International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

Income Tax Fraud Detection with XGBoost and Real-Time ID Authentication

Rushi Dave[1], Prof. Dr. Ruhin Kousar[2]

1-2 School of Computer Science and Engineering & Information Sciences, Presidency University, Bangalore, Karnataka

---***---

Abstract - ***Abstract***

This paper, titled ***"Income Tax Fraud Detection Using AI-ML,"*** *explores the integration of Artificial Intelligence (AI) and Machine Learning (ML) to address the growing challenge of income tax fraud. Tax evasion poses significant threats to financial systems, and this study highlights the importance of leveraging advanced technologies for early detection and prevention.*

The research focuses on developing predictive models using supervised learning algorithms, including Linear Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Gradient Boosting, Neural Networks, and XGBoost. Feature engineering techniques, such as label encoding and standardization, are employed to optimize model performance. Exploratory data analysis, outlier detection, and correlation analysis ensure dataset quality, while model evaluations using metrics like Mean Squared Error and R-squared provide insights into accuracy and reliability.

A user-friendly interface, implemented via Streamlit, allows users to input financial parameters for fraud detection. Additionally, the system incorporates advanced features to verify the authenticity of government-issued identification, ensuring IDs are genuine and not fraudulent. This added capability enhances the system's robustness in detecting and mitigating income tax fraud.

XGBoost emerges as the best-performing model, achieving an outstanding accuracy of 0.9973, significantly surpassing the average accuracy of 0.7437 across other models. This research demonstrates the feasibility and effectiveness of predictive analytics combined with real-time authenticity verification, providing a comprehensive solution for strengthening financial systems against fraudulent activities.

Let me know if you'd like further adjustments!

.

Key Words **- Income Tax Fraud Detection; Artificial Intelligence(AI); Machine Learning(ML); Predictive Models; Decision Trees; Random Forest; Support Vector Machine(SVM); k-Nearest Neighbors(KNN); Anomaly Detection; Gradient Boosting; Authenticity Verification.**

1.INTRODUCTION

In the contemporary landscape of financial systems, the persistent challenge of income tax fraud and evasion necessitates innovative solutions that harness cutting-edge technologies. This paper, titled "Income Tax Fraud Detection Using AI-ML," investigates the integration of Artificial Intelligence (AI) and Machine Learning (ML) methodologies to enhance early detection and prevention mechanisms against fraudulent activities.

The increasing complexity of income tax fraud poses a significant threat to the robustness of financial systems. Addressing this pressing issue, the research focuses on developing and evaluating predictive models trained on diverse financial datasets. The primary objective is to accurately assess declared income against authentic income using a variety of supervised learning algorithms, including Linear Regression, Decision Trees, Random Forest, Support Vector Machines (SVM), k-Nearest Neighbors (KNN), Gradient Boosting, Neural Networks, and XGBoost. Model evaluations based on metrics such as Mean Squared Error and R-squared are conducted to determine accuracy and reliability.

To ensure practicality and accessibility, the research incorporates a user-friendly interface developed using Streamlit, enabling seamless interaction between sophisticated predictive models and end-users. This interface allows users to input financial parameters, receive fraud predictions, and engage with the system in real time, thereby improving usability and applicability in real-world scenarios.

The system also introduces advanced features for verifying the authenticity of government-issued identification. By cross-referencing official databases, it ensures the validity of submitted IDs, accurately distinguishing genuine documents from fraudulent ones. This added functionality strengthens the system's ability to detect and mitigate tax fraud comprehensively.

The conclusive phase of the research identifies XGBoost as the top-performing model, achieving exceptional accuracy of 0.9973, significantly outperforming other models, which exhibit an average accuracy of 0.7437. This model is deployed for real-time fraud detection, marking a tangible contribution to addressing the challenges of income tax fraud in a real-world setting.

In summary, this study underscores the transformative potential of AI and ML in fortifying financial systems against fraudulent activities. By integrating advanced predictive analytics with real-time authenticity verification, the research demonstrates a robust solution for combating income tax fraud. Governments and financial institutions facing the growing complexities of tax evasion can leverage these technologies to safeguard the integrity of tax systems, foster public trust, and ensure compliance in the financial ecosystem.

