

# Developing Eligibility Predicting Model for Applicants of Oromia Micro and Small Enterprises Agency using Machine Learning Approach

Naol Bakala Defersha

Lecturer, Head of Department, Computer Science, Institute of Technology, Ambo University, Ethiopia

## ABSTRACT

Micro and Small Enterprise is the techniques that Oromia Regional state is implementing to reduce jobless people by organizing people under into different groups depending on the eligibility of the individuals. Currently, the eligibility of individuals desired to be organized as enterprise is identified manually. For this study, Data collected from Oromia Micro and Small Enterprise office in hardcopy format. Researcher converted data into electronic text document. Since data collected is large in size, has important data, researcher implement feature Extraction and feature selection techniques to prepare data set. From feature extraction researcher observed that eight attributes Applicant.ID, Education\_status, Employment\_status, Jobless\_Certificate, organization\_name, sectors, and types are important whereas applicants.name, Efficiency, Id given\_from and Interest are attributes are unimportant. Dataset prepared from attributes confirmed as important and saved by MSDS.csv to make it more supportable format. Researcher implemented machine learning algorithm such as SMO, Naïve Bayes, and Bayes Net to build model that predicts eligibility applicants for Micro and Small Enterprise. Experiment shown SMO algorithm scored high accuracy. Therefore, finally the eligibility prediction prototype for Micro and Small Enterprise of Oromia built by using SMO algorithm.

**Keywords :** Micro and Small Enterprise, prediction model, Machine learning, Dataset.

## I. INTRODUCTION

Micro and small Enterprise is the strategy that used by Oromia regional state as new approach to reduce the jobless and to increase the development of the communities. Under Micro and small Enterprise, Persons organize themselves to obtain services (license, fund and working place) depending on their interest, professions, skills and capacity only if they are eligible. Eligibility (eligible and ineligible) identified as per a working principle of MSE office of Oromia stated in 2004EC (Mootummaa, 2004). Nowadays, the eligibility of each applicant is identifying manually. Manual identification is tedious, prone to corruption, prone to error, and resources consumer. As a solution, it is essential to design model

that automatically identifies the eligibility of applicants. Therefore, the Researcher desired to develop prototype model that automatically identifies eligibility of applicants using machine learning approach.

To build prototype model, data will be collected from Oromia Regional state of MSE agency and Human Resource office in word and excel format. After collecting data, Researcher performs text preprocessing, text document representation, and feature extraction to prepare data set for machine learning. Machine learning algorithms will be applied on train set to build eligibility prediction model. Finally, researcher tests the accuracy of each algorithm to select algorithm with highest

performance to design final eligibility prediction model. To test performance of model, Researcher load test data set into it prediction model. Finally, the performance of prediction model is evaluated by using performance evaluation parameters.

### **STATEMENT OF THE PROBLEM**

Many applicants apply to MSE of Oromia office to organize themselves and to get services. This office is identifying the eligibility of applicants manually depending on its working principles. Such manual techniques is prone to corruption, prone to error, consumes resource, tedious and traditional approach. Automating eligibility prediction system for MSE of Oromia is essential. Therefore, Researcher aimed to develop eligibility prediction by utilizing machine learning approach.

### **Objectives**

#### **General Objectives**

The general objective of this work was to design a model that predicts the eligibility of applicants for Oromia MSE using machine learning approach.

#### **Specific Objectives**

- ✓ To develop corpus
- ✓ To establish the appropriate data sets depending on corpus
- ✓ To design Architecture of model that predicts eligibility of applicants
- ✓ To compare performance of some supervised machine learning algorithms
- ✓ To develop final prototype model that predict eligibility of applicants
- ✓ To evaluate performance of eligibility prediction model

### **Scope and limitation of study**

This work limited to designing prototype model that predicts eligibility of applicants for MSE in case of Oromia. Building eligibility prediction model depends on working principal of Oromia MSE. Researcher, focused only on text document information of applicants. Researcher also collect data in Afan Oromo Language from Oromia Region MSE agency and Oromia Human Resource office in text documents excluding video, image, audio and graphics data.

