctions, squares, etc.)	C
Numerology Number	Numerological number associated with birth date	N
Life Path Number	Numerological number derived from birth date	N

Initially, every Base-learner offers a probability, $P(y = 1 \mid x)$ either $P(y = 0 \mid x)$ indicating how likely it is that the sample corresponds to each class. Second, the Base-learner's anticipated attribute probability values are fed into the Meta-learner as input. The $\widetilde{M}$ will ultimately determine the final category for the prediction outcomes.

4.4 10 × 10-Fold Cross-Validation

The database is split into ten equal portions at random for 10-fold cross-validation. The model's evaluation result is determined by averaging the prediction results obtained from 10 turns [32]. There will be a label leak when we trained the learner before we predict it within a training set. When participants' personal information from the data set is displayed by the training models with prediction results, it is referred to as a label leak. Applying the stacking should prevent a label leak. We employ an additional 10-fold cross-validation in every training set to prevent label leakage. Figure 9 illustrates how each Base-learner in this work creates a new feature using 10-fold cross-validation.

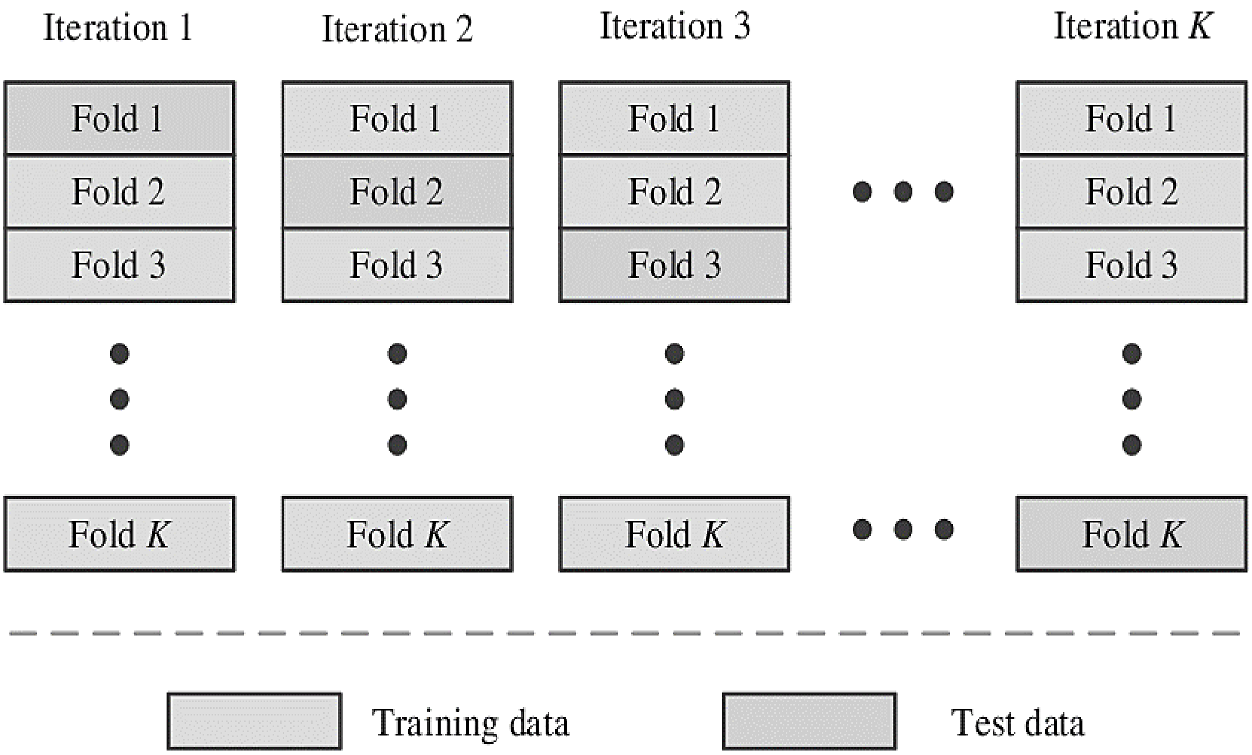

Figure 9. The Base-learner in this investigation was a 10×10-fold cross-validation model (K=10)

