



# International Journal of Advanced Research in Education and Technology (IJARETY)

Volume 11, Issue 6, November-December 2024

Impact Factor: 7.394



INTERNATIONAL  
STANDARD  
SERIAL  
NUMBER  
INDIA



# Fraud Detection in Internet Banking Using Machine Learning

Mr. E. Raju<sup>1</sup>, T. Anvesh<sup>2</sup>, T. Bhavana<sup>3</sup>, T. Siddesh<sup>4</sup>

Assistant Professor, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India<sup>1</sup>

Student, Department of CSE, Guru Nanak Institute of Technology, Hyderabad, Telangana, India<sup>2,3,4</sup>

**ABSTRACT:** Financial transactions or bank accounts that involve fraudulent or unlawful activity are referred to as banking fraud transactions. To identify these fraudulent actions, a variety of machine learning methods might be used. This study looks at a number of methods that can be used to categorize transactions as either valid or fraudulent. The Banking Fraud Transactions dataset, which is frequently characterized by substantial imbalance, is used in the study. We are using a number of machine learning methods, including Random Forest, K-Nearest Neighbor, and Decision Tree, to address this problem. The dataset is separated into training and test sets, and feature selection approaches are also used. KNN and Random Forest are two of the algorithms that were assessed in the study. The results show that every algorithm detects financial fraud transactions with high accuracy. The suggested model has potential for identifying further anomalies.

## I. INTRODUCTION

In recent years, the financial sector has seen a significant rise in fraudulent activities such as credit card fraud, identity theft, and money laundering, leading to considerable financial losses and a loss of trust in banking systems. To combat these threats, there is an increasing need for advanced technological solutions capable of detecting and preventing fraud in real-time. Machine learning, with its ability to analyze large datasets and uncover complex patterns, has become a key tool in identifying fraudulent transactions.

This project explores use of various machine learning algorithms, including Random Forest, K-Nearest Neighbors (KNN), and Logistic Regression, to develop a robust fraud detection model. These algorithms, each with unique strengths and trade-offs, are well-suited to address different aspects of the fraud detection task. Additionally, feature selection techniques are used to identify the most relevant transaction attributes, improving the model's accuracy and ability to generalize.

A key challenge in fraud detection is the class imbalance in datasets, where fraudulent transactions are much less common than legitimate ones. To overcome this, the project carefully partitions data into training and test sets and uses performance metrics such as precision, recall, F1-score, and AUC-ROC to evaluate the models. By developing and assessing these methodologies, the project aims to contribute to more effective fraud detection systems that can better protect financial institutions and customers from fraudulent activities.

## II. LITERATURE SURVEY

**M. Jullum et. al(2020)** The purpose of this paper is to develop, describe and validate a machine learning model for prioritising which financial transactions should be manually investigated for potential money laundering. The model is applied to a large data set from Norway's largest bank, DNB. Design/methodology/approach A supervised machine learning model is trained by using three types of historic data: "normal" legal transactions; those flagged as suspicious by the bank's internal alert system; and potential money laundering cases reported to the authorities. The model is trained to predict the probability that a new transaction should be reported, using information such as background information about the sender/receiver, their earlier behaviour and their transaction history. Findings The paper demonstrates that the common approach of not using non-reported alerts (i.e. transactions that are investigated but not reported) in the training of the model can lead to sub-optimal results. The same applies to the use of normal (uninvestigated) transactions. Our developed method outperforms the bank's current approach in terms of a fair measure of performance. Originality/value This research study is one of very few published anti-money laundering (AML) models for suspicious transactions that have been applied to a realistically sized data set. The paper also

presents a new performance measure specifically tailored to compare the proposed method to the bank's existing AML system.

**L. Keyan and Y. Tingting(2019)** The selection of the parameters of SVM model will affect the identification effect of suspicious financial transactions, this paper proposes the cross-validation method to find the optimal SVM classifier parameters to solve this problem. Cross validation method finds the optimal parameters based on the highest classification accuracy rate through grid search, it can effectively avoid the state of over-learning and less learning and greatly improves the overall performance of the classifier.

**R. Liu et. al(2020)** This paper presents a core decision tree algorithm identify money laundering activities. The clustering algorithm is the combination of BIRCH and K-means. In this method, decision tree of data mining technology is applied to anti-money-laundering filed after research of money laundering features. We select an appropriate identifying strategy to discover typical money laundering patterns and money laundering rules. Consequently, with the core decision tree algorithm, we can identify abnormal transaction data more effectively.

