



IMPROVE MACHINE LEARNING RESULTS FOR SEMEN ANALYSIS USING ENSEMBLE META CLASSIFICATION

Dr. N.Priya

Assistant Professor, Dept. of Computer Science
SDNB Vaishnav College for Women, Chennai 600 044
Tamil Nadu, India

Abstract: Numerous studies have found that, declining male fertility around the world in the past 50 years are still falling at a rate of two per cent every year. Reasons for this decline range from our increasingly stressful lifestyles, poor diet, drinking and smoking, or a lack of exercise and environmental factors. Applications of machine learning techniques have been implemented in many fields, including health care. This paper have experimented ensemble meta classification techniques such as bagging, boosting and stacking for improving the performance of the Machine Learning algorithms such as Decision-tree (J48), IBK and Naïve-Bayesian (NB) classification for the characterization of seminal quality. The performance of general machine learning classifier is compared with ensemble classifier models and also been verified with their accuracy and error rate.

Keywords : Semen analysis, Machine Learning, Meta Classifier, Ensemble Classification, Decision Tree classification, Naïve-Bayesian Classification, IBK.

I. INTRODUCTION

Over recent years there has been an increasing concern that a decline in male fertility is occurring. There have been many suggestions for the decline in developed countries. Climatic conditions, less exercise, excess weight gain, poor food choices, alcohol and tobacco use; these are just some of the environmental and life style factors that are proven to adversely impact fertility [1][2]. Hence the researchers carried out the semen analysis which is important for the assessment of male fertility potential and it can also be used for the assessment of sperm donors.

Machine learning has been incorporated in various areas including help with medical tasks such as disease identification and diagnosis, personalized treatment and behavioral modification [3][4]. This is run mainly by supervised learning that allows physicians to decide from limited sets of diagnoses or estimate patient risks based on symptoms and genetic information [5]. Machine learning application also helps pharmaceutical companies in their pursuit to find better ways of discovering drugs and manufacturing them [6][7][8].

The main objective of this paper is to improve the machine learning results using ensemble Meta classifier model, as applied to the problem of semen quality categorization. Here male fertility data set has been taken from the UCI data set repository. The rest of the paper is organized as follows. In section 2 data set have been discussed, in section 3 the methodology for ensemble meta classification have been discussed with experiment's design, section 4 shows the analysis of the results. Finally, section 5 contains conclusion with the direction of future work.

II. DATA SET

The fertility data set is taken from the University of California Irvine (UCI) dataset repository. It consists of semen samples, obtained from 100 volunteers, and analyzed

according to WHO 2010 criteria. The data set attributes are based on the fact that sperm concentration is affected by the social demographic and environmental factors, health status and life style habits. The data set can be summarized as follows:

- Number of Attributes: 9 plus the class attribute
- Number of instances: 100
- Missing attribute values: None
- Class distribution: There are 88 normal samples (88%) and 12 altered samples (12%)

Table I presents a description of the attributes and the range of their domain values.

Table I. Attribute description and domain value range

Attribute/Feature	Domain Values
Season in which the analysis was performed	Winter = -1, Spring = -0.33, Summer = 0.33, Fall = 1
Age at the time of analysis (18-36 range normalized to [0,1] range)	18-36 = 0, 36 = 1 and the values in between are calculated as a fraction of 36
Suffered from childhood diseases (e.g. chicken pox, mumps, measles, polio)	Yes = 0, No = 1
Suffered accident or serious trauma	Yes = 0, No = 1
Surgical intervention	Yes = 0, No = 1
Suffered high fever in the last year	Less than 3 months ago = -1, More than 3 months ago = 0, No = 1
Frequency of alcohol consumption	Several times a day = 0.0, Every day = 0.2 Several times a week = 0.4, Once a week =

	0.6 Hardly ever = 0.8, Never = 1.0
Smoking habit	Never = -1, Occasional = 0, Daily = 1
Number of hours spent sitting per day (a day is assumed to consist of 16 hours)	hour = 0, 16 = 1 and the values in between are calculated as fractions of 16
Diagnosis	(N for Normal, O for Altered)

III. METHODOLOGY

A. Ensemble Learning Methods

Ensemble is a technique of combining two or more algorithms of similar or dissimilar types called base learners which incorporates the predictions [3][4]. This will make our final decision more robust, accurate and less likely to be biased.

