Applying Government Schemes in Rural Sectors Prediction System for Evaluation of Data Science Algorithm

S. Maheswari, S. Kalaiselvi, D. Thamarai Selvi, M. Manochitra

***Abstract*: The administration dispatches different aggressive projects attempting to make the nation more prosperous, yet what they bomb is in fruitful execution and coming to recipients. The fundamental explanation for this issue is the absence of mindfulness among rustic individuals. This paper is to give an answer for this uninformed circumstance. Through this framework the rustic understudies will be instructed such that they can become acquainted with about what are the different plans that are outfitted by the administration and what are the plans they are qualified for. On the off chance that the country understudies came to know and get mindful of the apparent multitude of legislative plans gave by the Government of India for the government assistance of the provincial understudies, at that point their life would venture into next level. At first this framework will investigate the accessible government plans in the instructive for the government assistance of country understudies. Next, the understudy's information ((i.e.) name, age, station, occupation, annualincome.etc) are accumulated. At that point; both the datasets are brought into the Anaconda Navigator. At that point, investigation and grouping dependent on networks (SC, ST, BC and MBC) of the understudies and the plans are performed. At that point utilizing the forecast calculations (Naïve Bayes, Random Forest and Support Vector Machine (SVM)) what are generally the plans the specific understudy is qualified for are anticipated. An investigation is made on the proficiency of the three calculations. The exactness of the three calculations is broke down and the effective calculation which creates the outcome with most elevated precision is at last used to play out the forecast of the plans that a specific understudy is qualified for. At long last, the anticipated plans anticipated utilizing the most elevated effective calculation among the three calculations will be gotten back to the understudies. Hence, through this undertaking the rustic understudies will come to think about different recipient plans gave by government and they can use those plans for the improvement of the country environmental factors*

***Keywords*: Prediction System, Naïve Bayes, Random Forest, Prediction Algorithm, Anaconda Navigator.**

I. INTRODUCTION

Presently a days, government dispatches different driven projects attempting to make the nation more prosperous, yet what they bomb is in fruitful usage and coming to recipients. The principle explanation for this issue is the absence of mindfulness among rustic individuals. This paper is to give an answer for this ignorant circumstance. Through framework the country understudies will be instructed such that they can become more acquainted with about what are the different plans that are outfitted by the legislature and what are the plans they are qualified for On the off chance that the provincial understudies came to know and get mindful of the apparent multitude of legislative plans gave by the Government of India for the government assistance of the rustic understudies, at that point their life would venture into next level. At first this framework will investigate the accessible government plans in the instructive for the government assistance of country understudies. Next, the individuals' information ((i.e.) name, age, standing, occupation, annualincome.etc) are assembled. At that point; both the datasets are brought into the Anaconda Navigator. At that point, investigation and arrangement dependent on networks (SC, ST, BC and MBC) of the clients and the plans are performed. At that point utilizing the three forecast calculations (Naïve Bayes, Random Forest and Support Vector Machine (SVM)) what are generally the plans the specific understudy is qualified for are anticipated. An examination is made on the productivity of the three calculations (Naïve Bayes, Random Forest and Support Vector Machine (SVM)). The precision of the three calculations (Naïve Bayes, Random Forest and Support Vector Machine (SVM)) is investigated and the proficient calculation which creates the outcome with most noteworthy exactness is at long last used to play out the forecast of the plans that a specific understudy is qualified for. At last, the anticipated plans anticipated utilizing the most noteworthy proficient calculation among the three (Naïve Bayes, Random Forest and Support Vector Machine (SVM)) will be gotten back to the understudies. Consequently, through this venture the rustic understudies will come to think about different recipient plans gave by government and they can use those plans for the advancement of the country environmental factors.

A. Naïve Bayes Algorithm

This calculation is a simple technique for building classifiers: models that set class marks to issue occurrences, spoken to as vectors of highlight esteems, where the class names are stressed from some limited set. There is anything but a lone calculation for preparing such classifiers, however a group of calculations dependent on a comparative standard: all credulous Bayes classifiers assume that the estimation of a particular characteristic is free of the estimation of some other property, given the class variable.

Revised Manuscript Received on December 30, 2020.

***** Correspondence Author

Dr.S.Maheswari, CSE, National Engineering College, Kovilpatti, Thothukoodi, India. Email: maheswaricse@nec.edu.in

S.Kalaiselvi, CSE, National Engineering College, Kovilpatti, Thothukoodi, India. Email: sks@nec.edu.in

D.Thamarai Selvi, CSE, National Engineering College, Kovilpatti, Thothukoodi, India. Email: dts-cse@nec.edu.in

M.Manochitra, CSE, National Engineering College, Kovilpatti, Thothukoodi, India. Email: manochitra-cse@nec.edu.in

For example, an organic product might be viewed as an orange in the event that it is orange in shading, round, and around 10 cm in distance across. A Naive Bayesian calculation thinks about every one of these traits to give unequivocally to the likelihood that this natural product is an orange, paying little heed to any possible connections between's the shading, roundness, and measurement highlights. For certain sorts of likelihood models, Naive Bayes classifiers can be up gifted efficiently in a managed getting the hang of setting. In various continuous applications, variable assessment for innocent Bayes models uses the method of most extreme probability; in extra words, a solitary can work with the credulous Bayes model without gaining Bayesian likelihood or utilizing any Bayesian procedures.

