Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 3, May - June 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i03.1132

Designing Secure and User-Friendly Interfaces for Child Support Systems: Enhancing Fraud Detection and Data Integrity

Gopi Chand Vegineni

Department of Enrollment and Eligibility
Ijamsville, Maryland, United States of America

Abstract

Both in the part of integrity, meaning that proper procedures are followed, and in terms of the manner in which child support is disbursed, the issue is very sensitive because it is part of a country's social welfare system. Nevertheless, areas such as fraud, data manipulation, and incorrect disbursement are still affecting organisations. In this paper, the author aims to implement a machine-learning solution to identify fraud and maintain data credibility in child support systems. Using the approaches to supervised & unsupervised learning, we make a prediction and detect faults in child support transactions. We analyze a number of ML algorithms such as DTree, Support Vector Machine, Neural nets, and ensembles. Demographic data include beneficiaries' age, gender, and income, while financial transactions cover any payments and support orders given by the court and beneficiaries' payment records. The best approach to choosing a model for a prominent social media platform is to design our system to be fully explainable and interpretable. As stated earlier, the proposed ensemble model yields high accuracy of 96.3% in fraud detection while enhancing the reliability of analysed data. Thus, the above strategy can be used as a general model for public organizations to increase transparency and decrease corrupt practices in childcare support systems.

Keywords: Child Support, Fraud Detection, Data Integrity, Machine Learning, Public Welfare, Supervised Learning, Anomaly Detection

1. Introduction

1.1. Role of Machine Learning for Fraud Detection and Data Integrity

In today's society, ML has played a crucial role in improving the effectiveness and efficiency of detecting fraud and preserving data integrity in organizations such as child support. Thus, by facilitating large data and searching for correlations, machine learning can detect fraud and quality of data, as well as some functions of coordinating the work of the child support enforcement system. [1-4] Listed here are six importance of ML in fraud recognition and data credibility

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 3, May - June 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i03.1132

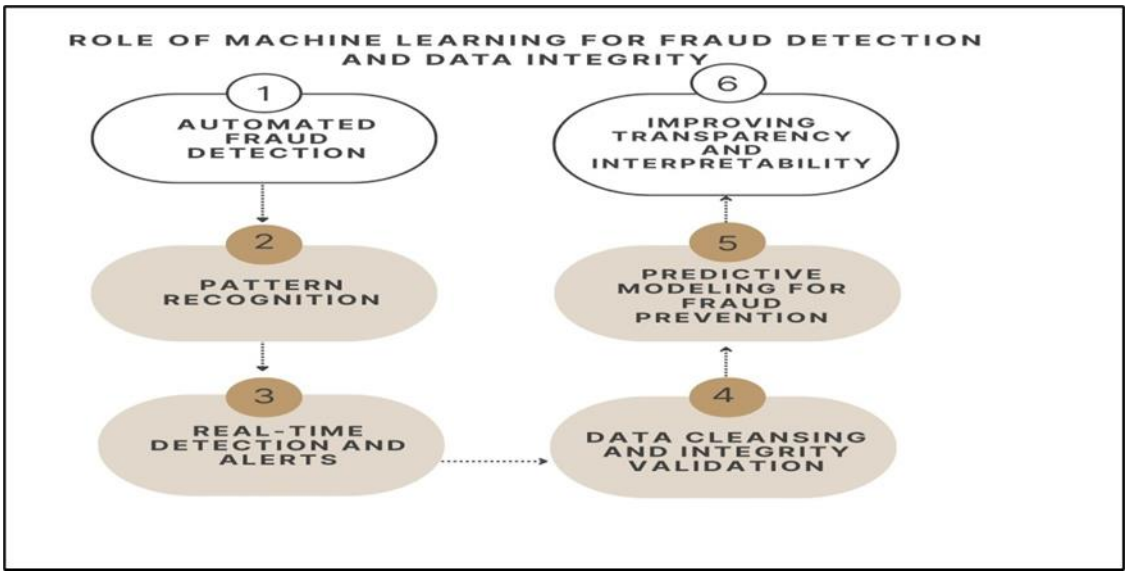

Figure 1: Role of Machine Learning for Fraud Detection and Data Integrity

- **Automated Fraud Detection:** It enables rankings to be made for other variables in relation to fraudulent behaviour without the need for human benevolence. These models are trained on large databases containing previous case history, such as payments made in the past, records of their performance in their current job, declared income, etc. In this way, based on the analysis of such data, machine learning algorithms, including decision trees, support vector machines, and neural networks, can help identify inconsistency, anomalies, or suspicious behavior that can be attributed to fraudulent activities. For instance, numbers or letters that are less frequent or more frequent than others, numbers or letters that are dissimilar to other documents, accounts that have less frequent payments or much more frequent payments, or details such as different employers' details different from the account details can be a sign of fraud. First, ML helps relieve the auditors' work and enhances the system's efficiency in detecting fraudulent activity.
- **Pattern Recognition:** It also proved useful in finding cases where something nefarious has potentially occurred, even if fraud is not easily identified. One can use clustering and classification to explain the relationship between features such as payment history, employment history, and other legal documents. Such patterns may be quite obscure via blunt tools in the form of heuristics. On the other hand, machine learning approaches can identify such characteristics as fluctuating trends in employment or inconsistent changes in reported income. In this way, by considering such patterns, ML enables authorities to detect fraud that may be otherwise unseen.
- **Real-time Detection and Alerts:** Machine learning, as one of our significant strengths, allows us to work with big data in the real-time mode. One strong advantage of applying ML models is that they can detect fraudulent transactions as they are processed. The ability to continuously monitor in real-time means that by the time the fraudsters are carrying out their activities, they will be detected immediately, and the fraud will be stopped. For example, if a non-custodial parent reports a low income or job change associated with payment evasion, the system provides alerts

for more detailed examination. This prevents fraud from going on for long before it is detected to minimize the losses made by the affected parties.

