Journal of Information Technology Management
Print ISSN: 2008-5893
Online ISSN: 2423-5059

Exploring the Influence of Microfinance on Entrepreneurship using machine learning techniques

Zeba Malik

Department of commerce and Business Studies, Jamia Millia Islamia, New Delhi, India. E-mail: zeba.11april@gmail.com

Naseeb Ahmad

Department of commerce and Business Studies, Jamia Millia Islamia, New Delhi, India. E-mail: nahmad3@jmi.ac.in

Waseem Ahmed*

*Corresponding Author, Department of Computer Engineering, Jamia Millia Islamia, New Delhi, India. E-mail:waseem.ahmed86@gmail.com

Abstract

Microfinance institutions in India provide a set of financial services to the economically weaker sections. Recently, a large number of microfinance institutions have emerged in India and they have favorable impact for poverty reduction. The impact of these institutions on entrepreneurship and society, needs to be explored in greater depth. The objective of this study is to apply machine learning techniques to explore this impact. The research uses a MIX dataset for three successive years, namely 2017, 2018, and 2019. This dataset comprises eight variables centered on gross loan portfolio. Principal Component Analysis (PCM) has been applied on the sample dataset for dimensionality reduction, resulting in two main components and each component consist of fraction from eight variables. Then, the sample dataset has been labelled with the help of clustering using K-means clustering technique. Further, classification models based on K-Nearest Neighbors (KNN) algorithm and Support Vector Machine (SVM) are applied to predict the appropriate category of entrepreneurship. The experiment result shows that the machine learning techniques have been found effective and useful tools for estimating the impact of microfinance on entrepreneurship in India.

Keywords: Microfinance, Entrepreneurship, Principal Component Analysis (PCM), K-means clustering, K-Nearest Neighbors (KNN), Support Vector Machine (SVM)

Journal of Information Technology Management, 2023, Vol. 15, Special Issue, pp. 139- 156
Published by University of Tehran, Faculty of Management
doi: https://doi.org/ 10.22059/jitm.2023.95250
Article Type: Research Paper

Received: July 06, 2023
Received in revised form: August 24, 2023
Accepted: November 09, 2023
Published online: December 24, 2023

Introduction

Microfinance is elucidated as a setup of parsimony, credit services and fiscal services for helping the poor, in earning income and achieving progress (Dasgupta, 2005; Karmakar, 2008; Madar, 2019). The pertinent issue related to microfinance is its influence on promoting entrepreneurship. Given the fact that not all microfinance institutions are sustainable, it is customary to investigate the factors that influence entrepreneurship. In the Indian context, the focus needs to be on the development of the microfinance institutions and examine their role in fostering economic growth within the country. One of the objective of microfinance institutions is to alleviate poverty and help the poor by enabling them to raise funds in emergencies and access the benefits provided by these institutions. Microfinance institutions offer credit in small proportions and provide training and other relevant services to the poor.

A number of factors influence the microfinance activity (Karmakar, 2008):

i) Entrepreneurship and self-employing ability.

ii) Limitation imposed by lack of financial resources.

iii) Option for savings for the poor.

Microfinance societies and institutions emerged with the aim of assisting farmers. It was in the 1960s that social banking concepts started to gain prominence. However, rural areas continued to grapple with the problem of high interest rates. As a solution, microfinance was introduced as a means of providing production credit.

The concept of microfinance has existed since the beginning of the last century. During that time, the microfinance societies and institutions were emerge to help the farmers. It was in the 1960s that Social banking concepts emerged, however, rural areas still faced the issue of paying high interest rates. Later, microfinance was introduced as a means of providing production credit. The NABARD (National Bank for Agriculture and Rural Development) program that was started in 1992, offered loans without collateral and by following the repayment norms. The introduction of production-cum-consumption loans was an innovative change and thus, rural people get benefits from these microfinance institutions for their financial needs. An overview of the microfinance institutions in the Indian context is presented in (Manzoor, 2021). This study emphasizes higher financial inclusion for the sustainable microfinance institutions. An overview of the productive change in the Indian microfinance institutions for the period 2014 to 2016 using an indexing method is presented in (Ambarkhane et. al., 2019).

The following services are offered by the microfinance institutions:

i) Savings
ii) Credit – production, trade, household loans and investment credit
iii) Insurance or risk fund services
iv) Pension
v) Fund transfer

The emergence of microfinance needs to be analyzed in the view of the modern trends and growth requirements. Since, a large number of poor people reside in India, there is a huge demand for microfinance institutions in India over the coming years.

Microfinance, introduced by Mohammad Yunus in Bangladesh in 1976, received substantial promotion and significant attention for its role in poverty alleviation following the success of Gramin Bank. Gramin Bank started being recognized in Bangladesh as a means of poverty alleviation by promoting entrepreneurship among the rural poor (Islam et. al., 2012). Microfinance once began to be regarded as an important tool for poverty alleviation by providing small-sized loans and encouraging entrepreneurship among the poor people (Bateman, 2010).

There is a specific body of microfinance literature, demonstrating the positive impact of microcredit on development (Morduch, 2016), nonetheless the outcomes lack transformative impact as they remain confined to local areas and specific markets (Banerjee et. al., 2015). Small business investments and existing businesses are experiencing an increase in profitability. However, there is not a substantial rise in household income and consumption. Microcredit programs introduced in Malaysia have effectively reduced poverty and increased income levels among the most economically vulnerable clients (AL-Mamun et. al., 2012). Several studies indicate a lack of correlation between increased borrowing and socio-economic factors, including income. Microfinance is widely popular in many countries; however, it has not been successful in effectively promoting entrepreneurship. The majority of individuals who avail microfinance, do so, primarily for necessity entrepreneurship purposes due to which people often get trapped in poverty cycle (Brutin et. al., 2015). Microfinance aids in poverty alleviation through entrepreneurship, but this is applicable only in the case of opportunity-driven entrepreneurship under specific conditions (Kiru, 2007). It has been observed that microfinance plays a supportive role in facilitating community building as per some literature (Nega & Schneider, 2014). This can serve as a pathway towards entrepreneurship. However, the success of microfinance in fostering entrepreneurship is influenced by the socio-economic conditions of the countries involved. According to Singh. J. et. al. (2022), if the multiple group members take the loan together, to fulfill the sustenance need then the loan impact may be more effective.