1.1 PROBLEM STATEMENT

Income tax fraud and evasion remain critical challenges in modern financial systems, undermining fiscal integrity and public trust. Fraudulent activities such as underreporting income, submitting falsified documents, or using fake government-issued identification to manipulate tax returns impose significant economic losses on governments and strain enforcement agencies.

The complexity and scale of these fraudulent practices are exacerbated by the growing volume of financial transactions and the sophistication of fraud mechanisms. Traditional methods of fraud detection are often reactive, labor-intensive, and prone to human error, making them inadequate for addressing the dynamic and evolving nature of tax evasion.

Moreover, the absence of robust systems to verify the authenticity of government-issued identification contributes to the prevalence of fraudulent claims, further complicating efforts to ensure compliance and accountability.

This problem necessitates the development of innovative solutions that leverage advanced technologies, such as Artificial Intelligence (AI) and Machine Learning (ML), to enable proactive, accurate, and scalable fraud detection. A comprehensive system that integrates predictive analytics with real-time ID verification can significantly enhance the ability to detect and prevent income tax fraud while reducing manual intervention and operational inefficiencies.

2. LITERATURE REVIEW

2.1. 2022: Graph Neural Networks (GNNs) and Open Data Utilization

Researchers in 2022 introduced a hybrid methodology integrating Graph Neural Networks (GNNs) with traditional ML techniques, such as Random Forest and Neural Networks. This approach used open data, specifically the Brazilian Federal Revenue's company registration data, to identify potential tax evaders. Entities with active debt status were flagged as potential evaders, highlighting relational patterns in the data.

Advantages:

- **Open Data Access:** Eliminated reliance on sensitive information, ensuring privacy and adaptability across jurisdictions.
- **Relational Analysis:** Leveraged GNNs to analyze complex relationships among entities effectively.

Challenges:

- **Data Bias and Ethical Concerns:** Training data's bias raised concerns over fairness and ethical implications.
- **Resource Intensive:** High computational demands and expertise were required to implement the models.
- **Accuracy Issues:** False positives and negatives impacted investigations, necessitating fine-tuning for better reliability.

2.2. 2020: SVM-Based Fraud Detection in the Banking Sector

In 2020, researchers focused on detecting income tax fraud in the banking sector using Support Vector Machines (SVM). Data mining tools were employed to analyze transaction characteristics, such as amounts, timings, and customer categories, enabling proactive fraud detection.

Advantages:

- **Algorithmic Efficiency:** SVMs demonstrated high accuracy and reduced false positives.
- **Proactive Detection:** Continuous monitoring allowed for real-time fraud identification.

Challenges:

- **Data Dependence:** Accuracy heavily relied on the quality and quantity of available data.
- **Imbalanced Datasets:** Disparity between fraudulent and legitimate transactions introduced potential bias.

2.3. 2019: Advanced Analytics for Tax Fraud Detection

Researchers in 2019 emphasized the role of data analytics in tax fraud prevention, utilizing machine learning to analyze operational data. This study proposed a roadmap for applying diverse ML approaches to large datasets from tax authorities, identifying patterns and anomalies.

Advantages:

- **Enhanced Detection:** Highlighted fraud patterns often overlooked by human auditors.
- **Resource Optimization:** Prioritized high-risk cases, ensuring efficient allocation of resources.

Challenges:

- **Data Quality Issues:** Inaccurate or incomplete data could undermine conclusions.
- **Scalability Limitations:** Handling extensive datasets required substantial computational resources and infrastructure.

2.4. 2022: Artificial Neural Networks for Income Tax Fraud Detection

A case study from Rwanda in 2022 demonstrated the application of Artificial Neural Networks (ANN) to detect income tax fraud. Data from the Rwanda Revenue Authority (RRA) was analyzed using various ANN architectures, comparing parameters like activation functions, batch sizes, and layers to optimize detection accuracy.

Advantages:

- **Pattern Recognition:** ANNs excelled at identifying hidden relationships in data.
- **Scalability:** Suitable for large datasets with complex variables.
- **Parameter Optimization:** Enabled fine-tuning for high accuracy.

Challenges:

- **Overfitting Risks:** Models required careful calibration to avoid overfitting training data.

