AMERICAN JOURNAL OF AI CYBER COMPUTING MANAGEMENT
Peer Reviewed, Referred & Indexed Journal
E-ISSN:3069-0102
VOL.6, NO. 2(1) (2026)

Intelligent Fraudulent Property Listing Classification using Ensemble Learning Techniques

P. Vyshali[1]*, Kandula Bhargavi[1], Suthoju Anusha[1], Kalvala Pallavi[1]
[1]Department of Computer Science and Engineering (DS), Kommuri Pratap Reddy Institute of Technology, Ghanpur, Ghatkesar, 501301, Telangana, India.
*Correspondence: P. Vyshali

ABSTRACT
The rapid digital transformation of the real estate sector has significantly increased the reliance on online property listing platforms, making the authenticity of listings a critical challenge. Fraudulent property advertisements not only lead to financial losses but also undermine user trust in these platforms. Traditionally, verification processes relied on manual moderation or simple rule-based techniques such as keyword filtering, price validation, and duplicate detection. While these methods offer basic screening, they are inefficient, time-consuming, and incapable of capturing complex fraud patterns in large-scale datasets. To overcome these limitations, this study proposes an intelligent hybrid ExtraNet framework for automated classification of real estate listings. The model integrates the Extra Trees Classifier (ETC) and customized Neural Networks (NN) to enhance detection accuracy. The system processes structured tabular data comprising essential property attributes, including price, location, area size, number of rooms, amenities, and textual descriptions. A robust preprocessing pipeline is employed to manage missing values, encode categorical variables, and normalize numerical features for improved model efficiency. For performance benchmarking, two baseline models Logistic Regression (LR) and K-Nearest Neighbors (KNN) are implemented and evaluated. Additionally, the system is deployed as an interactive desktop application using a Graphical User Interface (GUI) built with Tkinter and supported by a My Structured Query Language (MySQL) database for efficient data management. Experimental results demonstrate that the proposed hybrid ExtraNet model achieves a classification accuracy of 99.61%, significantly outperforming traditional models in identifying fraudulent listings and enhancing platform reliability.

Keywords: Real estate listings, fraud detection, digital transformation, property listing platforms, data preprocessing, tabular data analysis, feature normalization, categorical encoding, anomaly detection.

1. INTRODUCTION

Online real estate platforms provide important information for those planning to buy or rent real estate and affect customer behaviors and processes in property searches and purchases. As of 2023, the Turkish real estate industry has grown to USD 46 billion. With digitalization and online real estate platforms, real estate sales increase up to 15% annually. Fig. 1 demonstrates the role of artificial intelligence (AI) in the real estate industry. For instance, one-third of Turkey's real estate sale transactions were performed online. Properties are advertised on digital platforms a hundred times more often than they are advertised offline, and such advertisements encompass the entire property market. Visual elements in digital real estate portfolios, such as photos and virtual tours, are paramount for attracting customers. On these platforms, the photographs viewed by potential buyers or tenants often determine whether a property is worth buying, highlighting how important the photo quality is. On the other hand, the quality and classification of images are essential for attracting customers' attention and are critical in reducing editing time and increasing listing review time.

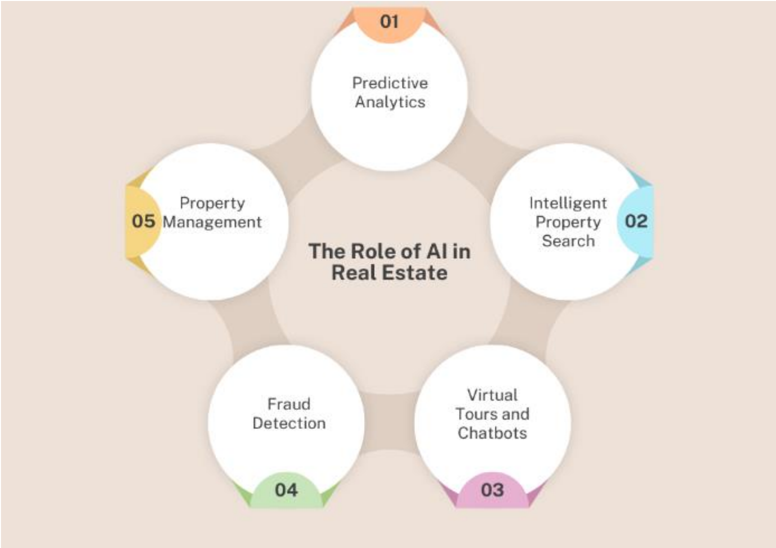

Fig. 1: AI in the real estate industry.

