IJCRT.ORG ISSN : 2320-2882

INTERNATIONAL JOURNAL OF CREATIVE RESEARCH THOUGHTS (IJCRT)

An International Open Access, Peer-reviewed, Refereed Journal

Fraud Detection In Banking Data By Machine Learning Techniques

M. Sasi Kumar
Assistant Professor Sr. Grade
Department of Computer Science and Design
Vel Tech Rangarajan Dr.Sangunthala R&D Institute of Science and Technology Chennai,India

R V Chaitanya
Department of Computer Science and Design
Vel Tech Rangarajan Dr.Sangunthala R&D Institute of Science and Technology Chennai,India

K.Madhu Sudhan Reddy
Department of Computer Science and Design
Vel Tech Rangarajan Dr.Sangunthala R&D Institute of Science and Technology Chennai,India

R Siva Jyothish Kumar Reddy
Department Computer Science and Design
Vel Tech Rangarajan Dr.Sangunthala R&D Institute of Science and Technology Chennai,India

Abstract: The study mostly focuses on the use of machine learning techniques to find fraudulent behavior in financial facts. This is the main challenge in the financial sector, where it is important to recognize and prevent fraud. Images are hyperparameters of tuning class as a method for increasing fraud detection. These settings improve fraud detection system by helping the version of the extra precisely distinguish between real and fraudulent transactions. The work deliberately uses three machine learning techniques: XGBoost, LightGBM, and CatBoost. Every method has sure advantages; their combined use seeks to improve the general fraud detection method performance. The research includes deep learning algorithms to adapt to hyperparameters. This connection improves the effectiveness and adaptability of fraud detection systems, and increases the efficiency of identifying modified fraud strategies. The effort employs actual data to conduct comprehensive analyses. The findings indicate that Lightgbm and XGBOOST outperformed the contemporary method when assessing numerous factors. This suggests that, among other strategies, the suggested one is more a success in spotting fraudulent behaviour. It incorporates a Stacking Classifier, which combines with precise parameters Random forest and LightGBM classifier predictions. Through using the strengths of numerous models, this ensemble technique improves prediction accuracy by means of a GradientBoostingClassifier as the final estimator.

"Index terms *- Bayesian optimization, data mining, deep learning, ensemble learning, hyper parameter, unbalanced data, machine learning".*

1. INTRODUCTION

Due to the expansion of financial institutions and the rise of "online e-commerce", the volume of financial transactions has recently surged. Online banking has increasingly encountered issues with fraudulent transactions; detecting fraud has consistently posed challenges. The evolution of credit cards has consistently influenced the "dynamics of credit card fraud. Credit card fraud" is continually evolving; scams aim to prioritise it as the primary concern. Individuals that perpetrate frauds attempt to masquerade as authentic. Their continual stimulation of fraud detection systems aids in identifying fraud when they seek to comprehend the functionality of these systems. Consequently, scholars are always looking for clean approaches or ways to enhance the performance of the current ones [3].

Typically, fraudsters exploit their objectives. Nonetheless, technology can serve as an instrument for preventing fraud. Finding the fraud early on after its occurrence will help to stop more possible fraud [5]. One definition of fraud is unlawful or mistaken dishonesty meant for either financial or private gain. Credit card fraud involves the illicit use of credit card information for tangible or virtual acquisitions. As cardholders frequently provide brief details such as expiration dates and card verification numbers via phone or online platforms, the fraud may occur during digital transactions over the line or network.

Two technologies designed to detect fruit fraud prevention can be employed to mitigate the damage connected with fraudulent activities. Proactively preventing frauds from inception is an effective strategy for their total eradication. Conversely, whereas a scam attempts a fraudulent transaction, it is imperative to identify fraud. In banking, fraud detection is regarded as a binary classification problem, resulting in data being categorised as either genuine or fraudulent [8]. Manually evaluating and spotting trends in scam transactions is either impractical or time-consuming due to the extensive quantity of economic data and the vast array of transaction datasets involved. Consequently, fraud detection and predictive machine learning algorithms are highly reliant on. Enhanced processing capability and machine learning methodologies facilitate the efficient management of substantial data volumes and the identification of fraudulent activities. 15 methods for deep learning and machine learning provide immediate and efficient solutions for real-world situations [10]. This letter suggests ways to recognize credit card fraud. This is evaluated on public data records where both "custom algorithms LightGBM, XGBoost, CatBoost, and logistic regression" are used individually and through majority selection including deep learning. An optimal fraud detection system must accurately identify instances of over-fraud; high precision in detecting fraudulent cases is essential to ensure that all effects are recognised, thereby fostering customer confidence in the bank. Conversely, the bank must avoid losses resulting from erroneous detections.