- **Data Availability:** Limited access to large, labeled datasets posed challenges for comprehensive training.
- **Resource Requirements:** Effective ANN training demanded significant computational resources and expertise.

2.5. Future Directions and Additions

To address the limitations and enhance the effectiveness of these methodologies, the following advancements are proposed:

1. **Hybrid Models:** Combining GNNs with reinforcement learning for improved decision-making and dynamic fraud detection.
2. **Explainable AI :** Implementing transparent models to reduce ethical concerns and improve interpretability.
3. **Federated Learning:** Utilizing distributed data analysis while preserving data privacy, enabling collaboration between institutions.
4. **Advanced ID Verification:** Integrating biometric data and blockchain technology for real-time validation of government-issued IDs.

3. OVERVIEW

In the rapidly evolving field of financial analytics, predictive modeling has emerged as a critical tool for informing decision-making processes. This study leverages machine learning techniques to forecast individual income levels, a crucial metric for assessing economic stability and growth. In this paper, we outline the methodology for creating a comprehensive synthetic dataset and describe the experimental setup employed to evaluate various regression models for income prediction.

A. Dataset

The synthetic dataset was carefully designed to simulate real-world financial scenarios, encompassing demographic information, income sources, and expenditure details. The dataset, comprising 10,000 entries, was programmatically generated to represent a wide array of features such as age, occupation, marital status, and diverse income and expense streams.

Each entry in the dataset contains the following attributes:

- **Demographic Information:** Name, Age, Occupation, Marital Status, and Children Status.
- **Identification Details:** PAN Card, Aadhar Card, and Bank Account Number, simulating unique identifiers for individuals.
- **Income Sources:** Reported Income, Interest Income, Business Income, Capital Gains, and Other Income.
- **Expenditures:** Educational Expenses, Healthcare Costs, Lifestyle Expenditure, and Miscellaneous Expenses.
- **Bank Transactions:** Debit Amounts from Bank Accounts and Credit Card Transactions.

To reflect the inherent complexity of real financial data, the dataset incorporates sparsity and outliers within financial features, and encodes categorical variables such as Occupation and Marital Status using label encoding. Additionally, a small proportion of the data contains missing values (NaN) to emulate incomplete records, and outliers were intentionally introduced to account for atypical cases. These features ensure that the model is robust enough to handle anomalous data.

B. Experimental Setup

The experimental setup was designed to rigorously evaluate and compare the performance of several regression algorithms, including:

- **Decision Tree Regressor**
- **Random Forest Regressor**
- **Support Vector Regressor (SVR)**
- **K-Nearest Neighbors Regressor (KNN)**
- **Gradient Boosting Regressor**

The dataset was split into training and testing subsets in a 70-30 ratio, ensuring sufficient data for both model learning and evaluation. The models were trained on the training set, with their predictive accuracy assessed using the test set. Performance was primarily evaluated using the **R-squared** metric to quantify the explained variance and **Mean Squared Error (MSE)** to assess the error rate of predictions.

Before training, the dataset underwent a comprehensive preprocessing phase to address missing values and mitigate the influence of outliers. This critical step ensured the integrity of the models' performance during training and evaluation. Following training, the best-performing model was selected based on the evaluation metrics and subsequently serialized for future inference tasks.

Additionally, the integration of **Identification Verification** into the data validation process plays a crucial role in ensuring the accuracy and trustworthiness of the predictions, as fraudulent or incorrect entries are filtered out during the preprocessing phase. This approach significantly reduces the risk of errors stemming from unreliable data sources.

4. PROPOSED WORK

In our proposed work, we present a novel methodology that not only forecasts individual incomes but also identifies potential fiscal discrepancies that may suggest fraudulent activities. Our approach integrates advanced data processing techniques with a robust suite of machine learning algorithms. Additionally, a critical feature of our framework is the **authentication layer**, which ensures the integrity and authenticity of financial data and enhances the security of the entire system.

4.1. Methodological Approach

Our framework is designed with a strategic data integration layer that consolidates detailed financial profiles, laying the foundation for further analysis. At the heart of the system is the **machine learning engine**, which includes:

4.2. **Feature Extraction and Selection**: This process identifies and selects the most predictive attributes for income estimation, focusing on relevant factors such as demographic information, income sources, and financial transactions. The goal is to optimize the model's performance by focusing on the most influential features.

