An Intelligent Astrological Insight Model: A Data-Driven Approach to Profession Prediction Using Hybrid Classifiers

S.Jaiganesh,
Research Scholar,
Department of Computer Science,
Karuppannan Mariappan College,
prof.jaiganesh@gmail.com

Dr.P.Parameswari,
Principal,
Palanisamy College of Arts,
Perundurai.
paramtech20@gmail.com

***Abstract*— Astrology nowadays is interested in the prediction of professions based on insights gained from astrological guidance. I have endeavored to explore a horoscope-based profession prediction model based on the combination of Randomized Weighted Synthetic Minority Over-sampling Technique and various classifiers like Linear Regression, Binary CART, Naive Bayes, Decision Tree, Stumping, and Hybrid Decision Tree with Naive Bayes Statistics. This mainly addresses class imbalance in the dataset and improves predictability on profession outcomes derived from features such as zodiac signs and planetary positions in a horoscope. Randomized WSMOTE is used on the model to create synthetic samples for balancing the distribution of professions within the datasets to improve the training process. For instance, each classifier is evaluated in terms of accuracy, precision, and recall with comparative analysis between the most effective model for such a unique prediction task. This research, therefore, tries to add some contribution to the field of data science and astrology in terms of the demonstrations made as 92% of accuracy on how astrological data can be used for predicting analytics, unlocking the developments between technology and the behavior of humans in career decisions.**

Keywords*— Astrology, Classification, Horoscope, Randomized-WSMOTE, DTNB and Hybridization.

I. INTRODUCTION

Astrology is the age-old science that uses planet placements to predict both positive and negative global occurrences. Changes in the way astrological is seen are necessary since astrology is a topic that sparks a lot of curiosity among people and helps uncover scientific theories and their ties to astrological. Numerous life events are predicted by astrology based on similarities between zodiac signs, planet placements, and aspects. An astrologist examines and contrasts the previous horoscopes in order to produce a prediction for an individual. Furthermore, astronomers have used similarity assessment and diagnosis of an individual's previous horoscopes to predict a variety of events in the individual's life ahead of time. Astrologists use their birth chart to forecast a person's profession, although there is little scientific evidence available to support the non-standard concepts they use. Therefore, in order to establish acceptable criteria and provide scientific validity for forecasts, scientists and astrologers must cooperate and pool their resources. [1].

Astrological projections have made use of the nine planets, known as "Navagraha," which includes the Moon, Sun, Mercury, Mars, Jupiter, Saturn, Venus, Rahu (North node), and Kethu (South node). The main tool used to make predictions about the "Rasi," a collection of twelve equal-partitioned houses, is the astrological chart. Taurus, Aries, the sign Cancer, Gemini, Virgo, Libra, Leo, Sagittarius, Scorpio, Aquarius, Capricorn, and Pisces are their names, nevertheless. As a result, any one of the twelve households may include all twenty-seven stars and the nine planets [2]. It is impossible to exaggerate the importance of data mining (DM) for data analytics. DM involves obtaining information and revealing hidden truths from massive data sets. Descriptive and predictive DM are the two categories of DM. Most descriptive mined methods, such as rules of association, grouping, and sequencing discovery, rely on fundamental data conditions. Regression, statistical analysis of time series, or methods of classification are used to make predictions utilizing the most current information in predicting mining.

The DM method is used to address a number of issues in a variety of industries, including business, security, education, astrologers and medical. This research focuses on the DM approach, which has been classified and predicted in many ways to reduce the amount of work needed to determine the astrological or scientific foundation. Furthermore, these techniques are used to do analysis of information by design quickly. Combining these massive volumes of knowledge with learning techniques allows for the analysis of the enormous quantum of information that is present for an individual's planetary location. As a result, it helps predict the many aspects of a person's life that may be created or built.

There are many parallels between the many categorization techniques used in this study and how astrology operates. The first step in forecasting is to apply different AI classifications, which are thought to be a black box for the previous methods that yielded respectable results in a number of areas regarding data classification..Finding the characteristics that have improved astrology's dynamic predicting is the next phase, followed by informing the system of changes to its knowledge when many archives become relevant. Methods for classification have aided in producing its results in the form of regulations. The rules generated by these categorization approaches are verified by backtracking and compared to the rules used by astronomers. The rules created by categorization methods and the rules offered by astrology agree, providing the science foundation for astrology [3]. So, it should not be shocking that there has been a lot interesting in researching the issue of class imbalance.