B. Random Forest Algorithm

A Random Forest calculation is a versatile, easy to use AI calculation that builds, even without hyper-boundary tuning, a prominent result a large portion of the cases. It is additionally one of the best used calculations, as its coherence and the component that it very well may be used for both order and relapse undertakings. Arbitrary Forest is a managed learning calculation. According to the name, it makes a woodland and makes it by one way or another irregular. The "backwoods" the calculation develops, is a gathering of Decision Trees, practically the entirety of the time up gifted with the "packing" technique. The general aim of the stowing strategy is that a mix of learning models expands the comprehensive outcome. Arbitrary woodland builds a few choice trees and consolidates them to get a more exact and stable forecast.

C. Support Vector Machince

Throughout AI, uphold vector machines (SVMs, likewise uphold vector organizations) are managed learning models with comparing learning calculations that look at information utilized for characterization and relapse examination. Expressed a bunch of preparing models, each recognized as having a place with either of two classes, a SVM preparing calculation builds a model that allots new cases to one class or the other, making it a non-probabilistic twofold direct classifier (despite the fact that strategies, for example, Platt scaling stay alive to use SVM in a probabilistic characterization setting). A SVM model is a depiction of the example as focuses in space, planned so the cases of the various classes are partitioned by an unmistakable hole that is as wide as could reasonably be expected. Most recent occurrences are then planned into that equivalent space and anticipated to have a place with a gathering dependent on which purpose of the chink they fall. In consideration to executing straight characterization, SVMs can profitably play out a non-direct arrangement utilizing what is known as the bit stunt, in a roundabout way planning their take in contributions to high-dimensional component spaces.

II. PROBLEM STATEMENT

As of now government dispatches different driven projects attempting to make our nation more prosperous, yet what they bomb is in effective usage and coming to recipients. A nation where ranchers actually end it all, absence of value training, and different danger of Indian culture – recounts a frustrating story that regardless of the presence of advancement plans, India is a long ways behind in making an amicable and prosperous society. The principle purpose for this issue is the absence of mindfulness among provincial individuals. Indian government spends almost Rs. 2 Lakh Cr on 100 leader programs zeroed in on scope of public assistance programs like MNREGS and Sarva Shiksha Abhiyan and other comparable plans. Practically 72% of the rustic family units in an overview held in 2011 for 7 states didn't know about the India's biggest lead program MGNREGS.

The legislature flops in making a mindful environment where individuals comprehend what the administration is accomplishing for them. On the off chance that little plans were reviewed, it may have delivered additionally frustrating outcomes. In the event that the rustic individuals came to know and get mindful of the apparent multitude of legislative plans gave by the Government of India for their government assistance, at that point their life would venture into next level. Consequently, this task is to give an answer for this unconscious circumstance by making the rustic individuals to get mindful about different recipient government plans and to use them in a successful manner.

III. PROBLEM STATEMENT

[1] This paper proposed a comparative analysis of four machine learning algorithms i.e., J48-Naïve Bayes, Random Forest and Multilayer Perceptron is presented in this paper. For the attribute reduction, the CFSSubset Eval is used. It has been shown that J48 is performing best among all the algorithms for the detection of Dementia. From the paper it is clearly concluded that J48 decision tree is the effective among other machine learning algorithms to perform prediction.[2] proposed a Comparative Study of Different Machine Learning Algorithms for Disease Prediction. Using different machine learning algorithms and further based on the performance of these algorithms on two medical datasets i.e. breast cancer and diabetes; we conclude which machine learning algorithm is better for disease detection at early stage. From the paper it is clearly concluded that J48 decision tree is the effective among other machine learning algorithms to perform prediction. [3] Compares three of the most popular ML techniques commonly used for breast cancer detection and diagnosis, namely Support Vector Machine (SVM), Random Forest (RF) and Bayesian Networks (BN). From the paper it is clearly concluded that Random Forest is the effective among other machine learning algorithms to perform prediction. [4] Proposed how to diagnose the disease asthma with Expert Systems, identifying using Machine learning algorithms such as Auto-associative memory neural networks, Bayesian networks, ID3 and C4.5. From the paper it is clearly concluded that J48 decision tree is the effective among other machine learning algorithms to perform prediction.[6]proposed a comparative study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random Forests.

From the paper it is clearly concluded that the objective is to introduce a random forests (RFs)-based prognostic method for tool wear prediction as is effective among others in prediction. [8] Compares different techniques that can be used for each of these three subtasks and discusses the main advantages and disadvantages of each technique. From the paper it is clearly concluded that alternative methods for algorithm selection, such as the combined F-test 5x2 cross-validations and nested cross-validation, are recommended for comparing machine learning algorithms when datasets are small.

