International Journal of Marketing and Technology
Vol. 16 Issue 01, January 2026
ISSN: 2249-1058 Impact Factor: 6.559
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

# Machine Learning-Enhanced Dea Framework for Performance Grouping and Prediction of Financial Institutions

Ashok Kumar Saka
Research Scholar
Department of Computer Science and Engineering
AcharyaNagarjuna University (ANU), Guntur (Dist)
Andhra Pradesh, India,
sakaashok@gmail.com

Dr.O.Nagaraju
Research Supervisor
Computer Science and Engg, ANU
Associate Professor,
APRDC, Nagarjunasagar-522439
profonr@gmail.com

**ABSTRACT**
This study develops a hybrid framework for evaluating the performance of financial institutions by integrating Data Envelopment Analysis (DEA), Fuzzy C-Means (FCM) clustering, and Machine Learning (ML) techniques. DEA cross-efficiency analysis is first applied using Deposits, Operating Expenses, and Borrowings as inputs, and Investments, Total Income, Interest Earned, and Advances as outputs, to derive peer-based efficiency scores. The normalized consensus score is then clustered via FCM into Low, Medium, and High-performance groups, capturing soft membership across categories. Logistic Regression and Random Forest classifiers are subsequently trained using the original financial indicators to predict these cluster labels. Results indicate that Random Forest achieves higher predictive accuracy and identifies key financial drivers of institutional performance, while Logistic Regression offers interpretability. The proposed DEA–FCM–ML methodology provides a systematic tool for benchmarking, classification, and predictive assessment of financial institutions, aiding regulators and managers in strategic decision-making.

## 1.INTRODUCTION

The performance evaluation of financial institutions has attracted continuous research attention, owing to their central role in mobilizing savings, allocating credit, and ensuring financial stability. Traditional ratio-based measures (e.g., return on assets, cost-to-income ratio) provide partial views and often overlook the multidimensional nature of banking operations. To address this, Data Envelopment Analysis (DEA) has been widely adopted as a non-parametric frontier method capable of handling multiple inputs and outputs simultaneously (Charnes, Cooper, & Rhodes, 1978). DEA has been extensively applied in banking efficiency studies across economies (Berger & Humphrey, 1997), and recent contributions have introduced advanced forms such as cross-efficiency and group DEA to improve discrimination and provide consensus-based rankings. For example, Yang (2025) developed a centralized cross-efficiency evaluation model to capture cooperative behaviors among units, while Kolahdoozi et al. (2024) extended group efficiency DEA to negative data scenarios in banking, highlighting the continuing evolution of DEA in financial contexts.While DEA provides powerful benchmarking, its descriptive nature limits predictive insight. Recent studies therefore integrate DEA with machine learning (ML) approaches to enhance explanatory and forecasting capabilities. Dong et al. (2025) proposed a machine learning-enhanced DEA framework (M-LASSO) that improved robustness in energy economics, while Perroni et al. (2024) demonstrated the benefits of combining relative efficiency models with ML algorithms in organizational performance analysis. Similarly, Zhang et al. (2022) reviewed DEA–ML integration strategies and emphasized their potential in financial applications. These studies confirm that ML can extend DEA from static benchmarking to dynamic prediction, making it more useful for regulators and practitioners.

To complement DEA, clustering techniques have been introduced to classify efficiency outcomes into performance strata. Among these, Fuzzy C-Means (FCM) is particularly relevant because it allows for soft membership, capturing the reality that financial institutions may simultaneously exhibit features of multiple performance categories. Bezdek (1981) first introduced FCM, and more recent developments have addressed stability and uncertainty in financial datasets. Wu et al. (2025) proposed a semi-supervised FCM to enhance clustering quality under limited labeled data, while Cardone et al. (2025) designed an entropy-based online FCM to handle evolving data streams—both demonstrating FCM's adaptability to modern banking datasets.

International Journal of Marketing and Technology
Vol. 16 Issue 01, January 2026
ISSN: 2249-1058 Impact Factor: 6.559
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

## 2.METHODOLOGY

This study employs a hybrid DEA–Machine Learning framework to evaluate and classify the performance of DMUs (e.g., MSMEs, banks). The methodology integrates Data Envelopment Analysis (DEA), Fuzzy C-Means (FCM) clustering, and supervised classifiers (Logistic Regression, Random Forest).

### Step 1: DEA Cross-Efficiency Analysis

The first step computes **cross-efficiency scores** to capture peer-evaluated efficiency of each Decision-Making Unit (DMU).

For each DMU $k$:

$$\max_{u_r, v_i} \; E_k = \sum_{r=1}^{s} u_r y_{rk}$$

Subject to:

$$\sum_{i=1}^{m} v_i x_{ik} = 1$$

$$\sum_{r=1}^{s} u_r y_{rj} - \sum_{i=1}^{m} v_i x_{ij} \leq 0, \quad \forall j = 1, 2, \ldots, n$$

$$u_r, v_i \geq 0$$

- $x_{ij}$: input $i$ for DMU $j$
- $y_{rj}$: output $r$ for DMU $j$
- $u_r, v_i$: weights for outputs and inputs
- $E_k$: efficiency of DMU $k$

The **cross-efficiency matrix** is then constructed:

$$CE_{kj} = \frac{\sum_{r=1}^{s} u_r y_{rj}}{\sum_{i=1}^{m} v_i x_{ij}}$$

where row $k$ represents weights optimized for DMU $k$ but applied to all DMUs $j$.