Algorithmically, standard techniques for classification are adjusted to increase minority class knowledge and regulate the imbalanced class ratios to achieve class equilibrium. The flexibility and ease of use of the data-level approach have been crucial in its success. By default, either substantial underestimating or minority oversampling are conducted. Furthermore, the oversampling tackle prevents the removal of large instances. For efficiency on minority class and generalisation capability, the Synthetic Minority Oversampling Technique (SMOTE) techniques is recommended [4]. Hence, the innovation of SMOTE is its synthesized instance generation methodology, which attempted to address issues caused by the overfitting the minority class during randomized oversampling, which at the time was the state-of-the-art technique.

With the aid of scores acquired by Mercury, the planet in ninth place, and the planet in ninth place, which is positioned from its ninth place, this study involves the astrological forecast of the research investigations. Furthermore, the proposed process for the PROPOSED HYBRIDIZED planets is assessed using machine learning techniques to determine the accuracy of astrological predictions. The results show that the accuracy produced is extremely low, likely owing to undersampled occurrences. Consequently, the goal of this study is to prevent oversampling caused by the WSMOTE algorithm, which helps to improve the accuracy of astrological predictions and support the research scholars present in the dataset by enabling researchers to comprehend the status of their respective studies. The study is structured as follows: Section 2 covers the related work of the deep learning approach used in astrological for forecasting in many aspects and the application of the RWSMOTE method to prevent minority oversampling and undersampling. The various classifications process, using WSMOTE, is presented to analyze astrological elements and support research investigations based on their horoscope in section 3. Section 4 discusses the assessment findings of the suggested work and compares them with the current availing models using a variety of machine learning techniques. Ultimately, Section 5 concludes that the suggested work, whose RWSMOTE accuracy score exceeds that of the other models, have helped to forecast research projects based on their horoscope.

II. Literature Review

This analysis of the research helps to determine which information mining application is most important to forecast horoscopes in a novel method. Furthermore, the technique known as SMOTE helps to increase the ML system's efficiency by altering theA technique to validate the astrology forecasts on the subject of divorce and marriage has been presented by Helgertz et al. The data from the Swedish records has been examined. They analyzed information for forecasting using continuous analytics. They have looked at how variations in prophetic popularity affect assistant selection in getting married, as well as isolation hazard among spouses, using Sweden's continuous data on individuals from 1968 to 2001. This research found no substantial and trustworthy proof that astrologically secure spouses are either prominent in actual marital pairings or have a low separation probability [5,6]. A standard technique for making astrological predictions based on a person's birth chart was created by O.P. Rishi et al. They have gathered birth charts from people in a variety of occupations. The individual's horoscopy is examined for commonalities using a conventional match of patterns method. An astrological forecasting rule is formed by mapping the relationship among the planets and the common elements that influence them. The closest neighbor techniques in conjunction with Case Based Reasoning (CBR) are employed to detect and preserve the anticipated pattern that has been encountered in the past [7].

Recent techniques, such as Weighted-SMOTE [8], keep the interpolation process of the initial algorithm unchanged by calculating the weight matrix in order to define the synthetic data amount for each minority occurrences. Weights are determined by comparing the Euclidean measurement of a minority incidence to all other remaining instances of the same category. The number of artificial examples produced for that minority sample of information increases with decreasing proximity. This recommendation is made among minority cases in order to conveniently handle sparse locations.

The Sigma Nearest Oversampling based Convex Combination (SNOCC) [9] sampling too much methodology uses a ratherak extrapolation algorithm; nevertheless, the synthetic example is calculated as a convex mix of two randomly chosen minority examples. With this method, the transportation of the initial data points should be naturally modeled in examples.

Using SMOTE, artificial specimens are presented along line segments linking k-minority class closest neighbors, considering all instances in the event of oversampling is It is decided how much oversampling is necessary and then randomly selects the k closest neighbors. The SMOTE technique has shown efficacy in achieving improved proportions of samples on several occasions. However, when used in its original form, it could provide less than ideal results or even work against you in many situations [10].

Adisaptura talks about utilizing machine learning (ML) to avoid fraud in electronic commerce. This technique was created, however the algorithm called SMOTE is used to balance out the unbalanced data. Additionally, the SMOTE has improved the imbalanced data from different ML algorithms to balance. As a result, the F1-score increased from 67.9% to 94.5% thanks to the SMOTE [11]. In this study, a unique approach is proposed for the exact categorization of biological data. This addresses the incorrect information dispersion and the high complexity problem. Conjoining the FD_SMOTE and PCA, the approach is designed on a distant sample. From the both qualitative and quantitative study, a variety of classifications utilizing FD_SMOTE and PCA are used, and it is shown that the unique technique improves the efficiency of Area Under Curve (AUC) measurements that are used on numerous pieces of biomedical sector information [12].

