Small and Medium Enterprises

IJSMEs VOL 6 NO 1 (2023) P-ISSN 2576-7712 E-ISSN 2576-7720

Available online at https://www.cribfb.com
Journal homepage: https://www.cribfb.com/journal/index.php/ijsmes
Published by CRIBFB, USA

DETERMINANTS OF NATURAL RESOURCES BASED MICROENTERPRISES PERFORMANCE IN INDIA'S WESTERN HIMALAYAN REGION: A NAÏVE BAYES CLASSIFIER ANALYSIS

Ina Bahuguna [(a)] *Ujjwal Kumar* [(b)1] *Kusum Arunachalam* [(c)] *Vijay Shridhar* [(d)] *Archana Sharma* [(e)]

[(a)] *Junior Research Fellow, School of Environment and Natural Resources, Doon University, Dehradun, India; E-mail: Ina_bahuguna@yahoo.com*
[(b)] *Assistant Professor, School of Environment and Natural Resources, Doon University, Dehradun, India; E-mail: ujjwalkumarin@yahoo.co.in*
[(c)] *Professor & HoD, School of Environment and Natural Resources, Doon University, Dehradun, India; E-mail: kusumdoon@gmail.com*
[(d)] *Assistant Professor, School of Environment and Natural Resources, Doon University, Dehradun, India; E-mail: vshridhardoon@gmail.com*
[(e)] *Assistant Professor, School of Environment and Natural Resources, Doon University, Dehradun, India; E-mail: doonarchana@gmail.com*

ARTICLE INFO

Article History:

Received: 5th April 2023
Revised: 29th May 2023
Accepted: 10th June 2023
Published: 24th June 2023

Keywords:

Naïve Bayes Classification, Microenteprises, Natural Resources, Performance, Factors

JEL Classification Codes:

O13, L26

ABSTRACT

The natural resources-based Microenterprises are the major part of the economy of the western Himalayan region of Uttara hand, India, as the region is predominantly covered with reserved forests. The present study evaluates the performance of Microenterprises and the factors affecting it in the region using the primary data enumerated from 110 microenterprises sampled under four major categories of microenterprises, viz, agro and allied, Animal and allied, handicrafts and handlooms, and miscellaneous. The Naïve Bayes classifier approach has been applied to evaluate the performances (Loss-making, breakeven, profit-making, or high-profit making) of these microenterprises based on their performance determining factors such as ease of raw material availability, level of training received, technological advancement, and the extent of market knowledge, and also on the type of ownership and the employee's number. The Naïve Bayes classification accuracy on the training dataset was 100%, while accuracy on the test dataset ranged from 93% to 100%. The results revealed that agro-based microenterprises have a greater probability (0.67) of making a profit/high profit, while animal product-based microenterprises have a high probability of running into losses. A higher level of Market Knowledge contributes to a high probability (0.89) of making high profits. The higher level of technology and training provides greater chances/probability (0.72, 0.72) of making high profits. Self-help groups (SHGs) have shown a better probability of making profits. The study suggests promoting SHGs in the region, wider dissemination of the market knowledge (marketing strategy), and leveling up the training/technology of the microenterprises.

INTRODUCTION

Microenterprises are categorized under small-scale businesses. World Bank (2005) defines a microenterprise as "an enterprise that has up to 10 employees, total assets of up to $10,000 and total annual sales of up to $100,000" (Thapa, 2015). U.S. Small Business Administration (2010) defines microenterprise as an enterprise that has "a sole proprietorship, partnership, limited liability corporation or corporation that has fewer than 5 employees, including the owner, and generally lacks access to conventional loans, equity, or other banking services." In India, a Government of India Act known as *Micro, Small, and Medium Enterprises Development (MSMED) Act (2006)* defines microenterprises as "enterprises (i) engaged in the manufacture or production of goods pertaining to any industry specified in the First Schedule to the *Industries (Development and Regulation) Act, 1951* (65 of 1951), where the investment in plant and machinery does not exceed twenty five lakh rupees; (ii) engaged in providing or rendering of services, where the investment in equipment does not exceed ten lakh rupees."

The natural resources-based microenterprises are the characteristics of the region dominated by the forests. Natural resources are the factors of production provided by nature (Andersen et al., 2018). They belong to what is traditionally referred to as the primary sector of the economy, which also encompasses the secondary (manufacturing) and tertiary (service) sectors (Andersen, 1982). The natural resources sector is primarily rural, and the resource-dependent communities

[1] *Corresponding author: ORCID ID: 0000-0003-3357-1128*

https://doi.org/10.46281/ijsmes.v6i1.2032

To cite this article: Bahuguna, I., Kumar, U., Arunachalam, K., Shridhar, V., & Sharma, A. (2023). DETERMINANTS OF NATURAL RESOURCES BASED MICROENTERPRISES PERFORMANCE IN INDIA'S WESTERN HIMALAYAN REGION: A NAÏVE BAYES CLASSIFIER ANALYSIS. *International Journal of Small and Medium Enterprises*, *6*(1), 27-40. https://doi.org/10.46281/ijsmes.v6i1.2032

traditionally struggle with persistent poverty (Shahidullah & Haque, 2014). Examining the strategies, performance, and sustenance of natural resources-based enterprises can harness local entrepreneurs' initiative to create sustainable and economically profitable businesses.

Many different approaches have been adapted to study the performance of microenterprises and the factors affecting their performance. Rosa et al. (1994) applied different multiple regression models to evaluate the performance of small businesses in Scottish and British areas. As considered in the study, the factors determining the performances were the number of owners, initial capital, firm age, gender, etc. Psaltopoulos et al. (2005) used a statistical analysis based on the simultaneous-to-bit model that showed significant factors influencing risk perception included the size of the new business and the sector of economic activity, as well as entrepreneurial experience and the location of the markets for the firm's (micro rural firms) output. Gulyani and Talukdar (2010) apply a multiple regression model to assess the performance of informal household microenterprises in Nairobi's slums. They conclude that better microenterprise performance is associated with certain "business-related" factors, such as sales area, time in, and sector of operation. But "living conditions"—residential tenure and infrastructure access—also strongly influence both creation and success of microenterprises. Masakure et al. (2009) employed a series of multiple regression equations to evaluate microenterprises' performance in terms of entrepreneurs' characteristics, enterprise sector, and market effects in Ghana. Boermans et al. (2012) apply a multiple regression model to test the effect of financial constraints and risk-taking on firm performance with Huber-White robust standard errors. Berrone et al. (2014) propose a theoretical multilevel framework that studies the determinants of microenterprises’ performance. They found that human capital (proxied by educational level and degree of dedication), innovation, and intensity, Thapa (2015) executes multiple linear regression models to identify the factors determining the performance of microenterprises in Nepal. Kamunge et al. (2014) also deploy multivariate regression analysis to establish the relationship between the performance and the factors affecting the performance of small and microenterprises in Limuru Town Market of Kiambu County, Kenya. Alom et al. (2016) employ multiple regression analysis to identify potential factors contributing to the overall growth of Malaysian microenterprises. They found that competition and the age of the enterprises negatively affect the overall performance of the microenterprises, whereas the age of the entrepreneurs, education, business training, and demand for the product/service, availability of physical space for business expansion in the city area, availability of financing and sufficiency of secured amount of finance pose positive impacts on the growth. Sohns and Revilla Diez (2018) use three-level binary-logistic random intercept models to analyze the effects of explanatory factors at different levels on micro-entrepreneurship in rural Vietnam. Their analyses show that identifying the motivation behind starting a microenterprise is a good way to split entrepreneurship into two groups: opportunity- and necessity-driven entrepreneurship, which are influenced by different explanatory factors at different levels. Vershinina et al. (2022) employ multiple regression analysis to assess the gendered regulations and SME performance in transition economies. Islam et al. (2018) formulate several hypotheses (H_1 to H_5) about the effect of usage of mobile phones on the performance of microenterprises. They apply regression analyses, regression-based path analysis, correlation analysis, and exploratory factor analysis to test those hypotheses. For example, to analyze hypotheses H_1 (Mobile phone usage is positively associated with the financial performance of MEs.) and H_4 (The impact of social capital on ME's business performance (both financially and non-financially) is positive and significant), multiple regression analyses were carried out. In order to test H_5 (The relationship between mobile phone use and MEs' performance is mediated by social capital), a set of hierarchical regressions were carried out. Correlation analysis was performed among social capital variables, financial performance, non-financial performance, and mobile phone use. Martin and Alejandro (2016) employ Cobb Douglas's productivity function to evaluate the role of human capital in the productivity performance of Mexican microenterprises.