Support Vector machine, Naive Bayes and Bayes Net used by researcher to build model that predict eligibility applicants.

### **Significance of the research**

The finding of this research can be used as input for the research that can be conducted in future. Result of this research also plays vital to make Afan Oromo language of Technology. Designed model is framework for developing system in the sector. Application of machine learning for macro and small enterprise of Oromia confirmed. Beneficiary of research are researchers, Macro and small enterprise of Oromia, Oromo Ethnic groups and scientific community. In general, implementing developed eligibility prediction model, in Oromia Mirco and Small Enterprise Agency reduces corruption, complain and save resource to identify eligible applicants from ineligible.

## **II. METHODS AND MATERIAL**

### **Literature Review**

Researcher reviewed different research article, journals, and thesis related to this work. Machine learning applied in different area of Customer profiling, credit risk analysis, medical diagnosis, market segmentation, targeted marketing, fraud detection and retail management are some application

areas of machine teaching (George et al., 2006). Machine learning also applied for business for Loan Approval (Kumar et al., 2016). But no one tried to apply machine learning to predict eligibility of applicants for Oromia Micro and Small Enterprise of agency.

Arutjothi & Senthamarai (2017) attempted to design model that predict of Loan Status in Commercial Bank using Machine Learning Classifier. Arutjothi & Senthamarai (2017) objective was implemented by using R tool packages. The machine learning classifier algorithms designed model that predict loan status in commercial banks with highest accuracy.

**Data source and data set preparation**

Researcher collected data from Oromia MSE office in text document format and hardcopy format. Hardcopy format of document changed into softcopy format then to in CSV format. From data collected, data set was prepared after Feature Extraction and feature selection applied on it. Data set split into training set and test set by utilizing R programming language.

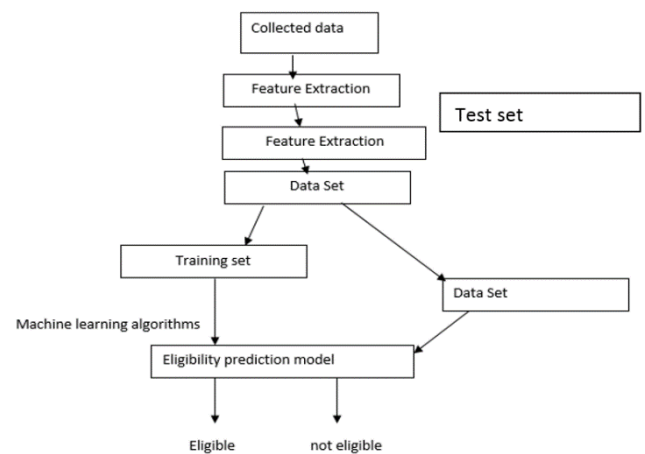
**Tools**

**1) Weka tool**

Researchers utilized weka tool for Text clustering, text document classification, feature extraction and feature selection (Shweta, 2014). In present study also I utilized weka tool for text for building and test Eligibility prediction model Macro and small enterprise of Oromia. Researcher selected weka tool in building eligibility prediction model due to familiarity with it and it is free available. Training data set and test data loaded into weka after entire data set divided into Training data set and test data. Researcher loaded train data set into weka tool to design Eligibility prediction model by applying SMO, Bayes Net and Naïve Bayes algorithms. Test data set also loaded to evaluate performance of the designed model.

**2) R programming language**

R is software that freely available with different help through internet (Paul & Claudia, 2014). As stated by Bernd et al. (2016) R is well known software that utilized for machine learning, data mining, and statistics. In this work, Researcher will use R programming language for data set preparation, conducting experiment and building eligibility prediction model.