The goal of ensemble learning methods is to construct a collection of individual classifiers to improve the machine learning results. If this can be achieved, then highly accurate classification decisions can be obtained by voting and averaging the decisions of the individual classifiers in the ensemble models. Three of the most popular techniques for constructing ensemble models are Bagging, Boosting and Stacking. In this paper, these methods operate by taking a base learning algorithm such as Decision Tree Algorithm, Naïve-Bayesian and IBK invoking it many times with fertility training sets.

a) Bagging

Bagging stands for bootstrap aggregation. Given a set, D , of d tuples, bagging works as follows. For iteration i ($i = 1, 2, \dots, k$), a training set, D_i , of d tuples is sampled with replacement from the original set of tuples, D . A classifier model, M_i , is learned for each training set, D_i . To classify an unknown tuple, X , each classifier, M_i , returns its class prediction, which counts as one vote. The bagged classifier, M^* , counts the votes and assigns the class with the most votes to X . Bagging can be applied to the prediction of continuous values by taking the average value of each prediction for a given test tuple.

b) Boosting

Boosting is a sequential technique in which, the first algorithm is trained on the entire dataset and the subsequent algorithms are built by fitting the residuals of the first algorithm, thus giving higher weight to those observations that were poorly predicted by the previous model.

c) Stacking

In stacking multiple layers of machine learning models are placed one over another where each of the models passes their predictions to the model in the layer above it and the

top layer model takes decisions based on the outputs of the models in layers below it. Two layers of machine learning models such as :

- Bottom layer models (d_1, d_2, d_3) which receive the original input features(x) from the dataset.
- Top layer model, which takes the output of the bottom layer models (d_1, d_2, d_3) as its input and predicts the final output.

B. Model Selection

In this paper, three models are selected such as Decision Tree, Naïve-Bayesian and IBK (K-Nearest Neighbor) for classification of fertility data. These are the different types of classification techniques work differently for different datasets. Some techniques give better efficiency for a dataset of very large size but it might not be the optimal technique to use for a dataset with higher number of attributes. Each model is a weak learner which might not be good for the entire dataset but is good for some part of the dataset. Thus, each model actually boosts the performance and accuracy by the ensemble meta classifiers. Hence, Table II shows the accuracy metrics for each classifier.

IV. RESULTS AND DISCUSSION

After improving the accuracy of the classifiers, the best model has been determined from 10-fold cross-validation. 10-fold cross validation is a measure to evaluate the accuracy of the classifiers or predictors in terms of error. It may seem intuitive to select the model with the lowest error rate; however, the mean error rates are just estimates of error on the true population of fertility data cases. Although the mean error rates obtained for three models may appear different, that difference may not be statistically significant.

The true positives, true negatives, false positives, and false negatives are also useful in assessing the costs and benefits (or risks and gains) associated with a classification model. The cost associated with a false negative (such as, incorrectly predicting that an Altered person is Normal) is far greater than that of a false positive (incorrectly yet conservatively labeling a Normal person as Altered Person). Loss functions measure the error between y_i and the predicted value, y_i' .

The most common loss functions are:

$$\text{Mean absolute error : } \sum_{i=1}^d \frac{|y_i - y_i'|}{d}$$

$$\text{Root Mean squared error : } \sqrt{\sum_{i=1}^d \frac{(y_i - y_i')^2}{d}}$$

Table II: Accuracy metrics for classifiers

Meta Classifiers	Percent - Correct	Percent - Incorrect	Mean _ absolute _ error	Root _ mean _ squared _ error	True _ positive _ rate	False _ positive _ rate	True _ negative _ rate	False _ negative _ rate
NB	86.8	13.2	0.21	0.34	0.99	1	0	0.01
IBK	83.6	16.4	0.18	0.38	0.92	0.73	0.27	0.08
J48-	87.6	12.4	0.21	0.33	1	1	0	0
NB- Bagging	80.9	19.1	0.22	0.39	0.91	0.89	0.11	0.09