2. LITERATURE SURVEY

The primary challenge in mitigating e-commerce fraud lies in the dynamic and diverse nature of fraudulent practices. Two novel technological fraud detection methods: link analysis and multilayer machine learning models. This paper presents a solution to successfully detect several forms of fraud. Link analysis helps create fraud islands by looking at the connections between several bogus entities and revealing the latent intricate fraud styles across the created network. The pretty varied character of fraud styles is addressed via a multi-layer model. These days, banks' fraud alert, customer chargeback requests, manual review agent rejection judgements, and declination choices of banks define Label deception via many routes. Considering the implementation of specialised fraud risk mitigation mechanisms—such as banking protocols, review teams, and fraudulent learning models—one might anticipate significantly different fraudulent patterns. Research indicated that the precision of fraud-related choices can be significantly enhanced by incorporating multiple diverse machine learning models trained on various scams [10].

The incidence of fraudulent billing escalates concurrently within government programs and privately funded health initiatives, exhibiting exponential growth. The intricate correlation among dynamic factors—such as physicians,

patients, and services—constituting fraudulent transactions inside the healthcare system is a significant endeavour. Consequently, sophisticated fraud detection algorithms are essential to initiate openness in healthcare programs, enabling the identification of deficiencies in existing processes to accurately uncover instances of medical invoice fraud. Furthermore, it is necessary to maximise the medical benefits for the client and the value load for the service provider. [2] This work introduces a new system-based fraud detection method based on series mining ideas to identify frauds connected to insurance claims in the healthcare sector. Recent studies emphasise quantity-based analysis or therapeutic efficacy against illness sequences rather than detecting fraud through the sequence generation of each feature's inner. The proposed technique generates consistent sequences across various pattern lengths. For each sequence, a trust level and values are computed. For every hospital's speciality, Sequence rules generate normal sequences and confidence values that are as compared with the real patient values [2, 7, 9]. This points out abnormalities since neither sequence would observe the rules engine's guidelines. Using transactional data from a local hospital spanning several documented instances of fraudulent behaviour, “the process-based fraud detection system is proven over last 5 years”.

The operation of credit cards has always increased in these years, as financial business continues to have constant growth. The fraud companies are likewise creating are rapid increasing. Under those conditions, fraud detection has become a more and more important issue. However, disproportionate data records reinforce this problem because their fraud share is much lower than that of talent transactions. This work [3] is mainly investigated the use of boosting strategies to the detection of credit card fraud, while also presenting various boosting methods alongside a brief comparison of [29, 30].

Credit card fraud is highly relevant all over the world. The use of machine learning technologies, including data mining tools, has recently attracted considerable attention in the recognition of credit card fraud. Still, there are other difficulties such poorly balanced class sizes, missing publically available data sets, variation dishonest behavior etc. [5] [This work evaluates three machine learning techniques. "Logistics regression in the identification of real data fraud by random forests, vector system support, and credit card transactions [20]." Apply a small sample to compensate for the class size. The use of incremental learning of specific “ML algorithms in experiments” can help address the problem of ever-changing fraud trends. Commonly accepted metric: precision and recall guides the evaluation of the techniques' performance.

In financial services, credit card fraud poses a major trouble. Each layer credit card theft causes losses of billions of dollars. Confidential entities undertake research studies to analyse authentic credit card data. This paper machine identifies credit card fraud utilising learning methodologies [10, 15, 20]. Initially, conventional models are employed. The predominant polishing techniques and hybrid methodologies are employed in conjunction with Ad boost. Utilizing publicly accessible credit card datasets allows for the validation of the model's efficacy. 4 A genuine “credit card data set from a financial institution is subsequently analyzed”. Furthermore included into the facts samples is noise to evaluate the algorithms' resilience even more. The testing results show favorably that the majority voting method detects fraud situations in credit cards with reasonable accuracy rates.

In the use, healthcare fraud is a costly white-collar crime with no victimless aspect. Costs related to fraud are passed on to the public as higher premiums or major injury to beneficiaries [2, 7]. Digital healthcare fraud detection technologies must develop in order to counteract this social hazard. Implementing digital improvements in healthcare is tough given the complicated, heterogeneous data infrastructures and different health fashions around thus. Detection of fraud in healthcare systems should provide investigators who will be tested in terms of recovery, recovery, or relocation to a responsible agency or authority in connection with recovery, recovery, or transfer. “In this article, fraud systems and techniques” discovered in the literature are presented and edited to healthcare systems [7]. The main goals, results, and data features of the tabular list of papers examined by experts are presented in these studies. When using such a system, possible holes are passed through

actual medical data. The authors propose several objects to cover these gaps for the next generation of scientists in this area.