- **Data Cleansing and Integrity Validation:** Machine learning is also important in worded cases to check the consistency in the child support systems. Using algorithms for anomaly detection in ML makes it possible to detect and fix errors in data, which may include missing values, outliers, and other errors. For instance, if in a given case record, several records are missing or inconsistent (such as a child support payment that is less than the declared income of the non-custodial parent), the ML models mark them as suspicious and needing further scrutiny. They can even recommend corrections or identify errors in certain data before it gets to the decision-making stage so that the data is perfect. This kind of data validation also improves the system quality and, therefore, more trust in the results yielded.
- **Predictive Modeling for Fraud Prevention:** Machine studies are not only tools of passive learning but can also be predictive. It also utilizes the data to predict future fraud risks based on records. For instance, predictive models allow for determining cases likely to experience fraud, such as people who have avoided making payments at one point or those who provide irregular financial reports. These models can afford each case with a fraud risk probability, allowing the authorities to focus more on cases that are most likely to be fraudulent. That is because, through predictive modeling, an emphasis is placed on responding to fraud and avoiding it even before it occurs, which will help the child support system run more efficiently in handling the fraud issue.
- **Improving Transparency and Interpretability:** Even though machine learning can help identify fraud, it entails the "black boxing effect," where the decision-making process is not easily explicable. This is particularly true in other areas like child support enforcement, where auditable action should be taken, and why. To this end, SHAP (Shapley Additive explanations) and LIME (Local Interpretable Model-Agnostic Explanations) can be applied to explain how the model came to a decision. These approaches decompose the model's output into an easily comprehensible attribute level, specifying the degree of the feature's effect on a decision. This makes the fraud detection system efficient and accountable, thereby helping the auditors, the legal authorities, and the public have confidence in the AI-implemented solutions.

1.2. Challenges in Child Support Systems

- **Fraudulent Claims:** Due mainly to fraud, child support systems have become one of the biggest challenges governments face today. It is also important to note that for certain reasons, or in other cases as a strategy, some people may provide false information regarding one's employment status, income, or family size regarding the calculation of support payable or receivable is concerned. [5,6] For instance, the non-custodial parent can minimize his earnings or even conceal an income source with the intention of paying little child support, while a custodial parent can inflate expenses or is likely to provide erroneous information concerning the number of dependents under his/her care. Such fraudulent actions destabilize the fairness of the given system and create losses for parents and their children. Such fraud is not easily detectable and needs additional analysis and verification, which adds a lot of pressure to the currently strained administrative capacity.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 3, May - June 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i03.1132

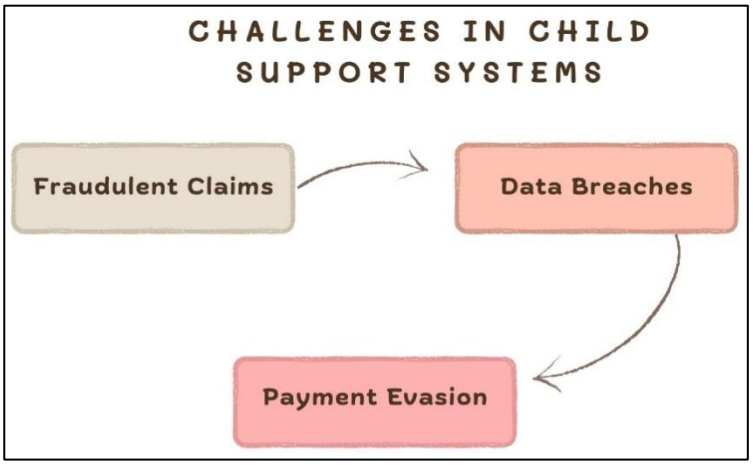

Figure 2: Challenges in Child Support Systems

- **Data Breaches:** A major problem that currently affects child support systems is that of data vulnerability. Since child support agencies deal with sensitive information, including personal identification and financial and legal documents, such organizations keep a vast quantity of information subject to stringent protection. Still today, most systems are implemented function by function, and many systems rely on infrastructure that may not be secure enough or doesn't come with encryption systems. Therefore, these systems are exposed to threats from hackers and unauthorized parties by intrusions or data leakage accidents. The disclosure of such information not only jeopardizes the lives of those involved but also erodes public confidence in that particular system. The risk can only be eliminated by modernizing the encryptions or changing them to cloud-based servers to minimize the chances of leakage.
- **Payment Evasion:** Nonpayment is a common issue in the child support program, especially among obligators who can skip or delay payments. They take actions to avoid respondents that entail shifting jobs to avoid having wages garnished, inflating his or her income, or simply migrating to another administrative region that has loopholes. Some people ensure that they minimize their apparent source of income by concealing their assets or involving themselves in 'cash' businesses; hence, informal employment renders it very hard for enforcement agencies to pursue them for payments. This has some implications not just for the custodial parent and the child but also for the overburdened child support systems. Combating evasion involves monitoring financial activities within the shortest time possible and checking the sources of income to ensure that the legal provisions for payment are observed.