Well-organized and presented images enhance users' experience and make the search for properties efficient. The more accurately these images appear, the more effectively and efficiently users utilize them. Therefore, estate agencies need to optimize images in their listings.

2. LITERATURE SURVEY

Sözdinler et al. [1] proposed an extended platform that examined the role of machine learning models and image enhancement methods in real estate technology. The authors also introduced an alternative solution that could be integrated into intelligent listing systems to enhance user experience and information accuracy. Their study demonstrated that artificial intelligence and machine learning could be effectively combined for cloud-distributed services, thereby paving the way for future innovations in the real estate sector and intelligent marketplace platforms.

Cristian Bogdan et al., [2] aimed to identify how fake news appeared in a world where access to information was simple and individuals could easily verify the truthfulness of information. The study also examined the barriers used to protect businesses from being affected and customers from being misinformed. Data for the study were collected through in-depth interviews with entrepreneurs from real estate businesses. The findings revealed that the increasing level of fake news negatively affected the trust-building process. The authors recommended that organizations analyze user behavior more thoroughly and utilize it in the decision-making process to reduce the negative effects of fake news. They also suggested that customers learn to identify quality brand communication, pay attention to details, and exercise caution when making decisions based on potentially fake news.

Dadkhah S, et al. [3] proposed and evaluated four image–text similarity measures, namely textual similarity, semantic similarity, contextual similarity, and post-training similarity. The textual and semantic similarities represented the original image text relationships in terms of textual content and image caption information, while the contextual similarity reflected the image–text similarity based on meaningful named entities. Consumers with higher ecological motivation and environmental awareness are more prone to engage in green purchase behavior. A study on green food purchase found that pro-environmental self-identity predicts green purchase intention [4]. The intention of the consumer to buy eco-friendly products is to support environmental sustainability, thus impacting pro-environmental self-identity. So, pro-environmental purchasing is more inclined to green purchase decisions [5].

The key to self-identity is perceiving oneself as possessing specific traits: characteristics, qualities, and abilities in combination with personality traits, physical attributes interests, hobbies, memories, moral qualities, social factors, and cognitive processes that are selectively chosen to identify themselves [6].

Karpook et al. [7] Investment in AI and machine learning (ML) was having a transformative effect on corporate and commercial real estate, providing unprecedented efficiencies in tasks such as market analysis, property valuation, document automation, and lease abstraction. These activities, which had always been tedious and time-consuming manual processes, could be done with great precision and speed using the emerging technologies, freeing workers for tasks that relied on human interaction and judgment. These innovations, which were overwhelmingly data-driven, were particularly well-suited to real estate, with its heavy reliance on property and market data. The promise of AI and ML was driving significant investment in an industry that had long stood out for its resistance to new technologies, with predictions of investment exceeding US$700bn by 2028. While exciting, this new territory was not without danger; the emergence of 'deep fakes' that used the same AI and ML technologies to enable fraud and theft threatened businesses that were just beginning to understand how to protect themselves.
Joel Paul et al. [8] The successful execution of financial and real estate machine learning (ML) prediction models required perfect performance balance to produce accurate results and reliable interpretations. The applications of financial ML models spanned algorithmic trading combined with credit risk evaluation as well as fraud detection, and real estate ML models concentrated on property evaluation, market trend analysis, and rental price forecasting. The existing models encountered obstacles because they experienced overfitting and encountered problems in data quality and had to follow regulatory guidelines. The author examined multiple techniques which enhanced model efficiency and explained the applications of feature engineering together with regularization combined with robust validation methods along with hybrid modeling approaches.
Patel, Diepesh Manish et al. [9] aimed to explore whether signs of property market bubble formation had become evident, and to create a pragmatic predictive model of Irish real-estate valuation. Acknowledging the challenge of the market, this study represented a holistic approach aware of the many limitations of traditional investigation that created a one-sided distinction between the factors influencing property values.
F. A. Sulistio et al. [10] proposed a multimodal deep learning model that combined textual features from Indonesia Bidirectional Encoder Representations from Transformers (IndoBERT) with visual features extracted from a Residual Network-18 (ResNet18) based Convolutional Neural Network (CNN) to classify listings as real or fake. The dataset comprised 2,632 entries, consisting of 1,632 real listings and 1,000 synthetically generated fake ones, and yielded a classification accuracy of 97.99%. This result outperformed unimodal baselines that used only text (97.15%) or image (90.51%) inputs.
Duy Nguyen et al. [11] proposed FADAML, a novel end-to-end machine learning system to detect and filter out fake online advertisements. Their system combined techniques in multimodal machine learning and automated machine learning to achieve a high detection rate. As a case study, the authors applied FADAML to detect fake advertisements on popular Vietnamese real estate websites. Their experiments showed that the system achieved 91.5% detection accuracy, which significantly outperformed three different state-of-the-art fake news detection systems.
Yuan et al. compared different approaches using regression and classification models for property price prediction [12]. determined the most effective method by evaluating different machine learning techniques. The result proved the usability and the performance of machine learning models for the real estate industry in forecasting property prices. Lemeš and Akagic predicted the cost of proprieties and assessed the performances of the algorithms. The prediction accuracy was increased by centering on feature engineering and model optimization [13]. Law et al. built prediction models for housing prices using deep learning models, street and satellite images, and economic factors [14]. According to the previous results, exterior and top-view real estate images are used to effectively evaluate properties. Semnani and Rezaei studied real estate price prediction using satellite images and deep learning techniques [15].