The **consensus efficiency** (priority score) for DMU $j$ is:

$$C_j = \frac{1}{n} \sum_{k=1}^{n} CE_{kj}, \quad \tilde{C}_j = \frac{C_j}{\sum_{j=1}^{n} C_j}$$

### Step 2: Fuzzy C-Means (FCM) Clustering

Consensus efficiency values $\tilde{C}_j$ are clustered into three groups (**Low, Medium, High**) using FCM.

The objective function:

$$J_m = \sum_{j=1}^{n} \sum_{c=1}^{C} u_{cj}^{m} \left\| x_j - v_c \right\|^2$$

International Journal of Marketing and Technology
Vol. 16 Issue 01, January 2026
ISSN: 2249-1058 Impact Factor: 6.559
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

where:

- $C$: number of clusters (here, $3$)
- $u_{cj}$: degree of membership of DMU $j$ in cluster $c$
- $v_c$: centroid of cluster $c$
- $m > 1$: fuzzifier (controls fuzziness, here $m = 2$)

Membership update:

$$u_{cj} = \frac{1}{\sum_{k=1}^{C} \left( \frac{\|x_j - v_c\|}{\|x_j - v_k\|} \right)^{\frac{2}{m-1}}}$$

Cluster assignment:

$$\text{Label}(j) = \arg\max_{c} \; u_{cj}$$

### Step 3: Supervised Classification

The cluster labels from FCM serve as **class labels** (dependent variable).
The **independent variables** are the original inputs and outputs of DMUs.

#### Logistic Regression

For $K$ classes, multinomial logistic regression models the probability:

$$P(y = j \mid \mathbf{x}) = \frac{\exp(\beta_j^T \mathbf{x})}{\sum_{k=1}^{K} \exp(\beta_k^T \mathbf{x})}, \quad j = 1, 2, \ldots, K$$

Decision rule:

$$\hat{y} = \arg\max_{j} P(y = j \mid \mathbf{x})$$

#### Random Forest

Random Forest is an **ensemble of decision trees**. Each tree predicts a class $\hat{y}_t$, and the final prediction is by majority vote:

$$\hat{y} = \text{mode}\{\hat{y}_1, \hat{y}_2, \ldots, \hat{y}_T\}$$

Feature importance is measured via the average decrease in **Gini impurity**:

$$Imp(f) = \frac{1}{T} \sum_{t=1}^{T} \Delta Gini_t(f)$$

where $\Delta Gini_t(f)$ is the reduction in impurity contributed by feature $f$ in tree $t$.

### Step 4: Performance Evaluation

Model performance is assessed using:

- **Confusion Matrix**
- **Accuracy**:

$$Acc = \frac{\text{Correct Predictions}}{\text{Total Predictions}}$$

- **Precision, Recall, F1-score** (per class)

## 3.CASE STUDY

The Indian banking sector, especially its public sector banks (PSBs), plays a crucial role in the country's economic landscape by mobilizing savings, providing credit, and supporting economic development across various sectors.

International Journal of Marketing and Technology
Vol. 16 Issue 01, January 2026
ISSN: 2249-1058 Impact Factor: 6.559
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

This case study focuses on analyzing the financial performance of 12 prominent public sector banks in India over a ten-year period, from 2015 to 2024. These banks include Bank of Baroda (BOB), Bank of India (BOI), Bank of Maharashtra (BOM), Central Bank of India (CB), Canara Bank (CBI), Indian Bank (IB), Indian Overseas Bank (IOB), Punjab & Sind Bank (PSB), Punjab National Bank (PNB), Union Bank of India (UCB), United Bank of India (UBI), and State Bank of India (SBI).The study evaluates seven critical financial factors—Deposits, Expenses, Borrowings, Investments, Total Income, Interest Earned, and Advances—using data gathered from secondary sources. Data on the financial factors is presented below.

| Banks | DEPOSITS(D) in Crores | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
| Bank of Baroda | 49,007 | 33,815 | 60,168 | 59,131 | 10,05,652 | 11,36,204 | 79,888 | 10,45,939 | 12,03,688 | 13,26,958 |
| Bank of India | 5,31,903 | 5,13,005 | 5,40,032 | 5,92,800 | 5,73,900 | 5,55,050 | 6,27,113 | 6,27,896 | 6,69,586 | 7,37,920 |
| Bank of Maharashtra | 1,22,119 | 1,38,990 | 1,39,053 | 1,38,981 | 1,40,650 | 1,50,066 | 1,74,006 | 2,02,294 | 2,34,083 | 2,70,747 |
| Canara Bank | 4,73,840 | 4,79,792 | 4,95,275 | 5,24,772 | 5,99,033 | 6,25,351 | 10,10,875 | 10,86,409 | 11,79,219 | 13,12,367 |
| Central Bank of India | 2,55,572 | 2,66,184 | 2,96,184 | 2,94,839 | 2,99,855 | 3,13,768 | 3,29,973 | 3,42,692 | 3,59,296 | 3,85,011 |
| Indian Bank | 1,69,225 | 1,78,286 | 1,82,509 | 2,08,294 | 2,42,076 | 2,60,226 | 5,38,071 | 5,93,618 | 6,21,166 | 6,88,000 |
| Indian Overseas Bank | 2,46,049 | 2,24,514 | 2,11,343 | 2,16,832 | 2,22,534 | 2,22,952 | 2,40,288 | 2,62,159 | 2,60,883 | 2,85,905 |
| Punjab & Sind Bank | 86,715 | 91,250 | 85,540 | 1,01,726 | 98,558 | 89,668 | 96,109 | 1,02,137 | 1,09,665 | 1,19,410 |
| Punjab National Bank | 5,01,379 | 5,53,051 | 6,28,170 | 6,42,226 | 6,96,030 | 7,03,848 | 11,06,382 | 11,46,218 | 12,81,163 | 13,69,713 |
| UCO Bank | 2,14,337 | 2,07,118 | 2,01,285 | 1,81,849 | 1,97,907 | 1,93,203 | 2,05,919 | 2,24,073 | 2,49,338 | 2,63,130 |
| Union Bank of India | 3,16,870 | 3,42,720 | 3,78,392 | 4,08,502 | 4,15,915 | 4,50,668 | 9,23,805 | 10,32,393 | 11,17,716 | 12,21,528 |
| State Bank of India (SBI) | 15,76,793 | 17,30,722 | 20,44,751 | 27,06,343 | 29,11,386 | 32,41,621 | 36,81,277 | 40,51,534 | 44,23,778 | 49,16,077 |