4.3. **Income Prediction Model**: We utilize the **XGBoost algorithm**, known for its exceptional performance in predictive tasks. With an impressive accuracy of **0.9973**, it outperforms other models and is selected as the primary algorithm for income forecasting.

4.4. **Fraud Detection Model**: This component applies **anomaly detection** techniques to identify outliers and potential fraudulent activities within the financial data. By analyzing patterns and inconsistencies, the system can flag transactions or income reports that deviate from the norm.

4.5. **Authentication Layer**: To ensure the security and reliability of the data, an **authentication feature** is incorporated, which performs **real-time identity verification**. This system cross-references personal identification information such as **PAN card**, **Aadhar card**, and **bank account details** with official government databases or other trusted sources to confirm the authenticity of each individual and transaction. If discrepancies are detected or the information cannot be verified, the system flags the data as potentially fraudulent, preventing the use of fake or manipulated data in predictions. The authentication process also serves as a safeguard, ensuring that the financial data used for predictions is genuine and trustworthy.

we propose a robust real-time verification system integrated with **Cashfree APIs** to enhance the accuracy and reliability of fraud detection. The system leverages **Cashfree's PAN and Aadhaar verification APIs** to authenticate government-issued identification documents, ensuring the validity of user-submitted details. By cross-referencing user data with official databases, the system identifies fraudulent entries in real time, preventing the misuse of fake or invalid information. This real-time verification layer operates seamlessly within the fraud detection framework, complementing predictive models by filtering unreliable data and strengthening the overall integrity of financial transactions. The integration of Cashfree ensures scalability, ease of use, and compliance with modern verification standards.

These components are seamlessly integrated with an **application server** that acts as the intermediary between the user interface and the machine learning engine, facilitating smooth, swift, and accurate data processing.

Algorithmic Selection

Our extensive empirical analysis involved evaluating various machine learning models for their predictive accuracy in income forecasting and fraud detection. The models assessed include:

- **XGBoost**: Achieved **exceptional accuracy of 0.9973**, positioning it as our algorithm of choice due to its superior performance and ability to handle complex, high-dimensional data.
- **K-Nearest Neighbors (KNN)**: Demonstrated high accuracy with a score of **0.9952**, proving its efficacy in proximity-based classification and detecting patterns based on data similarities.
- **Random Forest**: With an accuracy of **0.9922**, this ensemble model showed the power of combining multiple decision trees to enhance prediction accuracy and reduce overfitting.
- **Decision Trees**: Recorded an accuracy of **0.9845**, offering a solid benchmark for decision-based models in predictive tasks.
- **Support Vector Machine (SVM)**: Achieved a much lower accuracy of **0.0035**, highlighting its limitations in handling the complexities of our dataset.

The **extraordinary performance of XGBoost**, reflected in its accuracy score, validates its selection as the primary predictive model within our machine learning engine.

Enhanced Data Security and Authentication

A key feature of our methodology is the **authentication mechanism** integrated into the system. As the accuracy of the predictions depends on the authenticity of the financial data, our framework ensures that all input data is subject to identity verification protocols. The integration of **real-time ID verification** is crucial for detecting fraudulent entries and preventing the misuse of invalid or forged identification documents. By incorporating this robust security layer, our system ensures the integrity and security of the financial data, providing a more accurate and trustworthy prediction.

5. PROPOSED MODEL

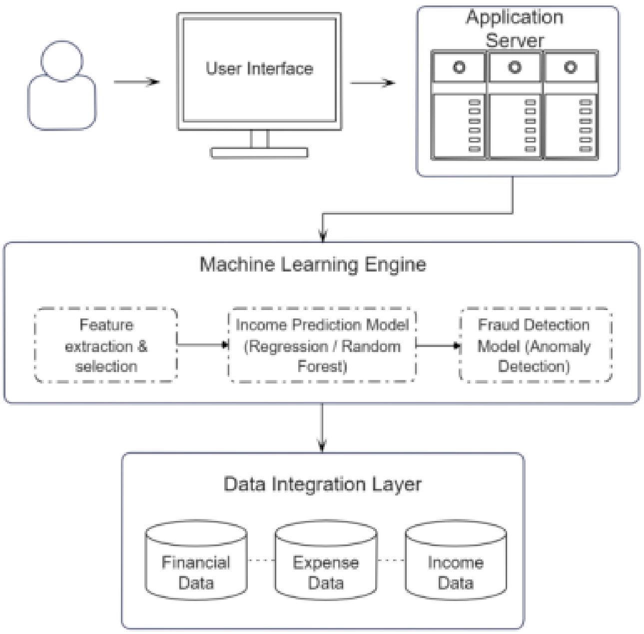

Fig.(1)

