JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

The Secure Framework to Develop Income Tax Fraud Detection using AI-ML Techniques

Prerana Bharati [1], Vishal D Hanchate [2], Manoj D [3], Sujit Gani [4], Shakkeera L [5]

[1,2,3,4,5] School of Computer Science and Engineering & Information Sciences, Presidency University, Bangalore, Karnataka

Abstract— The paper, titled "Income Tax Fraud Detection Using AI-ML," investigates the integration of Artificial Intelligence (AI) and Machine Learning (ML) to identify income tax fraud. With the rising challenge of tax evasion, advanced technologies are crucial for early detection. The study focuses on developing predictive models using supervised learning algorithms like Linear Regression, Decision Trees, Random Forest, SVM, KNN, Gradient Boosting, and Neural Networks. Feature engineering techniques, including label encoding and standardization, optimize model performance.

The report includes exploratory data analysis, outlier detection, and correlation analysis to ensure dataset quality. Model evaluations, using metrics like Mean Squared Error and R-squared, provide insights into model accuracy. The Income Tax Fraud Detection system's user interface is implemented through Streamlit, enabling users to input financial parameters for predictions. The report concludes by identifying the best-performing model, deployed for real-time fraud detection.

This research strengthens financial systems against fraud using AI and ML, providing valuable insights into the feasibility and effectiveness of predictive analytics for income tax fraud detection. Notably, XGBoost demonstrates exceptional accuracy, achieving 0.9973, surpassing all other models, which have a combined average accuracy of 0.7437.

Keywords – 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.

I. INTRODUCTION

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

The increasing intricacy of income tax fraud presents a significant menace to the robustness of financial systems. In response to this pressing issue, the research concentrates on creating and appraising predictive models that are trained on a range of financial datasets. The primary goal is to precisely evaluate declared income in comparison to authentic income, employing a diverse set of supervised learning algorithms such as Linear Regression, Decision Trees, Random Forest, Support Vector Machine (SVM), k-Nearest Neighbors (KNN), Gradient Boosting, and Neural Networks.

In order to enhance the effectiveness of these models, the study integrates advanced feature engineering techniques, including label encoding and standardization. This optimization endeavor is designed to elevate the precision and dependability of predictions by refining the representation of financial data.

Prior to delving into predictive modeling, the report underscores the importance of preliminary steps such as exploratory data analysis (EDA), outlier detection, and correlation analysis. These measures are crucial for ensuring the dataset's quality, establishing a robust foundation for subsequent modeling endeavors. The research underscores that the success of any predictive model relies heavily on the integrity and relevance of the underlying data.

The evaluation of models involves the use of standard metrics such as Mean Squared Error (MSE) and R-squared on a dedicated test set. This meticulous evaluation process yields valuable insights into the performance of each algorithm, facilitating the identification of the most effective model. These insights play a pivotal role in making well-informed decisions regarding model deployment and real-world applicability.

To ensure the practicality and accessibility of the research findings, the report introduces a user-friendly interface for the Income Tax Fraud Detection system. Developed through Streamlit, this interface allows users to input diverse financial parameters for predictions, acting as a seamless link between sophisticated predictive models and end-users. This approach enhances the technology's usability and applicability in real-world situations.

The conclusive section of the report highlights the top-performing model identified through evaluations, subsequently deploying it for real-time income tax fraud detection. This deployment in a real-world setting marks the culmination of the research, offering a tangible solution to the urgent issue of income tax fraud.

In summary, this research significantly contributes to the ongoing efforts to strengthen financial systems against fraudulent activities. Through the utilization of AI and ML capabilities, the study illustrates the viability and efficiency of predictive analytics in identifying and averting income tax fraud. The results emphasize the necessity of staying ahead in technological advancements to tackle the evolving challenges in the financial landscape. Governments and financial institutions, contending with the intricacies of tax evasion, recognize the integration of AI and ML as a powerful tool to uphold the integrity of income tax systems and maintain trust in financial transactions.