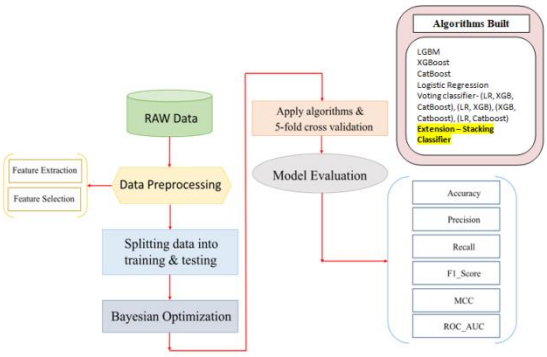
3. METHODOLOGY

i) Proposed Work:

The program employs “machine learning techniques” to deliver an advanced scam detection solution for banking data. “Utilizing the VAT setting and Biecian optimisation [29, 30, 31, 32] with Catboost, LightGBM, and XGBoost" enhances performance. The deep learning system enhances overall performance and ensures the effectiveness of evaluating critical criteria in real-world applications. Combining predictions from "RandomForest and LightGBM [17, 28] classifiers with certain settings, a stacking classifier has been developed". Using the strengths of several models, this ensemble technique improves prediction accuracy by means of a GradientBoosting Classifier as the last estimator. Furthermore produced is a user-friendly Flask framework connected with SQLite with sign-up and sign-in The green user test, therefore the accessibility and practical properties of the actual fraud detection system have been improved.

ii) System Architecture:

Starting with raw facts including credit card transaction specifics, the system uses labels and attributes suggesting fraud or validity. “To train data for machine learning”, it is essential to include functional extraction and selection operations. The dataset primarily consists of a test set for performance evaluation and a training set for model development. The “hyperparameter configuration of the machine learning algorithm is finalised via optimisation techniques”. Five-fold cross-validation is employed to “utilise machine learning algorithms such as CatBoost, LightGBM, and XGBoost”, which ensure the robustness of the training data model. We have also examined the utilisation of a stacking classifier to enhance the notion. The efficacy of algorithms in identifying credit card fraud is assessed by evaluating their lowered false positive rates using several metrics.



“Fig 1 Proposed architecture”

iii) Dataset collection:

“CREDIT CARD FRAUD DATASET”: Trained algorithms for “machine learning with Kaggle Credit Card Fraud Data Records”. The collection began with many properties related to transactions including "amount," "Time," and "V1" via "V28." Specific information regarding those original characteristics was omitted for privacy and security worries, therefore safeguarding sensitive facts while nevertheless allowing for efficient training in fraud prevention. Consequently, these credit cards represent the five leading categories concerning fraud statistics. Consequently, it comprises 32 columns; we present only a limited selection of them [6, 17].

V1	V2	V3	V4	V5	V6	V7	V8	V9	V10	...	V23	V
-0.611712	-0.769705	-0.149759	-0.224877	2.028577	-2.019887	0.292491	-0.523020	0.358468	0.070050	...	0.380739	0.0234
-0.814682	1.319219	1.329415	0.027273	-0.284871	-0.653985	0.321552	0.435975	-0.704298	-0.600684	...	0.090660	0.4011
-0.318193	1.118618	0.969864	-0.127052	0.569563	-0.532484	0.706252	-0.064966	-0.463271	-0.528357	...	-0.123884	-0.4956
-1.328271	1.016378	1.775426	-1.574193	-0.117696	-0.457733	0.681867	-0.031641	0.383872	0.334853	...	-0.239197	0.0099
1.276712	0.617120	-0.578014	0.879173	0.061706	-1.472002	0.373692	-0.287204	-0.084482	-0.696578	...	-0.076738	0.2587

‹ 32 columns

“Fig 2 NSL KDD dataset”

iv) Data Processing:

Data processing generates latent data for commercial value, hence creating knowledge. Data researchers often manage data by collecting, processing, validating, analysing, and converting it into comprehensible formats such as graphs or academic articles. Three strategies - Mainel, mechanical, and electronic touch - for managing the treatment of data. The objective is to elevate the knowledge fee and diminish

the selection presentation. This enables agencies to run better and make quick strategic decisions. that is exceptionally influenced by automated data processing technologies including computer software programming. huge facts among other volumes of data may be turned into valuable insights for quality control and decision-making.

v) Feature selection:

Feature selection as the scope and variety of datasets keep expanding, methodically shrinking their size is vital. Feature selection mostly aims to lower the computational fee of modelling and enhance the performance of a prediction model.