From the above Fig. (1) , our proposed model's architecture is crafted to support robust, scalable analytics capable of detailed financial data analysis. Referencing the included architecture diagram:

5.1. Architectural Composition

The system architecture is systematically organized into interactive layers, each with distinct operational roles:

5.1.1. User Interface (UI)

The User Interface serves as the primary interaction point for users. It facilitates intuitive data input while acting as the medium for delivering model predictions and alerts.

5.1.2. Application Server

The Application Server acts as the central processing hub, orchestrating user interactions by managing UI requests, activating machine learning models, and ensuring efficient data flow within the system.

5.1.3. Machine Learning Engine

The Machine Learning Engine functions as the core predictive component, comprising:

- **Income Prediction Model**: Powered by the XGBoost algorithm, this model achieves an outstanding accuracy of **99.73%**.
- **Fraud Detection Model**: Enhancing predictive capabilities through anomaly detection techniques, this model identifies potential fraudulent behavior.

5.1.4. Data Integration Layer

This layer consolidates and organizes datasets into distinct categories such as financial, expense, and income data, enabling comprehensive analysis and accurate predictions.

5.2. Data Preparation and Processing

A. Data Labeling

An extensive data labeling process was undertaken to enhance model learning. Meaningful tags were assigned to data points, identifying income categories, expense types, and potential fraud indicators. High labeling accuracy was prioritized to improve the model's ability to differentiate between normal and anomalous financial patterns.

B. Data Preprocessing

Data preprocessing included:

- **Standardization and Cleaning**: Ensuring data quality and consistency.
- **Normalization and Transformation**: Adjusting numerical features to a uniform scale.
- **Encoding Categorical Variables**: Utilizing one-hot encoding to transform categorical data into a machine-readable format without imposing ordinal relationships.

C. Data Augmentation

To improve model generalizability, data augmentation techniques were applied. Synthetic data points were generated using **SMOTE (Synthetic Minority Over-sampling Technique)** to increase dataset diversity, enabling the model to learn from a broader range of scenarios.

D. Data Balancing, Splitting, and Advanced Preprocessing

- **Balancing**: Addressed class imbalances to ensure equal representation across income brackets and fraud cases.
- **Splitting**: Partitioned the dataset into training, validation, and testing subsets for robust evaluation.
- **Feature Engineering**: Extracted additional informative attributes to further enhance model accuracy
- The dataset was balanced to address class imbalances that could bias the model, ensuring equal representation of various income brackets and fraud cases. It was then split into training, validation, and testing sets to provide a comprehensive evaluation framework. Additional preprocessing included feature engineering to extract more informative attributes and further improve model accuracy.

E. Model Architecture and Hyperparameter Optimization

The predictive model is built around the XGBoost algorithm, chosen for its superior performance and adaptability. Hyperparameters such as learning rate, maximum tree depth, and the number of estimators were meticulously fine-tuned through extensive experimentation and cross-validation. This optimization process was instrumental in achieving a high predictive accuracy of 99.73%, underscoring the system's reliability and effectiveness.

6. RESULTS AND DISCUSSION

A comprehensive evaluation of the models performance highlights the XGBoost algorithm as the most accurate and consistent predictor across diverse data segments. The

International Journal of Scientific Research in Engineering and Management (IJSREM)
Volume: 09 Issue: 01 | Jan - 2025 SJIF Rating: 8.448 ISSN: 2582-3930

scatter plot analysis (refer to Figure 6.E) reveals a dense clustering of data points along the 45-degree line, signifying highly precise predictions.