II. Literature Review

In 2022,[1] researchers employed a diverse set of machine learning methodologies, encompassing Random Forest, Neural Networks, and Graphs, to identify potential tax evaders through tax evasion detection. The methodology involved the integration of Graph Neural Networks (GNNs) with traditional machine learning techniques, utilizing open data sources such as the Brazilian Federal Revenue's company registration data. The active debt status in the state served as a crucial criterion for labeling entities as potential tax evaders.

Advantages of this approach included the utilization of open data, eliminating the need for sensitive information, and the development of a generic solution applicable to various regions or jurisdictions. The incorporation of Graph Neural Networks (GNNs) facilitated effective relational data analysis.

However, certain disadvantages were acknowledged. The reliance on potentially biased or incomplete training data raised concerns about biased predictions and ethical implications. The system's tendency to generate false positives and negatives was noted, impacting investigations and revenue outcomes. Additionally, the complexity of the employed models demanded substantial resources and expertise.

This innovative approach in 2022 aimed to leverage the power of artificial intelligence and open data for tax evasion identification, showcasing both its promising advantages and inherent challenges.

In 2020,[2] researchers focused on combating income tax fraud within the banking sector by introducing data mining tools, specifically relying on supervised machine learning through Support Vector Machines (SVM). The methodology involved the application of the SVM algorithm to data collected by banks, with the primary objective of identifying customers engaged in fraudulent transactions. This analysis considered various transaction parameters, including amounts, timings, customer categories, and money transmission rates. The insights obtained from the SVM algorithm were then leveraged to proactively combat banking fraud.

Notable advantages of this approach included the efficiency of fraud detection achieved through the use of SVM, a potent algorithm in the context of tax fraud detection within the banking sector. The employment of data mining tools ensured accurate results, thereby reducing the occurrence of false positives. The methodology also enabled timely detection by continuously monitoring transactions for proactive fraud prevention.

However, certain disadvantages were acknowledged in the implementation of this method. The reliance on the quality and quantity of data posed a potential challenge, as poor data could lead to inaccurate results. Additionally, the presence of imbalanced data, with variations between fraudulent and legitimate transactions, raised concerns about potential bias in the detection process.

This 2020 initiative showcased a focused and technologically advanced strategy for income tax fraud detection within the banking sector, emphasizing both the strengths and limitations of employing AI and data mining tools in the pursuit of financial security.

In 2019,[3] researchers delved into the realm of tax fraud detection by exploring the applications of machine learning and advanced analytics. The emphasis was on the pivotal role of data analytics in preventing fraud within taxation. The methodological foundation of the study aimed to provide insights into how tax authorities could leverage operational data, laying the groundwork for the effective application of advanced data analytics. Future research directions were outlined, focusing on the application of diverse machine learning approaches to extensive real data from tax authorities, with the goal of identifying optimal methods for detecting specific types of fraud.

Advantages of this approach included the enhanced detection accuracy achieved through the identification of patterns and anomalies that human auditors might overlook. The methodology also demonstrated efficient resource allocation, prioritizing high-risk cases and optimizing the use of resources for maximum impact.

Despite these advantages, the approach acknowledged certain disadvantages. The reliance on data quality emerged as a critical factor, as inaccurate or incomplete data could lead to flawed conclusions. Scalability challenges were noted, particularly in dealing with large volumes of data that demanded costly infrastructure. Moreover, the implementation costs associated with developing and maintaining machine learning systems were acknowledged as potentially expensive with uncertain returns.

This 2019 initiative showcased a forward-looking exploration of machine learning and advanced analytics in the context of tax fraud detection, highlighting the potential benefits and challenges inherent in leveraging cutting-edge technologies for enhanced fraud prevention within taxation.

In 2019,[4] a focused investigation into tax fraud detection was conducted through the application of advanced predictive tools, particularly neural networks, within a supervised learning framework. The study aimed to identify tax fraud in personal income tax returns filed in Spain, incorporating data mining techniques for comprehensive data processing tailored specifically for fraud analysis in the context of tax evasion detection. The proposed methodology demonstrated generalizability, intending to quantify the likelihood of any taxpayer committing various types of tax fraud, extending beyond the specific focus on personal income tax returns.