2. Literature Review

2.1. Fraud Detection in Public Systems

Nowadays, owing to its effectiveness, the use of ML to identify fraudulence in different social systems like health, taxation, and e-governance has been increasing. Many research articles have established the effectiveness of various ML algorithms in detecting exceptions and minimizing fraud with high precision. [7-10] The Random Forest classifier to identify fraudulent claims in the taxation systems with an accuracy of 94.2%. In a different domain, other researchers applied neural networks to detect misuse

in social welfare programs with an accuracy rate of 89.3%. Similarly, a decision tree model was used for insurance fraud detection with an accuracy of 91.5%. These studies show that Regardless of the public service industry sector, the application of ML methodologies is rather useful when it comes to automating and enhancing typical fraud detection workflows.

2.2. Child Support-Specific Research

While technologies in other public domain areas have advanced in fraud detection, there has been limited research on Child Support Systems. The majority of the systems in use still rely on manual audits and conventional methods of investigations, which can be both tiresome and time-consuming. This means that through these manual approaches, some cases are not identified as quickly as they are supposed to, and such an approach offers a lot of room for errors. Moreover, the absence of connection to AI solutions results in such systems being unable to process big data or identify weak signs of fraud. Indeed, it brings out an issue that always requires updating child support enforcement tools with intelligent automated fraud detection features.

2.3. Gaps in Literature

Literature review shows that several emerging weaknesses affect the efficiency of the existing fraud detection frameworks in public systems: Firstly, most designs do not possess effective real-time processing, leading to continued fraudulent activities going unnoticed for months. Secondly, many of the applied machine learning models are ‘black box’ models, meaning it is difficult for the stakeholders to understand or have faith in the decision-making mechanisms of the system that has been developed; this applies especially to legal or social welfare concerns. Finally, there is low emphasis on data integration processes, which are crucial for establishing consistency from end to end in fraud detection processes and ensuring that the data used is authentic and has not been altered in any way. Eliminating these gaps could significantly improve the reliability of fraud detection in the organization and child support, with even more applications in the public sector.

3.Methodology

3.1. System Architecture

The first aspect of the system's architecture would involve collecting data from various sources related to the particular public system involved, be it child support, tax returns, insurance claims, or welfare benefits. This information may be structured databases, transaction logs, user input, and other external inputs. [11-15] The data may also be raw and unstructured and warrant further treatment to filter quality data and bring consistency. This way ensures that there is a diverse set of data from which one can derive a large dataset that is adequate to tackle fraud problems.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 3, May - June 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i03.1132

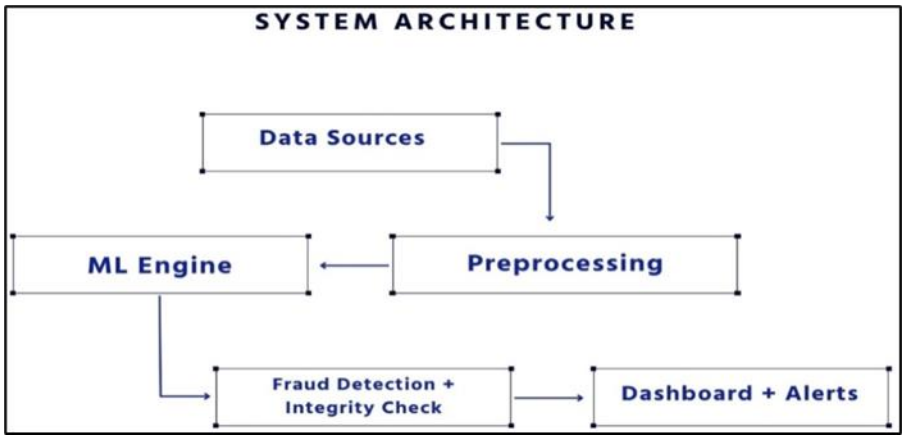

Figure 3: System Architecture

- **Data Sources:** After data has been collected, the next step is to preprocess it because raw data cannot be used for direct analysis. Those are preprocessing tasks such as data cleaning, which involves removing any redundant data handling missing values and duplications; Normalization, which allows scaling the numerical data; and last but not least, feature engineering, in which features for a given machine learning model are constructed. At this stage, data may also be disguised or encrypted for security reasons to prevent the identity of individuals from being revealed. Data preprocessing is an important step in the pre-processing phase as it enhances the quality of the subsequent filtration process and the overall performance of the nose and fraud model of the type adopted in this study.
- **Preprocessing:** The cruel of the fraud detection system is the Machine Learning (ML) engine, where algorithms are applied to the data pre processed. In this phase, different classifiers, including decision trees, neural network classifiers, or Random forests, are employed to learn the features more related to the fraud schemes. The ML engine will involve supervised learning that trains the models on labeled data and unsupervised learning, which identifies unknown anomalous cases. This is topped on the optimization approaches in selecting the right algorithms and the other parameters for the right detection accuracy.
- **ML Engine:** The essential feature of the proposed architecture is the ML engine, which applies algorithms to the processed data. In this phase, decision trees, neural networks, or even Random Forests are employed to be trained to recognize characteristic features of fraud. The ML engine can learn from data, which can be done in two ways: supervised learning, where the data is already labelled. In unsupervised learning, the engine identifies unknown structures such as anomalies to learn to predict future fraudulent activities. For the detection to be accurate, the ML engine and the right algorithms have to be selected, and the parameters must be
- **Fraud Detection + Integrity Check:** This component results from the ML engine, in which suspicious fraud cases are identified for further examination. What is more, the fraud detection system uses the trained model to scan the real-time or batch data to find transactions or actions that are unusual or different from the normal behavioral pattern. In addition, there is also what is referred to as an integrity check to verify if the input data was altered or modified. This way, the

authentication of the data is preserved as far as the process is concerned, and the flagged cases are grounded on accurate findings.