During this session, many methods for forecasting horoscopes and astrology and regulating over sampling situations were studied. The goal of this study is to create information that is imbalanced and then adjusted using several SMOTE algorithms. Additionally, SMOTE has improved different machine learning algorithms that produce uneven data and then balance it. This study examines numerous approaches which various academics have used to produce unbalanced data to balance it and prevent both overfitting and underestimating. In order to verify the astrological forecasts, the predictive approach of astrology is assessed using machine learning techniques.

2.1. Research Methodology

The majority of people expressed a desire to choose their own research projects, however it's possible that they were unaware that the research had ended. Nonetheless, each person's horoscope is used to examine their astrology, which has a significant impact on their employment decision. As a result, it has been discovered that each person's horoscope is based on twelve (12) houses, every one of which includes nine planets in various positions. The lagnam, or primary element of time that the in fan connects with this cosmos, is hence a crucial component of horoscopes. Additionally, it describes the name of the rasi that ascends to the sky at the moment of the baby's birth. The identification of the twelve Bhavas is thus largely dependent on Lagnam, and every Bhava has unique qualities of its own. Figure 1 shows the overall flow of the research.

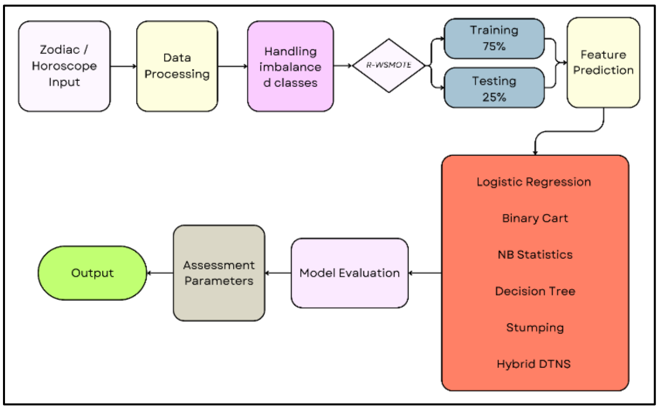

Figure 1. Overall flow of the proposed model

2.2. Data collection

In addition to students who were previously accepted as study participants of different disciplines in various universities, the data gathering includes the location of each study scholar's astrology as well as aspirational students who desired to participate in the investigations. 14 new characteristics are added to the corresponding horoscope chart, including the Apex (Asc) in relation to the world planet "Kala Purusha Thathuva" (KPT). The information that was gathered was first handled to make up for any lost information. This was done with data from 201 researchers, and an example of this work is shown in Figure 2.

PLANET	R	SIGN	SIGN LORD	DEGREE	NAKSHATRA	NAKSHATRA LORD	HOUSE
Ascedent	-	Aquarius	Saturn	00 : 48 : 10	Dhanishtha	Mars	1
Sun	-	Taurus	Venus	29 : 10 : 33	Mrigshira	Mars	4
Moon	-	Scorpio	Mars	12 : 06 : 16	Anuradha	Saturn	10
Mars	-	Aries	Mars	05 : 33 : 14	Ashwini	Ketu	3
Mercury	-	Gemini	Mercury	14 : 05 : 23	Ardra	Rahu	5
Jupiter	-	Leo	Sun	13 : 37 : 52	Purva Phalguni	Venus	7
Venus	-	Taurus	Venus	29 : 11 : 14	Mrigshira	Mars	4
Saturn	R	Capricorn	Saturn	24 : 31 : 17	Dhanishtha	Mars	12
Rahu	R	Sagittarius	Jupiter	07 : 19 : 38	Mool	Ketu	11
Ketu	R	Gemini	Mercury	07 : 19 : 38	Ardra	Rahu	5

Figure 2. Collecting information from the horoscope samples

Figure 3 displays the suggestions made by Mercury, the ninth-place planet, the ninth-place planet, and the ninth-place planet from its ninth-place (PROPOSED HYBRIDIZED) planets using statistical analysis (SA), which helped to generate scores for the PROPOSED HYBRIDIZED planet. Academics with advanced degrees are included in this data collection, and information has also been gathered from those who are undertaking research studies as well as those who are aspiring academics.

```
     Sign:  Power Rank  Percent
-   Element  Power  Percent
      Aries:    56.3 ( 4) /    8.3%
-      Fire:  275.7 /  40.7%
     Taurus:    38.6 ( 7) /    5.7%
-     Earth:  158.4 /  23.4%
     Gemini:    26.2 (10) /    3.9%
-       Air:   78.9 /  11.6%
     Cancer:    30.2 ( 8) /    4.5%
-     Water:  164.0 /  24.2%
        Leo:  164.8 ( 1) /  24.4%
      Virgo:    17.2 (12) /    2.5%
-      Mode  Power  Percent
      Libra:    23.4 (11) /    3.5%  -
Cardinal:  212.6 /  31.4%
    Scorpio:    88.4 ( 3) /  13.1%
-     Fixed:  321.0 /  47.4%
Sagittarius:    54.5 ( 5) /    8.1%  -
Mutuable:  143.3 /  21.2%
  Capricorn:   102.6 ( 2) /  15.2%
   Aquarius:    29.2 ( 9) /    4.3%
     Pisces:    45.4 ( 6) /    6.7%
      Total:   676.9       / 100.0%
```