The functional choice machine is a crucial component of functional technique, involving the selection of the most significant functions to integrate into the learning system. The elimination of unproductive or unnecessary functions diminishes the functional set to those most pertinent to machine learning models, whereas facilitation methods assist in reducing the amount of input variables. Machine learning models are predominantly used to ascertain the most pertinent aspects by selecting features prematurely.

vi) Algorithms:

- **LGBM "(Light Gradient Boosting Machine)":** LGBM is very effective and well-suited for huge datasets, functioning as a shield-seeking framework. It possesses speed and precision, rendering it appropriate for fraud detection tasks. The decision to optimise the transition to LGBM enhances convergence [28] and results in a collection of trees.

```
# create purpose function
def lgbm_cv(learning_rate, max_depth, num_leaves):
    model = LGBMClassifier(learning_rate = learning_rate,
                           num_leaves = int(round(num_leaves)),
                           max_depth = int(round(max_depth)),
                           class_weight = 'balanced'
                          )

    cv = StratifiedKFold(n_splits=5)
    scores = cross_validate(model, X_train, y_train, cv=cv, scoring= 'neg_log_loss')
    return np.mean(scores['test_score'])

# Interval to be explored for input values
params = {'learning_rate': (0.001, 0.2),
          'max_depth': (-1, 8),
          'num_leaves': (2, 250)
         }

from bayes_opt import BayesianOptimization
lgbmBO = BayesianOptimization(lgbm_cv, params)

start = time.time()
lgbmBO.maximize(init_points=5, n_iter = 8, acq='ei')

print('It takes %s minutes' % ((time.time() - start)/60))
params_lgbm = lgbmBO.max['params']
params_lgbm['max_depth'] = round(params_lgbm['max_depth'])
params_lgbm['num_leaves'] = round(params_lgbm['num_leaves'])
print(params_lgbm)
```

"Fig 3 LGBM"

- **XGBoost (Extreme Gradient Boosting):** A prevalent gradient boosting technique employed in numerous machine learning applications is XGBoost. It possesses both performance and addictive qualities. XGBOOST addresses unbalanced datasets in fraud detection with a regularization-promoting methodology.

```
def xgb_cv(learning_rate, max_depth, n_estimators):
    model = XGBClassifier(learning_rate = learning_rate,
                          max_depth = int(round(max_depth)),
                          n_estimators = int(round(n_estimators)),
                          scale_pos_weight = 592
                         )
    cv = StratifiedKFold(n_splits=5)
    scores = cross_validate(model, X_train, y_train, cv=cv, scoring='neg_log_loss')
    return np.mean(scores['test_score'])

# Interval to be explored for input values
params={'learning_rate': (0.001, 0.2),
        'max_depth': (3, 10),
        'n_estimators': (50, 100)
       }

from bayes_opt import BayesianOptimization
xgbBO = BayesianOptimization(xgb_cv, params)

start = time.time()
xgbBO.maximize(init_points=5, n_iter = 8, acq='ei')

print('It takes %s minutes' % ((time.time() - start)/60))

params_xgb = xgbBO.max['params']
params_xgb['max_depth'] = round(params_xgb['max_depth'])
params_xgb['n_estimators'] = round(params_xgb['n_estimators'])
params_xgb['learning_rate'] = round((params_xgb['learning_rate']),4)
print(params_xgb)
```

"Fig 4 XGBoost"

- **CatBoost (Categorical Boosting):** The gradient boost tool handled specifically for categorical facts is catboost. It simplifies management of categorical data by automating it, therefore facilitating working with such databases. Dealing with actual-world financial data, it is strong, manages "overfitting, and can be helpful [29, 30, 31, 32]".

```
# create purpose function
import catboost as cgb
from bayes_opt import BayesianOptimization
def cat_cv(learning_rate, depth, iterations):
    model = CatBoostClassifier(learning_rate = learning_rate,
                               depth = int(round(depth)),
                               iterations = int(round(iterations)),
                               class_weights = {0:1, 1:592},verbose=False
                               )
    cv = StratifiedKFold(n_splits=5)
    scores = cross_validate(model, X_train, y_train,verbose=False, cv=cv, scoring='neg_log_loss')
    return np.mean(scores['test_score'])

# Interval to be explored for input values
params={'learning_rate': (0.001, 0.2),
        'depth' : (6, 16),
        'iterations': (50, 200)
         }

from bayes_opt import BayesianOptimization
catBO = BayesianOptimization(cat_cv, params)
start = time.time()
catBO.maximize(init_points=4, n_iter = 8, acq='ei')

print('It takes %s minutes' % ((time.time() - start)/60))

params_cat = catBO.max['params']
params_cat['depth'] = round(params_cat['depth'])
params_cat['iterations'] = round(params_cat['iterations'])
print(params_cat)
```