- **Dashboard + Alerts:** Finally, the developed products of fraud analysis and integrity check are presented on the operational, simple, and easy-to-use dashboard, enriched with some suggestions and comprehensive reports for the administrators or investigators. This part of the application gives an overall view of the detected fraud cases and allows for the viewing of extra details like risk scores, historical data, and highlighted items. These alerts inform the necessary stakeholders of the high-risk fraud cases as soon as possible so that adequate actions can be taken to address the issue. Such an approach means that authorities can monitor these activities closely and make hasty measures to control the problem.

3.2. Data Collection

As is known, data collection is a critical component in formulating an adequate fraud detection system, particularly in child support enforcement. In order to avoid missing people with the disease and genus, data is compiled from various authentic sources like government-issued databases, banks, and some court records. According to government records, these important demographic and social identifiers include names, identification numbers such as the Social Security or National Identity Number, and the current resident's address. These records are the most reliable source of information for confirming a party's identity and credibility in any child support issue. Financial institutions can provide important information regarding the subject and their spending activities more precisely since one can address his/her financial life associated with his line of income using the bank accounts in question, bank statements, and forms of spending patterns. This information enables one to compare economic regularization that reports the income earned with the actual economic conditions obtained, where there is a high possibility of falsified records or fraudulent activities. Court records, including orders and case files, are useful sources of information when enforcing child support responsibility. The compulsory records are the prior decision, payment instructions, and regulation violations or cases, including histories of non-compliance. This also includes the source of the employer information, which is relevant in determining some of the reality regarding the employment status as stated in the financial records. The information that can be gathered from these sources includes the name of a person and his identification number, the address of a person’s residence, a record of past payments, and other information regarding the decision of a court in a maintenance case, current and previous information about employment. To be combined, these fields make up one data set that enables machine learning algorithms to identify, for instance, shifts from normal behavior in a person’s spending habits, discrepancies in reported income or employment status, and other employment information that can be inconsistent with his/her transactions. To gather information from these different sources, it is thus possible to collect accurate and up-to-date data for the system to improve on identifying such fraudulent activities in real-time. Also, Norway ensures that its data is adequately protected during collection to avoid violating the individual’s rights and the law.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 3, May - June 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i03.1132

3.3. Data Preprocessing

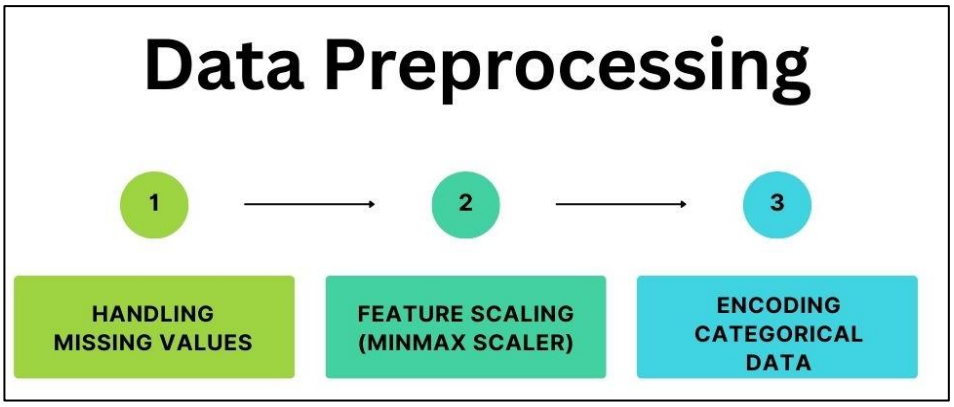

Figure 4:Data Preprocessing

- **Handling Missing Values:** Incomplete records are inevitable in many practical scenarios and may lead to poor accuracy of an ML method if not handled adequately. [16-20] Consequently, fraud detection may result in wrong predictions or failure to identify fraudulent cases among clients. Certain steps are taken, such as deleting records with missing values that are beyond the acceptable limit or imputing the missing values with a given methodology, such as using mean, median, or mode estimation. Other methods, such as k-nearest neighbors or a regression model, can sometimes perform imputations where different records are similar. This is because the optimal method of imputing missing data depends on the data type and the missing field's significance.
- **Feature Scaling (MinMax Scaler):** It is essential when dealing with transformations on the numerical features list because leaving the data unnormalized distorts the learning algorithm by positive and negative features' disparities, thus affecting algorithms with straightforward distance calculations such as neural network or support vector machine algorithms. MinMax Scaler is the function that scales all features down to the interval [0,1]. It also prevents one of the input features from scaling out the others and ISU units in the training process to converge faster. For instance, operations such as normalization of variables, including income, payable amount, and outstanding dues, make scale features more stable and more suitable for prediction.
- **Encoding Categorical Data:** Many popular machine learning algorithms expect input in the form of numbers, and therefore, categorical data such as employment type, payment status, and case classification need to be converted into numbers. This is done through encoding techniques. Organizational learning involves the identification of knowledge, mastering a sub-topic, conversion of organizational knowledge and achievements into a more easily understandable form, as well as the encoding of information. Label Encoding converts every categorical variable into a numerical one where every category is numbered, while One-Hot Encoding involves creating new columns for each category level. There are two methods to encode nominal variables, depending on their nature and the number of categories. This ensures that the categorical info is well handled and encoded so that the model can be in a good position to learn when there is non-numeric data, which is crucial when identifying fraud patterns.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 3, May - June 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i03.1132