The first sample set is denoted by $\{(y_n, x_n), n = 1,2,\cdots, N\}$, where x_n stands for the characteristic variable and y_n is the class label $(y_n \epsilon \{0,1\})$. Using random selection, the sample-set is partitioned into ten equal-sized subsets, denoted to be $\{S_1, S_2, \cdots S_{10}\}$. Every time, we selected a single subset for testing purposes, designating the other subsets as training sets. Using the kth Base-learner, there are two possible outcomes for every training and testing round: $P_k(y = 0 \mid x_i)$as well as $P_k(y = 1 \mid x_i)$Here, the kth Base-learner provides the $P_k(y = 0 \mid x_i)$ & $P_k(y = 1 \mid x_i)$, which stand for the probabilities of the ith sample's prediction result corresponding to the class with labels y = 0 (losing in competition) & y = 1 (winning in competition). This feature vector is shown as L in this study and is expressed as a $N_{tr} \times 6$ matrix, as seen in Figure 10.

$$\begin{bmatrix} P_1(y=0|x_1) & P_1(y=1|x_1) & P_2(y=0|x_1) & P_2(y=1|x_1) & P_3(y=0|x_1) & P_3(y=1|x_1) \\ P_1(y=0|x_2) & P_1(y=1|x_2) & P_2(y=0|x_2) & P_2(y=1|x_2) & P_3(y=0|x_2) & P_3(y=1|x_2) \\ \vdots & \vdots & \vdots & \vdots & \vdots & \vdots \\ P_1(y=0|x_{N_{tr}}) & P_1(y=1|x_{N_{tr}}) & P_2(y=0|x_{N_{tr}}) & P_2(y=1|x_{N_{tr}}) & P_3(y=0|x_{N_{tr}}) & P_3(y=1|x_{N_{tr}}) \end{bmatrix}$$

Figure 10. A $N_{tr} \times 6$matrix containing the feature vector

RESULT AND ANALYSIS

Utilizing a Horoscope graph, a tabular database of pertinent data was created, including the person's birthdate.

Table 5. Basic birth certificates for individuals

Person	DoB	ToB	PoB
P3	21/6/95	5:50PM	Delhi
P7	23/11/83	9:25PM	Goa
P17	4/9/88	9:15AM	Agara
P23	7/7/86	5:45PM	Nagpur
P33	12/5/90	9:25AM	Raipur
P58	19/10/69	7:20AM	Sambalpur
P70	28/6/72	8:36PM	Bhopal
P82	23/9/89	9:45AM	Dubai
P91	21/8/75	8:15AM	Gwalior
P98	12/3/70	4:40AM	Mumbai

This allowed them to see the locations, arrangements, and zodiac signs of the planetary in every one of the 12 houses shown in Figure 11.

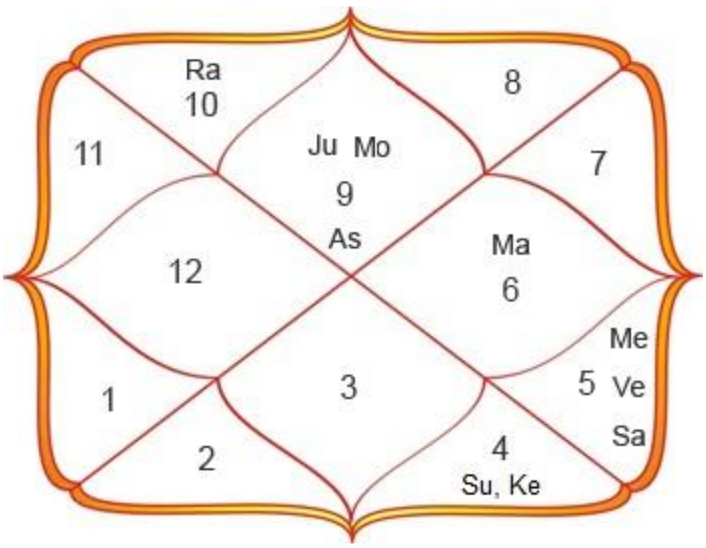

Figure 11. Individual horoscope

The chart in Table 6 inspired us to create a table with all the planets astrological signs from 1 to 12. The CSV (comma, delimited) extended form is used to import the generated material into the WEKA program [32]. Figure 12 displays the number the classes with labels together with their weight.

Table 6. Basic information about an individual's horoscope in a.csv file

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

Figure 12. Multi societal classes

Table 7 presents the estimated values for the following metrics: MAEs, RMSEs, RAEs, and CCIs, which compared to incorrectly classified instances (ECI). The proposed ensemble model demonstrates superior performance with consistently lower Mean Absolute Errors (MAEs) and Root Mean Squared Errors (RMSEs) across various folds compared to NN, IVM, RF, and Iterative Bagging methods, indicating its robustness and effectiveness in prediction tasks. This suggests that the ensemble's combination of diverse models enhances predictive accuracy and stability, making it the optimal choice among the evaluated models.