**Z. Gao(2019)** Financial institutions' capability in recognizing suspicious money laundering transactional behavioral patterns (SMLTBPs) is critical to antimoney laundering. Combining distance-based unsupervised clustering and local outlier detection, this paper designs a new cluster based local outlier factor (CBLOF) algorithm to identify SMLTBPs and use authentic and synthetic data experimentally to test its applicability and effectiveness...

**F. Anowar and S. Sadaoui(2020)** To tackle scalability challenges of fraud detection, we propose incremental classification approach using a neural network (MLP) and a memory model to address the stability-plasticity dilemma. This method adapts the fraud detection model with incoming data chunks, retaining past data to ensure continuity. Using a large credit card fraud dataset, we apply data sampling to address data skew and evaluate the model's performance after each incremental phase. The results show that our approach outperforms the non-incremental MLP in both efficiency and effectiveness.

### III. EXISTING SYSTEM

In this research, a fundamental decision tree algorithm for detecting money laundering operations is presented. The clustering algorithm is a blend of means and BIRCH. This approach applies data mining decision trees to the anti-money laundering domain following an investigation of money laundering characteristics. To find common money laundering patterns and money laundering regulations, we choose a suitable identifying technique. As a result, we can more successfully detect anomalous transaction data using the core decision tree technique.

#### DISADVANTAGES

- **Nature's instability:** The fact that decision trees are generally more unstable than other decision predictors is one of their drawbacks.
- **Less successful** in forecasting how a continuous variable will turn out.

### IV. PROPOSED SYSTEM

The proposed method leverages machine learning techniques to enhance the detection of fraudulent activity in banking transactions. It begins by collecting transaction data, which includes key information such as amounts, timestamps, merchant details, and customer information. This pre-processed data is then split into training and testing sets to train machine learning algorithms like Logistic Regression, Random Forest, and Decision Tree, allowing them to identify patterns indicative of fraudulent behaviour.

#### ADVANTAGES

- **Improved Accuracy:** The system enhances its ability to correctly identify fraudulent transactions by leveraging advanced machine learning algorithms and techniques, reducing false positives and negatives.
- **Real-time Detection:** By processing transaction data in real-time, the system can immediately flag suspicious activities as they occur, enabling faster intervention and minimizing potential fraud impact.
- **Adaptability to Evolving Threats:** The system continuously learns from new data and adjusts its detection models, allowing it to keep up with changing fraud patterns and emerging threats in the banking sector.

- **Enhanced Customer Trust:** With reliable and effective fraud detection, customers can feel more secure in their transactions, knowing that their financial information is protected, which strengthens their confidence in the banking system.

## V. SYSTEM ARCHITECTURE

The system architecture for analysing money laundering data begins with the collection of a comprehensive dataset that contains relevant transaction and behavioural information. This dataset is then subjected to a series of pre-processing steps to clean and transform the raw data, ensuring it is in a suitable format for analysis. Feature selection techniques are applied to identify the most significant attributes that contribute to detecting money laundering activities, which helps reduce dimensionality and improve the efficiency of the models.

Once the data has been pre-processed and key features selected, it is fed into machine learning models, including Random Forest (RF) classifiers and a neural network model. These models are trained to recognize patterns indicative of money laundering, and they generate predictions about the likelihood of each transaction being fraudulent. The system then evaluates the performance of the models using various metrics such as accuracy, precision, recall, and F1-score, providing insight into their effectiveness.

To further understand the models' behaviour, the results are visualized and analysed through graphical representations, such as confusion matrices, ROC curves, and other performance plots. This enables a deeper evaluation of how well the system detects money laundering, helping to fine-tune the models and improve their predictive capabilities over time.