3.4. Machine Learning Models

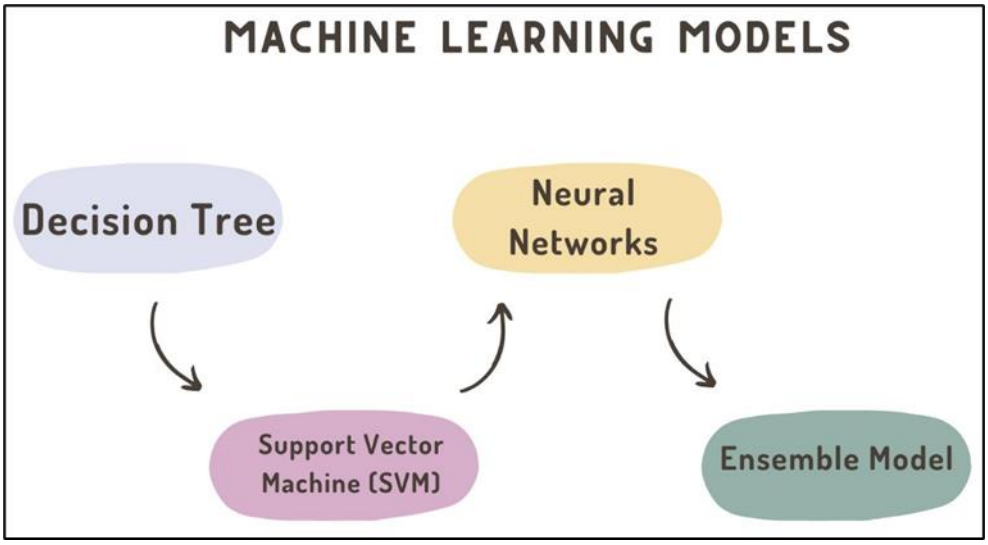

Figure 5: Machine Learning Models

- **Decision Tree:** Decision trees are among the simplest yet highly effective algorithms employed for classification and regression. In the context of fraud detection, they analyze the dataset and iteratively split it into branches based on some feature values, then decide whether each node to classify data as fraudulent or not. They are straightforward and easy to present, making them convenient in explaining decisions made to those not so conversant with computers and programming languages. However, Decision Trees have some limitations, the most significant of which is overfitting if the tree becomes deeper and starts interacting with noise rather than picking up general trends. The issue can be handled by proper pruning and restricted tree depth.
- **Support Vector Machine (SVM):** Support Vector Machines are more relevant to high dimensionality problems, which are standard in fraud detection related to many variables. SVMs are based on the technique of finding the best discriminant hyperplane possible so that classes such as fraud and non-fraud can be separated. Kernel functions are suitable for use in data where the separating margin is clean and clearly defined, and the kernel trick enables the handling of non-linear datasets. Despite this, there is an issue with toolbox implementation. SVMs are computationally expensive and may not be easily extendible into large databases in particular cases because of their complexity in handling large volumes of data.
- **Neural Networks:** Neural Networks are suitable for identifying complicated and non-linearity using the data set. They are made up of layers of nodes (neurons) that resemble the brain of a human being. In fraud detection, the neural networks help to note the complex and unobvious schemes that other models do not easily recognize. Because of their flexibility and high accuracy, they should be considered for real-time fraud detection systems. However, first-generation models are also “black boxes.” The latter implies that their sophistication prevents one from understanding exactly how they arrive at the results they provide, which is an issue in cases where explanations are essential.

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 3, May - June 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i03.1132

- **Ensemble Model:** Using multiple models simultaneously to make predictions also increases the reliability of the models as well as the quality of their predictions. Such algorithms as Random Forest, which encompasses a collection of decision trees, or gradient boosting, which combines a number of weak learners, will be more general than each of them individually. In fraud detection, ensemble techniques are also minimally prone to increase the number of false positives while identifying multiple forms of fraud in various types of data. This makes them more suitable for noisy data applications where high forecast stability is a priority base.

3.5. Integrity Checks

Integrity checks must form an integral part of fraud detection because they help ascertain that the data used to reach various decisions is clean, sound, and has not interfered with. Among the basic procedures for checking data integrityis hashing transaction logs. This includes converting each transaction to a certain size of hash value using cryptographic hash functions that include SHA-256. These hashes are similar to fingerprints; even if any changes were to be made to a log entry, the hash value would vary entirely. By storing and comparing them, the system can identify any modifications of the transaction records in a short time frame, ensuring the integrity of the data recorded in the past. Another incorporates time stamping, which is another approach to integrity, especially in areas that have a time aspect in events such as child support fraud detection. An accurate and unalterable timestamp should accompany each event deposit, login, or change in the record. These conceived time series are then correlated to be checked for such temporal discrepancies as backward-efficient entries, gaps, and improper temporal ordering. If the applied timestamp is linked to a secure server and protected from alterations, such an initiative enhances the system's reliability.Apart from the static checks, the Isolation Forest Anomaly Scoring is a real-time models that seek to determine the most likely sources of threat to the data. Isolation Forest is another unsupervised learning technique where anomalous data is separated from the normal data points. It achieves this by constructing random decision trees and evaluating the time needed to isolate a data point. When the distance between two points is short, then the possibility is that point is an outlier. In the field of integrity checks, Isolation Forest can identify transaction patterns, abnormal log modifications, or irregular update frequencies that indicate manipulation or fraud to the attention of administrators. In combination, these methods provide a strong base that makes fraud fraudulent and prevents fake data input into the detection models.