**Figure 1.** Architecture for Designing Applicants Eligibility Prediction Model

**Feature Extraction and Feature Selection**

To get accurate value from our experiment, normalizing data set is essential (Kamil, et al. 2013). The advantages of feature extraction is increasing performance of model and reducing number of features in dataset (Katzir & Elovici, 2017). In this work, researcher confirmed that eight attributes Applicant.ID, Education\_status, Employment\_status, Jobless\_Certificate, organization\_name, sectors, and types are important. Applicants.name, Efficiency, Idgiven\_fromand Interest are attributes are unimportant. Data set split into train and test set after feature extracted. Train set contain contained 80 percent of Data set and test set contain 20% Data set. Testing the performance of eligibility prediction model is essential. Researcher used Confusion matrix for evaluating actual and predicted value for testing

prediction model (kumari1, M, Vohra, R & Arora, A., 2014).

**III. RESULTS AND DISCUSSION**

**Model Performance Evaluation Metrics**

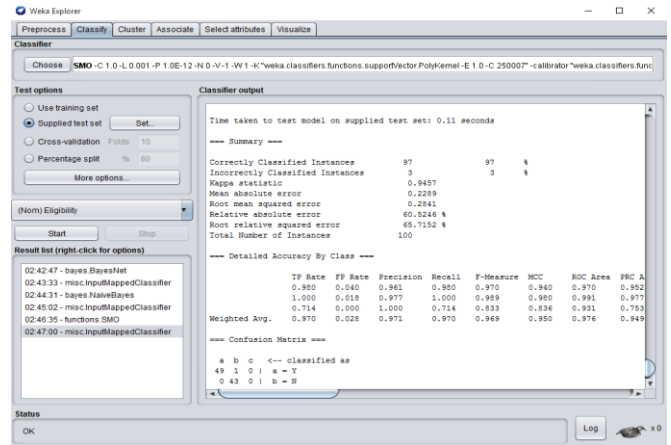
As stated in previous section the data set divided into two train and test data set. Train data set loaded into weka tool to build to prediction model. After train data set loaded into researcher selected SMO, Naive Bayes and Bayes Net algorithmsto build eligibility prediction model. Researcher utilized three algorithms SMO, Naive Bayes and Bayes Net to build Eligibility Prediction Model. All three applied algorithms recorded various accuracy in percent. SMO algorithm scored 89.45%, Naive Bayes scored 88.4058% and Bayes Net scored88.4058%of accuracy as stated in table1. By comparing the accuracy of three algorithms; SMO, Naive Bayes and Bayes Net used by researcher to build prediction model, SMO algorithm shown the highest performance comparing with the performance of Naive Bayes and Bayes Net. Therefore, the researcher utilized SMO to build eligibility prediction model.

**Table 1 :** The accuracy of used algorithms Testing Eligibility Prediction Model

Sno	Algorithm used	Accuracy
1	SMO	89.45%
2	NaiveBayes	88.4058 %
3	BayesNet	88.4058 %

As indicated in previous section 20 percent of data set divided as test set to evaluate the performance of

eligibility prediction model built by SMO. Test set prepared in appropriate manner by removing eligibility and ineligible classes. In test set eligible and ineligible replaced by unknown “?”. The researcher loaded this prepared test set into wekato evaluates performance of eligibility prediction model. Result shown table 2 displayed as result of evaluated prediction model tested by using test data set.



**Figure 1 :** Result Experiment

**DISCUSSION**

The performance of model was evaluated by using F-measure, percentage values of correctness, purity and entropy (Gottschalg & Ribeiro, 2014).

Designed eligibility prediction Model of Micro and Small Enterprise of Oromia evaluated by test data set. From detailed accuracy by class computed from confusion matrix in table 2, the model correctly predicts the result of accepted applicant as eligible 97.0% .