IV. PROPOSED DIAGRAM FOR THE PREDICTION SYSTEM

Right now, there are different gainful government conspires that provincial individuals don't know about. This imaginative thought proposition is to give an answer for this ignorant circumstance. Through this framework the rustic individuals will be instructed such that they can become more acquainted with about what are the different plans that are outfitted by the administration and what are the plans they are qualified for. In this venture, at first we'll investigate the accessible government plans for the government assistance of provincial individuals. At that point we'll make a dataset utilizing the subtleties of the plans gathered. Next, we'll accumulate the understudy's information [(i.e.) name, age, standing, occupation, yearly pay of their parent.etc]. At that point; the understudy's dataset is made utilizing those subtleties gathered. At that point; both the datasets are brought into the Anaconda Navigator. At that point, we perform examination and order dependent on networks (SC, ST, BC and MBC) of the clients and the plans. At that point utilizing three expectation calculations (Naïve Bayes, Random Forest and Support Vector Machine (SVM)) the forecast framework will foresee what the plans the specific understudy is qualified for are. An investigation is made on the productivity of the three calculations (Naïve Bayes, Random Forest and Support Vector Machine (SVM)). The exactness of the three calculations (Naïve Bayes, Random Forest and Support Vector Machine (SVM)) is examined and the effective calculation which creates the outcome with most noteworthy precision is at long last used to play out the expectation of the plans that a specific understudy is qualified for. At last, the anticipated plans anticipated utilizing the most noteworthy effective calculation among the three (Naïve Bayes, Random Forest and Support Vector Machine (SVM)) will be gotten back to the understudies. Subsequently, through this task the country understudies will come to think about different recipient plans gave by government and they can see what the plans everyone is qualified for are for the most part and they can use those plans for the improvement of the provincial environmental factors.

A. Module 1: Preprocessing

In this module the different government plans gave by the administration of India for the government assistance of country individuals (for example the country understudies) are investigated. At that point, utilizing those plans 2 datasets (understudy dataset which contain understudy subtleties and plans dataset which contain the subtleties of the administration plans) will be made and brought into the Anaconda Navigator. The understudy dataset comprises of traits as name, age, instructive classification, leading group of college, network, sexual orientation, parent's occupation ,parent's income,day researcher hosteller, genuinely tested, merit understudy and endeavors (the quantity of endeavors a specific understudy take to handle a class). The understudy dataset comprises of 13 ascribes and 510 examples. While, the plans dataset comprises of the qualities, for example, conspire name, instructive class, leading body of college, endeavors, parent's occupation, parent's pay, day researcher hosteller, additional inclination for incapacity, additional inclination for merit understudies, additional inclination for young ladies, grant sum and age. The plans dataset comprises of 13 ascribes and 40 occurrences. Later these put away two datasets will be utilized for examination, grouping and expectation framework.

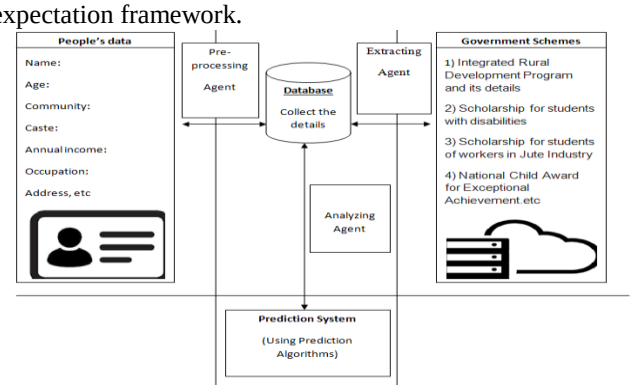

Fig. 1 Proposed diagram for the prediction system

B. Module 2: Data Transformation

Boa constrictor Navigator (utilizing the plan and understudy subtleties dataset) does the function of information change through the assistance of specialists. It has bidirectional relationship with gathering client subtleties and plans (dataset) and the expectation framework. Utilizing Anaconda Navigator, we perform investigation and order dependent on networks (SC, ST, BC and MBC) of the clients and the plans. It sends the information and returns the outcomes to the suitable modules.

C. Module 3: Prediction System

From the information got from the Anaconda Navigator (datasets), the expectation framework looks at the understudy subtleties from the understudy dataset with the plans from the plans dataset and predicts the resultant plans alongside the subtleties which the understudy is qualified for by utilizing the three forecast calculations.

- Naïve Bayes Algorithm
- Random Forest
- Support Vector Machine (SVM)

D. Module 4: Outcome Delivery

The plans anticipated by the forecast framework will be gotten back to the client through the examining specialist. At long last, the anticipated plans anticipated utilizing the most elevated proficient calculation among the three (Naïve Bayes, Random

Forest and Support Vector Machine (SVM)) will be gotten back to the understudies. The ultimate result comprising of the apparent multitude of qualified plans alongside its subtleties will be gotten back to the understudy

V. IMPLEMENTATION

A. Data Collection

Define abbreviations and acronyms the first time they are used in the text, even after they have been defined in the abstract. Abbreviations such as IEEE, SI, MKS, CGS, sc, dc, and rms do not have to be defined. Do not use abbreviations in the title or heads unless they are unavoidable.