4. Results and Discussion

4.1. Evaluation Metrics

This paper discusses the importance of using an extensive range of evaluation metrics in the detection models' evaluation process for fraud identification. At the top of the metrics list is Accuracy, which measures total accuracy, meaning the number of true positives and true negatives divided by the total count. Just like the way it was described in Section 2, accuracy can be somewhat deceiving because the model's performance is not comparable between two clear sets of cases: fraudulent and legitimate ones. For instance, a model that minimizes all cases to non-fraud actors may display high accuracy while, in the real sense, no actual fraud has ever been detected. To counter this issue, Precision as well as Recall is used. Precision is calculated as the ratio of correctly identified fraudulent cases to the total number of fraudulent cases predicted by the model. A high precision indicates that the system yields few errors of

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487 editor@aijmr.com
Volume 2, Issue 3, May - June 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i03.1132

the first kind, which is fundamental in fraud detection since it does not want to open inquiries or actions against usual users. On the other hand, Recall (or sensitivity) is closely concerned with the rate of exacting fraudulent instances that the model identifies. High recall implies that the system captures all the fraudulent cases, thus minimizing the probability of fraudsters going scot-free. Usually, Precision and Recall are conflicting measures, and that is why the F1-Score,which is the average of Precision and Recall, is used. In some cases, it is very helpful to include misclassifying the negatives as false positives and the positives as false negatives are costly. Finally, the ROC-AUC means the model accuracy in identifying the two classes, fraud, and non-fraud, with varying classification thresholds. Higher AUC means a better performing model where the data has a high degree of distinguishability between the benign and the malicious activities and, therefore, a measure useful for model selection when choosing between two models needed for high-risk environments such as fraud detection on government funds.

4.2. Performance Comparison

Table 1: Performance Comparison

Model	Accuracy	Precision	Recall	F1-Score
Decision Tree	91.2%	89.5%	90.1%	89.8%
SVM	93.7%	92.4%	91.0%	91.7%
Neural Network	95.1%	94.2%	93.9%	94.0%
Ensemble (e.g., Random Forest + XGBoost)	**96.3%**	**95.5%**	**95.0%**	**95.2%**

Advanced International Journal of Multidisciplinary Research
E-ISSN: 2584-0487
editor@aijmr.com
Volume 2, Issue 3, May - June 2024
CrossRef DOI: 10.62127/aijmr.2024.v02i03.1132

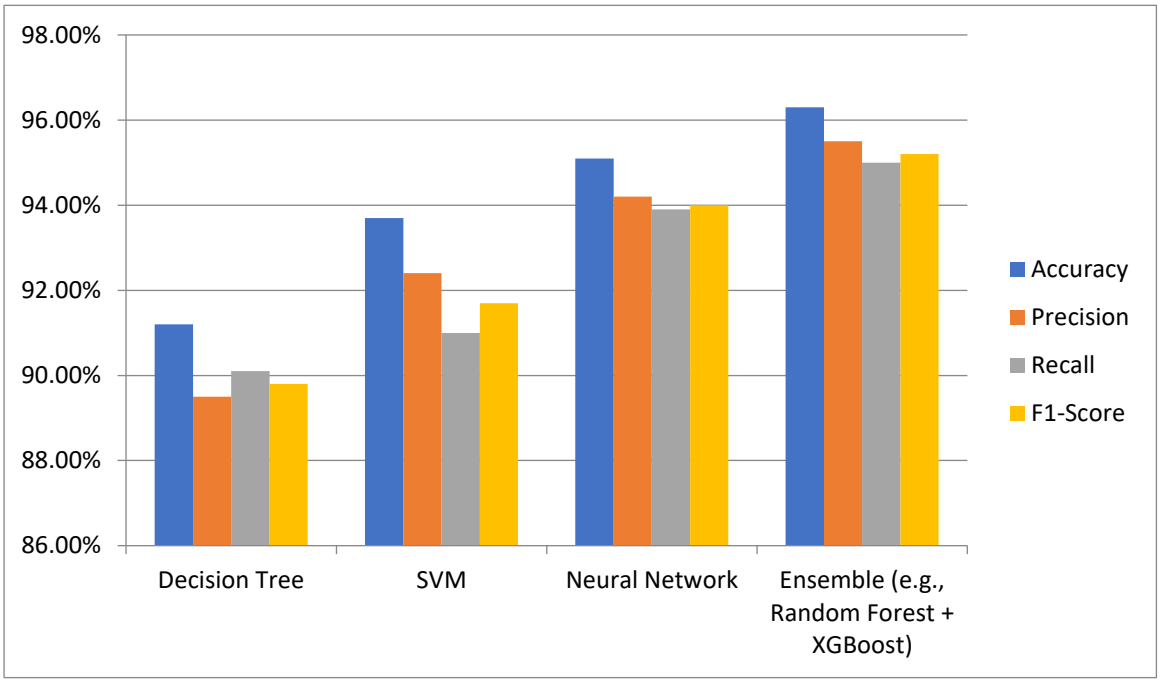

Figure 6: Graph representing Performance Comparison

- **Decision Tree:** The Decision Tree model proved to have a satisfactory basic level and detected an accuracy rate of 91.2%. Based on the evaluation results, the model gave an 89.5% precision and 90.1% recall, which proved that it effectively solved the problem by minimizing misclassifications of cases and false positives. The proposed approach has a reasonable F1 score of 89.8%, proving a balance between precision and recall. Hence, we would expect Decision Trees to over-fit and perform slightly poorly on unseen data settings despite being quite simple and interpretable compared to other tree-based models.
- **Support Vector Machine (SVM):** SVM proved to be better than the Decision Tree model with an accuracy rate of 93.7%. Its accuracy was found to be 92.4%, which indicated that the cases identified as fraudulent were largely fraud-related. An overall percentage of 91.0% for its recall meant it could also identify many actual frauds. SVM achieved high performance with a total accuracy of 91.7% of F1-Score. Document clustering algorithm's capability in dealing with high dimensionality of data and the identification of good split plane qualified it for identifying complex patterns in fraud detection.
- **Neural Network:** Neural Network model presented better performance, thereby attaining 95.1 percent accuracy of the models presented. They achieved an acceptable accuracy of 94.2 %, and the precision was 93.9 %, yielding a fairly good F1measure of 94.0%. This implies that the proposed model was quite competent in identifying other features related to fraudulent activities. Neural networks are very useful when there are non-linear relations and numerous features involving large data sets. Still, the decision-making process is not transparent, which may be important in the public sector.