3. PROPOSED SYSTEM

The proposed system, as demonstrated in Fig. 2, introduces a data-driven and automated approach to detect fraudulent property listings using ML techniques. Unlike traditional methods that rely on manual inspection, this system applies ML algorithms to learn complex patterns from property listing data. The core functionality of this research involves loading a dataset containing various listing attributes such as property price, location, area size, number of rooms, amenities, and other related features. The data is pre-processed, analyzed, and then used to train classification models capable of identifying fraudulent or misleading listings. The system integrates multiple algorithms, including LR, KNN, and the hybrid ExtraNet model, to improve prediction performance and robustness. The models are evaluated using key metrics such as accuracy, precision, recall, and F1-score to ensure reliability and effectiveness in detecting fake property listings.

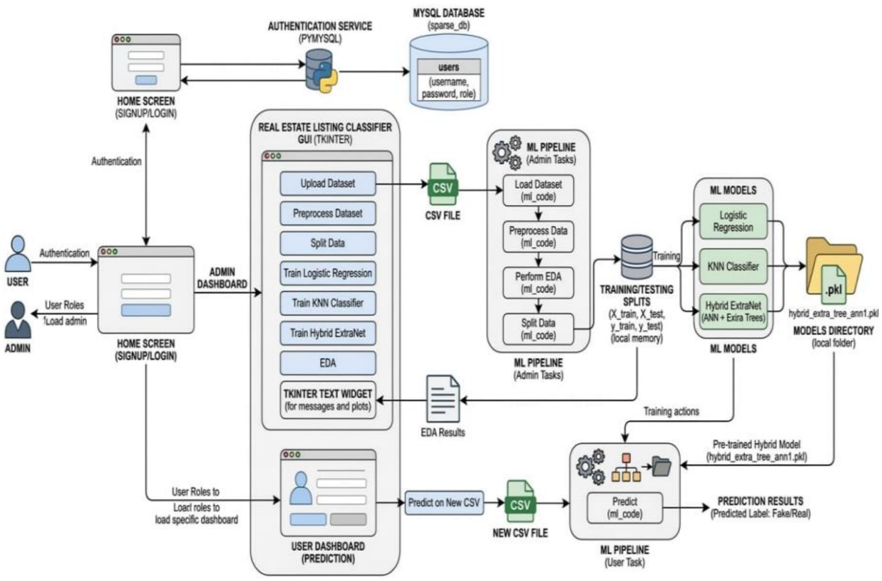

Fig. 2: Proposed System Architecture.

Dataset Collection: The first step in the proposed system involves collecting and loading a dataset that contains various real estate listing attributes. These attributes include information such as property price, location, area size, number of rooms, parking availability, and other related features. The dataset provides the foundational data required for training ML models. Proper dataset collection ensures that the system can learn meaningful patterns related to real and fake property listings.