At first in this paper the different government plans gave by the legislature of India for the improvement of the country understudies are inspected and utilizing those plans, two datasets are made. The first dataset is the understudy dataset which contains the traits, for example, name, age, instructive classification, leading body of college, network, sex, parent's occupation ,parent's income ,day researcher hosteller, genuinely tested, merit understudy and endeavors (the quantity of endeavors a specific understudy take to handle a class). The understudy dataset comprises of 13 ascribes and 510 occasions. Hardly any example cases of the first dataset (understudy dataset) are referenced beneath in table I.

The following dataset is the plans dataset which comprises of the qualities, for example, plot name, instructive class, leading body of college, endeavors, parent's occupation, parent's pay, and day researcher hosteller, additional inclination for handicap, additional inclination for merit understudies, and additional inclination for young ladies, grant sum and age. The plans dataset comprises of 13 credits and 40 cases. Hardly any example occasions of the second dataset (plans dataset) is referenced underneath in table II.

B. Data Transformation

The collected two datasets are then imported into the Anaconda navigator using the functions such as numpy, pandas, matplotlib. pyplot, seaborn and mpl_toolkits. While importing the two datasets (student dataset and schemes dataset) the datasets are encoded into Latin format.

These stored two datasets are used for analysis, classification and prediction system. In the analysis phase the mean value and standard deviation value are calculated using the following formulae.

$$\text{Mean} = (\Sigma x_i) / n \qquad \text{Equation(1)}$$

$$\text{Standard Deviation} = \sqrt{\sum} (x - \tilde{x}) / \sqrt{n} \qquad \text{Equation (2)}$$

The instances in the table 3 and 4, Mean and standard deviation are obtained using the equation (1) and equation (2).

The figure 2 envisions the analysis done based on the number of educational schemes a particular community consists of.

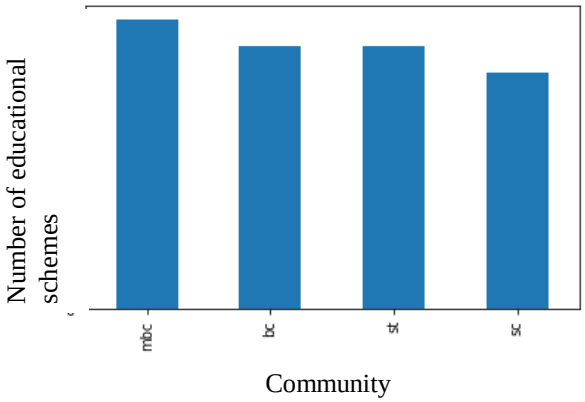

Fig.2 Analysis of number of educational schemes in each community

The x-axis of the figure 2 envisions the various communities of the students such as SC, ST, MBC and BC, whereas the y-axis of the figure 2 envisions the number of educational schemes for each community. From the figure 2 it is intelligible that there are around 10 educational schemes for each community students. The educational schemes for the SC and ST students are Pre-Matric Scholarship to the SC Students studying in classes IX & X, Pre-Matric Scholarships to the Children of those Engaged in occupations involving cleaning and prone to health hazards, Post-Matric Scholarship for SC students , Up gradation Of Merit Of SC Students, National Overseas scholarship, Central Sector Scholarship of Top Class Education for SC Students, National Fellowship for Scheduled Caste Students, BABU JAGJIVAN RAM CHHATRAWAS YOJANA, Free Coaching Scheme for SC and Credit Enhancement Guarantee Scheme for the Scheduled Castes.

Table: I Student dataset

name	age	educational category	board of university	Community	Gender	parents occupation	parents income	day scholar hosteller	physically challenged	merit student	attempts
Malika	14	Pre-Matric	Govt	SC	F	Match Industry	40000	Day scholar	No	Yes	1
Aspathi	8	Pre-Matric	Govt Aided	SC	F	Painter	50000	Day scholar	No	No	2
Aarthi	5	Pre-Matric	Govt	MBC	F	Colli	20000	Day scholar	No	No	1
Barani	16	Post-Matric	Govt	BC	M	Milk Buisness	70000	Hostellar	No	Yes	1
Pradeesh	17	Post-Matric	Govt	ST	M	Colli	15000	Dayscholar	Yes	No	1

Table: II Schemes dataset

Scheme Name	Educational category	Boardof University	Attempts	Parents Occupation	ParentsIncome	Day scholar / Hosteller	Extrapreference for disability	Extrapreference for merit students	Extra Preference for girls	Scholarship amount	Age
Pre-Matric Scholarship to the SC Students studying in classes IX & X	Pre-Matric	Govt and Govt Aided	1	Any Occcupation	250000	Any Category	No	No	No	1525	Not Mentioned
Pre-Matric Scholarships to the Children of those Engaged in occupations involving cleaning and prone to health hazards	Pre-Matric	Govt and Govt Aided	1	Manual Savangers,tanners,Flayers,Hazordous Cleaning	0	Any Category	No	No	No	1700	Not Mentioned
Incentives for Rural MBC/DNC girl students	Pre-Matric	Govt and Govt Aided	1	Any Occcupation	25000	Any Category	No	No	Yes	500-1000 per year	Not Mentioned
Boarding Grants	Pre-Matric	Private only	1	Any Occcupation	50000	Hostellar	No	No	No	2000 per year	Not Mentioned
Free Hostels Scheme	Pre-Matric	Govt and Govt Aided	1	Any Occcupation	50000	Hostellar	No	No	No	Free Hostel Fees	Not Mentioned