IBK Bagging	81.9	18.1	0.17	0.38	0.9	0.73	0.27	0.1
J48 Bagging	82.6	17.4	0.18	0.38	0.91	0.81	0.2	0.09
NB- Boosting	87.4	12.6	0.21	0.34	0.99	1	0	0.01
IBK Boosting	82.9	17.1	0.19	0.36	0.91	0.76	0.24	0.09
J48 Boosting	87.9	12.1	0.2	0.32	1	0.99	0.01	0
stacking	87.4	12.6	0.21	0.33	0.99	1	0	0.01

In addition, precision, recall and accuracy metrics to show the performance and comparison of the meta classifiers. Table III summarizes the results of these statistical tests. For NB Classifier, IBK Classifier and J48 Classifier without ensembling, the accuracy rate is 89%, 91% and 93%. But all the three ensemble methods do well against Bagged NB (90%), Bagged IBK (93%), Bagged J48 classifier (95%), Boosting NB (93%), Boosting IBK (93%), Boosting J48 (94%) and Stacking with 93%. Hence, table 3 clearly shows that the accuracy of the machine learning results has been increased due to ensemble meta classifiers.

Table III : Accuracy Comparisons with Machine Learning Classifiers

<i>Meta Classifiers</i>	<i>Precision</i>	<i>recall</i>	<i>Accuracy in %</i>
Naïve Bayesian (NB)	0.88	0.99	89
IBK	0.9	0.92	91
J48-	0.88	1	93
NB- Bagging	0.88	0.91	90
IBK Bagging	0.9	0.9	93
J48 Bagging	0.89	0.91	95
NB- Boosting	0.88	0.99	93
IBK Boosting	0.9	0.91	93
J48 Boosting	0.88	1	94
Stacking	0.88	0.99	93

V. CONCLUSION

In this paper, the popular machine learning techniques NB, IBK and J48 has been proposed for constructing ensemble classifiers using Bagging, Boosting

and Stacking to classify the quality of human semen. Here, the training set is used to evaluate the diagnosis accuracy between the Normal and Altered Person with respect to the relationship of life style habits and environmental factors.

VI. REFERENCES

- [1] David Gil, Jose Luis Girela et.al, "Predicting seminal quality with artificial intelligence methods", Expert Systems with Applications, Vol. 39(16),pp. 12564–12573, 2012.
- [2] Macmillan Simfukwe et.al, "Comparing Naive Bayes Method and Artificial Neural Network for Semen Quality Categorization ", International Journal of Innovative Science, Engineering & Technology, Vol. 2(7), pp. 689-694, 2015.
- [3] Thomas G. Dietterich, "An Experimental Comparison of Three Methods for Constructing Ensembles of Decision Trees: Bagging, Boosting, and Randomization", Vol. 40,pp. 139-157, 2000.
- [4] Quinlan, J. R. , "Bagging, boosting, and C4.5", In Proceedings of the Thirteenth National Conference on Artificial Intelligence, pp. 725–730, 1996.
- [5] Barrat C.L., Clements S., Kessopoulou E, " Semen Characteristics and fertility tests required for storage of spermatozoa", Human Reproduction (Oxford, England), vol. 13(Suppl. 2). pp. 1-11, 1998.
- [6] Koletti P.N, "Evaluation of the subfertile man", American Family Physician, vol. 67 (10), pp. 2165-2172, 2003.
- [7] WHO, WHO laboratory manual for the examination of human semen and sperm-cervical mucus interaction (4th ed.). Published on behalf of the World Health Organization by Cambridge University Press, Cambridge, UK, 1999.
- [8] Lisboa P.J, Taktak A.F.G, "The use of artificial neural networks in decision support in cancer: A systematic review", Neural Networks, vol. 19(4), pp. 408-425, 2006.