| Banks | OPERATING EXPENCES | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
| Bank of Baroda | 25,012 | 22,750 | 28,687 | 28,127 | 31,290 | 48,532 | 41,686 | 37,259 | 48,233 | 67,884 |
| Bank of India | 32,086 | 30,072 | 27,465 | 27,565 | 27,110 | 27,096 | 26,330 | 24,014 | 27,373 | 37,657 |
| Bank of Maharashtra | 8,790 | 9,174 | 8,887 | 7,707 | 7,116 | 6,971 | 7,217 | 6,975 | 8,158 | 10,673 |
| Canara Bank | 34,086 | 34,259 | 31,516 | 29,089 | 32,332 | 35,811 | 45,178 | 43,026 | 52,989 | 72,122 |
| Central Bank of India | 19,162 | 18,822 | 18,087 | 17,519 | 15,866 | 15,934 | 14,485 | 13,315 | 13,855 | 17,826 |
| Indian Bank | 11,392 | 11,798 | 10,894 | 10,850 | 12,167 | 23,440 | 13,799 | 22,128 | 24,717 | 32,341 |
| Indian Overseas Bank | 18,554 | 18,135 | 14,529 | 12,448 | 12,352 | 12,103 | 11,067 | 10,419 | 11,145 | 14,220 |
| Punjab & Sind Bank | 6,909 | 6,569 | 6,014 | 5,714 | 6,279 | 5,872 | 4,712 | 4,445 | 5,019 | 6,853 |
| Punjab National Bank | 29,760 | 32,113 | 32,253 | 33,073 | 34,154 | 36,362 | 17,273 | 46,185 | 50,652 | 66,819 |
| UCO Bank | 13,797 | 13,713 | 12,509 | 10,895 | 10,019 | 10,042 | 8,966 | 8,508 | 10,307 | 13,754 |
| Union Bank of India | 23,640 | 23,886 | 23,754 | 23,443 | 23,852 | 25,794 | 44,079 | 40,157 | 47,978 | 63,208 |
| State Bank of India (SBI) | 97,382 | 1,06,803 | 1,13,659 | 1,45,646 | 1,54,520 | 1,59,239 | 1,54,441 | 1,54,750 | 1,87,263 | 2,55,255 |

| BANKS | BORROWINGS | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
| Bank of Baroda | 35,264 | 33,472 | 30,611 | 62,572 | 67,201 | 93,069 | 66,848 | 1,03,899 | 1,01,910 | 94,402 |
| Bank of India | 40,057 | 51,083 | 39,406 | 43,589 | 44,241 | 39,752 | 32,464 | 26,760 | 64,979 | 80,924 |
| Bank of Maharashtra | 3,670 | 4,239 | 7,747 | 10,766 | 7,719 | 11,127 | 9,228 | 8,137 | 4,064 | 10,149 |
| Canara Bank | 25,672 | 26,873 | 39,504 | 38,809 | 40,992 | 42,762 | 49,984 | 46,285 | 58,090 | 57,592 |
| Central Bank of India | 25,974 | 9,208 | 9,282 | 5,706 | 5,239 | 5,787 | 5,469 | 7,474 | 8,119 | 19,806 |
| Indian Bank | 2,646 | 3,509 | 12,637 | 19,760 | 12,138 | 20,830 | 26,175 | 17,144 | 22,073 | 23,131 |
| Indian Overseas Bank | 3232. 41 08 | 7183.30 77 | 16,098 | 9,228 | 6,146 | 5,420 | 3,672 | 3,071 | 20,804 | 30,387 |
| Punjab & Sind Bank | 3,048 | 2,839 | 2,958 | 3,683 | 2,714 | 3,213 | 2,644 | 2,444 | 9,018 | 9,771 |
| Punjab National Bank | 39,326 | 50,225 | 50,225 | 42,840 | 42,840 | 45,681 | 51,292 | 51,292 | 50,430 | 50,430 |
| UCO Bank | 10,252 | 17,240 | 9,535 | 12,449 | 8,324 | 15,695 | 15,383 | 13,508 | 20,501 | 25,331 |
| Union Bank of India | 42,864 | 52,486 | 52,486 | 51,837 | 51,837 | 51,179 | 43,137 | 43,137 | 26,948 | 26,948 |
| State Bank of India (SBI) | 2,05,150 | 3,23,345 | 3,17,694 | 3,62,142 | 4,03,017 | 3,14,656 | 4,17,298 | 4,26,043 | 4,93,135 | 5,97,561 |