Microfinance institutions have been active in India for decades and their impact on reducing poverty has been shown in the literature (Mengstie, 2023; Ranabahu, et al., 2022; Ribeiro et al., 2022). The impact of microfinance institutions on entrepreneurship and society, needs to be explored in greater depth (Dzingirai et al., 2023). A model to obtain the entrepreneurship level from the values of the variables like Gross Loan Portfolio and control variables has been presented in the literature (Lahimer et al., 2013). This can calculate the amount or level of entrepreneurship, given the dataset containing these values of variables. Its efficacy has been also shown for various datasets of different countries.

The objective of this paper is to show the usage of machine learning techniques of classification for predicting a category of entrepreneurship, which will help to explore the impact of microfinance institutions on entrepreneurship. The study utilizes a MIX dataset spanning three consecutive years, specifically 2017, 2018, and 2019. This dataset comprises eight variables related to the gross loan portfolio, since Gross Loan portfolio has a direct impact on entrepreneurship, which has been elucidated in the literature.

Firstly, Principal Component Analysis (PCA) is applied to the sample dataset to reduce dimensionality, resulting in two main components that represent fractions from the eight variables. Subsequently, the sample dataset is labeled using K-means clustering, and classification models based on the K-Nearest Neighbors (KNN) algorithm and Support Vector Machine (SVM) are employed to predict the relevant entrepreneurship category. Finally, the experimental findings demonstrate the effectiveness and utility of machine learning techniques in estimating the impact of microfinance on entrepreneurship in India.

Rest of the paper is as follows: Section 2 discusses about Principal Component Analysis (PCA) to reduce dimensionality and K-means clustering to label the data. In section 3, KNN and SVM classification model are used to predict a category of entrepreneurship. Experiment results on MIX dataset for years 2019, 2018, and 2017 are shown in section 4. Section 5 concludes the paper.

Methodology

Data Exploratory Techniques based on Principal Component Analysis (PCA) and KMeans Clustering on the MIX datasets for the Year 2019, 2018, and 2017

In this section, set of microfinance institution data for the years 2019, 2018 and 2017 are used for analysis (Mix Datasets for Gross Loan Portfolio and related features (in Indian context) are used). A sample of the dataset used for the year 2019 has been shown in table 1.0. The dataset contains eight selected features that are related to the Gross Loan portfolio. The Gross Loan portfolio has been chosen since it has a direct impact on entrepreneurship, which has been elucidated in the literature.

Statistical information of the Mix Datasets for the selected eight features for the years 2019, 2018, and 2017 are shown in tables 2, 3 and 4 respectively. The statistical information that has been shown in tables, contains minimum value, maximum value, and standard deviation, total count of rows, 25% percentile, 50% percentile and 75% percentile value for the features used in the dataset.

Table 1. Sample dataset of Gross Loan portfolio features for the year 2019

index	Name of the microfinance Institute	Average gross loan Portfolio	Gross loan Portfolio	Gross loan Portfolio, Location, Rural	Gross loan Portfolio, Location, Urban	Gross loan Portfolio, Methodology, Village Banking SHG	Gross loan Portfolio, Methodology, Individual	No of active borrowers, Location, Rural	No of active borrowers, Location, Urban
0	Adhikar (101858)	30381610	35142736	..	15417349	..	..	84063	64092
1	Annapurna Cooperative (115904)	18404259	22764071	922184	21841887	922184	21841887	3443	59274
2	Annapurna Microfinance (163990)	3.34E+08	4.33E+08	3.67E+08	66519194	4.06E+08	27658714	1288720	192567
248	Madura (101205)	2.25E+08	2.55E+08	2.49E+08	5942811	..	1472065	946385	10458
249	M-power (157899)	30248227	32318164	22990995	9327170	32318164	..	82528	35062
250	Village Financial (100033)	1.27E+08	1.53E+08	1.16E+08	37063314	..	..	359841	115447

Table 2. Statistical information for the year 2019

	Average gross loan Portfolio	Gross loan Portfolio	Gross loan Portfolio, Location, Rural	Gross loan Portfolio, Location, Urban	Gross loan Portfolio, Methodology, Village Banking SHG	Gross loan Portfolio, Methodology, Individual	No of active borrowers, Location, Rural	No of active borrowers, Location, Urban
Count	2.50000e+02	2.50000e+02	2.50000e+02	2.50000e+02	2.50000e+02	2.50000e+02	2.50000e+02	2.50000e+02
Mean	3.002372e+08	3.290223e+08	7.430952e+07	3.509008e+07	5.927749e+07	1.063620e+07	3.134633e+05	1.1009956e+05
Std	5.330129e+08	5.695432e+08	7.573018e+07	3.77645e+07	2.359194e+07	6.054631e+06	2.413989e+05	9.691248e+04
Min	3.540850e+05	4.541480e+05	1.870420e+05	1.248000e+03	0.000000e+00	0.000000e+00	8.170000e+02	2.400000e+01
25%	2.115776e+08	2.704437e+08	7.430952e+07	3.509008e+07	5.927749e+07	1.063620e+07	3.134633e+05	1.109956e+05
50%	2.115776e+08	2.704437e+08	7.430952e+07	3.509008e+07	5.927749e+07	1.063620e+07	3.134633e+05	1.109956e+05
75%	2.115776e+08	2.704437e+08	7.430952e+07	3.509008e+07	5.927749e+07	1.063620e+07	3.134633e+05	1.109956e+05
Max	6.404966e+09	6.800472e+09	1.048148e+09	4.211077e+08	4.056911e+08	9.138559e+07	2.383269e+06	1.169785e+06