Data Preprocessing: After loading the dataset, the data is pre-processed to improve its quality and suitability for model training. This step involves handling missing values, encoding categorical attributes, and normalizing numerical features. Data preprocessing helps remove inconsistencies and prepares the dataset for efficient analysis. Proper preprocessing also improves the performance and accuracy of ML models.

Exploratory Data Analysis (EDA): EDA is performed to understand the structure and distribution of the dataset. Various visualizations and statistical techniques are used to analyze relationships between different property features. This step helps identify patterns, correlations, and potential anomalies in the data. EDA provides valuable insights that guide the model development process.

Data Splitting: The processed dataset is divided into two parts: training data and testing data. The training dataset is used to train the ML models, while the testing dataset is used to evaluate their performance. This separation ensures that the models are tested on unseen data, which helps measure their generalization capability. Proper data splitting is essential for obtaining reliable evaluation results.

Model Training: In this step, different ML classification algorithms are applied to the training dataset. The models learn patterns from the property listing features that help distinguish between real and fake listings. The algorithms implemented in the system include LR, KNN, and the hybrid ExtraNet model. Training the models allows them to capture important relationships within the data.

Model Evaluation: After training the models, their performance is evaluated using the testing dataset. Several evaluation metrics such as accuracy, precision, recall, and F1-score are used to measure the effectiveness of each model. These metrics help determine how well the models can identify fraudulent property listings. Model evaluation ensures that the most reliable model is selected for prediction.

Prediction of New Listings: Once the models are trained and evaluated, the system allows users to provide new property listing data for verification. The trained model analyzes the input features and predicts whether the listing is real or fake. This prediction process helps users quickly identify suspicious listings. It provides an automated way to verify the authenticity of property advertisements.

User Interaction through GUI: The entire system is integrated into a GUI developed using Tkinter to provide an interactive interface. The GUI allows administrators to upload datasets, preprocess data, perform EDA, and train ML models. Users can upload new property listing data and obtain predictions about listing authenticity. This interface makes the system easy to use and accessible for both admins and general users.

4. RESULTS ANALYSIS

The implementation of the research produced a fully functional, interactive, and intelligent system for classifying real estate listings as authentic or fraudulent. The results are presented through a sequence of graphical user interface screens and performance evaluation metrics of the implemented models. Each figure corresponds to a major component or stage of the research workflow, from data input to model prediction.

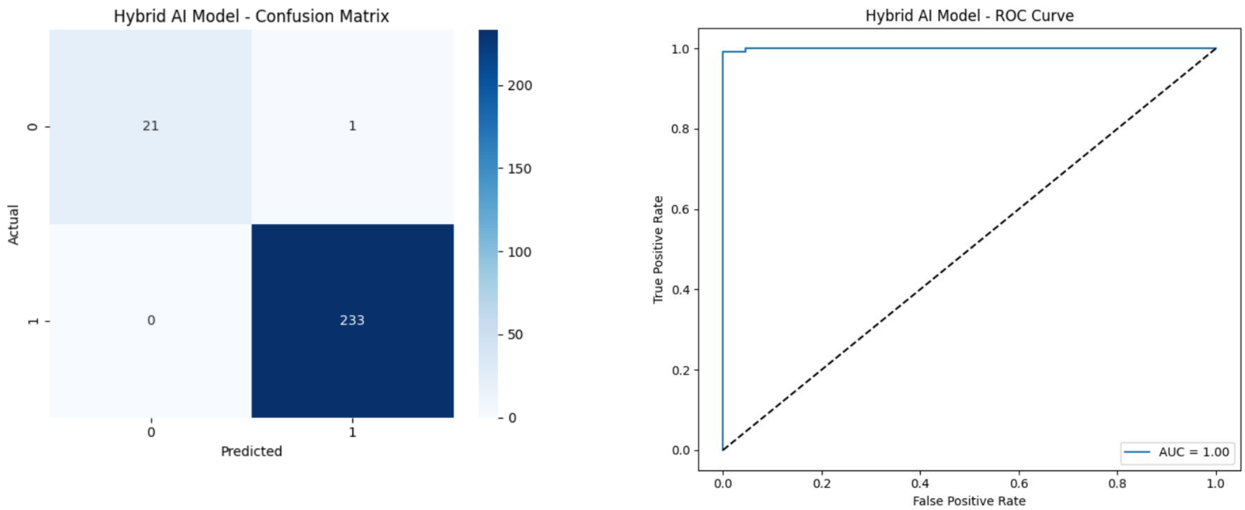

Fig. 3: Obtained confusion matrix and ROC curves from hybrid ExtraNet model.

