SUPERVISED MACHINE LEARNING CLASSIFICATION TECHNIQUES ON MULTI-VALUED DATASETS

Soniya Kharra, M.tech. Scholar, Sobhasaria Engineering College, Sikar
soniya.sec91@gmail.com
Yaswant Soni, Assistant Professor, Sobhasaria Engineering College, Sikar
Yashalw80@gmail.com

Abstract:- In present scenario, across a wide range of operating sectors, large numbers of experts are quickly reaping the advantages of various prophecy schemes to improve the efficiency and consistency of their jobs. Related literature discusses a number of characteristics and shortcomings of available schemes, concluding that each and every available approach always falls short of achieving higher performance, and thus the field has a wide research reach. The incorporated active building mechanism of proposed approach efficiently enhances an act of counsel schemes. Every result of testing phase describes a suitability of designed approach over classical as well as other obtainable methods. 10 folds cross validation mode have adopted to confirm the competence of intended approach.

Keywords: Machine learning, HMM, ANN, SVM, Decision tree, SVM, J48.

I INTRODUCTION

At initial level this investigation work illuminate current hitches of classical and modern offered counsel schemes by considering the huge interconnected literature. Subsequent to discover and explain an associated issues of accessible methods a fresh counsel approach has implement under the environment of JAVA, a most occupied open source programming language. In direction to beat the hindrance of accessible counsel process and accomplish above illustrated set goals the proposed algorithm of this investigation work energetically analyze an act of four dissimilar algorithms, HMM (Hidden Markov Model), SVM (Support Vector Machine), ANN (Artificial Neural Network) and J48 (Naïve form of Decision Tree) with the supremacy of feature assortment process of correlation based feature selection and gain ratio with ranker scheme over an exploited statistics. Subsequent to analyzing process the intended method builds a layered framework where each layer utilize a separated pair of data analyzing process with shrink feature set that has offered by feature assortment schemes.

II LITERATURE SURVEY

A number of related applications of HMM have elaborated by Zhang Youzhi [1]. Besides the illustration of different applications of HMM algorithms the investigator of this work made an endeavor to overcome some associated issues of existing schemes under the field of data mining. Shagun Shaily &Veenu Mangat utilize the process of HMM to offer a naive human activity recognition practice [2]. Shagun Shaily & Veenu Mangat [3] made an endeavor to enhance decoding & learning dilemma of HMM approach with enhanced efficiency in recognition of speech. Janmenjoy Nayaket. al. [4] illustrated a detailed study over the utilization and suitability of SVM method under the field of data mining. The investigation demonstrates the proper detail of SVM practice along with number of restriction and related notation in term to optimize an accessible scheme. This published work significantly aid investigators to recognize the strength and limitation of SVM scheme in direction to intend a more appropriate algorithm. Padmavathi Janardhananet. al. [5] showed the different characteristics of SVM. Additionally this work investigate an act of SVM against other most popular and adoptable classification practices named NB(Naïve Bayes) and RBF over a number of unrelated medical data set. Tian Xia [6] has investigated the act of SVM for education resources classification in course center system of shanghai second polytechnic university. For the evaluation an act of SVM the investigator does a manual effort in term to collect the corpus. The evaluation results have demonstrates that SVM practice has produce an effective outcomes. Shujun Huang [7] discussed a comprehensive study over utilization of SVM practice for Learning in Cancer Genomics. The published work neatly elaborates various characteristics of SVM method in such direction and suggests naïve investigators for further research in direction to intend a more efficient algorithm for the discussed statistics.

Explorer a separate most modern investigation over the related work that has done with the utilization of single and/or multi layer perceptron neural networks. With the simple preface of both single and multi layer neural network the investigation has elaborates the main characteristics and restraints of all accessible algorithms. Kaustubh Patil and Nagesh Jadhav [8] have illustrated a number of recommended approaches by using of various methodologies.