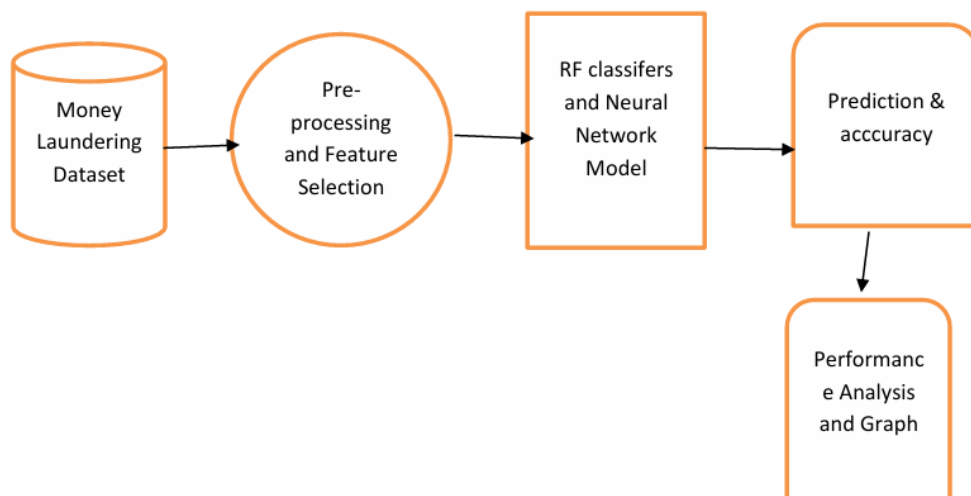


Fig.: SYSTEM ARCHITECTURE

## VI. METHODOLOGY

This project focuses on developing a robust machine learning system for detecting banking fraud using the Banking Fraud Transactions dataset, which is known for its significant class imbalance. To tackle this issue, the system incorporates multiple machine learning algorithms, such as Random Forest, K-Nearest Neighbors (KNN), and Logistic Regression. These algorithms are selected for their ability to handle different aspects of fraud detection, with the goal of improving the accuracy and reliability of fraud classification.

In order to enhance model performance, feature selection techniques are applied to identify the most relevant attributes in the dataset, which helps reduce dimensionality and increase the efficiency of the algorithms. The dataset is carefully divided into training and testing sets to allow for a fair evaluation of the models. The effectiveness of each algorithm is

measured by its ability to accurately classify transactions as fraudulent or legitimate, using a variety of performance metrics, including precision, recall, F1-score, and AUC-ROC, alongside overall accuracy.

Through rigorous testing and analysis, the project demonstrates that the proposed machine learning system can effectively identify fraudulent transactions, even in the face of class imbalance. The findings highlight the potential of this approach to not only improve fraud detection in banking but also to be applied in identifying irregularities across other financial transactions, contributing to more secure and reliable financial systems.

#### MODULES:

##### 1. Data Collection:

The first step in model development is collecting data, which is crucial for the model's performance. Better and more accurate data leads to better results. Data can be collected through various methods, such as web scraping or manual intervention.

##### 2. Dataset:

The dataset contains 1,048,576 rows and 11 columns, including attributes such as step, type, amount, nameOrig, oldbalanceOrg, newbalanceOrig, nameDest, oldbalanceDest, newbalanceDest, isFraud, and isFlaggedFraud.

##### 3. Data Preparation:

Data wrangling is performed to clean and prepare the data for training, including removing duplicates, fixing errors, handling missing values, and normalizing or converting data types. The data is randomized to eliminate biases from the order in which it was collected. It is then visualized to identify relationships or class imbalances, followed by splitting into training and evaluation sets.

##### 4. Model Selection:

A Neural Network was chosen to create the money laundering detection model, achieving an accuracy of 98.04% on the test set, leading to its implementation.

##### 5. Analysis and Prediction:

The model focuses on two main features: Amount (transaction details) and isFraud (indicating whether a transaction is fraudulent). These features are used to analyze and predict fraudulent transactions.

## VII. IMPLEMENTATION

#### ALGORITHM: -

##### 1. Support Vector Machine (SVM)

###### Algorithm:

- **Input:** A labelled dataset with features and corresponding class labels.
- **Objective:** Find the optimal hyperplane that separates different classes in the feature space.
- **Steps:**
  - Identify support vectors, which are the closest data points from each class.
  - Maximize the margin (distance between the hyperplane and the support vectors) to achieve better separation between classes.
  - If the data is linearly separable, find a linear hyperplane. Otherwise, apply a kernel function to map the data into a higher-dimensional space, making the classes more separable.

##### 2. Random Forest

###### Algorithm:

- **Input:** A labelled dataset with features and corresponding class labels.
- **Objective:** Build an ensemble of decision trees and aggregate their predictions for classification.
- **Steps:**
  - Randomly select subsets of the data and features to build multiple decision trees.
  - Each tree learns to classify instances based on the subset of features it was trained on.
  - After training, classify new data by aggregating the predictions from all the trees (majority voting for classification).