In this study, we propose to apply the Naïve Bayes classifier (Domingos & Pazzani, 1997; Webb et al., 2005) approach, a part of machine learning algorithms for classification, to evaluate the performance of natural resources-based microenterprises in the western Himalayan region of Uttarakhand in India and the factors affecting their performances. The probabilistic outcomes of the Naïve Bayes classification have been used to explain the influence and significance of factors affecting the performance of Microenterprises. The Naïve Bayes classifier approach might prove to be very useful and effective when many of the explanatory variables are qualitative in nature, and the outcomes are also qualitative in nature. A total of 110 Natural resources-based microenterprises (agriculture or allied-based, Animal based, handlooms and handicrafts based, and miscellaneous) have been surveyed. The performance of microenterprises has been categorized as loss-making (L), breakeven (BE), and profit-making (P), high profit-making (HP) (if profit is > 20%). Accordingly, their performances have been evaluated, and the influence of factors determining their performances has been discussed. The rest of the paper is organized as follows. Section 2 presents the details of the basis of selecting factors affecting the microenterprise performance. Section 3 explains the methodology adapted for this study. Section 4 contains the results and discussion, and section 5 concludes the study.

LITERATURE REVIEW

The western Himalayan region of India is predominantly forests, and the livelihood of people in this region is mainly dependent on agriculture or the natural resources-based economy. Being the Himalayan mountainous terrain, agriculture is based on terrace farming and mostly subsistence (Bhandari & Reddy, 2015; Mamgain & Reddy, 2017), and only a small section of the hill population practices commercial agriculture. The natural resources-based enterprises provide a viable alternative livelihood support for the population in the region. Hence, the enterprises working in hills are predominantly agriculture and natural resource-based and can broadly be considered to assume the characteristics such as *agro and related*, *Animal and allied-based handicrafts and handlooms,* and *miscellaneous*.

In *agro and related categor*ies, the microenterprises included food processing businesses like juice, pickle, and sauce and others such as MAP (Medicinal Aromatic Plant), Nursery Raising, Horticulture, and floriculture. Agro-based industries help in the diversification of the rural economy and reduce its extreme dependence on agriculture farming (Paramasivan & Pasupathi, 2016).

Animal-based enterprises include dairy, poultry, fishery, and livestock. The livestock sector alone contributes nearly 25.6 percent of the value of output at current prices of the total value of output in the agriculture, fishing & forestry sector in the Indian economy. The overall contribution of the Livestock sector in the total GSDP (Gross State Domestic Product) of Uttarakhand, India, is nearly 4.11 percent at current prices during 2012-13. In Uttarakhand livestock growth rate of GSDP at its current rate fluctuates. The 2013-14 growth rate was 11.99 percent, dropping to 9.04 percent by 2016-17 (Directorate of Economics and Statistics Uttarakhand 2017).

Handicrafts and handlooms are a major source of income for many parts of the rural areas in India. Handicrafts and handlooms are made of natural resources available. It is also related to art, creativity, culture, aesthetic values, festivals, religion, and social symbols. These traditional art products based on natural resources attract the regional public as well as foreigners. These products depict the story of the long old culture and dynasty. It provides ample employment opportunities even with low capital investments.

The microenterprises involved in more than two entrepreneurial activities, like agro and animal-based or involved in activities other than the above three, have been categorized as *miscellaneous* in the study as they constitute less than 10% of total microenterprises under the study.

Various factors, such as raw material availability, training level of personnel, technology adopted in the enterprise, marketing of products, etc., affect the performance of a microenterprise. A systematic step-wise approach was conducted to identify the factors affecting the performances of microenterprises. Initially, 11 factors were considered; however, in the second phase, the factors of irrigation facility, food processing unit, and cold storage were included in the technology factor. In the third phase, the electricity factor was also included in the technology factor. Similarly, the factors of market survey and marketing were merged into the factor market knowledge factor. Thus, finally, the major factors affecting the microenterprises remained were technology, market knowledge, training, raw material availability, no. of employees, and type of ownership of microenterprises.

Several studies highlight the importance of the above factors, viz., training, technology, raw material availability, and market knowledge, in determining the performance of a microenterprise. Sánchez et al. (2010) show empirical evidence of a significant relationship between training and the performance of microenterprise. Loader and Johnston (2003) examined the training provided in SMEs and showed that 69% of employees had gained experience and satisfaction in demand for their service. Ghafoor Khan et al. (2011) shows that on-job training improves the performance of an enterprise.

Parichatnon and Maichum (2018) show that improvement in technical efficiency has a significant influence on production efficiency. Stone and Deadrick (2015) highlight the use of technology to facilitate industry management. Tambunan (2008) shows how the Indonesian government's initiative of transferring technology to SMEs has led to positive growth of microenterprises in the country.

Datta and Bhattacharya (2016) indicate how India's microenterprises' performance was affected because of the raw material availability. Dhiman and Rani (2011) examine the degraded quality of products due to a lack of raw material availability.

Cardamone and Rentschler (2018) show how SMEs struggle to market products, some from extremely remote and isolated locations and others with few resources in urban areas, thereby affecting SMEs' performances. Kalita and Prasad (2016) demonstrate the problems faced by artisans in marketing their products. Deshpande and Farley (1998) show that market orientation scales and other measures improve the firm's performance.

Thus, based on the literature survey and the step-wise systematic approach as discussed, in order to evaluate the performance of microenterprises, the present study relied on the following factors presented in Table 1. Table 1 lists all the variables considered in the Naïve Bayes classification algorithm to evaluate the performance of Microenterprises.

MATERIALS AND METHODS

Table 1. List of Variables (Factors Affecting the Performance of Microenterprises)

Dependent Variable	Variable notation	Type of variable	Categories of a variable	Definition of the variable used in the model
Performance of Microenterprises (Income)	Inc	ordinal	L BE P HP	L: Microenterprises running into losses BE Microenterprises which are not making profits and neither in loss P: Microenterprises making profits less than 20% HP: Microenterprises making profits of more than 20%

Explanatory Variables	Variable notation	Type of variable	Categories of a variable	Definition of the variable used in the model
Characteristics of Microenterprises	chc_ME	nominal	AnimalB AgB	AnimalB: Animal-based enterprises are enterprises that are dependent on livestock, fisheries, beekeeping, and poultry. AgB: Agriculture-based microenterprises, commercial agriculture, MAP, Nursery raising, and forest-based HH: Handicrafts and handlooms enterprises are enterprises that are making shawls, sweaters, woolen clothes, carpets, and wooden, and bamboo items. Misc: Miscellaneous enterprises are those enterprises that

			HH Misc	are involved in more than two entrepreneurial activities like agro and animal-based.
Training	Tr	ordinal	1 2 3	Training received by the enterprise's owner and workers was consolidated into one value based on training received twice and more was coded as 3, training received once was 2, and no training or traditional knowledge was coded as 1.
Technology	Tech	ordinal	1 2 3	The level of technology available within micro enterprises is the consolidated value (No machinery= 1; hand run machinery= 2; Electricity or diesel run machinery = 3)
Market Knowledge	MK	ordinal	1 2 3 4 5	Information on market knowledge levels are market survey, product, branding, advertising, and distribution (1 Point for availability of each level, total points = 5)
Raw Material availability	RW	ordinal	1 2 3	Information on the availability of raw material availability in the area on the basis of distance consolidated into one value. The distance to the nearest area, i.e., forest, market, and wholesaler. (The consolidated value was assigned using a points system based on the distance to the facility: 1 if available outside the district, 2 outside a 10 km radius from enterprises, and 3 within 10 km of enterprises).
Employees No.	Empl	ordinal	m1 m2 m3	m1: No. of employees is one, i.e., the owner runs it alone m2: No. of employees are between 2 and 10 m3: No. of employees are more than 10
Type of Microenterprises	TyME	nominal	SP SHG Phip NGO coop	SP: Sole propriety SHG: Self-Help Groups Pship: Partnership NGO: Non-Governmental Organization Coop: Cooperative

Study Area

The study has been carried out for the natural resource-based microenterprises in the Himalayan state of India, Uttarakhand. Uttarakhand holds international importance as the state shares its borders with China in the north and Nepal in the east. The state comprises 13 districts and is divided into two administrative divisions, Garhwal and Kumuan (Anthwal et al., 2010). Geographically, 88% of the state is hilly, while the rest is plain. The 11 hills/semi-hills districts were selected, covering the maximum of the land areas of the state. The selected districts were Pauri Garhwal, Tehri Garhwal, Uttarkashi, Chamoli, Rudrprayag, Almora, Pithoragarh, Bageshwar, Nanital, Champawat, and Dehradun.