Table III Analysis of the student and schemes datas

```
In [1]: import numpy as np
        import pandas as pd
        import matplotlib.pyplot as plt
        import seaborn as sns
        import mpl_toolkits
        %matplotlib inline

In [2]: student = pd.read_csv("D:\student.csv",encoding='latin-1')

In [3]: schemes = pd.read_csv("D:\schemes.csv",encoding='latin-1')

In [4]: student.head()

Out[4]:
```

	sno	name	age	educationalcategory	boardofuniversity	community	gender	parentsoccupation	parentsincome	dayscholarhosteller	physi
0	1	Malika	14	Pre-Matric	Govt	SC	F	Match industry	40,000	Dayscholar	No
1	2	Aspathi	8	Pre-Matric	Govt Aided	SC	F	Painter	50,000	Dayscholar	No
2	3	Saranya	10	Pre-Matric	Govt	SC	F	Colli	15,000	Dayscholar	No
3	4	Malathi	12	Pre-Matric	Govt	SC	F	Colli	20,000	Dayscholar	No
4	5	Kaniraj	16	Post-Matric	Govt	SC	M	Colli	25,000	Dayscholar	No

```
In [5]: student.tail()

Out[5]:
```

	sno	name	age	educationalcategory	boardofuniversity	community	gender	parentsoccupation	parentsincome	dayscholarhosteller	p
97	98	Sasi Kumar	17	Post-Matric	Govt	BC	M	Colli	20,000	Dayscholar	N

Table IV Analysis of the student and schemes datasets

	Attempts	Parentsincome
Count	40	40,000,000
Mean	1	118750
Std	0	168110.73
Min	1	0
25%	1	0
50%	1	50000
75%	1	212500
Max	1	600000

The educational schemes for the MBC and BC students are Reimbursement of Tuition fees to students in Govt. /Govt. Aided English Medium Schools, Reimbursement of Special Fees, Examination Fees Reimbursement, Post-matric Scholarships, Free Education Schemes, Scheme for Hostellers, Incentives for Rural MBC/DNC girl students, Free Hostels Scheme, Boarding Grants and Pre-Matric scholarship scheme. The figure 3 envisions the result which is obtained from the analysis done on the different communities (SC, ST, MBC, BC) of the 102 students in the student dataset and their parent's annual income.

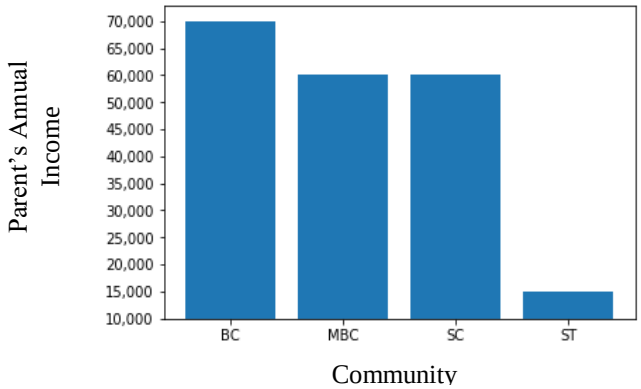

Fig. 3 Analysis of the parent's income in each community

The x-axis of the figure 3 envisions the various communities of the students such as SC, ST, MBC and BC, whereas the y-axis of the figure 3 envisions the annual income of the parents of the rural students.

C. Analysis of Prediction System

From the data obtained from the Anaconda Navigator (datasets), the prediction system compares the student details from the student dataset with the schemes from the schemes dataset and predicts the resultant schemes along with the details which the student is eligible for by using the two prediction algorithms (Naïve Bayes Algorithm and Random Forest).

The prediction system process can be explained with an instance. Consider the student Malika in the student dataset is going to apply for the government schemes in the educational domain. During this process in the prediction system phase initially the particular student's age who is going to apply for the government schemes in the educational category is obtained. Malika's age (14) is given to the prediction system. Then the particular student's community (SC/ST/MBC/BC) is obtained. Malika's community (SC) is obtained from the user. Followed by the community detail the gender (Male/Female) of that particular student and the number of attempts (1 or 2) the student take to pass that particular class are obtained. Malika's gender (Female) and attempts she took (1) is obtained from the user.