=== Detailed Accuracy By Class ===

TP Rate	FP Rate	Precision	Recall	F-Measure	MCC	ROC Area	PRC Area	Class
0.980	0.040	0.961	0.980	0.970	0.940	0.970	0.952	eligible
1.000	0.018	0.977	1.000	0.989	0.980	0.991	0.977	Not eligible

**Table Result of Confusion Matrix**

Number of class	Accuracy	Predicted class
1	97.0%	Elible
2	98.9%	Ineligible

**Table 3 :** Testing the eligibility applicant for MSE office using test data set results

**DISCUSSION**

Table 3 shown machine learning can be applied to identify eligible applicants from ineligible. As shown in table3 from 20% of test data set loaded into weka tool to evaluate performance of the built model, 97.0% of data set as eligible and 98.9% of test set predicted as ineligible. From this table we can easily concluded SMO is best algorithm is to identify the applicants as eligible and not eligible.

**IV.CONCLUSION**

Many Applicants applied to Micro and Small Enterprise of Oromia. It is very important to design model that identifies eligibility of the applicants. The proposed model developed to predict the eligibility of the applicants. This model enables the applicants to know their status on immediately within complaining to others. In addition this developed model also reduces corruption, save time, papers and human resources regarding to office Oromia at large. To design this model Machine learning techniques used. Machine learning techniques are essential for decision making and discovering the patterns of the data used. In this study, eligibility prediction tried to design model using SMO, Naive Bayes and Bayes Net machine learning algorithms. Among used Machine learning algorithms, SMO scored highest performance. As indicated during testing the model the built model

by SMO plays vital role to identify weather the applicant is eligible or not eligible.

**V. REFERENCES**

[1]. Arutjothi, G & Senthamarai, C.(2017). Prediction of Loan Status in Commercial Bank using Machine Learning Classifier. Department of Computer Applications Government Arts College (Autonomous) Salem India. Proceedings of the International Conference on Intelligent Sustainable Systems (ICISS 2017). IEEE Xplore Compliant - Part Number: CFP17M 19-ART, ISBN :978-1-5386-1959-9.

[2]. Kamil, A, Hassan, I & Abraham, A.(2013). Modeling Consumer Loan Default Prediction Using Neural Netware. Department of computer science Sudan University of Science and Technology, Sudan. Machine Intelligence Research Labs (MIR Labs) Scientific Network for Innovation and Research Excellence, WA, USA, ajith.abraham@ieee.org 2013 International Conference On Computing, Electrical And Electronic Engineering (Iccee)

[3]. Katzir Z & Elovici, Y. (2017).Quantifying the Resilience of Machine Learning Classifiers Used for Cyber Security. Deutsche Telekom Laboratories at BenGurion University, Department of Software and Information Systems Engineering, BenGurion University of the Negev.

[4]. kumari1, M, Vohra, R & Arora, A.(2014). Prediction of Diabetes Using Bayesian Network. College of Engineering Sector 3A, Sarai Aurangabad, Bhadurgarh. International Journal of Computer Science and Information Technologies, Vol. 5 (4) , 2014, 5174-5178

[5]. Liu B. (2011). Supervised machine learning approach applied for loan prediction. Web Data Mining: Exploring Hyperlinks, Contents, and Usage Data, Data-Centric Systems and

Applications, DOI 10.1007/978-3-642-19460-3\_3, © Springer-Verlag Berlin Heidelberg 201).

- [6]. Ribeiro, A. A., Gottschalg, C. D, (2014). Automated text clustering of newspaper and scientific texts in brazilianportuguese: analysis and comparison of methods. University of Brasília (Universidade de Brasília–UnB), Brasília, DF, Brazil. JISTEM J.Inf.Syst.Technol. Manag. vol.11 no.2 São Paulo May/Aug. 2014.

**Cite this article as :**

Naol Bakala Defersha, "Developing Eligibility Predicting Model for Applicants of Oromia Micro and Small Enterprises Agency using Machine Learning Approach", International Journal of Scientific Research in Computer Science, Engineering and Information Technology (IJSRCSEIT), ISSN : 2456-3307, Volume 5 Issue 5, pp. 29-34, September-October 2019.

Journal URL : <http://ijsrcseit.com/CSEIT19553>