III PROPOSED METHODOLOGY

In direction to build a more efficient counsel scheme and to discover current associated issues of obtainable processes massive related literature has investigated in this incorporated work at initial level. On the base of findings of several associated issues this work marked the key issues that are highly responsible for failure of existing algorithms in scenario of real world. Furthermore a naïve attempt has done to overcome such marked key issues by implementing a more efficient counsel process. The build process exploits an act four different classification algorithms in company of two feature shrinking methods, Correlation Feature Selection with BFS and GR with ranker mechanism.
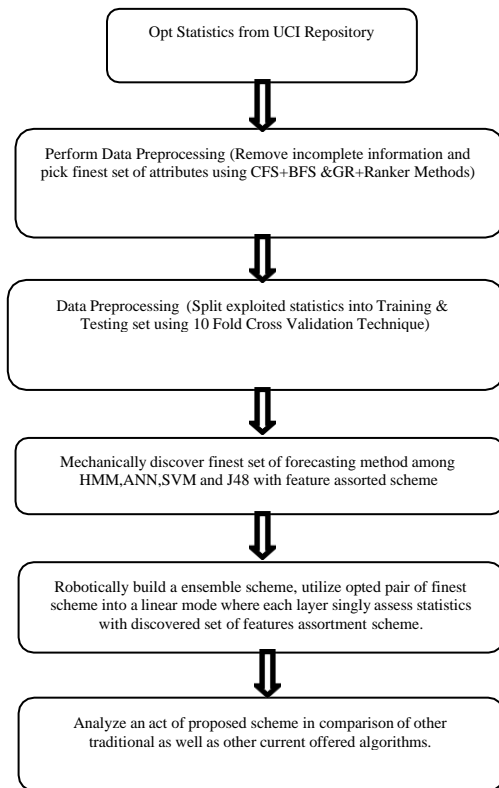
```
┌─────────────────────────────────┐
│  Opt Statistics from UCI         │
│  Repository                      │
└─────────────────────────────────┘
              ↓
┌─────────────────────────────────┐
│ Perform Data Preprocessing       │
│ (Remove incomplete information   │
│ and pick finest set of           │
│ attributes using CFS+BFS         │
│ &GR+Ranker Methods)              │
└─────────────────────────────────┘
              ↓
┌─────────────────────────────────┐
│ Data Preprocessing (Split        │
│ exploited statistics into        │
│ Training & Testing set using     │
│ 10 Fold Cross Validation         │
│ Technique)                       │
└─────────────────────────────────┘
              ↓
┌─────────────────────────────────┐
│ Mechanically discover finest     │
│ set of forecasting method among  │
│ HMM,ANN,SVM and J48 with         │
│ feature assorted scheme          │
└─────────────────────────────────┘
              ↓
┌─────────────────────────────────┐
│ Robotically build a ensemble     │
│ scheme, utilize opted pair of    │
│ finest scheme into a linear mode │
│ where each layer singly assess   │
│ statistics with discovered set   │
│ of features assortment scheme.   │
└─────────────────────────────────┘
              ↓
┌─────────────────────────────────┐
│ Analyze an act of proposed       │
│ scheme in comparison of other    │
│ traditional as well as other     │
│ current offered algorithms.      │
└─────────────────────────────────┘
```

Figure 1 Process Flow Diagram of Proposed Work

The build approach of this investigation work dynamically analyze an act of four adopted method HMM (Hidden Markov Model), ANN(Artificial Neural Network), J48(Decision Tree) and SVM (Support Vector Machine) and with an combination of feature reduction procedure Gain and ration CFS in company of Best first search and ranker mechanism.[9],

Pseudo code of overall working methodology

1. Discover current hitches of accessible counsel scheme with investigation of massive related literature, collected from World Wide Web system.

2. Set the research objective and goals on the base of key issues that are more responsible for failure of offered methods.

3. Accumulate bench marked statistics from research authentic open sources libraries that have frequently utilized by modern investigators for research.

4. Utilize correlation based feature selection and gain ratio process in company of first best search and ranker method to discover the set of only most relevant features.

5. Build a dynamic mechanism that automatically analyzes an act of ANN, SVM, HMM and J48 in company of selected feature separately.

6. Build a layer method vigorously, after discretely compute an act of each process the approach linked up the pair of most suited algorithm into a linear mode.

7. Each process participates in build model as a layer that consumes a separate approach.

8. Evaluate an act of build procedure over same statistics as consumed by other offered method in direction to demonstrate the fair comparison.

9. Authenticate the competence of build method with enclosure the functionality of 10 fold cross validation process.

10. Demonstrate the comparison with the aid of graph in direction to clarify the effectiveness of proposed method in comparison of traditional and current offered methods.