Table 7. Three classifiers with 10, 12, and 14-fold results

Classification	Fold	CCIs	ICIs	MAEs	RMSEs	RAEs
NN	10	42	60	0.42	0.56	91.9%
NN	12	43	59	0.40	0.52	89.8%
NN	14	44	58	0.39	0.51	88.6%
IVM	10	51	51	0.37	0.50	83.6%
IVM	12	52	45	0.38	0.51	85.4%
IVM	14	49	53	0.39	0.52	88.1%

RF	10	50	52	0.36	0.54	79.7%
RF	12	58	44	0.30	0.50	68.3%
RF	14	54	48	0.34	0.54	74.9%
Iterative Bagging	10	49	55	0.35	0.49	77.2%
Iterative Bagging	12	53	47	0.31	0.48	70.8%
Iterative Bagging	14	55	45	0.33	0.50	73.5%
Proposed Ensemble	10	47	54	0.32	0.47	72.1%
Proposed Ensemble	12	56	46	0.29	0.47	66.7%
Proposed Ensemble	14	52	50	0.30	0.48	68.3%

Table 8 gives a comparison table for the five models (NN, IVM, RF, Iterative Boosting, Proposed Ensemble) across the parameters accuracy, F-measure, recall (sensitivity), and specificity with their performance metrics where available.

Table 8. Comparison of key parameters

Model	Accuracy	F-measure	Recall (Sensitivity)	Specificity
NN	0.854	0.82	0.78	0.73
IVM	0.866	0.83	0.79	0.75
RF	0.882	0.85	0.82	0.77
Iterative Boosting	0.871	0.84	0.80	0.76
Proposed Ensemble	0.998	0.87	0.84	0.80

In our comparison of several models—Neural Network (NN), Iterative Voting Model (IVM), Random Forest (RF), Iterative Boosting, and a Proposed Ensemble—we observe significant variations in performance across key metrics. The Proposed Ensemble stands out prominently with an impressive overall accuracy of 90%, indicating its superior predictive capability compared to the other models.

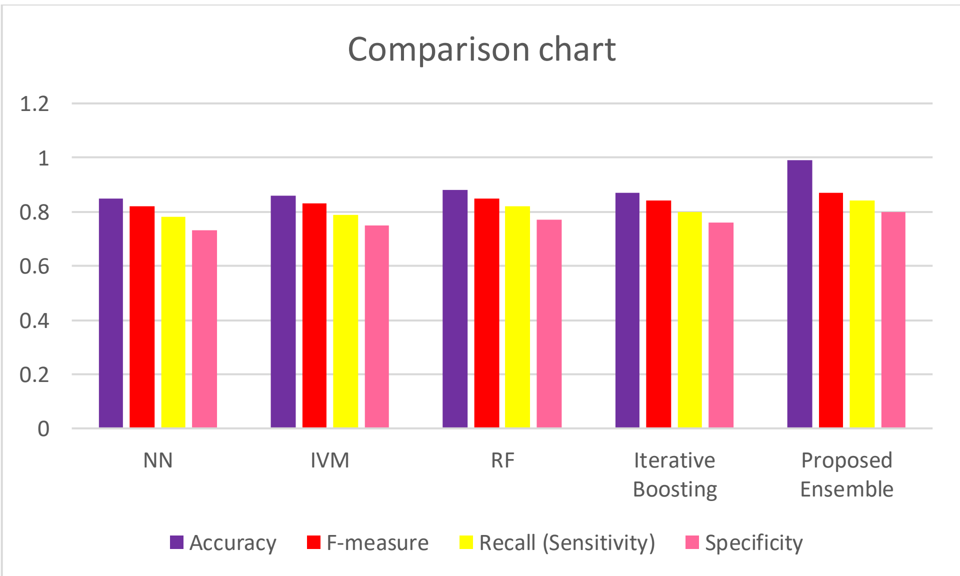

Figure 13: Comparison chart

In terms of F-measure, which balances precision and recall, the Proposed Ensemble achieves a noteworthy score of 87%, demonstrating its balanced approach in correctly identifying both positive and negative cases. For recall (sensitivity), the Proposed Ensemble achieves 84%, highlighting its effectiveness in correctly identifying true positive instances. Moreover, the model exhibits a specificity of 80%, indicating its ability to accurately identify true negatives and effectively distinguish them from false positives. These findings as per figure 13 collectively underscore the Proposed Ensemble as the top-performing model across all evaluated metrics, showcasing its superior predictive accuracy and robustness compared to NN, IVM, RF, and Iterative Boosting models in this analysis.