| BANKS | INVESTMENTS | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
| Bank of Baroda | 1,22,320 | 1,20,451 | 1,29,631 | 1,63,185 | 1,82,298 | 1,13,722 | 2,54,369 | 3,15,795 | 3,62,485 | 3,69,817 |
| Bank of India | 1,19,792 | 1,18,849 | 1,27,827 | 1,37,111 | 1,47,639 | 1,58,573 | 1,87,253 | 1,74,448 | 2,04,398 | 2,27,144 |
| Bank of Maharashtra | 36,715 | 36,231 | 38,590 | 43,623 | 59,697 | 57,741 | 68,112 | 68,590 | 68,867 | 68,274 |
| Canara Bank | 1,42,061 | 1,42,309 | 1,50,266 | 1,44,054 | 1,52,985 | 1,76,245 | 2,61,690 | 2,82,013 | 3,19,038 | 3,57,454 |
| Central Bank of India | 95,474 | 88,868 | 92,095 | 1,02,632 | 1,75,298 | 1,42,520 | 1,48,582 | 1,40,787 | 1,36,583 | 1,43,923 |
| Indian Bank | 45,899 | 53,089 | 67,552 | 71,398 | 64,992 | 81,242 | 1,76,537 | 1,74,559 | 1,85,988 | 2,12,554 |
| Indian Overseas Bank | 81,310 | 79,190 | 71,549 | 68,646 | 66,932 | 79,416 | 95,494 | 98,179 | 94,170 | 99,632 |
| Punjab & Sind Bank | 26,752 | 27,645 | 27,949 | 32,982 | 26,173 | 24,552 | 32,023 | 42,281 | 44,838 | 49,599 |
| Punjab National Bank | 1,51,282 | 1,59,846 | 1,86,725 | 2,00,306 | 2,02,128 | 2,40,466 | 39,983 | 3,72,168 | 3,95,997 | 4,20,318 |
| UCO Bank | 68,859 | 83,974 | 74,019 | 70,962 | 82,232 | 90,999 | 93,783 | 96,874 | 95,169 | 92,904 |
| Union Bank of India | 94,093 | 89,208 | 1,12,149 | 1,23,780 | 1,26,049 | 1,52,414 | 3,31,512 | 3,48,507 | 3,39,299 | 3,37,904 |
| State Bank of India (SBI) | 4,81,759 | 5,75,652 | 7,65,990 | 10,60,987 | 9,67,022 | 10,46,955 | 13,51,705 | 14,81,445 | 15,70,366 | 16,71,340 |

| Banks | TOTAL INCOME | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
| Bank of Baroda | 47,366 | 49,060 | 48,958 | 50,306 | 56,065 | 86,301 | 82,860 | 81,365 | 99,614 | 1,27,101 |
| Bank of India | 47,663 | 45,449 | 46,063 | 43,805 | 45,900 | 49,066 | 48,041 | 45,955 | 54,748 | 66,804 |
| Bank of Maharashtra | 13,671 | 14,072 | 13,570 | 12,602 | 12,397 | 13,145 | 14,494 | 15,672 | 18,179 | 23,493 |
| Canara Bank | 48,300 | 48,897 | 48,942 | 48,195 | 53,385 | 56,748 | 84,525 | 85,907 | 1,03,187 | 1,27,654 |
| Central Bank of India | 28,303 | 27,827 | 27,537 | 26,658 | 25,052 | 27,199 | 25,897 | 25,770 | 29,626 | 35,434 |
| Indian Bank | 17,216 | 18,025 | 18,251 | 19,519 | 21,068 | 24,717 | 45,185 | 45,772 | 52,085 | 63,482 |
| Indian Overseas Bank | 26,077 | 26,046 | 23,091 | 21,662 | 21,838 | 20,766 | 22,525 | 21,633 | 23,509 | 29,706 |
| Punjab & Sind Bank | 9,017 | 9,223 | 8,751 | 8,530 | 9,387 | 8,827 | 7,877 | 8,055 | 8,933 | 10,915 |
| Punjab National Bank | 52,206 | 54,301 | 56,227 | 56,877 | 58,688 | 63,074 | 93,562 | 87,199 | 97,287 | 1,20,285 |
| UCO Bank | 21,363 | 20,157 | 18,440 | 15,141 | 15,844 | 18,006 | 18,166 | 18,082 | 20,159 | 25,120 |
| Union Bank of India | 35,607 | 35,831 | 37,625 | 37,738 | 38,541 | 42,492 | 80,104 | 80,469 | 95,376 | 1,15,858 |
| State Bank of India (SBI) | 1,74,973 | 1,91,844 | 2,10,979 | 2,65,100 | 2,78,083 | 3,02,545 | 3,08,647 | 3,16,021 | 3,68,719 | 4,66,813 |

| Banks | INTEREST EARNED | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
| Bank of Baroda | 42,964 | 44,061 | 42,200 | 43,649 | 49,771 | 75,984 | 70,495 | 69,881 | 89,589 | 1,12,606 |
| Bank of India | 13,465 | 41,796 | 11,826 | 9,347 | 10,814 | 42,353 | 40,599 | 38,076 | 47,648 | 60,709 |
| Bank of Maharashtra | 9,821 | 9,935 | 8,470 | 7,073 | 6,567 | 6,409 | 7,154 | 13,019 | 15,898 | 20,495 |
| Canara Bank | 43,813 | 44,039 | 41,388 | 41,252 | 46,810 | 48,935 | 69,240 | 69,410 | 84,425 | 1,08,688 |
| Central Bank of India | 26,409 | 25,888 | 24,661 | 24,036 | 22,639 | 23,562 | 22,730 | 22,802 | 25,542 | 30,722 |
| Indian Bank | 12,074 | 11,924 | 11,461 | 11,857 | 13,984 | 15,933 | 27,455 | 38,856 | 44,942 | 55,615 |
| Indian Overseas Bank | 17,946 | 16,662 | 14,053 | 11,961 | 11,727 | 11,513 | 10,834 | 16,730 | 19,400 | 24,050 |
| Punjab & Sind Bank | 8,589 | 8,744 | 8,173 | 7,949 | 4,559 | 7,930 | 6,974 | 7,096 | 7,993 | 9,694 |
| Punjab National Bank | 46,315 | 47,424 | 47,276 | 47,996 | 51,310 | 53,808 | 80,749 | 74,880 | 85,144 | 1,06,902 |
| UCO Bank | 19,359 | 18,561 | 16,326 | 14,020 | 14,331 | 15,134 | 14,446 | 14,981 | 17,651 | 21,854 |
| Union Bank of India | 32,084 | 32,199 | 32,660 | 32,748 | 34,067 | 37,231 | 68,767 | 67,944 | 80,743 | 99,778 |
| State Bank of India (SBI) | 1,52,397 | 1,63,685 | 1,75,518 | 2,20,499 | 2,42,869 | 2,57,324 | 2,65,151 | 2,75,457 | 3,32,103 | 4,15,131 |