Next the educational category (Accounting and Finance/Agricultural Sciences and Medicine/Any Category/Commerce/Engineering and Management/Entrepreneurship/Humanities and Social Science/Law or Engineering/M Phil/Medical/PhD/Post Matric/Pre Matric/Post matric and Pre Matric/Pure Sciences and Applied Sciences/UG Only) of that particular student is obtained. Malika's educational category (Pre Matric) is obtained from the user. Next the board of university (Govt or Private/Institution of Excellence/Govt and Govt Aided/Govt and Govt Aided English Medium/Private only/Govt only) of the particular student is obtained. Malika's board of university (Govt only) is obtained from the user. Then the parent's occupation (Any Occupation/Manual Scavengers/tanners/Flayers/Hazardous Cleaning) of the particular student is obtained. Malika's parent is working in a match industry therefore; the parent's occupation is given as Any Occupation to the prediction system. Next the annual income of the parent of that particular student is obtained. The annual income of Malika's parent (40000) is given to the prediction system. The next detail the prediction system is in need of is whether the particular student is a day scholar or hosteller. Malika's detail (day scholar) is given to the prediction system. The next detail the prediction system is in need of is whether the particular student is physically challenged or not. Malika's detail (not physically challenged) is given to the prediction system. The next detail the prediction system is in need of is whether the particular student is from a farmer family or not.

Malika's detail (not from a farmer family) is given to the prediction system. Finally, the prediction system is in need of whether that particular student is girl student or not, if so she will get a girl preference quota. Since Malika is a girl student Malika's detail (girl student) is given to the prediction system.

Finally, the model is trained for the given inputs using the prediction algorithm and the output classes are predicted successfully. According to the input details given for the student Malika the predicted scheme for that student is Reimbursement of Tuition fees to students in Govt./Govt. Aided English Medium Schools and the scholarship amount predicted for that particular scheme is 25000.0 per year.

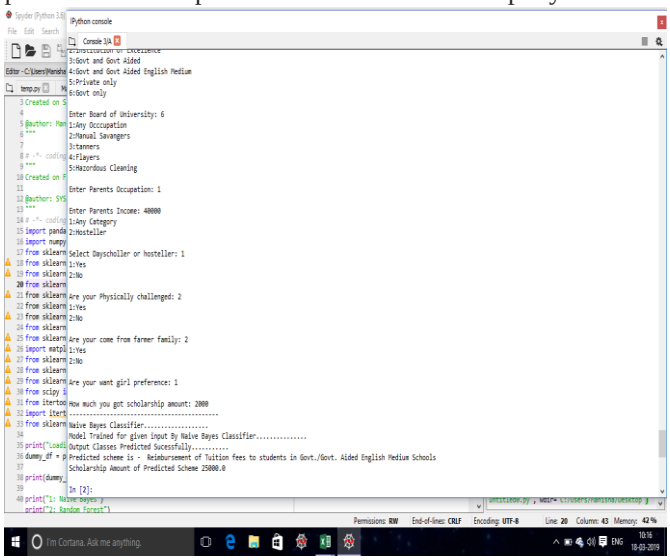

Fig. 4 Output of the Prediction System (Naïve Bayes Algorithm)

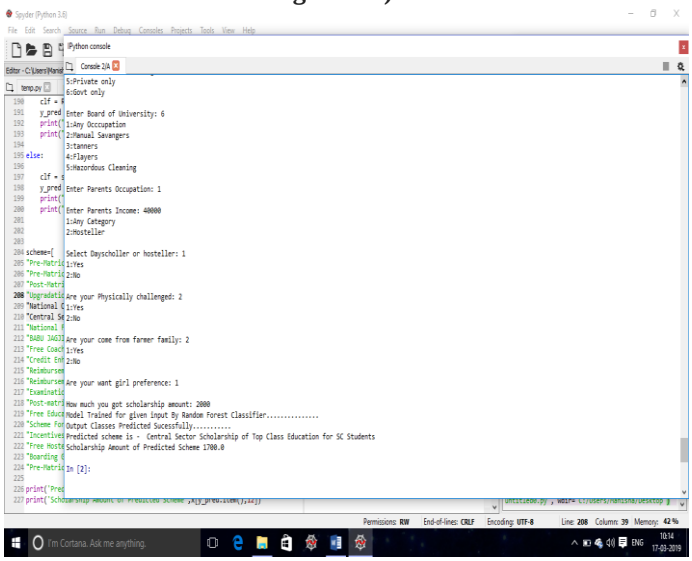

Fig. 5 Output of the Prediction System (Random Forest Algorithm)

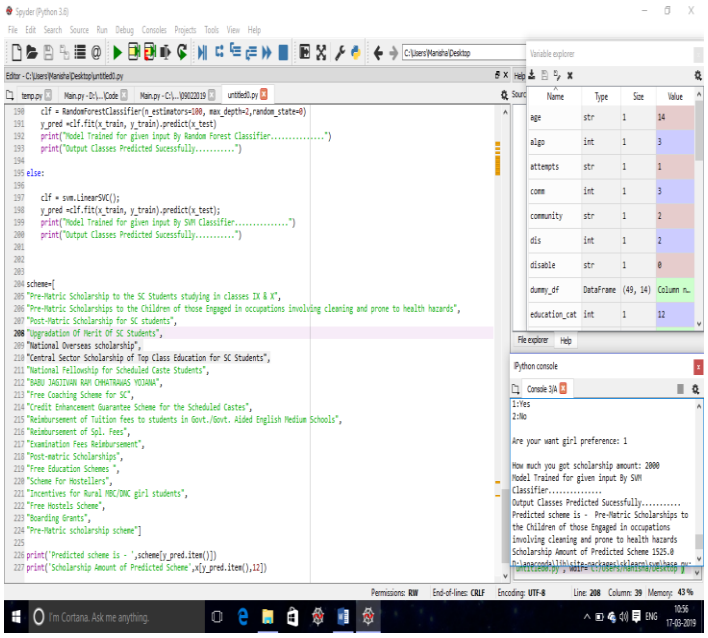

Fig. 6 Output of the Prediction System (Support Vector Machine (SVM) Algorithm)