Data Collection through Survey of Microenterprises

The data were collected from 110 natural resource-based microenterprises from 4 types of microenterprises (Table 2). The survey was conducted by interviewing the entrepreneurs/ owner of the enterprises using a structured questionnaire (Nardi, 2018). The information collected from the entrepreneurs included the background of the enterprise, size, primary operation, number of employees, details of training attended, use of technology and innovation, raw material availability, and market knowledge. The responses were measured using a structured set of ordinal scale**s** (Lee et al., 2010; Agresti, 1988).

Table 2. Number of units in selected microenterprises

S.No.	Type of Micro enterprises	Nature	Number of Micro units
1.	Agro and allied based	Commercial agriculture, MAP, Nursery raising and forest-based	47
2.	Animal and allied based	Dairy, poultry, fishery, beekeeping, Livestock	35
3.	Handicrafts and Handloom	Raw material extracted from forests and animals	17
4.	Miscellaneous	Entrepreneurs are working with agro and animal-based both. Ex: producer of fruit juices and owns poultry.	11

Naïve Bayes Classifier Approach

The Naïve Bayes classifier approach is a probabilistic approach to classification based on the Bayes theorem of probability**.** Bayes' Theorem tells us how to optimally predict the class of previously unseen examples, given a training sample (Duda et al., 1973). Naive Bayes is a conditional probability model: given a problem instance to be classified, represented by a vector $\boldsymbol{X} = (x_1, x_2, \dots, x_n)$ representing n features (independent variables), it assigns to this instance probabilities

$$P(C_k | x_1, \dots, x_n)$$

For each of K possible outcomes or classes C_k (Murty & Devi, 2011). Using Bayes' theorem, the conditional probability can be decomposed as

$$P(C_k|\boldsymbol{X}) = \frac{P(C_k)P(\boldsymbol{X}|C_k)}{P(\boldsymbol{X})}$$

The above equation is also conventionally written as

$$posterior = \frac{prior \times likelihood}{evidence}$$

Since the denominator does not depend on C and the values of features x_i are known, so the denominator is effectively constant. Therefore, the numerator is equivalent to the joint probability model

$$P(C_k, x_1, \dots, x_n)$$

Assuming that all features in $\boldsymbol{X}$ are mutually independent, conditional on category C_k, the joint model can be expressed as

$$P(C_k|x_1, \dots, x_n) \propto P(C_k, x_1, \dots, x_n)$$

$$\propto P(C_k)P(x_1|C_k)P(x_2|C_k) \dots$$

$$\propto P(C_k) \prod_{i=1}^{n} P(x_i|C_k)$$

The naïve Bayes classifier combines this model with a decision rule. One common rule is picking the most probable hypothesis to minimize the probability of misclassification, known as the maximum a posteriori or MAP decision rule. The corresponding classifier, a Bayes classifier, is the function that assigns a class label $\hat{y} = C$ for some k as follows:

$$\hat{y} = \underset{k \in \{1,\dots,K\}}{argmax}\, P(C_k) \prod_{i=1}^{n} P(x_i|C_k)$$

In the present study, the classifiers are the Performance measures of the microenterprises viz., L, BE, P, HP, and the features are MK, Tr, Tech, RW, Emp, chc_ME, andTyME (see table 2). The naïve Bayes algorithm has been implemented using R package e1071 (Meyer et al., 2014) on R-Studio software (R Core Team, 2021).

RESULTS AND DISCUSSIONS

In the present study, a total of 110 natural resources-based microenterprises were surveyed and analyzed to determine their performances and the factors affecting them.

General Analysis of factors affecting the Performance of Microenterprises

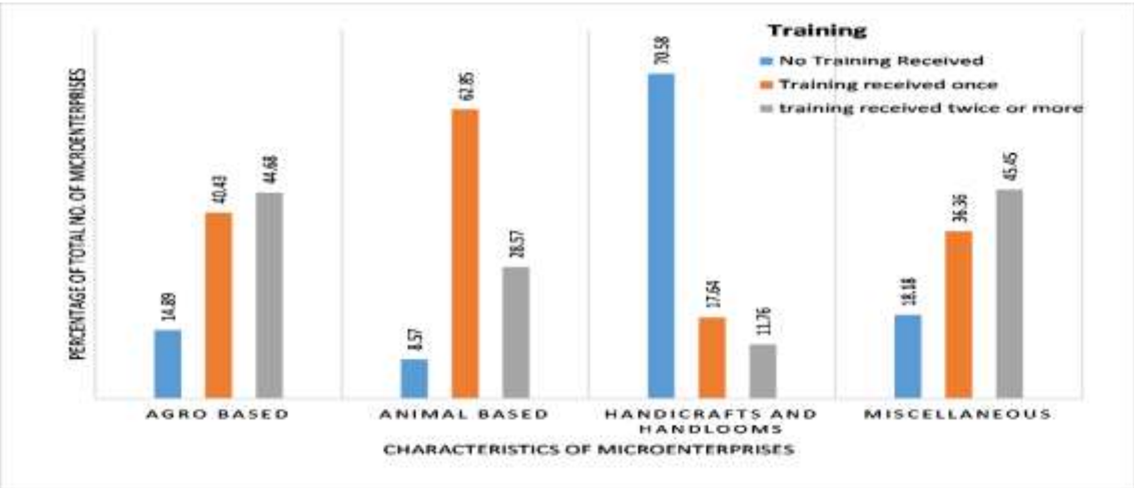

Figure 1. Percentage of microenterprises in which personnel received training

Figure 1 present's relative percentage of training received by personnel engaged in the surveyed 110 micro-enterprises. The factor training is coded according to a number of times entrepreneurs receive training. The findings show that 70 percent of respondents from handicrafts and handlooms either depended on traditional knowledge or did not receive any training to do the business (Figure 1).The reason is that mostly the knowledge is passed on from one generation to the next generation, which helps them in the preservation of culture and tradition, and the entrepreneurs engaged in handloom and handicrafts do not feel the need for newer training (Kathuria, 1986). For animal-based enterprises, the personnel's of 62.85% of enterprises received the training (at least once). In Agro based and miscellaneous, approx. 45% of respondents agreed that they used to receive training regularly, more than two times a year (Figure 1).

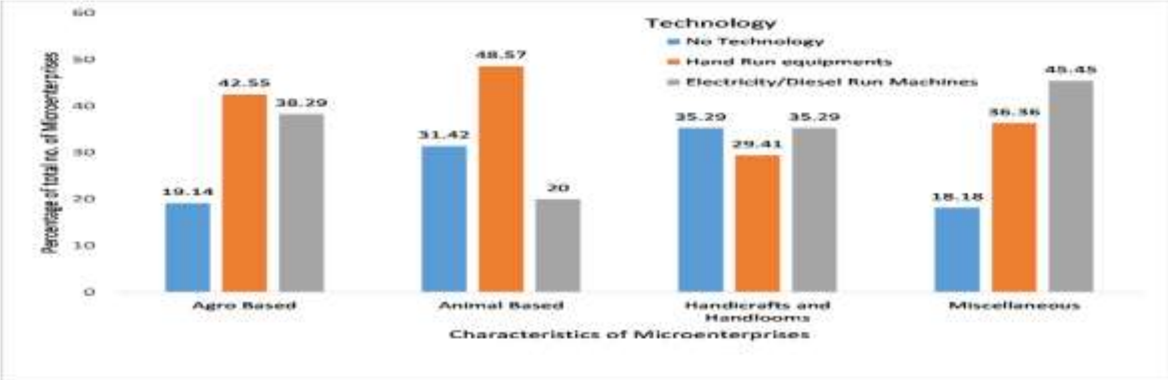

Figure 2. Relative percentage of availability of Technology indicator at different micro enterprises

Figure 2 shows the relative percentage of availability of technology among all the 110 microenterprises. The technology has been measured at three ordinal scales: no machinery, hand-run equipment, and electricity/diesel-run machinery. The analysis shows that the maximum number of respondents (42.55% in agro and allied-based and 48.57% in Animal and allied-based) agreed that they were using small machinery for the production of goods. In handicrafts and handlooms, 35% of the respondents conveyed that they were not using any kind of machinery, either hand run nor electricity based, while 35% said they were using electricity or diesel run machinery to make carpets, shawls, and for stitching. Approximately 30 percent of handicrafts and handlooms entrepreneurs used hand-run equipment like *hathkargha* (hand-run machines), and weaving machinery. In the category of miscellaneous micro-enterprises, 45.45% were using electricity/diesel based machinery (Figure 2).