Advantages of this methodology encompassed the identification of individuals prone to tax avoidance, the elimination of atypical values and multicollinearity issues, the normalization of variables within the model, and an enhancement of data confidentiality.

However, the approach acknowledged certain disadvantages. Effective training of the neural network demanded extensive data, raising concerns about the availability of ample labeled tax return data for tax authorities. Diverse data sources were deemed necessary to prevent overfitting and enhance the overall performance of the model.

This 2019 initiative showcased a targeted application of neural networks in the realm of tax fraud detection, emphasizing both its potential advantages and challenges, and offering insights into the intricate landscape of leveraging advanced predictive tools for enhanced fraud analysis in the context of personal income tax filings in Spain.

In 2016,[5] a comprehensive framework for detecting fraudulent activities within the tax collection system in EDO state was developed, employing investigative data mining techniques. The methodology centered on the utilization of neural networks, business classification, and decision trees as key tools for analyzing taxpayer data. This analysis involved classifying data into tiers and identifying patterns indicative of fraudulent activities, such as cash suppression and diversion. The application of investigative data mining methods aimed to enhance the tax collection system's capability to identify and combat various forms of fraudulent behavior among taxpayers.

Advantages of this framework included the utilization of data mining techniques for accurate fraud detection, the incorporation of the BIRGENT software agent with reasoning abilities for intelligent decision-making, and the analysis of transactions to prevent fraud, specifically addressing issues like cash suppression and diversion.

However, the framework acknowledged certain disadvantages. Concerns related to data privacy were highlighted, emphasizing the need to handle taxpayers' data with caution and comply with data protection regulations. The effectiveness of the framework was contingent on the accuracy and quality of the data, and the potential for inaccurate or incomplete data leading to incorrect fraud detection results was recognized.

This 2016 initiative showcased a tailored framework for detecting fraudulent activities in the EDO state tax collection system, emphasizing the integration of investigative data mining techniques and the challenges associated with maintaining data privacy and accuracy in the pursuit of more effective and intelligent fraud detection measures.

In 2022,[6] a focused investigation on fraud detection in the context of income tax was conducted through the implementation of Artificial Neural Networks (ANN), as outlined in the research paper titled "Fraud Detection Using Neural Networks: A Case Study of Income Tax." The methodology involved the utilization of income tax-related data from the Rwanda Revenue Authority (RRA), applying various neural network algorithms to identify factors and effectively detect instances of income tax fraud within Rwanda. A comparative analysis of different parameters, including activation functions, batch size, epochs, and layers, was undertaken to determine the optimal combination resulting in high accuracy for the detection of income tax fraud.

Advantages of this methodology included the ability to detect hidden relationships in data, versatile pattern recognition during training, and suitability for handling large datasets with discrete variables.

Nevertheless, the approach acknowledged certain disadvantages. The risk of overfitting training data was highlighted, emphasizing the importance of balancing model performance. The effective training of neural networks demanded extensive data, posing potential challenges related to the availability of large labeled datasets for tax authorities.

This 2022 initiative showcased a case study employing neural networks for income tax fraud detection, emphasizing both the advantages and challenges associated with utilizing ANN in a specific tax context. The methodology aimed to contribute insights into optimizing neural network parameters for accurate and effective fraud detection within the realm of income tax in Rwanda.

III. OVERVIEW

In the burgeoning field of financial analytics, predictive modeling has become a cornerstone for decision-making processes. Our research leverages machine learning techniques to forecast individual income levels, a metric pivotal for assessing economic stability and growth. In this paper, we delineate the methodology behind creating a comprehensive synthetic dataset and the experimental setup employed for the evaluation of various regression models.

A. *Dataset*

The synthetic dataset was meticulously engineered to simulate real-world financial scenarios, encapsulating demographic information, income sources, and expenditure details of individuals. The dataset, encompassing 10,000 entries, was generated programmatically to include a broad spectrum of features such as Age, Occupation, Marital Status, and various income and expense streams.

Each entry in the dataset consists of the following attributes:

Demographic Details: Name, Age, Occupation, Marital Status, and Children status.

Identification Numbers: PAN Card, Aadhar Card, and Bank Account Number to simulate unique identifiers for each individual.