Fig. 3 the confusion matrix highlights hybrid ExtraNet predictions, showing counts of real and fake listings classified correctly and incorrectly. True positives and true negatives confirm high accuracy, while false positives and false negatives indicate rare misclassifications. The ROC curve shows sensitivity versus specificity, and the AUC demonstrates the superior performance of hybrid ExtraNet, combining ETC feature selection and ANN non-linear pattern recognition for accurate classification.

Predictions:

	ID	Listing Confirmation Method	Deposit Amount	...	Listing Weekday	Days Since Listed	Predicted Label
0	TRAIN_0004	On-site Confirmation	346000000	...	Tuesday	382	Real
1	TRAIN_0012	Phone Confirmation	348500000	...	Monday	362	Real
2	TRAIN_0022	On-site Confirmation	24500000	...	Sunday	426	Fake
3	TRAIN_0031	On-site Confirmation	132000000	...	Wednesday	367	Fake
4	TRAIN_0047	On-site Confirmation	76000000	...	Tuesday	571	Fake
5	TRAIN_0048	On-site Confirmation	156000000	...	Friday	512	Fake
6	TRAIN_0093	Phone Confirmation	403500000	...	Wednesday	437	Real
7	TRAIN_0106	On-site Confirmation	12500000	...	Saturday	469	Fake
8	TRAIN_0117	On-site Confirmation	12000000	...	Sunday	419	Fake
9	TRAIN_0118	On-site Confirmation	11000000	...	Friday	449	Fake
10	TRAIN_0128	서류확인	69500000	...	Wednesday	360	Fake
11	TRAIN_0135	On-site Confirmation	406500000	...	Thursday	478	Real
12	TRAIN_0136	On-site Confirmation	135000000	...	Friday	449	Fake
13	TRAIN_0147	On-site Confirmation	154500000	...	Friday	337	Fake
14	TRAIN_0156	On-site Confirmation	8000000	...	Friday	736	Fake
15	TRAIN_0160	Phone Confirmation	11000000	...	Sunday	293	Fake
16	TRAIN_0164	On-site Confirmation	26000000	...	Saturday	406	Fake
17	TRAIN_0000	On-site Confirmation	402500000	...	Wednesday	276	Real

Fig. 4: Sample predictions on test data.

Fig. 4 showcases the final output of the system. It displays the dataset entries along with their predicted labels Real or Fake based on the model's analysis. The results are presented in a tabular format for easy interpretation, with additional options for saving or exporting the prediction report. This screen reflects the system's ability to process unseen data and deliver reliable classification outcomes instantly. It validates the overall research objective ensuring authenticity and transparency in online real estate listings through an intelligent AI-driven solution.

4.1 Comparative Analysis

Table 1: Performance comparison for the LR, KNN, and hybrid ExtraNet models.

Algorithms Name	**Accuracy**	**Precision**	**Recall**	**F1-Score**
LR Model	8.63	4.31	50.00	7.94
KNN Model	91.37	45.69	50.00	47.75
hybrid ExtraNet	99.61	99.79	97.73	98.73

Table 1 presents a detailed performance comparison of the three models implemented in the research LR, KNN, and hybrid ExtraNet. LR showed limited performance, achieving only 8.63% accuracy with low precision and F1-score, indicating its inability to capture complex, non-linear patterns in the real estate data. KNN performed significantly better, achieving 91.37% accuracy, demonstrating that neighborhood-based learning effectively identified similarities among authentic listings. However, moderate precision and recall reflect its limitations in consistently classifying boundary cases. The hybrid ExtraNet model outperformed both, achieving 99.61% accuracy, 99.79% precision, and 98.73% F1-score. The combination of ETC feature selection and ANN learning enabled the model to generalize effectively, distinguishing real listings from fraudulent ones with near-perfect reliability.