| Banks | ADVANCES | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|
| | 2015 | 2016 | 2017 | 2018 | 2019 | 2020 | 2021 | 2022 | 2023 | 2024 |
| Bank of Baroda | 3,94,612 | 3,80,382 | 3,83,259 | 4,27,432 | 4,68,819 | 6,90,121 | 4,69,330 | 7,77,155 | 9,40,998 | 10,65,782 |
| Bank of India | 4,02,026 | 3,59,189 | 3,66,482 | 4,00,176 | 4,09,154 | 4,16,521 | 4,10,436 | 4,20,842 | 4,85,900 | 5,63,145 |
| Bank of Maharashtra | 98,599 | 1,07,563 | 95,515 | 85,797 | 82,666 | 86,872 | 1,02,405 | 1,31,170 | 1,71,221 | 2,00,240 |
| Canara Bank | 6,94,945 | 7,19,810 | 3,42,009 | 3,81,703 | 4,27,727 | 4,32,175 | 6,39,049 | 7,03,602 | 8,30,673 | 9,31,613 |
| Central Bank of India | 1,88,478 | 1,80,010 | 1,39,397 | 1,56,542 | 1,46,525 | 1,51,101 | 1,54,579 | 1,68,174 | 2,02,984 | 2,43,406 |
| Indian Bank | 1,25,864 | 1,29,049 | 1,27,699 | 1,56,569 | 1,81,262 | 1,97,887 | 3,64,010 | 3,89,186 | 4,49,297 | 5,14,889 |
| Indian Overseas Bank | 1,71,756 | 1,60,861 | 1,40,459 | 1,32,489 | 1,32,598 | 1,21,333 | 1,27,721 | 1,44,244 | 1,78,053 | 2,13,319 |
| Punjab & Sind Bank | 63,870 | 63,916 | 58,335 | 66,569 | 69,176 | 58,412 | 60,902 | 63,627 | 76,819 | 82,736 |
| Punjab National Bank | 3,80,534 | 4,12,326 | 4,19,493 | 4,33,731 | 4,58,249 | 4,71,828 | 6,74,230 | 7,28,186 | 8,30,834 | 9,34,431 |
| UCO Bank | 1,47,351 | 1,25,905 | 1,19,724 | 1,07,470 | 99,314 | 1,01,174 | 1,11,355 | 1,22,784 | 1,55,870 | 1,82,022 |
| Union Bank of India | 2,55,655 | 2,67,354 | 2,86,467 | 2,88,761 | 2,96,932 | 3,15,049 | 5,90,983 | 6,61,005 | 7,61,845 | 8,70,776 |
| State Bank of India (SBI) | 13,00,026 | 14,63,700 | 15,71,078 | 19,34,880 | 21,85,877 | 23,25,290 | 24,49,498 | 27,33,967 | 31,99,269 | 37,03,971 |

**4.RESULTS AND DISCUSSION:** Python code is developed for the proposed steps and results obtained are presented below.

International Journal of Marketing and Technology
Vol. 16 Issue 01, January 2026
ISSN: 2249-1058 Impact Factor: 6.559
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

**4.1DEA Cross-Efficiency Matrix:** A cross-efficiency matrix CEkj was obtained, using the optimal weights of DMU k.This matrix captures both self-efficiency (diagonal values) and peer efficiency (off-diagonal values). The diagonal values were generally higher, as each DMU maximizes its own efficiency, while off-diagonal entries provided a more balanced peer perspective.

**Self Efficiency**

| Year | BoB | BoI | BoM | CB | CeB | IB | IoB | PSB | PNB | SBI | UCO | UBI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2015 | 1.0000 | 0.6341 | 1.0329 | 0.6320 | 0.3244 | 1.0000 | 1.0000 | 1.0231 | 0.6442 | 0.5928 | 1.0000 | 1.0000 |
| 2016 | 0.7361 | 0.3295 | 0.2850 | 0.1306 | 0.2805 | 0.3288 | 0.3275 | 0.3295 | 0.3150 | 0.2812 | 0.2740 | 0.1663 |
| 2017 | 0.7042 | 0.5310 | 0.5436 | 0.5310 | 0.5207 | 0.3747 | 0.4163 | 0.6322 | 0.5310 | 0.6408 | 0.5308 | 0.4840 |
| 2018 | 1.0000 | 0.5594 | 0.7021 | 0.5199 | 0.5635 | 0.5607 | 0.8878 | 0.8857 | 0.6198 | 0.3553 | 0.5198 | 0.5619 |
| 2019 | 0.5589 | 0.3687 | 0.2358 | 0.5320 | 0.2444 | 0.5167 | 0.5167 | 0.4229 | 0.5346 | 0.2444 | 0.2426 | 0.5161 |
| 2020 | 0.8376 | 0.4483 | 0.0552 | 0.1340 | 0.5006 | 0.5831 | 0.4551 | 0.1169 | 0.5095 | 0.0659 | 0.4213 | 0.5003 |
| 2021 | 0.5836 | 0.3497 | 0.4621 | 0.3979 | 0.3946 | 0.4598 | 0.4555 | 0.3936 | 0.4455 | 0.5890 | 0.4353 | 0.3703 |
| 2022 | 0.6710 | 0.6710 | 0.0919 | 0.4045 | 0.0905 | 0.5344 | 0.6592 | 0.3027 | 0.5469 | 0.1085 | 0.1076 | 0.5323 |
| 2023 | 0.6086 | 0.6085 | 0.5251 | 0.4161 | 0.5506 | 0.5545 | 0.3294 | 0.1013 | 0.5499 | 0.5133 | 0.6044 | 0.4522 |
| 2024 | 0.6377 | 0.5941 | 0.6118 | 0.6111 | 0.4577 | 0.1539 | 0.6109 | 0.5895 | 0.6377 | 0.1546 | 0.5669 | 0.5091 |