Table V Comparison of Naïve bayes algorithm, Random Forest algorithm and Support Vector Machine's (SVM) prediction system

Input Attributes from the user	Sample Input	Output of Naïve Bayes Algorithm	Output of Random Forest Algorithm	Output of Support Vector Machine
Age	14	Predicted Scheme is- Reimbursement of Tuition fees to students in Govt./Govt. Aided English Medium Schools	Predicted Scheme is- Central Sector Scholarship of Top class education for SC students	Predicted Scheme is- Pre Matric Scholarship for the children of those who engaged in occupations involving cleaning and prone to health hazards
Community	SC			
Gender	Female			
Attempts	1			
Educational Category	Pre Matric			
Board of University	Govt Only			
Parent's Occupation	Match Industry(Any Occupation)			
Parent's Income	40000	Scholarship Amount of Predicted Scheme – 25000.0	Scholarship Amount of Predicted Scheme – 1700.0	Scholarship Amount of Predicted Scheme – 1525.0
Day scholar or Hosteller	Day scholar			
Physically Challenged or Not	No			
Farmer Family or Not	No			
Girl Preference or Not	Yes			

Table VI ANOVA result

ANOVA						
Source of Variation	***SS***	***df***	***MS***	***F***	***P-value***	***F crit***
Rows	152	31	4.903226	6.333333	8.16E-07	1.822132
Columns	3600	1	3600	4650	2.4E-35	4.159615
Error	24	31	0.774194			
Total	3776	63				

The table IV envisions the comparison between the outputs of the three algorithms (Naïve Bayes, Random forest and Support Vector Machine (SVM)) while the same sample input is given to the prediction system of the three algorithms.

The output, predicted scheme predicted by the Naïve Bayes Algorithm is Reimbursement of Tuition fees to students in Govt. /Govt. Aided English Medium Schools and the scholarship amount of the predicted scheme is 25000.0 per year with an accuracy value 0.72. Similarly the output, predicted scheme predicted by the Random Forest Algorithm is Central Sector Scholarship of Top class education for SC students and the scholarship amount of the predicted scheme is 1700.0 per month with an accuracy value 0.69. Since the accuracy value of the Naïve bayes algorithm's output is greater than the Random Forest Algorithm's output and the Support Vector machine's (SVM) output, the Naïve Bayes algorithm is concluded as the most efficient algorithm for the prediction system among the three algorithms.

D. Data analysis

Table VII t-Test: Paired Two Sample for Means

	Variable 1	*Variable 2*
Mean	52.5	67.5
Variance	4.387096774	1.290322581
Observations	32	32
Pearson Correlation	0.867721831	

With the help of the sample dataset taken the working of the prediction algorithm is discussed here. To comparing the results of both the prediction algorithm AVONA and posthoc tests are used.

For the table 5 F ratio=6.333333 and F critical value=1.822132, if the F ratio is larger than the F critical value then there is a statistically significant difference between Random Forest and naïve bayes. To find which prediction algorithm is efficient posthoc test is used. The result of the tool for the values in table 6 is given above. If tstat is greater than tcritical .Therefore, Random forest is efficient .Similarly remaining algorithms are analysed using ANOVA and posthoc test. From the results obtained by using ANOVA and posthoc test it is clear that Naïve Bayes algorithm is efficient than Random Forest.

VI. CONCLUSION

The Indian government inaugurates numerous ambitious programs trying to make the country more harmonious, hence creating an aware situation among rural people is very important. Through this innovative proposal the maximum utilization of the government schemes will be possible. The rural people can effectively make use of all the schemes furnished by the government of India for their welfare to the maximum extent. The amount spent by government for the welfare schemes will yield a worthy outcome. Thus according to the saying "Developed Villages—Developed Nation", India will move towards a prosperous future. In future this project can be enhanced by performing the prediction process for every government school students in a whole district.