Table 3. Statistical information for the year 2018

	Average gross loan Portfolio	Gross loan Portfolio	Gross loan Portfolio, Location, Rural	Gross loan Portfolio, Location, Urban	Gross loan Portfolio, Methodology, Village Banking SHG	Gross loan Portfolio, Methodology, Individual	No of active borrowers, Location, Rural	No of active borrowers, Location, Urban
Count	2.50000e+02	2.50000e+02	2.50000e+02	2.50000e+02	2.50000e+02	2.50000e+02	2.50000e+02	2.50000e+02
Mean	1.655857e+08	2.093858e+08	6.016222e+07	3.489767e+07	8.283992e+06	1.074791e+07	1.788838e+05	1.031266e+05
Std	3.349745e+08	4.983069e+08	6.506691e+07	3.769957e+07	1.798486e+07	1.027627e+07	2.054890e+05	1.112314e+05
Min	2.464430e+05	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
25%	2.481394e+07	2.588052e+0	6.016222e+07	3.489767e+07	8.283992e+06	1.074791e+07	1.788838e+05	1.031266e+05
50%	1.655857e+08	2.093858e+08	6.016222e+07	3.489767e+07	8.283992e+06	1.074791e+07	1.788838e+05	1.031266e+05
75%	1.655857e+08	2.093858e+08	6.016222e+07	3.489767e+07	8.283992e+06	1.074791e+07	1.788838e+05	1.031266e+05
Max	3.575803e+09	6.009458e+09	8.410332e+08	5.286614e+08	2.885968e+08	1.188044e+08	2.e+06	1.e+06

Table 4. Statistical information for the year 2017

	Average gross loan Portfolio	Gross loan Portfolio	Gross loan Portfolio, Location, Rural	Gross loan Portfolio, Location, Urban	Gross loan Portfolio, Methodology, Village Banking SHG	Gross loan Portfolio, Methodology, Individual	No of active borrowers, Location, Rural	No of active borrowers, Location, Urban
Count	2.50000e+02	2.50000e+02	2.50000e+02	2.50000e+02	2.50000e+02	2.50000e+02	2.50000e+02	2.50000e+02
Mean	1.251655e+08	1.371371e+08	4.746266e+07	4.252699e+07	4.230496e0+07	9.126956e+06	1.689743e+05	1.515124e+05
Std	2.232705e+08	2.424641e+08	6.963116e+07	1.242018e+08	1.159248e+08	1.553794e+07	2.248482e+05	3.765473e+05
Min	3.586800e+04	5.194000e+03	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00	0.000000e+00
25 %	2.305086e+07	2.142100e+07	3.902071e+07	2.704252e+07	4.230496e+07	9.126956e+06	1.579012e+05	1.070505e+05
50 %	1.251655e+08	1.371371e+08	4.746266e+07	4.252699e+07	4.230496e+07	9.126956e+06	1.689743e+05	1.515124e+05
75 %	1.251655e+08	1.371371e+08	4.746266e+07	4.252699e+07	4.230496e+07	9.126956e+06	1.689743e+05	1.515124e+05
Max	2.471427e+09	2.590194e+09	8.287055e+08	1.966335e+09	1.762077e+09	2.126585e+08	2.861842e+06	5.863518e+06

Figure 1 and Figure 2 presents a comparison of the gross loan portfolio (GLP) mean values and maximum values for the years 2019, 2018, and 2017. It can be seen from figure 1, the mean values of the GLP has an ascending trend with highest GLP mean value in the year 2019. Figure 2 shows that the highest value for the GLP maximum value is in the year 2019.