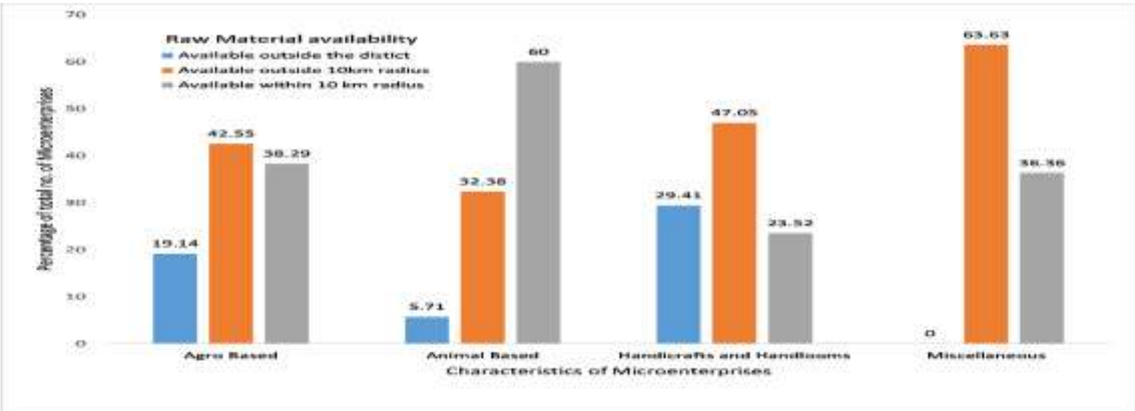

Figure 3. Relative percentage of raw material availability indicator at different micro enterprises

The raw material availability on the basis of distance from the micro-enterprises is shown in Figure 3. Raw material availability has been studied at three ordinal scales: available outside the district, outside 10 km, and within 10 km. Figure 3 shows that almost 60 percent of the microenterprises in Animal and allied based had reached raw material. The raw material availability outside the 10 km radius of the miscellaneous and agro-based enterprises is 63.63% and 42.55%, respectively. In handicrafts and handlooms, 29.41% of respondents purchase raw materials outside the district or the state.

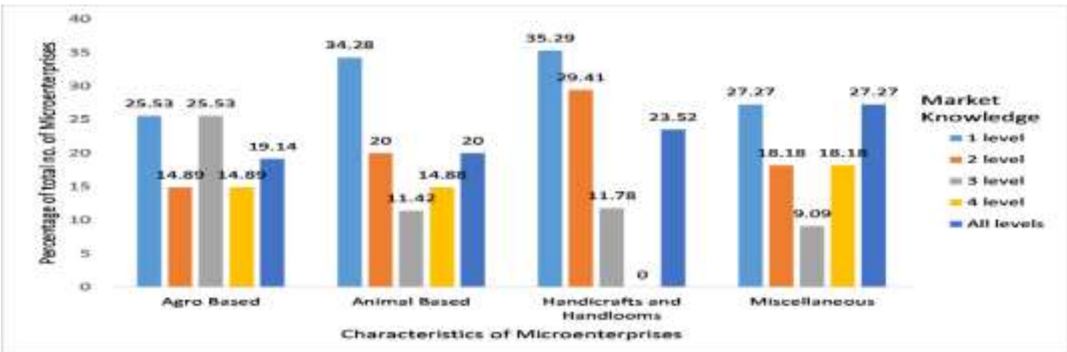

Figure 4. Relative percentage of market knowledge indicator at different micro enterprises

Figure 4 shows the percentage of microenterprises having market knowledge of different levels in different microenterprises. Market knowledge has been studied at five ordinal scales: market survey, product, branding, advertising, and distribution. The study shows that all four microenterprises use just one level of market knowledge (Figure 4). It is to be noted that handicrafts and handlooms and animal-based and allied have 35.29% and 34.29% respondents, respectively, with just one level of market knowledge, while for agro-based and allied, the maximum number of respondents had 1 level or 3 levels in market knowledge (Figure 4). The agro and allied-based (19.14%), Animal and allied-based (20%), handicrafts and handlooms (23.52%), and miscellaneous (27.27%) consisted of all five levels of market knowledge.

Correlation between factors and the performance (income) of microenterprises

Table 3. Correlation between Income (performance) and the factors affecting the performance of the microenterprises

	Income
Income	1
Training	.664**
Tech	.653**
MK	.889**
RW	.340**

**. Correlation is significant at the 0.05 level (2-tailed)

Pearson's correlation coefficients have been calculated to study the degree of relationship between the factors affecting the performances and the performance (income) of the microenterprises (Table 3). Table 3 shows that market knowledge is strongly correlated (.889) with income, indicating that the income of the microenterprises increases with the improvement in Market knowledge. Among four types of micro-enterprises and out of 110 microenterprises, it was observed that most of them possessed only one level of market knowledge. These microenterprises have products but need to market them (Figure 4). The income of these microenterprises has also been observed to be low. A strong correlation between income and market knowledge has also been observed in the study by Uematsu and Mishra (2011). Uematsu and Mishra (2011) explain that the intensity of the adoption of marketing strategies influences the income of the industry.

Income also shows a positive correlation with training (.66) and technology (.65). This indicates that when training activities aim to instill specific skills in unskilled labor, there is an improvement in quality, income, and productivity (Aragón-Sánchez et al., 2010). In the case of technology, productivity can be increased through the development of new technology and improved knowledge (Tambunan, 2008). A poor correlation (0.34) between raw material availability and income has been observed. One key reason could be that most enterprises were set up within a 10 km radius of the place of raw material availability. It has often been observed that SMEs, especially those dependent on natural resources, do not rely on imports of raw materials, and for the production cycle, raw material supply is important (Gandhi et al., 1999).

Naïve Bayes Classifier Approach

As discussed in Section 2 and Section 3, the performance of microenterprises has been classified as Loss-making (L), breakeven (BE), profit-making (P), and high profit-making (HP). Each of the factors influencing the performance of microenterprises has been measured on either an ordinal scale or a nominal scale, as detailed in Table 1. The factors are characteristics of microenterprises (agro-based or animal-based or handloom & handicrafts and miscellaneous), type of microenterprises (sole propriety, partnership, SHG, NGO, Cooperative), training, technology, no. of Employees and the market knowledge (Please see table 1). The data of all microenterprises have been split into two random sets: training and test data set in the ratio of 75:25. The naïve Bayes simulation experiment was performed on the training dataset, and the test dataset was kept out of the simulation experiment. Once the configuration of the Naïve Bayes model was finalized, it was tested on the test data set. In the following simulation experiment, again a new set of training and test data set was created by dividing the datasets into 75:25 ratio randomly, and then again the naïve Bayes simulation experiment was performed and the configured Naïve Bayes model was tested on the test dataset which was out of simulation experiment. This way, several simulation experiments were performed and tested. All the simulations have been performed using R software.

In each experiment, the naïve Bayes model was evaluated for its predicted classification accuracy on both the training and test dataset using the confusion matrix (Sokolova & Lapalme, 2009). The confusion matrix for training and test datasets of a naïve Bayes simulation experiment has been shown in Tables 4a & 4b. The Pearson's χ^2- test for all the table factors (L, BE, P, HP) shows p–value close to zero, therefore, the table factors (L, BE, H, HP) considered are highly significant. Similarly, Pearson's χ^2-test performed on test dataset results also exhibit a p-value close to zero. Hence, the factors considered are statistically significant.