Income Streams: Reported Income, Interest Income, Business Income, Capital Gains, and Other Income.

Expenditures: Educational Expenses, Healthcare Costs, Lifestyle Expenditure, and Other Expenses.

Bank Transactions: Bank Debited and Credit Card Debited amounts.

To introduce realistic complexity, the dataset includes sparsity and outliers within financial features and encodes categorical variables such as Occupation and Marital Status. A label encoding technique was applied to transform these categorical variables into a machine-readable format. Moreover, a small percentage of the data was deliberately assigned NaN values to mimic missing information, and outliers were introduced to represent atypical cases, thus ensuring the model's robustness in handling anomalous data.

B. *Experimental Setup*

Our experimental setup is constructed to rigorously assess and compare the performance of several regression algorithms. The models considered include Decision Tree Regressor, Random Forest Regressor, Support Vector Regressor, K-Nearest Neighbors Regressor, and Gradient Boosting Regressor.

The dataset was partitioned into training and testing sets with a 70-30 split, ensuring a substantial amount of data for model learning while reserving an adequate portion for evaluation purposes. The models were trained on the training set, and their predictive accuracy was gauged using the test set. Model performance was primarily evaluated using the R-squared metric and Mean Squared Error to quantify the prediction accuracy and error rate, respectively.

Prior to model training, the dataset underwent a rigorous preprocessing phase to handle missing values and reduce the influence of outliers. This step was crucial to maintain the integrity of the predictive models' performance. Post-training, the best-performing model was identified based on the evaluation metrics and was subsequently serialized to disk for future inference tasks.

IV. PROPOSED WORK

In our proposed work, we delineate a novel methodology for forecasting individual incomes and identifying fiscal discrepancies, which may suggest fraudulent activities. Our approach integrates sophisticated data processing with an advanced suite of machine learning algorithms.

Methodological Approach
Our framework comprises a strategic data integration layer, which consolidates comprehensive financial profiles, serving as the groundwork for analysis. At the core of our system lies the machine learning engine, which includes:

A process for feature extraction and selection, determining the most predictive attributes for income estimation.
An income prediction model, utilizing the XGBoost algorithm, which delivered an impressive accuracy of 0.9973, outperforming other models.

A fraud detection model employing anomaly detection to identify outliers and potential fraud within the financial data.
These components are cohesively integrated with an application server that mediates between the user interface and the machine learning engine, ensuring swift and accurate data processing.

Algorithmic Selection
Through extensive empirical analysis, we evaluated various models, obtaining the following accuracies:

XGBoost: Exceptional accuracy of 0.9973, leading the selection as our algorithm of choice for its superior predictive power.

KNN: Exhibited high accuracy, achieving a 0.9952 score, indicative of its robustness in proximity-based classification.
Random Forest: With an accuracy of 0.9922, this model showcased the strength of ensemble strategies in predictive accuracy.

Decision Trees: Recorded an accuracy of 0.9845, serving as a benchmark for decision-based algorithms.

SVM: Achieved a lower accuracy of 0.0035, demonstrating limitations in the context of our dataset's complexity.

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

V. PROPOSED MODEL

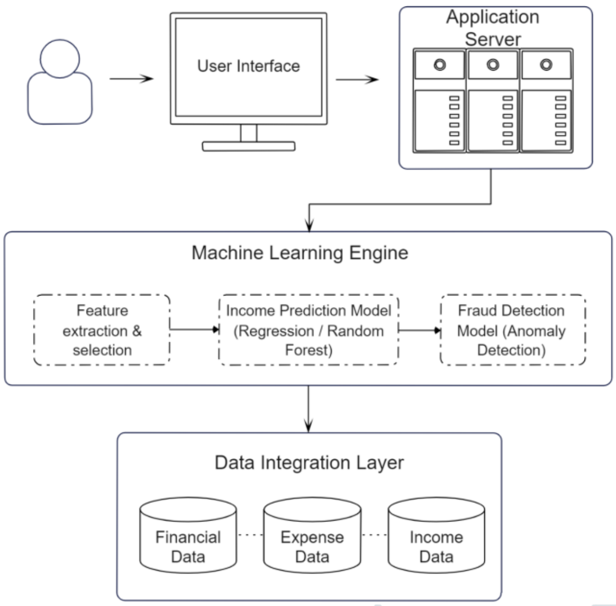

Fig.(1)

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:

Architectural Composition
The system's architecture is systematically divided into interactive layers, each with defined operational roles:

User Interface (UI): The UI is the gateway for users, allowing easy data input and acting as the medium for receiving model predictions and alerts.