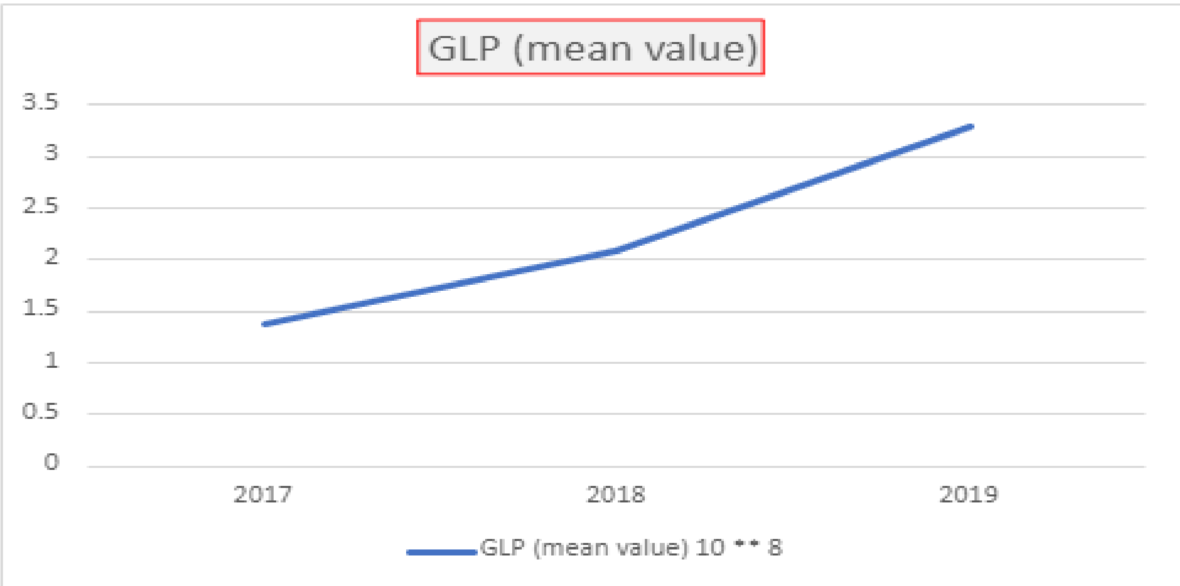

Figure 1. Gross Loan Portfolio (mean values) for the years 2017, 2018 and 2019

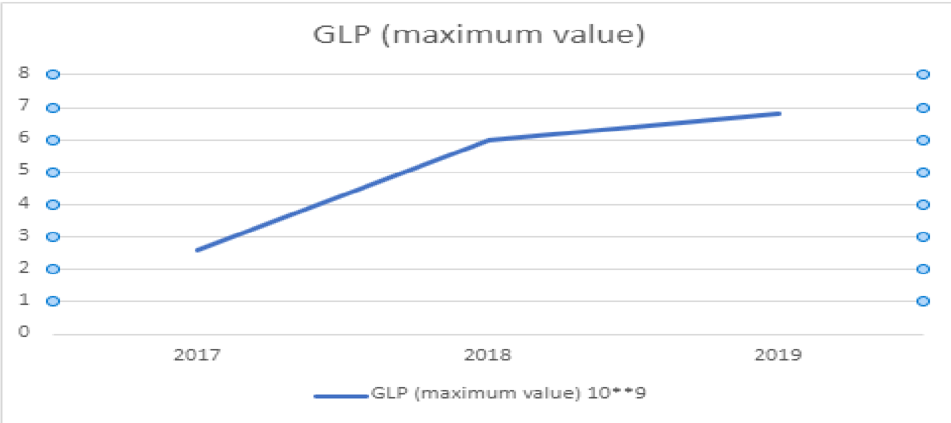

Figure 2. Gross Loan Portfolio (maximum values) for the years 2017, 2018 and 2019

Now, using the GLP dataset that has eight chosen features, the machine learning algorithms, Principal Component Analysis and K-Means clustering are applied to obtain the labels for each sample in the dataset. Then, classification algorithm SVM and KNN algorithms has been applied to the labeled dataset.

Figure 3 gives the proposed workflow used in the paper. Algorithm 1 gives step by step procedure that has been used in this work.

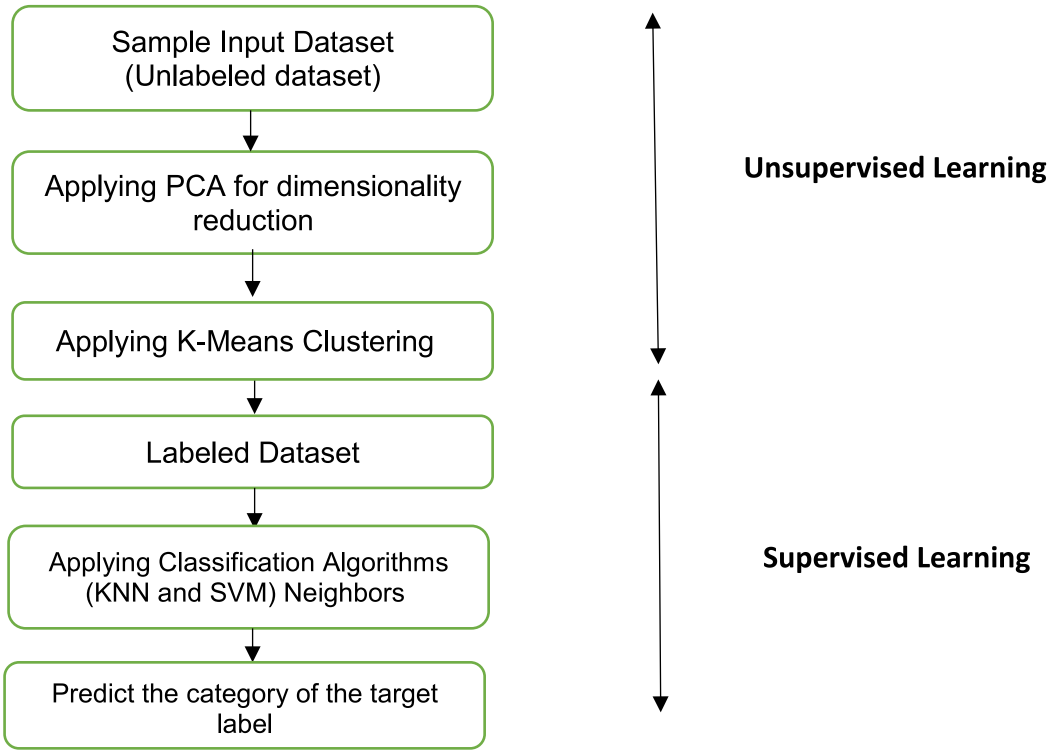

Figure 3. Proposed Methodology

Algorithm 1: Proposed Algorithm	
Step 1: Read the MIX dataset containing 8 features.	
Step 2: PCA is applied to obtain two main principal components, each containing fraction of eight features.	
Step 3: Apply K-Means Clustering on the reduced dataset. Clustering has been used to obtain the labels for each sample in the data set.	
Numbers of clusters obtained= 3 (Label 0, 1, and 2)	
The labels signify the following categories:	
Label 0	High entrepreneurship category,
Label 1	Medium level of entrepreneurship category
Label 2	Low entrepreneurship value
Step 4: KNN and SVM Classification algorithms has been applied on the labeled dataset.	
Step 5: Predict the category of the target label	

II A - Applying PCA (Principal Component Analysis) technique for dimension reduction

Principal Component Analysis (PCA) is a powerful machine learning technique, and has been widely used in machine learning applications for dimensionality reduction (Wold, 1987; Ahmed, 2020). PCA technique has been applied to the sample dataset consisting of eight features, and the output is two principal components- first principal component and second principal component. Scatter plots are used in machine learning applications for visualizing the features graphically. Figure 4 presents a scatter plot showing the first PCA component and second PCA component for the sample dataset of the year 2019.