"Fig 5 Catboost"

- **Logistic Regression:** One essential binary type method is logistic regression. While a group of artists, such as those involved in boosting, is not evaluated, it establishes a baseline for identifying fraud. It facilitates comprehension of convenience in relation to relevance and is really straightforward.

```
log_reg = LogisticRegression(class_weight='balanced')
cv_results(log_reg, output_type='dict')
```

"Fig 6 Logistic regression"

- **Voting Classifier:** Combining the predictions of several "machine learning models—including Logistic Regression, XGBoost, and Catboost—the voting Classifier generates a single forecast". often producing better accuracy and resilience, this ensemble technique uses the pooled intelligence of several models. we have developed voting classifiers combining several methods [19, 24].

```
from sklearn.ensemble import StackingClassifier

estimators = [('rf', RandomForestClassifier(n_estimators=1000, random_state=4000)),('lgbm', LGBMClassifier(learning_rate='0.182'

clf = StackingClassifier(estimators=estimators, final_estimator=GradientBoostingClassifier(n_estimators=1000, learning_rate=1.0,

#HYPER_PARAMETR
lightgbm = lgb.LGBMClassifier(learning_rate='0.182', max_depth= '8', num_leaves= '33', class_weight='balanced')
xgboost = XGBClassifier(scale_pos_weight = 592, learning_rate= 0.1109, max_depth=9, n_estimators= 98)
catboost = CatBoostClassifier(scale_pos_weight = 592,verbose=False)

#ENSEMBLE
Model1 = [('lightgbm', lightgbm), ('xgboost', xgboost), ('catboost', catboost)]
Model2 = [('lightgbm', lightgbm), ('xgboost', xgboost)]
Model3 = [('catboost', catboost), ('xgboost', xgboost)]
Model4 = [('lightgbm', lightgbm), ('catboost', catboost)]

voting1 = VotingClassifier(estimators=Model1,voting='soft')
voting2 = VotingClassifier(estimators=Model2,voting='soft')
voting3 = VotingClassifier(estimators=Model3,voting='soft')
voting4 = VotingClassifier(estimators=Model4,voting='soft')
```

"Fig 7 Voting classifier"

- **Neural Network:** A neural network, inspired by the human brain, is a sophisticated learning model. In this regard, it can eliminate complex computer connections and configurations. Particularly in a wide range of "datasets, neural networks are used for their ability to recognize complex fraud" formulas.

```
def generate_model(batch_size, epochs, neuronPct):

    model = Sequential()
    neurons = int(neuronPct * 100)
    # So long as there would have been at least 20 neurons and fewer than 5layers, create a new layer.
    layer = 0
    while  round(neurons)>20 and layer <5:
        # The first (0th) layer needs an input input_dim(neuronCount)
        if layer==0:
            model.add(Dense(neurons,input_dim=31 , activation= 'relu', kernel_initializer='he_uniform'))
        else:
            model.add(Dense(neurons, activation='relu'))

        layer += 1
        neurons = round((neurons +1)/2)

    model.add(Dense(1,activation='sigmoid')) # Output
    return model
```

"Fig 8 Neural network"

- **Stacking classifier:** As a supplement, we developed a stacking classifier. The file method, the stacking classifier combines with specific settings, predictions from two basic classifiers - "random forest and lightGBM". As a last estimate, he uses the "Gradientostingclassifier" as the last estimate that combines the strengths of several models in the learning of the file, improving the accuracy of prediction.

```
#Extension
from sklearn.ensemble import RandomForestClassifier
from sklearn.tree import DecisionTreeClassifier
from sklearn.ensemble import GradientBoostingClassifier

from sklearn.ensemble import StackingClassifier

estimators = [('rf', RandomForestClassifier(n_estimators=1000, random_state=4000

clf = StackingClassifier(estimators=estimators, final_estimator=GradientBoosting
```

"Fig 9 Stacking classifier"

4. EXPERIMENTAL RESULTS

Precision: Accurate metrics designated as positivity are categorised as correctly classified events or snippets. The formula for calculating accuracy is as follows:

"Precision = True positives/ (True positives + False positives) = TP/(TP + FP)"

$$\text{Precision} = \frac{True\ Positive}{True\ Positive + False\ Positive}$$

Recall: In "Machine Learning", the memory of a metric This measures the ability of the model to identify all relevant cases in a particular category. It brings information about the accuracy of the model in relation to precisely positive comments on the overall real positivity.