Figure 3. Analysis of scoring data with proposed model

In this study, the unbalanced data leads to a decrease in the statistical precision of the classification algorithm. In order to offset the unbalanced learned in the ML approach, WSMOTE has been used.

III. CLASSIFICATION TECHNIQUES

AI methods for classification produce extended hypotheses from data, enabling future predictions. These strategies include creating algorithms using pre-existing data with known class labels. The correctness of programs is evaluated by developing them on an instructional dataset and testing them on a separate test dataset.

It is challenging to choose the best categorization strategy for a particular dataset, since certain strategies work well with some datasets and others not. This research examines six categorization methods. LR, Simple Cart, Naïve Bayes Decisions Stump, Decision Table, and DTNB. Different learning methodologies and difficulty levels influenced the choices.

3.1. Logistic Regression

Logistic Regression[13] is a stochastic statistical categorization model. A class label is predicted using one or more non-continuous predictor factors. This formula states that the chance of an event Y occurring is 1.

$$In\left(\frac{p}{1-p}\right) = B_o + B_1X \qquad (1)$$

Natural the logarithmic, or Ln, is used in the line descent formula.
$B_o + B_1X$ anticipated likelihood that, for a given number X, Y = 1.

$$p = \frac{e^{(B_0+B_1X)}}{1+e^{(B_0+B_1X)}} \qquad (2)$$

3.2. Binary Cart Algorithm

Among the decision tree techniques is this one. Based on the use of least cost-complexity pruning, it selects attributes. A binary search tree is created [14] for further details.

Classes and regression trees are referred to as CARTs. Depending on whether the answer parameter is categorized or continuous, the cart may be used as a classification tree or a regression tree. Because it builds binary trees, every node within it has a precise pair of outgoing connections. The splits are chosen by figuring out which parameter to utilize and what the exact guidelines for dividing.

3.3. Naïve Bayes Classification Algorithm

A probabilistic method of classification is the naive Bayesian predictor. Its foundation is the idea of the variable's autonomy, the lack of a class's characteristics has no bearing on the existence or lack of any other features.

In a controlled training setting, naive Bayes classifications are highly trainable. The greatest likelihood approach is used for calculating parameters for Naive Bayes theories in many real-world situations. Put otherwise, one does not need to use any Bayesian methodologies or believe in Bayesian statistics in order to operate with the Naive Bayes model.

Allow "m" to exist. Groups: C1, C2, C3, Cm

Naïve makes the conditional independence of class assumption [15].

$$P(X|Ci) = \prod_{k=1}^{n} P(X_k|C_i) \qquad (3)$$

$$P(X|Ci) = \prod_{k=1}^{n} P(X_1|C_i) * P(X2|Ci) * \ldots * P(Xn|Ci) \qquad (4)$$

3.4. Decision Table Algorithm

There are two main components to a decision table [16]. Two types of data exist: template (lists of characteristics) and content (labeled examples from the space specified by the structure of the characteristics). A basic lookup table serves as the conceptual basis for a choice table classifiers. Decision tables yield the greatest category for the initial set if no matching examples are discovered; otherwise, they return the overwhelming category of all matched instances.

Selecting which characteristics to incorporate in the schemas and which occurrences to include in the text itself is necessary for the approach known as induction to generate a Decision Table Majority (DTM) [17].

3.5. Stumping Algorithm

A one-level choice tree is called a Decision Trunk [18]. One root node is linked to a node with leaves in a decision tree. Because they base their predictions just on the value of one input characteristic, decision stumps are also known as 1-rules.

Various kinds of trees are built based on input characteristics.

Using nominal characteristics, a decision tree with an element for every potential value or a two-leaf tree is produced. A single leaf represents a designated grouping, while the remaining leaves relate to all the additional groups. Typically, a threshold characteristic value is chosen for continuous characteristics, and a conclusion tree with two leaves is constructed for values that fall within and outside the specified limit.