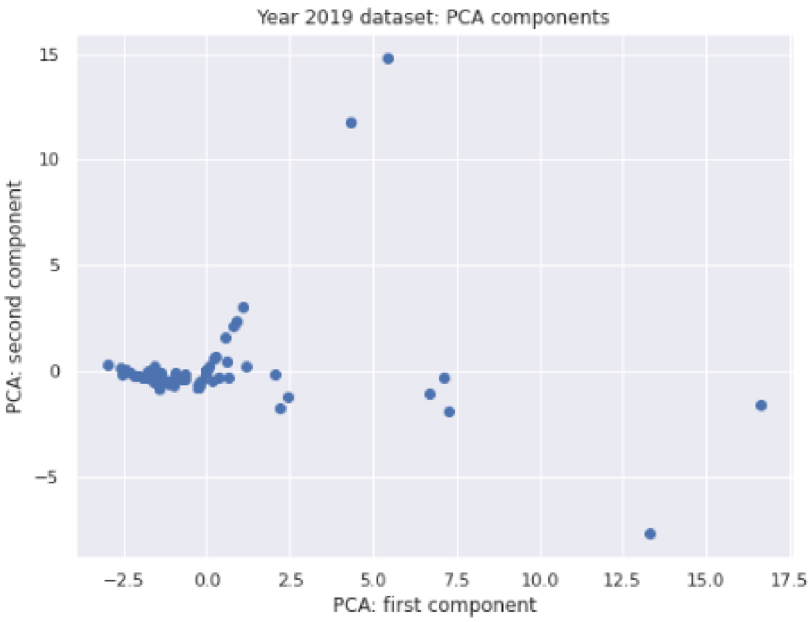

Figure 4. Scatter plot for PCA components of the year 2019

The two main components obtained from the PCA technique for the sample data of GLP features for the year 2019 has been shown in the table 5. Table contains the percentage of the features that are used in each PCA component of the data of 2019.

Table 5. PCA components (Percentage of each feature) year 2019

	Average gross loan Portfolio	Gross loan Portfolio	Gross loan Portfolio, Location, Rural	Gross loan Portfolio, Location, Urban	Gross loan Portfolio, Methodology, Village Banking SHG	Gross loan Portfolio, Methodology, Individual	Number of active borrowers, Location, Rural	Number of active borrowers, Location, Urban
0	0.233574	0.243398	0.40445	0.471029	0.163387	0.299727	0.394145	0.478409
1	0.650794	0.645468	-0.037169	-0.213114	-0.051021	-0.254711	-0.018490	-0.212643

The second principal component uses the first two features in majority. The heatmap has been used for depicting the correlation among the two variables. Figure 5 shows the heatmap for each of the PCA components for the dataset of the year 2019.

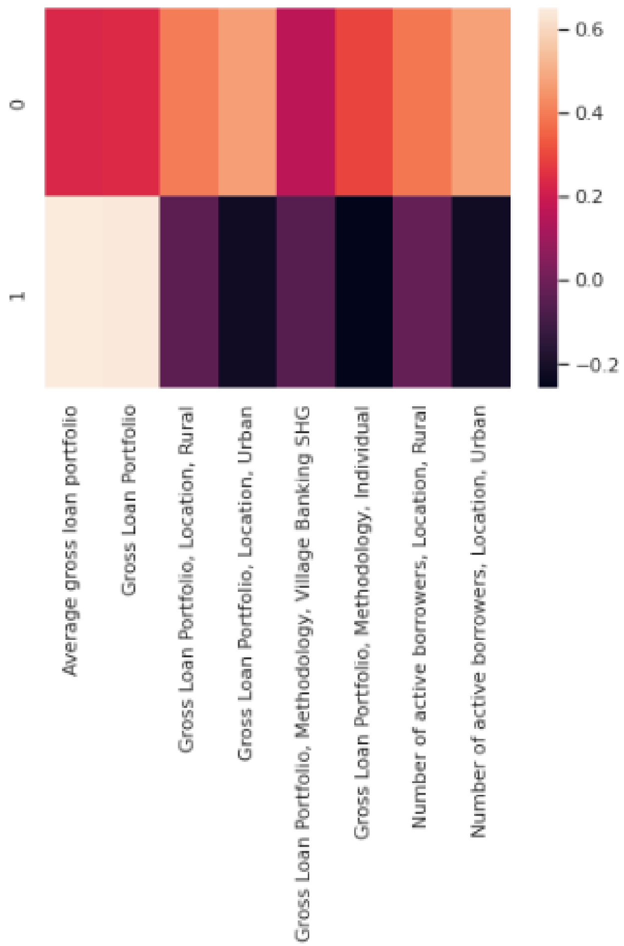

Figure 5. Heatmap for PCA components (Year 2019)

The first principal component has positive correlation among eight features, while the second principal component has positive correlation only for the first two features.

Next, the clustering has been applied next to identify groups based on the values of these two components.

Applying K-Means algorithm for clustering of data

Clustering algorithms use input data for generating different clusters of data points. Next, when a new data point is given as input, this will identify the suitable cluster for this point. This machine learning algorithm learns the attributes from the given data set and puts a new data point using these attributes. All the data points that have similar attributes are grouped together in a distinct cluster. The output commonly has multiple clusters. There could be a single cluster if all the points are similar. The clusters need to be put in different classes. Although, classification involves a training phase and a prediction phase. The output of clustering consists of only the prediction phase.

The clustering has been used to identify clusters. The three clusters that are formulated using the K-Means clustering algorithm have been shown with their clusters centers in the figure 6 (k=3 is obtained using elbow method). The output of the clustering shows a set of labels that are obtained from K-Means technique on the dataset of the year 2019 and these have been shown in Figure 7.

The labels of category 0, or 1, or 2 have been obtained after applying the clustering algorithm. The labels signify the following categories:

Label 0 High entrepreneurship category,

Label 1 Medium level of entrepreneurship category

Label 2 Low entrepreneurship value

$$\begin{bmatrix} [-0.25022275, & -0.05808621] \\ [10.2018271, & -2.50037341] \\ [4.89749592, & 13.30840775] \end{bmatrix}$$

Figure 6. Cluster centers on the data of the PCA components for the year 2019

Figure 7 shows a scatter plot for the clusters obtained after applying K-Means technique (Year, 2019). Label 0, 1 and 2 has been shown in blue, red and green color respectively.