$$Recall = \frac{TP}{TP + FN}$$

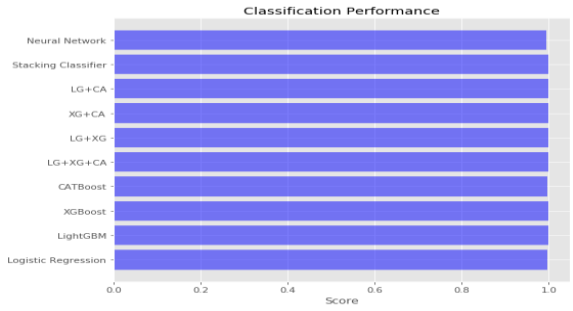
Accuracy: In a classification task, accuracy denotes the proportion of correct predictions, hence assessing the overall efficacy of a model.

$$Accuracy = \frac{TP + TN}{TP + FP + TN + FN}$$

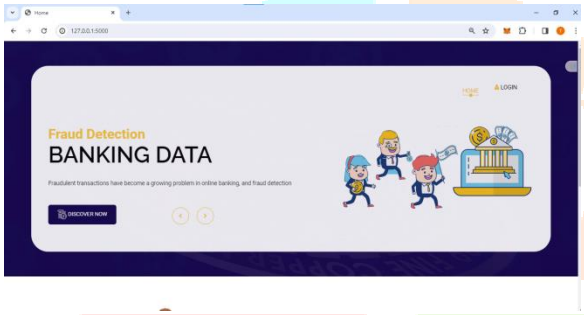
F1 Score: "F1 score" is appropriate for unbalanced datasets, as it measures accuracy and harmonises Accuracy and recall

that provide a balanced rating declare both “false positives and false negative” statements.

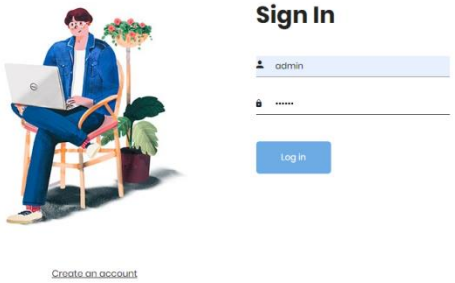
$$\text{F1 Score} = 2 * \frac{\text{Recall} \times \text{Precision}}{\text{Recall} + \text{Precision}} * 100$$



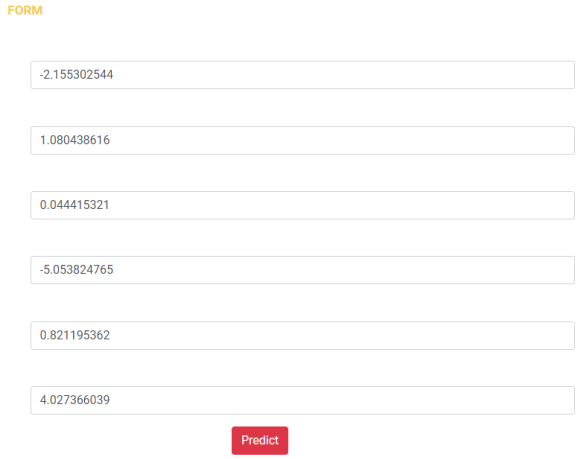
“Fig 10 Performance Evaluation”



“Fig 11 Home page”



“Fig 12 Signin page”



“Fig 13 User input”

“Fig 14 Predict result for given input”

5. CONCLUSION

The stacking classification demonstrated superior accuracy compared to all other models, proving highly effective in fraud detection. Emphasising its flexibility, the project presented strong performance across a range of “machine learning models, including LightGBM, XGBoost, Catboost [29, 30, 31, 32], voting classifiers and neural networks”. Emphasising their relevance, using several sample and scaling methods greatly helped to increase fraud detection accuracy. Emphasising its efficiency, Employ a contingency of artists to enhance the accuracy of fraud detection through stratified classification. An accessible flask via front-end, personalised testing, and certification is simplified, so guaranteeing accessibility and practicality. Testing the machine in Flask, using input, confirms its usability and functionality. [1, 2, 3] The effects of the study display the possibilities of sophisticated machine learning methods in tackling problems of “fraud detection in the financial industry”, therefore opening the path for further uses. Through investigating different ensemble approaches and optimisation strategies, the results of the study create

chances for ongoing development. in the end, the outcomes of the project assist the banking sector by improving general security and confidence, lowering financial losses, and boosting "fraud detection tools", so guaranteeing safe transactions.

6. FUTURE SCOPE

"Future work examines the merger of other hybrid models with Catboost [29]" to improve the accuracy and resistance of fraud detection. Future research will be a great adjustment Cat Boost's hyper parameters, with a watch towards specifically maximising the tree "count to increase the model's efficiency [33]". Studies will concentrate on ways to alter to always shifting fraud trends so that the model stays useful in spotting newly occurring fraudulent behaviour. Real-time data inclusion into ongoing studies seeks to increase system responsiveness and adaptability, therefore enabling faster reactions to new hazards. The future initiative will focus on the simplification of the model production process to make its thinking better understand and therefore promote the trust and strengthening of fraud detection techniques.