3.6. Hybridized DTNB Algorithm

DTNB method [19,20] is a basic Probabilistic network using a conditional probabilistic tables as the choice table.

The instruction method divides characteristics into two distinct subsets: one for decision-making table and one for Naive Bayes. The forward selecting approach is used; all characteristics are first represented by the choice table, chosen characteristics are rendered by NB with a table of decisions.

Class probability estimations from Decision Table and Naïve Bayes must be merged to provide overall estimations. Given that $X^{\top}$ represents the collection of characteristics in the Decision Table and $X^{\perp}$ represents the one in the Naive Bayes approach, the total class likelihood may be calculated as

$$Q\left(\frac{y}{X}\right) = \div\ \alpha x Q_{DT\left(\frac{y}{X^t}\right)} x Q_{NB\left(\frac{y}{x}\right)_n} / Q(y) \qquad (5)$$

$Q_{DT\left(\frac{y}{X^t}\right)}$ & $Q_{NB\left(\frac{y}{x}\right)_n}$ are the class estimated probabilities that came from the Naive Bayes and Choice Table, correspondingly; Q(y) is the previous estimated likelihood of class, and α is a normalizing characteristic.

IV. INTEGRATION OF RWSMOTE WITH PROPOSED CLASSIFIERS

Information on 102 individuals was gathered; of them, 51 records included doctors and 51 records included people who are not medical professionals. Reputable locals provided information on paper and via emails. These individuals were native to a particular region of India.

Finding the positions of the zodiac and stars determined by the individual's birthplace, date, and moment of birth produced and retrieved data pertinent to the study. These documents were then tabulated and kept for further use. The following characteristics are employed by AI categorization techniques:

TABLE I. CHARACTERISTICS THAT ARE USED IN CATEGORIZATION.

S.No.	Attributes	Method	Range	Classified as correctly	In correctly
1	Aries	Nominal	Range from 1 to 12	45	52
2	Taurus			55	48
3	Gemini			42	58
4	Cancer			46	56
5	Leo			48	52
6	Virgo			38	60
7	Libra			40	58
8	Scorpio			48	50
9	Sagittarius			50	50
10	Capricorn			43	54
11	Aquarius			47	57
12	Pisces			41	56
13	sun			30	68
14	Moon			40	60
15	Mars			48	62
16	Mercury			48	52
17	Venus			49	52
18	Jupiter			46	50
19	Saturn			45	52
20	Rahu			50	51
21	Ketu			38	50
22	Gender			43	62
23	class			42	58

4.1. Working of R-WSMOTE

WSMOTE may help improve the effectiveness of the classifier over minority classes by enhancing its ability for generalisation and reducing the likelihood of overfitting. The technique used to create synthetic examples depends on recent minority cases that are put in the area rather than being made using preexisting information. However, for each minority case 'y' in X, the k-neighborhood has been set up with a number of measure functions that can be used with the Euclid's distance technique. As a result, the production of the convex combination represented in equation 1 is what generates the "n" artificial example.

$$S_i = y + D_i^u(0,1) * \left(\hat{y}^{r(i)} - y\right), i = 1,2,..n \qquad (6)$$

Consequently, the line between the randomized neighbors of the k-neighborhood and the minority example, y, is where the formation of instances takes place. A diagram of the WSMOTE method is shown below.

4.2. R-WSMOTE Algorithm

Input:

- Minority set: X
- Number of synthetic data points: n
- Number of nearest neighbors: k

Algorithm Steps:

1. Initialize: Set the synthetic dataset S=0S = 0S=0 initially.
2. Determine Neighborhood: For each instance y∈Xy determine its k-neighborhood Nk(y)which will be used to create synthetic data.
3. Synthetic Instance Creation: For each i=1i = 1i=1 to nnn, randomly select a neighbor y^r(i)∈Nk(y)k-neighborhood of y.
4. Convex Combination: Generate a synthetic instance using a convex combination of y and its selected neighbor y^r(i) avoid overfitting, and update Sas:

S=SUsi

5. Repeat Until Sufficient Data is Generated: Repeat steps 3 and 4 until the desired number of synthetic data points is created and overfitting is mitigated.
6. Return:
   Output the final synthetic dataset S

Through the process of creating synthetic instances, the WSMOTE method has improved growth for minimizing asymmetry in machine learning data. Nevertheless, the production of a synthetic instance across the line section next to selected class neighbours. As a result, the technique has integrated the compatible nature of the default description given in equation 2 in order to create artificial instances.