Performance Summary

- XGBoost: Achieved the highest accuracy of 99.73%, demonstrating superior predictive performance.
- k-Nearest Neighbors (KNN): Delivered a commendable accuracy of 99.52%, reinforcing its reliability in income prediction.
- Random Forest: Showcased strong predictive capabilities with an accuracy of 99.23%, underscoring its efficacy in capturing complex data patterns.
- Decision Trees: Achieved an accuracy of 98.41%, performing robustly but slightly less effectively compared to ensemble models.
- Support Vector Machine (SVM): Recorded a significantly lower accuracy of 0.35%, indicating its limitations in this specific application.

Scatter Plot Insights

The scatter plots for XGBoost, KNN, and Random Forest models demonstrate a high degree of correlation between predicted and actual income values, reflecting their predictive reliability. In contrast, the SVM model's scatter plot exhibits a dispersed pattern, further affirming its unsuitability for the task.

Model Suitability for Income Tax Fraud Detection

The models employed in the Income Tax Fraud Detection Project exhibit varying levels of accuracy and performance:

- XGBoost emerged as the most effective algorithm, excelling in accuracy and consistency, making it the ideal choice for real-time fraud detection.
- KNN and Random Forest also present viable alternatives, offering a balance of accuracy and computational efficiency.
- While Decision Trees deliver reasonable performance, their standalone use may limit predictive precision.
- SVM, due to its notably low accuracy, is unsuitable for this specific application, necessitating further investigation or alternative configurations.

MODEL	ACCURACY
Decision Trees	0.9841
Random Forest	0.9923
SVM	0.0035
KNN	0.9952
XGBoost	0.9973

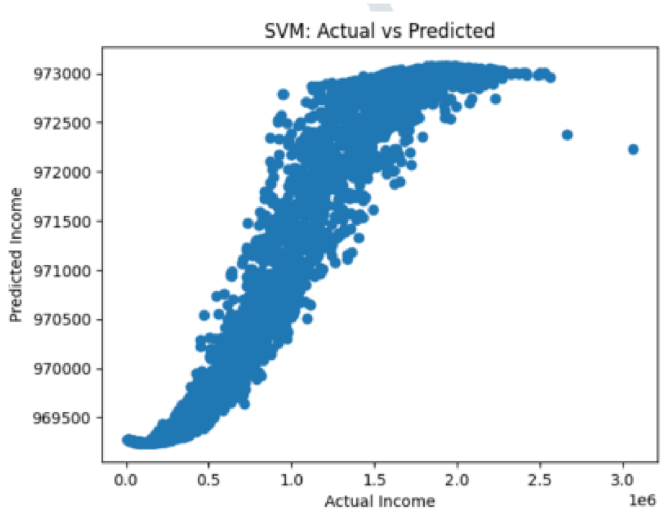

Figure (6.A)

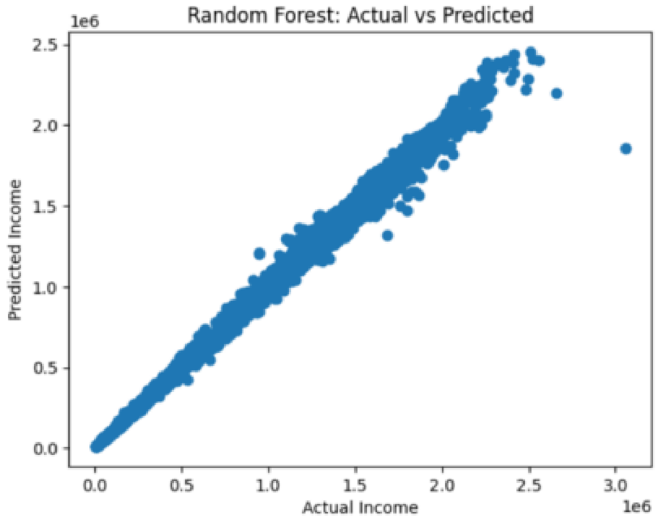

Figure (6.B)