Table 4a. Confusion Matrix for the training dataset

Naïve Bayes Predicted	Actual				
	L	**BE**	**P**	**HP**	**Row total**
L	15	0	0	0	15
BE	0	16	0	0	16
P	0	0	18	0	18
HP	0	0	0	18	18
Column total	15	16	18	18	67

Pearson's χ^2 $test$ for all table factors:
$\chi^2 = 201$, **deg of freedom = 9, p = 2.0444E-38**

Table 4b.Confusion Matrix for the test dataset

Naïve Bayes Predicted	Actual				
	L	**BE**	**P**	**HP**	**Row total**
L	4	0	0	0	5
BE	1	6	1	0	7
P	0	0	5	0	5
HP	0	0	0	6	6
Column total	5	6	6	6	23

Pearson's χ^2 $test$ for all table factors:
$\chi^2 = 69$, **deg of freedom = 9, p = 2.0444E-11**

Table 5. Accuracy of Naïve Bayes classification on training and test datasets on several simulation experiments

Simulation Experiment S. No.	Accuracy of the training dataset	Accuracy of test dataset
Expt 1	100%	100%
Expt 2	100 %	100 %
Expt 3	100 %	93.1 %
Expt 4	100 %	95.6 %
Expt 5	100 %	95.6 %
Expt 6	100 %	95.6 %
Expt 7	100 %	100 %
Expt 8	100 %	93.1 %
Expt 9	100 %	93.1 %
Expt 10	100 %	100 %

The confusion matrix results of the performance of naïve Bayes on the training and test data over 10 experiments have been presented in Table 5. The accuracy results for both the training and test datasets shown in table 6 clearly demonstrates that Naïve Bayes classifier has worked remarkably well (test datasets accuracy > 93.1 %; training datasets accuracy ≈ 100 %) in classifying the performances of microenterprises in the defined categories of L (loss-making), BE (breakeven), P (profit-making), HP (high profit making) as a Bayesian probabilistic function of the factors viz. characteristics of microenterprises, type of microenterprises, training, technology, Employees no., raw material availability and the market knowledge.

The Naïve Bayes conditional probabilities for different factors ($P(Market\ Knowledge\ |Income)$, $P(Training\ |\ Income)$,($P\ Technology\ |Income)$, $P(characteristics\ of\ microenterprises\ |\ Income)$ $(Type\ of\ microenterprises\ |Income)$, $P(Employees\ No.\ |\ Income)$) have been presented in Figure 5 to Figure 10.

Figure 5. Bayes conditional probability for 5 different levels of Market knowledge: *P (Market Knowledge | Income)*: boxes' heights denote the probabilities' values in relative terms.

Figure 5 shows the Bayes conditional probability for five different levels of market knowledge. It is clear from Figure 5 that the microenterprises earning high profits (HP) have the greatest probability for a higher level (5th level) of market knowledge ($P(MK_5|HP) = 0.89$), and the microenterprises running into losses have the highest probability of possessing the lowest level (1st level) of market knowledge (($P(MK_1|L) = 0.99$). The microenterprises running into profits have a higher probability for 3rd level of market knowledge ($P(MK_3|P) = 0.61$), while microenterprises under breakeven condition have a higher probability for the 1st and 2nd level of market knowledge ($P(MK_2|HP) = 0.62$). Thus, the performance of enterprises is governed by the level of market knowledge possessed by them.

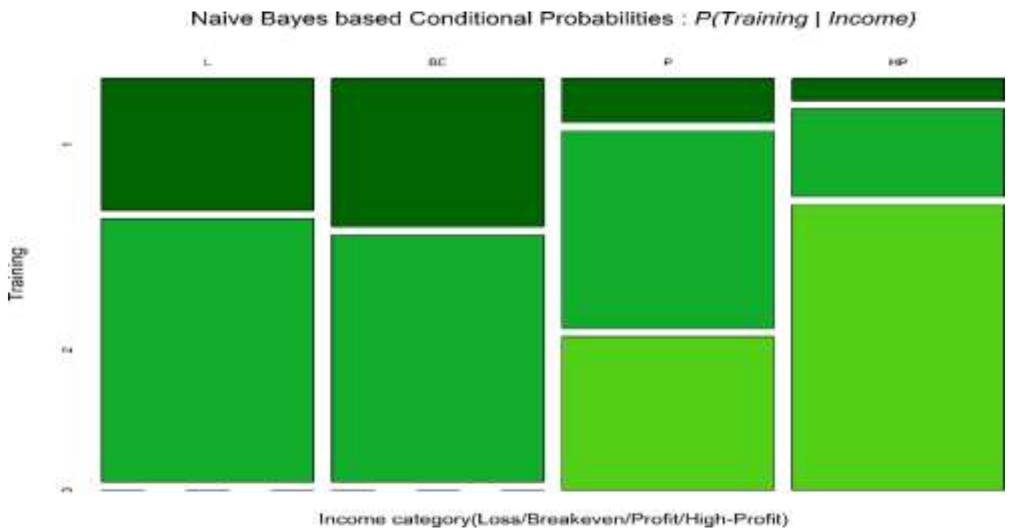

Figure 6. Bayes conditional probability for 3 different levels of training received by the personnels in the microenterprises: *P(Training | Income)*: heights of the boxes denote the values of the probabilities in relative terms.

Figure 6 shows the Bayes conditional probability for 3 different levels of training received by the personnel's in the microenterprises. The microenterprises earning high profits have the greatest probability for the highest level of training (3rd level) ($P(Tr_3|HP) = 0.89$), the microenterprises earning the profits (P) have the higher probability for 2nd and 3rd level of training received ($P(Tr_2|P) = 0.50$; $P(Tr_3|P) = 0.38$) while the microenterprises running into losses or under breakeven condition have the higher probability for the lower level, or no training (1st / 2nd level) received ($P(Tr_1|L) = 0.4$; $P(Tr_2|L) = 0.6$). Thus, the level of training received by the personnel in the microenterprises is also an important factor affecting the performance of microenterprises.

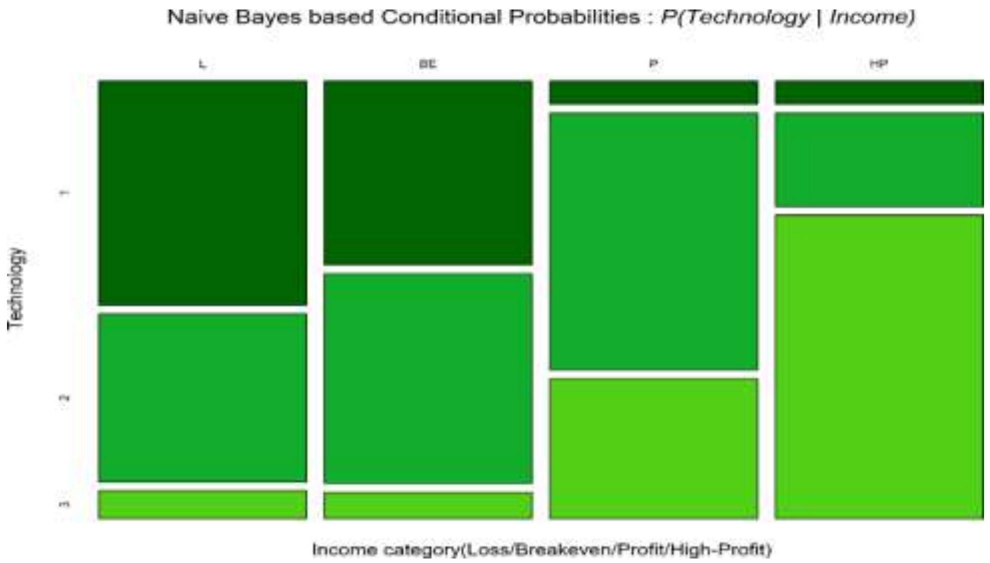

Figure 7. Bayes conditional probability for 3 different levels of technology used by the microenterprises: *P(Tech | Income level)*: heights of the boxes denote the values of the probabilities in relative terms.

Figure 7 shows the Bayes conditional probability for the 3 levels of technology possessed by the microenterprises. The microenterprises earning high profits have the greatest probability for the highest level of technology (3rd level) ($P(Tech_3|HP) = 0.78$), the microenterprises earning the profits (P) have the higher probability for 2nd and 3rd level of technology ($P(Tech_2|P) = 0.61$; $P(Tech_2|P) = 0.39$) while the microenterprises running into losses or under breakeven condition have the higher probability for the lower level of technology (1st / 2nd level) received ($P(Tech_1|L) = 0.53$; $P(Tech_2|L) = 0.4$). Therefore, the level of technology used by microenterprises significantly affects their performances.