REFERENCES

[1] J. Nanduri, Y.-W. Liu, K. Yang, and Y. Jia, ''Ecommerce fraud detection through fraud islands and multi-layer machine learning model,'' in Proc. Future Inf. Commun. Conf., in Advances in Information and Communication. San Francisco, CA, USA: Springer, 2020, pp. 556–570.

[2] I. Matloob, S. A. Khan, R. Rukaiya, M. A. K. Khattak, and A. Munir, ''A sequence mining-based novel architecture for detecting fraudulent transactions in healthcare systems,'' IEEE Access, vol. 10, pp. 48447–48463, 2022.

[3] H. Feng, ''Ensemble learning in credit card fraud detection using boosting methods,'' in Proc. 2nd Int. Conf. Comput. Data Sci. (CDS), Jan. 2021, pp. 7–11.

[4] M. S. Delgosha, N. Hajiheydari, and S. M. Fahimi, ''Elucidation of big data analytics in banking: A four-stage delphi study,'' J. Enterprise Inf. Manage., vol. 34, no. 6, pp. 1577–1596, Nov. 2021.

[5] M. Puh and L. Brkić, ''Detecting credit card fraud using selected machine learning algorithms,'' in Proc. 42nd Int. Conv. Inf. Commun. Technol., Electron. Microelectron. (MIPRO), May 2019, pp. 1250–1255.

[6] K. Randhawa, C. K. Loo, M. Seera, C. P. Lim, and A. K. Nandi, ''Credit card fraud detection using AdaBoost and majority voting,'' IEEE Access, vol. 6, pp. 14277–14284, 2018.

[7] N. Kumaraswamy, M. K. Markey, T. Ekin, J. C. Barner, and K. Rascati, ''Healthcare fraud data mining methods: A look back and look ahead,'' Perspectives Health Inf. Manag., vol. 19, no. 1, p. 1, 2022.

[8] E. F. Malik, K. W. Khaw, B. Belaton, W. P. Wong, and X. Chew, ''Credit card fraud detection using a new hybrid machine learning architecture,'' Mathematics, vol. 10, no. 9, p. 1480, Apr. 2022.

[9] K. Gupta, K. Singh, G. V. Singh, M. Hassan, G. Himani, and U. Sharma, ''Machine learning based credit card fraud detection—A review,'' in Proc. Int. Conf. Appl. Artif. Intell. Comput. (ICAAIC), 2022, pp. 362–368.

[10] R. Almutairi, A. Godavarthi, A. R. Kotha, and E. Ceesay, ''Analyzing credit card fraud detection based on machine learning models,'' in Proc. IEEE Int. IoT, Electron. Mechatronics Conf. (IEMTRONICS), Jun. 2022, pp. 1–8.

[11] N. S. Halvaiee and M. K. Akbari, ''A novel model for credit card fraud detection using artificial immune systems,'' Appl. Soft Comput., vol. 24, pp. 40–49, Nov. 2014.

[12] A. C. Bahnsen, D. Aouada, A. Stojanovic, and B. Ottersten, ''Feature engineering strategies for credit card fraud detection,'' Expert Syst. Appl., vol. 51, pp. 134–142, Jun. 2016.

[13] U. Porwal and S. Mukund, ''Credit card fraud detection in e-commerce: An outlier detection approach,'' 2018, arXiv:1811.02196.

[14] H. Wang, P. Zhu, X. Zou, and S. Qin, ''An ensemble learning framework for credit card fraud detection based on training set partitioning and clustering,'' in Proc. IEEE

SmartWorld, Ubiquitous Intell. Comput., Adv. Trusted Comput., Scalable Comput. Commun., Cloud Big Data Comput., Internet People Smart City Innov. (SmartWorld/SCALCOM/UIC/ATC/CBDCom/IOP/SCI), Oct. 2018, pp. 94–98.

[15] F. Itoo, M. Meenakshi, and S. Singh, ‘‘Comparison and analysis of logistic regression, Naïve Bayes and knn machine learning algorithms for credit card fraud detection,’’ Int. J. Inf. Technol., vol. 13, no. 4, pp. 1503–1511, 2021.

[16] T. A. Olowookere and O. S. Adewale, ‘‘A framework for detecting credit card fraud with cost-sensitive meta-learning ensemble approach,’’ Sci. Afr., vol. 8, Jul. 2020, Art. no. e00464.