Application Server: This central component processes UI requests, activates the machine learning models, and manages the data flow within the engine.

Machine Learning Engine: As the predictive powerhouse, it is composed of:

An income prediction model, founded on the XGBoost algorithm, which boasts a stellar accuracy of 0.9973.

A fraud detection model, which complements the predictive model by applying anomaly detection techniques for fraud potential.

Data Integration Layer: This repository houses the data, categorized into financial, expense, and income data sets, enabling comprehensive analysis for precise prediction.

A. Data Labelling

For our model to understand and learn from the dataset, an extensive data labelling process was undertaken. This step involved assigning meaningful tags to various data points, which serve as identifiers for income categories, expense types, and potential fraud indicators. Labelling accuracy is critical as it directly impacts the model's ability to discern between normal and anomalous financial behaviour.

B. Data Preprocessing

Preprocessing involved standardizing and cleaning the dataset to improve model performance. Techniques such as normalization and transformation were applied to numerical features to bring them onto a similar scale, while categorical variables were encoded using one-hot encoding to convert them into a machine-readable format without introducing ordinal relationships.

C. Data Augmentation

To bolster the dataset and enhance the model's generalizability, data augmentation techniques were employed. These techniques included generating synthetic data points using methods like SMOTE (Synthetic Minority Over-sampling Technique) to introduce variability, thereby enabling the model to learn from a more diverse set of scenarios.

D. Data Balancing, Splitting & Further Preprocessing

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 Hyperparameters

Our model's architecture is built around the XGBoost algorithm, selected for its exceptional performance and versatility. We fine-tuned hyperparameters such as learning rate, max depth of trees, and the number of estimators through a series of experiments and cross-validation to find the optimal configuration that maximizes predictive accuracy. Specifically, the high accuracy of 0.9973 achieved with XGBoost can be attributed to such meticulous hyperparameter optimization.

VI. RESULTS AND DISCUSSION

In-depth analysis of the models' performance reveals that the XGBoost algorithm not only achieved the highest accuracy but also maintained consistency across various segments of the data. The scatter plot analysis (Figure 6.E) highlighted a dense clustering of points along the 45-degree line, indicative of precise predictions. The KNN and Random Forest models also exhibited strong predictive capabilities, with their scatter plots showing a high degree of correlation between predicted and actual incomes.

The predictive models employed in the Income Tax Fraud Detection project exhibit varying levels of accuracy. Decision Trees and Random Forest demonstrate robust performance with accuracies of 0.9841 and 0.9923, respectively, showcasing their efficacy in predicting individuals' actual income based on reported financial information. Conversely, Support Vector Machine (SVM) yields a notably lower accuracy of 0.0035, indicating potential challenges in its suitability for the task at hand. Meanwhile, k-Nearest Neighbors (KNN) and XGBoost exhibit commendable accuracy scores of 0.9952 and 0.9973, with XGBoost standing out as the highest-performing model among the evaluated algorithms. These accuracy metrics serve as valuable insights into the strengths and limitations of each model, guiding the selection of the most effective algorithm for real-time income tax fraud detection.