Table 2: Class-wise Performance of LR Model

Category	**Precision (%)**	**Recall (%)**	**F1-Score (%)**	**Support**
Fraudulent Listing	0.00	0.00	0.00	22
Genuine Listing	91.00	100.00	95.00	233
Accuracy			91.37%	255
Macro Average	45.69	50.00	47.75	255

Weighted Average	83.00	91.00	87.00	255

Table 2 presents the class-wise performance of the LR model for detecting fraudulent and genuine real estate listings. The results show that the model achieved an overall accuracy of 91.37% on the testing dataset. For genuine listings, the model performed very well with 91.00% precision, 100.00% recall, and an F1-score of 95.00%, correctly identifying most of the legitimate property entries. However, the model failed to correctly detect fraudulent listings, resulting in 0.00% precision, recall, and F1-score for that category. This indicates that the model predicted almost all samples as genuine due to the class imbalance in the dataset, where genuine listings significantly outnumber fraudulent ones. The macro average values (Precision: 45.69%, Recall: 50.00%, F1-score: 47.75%) further highlight the model's limited capability in handling minority class predictions. These results suggest that while LR provides a strong baseline accuracy, it is not sufficiently effective for identifying fraudulent listings in imbalanced real estate datasets.

Table 3: Class-wise Performance of KNN Model

Category	Precision (%)	Recall (%)	F1-Score (%)	Support
Fraudulent Listing	100.00	5.00	9.00	22
Genuine Listing	92.00	100.00	96.00	233
Accuracy			91.76%	255
Macro Average	96.00	52.00	52.00	255
Weighted Average	92.00	92.00	88.00	255

Table 3 presents the class-wise performance of the KNN model for detecting fraudulent and genuine real estate listings. The model achieved an overall accuracy of 91.76% on the testing dataset. For genuine listings, the model performed strongly with 92.00% precision, 100.00% recall, and an F1-score of 96.00%, indicating that most legitimate listings were correctly identified. For fraudulent listings, the model obtained 100.00% precision but only 5.00% recall, resulting in a low F1-score of 9.00%, which means that only a small portion of fraudulent listings were detected. This performance is mainly due to the class imbalance in the dataset, where genuine listings significantly outnumber fraudulent ones. The macro average values (Precision: 96.00%, Recall: 52.00%, F1-score: 52.00%) further indicate that the model struggles to effectively capture minority class instances.

Table 4: Class-wise Performance of hybrid ExtraNet Model

Category	Precision (%)	Recall (%)	F1-Score (%)	Support
Fraudulent Listing	100.00	95.00	98.00	22
Genuine Listing	100.00	100.00	100.00	233
Accuracy			99.61%	255
Macro Average	100.00	98.00	99.00	255
Weighted Average	100.00	100.00	100.00	255

Table 4 presents the class-wise performance of the proposed hybrid ExtraNet model for detecting fraudulent and genuine real estate listings. The model achieved a very high accuracy of 99.61%, demonstrating its strong ability to classify property listings correctly. For fraudulent listings, the model achieved 100.00% precision, 95.00% recall, and an F1-score of 98.00%, indicating that most fraudulent listings were successfully detected. For genuine listings, the model achieved 100.00% precision, recall, and F1-score, showing perfect classification of legitimate property listings. The macro average values (Precision: 100.00%, Recall: 98.00%, F1-score: 99.00%) further confirm the balanced performance across both classes. These results demonstrate that the hybrid ExtraNet model significantly outperforms the LR and KNN models, providing a highly accurate and reliable solution for detecting fraudulent real estate listings.

5. CONCLUSION

The proposed real estate fraud detection framework represents a significant advancement in identifying fraudulent property listings by effectively combining structured property data with intelligent machine learning models. The system addresses the growing challenge of fake listings on online property platforms by introducing a hybrid ExtraNet architecture that integrates ETC and ANN. This hybrid approach enables the model to capture both important feature interactions and complex non-linear patterns within the dataset, leading to highly reliable prediction performance. A major challenge addressed in this research was the presence of noisy and imbalanced data commonly found in real estate listings. The implemented preprocessing pipeline, which includes handling missing values, encoding categorical attributes, normalization of numerical features, and outlier removal, significantly improved the quality of the input data. In addition, EDA provided valuable insights into relationships among features such as rent, room distribution, parking availability, and property size, which contribute strongly to accurate model predictions. The experimental evaluation demonstrated that the proposed hybrid ExtraNet model consistently outperformed baseline models such as LR and KNN. Achieving a classification accuracy of 99.61%, the model effectively distinguishes genuine listings from fraudulent ones while maintaining strong precision, recall, and F1-score values. Ultimately, the developed framework provides a scalable and reliable solution for real estate fraud detection. Through its integration with a Tkinter-based GUI and MySQL database support, the system enables efficient dataset management, model training, and real-time prediction, offering a practical tool to enhance trust, transparency, and security in modern digital real estate platforms.