VIII. EXPERIMENTAL OUTCOMES

1. HOME PAGE:

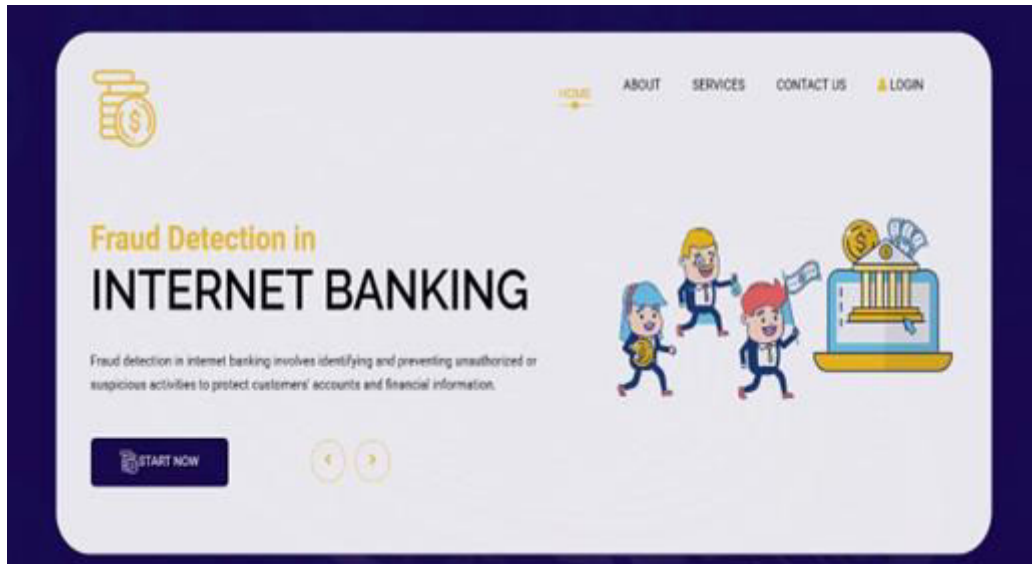


Fig.: Home Page

A website's home page is its primary landing page and the starting point for users to explore its features and content. It usually has a kind greeting and a simple, eye-catching layout that captures the essence of the business.