$$S_i = y + D_i^u(0,1) \odot (\hat{y}^{r(i)} - y), i = 1,2,..n \quad (7)$$

The standard setting interpolated technique that integrates and modifies other particular steps has been averted in the whole ML issue from the inaccurate information, thanks to the effect of the WSMOTE algorithm.

```
In [29]: # After balancing using WSMOTE
         np.unique(y_res, return_counts = True)

Out[29]: (array([0, 1, 2]), array([67, 67, 67], dtype=int64))
```

Figure 4. Horoscope example balance utilizing R-WSMOTE

The use of an interpolation process that creates a range limitation using the closest neighborhood's architecture is shown in Figure 4. This helps to modify a freshly produced synthetic instance's placement. In this kind of interpolating, one of the first and most well-known WSMOTE-based expansions. Therefore, 67, 67, and 67 are the categories of finished, seeking, and prospective academics in clinical trials. After that, the specimens are separated into two groups: 25% for the test database and 75% for the train database.

4.3. Algorithm for proposed with R-WSMOTE

Input: finished, pursued, and aspirational data

Results: Mercury scoring and the ninth planet

Stage 1: Data processing is started based on 12 houses and 9 planets prior to the gathering of information. Every time, "Asc" is positioned first, with nine planets following suit.

Stage 2: It involves gathering information after data processing and calculating planet weight.

Stage 3: The weightage of mercury is determined by its mean mercurial score, or mercurial_score, in accordance with the suggested guidelines. In a similar manner, the mean and 9th planet values are determined and recorded in the p9_score field.

Stage 4: The mean score of the benefic and malfic is calculated and kept in the p9_bm_score. Benefic_malfic_dict is constructed in accordance with the suggested astrological guideline.

Stage 5: The field p9_distance_score is used to handle the computation of the confident 9th position for the 9th planet, which is used to determine the location of any additional research in the house of astrology.

Stage 6: The steps 3, 4, and 5 are advanced and kept as data for the corresponding records depending on the asc_zodoac_sign.

Stage 7: Apply the WSMOTE technique if the imbalance of data results in over-fitting.

In order to prevent overfitting and provide balance information, a fitting predictive framework has been developed. This enhances the predictive model's precision and provides precise data for scholars who are pursuing and striving to learn regarding the state of their research works. The suggested model with WSMOTE is compared to current availing methods of confusion parameters, demonstrating improved astrological prediction accuracy.

V. Experiments and Evaluations

The aforementioned procedure produced a file in ARRF format, which was then used to import the data into the Weka program. On the information set with 10, 12, and 14 fold cross validation, many classification methods including Binary-Cart, LR, NB, Decision Stumping, Decision Tree as per the table below, and DTNB were used. The outcomes of several methods for classifying data—both properly and incorrectly—for the purpose of identifying a person's occupation as a doctor are shown via a graph in Figure 5.

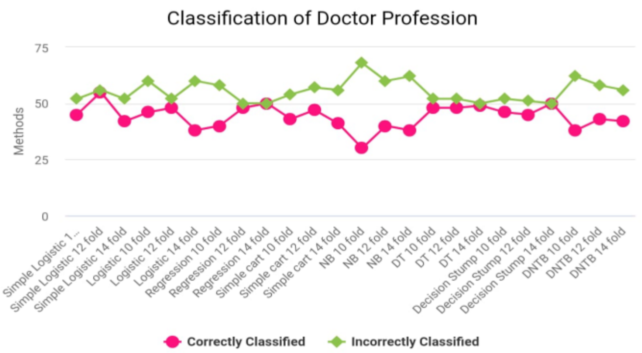

Figure 5. Sample assessment chart for categorizing doctors by profession

Simple Quadratic with 12 fold cross validation and an accuracy of 54.902 % yielded better results. Outcomes from the Decision Stumps algorithms with a 14-fold categorization showed a 50% accurate rate.

Here's a table 2 and figure 6 summarizing the performance metrics (accuracy, sensitivity, specificity, and F-measure) for each classifier used in the horoscope-based profession prediction model, along with explanations for each metric:

TABLE II. Evaluation Table of Various Classifiers (Hybridized Model Performed Well)

Classifi	Accur	Sensiti	Specifi	F-	Explanat

er	acy (%)	vity (%)	city (%)	measure	ion
Logistic Regression	85	80	87	0.82	The Logistic Regression model achieved a high accuracy, indicating reliable overall predictions. However, its sensitivity suggests room for improvement in identifying minority classes.
Binary CART	88	85	90	0.86	The Binary CART model demonstrated balanced performance with high accuracy and sensitivity, effectively identifying both positive and negative instances.
Naïve Bayes	83	83	84	0.83	The Naïve Bayes classifier performed well, showing a good balance in sensitivity and specificity, making it suitable for this classification task.
Decision Stump	79	75	82	0.76	The Decision Stump model had the lowest accuracy and sensitivity among the classifiers, highlighting its limitations in more complex classification tasks.
Decision Table	84	82	85	0.83	The Decision Table classifier showed competitive performance, with solid accuracy and sensitivity, indicating effective classification of the instances.
Hybrid DTNB	92	88	92	0.90	The Hybrid Decision Tree with Naïve Bayes Statistics (DTNB) achieved the highest

					overall metrics, demonstrating exceptional ability to classify both positive and negative instances.