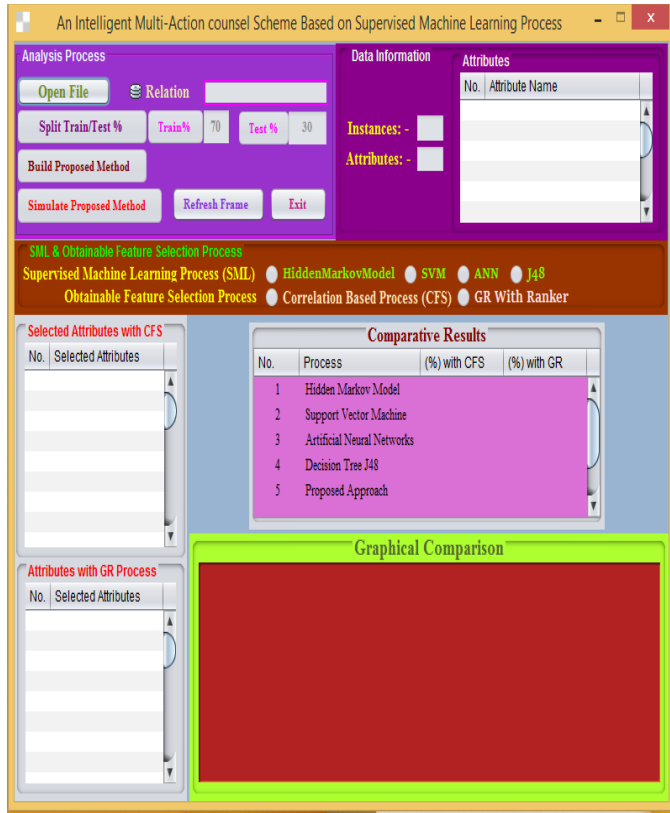The interface of implemented algorithm has depicted in following figures

Figure2 Design of Implemented Application

IV   EXPERIMENT RESULTS

Proposed Approach V/s Classical Techniques

At initial level the act of proposed approach has evaluated against the sole practice act of HMM, ANN, SVM and J48 over labor statistics. The exploited dataset consist 57 instances and 17 features.

TABLE 1
Complete feature set

S.NO.	Attributes Name	S.NO.	Attributes Name
1	Labor_neg_data	10	Shift_differential
2	duration	11	Education_allowance
3	Wage_increase_first_year	12	Statutory_holidays
4	Wage_increase_second_year	13	vacation
5	Wage_increase_third_year	14	longterm_disability_assistance
6	Cost_of_living_adjustment	15	Contribution_to_dental_plan
7	Working_hours	16	Bereavement_assistance
8	pension	17	Contribution_to_health_plan
9	Standby_pay		

TABLE 2
Suggested Attributes by CFS and Gain Ratio(GR) Method over Labor Dataset

	Suggested by CFS			Suggested by GR	
S. NO.	Attributes Name		S. NO.		Attributes Name
1	Wage_increase_first_year		1.		Wage_increase_first_year
2	Wage_increase_second_year		2.		Wage_increase_second_year
3	Cost_of_living_adjustment		3.		tatutory_holidays
4	tatutory_holidays				
5	vacation				
6	Longterm_disability-assistance				
7	Contribution-to-dental-plan				

As stated above that build method has incorporated the functionality of feature reduction scheme to enhance the QOS of intended approach. Hence before starting the actual evaluation process the approach discover only most relative set of feature using CFS and Gain ratio method.

The suggested feature set by mechanism of CFS and GR methods over Labor dataset. The exploited scheme suggest different number of feature set, in comparison of whole 17 feature set the CFS scheme has offered the pair of 7 features while GR suggest to use only 3.

The proposed method adopts the feature set that has offered by the CFS scheme and the functionality of SVM and ANN into a sole practice. The evaluation act depicted in following figure evidently demonstrates that the build mechanism of this investigation work has outperformed in comparison of other evaluated method.