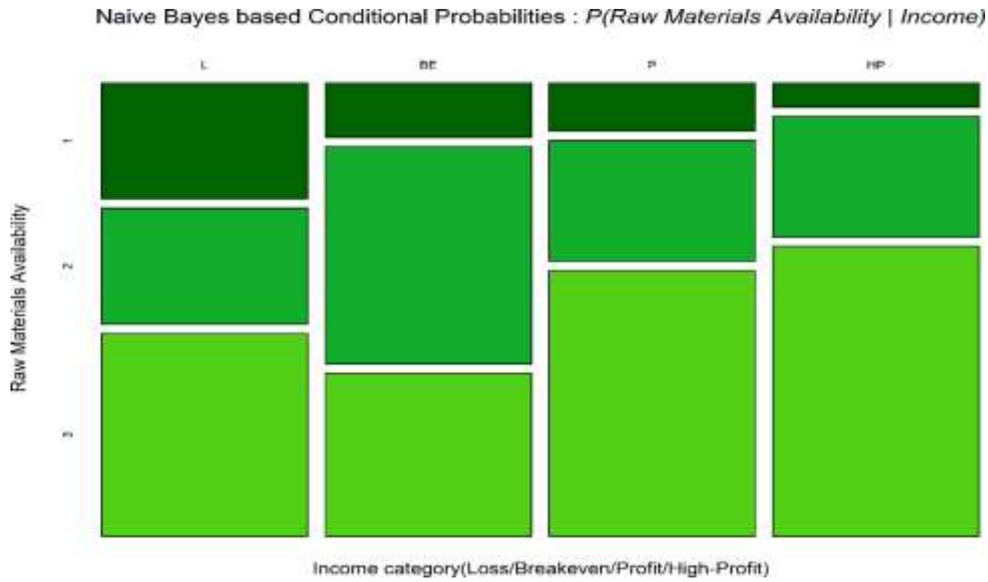

Figure 8. Bayes conditional probability for 3 different distances for raw material availability, 3 indicates the nearest and 1 indicates the farthest from the microenterprises: *P(RW | Income level)*: boxes heights denote the probabilities' values in relative terms.

In the case of raw material availability (Figure 8), the microenterprises earning high profits or profit have a higher probability of them being near the place of raw material availability ($P(RW_3|HP) = 0.72$; $P(RW_3|P) = 0.67$); however, the microenterprises running into losses and under breakeven condition have probability of being close to raw material availity places nearly 50% ($P(RW_3|BE) = 0.50$; $P(RW_3|L) = 0.53$. This is because most of the microenterprises have been started and located near the place of their raw material availability.

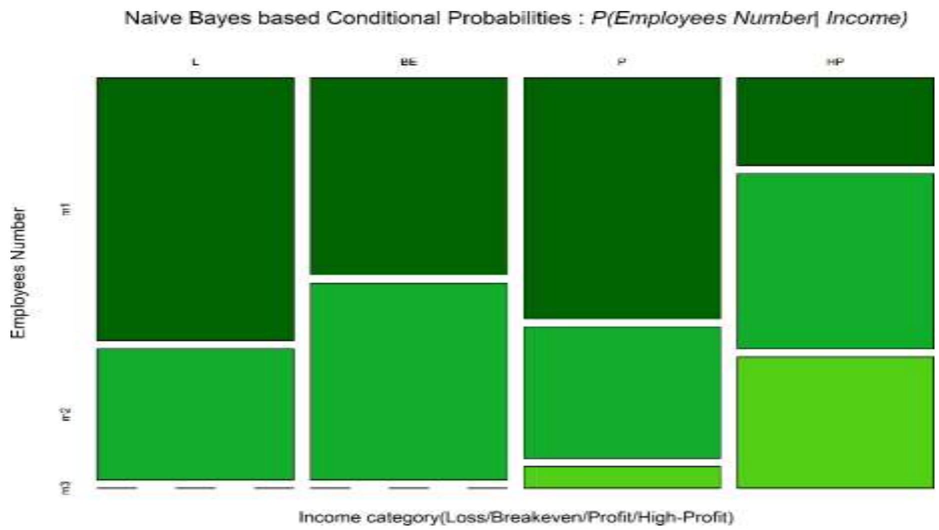

Figure 9. Bayes conditional probability for 3 different ranges of the number of employees in the microenterprise, *m1* indicates single employee, *m2* indicates no. of employees between 2 and 10, *and m3* indicates no. employees greater than 10: *P (Employees No. | Income level)*: the boxes' heights denote the probabilities' values in relative terms.

Regarding no. of employees (Figure 9), the microenterprises running into losses have shown a higher probability for low no. of employees or single employees ($P(m_1|L) = 0.72$). Microenterprises earning high profits have shown a higher probability for moderate no. of employees ($P(m_2|P) = 0.44$).

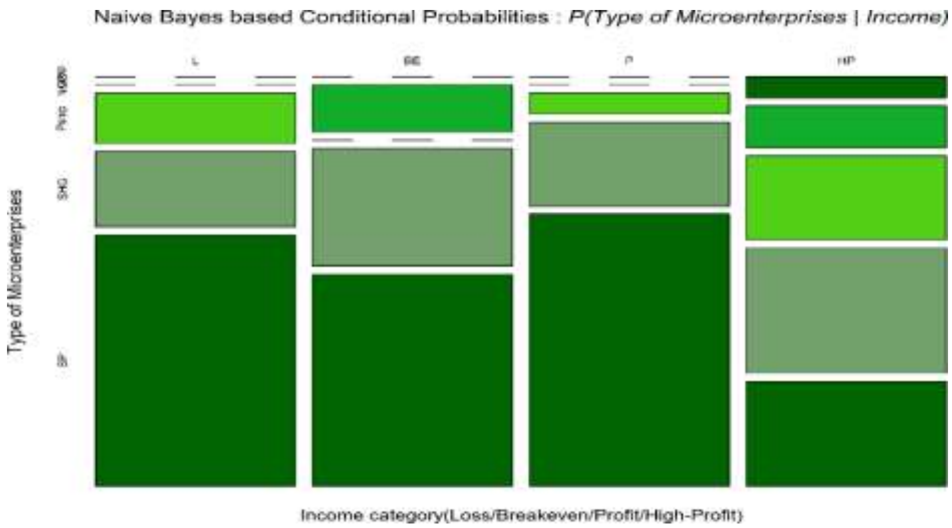

Figure 10. Bayes conditional probability for 3 different types of ownership in the microenterprise, SP: sole propriety, SHG: Self Help group, Pship: Partnership, NGO, coop: cooperative; *P(Type of Microenterprises | Income level)*: heights of the boxes denote the values of the probabilities in relative terms.

It is to be noted that the type of ownership of microenterprises (Figure 10), the microenterprises earning profits show a high probability for either sole propriety (SP) or the self-help groups (SHG) ($P(SP|HP) = 0.44$; $P(SHG|HP) = 0.28$; $P(SP|P) = 0.66$; $P(SHG|P) = 0.22$). However, microenterprises running into losses have also shown higher probability for sole propriety(SP) or self-help group (SHG) ($P(SP|L) = 0.66$; $P(SHG|L) = 0.20$).

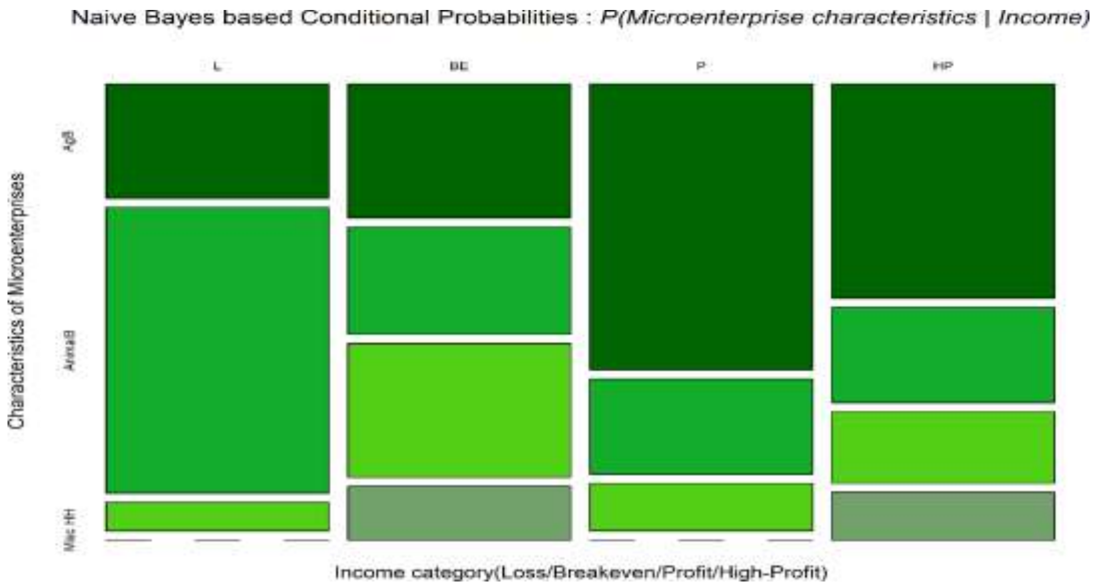

Figure 11. Bayes conditional probability for 4 different microenterprise characteristics, AGB: Agro-based, AnimalB: animal-based, HH: Handloom & Handicrafts, Misc: Miscellaneous; *P(chc_mic | Income level)* : heights of the boxes denote the values of the probabilities in relative terms.