2. LOGIN PAGE:

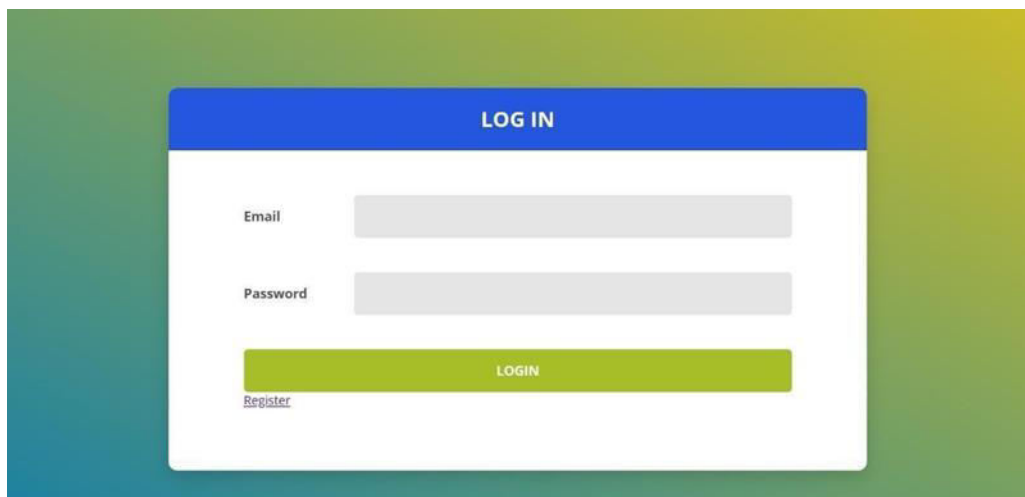
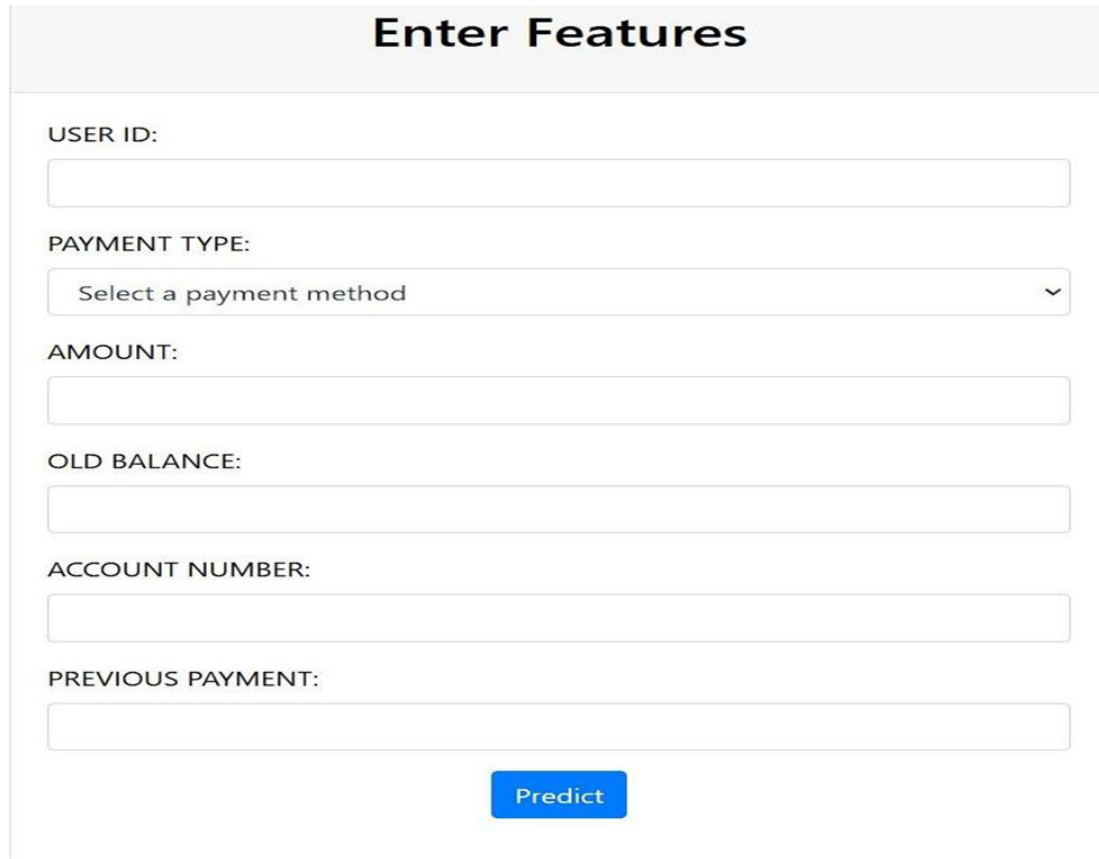


Fig.: Login Page

A user interface called the login page enables users to safely access their accounts. Input areas for entering a username and password are usually included, along with a button to submit the login information. Options for recovering a forgotten password or setting up a new account might also be available on the page.

3. TEST RESULT PAGE:



The screenshot shows a web form titled "Enter Features". It contains the following fields from top to bottom: "USER ID:" with a text input box; "PAYMENT TYPE:" with a dropdown menu showing "Select a payment method"; "AMOUNT:" with a text input box; "OLD BALANCE:" with a text input box; "ACCOUNT NUMBER:" with a text input box; and "PREVIOUS PAYMENT:" with a text input box. At the bottom center of the form is a blue button labeled "Predict".

Fig.: Test Result Page

Perhaps connected to a research or assessment of frauds, this context seems to be a structured data entry or submission for banking information.

### IX. CONCLUSION

This study concludes by showing how well machine learning algorithms, such as Random Forest, K-Nearest Neighbours (KNN), and Logistic Regression, can identify financial fraud transactions. We have created strong models that can reliably identify transactions as either valid or fraudulent by tackling the problem of class imbalance with meticulous algorithm selection, feature enhancement strategies, and thorough assessment metrics. In light of the fact that every algorithm has unique strengths, the results underscore the need of utilizing a variety of methodologies.

### X. FUTURE ENHANCEMENT

In machine learning, feature augmentation refers to raising the calibre and applicability of the features that are used to train models, which eventually improves generalization and predictive performance. Feature transformation, feature engineering, and feature selection are some of the methods for improving features. Finding the most informative subset of features from the original feature space is the goal of feature selection, which lowers computational cost and dimensionality while maintaining or even increasing model accuracy.