**Cross Efficiency**

| Year | BoB | BoI | BoM | CB | CeB | IB | IoB | PSB | PNB | SBI | UCO | UBI |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 2015 | 0.0351 | 0.5203 | 0.3766 | 0.5146 | 0.3054 | 0.3894 | 0.3504 | 0.3698 | 0.4731 | 0.5185 | 0.4836 | 0.5198 |
| 2016 | 0.3365 | 0.4527 | 0.5639 | 0.5381 | 0.5883 | 0.4545 | 0.4504 | 0.4529 | 0.4715 | 0.5476 | 0.5420 | 0.5378 |
| 2017 | 0.5012 | 0.5605 | 0.3768 | 0.5318 | 0.5220 | 0.3798 | 0.4542 | 0.4139 | 0.5318 | 0.4198 | 0.5603 | 0.5481 |
| 2018 | 0.3659 | 0.5116 | 0.3152 | 0.5620 | 0.5327 | 0.5084 | 0.3924 | 0.3764 | 0.3420 | 0.4610 | 0.5615 | 0.5342 |
| 2019 | 0.5570 | 0.5450 | 0.4656 | 0.5116 | 0.4638 | 0.5236 | 0.5236 | 0.4535 | 0.5139 | 0.4638 | 0.4655 | 0.4984 |
| 2020 | 0.2819 | 0.4949 | 0.3656 | 0.3960 | 0.5638 | 0.5421 | 0.5134 | 0.4780 | 0.5275 | 0.2745 | 0.5145 | 0.5636 |
| 2021 | 0.4233 | 0.5138 | 0.5442 | 0.5277 | 0.4667 | 0.5397 | 0.5283 | 0.3725 | 0.5698 | 0.4216 | 0.5019 | 0.6110 |
| 2022 | 0.4781 | 0.4781 | 0.4808 | 0.4564 | 0.4782 | 0.5600 | 0.5103 | 0.3512 | 0.5300 | 0.4234 | 0.3989 | 0.5595 |
| 2023 | 0.5169 | 0.5170 | 0.4864 | 0.4603 | 0.5095 | 0.4489 | 0.4544 | 0.2906 | 0.5336 | 0.5227 | 0.5676 | 0.5328 |
| 2024 | 0.5258 | 0.4866 | 0.5169 | 0.5166 | 0.4496 | 0.4712 | 0.5514 | 0.5156 | 0.5258 | 0.4753 | 0.5088 | 0.4391 |

Bank of India and SBI emerge as the role models of efficiency. Much like students who score consistently well whether in their own practice tests or in external examinations, these banks demonstrate stability and reliability. Their self-efficiency scores are closely aligned with peer evaluations, which means their performance is not exaggerated but validated across the sector. This builds trust among regulators, investors, and customers alike, making them benchmarks of steady, low-risk banking operations.

**Bank of Baroda and Bank of Maharashtra show that they have the strength and resources to perform at the top level, but their performance is uneven.** When they set their own criteria, they appear very efficient — which means they know how to use their advantages well. But when their results are compared against common standards across banks, their efficiency drops. In practice, this means they can deliver excellent outcomes in certain years or under favorable conditions, but they struggle to maintain the same efficiency consistently. For regulators and customers, this points to a need for more stability; for managers, it means focusing less on one-time peaks and more on building systems and strategies that can keep performance high year after year.

UCO Bank and Union Bank of India appear highly efficient when they use their own yardstick, but when judged by common benchmarks, they fall behind — meaning they may look stronger on paper than they actually are in practice.

**4.2 Consensus Efficiency Scores: The** average of each DMU's cross-efficiency across all peers was calculated to obtain a consensus efficiency score. Scores were normalized to ensure comparability across DMUs. These scores represent the relative performance priorities of financial institutions.

**4.3 FCM Clustering of Consensus Scores:** The consensus efficiency vector was subjected to Fuzzy C-Means (FCM) clustering into three performance groups: Low, Medium, and High. Each institution was assigned both a hard label (cluster with highest membership) and soft membership probabilities.The**cluster centroids** confirmed that the High-performance group had significantly higher consensus efficiency scores.

**4.4 Logistic Regression Classification:** Logistic Regression was trained using the original inputs (Deposits, Operating Expenses, Borrowings) and outputs (Investments, Total Income, Interest Earned, Advances) as predictors, with FCM cluster labels as the target.Resultsinclude Predicted probabilities for each class.A**confusion matrix** summarizing classification accuracy.