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

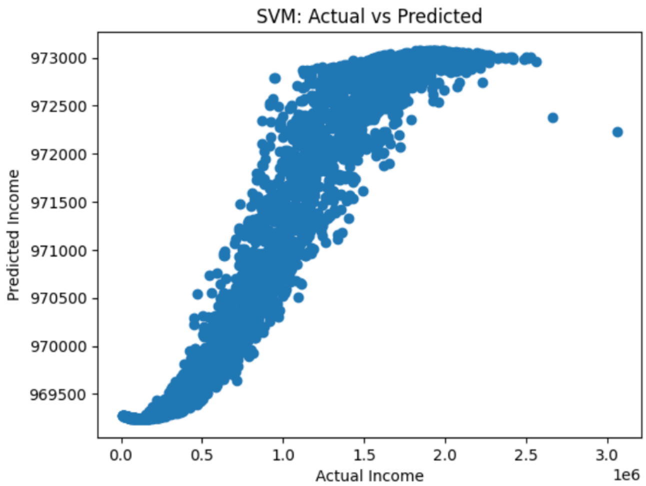

Figure (6.A)

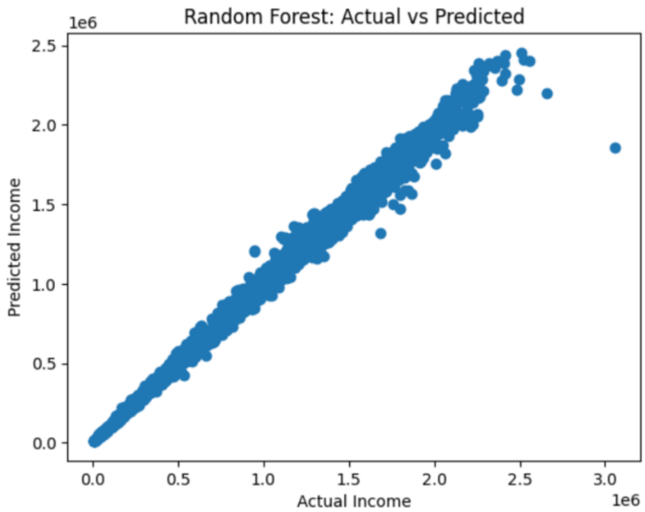

Figure (6.B)

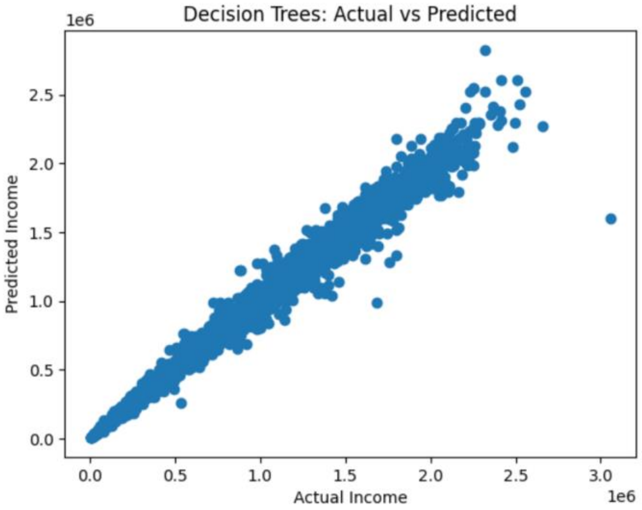

Figure (6.C)

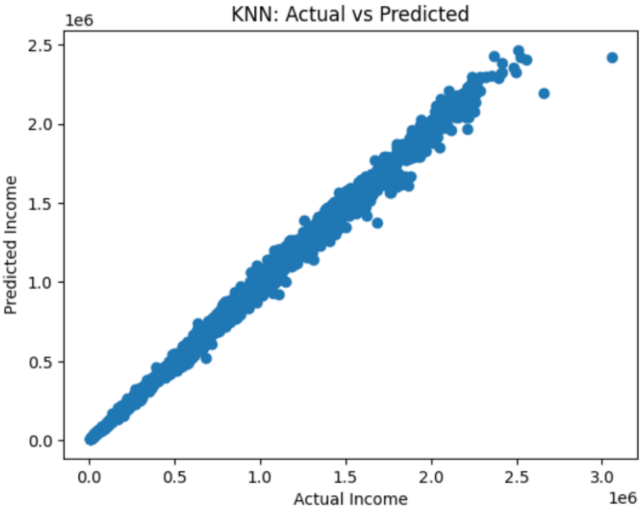

Figure (6.D)

The outlier performance in the SVM model prompted a re-evaluation of the feature space and hyperparameter settings, suggesting that SVM may require a different approach to data preparation or parameter tuning to achieve competitive results.