[17] A. A. Taha and S. J. Malebary, ‘‘An intelligent approach to credit card fraud detection using an optimized light gradient boosting machine,’’ IEEE Access, vol. 8, pp. 25579–25587, 2020.

[18] X. Kewei, B. Peng, Y. Jiang, and T. Lu, ‘‘A hybrid deep learning model for online fraud detection,’’ in Proc. IEEE Int. Conf. Consum. Electron. Comput. Eng. (ICCECE), Jan. 2021, pp. 431–434.

[19] T. Vairam, S. Sarathambekai, S. Bhavadharani, A. K. Dharshini, N. N. Sri, and T. Sen, ‘‘Evaluation of Naïve Bayes and voting classifier algorithm for credit card fraud detection,’’ in Proc. 8th Int. Conf. Adv. Comput. Commun. Syst. (ICACCS), Mar. 2022, pp. 602–608.

[20] P. Verma and P. Tyagi, ‘‘Analysis of supervised machine learning algorithms in the context of fraud detection,’’ ECS Trans., vol. 107, no. 1, p. 7189, 2022.

[21] J. Zou, J. Zhang, and P. Jiang, ‘‘Credit card fraud detection using autoencoder neural network,’’ 2019, arXiv:1908.11553.

[22] D. Almhaithawi, A. Jafar, and M. Aljnidi, ‘‘Example-dependent costsensitive credit cards fraud detection using SMOTE and Bayes minimum risk,’’ Social Netw. Appl. Sci., vol. 2, no. 9, pp. 1–12, Sep. 2020.

[23] J. Cui, C. Yan, and C. Wang, ‘‘Learning transaction cohesiveness for online payment fraud detection,’’ in Proc. 2nd Int. Conf. Comput. Data Sci., Jan. 2021, pp. 1–5.

[24] M. Rakhshaninejad, M. Fathian, B. Amiri, and N. Yazdanjue, ‘‘An ensemble-based credit card fraud detection algorithm using an efficient voting strategy,’’ Comput. J., vol. 65, no. 8, pp. 1998–2015, Aug. 2022.

[25] A. H. Victoria and G. Maragatham, ‘‘Automatic tuning of hyperparameters using Bayesian optimization,’’ Evolving Syst., vol. 12, no. 1, pp. 217–223, Mar. 2021.

[26] H. Cho, Y. Kim, E. Lee, D. Choi, Y. Lee, and W. Rhee, ‘‘Basic enhancement strategies when using Bayesian optimization for hyperparameter tuning of deep neural networks,’’ IEEE Access, vol. 8, pp. 52588–52608, 2020.

[27] F. N. Khan, A. H. Khan, and L. Israt, ‘‘Credit card fraud prediction and classification using deep neural network and ensemble learning,’’ in Proc. IEEE Region 10 Symp. (TENSYMP), Jun. 2020, pp. 114–119.

[28] W. Liang, S. Luo, G. Zhao, and H. Wu, ‘‘Predicting hard rock pillar stability using GBDT, XGBoost, and LightGBM algorithms,’’ Mathematics, vol. 8, no. 5, p. 765, May 2020.

[29] S. B. Jabeur, C. Gharib, S. Mefteh-Wali, and W. B. Arfi, ‘‘CatBoost model and artificial intelligence techniques for corporate failure prediction,’’ Technol. Forecasting Social Change, vol. 166, May 2021, Art. no. 120658.

[30] J. Hancock and T. M. Khoshgoftaar, ‘‘Medicare fraud detection using CatBoost,’’ in Proc. IEEE 21st Int. Conf. Inf. Reuse Integr. Data Sci. (IRI), Aug. 2020, pp. 97–103.

[31] B. Dhananjay and J. Sivaraman, ‘‘Analysis and classification of heart rate using CatBoost feature ranking model,’’ Biomed. Signal Process. Control, vol. 68, Jul. 2021, Art. no. 102610.

[32] Y. Chen and X. Han, ‘‘CatBoost for fraud detection in financial transactions,’’ in Proc. IEEE Int. Conf. Consum. Electron. Comput. Eng. (ICCECE), Jan. 2021, pp. 176–179.

[33] A. Goyal and J. Khiari, ‘‘Diversity-aware weighted majority vote classifier for imbalanced data,’’ in Proc. Int. Joint Conf. Neural Netw. (IJCNN), Jul. 2020, pp. 1–8.

[34] A. Roy, J. Sun, R. Mahoney, L. Alonzi, S. Adams, and P. Beling, ‘‘Deep learning detecting fraud in credit card transactions,’’ in Proc. Syst. Inf. Eng. Design Symp. (SIEDS), Apr. 2018, pp. 129–134.