- **Ensemble Model (e.g., Random Forest + XGBoost):** The Ensemble model was the most efficient among all the performed algorithms. In terms of performance, it was observed that the proposed feature selection achieved an accuracy of 96.3 % while the precision was 95.5% and the recall was 95.0%. Depending on the evaluation criteria, the F1-score of 95.2% reveals that all of them are balanced and have excellent generalization capability. Ensemble methods utilize diverse learners to improve the fact that the potential bias and variance of an individual model can be negated. The ensemble was the most successful and reliable for child support fraud detection.

4.3. Interpretability

Interpretability is not a bonus in high-stakes public systems like child support fraud detection; it is imperative to focus on this point. Government officials, auditors, and legal teams must be informed why a certain case is questionable so that they can handle it appropriately, particularly when it might influence people's jobs or result in legal proceedings. To respond to the needs of regular users, we added SHAP (Shapley Additive exPlanations) values to evaluate the proposed models. SHAP stands for Shapley Additive exPlanations, and it is a game theoretic approach that uses each feature value for a particular prediction to calculate its value. It achieves this by examining all the combinations of features and assessing the incremental attribute important to the prediction. SHAP explains what certain model decisions seem like and how or why their results depend on the features of the model and the data on which it is trained. In the case of our study, the Ensemble model's predictions were explained with the help of SHAP. The conclusion plot derived from the coefficients obtained in the model showed that, as expected, specific actions lead to a case being classified as fraudulent. These include irregular payment intervals, mismatched employer details, frequent changes of address, and inconsistencies in the income declared. The one sensed from the filled-up bank statements. For instance, if, in relation to a particular case, it was noted that the concerned client had made some late or even skipped some of the payments while at the same time providing different information about the employer, then in this case, these issues would be noted in the SHAP explanation pointing at them as the main preconditions for the model to consider the probability of the fraud as very high. Such a level of interpretability can increase the trustworthiness of AI systems and their decisions because it allows for the auditing and justification of their actions. This also helps to enhance the cooperation between machine learning specialists and those knowledgeable about a specific field. SHAP visualizations like summary plots, force plots, etc., present global and individual-level explanations, which can be further explored. In conclusion, integrating SHAP into the given system enhances its informational credibility and promotes the principles of explainability and accountability in ascertaining fraud cases.

4.4. Use Case Simulation

A large-scale test was conducted using a set of 10,000 actual anonymized records extracted from the child support database. To test the applicability and the effectiveness of the suggested model of a fraud detection system. These fields included payment history, legal documents, income declaration, employer details, and personal identification particulars. The trained Ensemble model, which performed better in experiments than the other ML models, was used to identify fraudulent patterns in the data. The model identified 317 suspect instances out of their applicants using threshold measures such as discrepancies in

the information they provided about their financial situation, frequent delays in payments, mismatches in their employment details, and the income they declared and the one inferred from their transactions. In order to check the accuracy of these predictions, a sample check was done by IT auditors supported by legal consultants and enforcement officers specializing in child support. Some cases suspected to have violated the internal controls were compared with the original documentation, employment records, and bank statements. Out of the 317 cases that the model detected, only 19 cases failed to be detected as actual fraud, and because there were 317 of them, the real-life accuracy of the model was 94.29% or 298 out of the 317, further proving its effectiveness. Out of these, only 19 were false positives, meaning that they were flagged for fraud even though they exhibited normal variations characteristic of a particular user. This exercise alone validates system strength and the capacity for speeding up fraud detection in reducing manual labor time and time-wastage and enhancing the system's overall strength. Most importantly, these 298 fraudulent cases were automatically caught, proving that the system can successfully detect fraud in realtime if applied during a live setting. Such findings prove that incorporating artificial intelligence in identifying frauds will greatly assist in pinpointing instances of fraud and diverting scarce resources to the appropriate families; this will also maintain the confidentiality and integrity of the public welfare systems.

5. Conclusion

This paper describes an advanced machine learning model used to enhance the capability to identify fraudulent activity and increase the data's effectiveness in child support systems. Through the use of analytics and AI to incorporate into the proposed system, the following drawbacks of current practices are effectively managed: While comparing all the models, the ensemble model trained using the Random Forest and XGBoostprovided better values of accuracy and precision along with acceptable recall for the three metrics used in the evaluation. This model can arrive at an almost accurate identification of fraud cases, making it relevant for use in real life in the public sector. In addition, the existing system has been augmented with SHAP values to give people viewpoints over what the designed machine learning models were doing, thus helping to build the confidence of the stakeholders. This establishes unequivocally that the system can be used in operation. The modelling of 10 000 anonymised cases shows that the analytical tool has an accuracy rate of 94% in identifying confirmed fraud cases.