International Journal of Marketing and Technology
Vol. 16 Issue 01, January 2026
ISSN: 2249-1058 Impact Factor: 6.559
Journal Homepage: http://www.ijmra.us, Email: editorijmie@gmail.com
Double-Blind Peer Reviewed Refereed Open Access International Journal - Included in the International Serial Directories Indexed & Listed at: Ulrich's Periodicals Directory ©, U.S.A., Open J-Gate as well as in Cabell's Directories of Publishing Opportunities, U.S.A

**4.5 Random Forest Classification:** A Random Forest (RF) classifier was applied using the same predictors.Results include Improved classification accuracy compared to Logistic Regression. Feature importance rankings, showing which financial indicators (e.g., Advances, Total Income, Deposits) most strongly drove classification. A confusion matrix visualizing true vs. predicted group memberships.

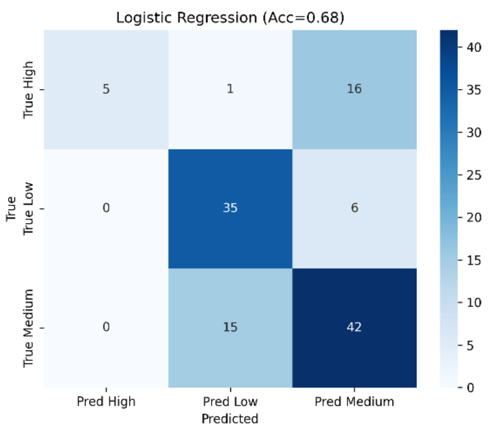
The analysis clearly shows that **Random Forest is more reliable than Logistic Regression** for diagnosing bank performance classifications, as it consistently aligns more closely with true labels. **High-performing banks such as SBI, UCO, and Punjab & Sind** are often underestimated by Logistic Regression, which tends to downgrade their classifications, whereas Random Forest provides more accurate recognition of their strong performance. On the other hand, **inconsistent banks like Bank of Baroda, Bank of Maharashtra, and Indian Overseas Bank** demonstrate occasional spikes in efficiency but struggle to sustain these levels over time, highlighting the need for greater consistency. Meanwhile, **mid-tier banks such as Canara, Central Bank, Union Bank, Indian Bank, and Punjab National Bank** show stable performance with predictable behavior, making them relatively safer bets in performance modeling and strategic planning.

**4.6 Comparative Evaluation:** in case of Accuracy, RF > Logistic Regression. Discrimination power of RF better separated Medium vs. High institutions, while Logistic Regression sometimes misclassified borderline cases. The hybrid approach thus balances efficiency measurement (DEA), grouping (FCM), and prediction (ML).

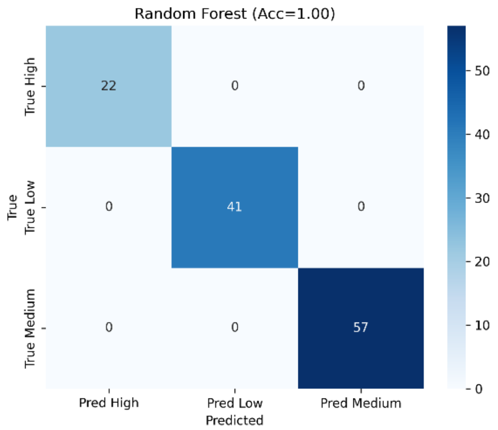
| Metric | Class (0 = High) | Class (1 = Low) | Class (2 = Medium) | Accuracy | Macro Avg | Weighted Avg |
|---|---|---|---|---|---|---|
| **Precision (Logit)** | 1 | 0.686 | 0.656 | 0.683 | 0.781 | 0.73 |
| **Precision (RF)** | 1 | 1 | 1 | 1 | 1 | 1 |
| **Recall (Logit)** | 0.227 | 0.854 | 0.737 | 0.683 | 0.606 | 0.683 |
| **Recall (RF)** | 1 | 1 | 1 | 1 | 1 | 1 |
| **F1-score (Logit)** | 0.37 | 0.761 | 0.694 | 0.683 | 0.608 | 0.658 |
| **F1-score (RF)** | 1 | 1 | 1 | 1 | 1 | 1 |
| **Support** | 22 | 41 | 57 | 120 | 120 | 120 |

The results show that **Random Forest (RF) achieves perfect performance**, with precision, recall, and F1-scores all equal to 1.0 across every class, indicating strong predictive power and flawless classification. In contrast, **Logistic Regression (Logit) performs moderately**, reaching an overall accuracy of **68.3%**, but showing clear weaknesses in certain areas. Specifically, Logit struggles with **class 0 (High performers)**, where recall drops sharply to **0.227**, meaning it fails to correctly identify most of the high-performing cases. This limitation highlights its tendency to underestimate strong performers. Overall, **RF is far superior**, as it accurately identifies all categories without any misclassification, making it a much more reliable model for bank performance classification.

**4.6.1 Confusion matrix-Logistic**



**4.6.2Confusion matrix-Random Forest**




**4.7 Discussion of Results:**

The cross-efficiency matrix highlighted notable variations across institutions, with some achieving consistently high peer evaluations while others showed efficiency dependent on their own weight choices. This finding is consistent with prior evidence that DEA cross-efficiency provides a more discriminating view than self-efficiency scores alone (Doyle & Green, 1994; Yang, 2025). Institutions with strong consensus efficiency are likely to have well-balanced financial structures, while lower consensus performers may be constrained by inefficiencies in input utilization or weaker output generation.The application of FCM enabled grouping of institutions into Low, Medium, and High performance categories, while preserving soft membership. This is especially relevant for financial institutions near the performance boundary, as they may exhibit features of more than one group. For example, some banks showed partial membership in both Medium and High clusters, suggesting transitional performance status. This aligns with Wu et al. (2025), who noted that soft clustering better captures uncertainty and overlap in financial performance data than crisp partitions.