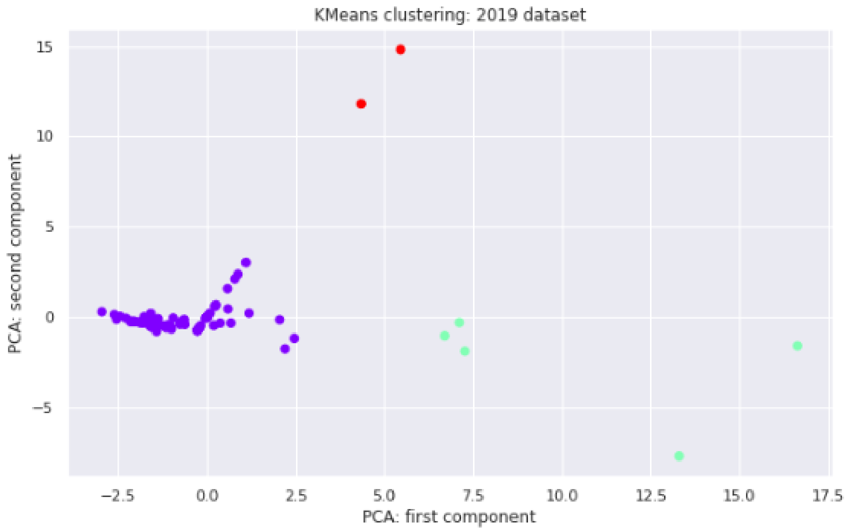

Figure 7. Scatter plot for clusters (Year 2019 data)

Similarly, figure 8 presents a scatter plot showing the first PCA component and second PCA component for the year 2018. Table 6 contains the percentage of the features that are used in each PCA component for the 2018 dataset. Figure 9 shows the heatmap for each of the PCA component for the dataset of the year 2018.

Figure 10 shows the cluster centers and figure 11 shows a scatter plot for the clusters obtained after applying K-Means technique (Year, 2018).

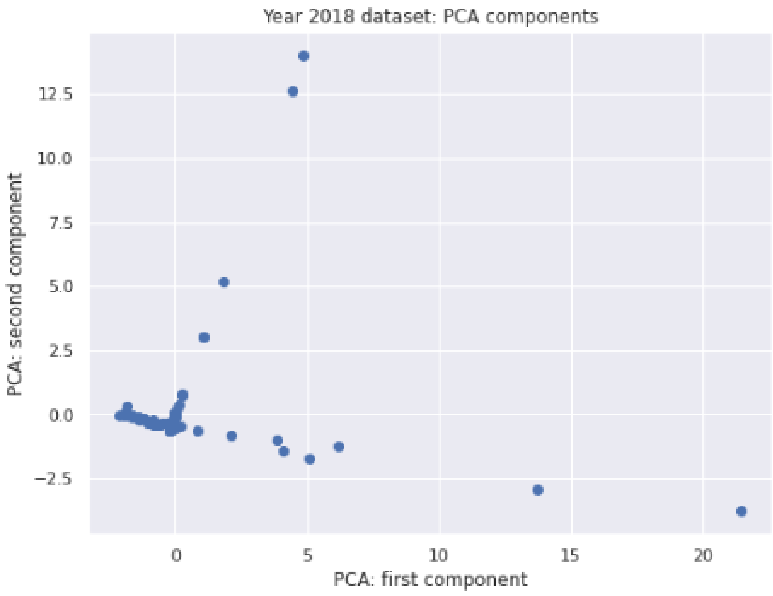

Figure 8. Scatter plot for PCA components of the year 2018

Table 6. PCA components (percentage of each feature) year 2018

	Average gross loan Portfolio	Gross loan Portfolio	Gross loan Portfolio, Location, Rural	Gross loan Portfolio, Location, Urban	Gross loan Portfolio, Methodology, Village Banking SHG	Gross loan Portfolio, Methodology, Individual	Number of active borrowers, Location, Rural	Number of active borrowers, Location, Urban
0	0.242313	0.222924	0.486922	0.476785	-0.038546	0.029855	0.476084	0.445129
1	0.660786	0.674484	-0.163727	-0.170430	0.007396	-0.016927	-0.157412	-0.165711

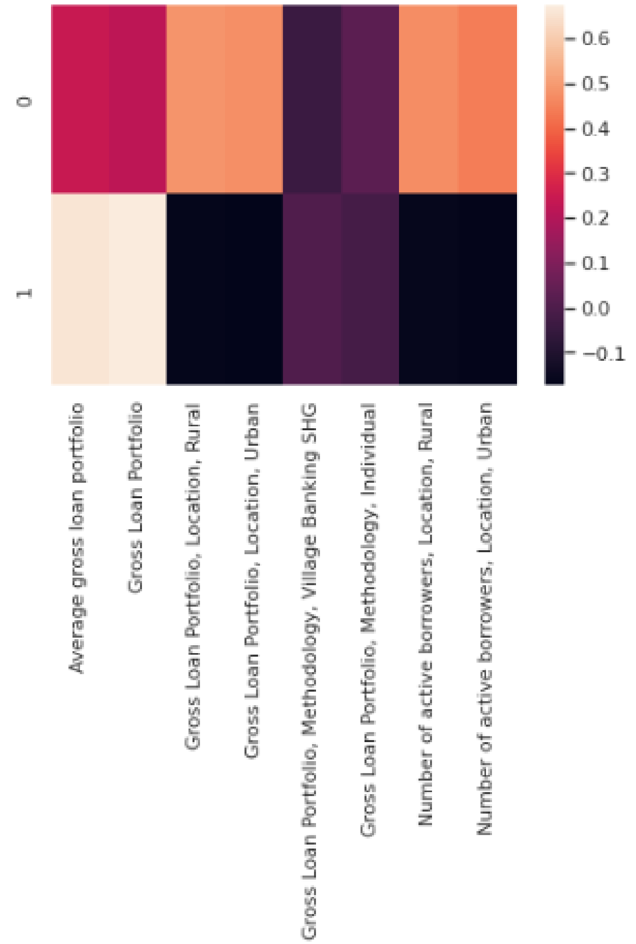

Figure 9. Heatmap for PCA components (Year 2018)

$$\begin{bmatrix} [-0.18001196, & -0.08093001] \\ [17.59259713, & -3.33486231] \\ [4.63888026, & 13.32971852] \end{bmatrix}$$