REFERENCES

[1] Küp, E.T.; Sözdinler, M.; Işık, A.H.; Doksanbir, Y.; Akpınar, G. Enhancing Real Estate Listings Through Image Classification and Enhancement: A Comparative Study. Eng. Proc. 2025, 92, 80. https://doi.org/10.3390/engproc2025092080

[2] Cristian Bogdan ONETE The Department of Business, Consumer Sciences and Quality Management, Faculty of Business and Tourism, Bucharest University of Economic Studies, Bucharest, Romania cristian.onete@com.ase.ro

[3] Zhang X, Dadkhah S, Weismann AG, Kanaani MA, Ghorbani AA. 2024. Multimodal fake news analysis based on image-text similarity. IEEE Transactions on Computational Social Systems 11(1):959-972

[4] Bajar, R.G.; Ong, A.K.; German, J.D. Determining sustainable purchase behavior for green products from name-brand shops: A Gen Z perspective in a developing country. Sustainability 2024, 16, 3747.

[5] Lopes, J.M.; Gomes, S.; Trancoso, T. Navigating the green maze: Insights for businesses on consumer decision-making and the mediating role of their environmental concerns. Sustain. Account. Manag. Policy J. 2024, 15, 861–883.

[6] Rudnev, M.; Barrett, H.C.; Buckwalter, W.; Machery, E.; Stich, S.; Barr, K.; Bencherifa, A.; Clancy, R.F.; Crone, D.L.; Deguchi, Y.; et al. Dimensions of wisdom perception across twelve countries on five continents. Nat. Commun. 2024, 15, 6375.

[7] Karpook, Corporate Real Estate Journal, Volume 14 / Number 3 / Spring 2025, pp. 210-221(12), Henry Stewart Publications DOI: https://doi.org/10.69554/TQFL5101

[8] Joel Paul Balancing Performance in Financial and Real Estate Machine Learning Models, February 2025.

[9] Mahtabi, M., Roshan, M., Muhit, M. M. I., Behvar, A., & Haghshenas, M. (2026). Cryogenic ultrasonic fatigue: Mechanisms, advancements, and insights. Cryogenics, 153, 104257. https://doi.org/10.1016/j.cryogenics.2025.104257

[10] Patel, Diepesh Manish (2024) Stacking Ensemble Approach towards Predicting Irish-Real Estate. Masters thesis, Dublin, National College of Ireland.

[11] F. A. Sulistio, S. Suakanto, N. Ambarsari and S. C. R. Rijadi, "Detecting Fake Boarding House Listings Using Multimodal Deep Learning," 2025 IEEE International Conference on Industry 4.0, Artificial Intelligence, and Communications Technology (IAICT), Bali, Indonesia, 2025, pp. 330-337, doi: 10.1109/IAICT65714.2025.11101391.

[12] Duy Nguyen, Fake Advertisements Detection Using Automated Multimodal Learning: A Case Study for Vietnamese Real Estate Data, Datal, Ho Chi Minh City, Vietnam duyndv@datal.com.vn.

[13] Mudusu, S. (2025). Health Insurance Fraud Detection: The Role Of Advanced It Systems In Preventing And Identifying Fraud. International Journal, 16(1), 3769-3777

[14] Yuan, F.; Wu, J.; Wei, Y.D.; Wang, L. Policy change, amenity, and spatiotemporal dynamics of housing prices in Nanjing, China. Land Use Policy, 75, 225–236.

[15] Lemeš, L.; Akagic, A. Prediction of Real Estate Market Prices with Regression Algorithms; Lecture Notes in Networks and Systems; Springer: Cham, Switherland, 2023; Volume 539, pp. 401–411.

[16] Law, S.; Paige, B.; Russell, C. Take a look around: Using street view and satellite images to estimate house prices. ACM Trans. Intell. Syst. Technol. 2019, 10, 54.

[17] Semnani, S.J.; Rezaei, H. House Price Prediction using Satellite Imagery. arXiv 2021, arXiv:2105.06060.