In terms of characteristics of microenterprises (Figure 11), the microenterprises earning profits have shown a higher probability for the agro-based enterprises ($P(AgB|P) = 0.72$; $(AgB|HP) = 0.55$;) while microenterprises running into losses have shown a higher probability for being animal-based ($P(AnimalB|L) = 0.6$). Thus, agro-based enterprises have sustained better performance.

CONCLUSIONS

The present study has identified the factors affecting the performances of natural resources-based microenterprises in the Western Himalayan region of Uttarakhand, India, and evaluated their performances based on the Naïve Bayes classifier approach. To apply the Naïve Bayes classification, the collected data were randomly divided into 75:25 ratios as training and test dataset, and several experiments were carried out on each of these random sample datasets. The Naïve Bayes classification accuracy on the training dataset was found to be 100%, while accuracy on the test dataset ranged from 93% to 100%. The results revealed that agro-based microenterprises have a greater probability (0.67) of making a profit/high profit, while animal product-based microenterprises have a high probability of running into loss. A higher level of Market Knowledge contributes to a very high probability (0.89) of making high profits, while low levels of Market knowledge means there is a high probability (0.87) of making a loss. The higher level of technology and training provides greater chances/probability (0.72, 0.72) of making high profits. Self-help groups (SHGs) have shown a better probability of making profits, while sole propriety has also shown risks of making losses.

The study suggests promoting SHGs in the region and disseminating market knowledge (marketing strategy) among entrepreneurs. The findings also suggest that it is important to invest in leveling up the technology adapted by the microenterprises.

Natural Resources are in abundance in the western Himalayan region. Therefore, their proper use and establishment of more micro-enterprises will help in building human capacity. The western Himalayan region, being diverse in topography, and biodiversity, has enormous prospects for sectors like Agro-based industry as food processing, floriculture, horticulture, processing of honey, herbal and medicinal plants, Khadi, Forest based enterprises along with Handicrafts, Handlooms, wool-based industry, adventure sports, Hotels, Tour and travels. The present study provides an outlook on how these micro enterprises can be strengthened in the region.

Author Contributions: Conceptualization, I.B., U.K., K.A. V.S. and A.S.; Methodology, I.B.; Software, I.B.; Validation, I.B.; Formal Analysis, I.B., U.K., K.A. V.S. and A.S.; Investigation, I.B., U.K., K.A. V.S. and A.S.; Resources, I.B., U.K., K.A. V.S. and A.S.; Data Curation, I.B.; Writing – Original Draft Preparation, I.B., U.K., K.A. V.S. and A.S.; Writing – Review & Editing, I.B., U.K., K.A. V.S. and A.S.; Visualization, I.B.; Supervision, K.A.; Project Administration, K.A.; Funding Acquisition, I.B., U.K., K.A. V.S. and A.S. Authors have read and agreed to the published version of the manuscript.

Institutional Review Board Statement: Ethical review and approval were waived for this study due to the research does not deal with vulnerable groups or sensitive issues.

Funding: The research work is funded by the Ministry of Environment, Forests, and Climate Change (MoEFCC), Government of India through GB Pant National Institute of Himalayan Environment (GBPNIHE), India under the aegis of National Mission on Himalayan Studies (NMHS).

Acknowledgments: Authors would like to acknowledge the financial support given by the Ministry of Environment, Forests, and Climate Change (MoEFCC), Government of India through GB Pant National Institute of Himalayan Environment (GBPNIHE), India under the aegis of National Mission on Himalayan Studies (NMHS), for conducting this study and Doon University for providing all the necessary facilities required of a research center. Most importantly the authors would like to thank the owners of micro enterprises from all across the study areas and all respondents of the study for being welcoming and forthcoming in responding to the researchers' queries and giving their valuable time for the survey. Without their support and cooperation this research work wouldn't have been possible.

Informed Consent Statement: Informed consent was obtained from all subjects involved in the study.

Data Availability Statement: The data used in the study has been made publicly available through the DOI: 10.4121/22010042. The R codes used for the analysis of the data has been made publicly available via DOI: 10.4121/22010111.

Conflicts of Interest: The authors declare no conflict of interest.

REFERENCES

Agresti, A. (1988). A Model for Agreement Between Ratings on an Ordinal Scale. *Biometrics*, *44*(2), 539–548. https://doi.org/10.2307/2531866

Alom, F., Abdullah, M. A., Moten, A. R., & Azam, S. M. F. (2016). Success factors of overall improvement of microenterprises in Malaysia: an empirical study. *Journal of Global Entrepreneurship Research*, *6*(1), 1–13. https://doi.org/10.1186/s40497-016-0050-2

Andersen, A. D., Marìn, A., & Simensen, E. O. (2018). Innovation in natural resource-based industries: A pathway to development? Introduction to the special issue. *Innovation and Development*, *8*(1), 1–27. https://doi.org/10.1080/2157930X.2018.1439293

Anderson, D. (1982). Small industry in developing countries: A discussion of issues. *World Development*, *10*(11), 913–948. https://doi.org/10.1016/0305-750X(82)90034-1

Anthwal, A., Gupta, N., Sharma, A., Anthwal, S., & Kim, K. H. (2010). Conserving biodiversity through traditional beliefs in sacred groves in Uttarakhand Himalaya, India. *Resources, Conservation, and Recycling*, *54*(11), 962–971. https://doi.org/10.1016/j.resconrec.2010.02.003

Aragón-Sánchez, A., Barba-Aragón, I., & Sanz-Valle, R. (2010). Effects of training on business results. *The International Journal of Human Resource Management*, *14*(6), 956–980.

Berrone, P., Gertel, H., Giuliodori, R., Bernard, L., & Meiners, E. (2014). Determinants of Performance in Microenterprises: Preliminary Evidence from Argentina. *Journal of Small Business Management*, *52*(3), 477–500. https://doi.org/10.1111/jsbm.12045

Boermans, M. A., Willebrands, D., & Willebrands, D. (2012). Financial Constraints, Risk Taking and Firm Performance: Recent Evidence from Microfinance Clients in Tanzania (November 19, 2012). De Nederlandsche Bank Working Paper No. 358. http://dx.doi.org/10.2139/ssrn.2177842

Bhandari, G., & Reddy, B. V. (2015). Impact of Out-Migration on Agriculture and Women Work Load: An Economic Analysis of Hilly Regions of Uttarakhand India. *Indian Journal of Agricultural Economics*, *70*(3), 1-10.

Cardamone, M., & Rentschler, R. (2018). Indigenous innovators: The role of web marketing for cultural micro-enterprises. *International Journal of Nonprofit and Voluntary Sector Marketing*, *11*(4), 347–360. https://doi.org/10.1002/nvsm.278

Dhiman, P. K., & Rani, A. (2011). Problems and prospects of small scale agro-based industries: An analysis of Patiala district. *International journal of multidisciplinary research*, *1*(4), 129-142.

Deshpandé, R., & Farley, J. U. (1998). Measuring market orientation: generalization and synthesis. *Journal of market-focused management*, *2*, 213-232. https://doi.org/10.1023/A:1009719615327

Directorate of Economics and Statistics Uttarakhand. (2017). State Domestic Product of Uttarakhand (2011-12 to 2016-17th (Disaggregated Data 2011-12 to 2015-16th) with Base Year, 2011-12), Dept of Planning, Govt of Uttarakhand Retrieved from https://des.uk.gov.in/files/GSDP_BOOK_2016-17.pdf

Domingos, P., & Pazzani, M. (1997). On the optimality of the simple Bayesian classifier under zero-one loss. *Machine learning*, *29*, 103-130. https://doi.org/10.1023/A:1007413511361

Duda, R. O., Hart, P. E., & Stork, D. G. (1973). *Pattern classification and scene analysis* (Vol. 3, pp. 731-739). New York: Wiley.