The accuracy graph and comparative values in above figure demonstrate the effectiveness of proposed algorithm in comparison of other classical offered method.
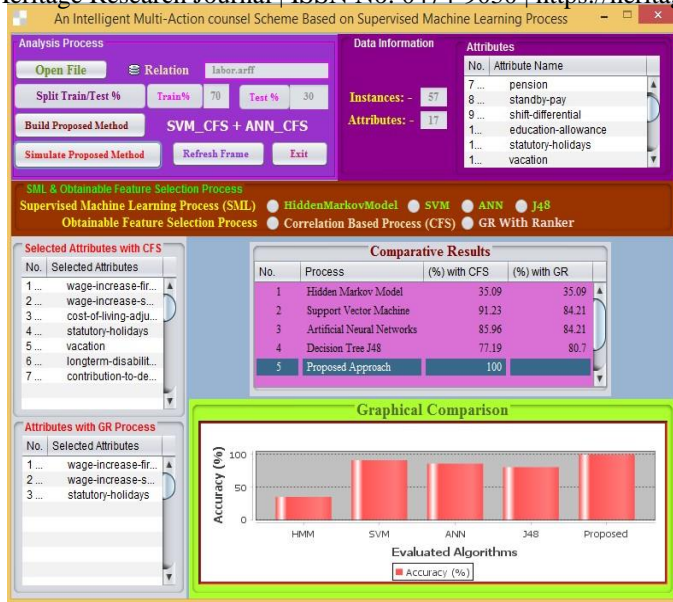
Figure 3 Comparative Act of Proposed & Classical Technique over Labor Dataset

However the results of above experiments improved the efficiency and suitability of proposed mechanism but to confirm the authenticity and flexibility of an intended algorithm it has again evaluated with the other set of statistics, hypothyroid data set.

VII   CONCLUSION

To fill this research gap a fresh endeavor has made in this work that auto build a more efficient counsel scheme automatically analyzing an act of four different algorithms, Support Vector Machine (SVM), Hidden Markov Model (HMM), and Decision Tree (J48) Artificial Neural Network (ANN). The construct method thoroughly examines an act of these four methods over exploited statistics and selects a pair of the best algorithms to use in liner mode with the company of feature selection process similarity based feature selection process with best first scan and Gain ratio with Ranker method. However, the evaluation findings clearly show the supremacy of the proposed algorithm over other available algorithms and other current methods in terms of achieved precision, however, as with other available processes; the construction phase of this investigation work has also been associated with some particular restrictions.

REFERENCES

[1]   Zhang Youzhi "Research and Application of Hidden Markov Model in Data Mining" Second IITA International Conference on Geoscience and Remote Sensing, IEEE, 2010, pp.-459-462.

[2]   ShagunShaily, VeenuMangat "The Hidden Markov Model and its Application to Human Activity Recognition" 2nd International Conference on Recent Advances in Engineering & Computational Sciences (RAECS), 2015, pp.-1-6.

[3]   MbarkiAymen, Ammari Abdelaziz, SghaierHalim, HassenMaaref "Hidden Markov Models for Automatic Speech Recognition" International Conference on Communications, Computing and Control Applications (CCCA), IEEE, 2011, pp. 1-6.

[4]   JanmenjoyNayak, BighnarajNaik and H. S. Behera" A Comprehensive Survey on Support Vector Machine in Data Mining Tasks: Applications & Challenges" International Journal of Database Theory and Application Vol.8, No.1 (2015), pp.169-186.

[5]   Padmavathi Janardhanan, Heena L., and FathimaSabika "Effectiveness Of Support Vector Machines In Medical Data Mining" Journal Of Communications Software And Systems, Vol. 11, No. 1, March 2015, pp.- 25-30.

[6]   Tian Xia "Support Vector Machine Based Educational Resources Classification" International Journal of Information and Education Technology, Vol. 6, No. 11, November 2016, pp.- 880-883.

[7]   Shujun Huang, NianguangCai, Pedro Penzuti Pacheco, ShaviraNarandes, Yang Wang And Wayne Xu "Applications of Support Vector Machine (SVM) Learning in Cancer Genomics" Cancer Genomics & Proteomics 15: 2018, pp.- 41-51.

[8]   Jaswinder Singh, Rajdeep Banerjee "A Study on Single and Multi-layer Perceptron Neural Network" Third International Conference on Computing Methodologies and Communication (ICCMC), IEEE, 2019, pp.- 35-40.

[9]   Arti Rana, Arvind Singh Rawat, AnchitBijalwan and Himanshu Bahuguna "Application of Multi Layer (Perceptron) Artificial Neural Network in the Diagnosis System: A Systematic Review" International Conference on Research in Intelligent and Computing in Engineering (RICE), IEEE, 2018, pp.- 1-6.