Logistic Regression achieved moderate predictive accuracy, confirming its usefulness as a transparent and interpretable baseline model. The sign and magnitude of coefficients suggested that Advances, Total Income, and Deposits were the most influential variables in predicting performance groups. These findings echo earlier studies in banking where profitability and lending activity were found to be central determinants of efficiency (Berger & Humphrey, 1997; Shi et al., 2025). However, the linearity of Logistic Regression limited its ability to capture more complex relationships, leading to misclassification between Medium and High performers.The Random Forest model outperformed Logistic Regression in predictive accuracy, particularly in discriminating between Medium and High performance groups. Feature importance analysis indicated that Advances and Total Income were the strongest drivers, followed by Deposits and Investments. This confirms prior findings that ensemble learning methods capture nonlinear dependencies and interactions more effectively than linear models (Reiff et al., 2025; Sundaravadivel et al., 2025). Importantly, Random Forest provides not only higher accuracy but also robustness against noise, which is crucial for financial datasets that are often volatile.

Together, the results demonstrate that the hybrid DEA–FCM–ML approach offers a richer analytical perspective than standalone DEA or ML methods. DEA provides a peer-based benchmarking tool, FCM adds interpretability through fuzzy categorization, and ML classifiers enable prediction for new institutions. For regulators and policymakers, this implies that institutions flagged as Low performers can be identified with greater confidence, while borderline cases can be closely monitored for early intervention. For financial managers, the feature importance results highlight key levers—particularly advances and income generation—that can be targeted for performance improvement.

**5. CONCLUSIONS**

This study proposed and implemented a hybrid DEA–Fuzzy C-Means–Machine Learning framework to evaluate and predict the performance of financial institutions. By employing DEA cross-efficiency, the analysis moved beyond self-evaluation to provide peer-based benchmarking and derive a consensus priority score for each institution. The integration of FCM allowed institutions to be grouped into Low, Medium, and High performance categories, with fuzzy memberships capturing the uncertainty inherent in financial performance boundaries. Logistic Regression offered interpretability in identifying performance drivers, while Random Forest achieved superior predictive accuracy, highlighting the importance of advances, deposits, and total income as key determinants of institutional efficiency.The results underscore the value of combining traditional efficiency analysis with modern ML tools. The proposed framework provides regulators, policymakers, and financial managers with a practical decision-support system for benchmarking, monitoring, and forecasting institutional performance**.**Integrate the model with scenario-based simulations to evaluate the impact of regulatory changes (e.g., Basel III norms, monetary policy shifts) on institutional efficiency and classification outcomes.The Basel III framework, introduced by the Basel Committee on Banking Supervision (BCBS) in response to the 2008 global financial crisis, establishes stringent regulations to enhance the resilience of banks and reduce systemic risks.

**REFERENCES**

1. Berger, A. N., & Humphrey, D. B. (1997). Efficiency of financial institutions: International survey and directions for future research. *European Journal of Operational Research, 98*(2), 175–212.
2. Bezdek, J. C. (1981). *Pattern recognition with fuzzy objective function algorithms*. Springer Science & Business Media.
3. Breiman, L. (2001). Random forests. *Machine Learning, 45*(1), 5–32.
4. Cardone, B., et al. (2025). A fuzzy-entropy based online fuzzy c-means algorithm for evolving data streams. *Journal of Ambient Intelligence and Humanized Computing.* https://doi.org/10.1007/s12652-025-07435-1
5. Charnes, A., Cooper, W. W., & Rhodes, E. (1978). Measuring the efficiency of decision making units. *European Journal of Operational Research, 2*(6), 429–444.
6. Dong, H., He, Y., & Zhang, Y. (2025). A machine learning-enhanced DEA (M-LASSO) model for performance measurement. *Applied Energy.* https://doi.org/10.1016/j.apenergy.2025.122345
7. Kolahdoozi, L., Khalili-Damghani, K., &Aryanezhad, M. B. (2024). Advancing group efficiency evaluation in DEA with negative data: Application to banking. *RAIRO – Operations Research, 58*(6), 2475–2495.
8. Ohlson, J. A. (1980). Financial ratios and the probabilistic prediction of bankruptcy. *Journal of Accounting Research, 18*(1), 109–131.
9. Perroni, M. G., Mariani, M., & D'Ascenzo, F. (2024). Integrating relative efficiency models with machine learning: A hybrid framework. *SAGE Open, 14*(3), 21582440241234567.
10. Reiff, A., et al. (2025). Predicting IPO withdrawals: A comparison of random forest and logistic regression. *Intelligent Systems in Accounting, Finance and Management.* https://doi.org/10.1002/isaf.1234
11. Seiford, L. M., & Thrall, R. M. (1990). Recent developments in DEA: The mathematical programming approach to frontier analysis. *Journal of Econometrics, 46*(1–2), 7–38.
12. Sundaravadivel, P., et al. (2025). Credit card fraud detection using random forests: A comparative analysis. *Scientific Reports, 15*(1), 22345. https://doi.org/10.1038/s41598-025-22345-7
13. Wu, C., Chen, J., & Zhang, H. (2025). Semi-supervised fuzzy c-means clustering with improved stability for financial applications. *Expert Systems with Applications, 239,* 122317. https://doi.org/10.1016/j.eswa.2025.122317
14. Yang, M. (2025). A novel centralized cross-efficiency evaluation model in DEA: Application to cooperative banking. *European Journal of Operational Research, 314*(2), 625–639.
15. Zhang, Z., Wang, Y., & Li, X. (2022). Integrating data envelopment analysis and machine learning: A review and applications. *Mathematics, 10*(12), 2114. https://doi.org/10.3390/math10122114