REFERENCES

1. Deepika Bansal, Rita Chhikara, Kavitha Khanna, Poonam Gupta: "Comparative Analysis of Various Machine Learning Algorithms for Detecting Dementia"© 2018 The Authors. Published by Elsevier Ltd.
2. Anantvir Singh Romana: "A Comparative Study of Different Machine Learning Algorithms for Disease Prediction", International Journals of Advanced Research in Computer Science and Software Engineering ISSN: 2277-128X (Volume-7, Issue-7) (July 2017).
3. RaedShubair: "Comparative Study of Machine Learning Algorithms for Breast Cancer Detection and Diagnosis",The 2016 IEEE 5th International Conference on Electronic Devices, Systems, and Applications (ICEDSA'2016).
4. Krishna Prasad, B.D.C.N. Prasad, SagarYeruva: "A Comparative Study of Machine Learning Algorithms as Expert Systems in Medical Diagnosis (Asthma)",Purdue University, West Lafayette, USA.
5. GiorgioBiagetti`,PaoloCrippa, LauraFalaschetti, GiuliaTanoni, ClaudioTurchetti: "A comparative study of machine learning algorithms for physiological signal classification",Under a Creative Commons license.
6. Dazhong Wu, Connor Jennings, Janis Terpenny, Robert X. Gao and Soundar Kumara: "A Comparative Study on Machine Learning Algorithms for Smart Manufacturing: Tool Wear Prediction Using Random" ,The American Society of Mechanical Engineers.
7. SonghuiRyu, Baijian Yang: "A Comparative Study of Machine Learning Algorithms and Their Ensembles for Botnet Detection", Purdue University, West Lafayette, USA.
8. Sebastian Raschka : "Model Evaluation, Model Selection, and Algorithm Selection in Machine Learning" , University of Wisconsin–Madison, Department of Statistics, November 2018.
9. Sabyasachi Mohanty, Amit K.Mishra, Dhruba C. Panda : " IgoSA- A novel framework for analysis of and facilitating government schemes, 2015 IEEE 2nd International Conference on Recent Trends in Information Systems (ReTIS).
10. R. S. Mohanty: "Rural development programmes in panchayati raj institutions (pris): An overview", Editors Note, p. 16, 2014.
11. (2015) Schemes — national portal of india. [Online]. Available: http://www.india.gov.in/my-government/schemes/
12. R. Nayak, N. Saxena, and J. Farrington, Reaching the poor: "The influence of policy and administrative processes on the implementation of government poverty schemes in India." , Overseas Development Institute London, 2002.
13. G.-H. Kim, S. Trimi, and J.-H. Chung, "Big-data applications in the government sector," Communications of the ACM" ,vol. 57, no. 3, pp. 78–85, 2014.
14. Patrick Hall, Navdeep Gill, Mark Chan: "Practical Techniques for Interpreting Machine Learning Models: Introductory Open Source Examples Using Python, H2O, and XGBoost" ,H2O.ai, Mountain View, CA February 3, 2018.
15. Marco Tulio Ribeiro, Sameer Singh, and Carlos Guestrin.: "Why should I trust you?: Explaining the predictions of any classifier" ,In Proceedings of the 22nd ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, pages 1135–1144.,ACM, 2016., URL http://www.kdd.org/kdd2016/papers/files/ rfp0573-ribeiroA.pdf.
16. Bharat R.Naiknaware, Seema Kawathekar, Sachin N.Deshmukh: "Sentiment Analysis of Indian Government Schemes Using Twitter Datasets" , IOSR Journal of Computer Engineering (IOSR-JCE) e-ISSN: 2278-0661,p-ISSN: 2278-8727, PP 70-78 www.iosrjournals.org.
17. Paul Akangah, Leotis Parrish, Andrea Ofori-Boadu and Francis Davis: "Predicting Academic Achievement in Fundamentals of Thermodynamics using Supervised Machine Learning Techniques", North Carolina A&T State University/ Kwame Nkrumah University of Science & Technology.
18. Nithyassik, B., and Nandhini, D. E. C.: "Classification Techniques in Education Domain", International Journal on Computer Science and Engineering, 2010, pp. 1647-1684.
19. Martin Grohe RWTH Aachen: "Machine Learning and Algorithmic Model Theory", Rwthaachen University.
20. Harshad M. Kubade: "The Overview of Bayes Classification Methods", Published in International Journal of Trend in Scientific Research and Development (ijtsrd), ISSN: 2456-6470, Volume-2 | Issue-4, June 2018, URL: http://www.ijtsrd.com/papers/ijtsrd15750.pdf.

AUTHORS PROFILE

Dr.S.Maheswari is working as an Associate Professor at National Engineering College, Kovilpatti, India. She has received PhD from, Anna University; Chennai.Her area of interest includes Semantic web service selection and machine learning concepts. Her current research focuses on Semantic web service selection, Machine learning and Computer Networks. She has published papers in national/international journals and conferences and a Reviewer of international journals.

Mrs.S.Kalaiselvi received her M.E Degree in Computer Science and Engineering from National Engineering College, Kovilpatti. She is pursuing her Ph.D in Anna University, Chennai. Her research interests include Iage analysis, Remote sensing, Machine Learning and soft computing.

Mrs.D.Thamarai Selvi is working as an Assistant Professor at National Engineering College, Kovilpatti, India. Her area of interest includes Machine and Deep learning concepts. She has published papers in national/international conferences.

Mrs.M.Manochitra is working as an Assistant Professor at National Engineering College, Kovilpatti, India. She has received M.E from, Anna University; Chennai.Her area of interest in machine learning concepts. She has published papers in national/International journals and conferences.