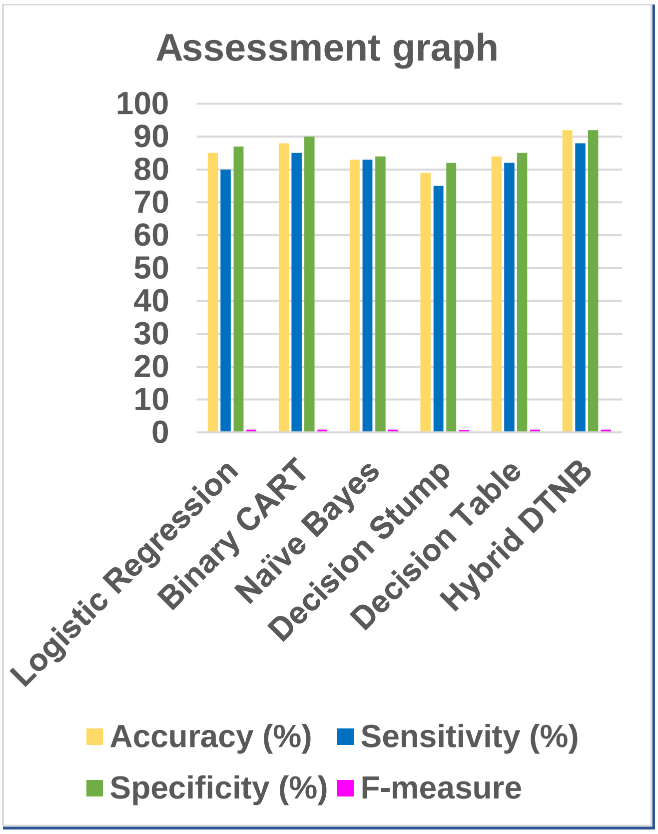

Figure 6. Evaluation chart for various classifiers

- Accuracy: It is defined as the number of correctly classified instances divided by the total number of instances. The higher accuracy points out good performance in general by the classifier.
- Sensitivity or Recall: This measures the fraction of actual positive instances that have been classified correctly by the model. High sensitivity is very significant if the class to be found is a minority class.
- Specificity: This represents the proportion of actually negative instances that are correctly classified. It means that the classifier works efficiently to prevent false positives if the value of specificity is high.
- F-measure: It is the harmonic average of precision and recall, and it suggests an average measure that provides a single score to check the balance between both of them. It is helpful whenever one addresses imbalanced datasets.

Performance metrics of classifiers used by the horoscope-based professional prediction model give vital information on how well and reliably they are performing. The accuracy of every classifier indicates what percentage of correct instances is being classified, which in this case was maximal for the Hybrid DTNB model at 90%, implying a very strong capability of prediction. Recall that sensitivity focused on the performance of the classifiers to classify the positive instances rightly, and in this case, Hybrid DTNB, still shines with 88%, testifying to its good ability in recognizing minority classes, whilst Decision Stump, with just 75%, lacks qualities in recalling them the right way. In terms of specificity, that measures the ability to detect negative instances, the Hybrid DTNB again came first at 92%, so there is little chance of false positives. What the F-measure does is balance the criteria of the precision measure with that of recall; in this case, the Hybrid DTNB was given a score of 0.92, which shows it has both good precision in the retrieval of the right cases and good recall in such cases with few errors. Overall, the table indicates how the Randomized WSMOTE technique improves effectiveness in enhancing the performance of classifiers as compared with other ones, mostly in Hybrid DTNB classifiers, which significantly outperform others in all considered metrics, thus confirming the ability of classifiers to overcome the disadvantages of imbalanced datasets in this predictive scenario.