REFERENCES

1. M. Jullum, A. Løland, R. B. Huseby, G. A° nonsen, and J. Lorentzen, “Detecting money laundering transactions with machine learning,” *Journal of Money Laundering Control*, vol. 23, no. 1, pp. 173–186, jan 2020.
2. L. Keyan and Y. Tingting, “An improved support-vector network model for anti-money laundering,” in 2011 Fifth International Conference on Management of e-Commerce and e Government. IEEE, 2011, pp. 193– 196.
3. R. Liu, X.-l. Qian, S. Mao, and S.-z. Zhu, “Research on anti-money laundering based on core decision tree algorithm,” in 2011 Chinese Control and Decision Conference (CCDC). IEEE, 2011, pp. 4322– 4325.
4. Z. Gao, “Application of cluster-based local outlier factor algorithm in anti-money laundering,” in 2009 International Conference on Management and Service Science. IEEE, 2009, pp. 1–4.
5. J. de Jes’us Rocha Salazar, M. Jes’us Segovia-Vargas, and M. del Mar Camacho-Mi~nano, “Money laundering and terrorism financing detection using neural networks and an abnormality indicator,” *Expert Systems with Applications*, p. 114470, dec 2020. [Online]. Available: <https://linkinghub.elsevier.com/retrieve/pii/S0957417420311209>
6. E. L. Paula, M. Ladeira, R. N. Carvalho, and T. Marzagao, “Deep learning anomaly detection as support fraud investigation in Brazilian exports and anti-money laundering,” in 2016 15th IEEE International Conference on Machine Learning and Applications (ICMLA). IEEE, 2016, pp. 954–960.
7. F. Anowar and S. Sadaoui, “Incremental Neural-Network Learning for Big Fraud Data,” in *IEEE Transactions on Systems, Man, and Cybernetics: Systems*, vol. 2020-Octob. Institute of Electrical and Electronics Engineers Inc., oct 2020, pp. 3551–3557.
8. G. A. Carpenter and S. Grossberg, “A massively parallel architecture for a self-organizing neural pattern recognition machine,” *Computer vision, graphics, and image processing*, vol. 37, no. 1, pp. 54–115, 1987.
9. G. Carpenter, “An adaptive resonance algorithm for rapid category learning and recognition,” *Neural Networks*, vol. 4, pp. 439–505, 1991.
10. T. Kohonen, “Self-organized formation of topologically correct feature maps,” *Biological cybernetics*, vol. 43, no. 1, pp. 59–69, 1982.
11. A. Ultsch, “Kohonen’s self-organizing feature maps for exploratory data analysis,” *Proc. INNC90*, pp. 305–308, 1990.
12. Senator T E, Goldberg H G, Wooton J, etal. The Financial Crimes Enforcement Network AI System (FAIS)-identifying Potential Money Laundering from Reports of Large Cash Transactions [ J]. *AI Magazine*, 1995, pp. 21-39.
13. Zdanowicz John S. Detecting Money Laundering and Terrorist Financing Via Data Mining [ J]. *Communications of the ACM*, 2004, pp.53-55.
14. Bolton R J, Hand D J. Statistical Fraud Detection [ J]. *Statistical Science*, 2002, pp. 235 254.
15. Zhang Yan, Ouyang Yiming, Wang Hao, Wang Xidong, Application of Data Mining in the Financial Field *Computer Engineering and Applications*, vol.18, pp.208-211, 2004
16. Yang Sheng-gang, Wang Peng, He Xue-hui, Exploring Decision Trees as a Tool to Investigate Money Laundering. *Journal of Hunan University Social Sciences*, vol.20, No.1, Jau. 2006, pp.65-71.
17. Tian Zhang □Raghu Ramakrishman □Miron Livny. BIRCH: An Efficient Data Clustering Method for Very Large Databases. In: H. V. Jagadish□Inderpal Sinhg Mumick eds. *Proceedings for the 1996 ACM SIGMOD International Conference on Management of Data (SIGMOD96)*. Monteral □ Canada. 1996. NewYork: ACMPerss, pp.103□114,1996.
18. Eui-Hong Han. Text Categorization Using Weight Adjusted k-Nearest Neighbor Classification. PhD thesis□University of Minnesota□1999.





## International Journal of Advanced Research in Education and Technology

ISSN: 2394-2975

Impact Factor: 7.394