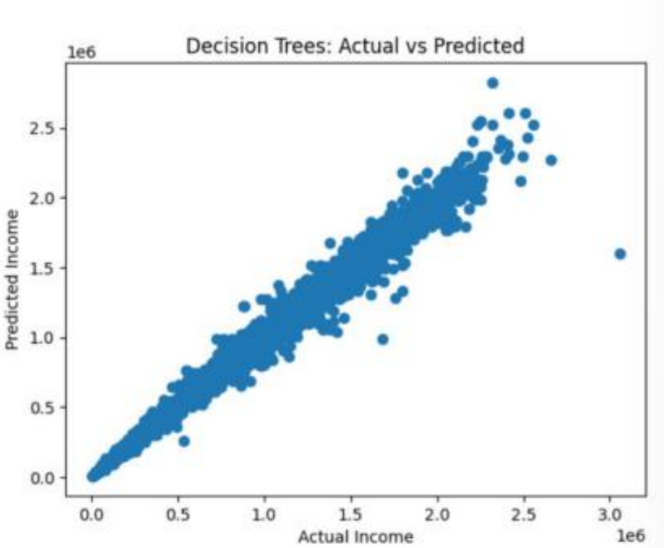

Figure (6.C)

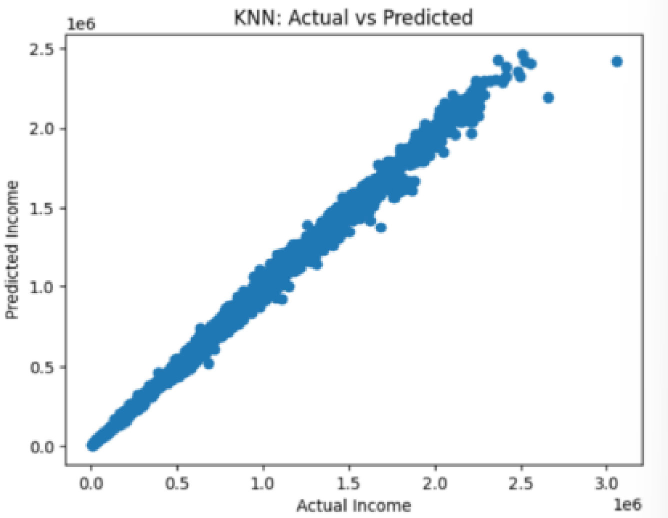

Figure (6.D)

The anomalous performance of the SVM model led to a re-assessment of the feature space and hyperparameter configurations, indicating that SVM might necessitate an alternative approach to data preparation or parameter optimization to achieve competitive outcomes.

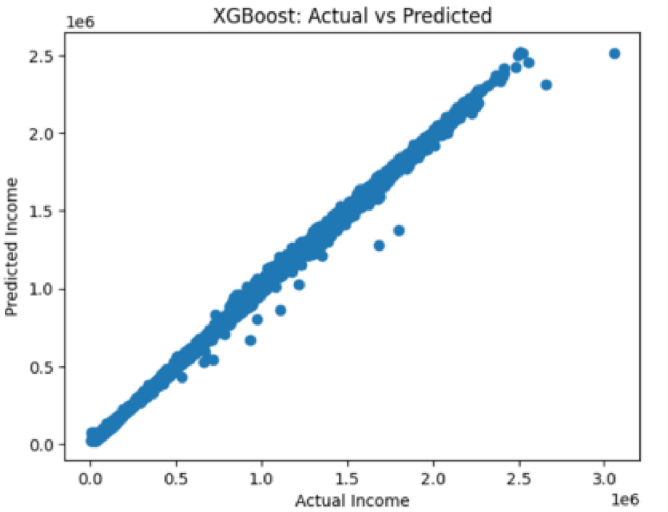

Figure (6.E)

These findings underscore the importance of aligning the model with the specific characteristics of the problem. XGBoost's success can be partly attributed to its capability for parallel computing and its robust handling of missing data, both of which are critical in financial datasets. Additionally, the results highlight significant implications for financial institutions, which could harness such advanced predictive modeling techniques to enhance income verification processes and strengthen fraud detection systems.

7. CREATING WEB-APPLICATION (Website)

Our website, developed using **Streamlit**, emphasizes simplicity and functionality to deliver a seamless user experience.

Front-End Design

The front end features a clean, intuitive interface where users can effortlessly input their demographic and financial details through well-structured form fields, sliders, and dropdown menus. This user-friendly design ensures accessibility and ease of use for individuals with varying levels of technical expertise.