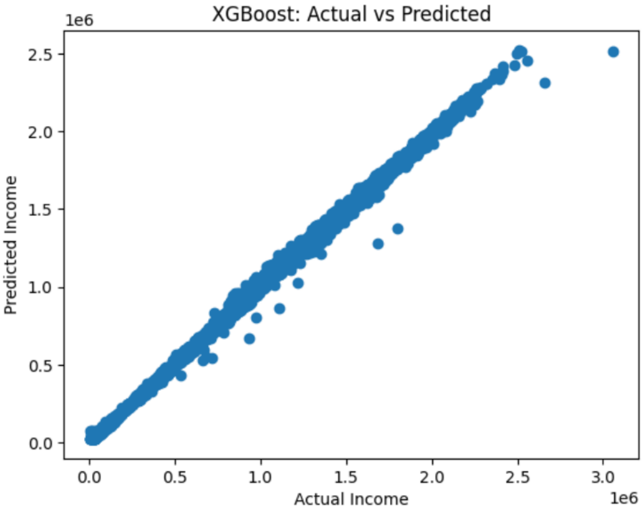

Figure (6.E)

These findings prompt a discussion on the importance of matching the model to the problem's characteristics. XGBoost's success is partially attributed to its ability to perform parallel computing and handle missing data, which are prevalent in financial datasets. Our discussion also considers the implications of these results for financial institutions, which could leverage such predictive modeling to enhance their income verification processes and fraud detection systems.

VII. CREATING WEBSITE

Our website, designed using Streamlit, reflects a commitment to simplicity and functionality. The front end presents a clean and navigable interface, where users can effortlessly input their demographic and financial details through form fields, sliders, and dropdown menus. The backend, powered by our machine learning model, performs real-time calculations to estimate income and evaluate the likelihood of fraud.

Technical details of the website include the use of Label Encoders to transform categorical data into a format suitable for our models and the deployment of a pre-trained Linear Regression model to establish baseline income predictions. The inclusion of randomness in income fields simulates a dynamic financial environment, providing users with a realistic assessment of income variance.

Upon submission, the application employs our "best model," loaded dynamically via joblib, to predict income based on user input. The fraud classification mechanism then compares this prediction against the user-reported income, factoring in a percentage-based threshold to determine the fraud status.

The website serves as a tangible product of our research, translating complex algorithms into a practical tool for end-users. By bridging the gap between theoretical research and real-world application, the website stands as a testament to the potential of machine learning in everyday financial decision-making.

CONCLUSION

Our research conclusively demonstrates the effectiveness of machine learning algorithms for income prediction and fraud identification. The XGBoost algorithm, in particular, has proven to be exceptionally accurate, achieving a 99.73% match in predictions versus actual income figures. This striking accuracy, highlighted by our scatter plot visualizations, emphasizes the capabilities of ensemble learning in economic data analysis.

Additionally, our findings show that machine learning algorithms adeptly navigate the complexities of financial data, with KNN and Random Forest models also yielding robust performances. In contrast, the lesser accuracy of the SVM model underscores the importance of matching algorithm strengths to data set features.

Moving from theory to application, we developed a user-centric web application utilizing Streamlit, which operationalizes our models' capabilities for real-world user scenarios. This platform allows users to input financial data and promptly receive income predictions and fraud risk evaluations. Notably, our fraud detection framework is designed with adjustable thresholds, allowing for tailored risk management according to different operational environments.

The incorporation of our research findings into a practical application reinforces the significant role machine learning can play in the financial industry, enhancing income verification processes and fraud detection mechanisms. The deployment of our user-friendly web application illustrates the practicality of machine learning, effectively transforming complex algorithms into accessible tools that empower financial decision-making.

Overall, our work not only advances the field of financial analytics through innovative predictive modeling but also emphasizes the importance of making such advancements available and understandable to end-users, thus fostering a culture of informed and technology-empowered financial practices.

REFERENCES

[1] Murorunkwere, Belle Fille, Origene Tuyishimire, Dominique Haughton, and Joseph Nzabanita. "Fraud detection using neural networks: A case study of income tax." Future Internet 14, no. 6 (2022): 168.