Figure 10. Cluster centers for the PCA components of the year 2018

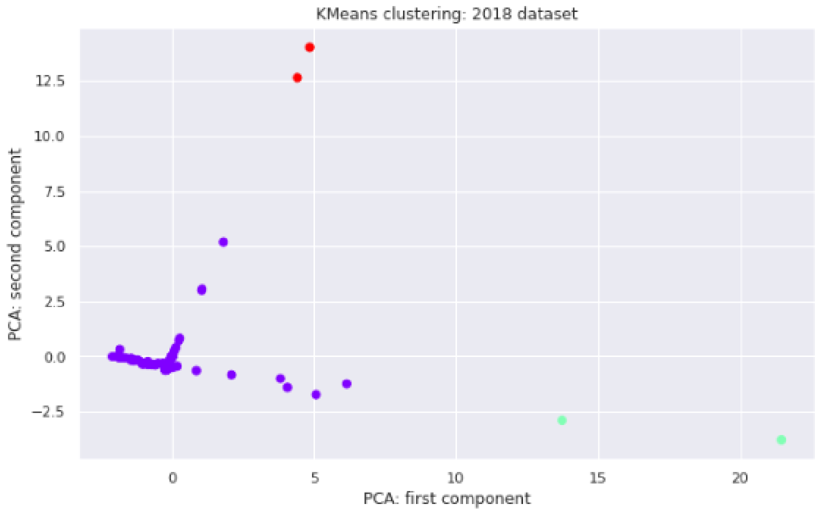

Figure 11. Scatter plot for clusters (Year 2018 data)

Using K-Nearest Neighbors Classifier to predict a category

In the previous section, we presented an overview of microfinance and its influence on entrepreneurship. In this section, we will focus on applying machine learning techniques on a sample dataset. The sample dataset comprises 250 observations which have eight features based on Gross Loan portfolio (GLP) values in the Indian context for year 2019. The database used is Microfinance Information Exchange (MIX) online database. This is composed of microfinance data that provides a comprehensive information of microfinance institutions. A sample set of observations for India, are shown in Table 1.

Results

This section presents the outputs of the machine learning techniques that are used in the study. Machine Learning techniques are applied to predict the category of entrepreneurship. As it can be seen from figure 7 and 11, most of the sample data points are labelled 0, so more sample data points have been added in the datasets for the Labels 1 and Label 2 for balancing the sample data. Label 0 has 244 samples, Label 1 has 70 samples and Label 2 has 59 samples. The data set is split into training and testing data sets, with 70% as training data and 30% as testing data. The confusion matrix for the classifier based on K Nearest Classifier with k=3, for the sample testing data is shown in tables 7 and 8, and the classification report is shown in table 9.

Table 7. Confusion matrix for testing data

86	1	0
0	14	0
0	0	11

Table 8. Precision, Recall and F1-score using KNN

Label	Precision	Recall	F1-score	Support
0	1.00	0.99	0.99	87
1	0.93	1.00	0.97	14
2	1.00	1.00	1.00	11

Table 9. Classification report for classification model on testing data using KNN

Accuracy			0.99	112
Macro avg	0.98	1.00	0.99	112
Weighted avg	0.99	0.99	0.99	112

As seen from classification report the KNN classification model is capable of performing quite well with the given data set. This can predict the output label accurately.

Next, we are using the SVM (Support Vector Machine) linear model to find the class of entrepreneurship.

Tables 10 and 11 shows the confusion matrix, while classification report of the SVM classifier model is shown in table 12.

Table 10. Confusion matrix for SVM linear model classifier

84	1	0
2	14	0
0	0	11

Table 11. Precision, Recall and F1-score using KNN

Label	Precision	Recall	F1-score	Support
0	0.98	0.99	0.98	85
1	0.93	0.88	0.90	16
2	1.00	1.00	1.00	11

Table 12. Classification report of SVM linear model classifier

Accuracy			0.97	112
Macro avg	0.97	0.95	0.96	112
Weighted avg	0.97	0.97	0.97	112

The obtained results illustrate that the K Nearest Classifier model has better prediction outputs than the linear model of SVM classifier. Moreover, the classification model can predict the label accurately. The samples consist of Label 0, Label 1 and Label 2 placed in non-linear fashion. The linear model of SVM has not been able to predict the correct category in some of these cases. Therefore, the K Nearest Classifier has been able to predict more correct cases for this sample dataset.

The classifiers used in this study are capable of predicting a correct category for entrepreneurship based on selected features in the data set. Therefore, machine learning techniques are a suitable choice for predicting a category of entrepreneurship based on appropriately chosen features. Therefore, the models used in this section are beneficial to find the impact of GLP, a microfinance variable on entrepreneurship.

Conclusion

Microfinance institutions have been active in India for decades and their impact on reducing poverty has been shown in the literature. There has been a need to explore its impact on entrepreneurship. A model to obtain the entrepreneurship level from the values of the variables like Gross Loan Portfolio and control variables has been presented in the literature. This can calculate the amount or level of entrepreneurship, given the dataset containing these values of variables. Its efficacy has been also shown for various datasets of different countries. This paper has shown the usage of machine learning techniques of classification for predicting a category of entrepreneurship. Firstly, the Gross Loan portfolio features are chosen for the dataset of Indian microfinance institutions. A labeled dataset has been used in the study that has been obtained after applying PCA and K-Means clustering machine learning techniques. PCA machine learning technique has been widely employed in applications for dimensionality reduction. The chosen eight features are reduced to two main components. Clustering has been applied next to get 3 clusters, in order to get the labelled data. Further, KNN and SVM classifier has been applied on the labeled dataset, for predicting the category of entrepreneurship in the Indian context.

This paper has presented an application of machine learning techniques for building a classification model using different classification algorithm, to predict a category of entrepreneurship from microfinance institutions dataset.