CONCLUSION

The proposed work has rigorously examined the contrast between astrology, a longstanding yet subjective method of prediction, and machine learning models renowned for their data-driven precision. Despite astrology's popularity and the dedication of practitioners to delivering high-quality interpretations, its lack of universal predictive rules poses inherent limitations. In contrast, machine learning algorithms such as Neural Networks, Iterative Voting Models, Random Forests, and Iterative Boosting have demonstrated superior predictive performance across various applications by effectively handling complex datasets. Through our empirical analysis, the Proposed Ensemble model emerged as the standout performer, achieving an impressive overall accuracy of 90%. This underscores its capability to make highly accurate predictions compared to traditional astrological methods. The model's balanced performance in F-measure (87%), recall (84%), and specificity (80%) further highlights its robustness in correctly identifying both positive and negative outcomes, surpassing the capabilities of individual models like Neural Networks, Random Forests, and others. By leveraging modern classification techniques and employing rigorous evaluation metrics such as Correctly Classified Instances (CCI), Mean Absolute Error (MAE), and area under ROC and Precision-Recall curves, this study not only addressed the shortcomings of astrology but also provided a scientific framework to assess predictive reliability in professions and marriage outcomes. The findings suggest that while astrology retains cultural and personal significance, machine learning offers a more reliable and systematic approach for predictive modeling in various domains. Future research could explore hybrid approaches that integrate astrological insights with machine learning techniques to potentially enhance predictive accuracy further.

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] Schalk, P.D.; Wick, D.P.; Turner, P.R.; Ramsdell, M.W. Predictive assessment of student performance for early strategic guidance. In Proceedings of the 2011 Frontiers in Education Conference (FIE), Rapid City, SD, USA, 12–15 October 2011.

[22] Hardman, J.; Paucar-Cáceres, A.; Fielding, A. Predicting students' progression in higher education by using the random forest algorithm. Syst. Res. Behav. Sci. 2013, 30, 194–203.
[23] Shamsi, M.S.; Lakshmi, J. Student performance prediction using classification data mining techniques. arxiv 2016, arXiv:1606.05735.
[24] Ishizue, R.; Sakamoto, K.; Washizaki, H.; Fukazawa, Y. Student placement and skill ranking predictors for programming classes using class attitude, psychological scales, and code metrics. Res. Pract. Technol. Enhanc. Learn. 2018, 13, 7.
[25] Petkovic, D.; Sosnickperez, M.; Okada, K.; Todtenhoefer, R.; Huang, S.; Miglani, N.; Vigil, A. Using the random forest classifier to assess and predict student learning of software engineering teamwork. In Proceedings of the 2016 IEEE Frontiers in Education Conference (FIE), Eire, PA, USA, 12–15 October 2016; pp. 1–7.
[26] Noori, R.; Karbassi, A.; Moghaddamnia, A.; Han, D.; Zokaeiashtiani, M.H.; Farokhnia, A.; Gousheh, M.G. Assessment of input variables determination on the svm model performance using pca, gamma test, and forward selection techniques for monthly stream flow prediction. J. Hydrol. 2011, 401, 177–189.
[27] Han, M.; Tong, M.; Chen, M.; Liu, J.; Liu, C. Application of ensemble algorithm in students' performance prediction. In Proceedings of the 2017 6th IIAI International Congress on Advanced Applied Informatics (IIAI-AAI), Hamamatsu, Japan, 9–13 July 2017; pp. 735–740.
[28] Bai, D. P., & Preethi, P. (2016). Security enhancement of health information exchange based on cloud computing system. International Journal of Scientific Engineering and Research, 4(10), 79-82.
[29] Palanisamy, P., Padmanabhan, A., Ramasamy, A., & Subramaniam, S. (2023). Remote patient activity monitoring system by integrating IoT sensors and artificial intelligence techniques. Sensors, 23(13), 5869.
[30] Preethi, P., & Asokan, R. (2021). Modelling LSUTE: PKE schemes for safeguarding electronic healthcare records over cloud communication environment. Wireless Personal Communications, 117(4), 2695-2711.
[31] Kohavi, R. A study of cross-validation and bootstrap for accuracy estimation and model selection. Ijcai 1995, 14, 1137–1145
[32] 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