[2] Kavitha, Ms VR, Ms SR Krithika, Ms M. Tejaswini, and Ms N. Nihitha. "AN INCOME TAX FRAUD DETECTION USING AI."

[3] Xavier, Otávio Calaça, Sandrerley Ramos Pires, Thyago Carvalho Marques, and Anderson da Silva Soares. "Tax evasion identification using open data and artificial intelligence." Revista de Administração Pública 56 (2022): 426-440

[4] De Roux, Daniel, Boris Perez, Andrés Moreno, Maria del Pilar Villamil, and César Figueroa. "Tax fraud detection for under-reporting declarations using an unsupervised machine learning approach." In Proceedings of the 24th ACM SIGKDD International Conference on Knowledge Discovery & Data Mining, pp. 215-222. 2018.

[5] Pérez López, César, María Jesús Delgado Rodríguez, and Sonia de Lucas Santos. "Tax fraud detection through neural networks: An application using a sample of personal income taxpayers." Future Internet 11, no. 4 (2019): 86.

[6] Phill Ostwalt, Global Data & Analytics, Trusted Analytics article series, KPMG International, July 2016

[7] P. D. Meo, E. Ferrara, G. Fumaria, and A. Provetti, “Generalized Louvain method for community detection in large networks,” IEEE International Conference on Intelligent Systems Design and Applications, vol.79, p. 88-93, 2012

[8] Denys A. Flores, Olga Angelopoulos, Richard J. Self,” Design of a Monitor for Detecting Money Laundering and Terrorist Financing”, International Journal of Computer Networks and Applications 2014.

[9] Mena J(2003b) “ Investigative Data Mining for Security and Criminal Detection”, Butterworth Heinemann, MA, USA.

[10] He H, Wang J, Graco W and Hawkins S (1998) “Application of Neural Networks to Detection of Medical Fraud”, Expert Systems with Applications, 13, pp329-336

[11] Bethencourt, C., & Kunze, L., (2019, April). Tax evasion, social norms, and economic growth. Journal of Public Economic Theory, 21(2), 332-46. Retrieved from https://doi.org/10.1111/jpet.12346

[12] Herwartz, H.; Sardá, J.; Theilen, B. Money demand and the shadow economy: Empirical evidence from OECD countries. Empir. Econ. 2016, 50, 1627–1645.

[13] Castellón, P.; Velásquez, J.D. Characterization and detection of taxpayers with false invoices using data mining techniques. Expert Syst. Appl. 2012, 40, 1427–1436.

[14] Tosun, E.; Aydin, K.; Bilgili, M. Comparison of linear regression and artificial neural network model of a diesel engine fueled with biodiesel-alcohol mixtures. Alex. Eng. J. 2016, 55, 3081–3089.

[15] Castellón, P.; Velásquez, J.D. Characterization and detection of taxpayers with false invoices using data mining techniques. Expert Syst. Appl. 2012, 40, 1427–1436.

[16] Anyaeche, C.O.; Ighravwe, D.E. Predicting performance measures using linear regression and neural network: A comparison. Afr. J. Eng. Res. 2013, 1, 84–89.

[17] Garavaglia, S.; Sharma, A. A smart guide to dummy variables: Four applications and a macro. In Proceedings of the Northeast SAS Users Group Conference, Pittsburgh, PA, USA, 4–6 October 1998; Volume 43.

[18] Kull, M.; Silva, F.; Telmo, M.; Flach, P. Beyond Sigmoids: How to Obtain Well-Calibrated Probabilities from Binary Classifiers with Beta Calibration. Electron. J. Stat. 2017, 11, 5052–5080.

[19] Agostinelli, F.; Hoffman, M.; Sadowski, P.; Baldi, P. Learning Activation Functions to Improve Deep Neural Networks. arXiv 2014, arXiv:1412.6830.

[20] Sharma, S.; Sharma, S.; Athaiya, A. Activation Functions in Neural Networks; Towards Data Science, UK, 2017; Volume 12, pp. 310–316.