Both these machine learning techniques have been found effective for identifying a category of entrepreneurship and determining the amount of entrepreneurship. Thus, these can prove beneficial and useful tools for estimating the impact of microfinance on entrepreneurship in India.

Conflict of interest

The authors of this paper state that they do not have any competing financial interests or personal relationships that could have influenced their work. We would like to verify that there are no conflicts of interest related to this publication and that no significant financial support has been received that could have influenced the outcome of the research. This statement indicates that the authors have taken steps to ensure that their work is unbiased and free from any undue influence.

Funding

The author(s) received no financial support for the research, authorship, and/or publication of this article.

References

Ahmed, W., Sufyan Beg, M. M., & Ahmad, T. (2018). Entropy based feature selection for fuzzy set-valued information systems. *3D Research*, *9*, 1-17.

Ahmed, W., Beg, M. S., & Ahmad, T. (2020). A fuzzy rough set based fitting approach for fuzzy set-valued information system. *International Journal of Information Technology*, *12*, 1355-1364.

Al-Mamun, A., Malarvizhi, C. A., Hossain, S., & Tan, S. H. (2012). Examining the effect of microcredit on poverty in Malaysia. *ASEAN Economic Bulletin*, 15-28.

Ambarkhane, D., Singh, A. S., & Venkataramani, B. (2018). Measuring total factor productivity change of microfinance institutions in India using Malmquist productivity index. *Indian Growth and Development Review*.

Banerjee, A., Karlan, D., & Zinman, J. (2015). Six randomized evaluations of microcredit: Introduction and further steps. *American Economic Journal: Applied Economics*, *7*(1), 1-21.

Banerjee, A. V., Breza, E., Duflo, E., & Kinnan, C. (2017). Do credit constraints limit entrepreneurship? Heterogeneity in the returns to microfinance. *Heterogeneity in the Returns to Microfinance (September 1, 2017).*

Bateman, M. (2012). Let's not kid ourselves that financial inclusion will help the poor. *The Guardian*, 8.

Bhattacharya, R., & Santra, S. (2018). Financial Inclusion-Microfinance-Current Scenario, It's Crisis and Aftermath. *BULMIM Journal of Management and Research*, *3*(2), 46-54.

Bruton, G., Khavul, S., Siegel, D., & Wright, M. (2015). New financial alternatives in seeding entrepreneurship: Microfinance, crowdfunding, and peer–to–peer innovations. *Entrepreneurship theory and practice*, *39*(1), 9-26.

Dasgupta, R. (2005). Microfinance in India: Empirical evidence, alternative models and policy imperatives. *Economic and Political weekly*, 1229-1237.

Dzingirai, M., & Baporikar, N. (2023). Role of microfinance for entrepreneurial success. In *Research Anthology on Microfinance Services and Roles in Social Progress* (pp. 193-210). IGI Global.

Islam, J., Mohajan, H., & Datta, R. (2012). Aspects of microfinance system of Grameen Bank of Bangladesh.

Karmakar, K. G. (Ed.). (2008). *Microfinance in India*. SAGE Publications India.

Kiiru, J. M. M. (2007). Microfinance: getting money to the poor or making money out of the poor?. *Finance ET Bien Commun*, (2), 64-73.

Kiiru, J. M. M., & Machakos, K. (2007). The impact of microfinance on rural poor households' income and vulnerability to poverty: case study of makueni district, kenya.

Lahimer, N., Dash, S., & Zaiter, M. (2013). Does microfinance promote entrepreneurship and innovation? A macro analysis. *African Journal of Science, Technology, Innovation and Development*, *5*(1), 19-29.

Mader, P. (2013). Rise and fall of microfinance in India: The Andhra Pradesh crisis in perspective. *Strategic Change*, *22*(1-2), 47-66.

Mader, P., & Morvant-Roux, S. (2019). Financial inclusion and microfinance. In *Handbook of Social Policy and Development* (411-430). Edward Elgar Publishing.

Manzoor, A. (2021). The Quest for Financial Inclusion and Microfinance Institutions in India. In *Research Anthology on Personal Finance and Improving Financial Literacy* (358-372). IGI Global.

Mengstie, B. (2023). Entrepreneurship Development through Microfinance: Evidence from Amhara Regional State of Ethiopia. *Journal of the Knowledge Economy*, 1-19.

Morduch, J. (2016), Microfinance and financial inclusion, International Monetary Fund, November.

Nega, B., & Schneider, G. (2014). Social entrepreneurship, microfinance, and economic development in Africa. *Journal of Economic Issues*, *48*(2), 367-376.

Ranabahu, N., & Tanima, F. A. (2022). Disabled women entrepreneurs and microfinance: a road less travelled (for a reason)?. *Research Handbook on Disability and Entrepreneurship*, 196-207.

Ribeiro, J. P. C., Duarte, F., & Gama, A. P. M. (2022). Does microfinance foster the development of its clients? A bibliometric analysis and systematic literature review. *Financial Innovation*, *8*(1), 34.

Roy, M. (2018). MUDRA: The transformation of microfinance in India: Review, experiences and future prospect. *Advances in Finance & Applied Economics*, 93-104.

Singh, J., Dutt, P., & Adbi, A. (2022). Microfinance and entrepreneurship at the base of the pyramid. *Strategic Entrepreneurship Journal*, *16*(1), 3-31.

Tarozzi, A., Desai, J., & Johnson, K. (2015). The impacts of microcredit: Evidence from Ethiopia. *American Economic Journal: Applied Economics*, *7*(1), 54-89.

Wold, S., Esbensen, K., & Geladi, P. (1987). Principal component analysis. *Chemometrics and intelligent laboratory systems*, *2*(1-3), 37-52.

Bibliographic information of this paper for citing:

Malik, Zeba; Ahmad, Ahmad & Ahme, Waseem (2023). Exploring the Influence of Microfinance on Entrepreneurship using machine learning techniques. *Journal of Information Technology Management,* 15 (Special Issue), 139-156. https://doi.org/10.22059/jitm.2023.95250