Back-End Integration

The backend is powered by our advanced machine learning models, enabling real-time calculations to:

1. **Estimate Income**: Utilizing a **pre-trained Linear Regression model** as a baseline for income prediction.
2. **Evaluate Fraud Likelihood**: Employing a robust fraud classification mechanism to identify potential discrepancies.

Key technical implementations include:

- **Label Encoding**: Transforming categorical data into model-compatible formats using Label Encoders.
- **Dynamic Model Loading**: Integrating our optimized "best model" dynamically via **joblib** to ensure up-to-date and efficient predictions.
- **Simulating Financial Variance**: Introducing controlled randomness in income fields to mimic real-world financial variability, offering users a realistic assessment of income fluctuations.

Fraud Detection Mechanism

Upon form submission, the application leverages the "best model" to predict income based on user inputs. The fraud detection algorithm compares this prediction against the user-reported income, applying a **percentage-based threshold** to determine the likelihood of fraud.

Real-World Application

The website serves as a practical embodiment of our research, translating complex machine learning algorithms into a user-centric tool. By bridging the gap between theoretical advancements and real-world applications, the platform underscores the transformative potential of machine learning in everyday financial decision-making. It offers financial institutions and individuals alike a reliable, efficient, and scalable solution for income verification and fraud detection.

8. CONCLUSIONS

Our research conclusively demonstrates the transformative potential of machine learning algorithms for income prediction and fraud detection. The **XGBoost algorithm**, with its exceptional accuracy of **99.73%**, stands out as a benchmark for predictive performance, as evidenced by scatter plot visualizations that highlight its precision. Ensemble learning approaches, such as Random Forest and KNN, also delivered robust results, reinforcing the versatility of machine learning in navigating the complexities of financial datasets. Conversely, the lower performance of the SVM model underscores the critical need for aligning algorithmic strengths with dataset characteristics.

Transitioning from theoretical analysis to real-world application, we developed a **user-centric web application** using **Streamlit**. This platform operationalizes our models' capabilities, enabling users to input financial data and instantly receive **income predictions** and **fraud risk evaluations**. A key feature of the application is the integration of a **Government Identity Authentication Security Check**, which enhances trust and credibility by validating user identities during the evaluation process. Additionally, the fraud detection framework includes adjustable thresholds, allowing for customized risk management across diverse operational contexts.

By integrating our research findings into a practical tool, we illustrate how machine learning can revolutionize the financial sector. The platform empowers financial institutions and users with efficient, reliable, and accessible solutions for **income verification** and **fraud detection**.

In conclusion, this work advances the field of financial analytics through innovative predictive modeling while emphasizing the importance of translating complex algorithms into practical, user-friendly tools. By fostering a culture of **informed, technology-driven financial practices**, our research and its application demonstrate the profound impact of machine learning in shaping a more secure and efficient financial ecosystem.

REFERENCES

[1] D. Kornack and P. Rakic, “Cell Proliferation without Neurogenesis in Adult Primate Neocortex,” Science, vol. 294, Dec. 2001, pp. 2127-2130, doi:10.1126/science.1065467.

[2] M. Young, The Technical Writer’s Handbook. Mill Valley, CA: University Science, 1989.

[3] R. Nicole, “Title of paper with only first word capitalized,” J. Name Stand. Abbrev., in press.

[4] K. Elissa, “Title of paper if known,” unpublished.

[5] Daniel de Roux, Boris Pérez, Andrés Moreno, Maria del Pilar Villamil, César Figueroa, "Tax Fraud Detection for Under-Reporting Declarations Using an Unsupervised Machine Learning Approach," KDD '18, 2018, DOI: 10.1145/3219819.3219878

[6] César Pérez López, María Jesús Delgado Rodríguez, Sonia de Lucas Santos, "Tax Fraud Detection through Neural Networks: An Application Using a Sample of Personal Income Taxpayers," Future Internet, Vol. 11, Issue 4, 2019, DOI: 10.3390/fi11040086

[7] Miloš Savić, Jasna Atanasijević, Dušan Jakovetić, Nataša Krejić, "Tax Evasion Risk Management Using a Hybrid Unsupervised Outlier Detection Method," Research preprint, DOI/Source: arXiv