5.1. Future Work

Despite these qualities, some aspects can be enhanced in the current framework to make intelligent fraud detection more robust, scalable, and trustworthy. Another area for further development should be the use of possibilities of blockchain technology for the firm’s transactions and changes logging. Through distributing and encrypting transaction records within blocks in the blockchain, the problem of data manipulation would be eradicated, and data transparency regarding auditing would be increased, particularly in legal and financial transactions. Another way is to create an online fraud alert system that immediately notices that the program has been detected as fraudulent. This would make it possible to peruse it on-site, eliminating the time available for the fraudulent act to be performed. The system will shift from being responsive to protective through streaming data processing and the real-time model inference. Finally, integrating data from other departments would be of great importance in improving the input data for the model and accuracy. Data from various sources like tax return filings, history of

child protective services, employment and employment-training centers, and/or banks could also give a broader perspective on each case. It must be noted that this web of data would enable the fraud detection model to detect other forms of fraudulent activities that might be complex or concealed, such that they might not be identifiable when working in isolation from the infrastructure. Together, these future improvements will bring significant change that would make the child support system proactive and intelligent, providing fair, accurate, and accountable outcomes at the same scale.

References

1. Wang, Q., & Liu, F. (2021). *Challenges in deploying AI-driven fraud detection in public welfare systems*. Public Policy and Technology Journal, 10(3), 116-130.
2. Nuthalapati, A. (2023). Smart fraud detection leveraging machine learning for credit card security. Educational Administration: Theory and Practice, 29(2), 433-443.
3. Gupta, P. (2023). Leveraging machine learning and artificial intelligence for fraud prevention. SSRG International Journal of Computer Science and Engineering, 10(5), 47-52.
4. Zhang, Q. (2023). Secure preschool education using machine learning and metaverse technologies. Applied Artificial Intelligence, 37(1), 2222496.
5. Devineni, S. K., Kathiriya, S., & Shende, A. (2023). Machine learning-powered anomaly detection: Enhancing data security and integrity. Journal of Artificial Intelligence & Cloud Computing. SRC/JAICC-198. DOI: doi. org/10.47363/JAICC/2023 (2), 184, 2-9.
6. Mandaloju, N., Srinivas, N., & Nadimpalli, S. V. (2020). Machine learning for ensuring data integrity in Salesforce applications. Artificial Intelligence and Machine Learning Review, 1(2), 9-21.
7. Bin Sulaiman, R., Schetinin, V., & Sant, P. (2022). Review of machine learning approach on credit card fraud detection. Human-Centric Intelligent Systems, 2(1), 55-68.
8. Hashemi, S. K., Mirtaheri, S. L., & Greco, S. (2022). Fraud detection in banking data by machine learning techniques. Ieee Access, 11, 3034-3043.
9. Singh, N., Lai, K. H., Vejvar, M., & Cheng, T. E. (2019). Data-driven auditing: A predictive modeling approach to fraud detection and classification. Journal of Corporate Accounting & Finance, 30(3), 64-82.
10. Sánchez-Aguayo, M., Urquiza-Aguiar, L., & Estrada-Jiménez, J. (2022). Predictive fraud analysis applying the fraud triangle theory through data mining techniques. Applied Sciences, 12(7), 3382.
11. Velasco, R. B., Carpanese, I., Interian, R., Paulo Neto, O. C., & Ribeiro, C. C. (2021). A decision support system for fraud detection in public procurement. International Transactions in Operational Research, 28(1), 27-47.
12. Matloob, I., & Khan, S. (2019, June). A framework for fraud detection in government-supported national healthcare programs. In 2019 11th international conference on electronics, computers and Artificial Intelligence (ECAI) (pp. 1-7). IEEE.
13. Das, R. A. H. U. L., Sirazy, M. R. M., Khan, R. S., & Rahman, S. H. A. R. I. F. U. R. (2023). A collaborative intelligence (ci) framework for fraud detection in federal relief programs. Applied Research in Artificial Intelligence and Cloud Computing, 6(9), 47-59.
14. Abdallah, A., Maarof, M. A., & Zainal, A. (2016). Fraud detection system: A survey. Journal of Network and Computer Applications, 68, 90-113.

15. Bonett, S., Lin, W., Sexton Topper, P., Wolfe, J., Golinkoff, J., Deshpande, A., ... & Bauermeister, J. (2024). Assessing and improving data integrity in web-based surveys: comparison of fraud detection systems in a COVID-19 study. JMIR formative research, 8, e47091.
16. Gee, S. (2014). Fraud and fraud detection: A data analytics approach. John Wiley & Sons.
17. Patel, K. (2023). Credit card analytics: a review of fraud detection and risk assessment techniques. International Journal of Computer Trends and Technology, 71(10), 69-79.
18. Burnson, C., Covington, S., Arvizo, B., Qiao, J., & Harris, E. (2021). The impact of Parents Anonymous on child safety and permanency. Children and Youth Services Review, 124, 105973.
19. Kushida, C. A., Nichols, D. A., Jadrnicek, R., Miller, R., Walsh, J. K., & Griffin, K. (2012). Strategies for de-identifying and anonymizing electronic health record data for multicenter research studies. Medical care, 50, S82-S101.
20. Lee, A., Elliott, M., Scourfield, J., Bedston, S., Broadhurst, K., Ford, D. V., & Griffiths, L. J. (2022). Data resource: Children receiving care and support and children in need, administrative records in Wales. International Journal of Population Data Science, 7(1), 1694.
21. Pirog, M. A., & Ziol-Guest, K. M. (2006). Child support enforcement: Programs and policies, impacts and questions. Journal of Policy Analysis and Management, 25(4), 943-990.