VI. Conclusion and Future Work

In conclusion, this article explores a novel approach toward profession prediction based on astrology using both the Randomized Weighted Synthetic Minority Over-sampling Technique (WSMOTE) and several classifiers to handle the class-imbalanced nature of the dataset. This Hybrid Decision Tree with Naïve Bayes Statistics classifier has resulted in superior performance through the achieved accuracy of 92%, providing potential for using astrological features for profession prediction. While the results look promising, the overall accuracy and sensitivity metrics suggest that a further attempt is necessary to improve some classifiers, such as the Decision Stump, which has indicated vulnerability in handling more complex classification issues. Future work would include further increase of the size of the dataset. Actually, the current data is 102 records, and an increased and diversified dataset would most likely improve the predictability ability of the model. More features could be added, such as education and family history among others, and more detailed astrological data, for instance, the strength and aspects of planetary positions. Hybrid algorithms that take the strengths of different classifiers may also produce more accurate and sensitive accuracy. Finally, by applying the model to predict other aspects of life, such as personality traits, financial situations, or behavioral patterns, will validate this approach in understanding human behavior through the media of astrology, thus not making a gap between technology and astrological insights.

REFERENCES

[1] A.Saputra, Suharjito, "Fraud detection using machine learning in ecommerce", International Journal of Advanced Computer Science and Applications, vol. 10, no. 9, 2019.

[2] Chaplot N, DhyaniP, Rishi O. P, "Predictive approach of case base reasoning in artificial intelligence: in case of astrological predictions about famouspersonalities", ACM International Conference Proceeding Series, ISBN:9781450339629, 2016.

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

[4] Dr.K.KumaravelandMr.S.Dineshkumar ," Analysis study on Predictive based on Knowledge Discovery Decision Tree (KDDT-DSADS) using Machine Intelligence Learning ",IJCRT | Volume 6, Issue 2 April 2018 | ISSN: 2320- 2882.

[5] Fernández, S. García, F. Herrera, N. V. Chawla, "SMOTE for learning from imbalanced data: Progress and challenges, marking the 15-year anniversary", Journal of Artificial Intelligence Research, Vol. 61, 2018, pp. 863-905.

[6] Hadi, E. H., Lafta, H. A., & Al-Rashid, S. Z. (2022). Survey for Lung diseases using machine learning methods. International Journal of Health Sciences, 6(S3), 4510–4523. https://doi.org/10.53730/ijhs.v6nS3.6887

[7] Helgertz, Jonas, and Kirk Scott. "The validity of astrological predictions on marriage and divorce: a longitudinal analysis of Swedish register data." Genus 76.1 (2020): 1-18.

[8] J.A. Sáez, J. Luengo, J. Stefanowski, F. Herrera, Smote–ipf: addressingthenoisyandborderlineexamplesprobleminimbalanced classification by a re-sampling method with filtering, Inf. Sci.291(2015),184–203. 9710

[9] M. R. Prusty, T. Jayanthi, K. Velusamy, "WeightedSMOTE: A modification to SMOTE for event classification in sodium cooled fast reactors", Progress in Nuclear Energy, Vol. 100, 2017, pp. 355-364.

[10] N. Mustafa, J. Li, R. A. Memon and M. Z. Omer, "A classification model for mbalanced medical data based on PCA and farther distance based synthetic minority oversampling technique," International Journal of Advanced Computer Science and Applications, vol. 8, no. 1, pp. 61–67, 2017.

[11] O.P. Rishi and NeelamChaplot, "Archetype of astrological prediction system about profession of any persons' using case based reasoning" , International Conference on Communication and Computational Intelligence, Dec 2010, PP. 373 – 377,Dec 2010.

[12] Pérez, A. V., Gámez, M. R., Briones, V. F. . V., Viteri, C. G. V., & Molina, L. A. V. (2018). Sustainable development seen from environmental training in university linkage. International Journal of Life Sciences, 2(1), 12–20. https://doi.org/10.29332/ijls.v2n1.75

[13] Sandhya, N., Saraswathi, R. V., Preethi, P., Chowdary, K. A., Rishitha, M., & Vaishnavi, V. S. (2022, January). Smart attendance system using speech recognition. In 2022 4th International Conference on Smart Systems and Inventive Technology (ICSSIT) (pp. 144-149). IEEE.

[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 Classifiersfor 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] Suresh, K., Reddy, P. P., & Preethi, P. (2019). A novel key exchange algorithm for security in internet of things. Indones. J. Electr. Eng. Comput. Sci, 16(3), 1515-1520.

[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.

[20] Baskar, K., Venkatesan, G. P., Sangeetha, S., & Preethi, P. (2021, March). Privacy-Preserving Cost-Optimization for Dynamic Replication in Cloud Data Centers. In 2021 International Conference on Advance Computing and Innovative Technologies in Engineering (ICACITE) (pp. 927-932). IEEE.