Datta, D. B., & Bhattacharyya, S. (2016). An Analysis of Problems and Prospects of Indian Handicraft Sector. *Asian Journal of Management*, *7*(1) 5–16. https://doi.org/10.5958/2321-5763.2016.00002.0

Gandhi, V., Kumar, G., & Marsh, R. (1999). Agroindustry for rural and small farmer development: issues and lessons from India. *The International Food and Agribusiness Management Review*, *2*(3-4), 331–344. https://doi.org/10.1016/S1096-7508(01)00036-2

Ghafoor Khan, R., Ahmed Khan, F., & Aslam Khan, M. (2011). Impact of Training and Development on Organizational Performance. *Global Journal of Management and Business Research*, *11*(7), 63–68. Retrieved from https://www.google.com/url?sa=i&rct=j&q=&esrc=s&source=web&cd=&cad=rja&uact=8&ved=0CAIQw7AJahcKEwigqorD7Nv_AhUAAAAAHQAAAAAQAg&url=https%3A%2F%2Fglobaljournals.org%2FGJMBR_Volume11%2F8-Impact-of-Training-and-Development-on-Organizational-Performance.pdf&psig=AOvVaw1iSgOgXFQWq3WYyoFBB-KB&ust=1687694077241085&opi=89978449

Gulyani, S., & Talukdar, D. (2010). Inside Informality: The Links between Poverty, Microenterprises, and Living Conditions in Nairobi's Slums. *World Development*, *38*(12), 1710–1726. https://doi.org/10.1016/j.worlddev.2010.06.013

Industries (Development and Regulation) Act. (1951). Retrieved from https://legislative.gov.in/sites/default/files/A1951-65.pdf

Islam, M. M., Habes, E. M., & Alam, M. M. (2018). The usage and social capital of mobile phones and their effect on the performance of microenterprise: An empirical study. *Technological Forecasting and Social Change*, *132*, 156-164. https://doi.org/10.1016/j.techfore.2018.01.029

Kathuria, S. (1986). Handicrafts Exports: An Indian Case Study. *Economic and Political Weekly*, *21*(40), 1743–1755. Retrieved from https://www.jstor.org/stable/4376184

Kamunge, M. S., Njeru, A., & Tirimba, O. I. (2014). Factors affecting the performance of small and micro enterprises in Limuru Town Market of Kiambu County, Kenya. *International journal of scientific and research publications*, *4*(12), 1-20.

Kalita, M., & Prosad, P. (2016). Impact of Globalisation on Bell Metal Industry of Sarthebari Barpeta (with special reference to local artisans). *Apeejay- Journal of Management Sciences and Technology*, *4*(1), 72-80.

Loader, K., & Johnston, K. (2003). Encouraging SME participation in training: Identifying practical approaches. *Journal of European Industrial Training*, *27*(6), 273–280. https://doi.org/10.1108/03090590310479901

Lee, S. K., Cho, Y. H., & Kim, S. H. (2010). Collaborative filtering with ordinal scale-based implicit ratings for mobile music recommendations. *Information Sciences*, *180*(11), 2142–2155. https://doi.org/10.1016/j.ins.2010.02.004

Mamgain, R. P., & Reddy, D. N. (2017). Out-migration from the hill region of Uttarakhand: Magnitude, challenges, and policy options. *Rural labor mobility in times of structural transformation: Dynamics and perspectives from Asian economies*, 209–235. Retrieved from https://link.springer.com/chapter/10.1007/978-981-10-5628-4_10

Martin, R. U., & Alejandro, M. (2016). The Role of Education and Learning by Experience in the Performance of Microenterprises. *Procedia - Social and Behavioral Sciences*, *228*(June), 523–528. Retrieved from https://doi.org/10.1016/j.sbspro.2016.07.080

Masakure, O., Henson, S., & Cranfield, J. (2009). Performance of microenterprises in Ghana: A resource-based view. *Journal of Small Business and Enterprise Development*, *16*(3), 466–484. https://doi.org/10.1108/14626000910977170

Meyer, D., Dimitriadou, E., Hornik, K., Weingessel, A., Leisch, F., Chang, C. C., & Lin, C. C. (2014). e1071: Misc functions of the Department of Statistics (e1071), TU Wien. *R package version*, *1*(3), 9. Retrieved from https://CRAN.R-project.org/package=e1071

Micro, Small, and Medium Enterprises Development (MSMED) Act (2006). MSME development institute, Ministry of MSME, & GOI. (2017). *Annual Report 2016-17*. Retrieved from https://legislative.gov.in/sites/default/files/A2006-27.pdf

Murty, M. N., & Devi, V. S. (2011). *Pattern recognition: An algorithmic approach*. Springer Science & Business Media.

Nardi, P. M. (2018). *Doing Survey Research* (4th ed.). Taylor and Francis.

Paramasivan, C., & Pasupathi, R. (2016). Performance of agro-based industries in India. *National Journal of Advanced Research*, *2*(6), 25-28.

Psaltopoulos, D., Stathopoulou, S., & Skuras, D. (2005). The location of markets perceived entrepreneurial risk, and start-up capital of micro rural firms. *Small Business Economics*, *25*(2), 147–158. https://doi.org/10.1007/s11187-003-6456-6

Parichatnon, S., & Maichum K, P. (2018). Measuring technical efficiency of Thai rubber production using the three-stage data envelopment analysis. *Agricultural Economics*, *64*(5), 227–240. https://doi.org/10.17221/19/2016-AGRICECON.

R Core Team (2021). R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. Retrieved from https://www.R-project.org

Rosa, P., Carter, S., & Hamilton, D. (1994). Gender as a Determinant of Small Business Performance: Preliminary Insights from a British Study. *National Small Firms Policy and Research Conference*, 271–288.

Sohns, F., & Revilla Diez, J. (2018). Explaining micro-entrepreneurship in rural Vietnam—a multilevel analysis. *Small Business Economics*, *50*(1), 219–237. https://doi.org/10.1007/s11187-017-9886-2

Stone, D. L., & Deadrick, D. L. (2015). Challenges and opportunities affecting the future of human resource management. *Human Resource Management Review*, *25*(2), 139–145. https://doi.org/10.1016/j.hrmr.2015.01.003

Shahidullah, A. K. M., & Haque, C. E. (2014). Environmental orientation of small enterprises: Can microcredit-assisted microenterprises be "green"? *Sustainability*, *6*(6), 3232-3251. https://doi.org/10.3390/su6063232

Sokolova, M., & Lapalme, G. (2009). A systematic analysis of performance measures for classification tasks. *Information Processing and Management*, *45*(4), 427–437. https://doi.org/10.1016/j.ipm.2009.03.002

Sánchez, A. A., Aragón, I. B., & Valle, R. S. (2003). Effects of training on business results, *The International Journal of Human Resource Management, 14*(6), 956–980. https://doi.org/10.1080/0958519032000106164

Tambunan, T. (2008). The Role of Government in Technology Transfer to SME Clusters in Indonesia: Micro-level Evidence from the Metalworking Industry Cluster in Tegal (Central Java). *Asian Journal of Social Science*, *36*, 321–349.

Thapa, A. (2015). Determinants of microenterprise performance in Nepal. *Small Business Economics*, *45*(3), 581–594. https://doi.org/10.1007/s11187-015-9654-0

U.S. Small Business Administration. (2010). Program for investment in microentrepreneurs act (‘‘PRIME’’). Retrieved from http://www.sba.gov/sites/default/files/files/serv_fa_2010_ primetrack123.pdf

Uematsu, H., & Mishra, A. K. (2011). Use of Direct Marketing Strategies by Farmers and Their Impact on Farm Business Income. *Agricultural and Resource Economics Review*, *40*(1), 1–19. https://doi.org/10.1017/S1068280500004482

Vershinina, N., Markman, G., Han, L., Rodgers, P., Kitching, J., Hashimzade, N., & Barrett, R. (2022). Gendered regulations and SME performance in transition economies. *Small Business Economics*, *58*(2), 1113–1130. https://doi.org/10.1007/s11187-020-00436-7

Webb, G. I., Boughton, J. R., & Wang, Z. (2005). Not so naive Bayes: aggregating one-dependence estimators. *Machine learning*, *58,* 5–24. https://doi.org/10.1007/s10994-005-4258-6